PCI-Express a Uniform Protocol Across Applications | 2010

Brian Stark
Lead Software Engineer

Mission Technology Group Inc
dba Magma
9918 Via Pasar
San Diego, CA 92126

A bus can be defined as a shared medium for communication between multiple components of a computer. The architecture of today's computers have a local bus that is used to transmit data among the components "local" to the CPU. Typically serial interfaces are used to communicate with peripheral devices. This paper will discuss the advantages of the acceptance of PCI-Express - a serialized version of the PCI bus architecture and the potential markets that are opened as they relate to Expansion. We define Expansion to mean the expansion of the local bus beyond the mother board. The expansion market has traditionally been a niche market for applications which required multiplexing and extending I/O data channels.

With the serialization of the local bus there lies the potential for an industry changing influence on the architecture of the computing world that includes tighter communication between platforms, resource sharing of platforms, more efficient and cost effective parallel computing, and more. There will also be a market for supporting legacy hardware in the field with the new local bus.

Parallel vs. Serial

A Parallel bus is one in which data with the width of the bus 16, 32, 64 is transferred in one or two clock cycles (usually the address and data are multiplexed, and usually takes more than one clock for a transfer, typically two excepting for burst transfers that take one cycle). A parallel bus can have multiple devices sharing the lines of the bus. Bus arbitration is handled either by a central master, or devices on the bus that can be scheduled to have control of the bus. A serial bus sends packetized data from one point to another at a bit per clock cycle. Devices on serial buses are connected via a series of switches or hubs that route the packets based on an address embedded in the header of the packet.

Intuitively the advantage of a parallel bus would be speed: the greater the width and faster the clock, more the bandwidth. There is less overhead on a parallel bus as there is no need to transmit packet information. The disadvantage is the number of wires or traces needed to support a parallel bus.

The advantage of a serial connection is portability. There are fewer wires needed to transfer the data from point to point, so there is greater potential to transmit data "outside the box" to peripheral devices. Cabling is less dense and more flexible for serial interfaces. Advances in technology including clock and processing speed offset the cost of the extra overhead associated with a serial bus. With larger buffers the percentage of bandwidth used for overhead is further reduced. It is also easier to take advantage of higher clock speeds as each device can maximize its clock speed rather then using the clock speed of the slowest device on the bus. It is this advantage that makes the serial architecture more open to greater bandwidth requirements as technology advances. Slower components will not throttle the bus when advancements are made in bus speed as each link will be negotiated to its maximum potential.

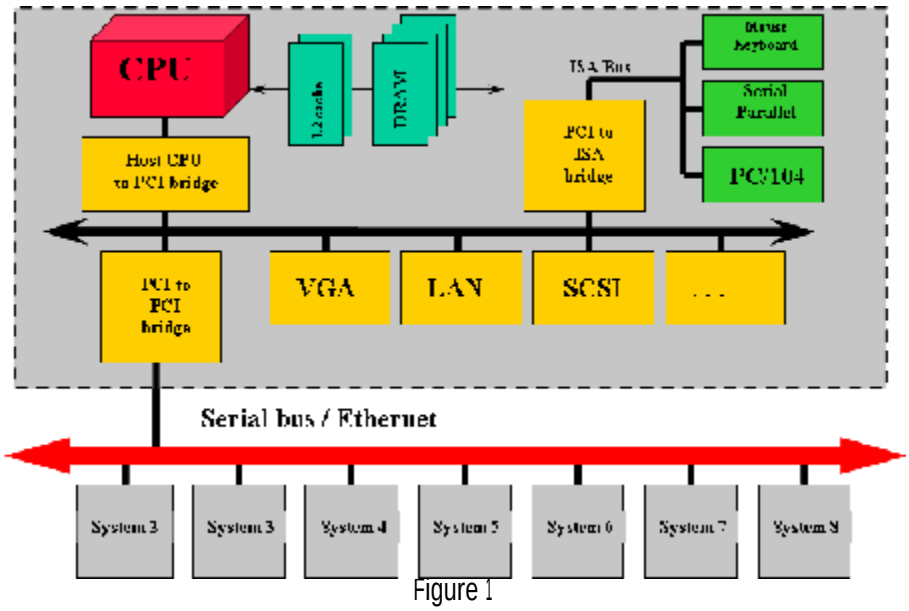

Figure 1

The architecture of I/O in the PC world has been to have a parallel bus (for example PCI, PCI-X etc.) for transmitting data locally within a platform, while using serial protocols for transmission outside of a platform (USB, Ethernet etc.). However once data is read or written to peripheral devices or other platforms via serial protocols this data is still made parallel on the local bus for use by the CPU via plug in cards or embedded controllers on the motherboard.

PCI-Express was developed as a serial local bus that maintained compatibility with the industry accepted PCI bus architecture. By maintaining compatibility with the configuration space of the PCI bus, PCI-Express based platforms could be developed and deployed without any changes to existing Operating Systems. This transition has been accomplished with unprecedented efficiency. Most end users could not tell you the difference between their new PCI-Express based platforms and the PCI based platforms they replaced with the exception of speed and performance. Because of this PCI-Express has laid a foundation for accomplishing what previous protocols and standards have been unable to achieve, a uniform protocol across all applications. Bandwidth can be increased on the PCI-Express bus by adding "lanes". The amount of bandwidth dedicated to a specific device is defined by the number of lanes a device uses. Adding lanes does increase the number of wires needed for an interface thus decreasing the portability for a high bandwidth device. Yet another method of increasing bandwidth would be an increase in the clock speed.

The industry acceptance of PCI-Express has created many markets for computer development. Magma is currently positioned to take advantage of two of those markets based on its current product offering and engineering experience. The first market is PCI-Express to PCI expansion. Given the multitude of PCI plug-in cards deployed, being sold, and in development, there is a market for providing the means for these cards to maintain their viability in the computer industry. The second market is PCI-Express to PCI-Express expansion. The more exciting of the two markets however may be the modularization of computer resources via the PCI-Express to PCI-Express expansion market.

PCI-Express to PCI

As platforms transition from PCI to PCI-Express there will be less PCI slots. This opens a natural market as there remains a large number of PCI plug in cards that are used in today's applications. The industry is slower to change then technology. It is likely that older Operating Systems and applications using PCI based devices will make up a significant percentage of the market for the next 5 years or so.

Magma is well positioned to take advantage of this market, as it has been the innovator in cabled expansion for the past 5 years. The strategy employed by Magma has been to use its existing product line as a basis for expansion while only replacing the PCI to PCI Bridge in its previous architecture with a PCI Express to PCI Bridge. Magma has also taken advantage of the increased bandwidth of PCI Express to provide other products to support this market, for example the PE6R4 a 6 slot PCI-X expansion chassis using a fiber cable to transmit bus data.

Figure 2

The PCI-Express to PCI market should hold long enough and may even increase in the short term to enable Magma to maintain its viability while increasing its effort in the PCI-Express to PCI-Express expansion market. While Magma no longer has the exclusivity of its IP patents for PCI Express to PCI Express expansion, its experience in the expansion market in both engineering and marketing should enable it to be a solid player in what is likely to be a very large market.

PCI-Express to PCI-Express

PCI to PCI expansion was a niche market. The market was supported by applications that required multiple IO channels synchronized on a common clock (for example multi-channel recording and reconstruction), the ability to use PCI plug in cards with laptop computers using the card bus form factor and a few other applications. With expandability built into the PCI-Express specification, including hot plug requirements, the market for expansion should go from a niche market to a major market. This section will explore two potential applications for local bus expansion.

Parallel Computing Using Server Clusters

In the early nineties there was push for massively parallel architectures. The "super computer" is an example of this architecture. However the cost of developing these architectures was too high to be supported by the commercial market. This coupled with the exponential growth of processor speeds slowed the acceptance of a massively parallel architecture, as PC computers could easily keep up with the market's computing demands at a much lower cost and development time.

Probably the largest growth sector of the technology industry in the last ten years has been in Internet search. In the beginning of Internet search the strategy was to focus on a manageable percentage of the Internet for search queries. This was done by manually researching "quality sites". Google was the first search company to attempt to index the entire web. This was only achievable using a massively parallel architecture. In order to make the endeavor cost affective, Google used a cluster of inexpensive servers running the Linux operating system networked via Ethernet to create a cost affective massively parallel environment.[1] With PCI-Express it may now be possible to marry the speed advantages of local bus communication, with the clustering architecture employed by the Google File System. Essentially, "super computers" can be built cost effectively using existing networking techniques

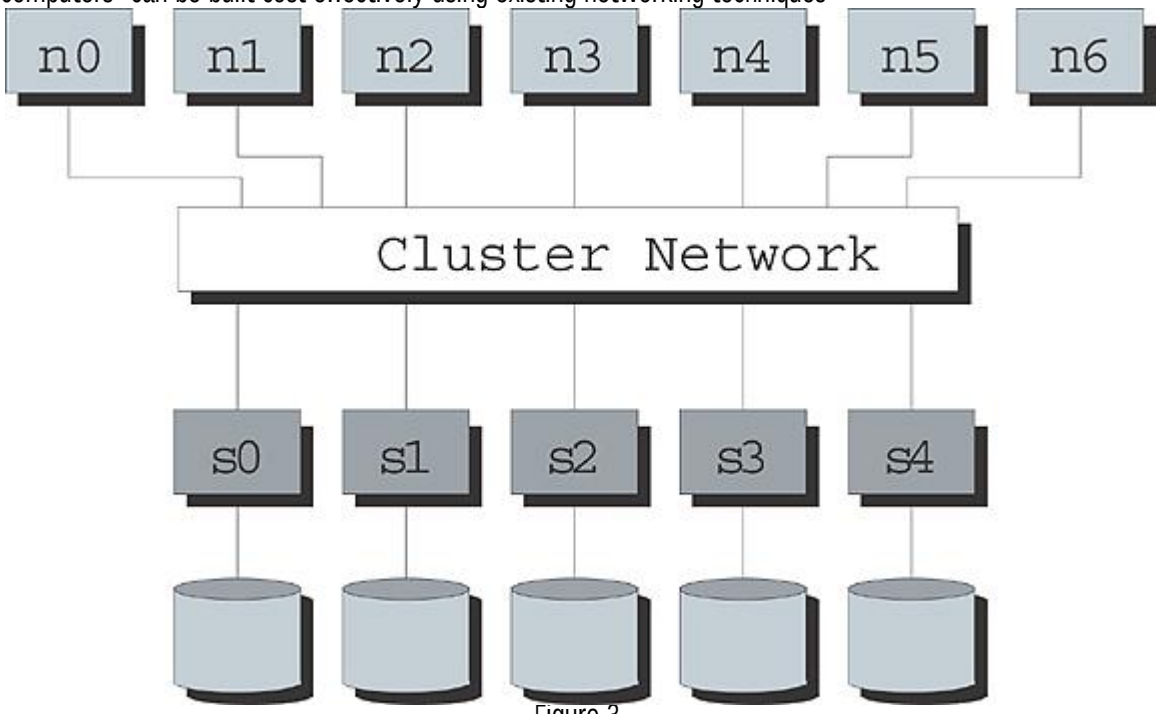

Figure 3

The GPU and Expansion

One of the first markets to adapt to PCI-Express was the graphics market. The power and cooling requirements of the Graphics cards make them natural candidates for expansion, extra power and cooling is more easily applied to the cards in an external chassis. This is especially true in the laptop environment. The ExpressCard form factor found on laptops today provide one lane of PCI-Express for expansion. Currently it is assumed that a minimum of eight lanes should be dedicated to a GPU. However the 3DMark benchmark implies that there can be tremendous value added to a laptop with only a single lane. Monitoring an ATI 1900X board using an Agilent analyzer showed that the actual traffic on the PCI-Express bus was less then half of the 2.5 Gigabit limit of the bus. Running the test showed less then 5 % degradation to one lane from eight lanes. While an NVIDIA card running the same benchmark showed a much greater degradation and higher bandwidth usage, the first test implies that the problem may be more software related then hardware related. This is not to imply that there are not other applications that would require the greater bandwidth, but it seems likely that there is a greater potential for the graphics market using the ExpressCard form factor then originally estimated.

The current push within the graphics industry is to take advantage of the computational strengths of the GPU in areas other then graphics. If this idea comes to fruition, then the GPU will be treated more like a specialized CPU then a plug in card. This should fit in well with the parallel computing model discussed earlier.

The local bus expansion market should become the backbone of the industries push to a more modularized view of resources. Another industry trend that fits in well with these concepts is virtualization. With virtualization an application should be able to take advantage of whatever resources are available. Expansion of the local bus is also a solid fit with the virtualization model as networks of computers become more tightly coupled.

Common Pitfalls in Expansion

Older Operating Systems treat the local bus to be static in nature. Some devices may be introduced after boot, however the infrastructure of the bus is generally considered to be static. This is typical of pre-boot environments as well. Given the niche nature of expansion before PCI-Express, this is completely understandable. It does not make sense to expend resources on supporting the complexity of a dynamically changing bus (rebalancing resources and re-enumerating the bus) to support a tiny fragment of the market.

This view of the local bus was sufficient for PCI expansion in desktop/server platforms where a hot-plugging PCI bridge is not typical. The BIOS support for these platforms was in general sufficient.

The Card bus support on laptops was not designed for PCI expansion. In order to facilitate expansion on laptop platforms, modifications needed to be made to an Operating System's local bus configuration. In Unix style Operating Systems and Apple Operating Systems this was done by modifying the kernel to support the resources that needed to be supplied downstream. Under Windows based platforms a "Bus Filter Driver" was employed to make changes to the local bus configuration during the systems initialization. A paper was published on the Microsoft Developers Network (MSDN) that illustrates this issue quite well.[2]

Windows Vista has added dynamic bus rebalancing into its driver supporting the PCI bus which essentially replaces the functionality of the Bus Filter Driver, however testing has shown that Vista may or may not rebalance the PCI bus depending on how the pre-boot environment has initialized the local bus. This causes problems when an architecture downstream is introduced that requires resource usage greater then that which is provided upstream. Vista also does not re-enumerate the PCI bus after boot. This can also cause problems if an architecture is introduced downstream that has a deeper bus level then is available (see figure 4) for examples.

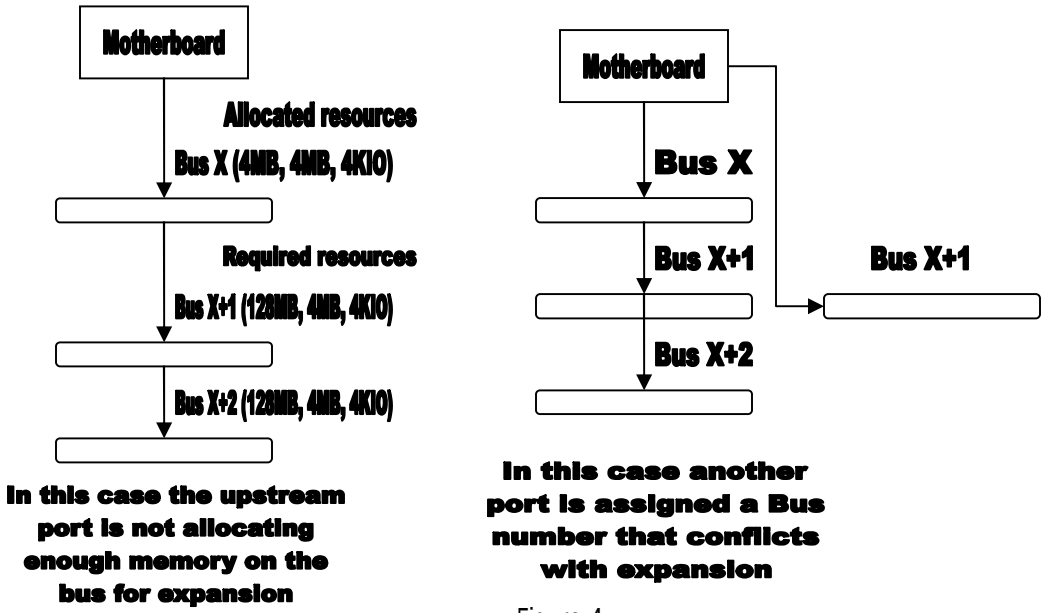

Figure 4

Most manufacturers of PCI-Express plug in cards supported their applications with the PCI bus prior to developing PCI-Express cards. Typically these new cards use the same drivers as their PCI versions, with little or no modifications. Since hot-plugging was not an issue for these cards many of their drivers do not fully support hot-plugging. As an example in the Windows environment, a message is sent to these devices when a hot-plug event takes place. There are two messages that may be sent: one requests that a driver stops, the other informs the driver that a surprise removal event has taken place. If these messages are not handled correctly by the driver the applications supported by the hardware may fail in some manner. In the more serious cases this failure may lead to a system fault (Blue Screen Of Death BSOD) or a system lock up. What this exemplifies is that hot-plugging is a software issue as well as a hardware issue.

Given the niche market of expansion using the Cardbus form factor, there was never any support put into notebook computers for local bus expansion. As a result the pre-boot environments of notebook computers tend to focus on speed rather then expandability. As a result the philosophy of most notebook pre-boot environments is only on devices that are built into the local bus. The ExpressCard form factor provides a local bus expansion slot to the outside world. Support for this expansion has been slow among the notebook system designer's.

The pre-boot environment of the aforementioned platforms may not initialize the resources upstream to support the requirements downstream. This is the same issue described above related to Windows Vista and illustrated in figure 4. On platforms using earlier Operating Systems this limitation can usually be overcome by a Bus Filter Driver.

Another common problem with pre-boot environments arises when a graphics card is introduced. Since the designers of these platforms did not anticipate the introduction of external graphics into their system the pre-boot environment may become "confused" as to which graphical device to use during initialization. A typical result of this is a blank screen, or the system may fail to initialize.

In general the introduction of external graphic cards or multiple graphic cards is new to the industry. There is currently a push within the industry for multi-GPU environments (SLI and Crossfire). While this fits in well with PCI-Express expansion, it has also created compatibility problems introduced by using different vendors within the same platform, or multiple graphic cards that were not designed for a multi-GPU environment. Apple's philosophy has been to qualify graphics card before they are allowed to be used in their system, while Microsoft has restricted the use of their new driver model under Vista to one executable.

The implication of the latter is that any GPU introduced into a system must be supported by the one driver used by that system. Unless a special driver is written that will support the cards of multiple vendors (none exist so far), each system will be effectively restricted to one vendor's (ATI, NVIDIA, Intel etc.) hardware. Further it is up to the graphic vendors to support all the models used in their driver. The implication of this in the notebook world is that only external graphics cards that are supported by the driver written for the embedded chipset may be introduced into the system. As with any significant changes in a market the earliest video drivers were not as robust, but significant improvement is being made as Vista gains industry acceptance.

To summarize the above, while electrically most issues involving expansion of the local bus both at boot and after OS initialization are resolved, there remains a lot of work in the pre-boot and initialization software environments to take advantage of the capabilities of PCI-Express based platforms. Furthermore the nature of these issues will require cooperation among many different developers and stages of software.

Call to Action

- Developers of the pre-boot environments BIOS and UEFI, need to support greater configurability of their local bus in order to take advantage of the characteristics of PCI-Express. For example it would simplify expansion if the resource windows used by the ExpressCard port were configurable in the BIOS setup.
- Different Vendors of CPUs and GPUs should make a push towards standardization in their respective markets. While this may create an initial fluctuation in the market share of the respective vendors, the innovation that would result could potentially increase the overall market size. Competition will be driven by innovation and quality more then by existing market share.
- Current graphics benchmark scores do not necessarily scale with bandwidth availability, graphics vendors need to research how much bandwidth is actually needed for specific graphic applications and if portability of higher graphic applications can be achieved with software efficiency.
- Increased support for the modular architecture in existing Operating Systems through virtualization support. Operating Systems should become more network centric rather then platform centric.
- Plug in card manufacturers need to design their drivers such that they are aware that they may be entering an existing environment, or leaving a running environment.
- PCI-Express aware Operating Systems need to clearly define the algorithm used in determining whether or not to re-balance or re-enumerate the local bus.

---

[1] "The Google File System" - Sanjay Ghenawat, Howard Gobioff, Shun-Tak Leung

[2] "PCI-to-PCI Bridges and CardBus Controllers on Windows", http://www.microsoft.com/whdc/system/bus/PCI/pcibridge-cardbus.mspx, May 7, 2004

About Magma:

Since its founding in 1987, Magma's flagship product has been the patented PCI Expansion System which provides the capability to add additional PCI slots to desktop computers, servers, PC notebooks, and Apple PowerBooks. As computer technology evolved, so has Magma, releasing innovative PCI-X, and now, PCI Express (PCI-E) compatible products.

Magma's key markets include: audio/video and music production, test and measurement, federal and state government, and the defense industry. Within the last year Magma has increased its presence in telecommunications and IT infrastructure. Magma is pleased to be a part of an ever changing and challenging market. We continue to grow and plan to remain at the forefront of the computer hardware industry.