

# PCI Express: Ever-faster graphics pipe serves many masters

THE NEW PCI EXPRESS SPEC SIGNIFICANTLY IMPROVES DESKTOP-PC GRAPHICS. DEVELOPERS ARE NOW WORKING ON GENERATION 2, WHICH WILL FURTHER EXPAND THE GRAPHICS PIPE.

The new PCI (Peripheral Component Interconnect) Express spec provides the biggest improvement in more than a decade in I/O performance for computation systems, significantly improving graphics in desktop PCs and workstations. Intel initially launched the spec in its chip sets in mid-2004, and the technology has become mainstream in high-end systems. But PCI Express is far more than an avenue to better games or video. As have many other PC innovations, PCI Express will enable significant applications, such as medical imaging, and serve in industrial control and many other embedded-system roles.

Serial PCI Express is usurping the parallel AGP (accelerated graphics port) in graphics as just one part of a broad industry trend toward serialization. For example, USB replaced the parallel and serial ports, and SATA and SAS (serial attached SCSI) are replacing parallel ATA and SCSI in storage interfaces. And, like parallel PCI, PCI Express will handle far more than graphics, providing a flexible and scalable data highway for all types of performance-centric add-on functions.

You might ask whether a replacement for PCI-derived AGP is necessary. AGP8× pushed the evolutionary limits of what was possible with a parallel bus—at least at a price point sustainable for volume PC production. Clock skew was the problem plaguing AGP, like other parallel buses touting higher performance. A single clock defines the data-valid period across AGP’s 32 data lanes. But the faster the clock runs, the narrower the data-valid window becomes, and the tougher the design challenge becomes (see sidebar “History of the graphics pipe”).

So, the graphics pipeline became serialized with PCI Express. This serialization means that data carries with it an embedded clock, which a PLL recovers in the receiver circuits. Multiple lanes can carry data in parallel, providing scalable performance, but each lane has its own embedded clock (Figure 1). At the receiver end, the data resynchronizes. PCI Express can tolerate as much as four symbols of skew between lanes, dramatically easing the constraints for motherboard and graphics-controller design.

PCI Express is more than just the follow-on to AGP. PCI

Express is also the migration point for applications currently on conventional PCI and PCI-X as they move to higher speeds and performance. These legacy buses will not go away overnight, but, like ISA (Industry Standard Bus), they eventually will. PCI Express reunites the forks that broke away from the original PCI with a common protocol and electrical interface. Moreover, PCI Express technology will make its way to the embedded-system world in CompactPCI implementations.

For graphics implementations, the PCI Express bus comprises 16 lane pairs. Each lane pair has four data wires—a transmitting differential pair and a receiving differential pair. Because PCI Express is dual-simplex, data can flow in both directions unimpeded by data going the other direction, unlike AGP (Figure 2). AGP was point-to-point like PCI Express, but AGP

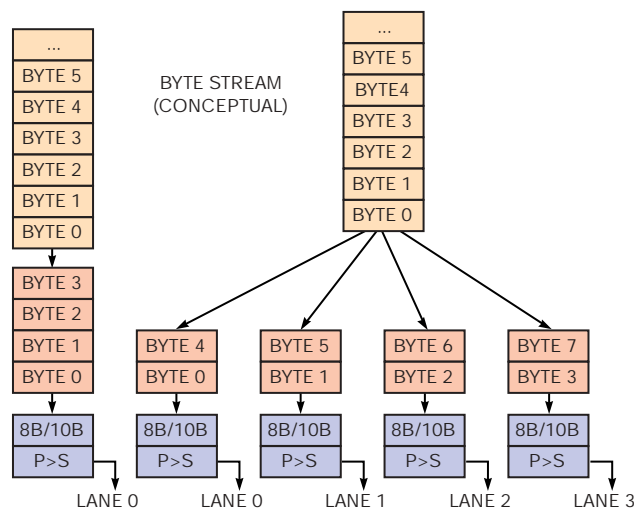


Figure 1 The PCI Express offers scalability through the concept of lanes. Designers can linearly extend bandwidth by adding lanes. In a one-lane system, such as Lane 0, the data bytes flow sequentially as expected. In a multilane scenario, the bytes are interleaved across the lanes for transmission.

is full-duplex: Data can flow in only one direction at a time. Furthermore, AGP never offered symmetric bandwidth in its implementations, even though nothing about the AGP protocols precludes symmetry. AGP8× achieves 2.1-Gbyte/sec peaks in data flows from the host CPU to graphics controllers, which is the primary direction for graphics traffic. However, in typical system implementations, the real available “back-channel” bandwidth is about one-tenth of that figure. In contrast, PCI Express graphics offers 4-Gbyte/sec peak bandwidth simultaneously in both directions and, in typical implementations, delivers the bulk of that bandwidth in both directions. This highly symmetric bandwidth leads to some interesting new capabilities for graphics based on PCI Express and in many other I/O applications.

AGP implementations yielded weak back-channel bandwidth for several reasons. To achieve the best CPU-to-graphics controller performance, AGP worked from uncached address spaces (the AGP “aperture”). For reads, this assumption makes perfect sense, because caching requires snooping, degrading performance. However, chip-set designers simply did not optimize their chip sets for large data-set writes to this space from the graphics controller because the cost of providing a large number of write-posting buffers was prohibitive. This trade-off made sense

IN THE DEFINITION OF PCI EXPRESS, THE GRAPHICS INDUSTRY WAS UNANIMOUS IN THE OPINION THAT IT DIDN'T WANT SIMILAR “HELP” IN THE NEW I/O INTERFACE.

to the chip-set architects, because graphics traffic is primarily in the forward direction.

A second reason for the relative weakness of AGP’s back-channel bandwidth was a limitation in the GART (graphics-address-remapping-table) memory-management system that AGP provided to assist in the graphics controller’s task of managing physical- and virtual-address translations for access to uncachable system-memory space. Again, the theory sounded great, but practical design considerations led to suboptimal graphics performance, because real chip sets never implemented enough TLBs (translation-look-aside buffers). Each 4 kbytes of memory requires a new TLB because 4 kbytes is the default page size in Windows. But even two dozen TLBs support only about 100 kbytes of memory before the onset of TLB “thrash.”

## HISTORY OF THE GRAPHICS PIPE

The demands of graphics were the primary driving force behind the PCI (Peripheral Component Interconnect) bus. The developers of the original ISA (Industry Standard Bus) based it on the PC AT’s Intel 286 processor bus, which debuted in 1984. ISA delivered only 16 Mbytes/sec at 8 MHz, and was woefully inadequate for the fire hose of data that 3-D graphics requires. The VESA (Video Electronics Standards Association) developed the VL (Video Local) Bus as a proposed alternative to ISA. EISA (Extended ISA) was another. The VL Bus designers were clever in getting more than 100-Mbyte/sec bandwidth from the CPU to the graphics controller. However, the bus worked only on graphics because it did not support multiple devices.

Furthermore, it tied only to a specific processor’s bus and, thus, could not adapt to future technologies. These facts explain why PCI, with its 133-Mbyte/sec bandwidth, triumphed over a 32-bit, shared, “multidrop” bus. Intel in 1991 proposed PCI as a scalable replacement for ISA and helped form the PCI-SIG (special-interest group), which in 1993 released the first specification.

You’d think the graphics industry would happily chew on this order-of-magnitude increase in bandwidth for a while, but that scenario didn’t occur. During that era, Microsoft (www.microsoft.com) moved from DOS to Windows, causing a discontinuity in demand for 3-D-graphics capabilities in PCs. Intel in 1997—only three years after PCI

entered production—introduced its Pentium II processor with the AGP (accelerated-graphics port). AGP borrows heavily from PCI technology. AGP is also 32 bits wide, and its protocols build on top of PCI protocols. But AGP is a nonshared, point-to-point bus. It uses a different connector from the one that PCI uses. And, although PCs would commonly have four to six PCI slots, they would have only one AGP slot.

AGP went through many speed bumps to achieve 2.1 Gbytes/sec in its AGP8× version, which its developers released in 2002, a remarkable evolution from its roots in PCI. You’d think the graphics community would be happy now, having fattened the pipe nearly three orders of magnitude in only two decades.

However, graphics vendors believed they’d be taxing this bandwidth within a few short years.

PCI Express began life as 3GIO (third-generation I/O) in the Arapahoe Work Group but moved quickly to become a specification that the PCI-SIG owned. That group in July 2002 released the 1.0 Version. First desktop, workstation, and server products from Intel went into production mid-2004, and the first chip sets for notebooks debuted early in 2005. Now, a range of PCI Express products are available in many categories of devices (Reference A).

### REFERENCE

■ Intel Developer Network for PCI Express Architecture, www.pciexpressdevnet.org/kshowcase.

A cache thrashes when its miss rate is too high, and it spends most of its time servicing misses. Thrashing, particularly virtual-memory thrashing, is bad for performance, because the relative cost of a miss is so high: It may slow a machine down by a factor 100 or more.

In the definition of PCI Express, the graphics industry was unanimous in the opinion that it didn't want similar "help" in the new I/O interface. With PCI Express graphics, all memory management occurs within the graphics controller. Memory-management performance is under the control of the graphics vendors, who are more economically motivated in general than are chip-set vendors to spend gates on graphics performance.

At this point, it is reasonable to ask why continuously increasing the bandwidth of the pipe from the host CPU to the graphics controller is so important. More prosaically, an end user may reasonably want to know what he can do with a notebook, desktop PC, or workstation with PCI Express graphics that is not possible with AGP8× or, for that matter, what PCI Express offers in embedded-system roles. Only a few of today's applications are starting to top the limits of what AGP8× can deliver. It is not difficult to create a demo application that uses the full bandwidth of PCI Express, and the suppliers of graphics controllers use this approach to show off their latest products. But broadly available commercial applications rarely show any advantage just from the fatter pipe when the pipe enters.

The PC world still needs PCI Express graphics, and end users have good reasons for desiring them. To understand this concept, consider the perspective of a developer of graphics-intensive software, such as that for video games or CAD. These developers write their programs to the capabilities of the least-common-denominator hardware in their target customer base. For leading-edge video games and high-end-workstation applications, these hardware units are highly capable, recently introduced systems, but more generally the target is likely to be systems the vendors released over the last two to four years.

The only way to get the software community to continuously raise the bar in graphics capabilities is by continuously increasing the capabilities, including bandwidth, of the client platforms. The impact of AGP8× on software development is happening now. The work to bring PCI Express to market will have its major impact on software applications in the future.

But the end user still has plenty of incentive to purchase PCI Express graphics today. A buyer of PCI Express graphics is well-prepared for the getting the most from the new applications that emerge over the life of that PC. And, second, the suppliers of

---

**THE ONLY WAY TO GET THE SOFTWARE COMMUNITY TO CONTINUOUSLY RAISE THE BAR IN GRAPHICS CAPABILITIES IS BY CONTINUOUSLY INCREASING THE CAPABILITIES, INCLUDING BANDWIDTH, OF THE CLIENT PLATFORMS.**

---

graphics controllers are focusing their latest developments on PCI Express products. So, the client with PCI Express graphics will likely significantly outperform previous generations of products, even if PCI Express cannot, in general, now take credit for that performance advantage.

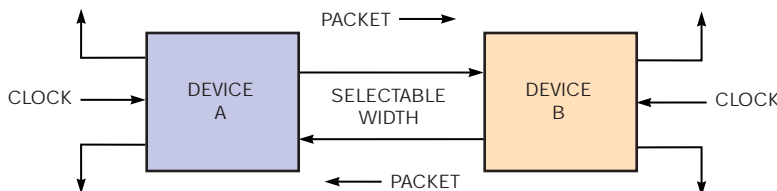
The performance advantages of PCI Express graphics may emerge more quickly than with previous transitions because of the tremendous increase of CPU performance and memory bandwidth just coming to market. For example, the latest Intel workstation platform exploits as much as 64 Gbytes of fully buffered DIMM with a peak bandwidth of 21 Gbytes/sec. Conjoin those features with additional capabilities in the most recent graphics controllers, and you can more fully exploit the bus.

PCI Express affords system designers an extremely broad range of capabilities. In the graphics world, PCI Express dual-graphics controllers are popular with gamers. Some designers are even contemplating externally cabled PCI Express subsystems (see sidebar "Express outside the box" and the online sidebar "Dual controllers accelerate rendering" at [www.edn.com/ms4201](http://www.edn.com/ms4201)).

## REAL-WORLD ADVANTAGES

PCI Express is a key enabler in some lifesaving applications. For example, consider Vital Images ([www.vitalimages.com](http://www.vitalimages.com)), a leading provider of enterprisewide advanced visualization and analysis software for use in disease-screening applications, clinical diagnosis, and therapy planning. The company's technology gives radiologists, cardiologists, oncologists, and other medical specialists timesaving productivity and communications tools for easy use in the day-to-day practice of medicine. Vital Images' software products include a medical-diagnostic tool that allows physicians to use PCs or notebook computers to gain remote access to 2-, 3-, and 4-D advanced visualization. The software enables users to measure, rotate, analyze, and segment images.

One technical challenge for this medical application stems from the size of data sets that volumetric visualization requires. According to Karel Zuiderveld, PhD, director of technology research at Vital Images, "PCI Express is especially beneficial when dealing with large data sets that do not fit into graphics memory. The size of modern medical data sets, [which the company obtained using computer-tomography], range from hundreds of megabytes to several gigabytes. In addition to a vast amount of CPU and GPU [graphics-processing-unit] resources, fast rendering of such data sets also



**Figure 2** PCI Express implements a full-duplex link in which traffic flows symmetrically and simultaneously in each direction over serial links with the clock embedded in the data.

requires high transfer speeds to the GPU.” With PCI Express graphics, medical professionals can view 3-D images from alternative perspectives with reasonable response times.

In the past, the graphics bus has not been the bottleneck. When uploading textures to OpenGL, the driver usually “swizzles” the texture—that is, swaps the pixels around so that they are stored in the same way as the original format. It then creates a copy in system memory, according to the OpenGL spec, and writes a copy to the graphics memory. Until recently, the CPU usually performed texture swizzling, resulting in low texture-upload speeds. Each texture requires a read-swizzle-write memory DMA to the graphics card; this approach involves at least two reads and one write to system memory. With the latest workstations, Vital Images’ target applications may fully exploit the bandwidth of PCI Express, says Zuiderveld. Though medical imaging may be a leading application for taxing the PCI Express bus, he believes that the trend toward using resources such as the virtual texture maps that Microsoft’s (www.microsoft.com) DX10 supports, will drive steeply increasing usage of PCI Express’s graphics bandwidth into mainstream applications.

### EXPRESS VIDEO EDITING

One application that can now take advantage of PCI Express’ unique capabilities is video editing. In video applications, PCI Express affords dramatically better back-channel bandwidth than AGP. Back-channel rates are important, because main memory must store intermediate and temporary video-processing results. The files are just too large for the graphics controller’s local store. However, the memory must preserve this data without lossy compression, because the cumulative effects of repeated compressions would visibly degrade the end result. With AGP, these writes of uncompressed data back to main system memory are major performance bottlenecks that PCI Express relieves. Watch for video editors that take advantage of PCI Express to enter the market.

PCI Express is backward-compatible with PCI protocols but

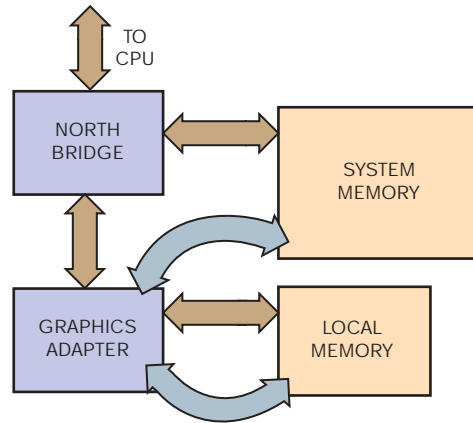


Figure 3 Nvidia and ATI with their TurboCache and HyperMemory technologies, respectively, use the PCI Express bus back channel to effectively cache their local memory (courtesy ATI).

offers numerous features that go beyond the PCI protocols. One feature—*isochrony*, or equality in length of time—promises to further aid video editing and other heavily multithreaded applications. PCI and AGP provide no guarantee for worst-case latency. Particularly in commercial-scale video editing for broadcast and film, this lack of guarantee for data delivery creates difficult challenges when trying to maintain output at a given frame rate. Isochrony could also ensure that systems running many multithreaded concurrent applications don’t drop display frames. As chip-set and device-hardware vendors and operating-system upgrades add isochrony support, PCI Express will provide this guarantee that AGP could not.

Some graphics vendors have already figured out how to exploit PCI Express’ back-channel bandwidth to create a new class of products. Before PCI Express, graphics memory was either in system memory—for chip sets with integrated graphics—or on an external card with an external graphics controller. Though AGP’s developers intended it to support heavy use of system

## EXPRESS OUTSIDE THE BOX

An ongoing development in the PCI-SIG (Peripheral Component Interconnect Special Interest Group) that potentially relates to PCI Express graphics is the definition of a PCI Express cable. PCI Express intends to provide an I/O-attachment point for a host, not to be that I/O itself. The developers of PCI Express had no intention of competing with cabled-I/O interfaces, such as USB, FireWire, Ethernet, InfiniBand, and others.

However, a cable allows the extension of the I/O-attachment point to be remote from a host system. A cable adapter card could plug into a 16-card slot and cable externally to a remote 16-card slot. Depending on distances and adherence to all the PCI Express design rules, this card might get away with containing no active circuitry, or it might contain a PCI Express switch or bridge that acts as a signal repeater.

One intriguing possibility would be a cable adapter card connecting, through a PCI Express cable, to a remote box containing a switch that supports two 16-card slots. Using a 48-lane switch, you could fully provision both of these slots and deliver a spec-compliant option. Such a remote dual-graphics box might afford some additional advantages. For example, by removing the graphics controllers from the base system, it also

removes their power and thermal requirements from burdening the base system. Also, any client with just one 16-card slot becomes “dual-graphics capable.” Such a remote dual-graphics box might prove attractive to OEMs that do not perceive dual graphics to be a mainstream requirement but want to provide their customers the capability and an upgrade path. It also might make an attractive after-market product.

memory, especially for textures, actual AGP implementations failed to deliver sufficient performance. The performance gap was too great between implementations with local frame memory on the graphics card and implementations using the AGP bus to access system memory. So, either you integrated graphics in a chip set, or you put a lot of graphics RAM on the external controller. A significant price and performance difference exists between these two approaches. Nvidia ([www.nvidia.com](http://www.nvidia.com)) and ATI ([www.ati.com](http://www.ati.com)) with their TurboCache and HyperMemory technologies, respectively, use the PCI Express bus back channel to effectively cache their local memory (Figure 3).

This method provides lower performance than that of a large local memory store on the graphics card, although the performance decrease does not approach the degradation that would occur on AGP. Still, these caching technologies allow the removal of significant amounts of RAM from the graphics-controller card. Instead of, say, eight 8-Mbit $\times$ 16-word DDR DRAMs for a traditional, state-of-the-art graphics controller, the controller card using caching over PCI Express could use just a single 4-Mbit $\times$ 32-word DDR DRAM. Memory costs would drop from approximately \$16 in today's prices to \$3.50, and performance would still be better than that of integrated graphics.

The future of PCI Express graphics does not end here. The PCI-SIG (special-interest group) has announced work on a

**MORE AT EDN.COM** 

 Go to [www.edn.com/ms4201](http://www.edn.com/ms4201) and click on Feedback Loop to post a comment on this article.

Generation 2 version of PCI Express. Though the specification was under development at press time, the PCI-SIG has announced key aspects of the new version. It will double the clock rate to 5 GHz. The group doesn't plan significant protocol enhancements over PCI Express 1.1, and Generation 2 will be backward-compatible with PCI Express 1.1. The PCI-SIG suggests that Generation 2 could be in production in 2007. When it does arrive, Generation

2 PCI Express will continue the grand tradition of regularly expanding the graphics pipe between host and graphics controller that you have witnessed from ISA to PCI to AGP and now to PCI Express. **EDN**

---

#### AUTHOR'S BIOGRAPHY

*David L Fair is enterprise-I/O-technology-initiatives manager at Intel Corp's Server Platforms Group Marketing Division (Santa Clara, CA). He is responsible for driving initiatives such as PCI Express for Intel's server and workstation businesses and managing the independent-hardware-vendor-enabling team. He has a bachelor's degree in physics from Pomona College (Claremont, CA) and a doctorate in the philosophy of science from Princeton University (Princeton, NJ). His personal interests include making broad technology adoption successful, road biking, "lunatic" high-end audio, quantum discontinuities in the ether, and trees that fall in the forest that no one hears.*

## DUAL CONTROLLERS ACCELERATE RENDERING

Among the first commercial successes for PCI (Peripheral Component Interconnect) Express, multiple graphics controllers in the same client can boost 3-D performance in demanding gaming and technical applications. PCI graphics could use multiple controllers, but AGP lacked this ability. Although it made economic sense for chip-set and system vendors to support multiple conventional PCI slots, no one ever built a system with more than one AGP slot, because there was too little market demand.

Two classes of applications might enjoy the use of more than one graphics controller. The first of these applications are those that require multiple monitors. Today's mainstream graphics controllers often support two independent monitors, and some applications require more than two. Financial trading is a key example; traders need to track activities in real time simultaneously on as many as eight monitors. It is easy to imagine many other examples, such as process monitoring, in which two screens are inadequate. At least in theory, PCI Express systems could provide support for

applications requiring more than two or three monitors.

An end user might also desire multiple graphics controllers to use them in parallel to boost the graphics performance on a single monitor beyond the current state of the art. The first company to take this approach, the now-defunct 3Dfx, offered PCI Voodoo2 cards, which users could gang to perform SCI (scan-line interleaving) on conventional, 32-bit PCI (Figure A). A private bus connected the two Voodoo2s, so the rendering data could exit a single connector. But each would compute alternate scan lines, theoretically permitting as much as two-times-better performance for applications such as action video games, which have rendering limitations.

Nvidia has revived many of these concepts for PCI Express. The company even preserved 3Dfx's acronym, SLI, which Nvidia owns though an acquisition of 3Dfx's intellectual property. Nvidia defines SLI as "scalable-link interface," which is more sophisticated than simple scan-line interleaving and is performing this task in the digital domain. Like the 3Dfx approach,

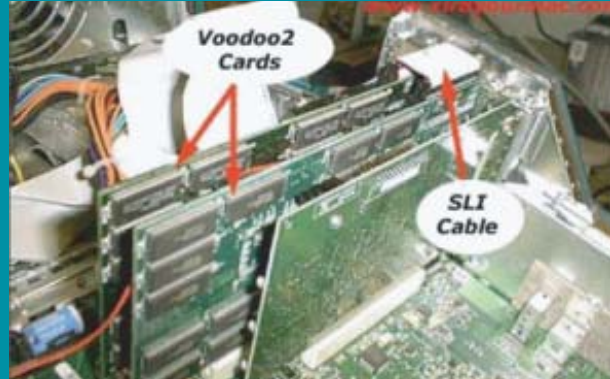


Figure A The now-defunct 3Dfx offered PCI Voodoo2 cards, which users could gang to perform SCI (scan-line interleaving) on conventional, 32-bit PCI.

Nvidia's PCI Express product uses a private bus over a dedicated physical bridge between controllers for passing rendering data. Unlike the 3Dfx approach, the Nvidia technology has advanced capabilities, such as scene-dependent load-balancing between the controllers to keep both equally busy.

ATI offers Crossfire, its version of cooperative dual graphics. Like SLI, Crossfire can boost the graphics-subsystem performance on heavily render-bound applications by cooperatively using the rendering engines in each controller, dividing the rendering task. Crossfire exploits the industry-standard DVI (Digital Visual

Interface) and PCI Express buses to speed rendering. The slave controller sends its rendered output to the master over a special DVI cable, on which it deposits its slave and master image data before transmission from the master to the display. The Crossfire cards need not to be adjacent, unlike with SLI.

Crossfire exploits yet another feature of PCI Express: the ability, if the chip set supports it, for peer-to-peer transactions to take place directly from one PCI Express bus to another. Crossfire masters and slaves can talk directly to each other under ATI Crossfire driver control and provide commands for synchronizing master- and

slave-rendered outputs. If one controller creates a texture map for use by both controllers, the render-to-texture data can also pass through peer-to-peer PCI Express.

Crossfire requires PCI Express peer-to-peer communication for synchronization. SLI also uses PCI Express peer-to-peer transfers in configurations requiring more bandwidth and in which it can provide more bandwidth. However, SLI can operate, perhaps at lower performance levels, even when PCI Express peer-to-peer capabilities are unavailable between buses on the host chip sets.

The benefit of parallelized graphics depends highly on the application. Generally, both controllers perform exactly the same geometry operations in parallel. The graphics task becomes split only in the rendering phase. Thus, a heavily geometry-bound application accrues no benefits. A heavily render-bound, fill-rate-limited application, such as a cutting-edge video game, could see nearly a doubling in frame rate. Crossfire supports an alternative-frame-rendering mode in which the controllers pingpong from

frame to frame. In that mode, Crossfire with one controller could theoretically provide performance improvements for application loads that are not render-bound.

Whether for multiple monitors or for a single monitor driven by dual graphics controllers, designing a system to support multiple PCI Express graphics controllers presents major challenges. Power consumption and heat represent two such challenges. Some of today's graphics controllers consume 150W each, and the trend is to go even higher. A full-speed dual-graphics system must include power-supply and cooling capabilities to accommodate requirements of 300W or greater just for the graphics subsystem. Worse, future graphics controllers should consume even more power. An effort is under way in the PCI-SIG to increase the maximum power consumption for every 16 graphics ports well beyond the current 150W maximum.

Nvidia recently launched quad SLI. It parallelizes four graphics controllers, connecting external power supplies directly to the graphics cards. Although nothing in the PCI Express

spec precludes this scenario, the architecture does present the designer with some significant power and thermal challenges.

A second challenge for dual graphics follows from that fact that a full-performance, dual-graphics client requires 16 ports to comply with the PCI-SIG's Revision 1.1 specification for PCI Express. However, few chip sets today have the requisite 32 lanes of PCI Express available to support graphics. Noncompliant systems have been hitting the market with one or both of the 16-port connectors electrically "plumbed" to handle only eight or even four ports. The lack of a 16-port connector, though it no longer violates the specification, risks situations in which a user may plug a card into such a slot, only to find that the card does not operate. The reason is that the PCI Express Revision 1.1 specification requires a 16-port card only to electrically support 16- and one-port configurations. Support for eight- and -four-port electrical connectivity is optional, although many currently available hosts and graphics controllers do support them. A multiplicity of provisioning options for the

16-port connector also risks end-user confusion and may lead to customer-support issues for OEMs.

Dual-graphics clients with electrically underprovisioned 16-port connectors may operate in with today's applications, because those applications do not tax the abilities of the 16-port PCI Express bus. Tomorrow's applications will, however. One of the worst consequences of underprovisioning, if it were to become prevalent, is that it might cause developers of graphics applications to hold back on pushing the graphics-performance envelope if a significant base of systems with eight or fewer ports existed. It could, in the worst case, undermine the point of going from AGP to PCI Express in the first place. Proponents of dual graphics argue that graphics developers want to take advantage of the increased graphics-processing power a dual-graphics system offers, counterbalancing this effect. If the effect is to cause performance shortfalls in systems with underprovisioned graphics slots, then market forces will simply eliminate these systems in favor of fully provisioned ones. The jury on this debate is still out.