Applications

PCI Express Expansion in the Industrial Marketplace

Market Needs and Applications of PCI Express Expansion Products

By: Jim Ison
Product Marketing Manager
One Stop Systems
2235 Enterprise Street
Escondido, CA 92029
Date: January 2005

Add-in card bus for the next 10 years

Approximately every ten years, a radical change in technology is required to grow computer system performance over the next ten years. With the add-in card bus PCI becoming a maximized technology, a new add-in card bus is required. PCI Express has been designed to scale in speed relative to the CPU technology over the next 10 years. In contrast to the traditional parallel PCI busses using 32 or 64 individual wires to both transmit and receive signals, PCI Express in its most basic form is a serial point-to-point, packet-based device interconnect using both a transmit (Tx) pair and receive (Rx) pair of wires. This four wire interconnect, known as a PCI Express lane, communicates at 2.5 Gb/s full duplex. Effective bandwidth of this x1 (pronounced "by one") PCI Express link is therefore 5 Gb/s. By using multiple lanes between two devices it is possible for the designer to increase speeds to 160 Gb/s with a x32 PCI Express link. In comparison, mainstream PCI-X 133 add-in boards are limited to 8 Gb/s bandwidth. Bandwidth comparisons between PCI Express and other well-known technologies are shown in Figure 1. By the end of 2006, second generation PCI Express will reach speeds of 320 Gb/s in a x32 link thanks to doubling the effective clock speed of the basic lane. In addition to increased speed, packets of PCI Express data contain several new features over traditional PCI. The primary enhancement is quality of service (QoS) features including flow control and error control. Additionally, PCI Express is fully backward compatible with PCI developed applications and operating systems which is typically a significant investment of the OEM.

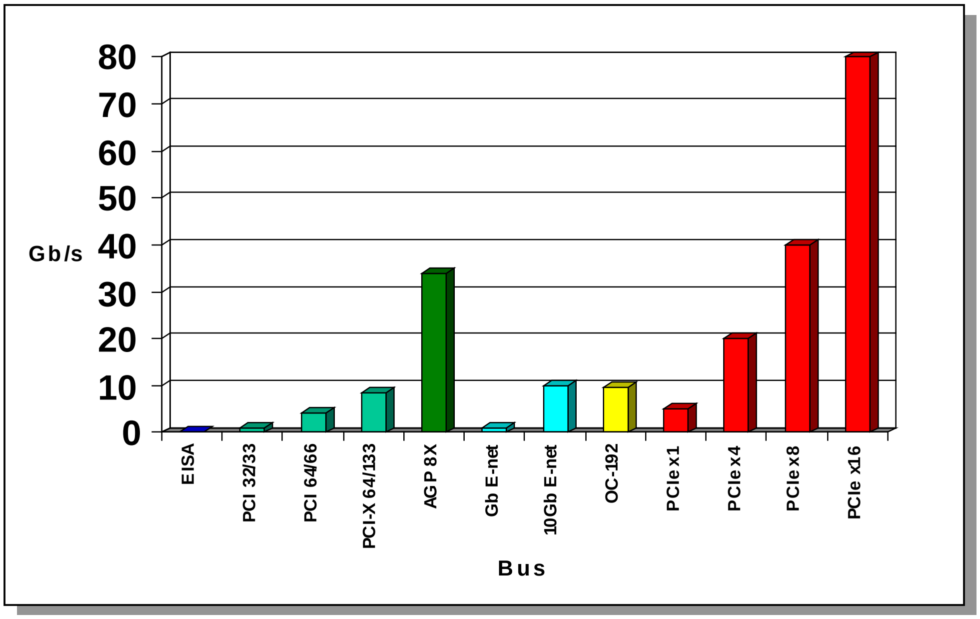

Figure 1: Bandwidth comparison

Market influence & the need for expansion

Motherboard manufacturers have had to address the issue of how many add-in card slots should be designed into their latest motherboard. Intel, IBM, and Sun, the largest of these manufacturers, have followed similar technologies in recent history for these add-in cards. In the late 1980's to early 1990's there was relatively little integrated functionality on the motherboard leaving video, sound, communications and even the simplest I/O functions to add-in cards. During this time when the ISA and EISA bus dominated the add-in card areas of motherboards, it was not uncommon to see at least 8 add-in card slots. Many of the add-in card slots were used for the basic functions mentioned above. In contrast, more modern PCI/PCI-X motherboards typically have larger amounts of integrated functionality depending on the target customer. The feature set of these motherboards varies from limited integration desktop units to extremely well integrated server motherboards with several Ethernet ports, several disk drive interconnect technologies (ATA, SATA & SCSI), traditional serial and parallel I/O, PS/2 serial, 4 or more USB ports, sound, etc. This has lead to less emphasis on the need for add-in card slots on the motherboard. In the case of the newest PCI-X and PCI Express motherboards there may be a dedicated high-speed video slot and only 2-5 add-in card slots. As the latest round of chip sets are introduced, add-in card slot counts may be decreased to only 2 or 3 due to the limit of the available busses from the host chip set.

Many of the decisions made by the motherboard manufacturer, especially related to add-in card slot count, are based on price pressure requirements of the high volume consumer and the availability of new PCI Express busses from the base processor chip set. These chip sets, especially for high end server CPUs, may have only two or three point-to-point interfaces to PCI Express add-in card slots. Unlike the parallel PCI bus that may allow several add-in cards to exist on the same bus, PCI Express will only allow one device per interface. Switching will be required for fan-out to additional add-in card slots. Typical switch chips may add from $40 to $120 per chip depending on the port count of the switch. This is a large cost burden for a typical mainstream manufacturer of motherboards to add to their product. At the very least an on-board switch may significantly lower the potential return on investment (ROI) of the product. With more and more consumer and commercial off-the-shelf (COTS) systems being utilized in the traditional rugged industrial, communications and military markets, these relatively low add-in board counts may not address the high I/O needs of the most demanding applications. The answer to these challenges lies in expansion products. More specifically, the answer is expansion products based on the latest native add-in card technology, PCI Express. PCI Express expansion using cables or fiber optics will address the traditional needs of the high add-in card count application as well as open new applications due to the extremely high bandwidth now available over a cable.

Basic Switched Expansion: The "tree" approach

The "tree" approach to bus expansion is a simple expansion of a single system host bus to more add-in cards than originally designed into the host system. In this configuration a single add-in card slot with a host interface card inserted can cable to an external expansion system with multiple add-in card slots that continue to be controlled by the master system (Figure 2). With a traditional parallel add-in card bus, such as PCI, the bus signal characteristics do not lend themselves to transmission over a cable to an expansion system. This lead to several vendors creating proprietary, and often costly, methods of bus expansion including reflective memory, Myrinet, and, more recently, switched fabric technologies such as Infiniband and StarFabric. Besides high cost, these technologies injected latencies detrimental to the throughput of the system; they required translation of the native add-in card bus to the expansion bus, then back

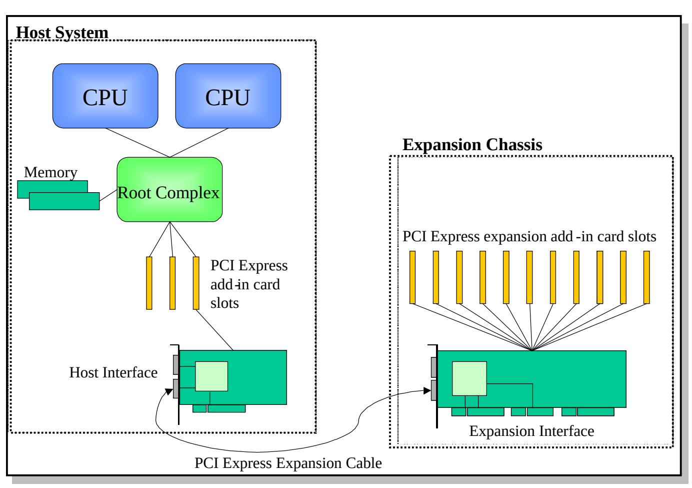

Figure 2: Basic expansion system

again. While Figure 2 shows a system where the host system, host interface, cable, expansion interface and expansion slots are all PCI Express, other switched fabrics have to translate the PCI signals into the proprietary or non-PCI fabric then back again using bridges. The adoption of PCI Express as the native add-in card bus of the next generation of motherboards and a multitude of industrial computer form factors has reduced both the relative cost of expansion systems and inherent latencies of traditional expansion technologies.

The PCI Express signal characteristics easily allow for the native bus to be carried over a cable. This capability can be used to enhance systems that have 2-3 PCI Express add-in card slots due to the limited connections provided by motherboard manufacturers. PCI Express implementations over cable are defined by the PCI-SIG standards organization. The specification is designed for implementations up to 80 Gigabits per second, or x16, in the first generation. Second-generation speeds are planned for up to 160 Gigabits per second in late 2006. Current applications will likely utilize the lower defined cabled data rates of x4 or x8 PCI Express at 20 or 40 Gb/s respectively since these speeds more closely match the data rates of contemporary chip technology. The higher speed x16 PCI Express cabled expansion becomes more compelling in the networking topology between high-end clustered servers.

Expanding a PCI Express host system consists of placing a host interface board (HIB) with a PCI Express switch acting as a repeater into an add-in card slot, cabling from the HIB to an expansion link board (ELB) in an expansion system which consists of a switch to provide fan-out to multiple expansion slots as shown in Figure 2. It is important to note that this is an extension of the native add-in card bus without translation to a secondary technology. Switches provide the end point connections that also allow for less latency than bridges of traditional expansion systems. PCI Express is uniquely suited to replace these traditional cabled expansion systems of legacy PCI or PCI-X system using PCI-X to PCI Express bridges in much the same fashion as StarFabric or Infiniband. Due to the inherent software compatibility of PCI Express to PCI, PCI Express over cable can be done very straight forwardly, at a lower cost, lower latency and at higher cabled data speeds than other expansion protocols.

There are several market segments that can benefit from the various topologies available with PCI Express expansion systems. For highly embedded applications an add-in card can be located in a machine up to 6 meters away from the host connected by cable. This topology would allow an I/O card to be located closer to a sensitive instrument while the main host computer remains in the rack portion of the equipment. In some industrial machinery applications with large swing-arm display panels using high resolution graphics, the host computer can remain in the rack portion of the equipment and allow the high speed video and user interface to be located several meters away at the end of the swing arm without the use of video extension devices. This is especially compelling since video, such as the latest x16 PCI Express video cards, operate at speeds up to 80 Gb/s.

In a control system or biotechnology clean room environment the host operator console can be located in a control room outside of a hazardous area while the I/O portion of the system resides up to 6 meters away, using copper cables, or—by the second half of 2005—up to several miles away at high speeds using fiber optics. This model can also be adapted to the office environment where the host system and networking functions reside in a central data rack in the network room and the user workstation consists of only the monitor, removable drives, and human interface devices in a single, quiet, expansion enclosure.

In addition to potentially revolutionizing the desktop PC, laptops can also benefit from PCI Express expansion. In this application the laptop computer can be fashioned with an external PCI Express slot that can accept a standard cable. This cable would be connected to a docking station that has a similar PCI Express external cable interface. The docking station may consist of the connectivity to the office network, local printers or plotters, high-resolution graphics adapter and monitor. This model decouples the mechanicals of the docking station from that of the particular model laptop. For the laptop manufacturer a new docking station design is not required every time a connector has to move or molded plastic enclosure changes. This allows a docking station design to remain unchanged through several generations of laptops. For the user, docking station features may become more flexible as docking stations from 3rd party vendors become available. Smaller docking station manufacturers would be able to provide more specialty features that large volume, mainstream manufacturers may not provide. This would also allow docking station vendors to offer enhanced features such as additional add-in card slots for special function I/O boards, built-in PDA, digital music or digital camera links.

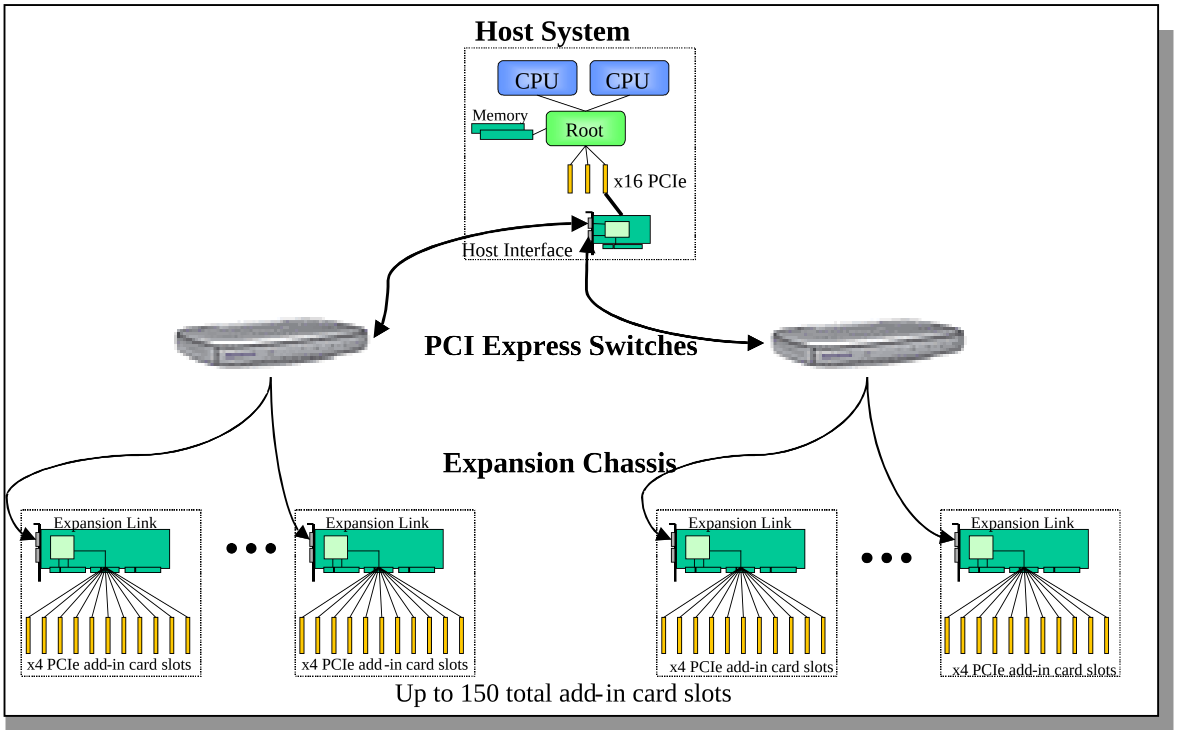

Figure 3: Large expansion system using basic switching

Industrial, converged communications and military markets have several opportunities to take advantage of PCI Express expansion. With basic switched expansion, a single host can control up to 150 add-in boards. This is accomplished with a combination of a single PCI Express cable host interface board in the host system attached to multiple expansion chassis. 1U PCI Express switches in the system can assist in keeping an efficient flat topology that provisions the PCI Express bandwidth. A system consisting of large numbers of add-in boards may consist of two PCI Express 1U switches and up to eight expansion chassis of 19 add-in boards each as shown in Figure 3. The host interface is inserted into a x16, 80 Gb/s PCI Express slot connected to the processor chipset or root complex. The host interface connects to each of the PCI Express switches with a x8 or 40 Gb/s cable connection. From here, the PCI Express switches fan out to multiple PCI Express expansion systems with a x4 or 20Gb/s cabled link and multiple x4 PCI Express add-in expansion slots per chassis. While this is a compelling system, care should be taken at the types of I/O boards used in such a system to not overburden the single host processor. Intelligent add-in boards that require a minimum of host processor cycles but require shared access to the main system memory are ideal for this topology. In addition, creative configuration of the topology

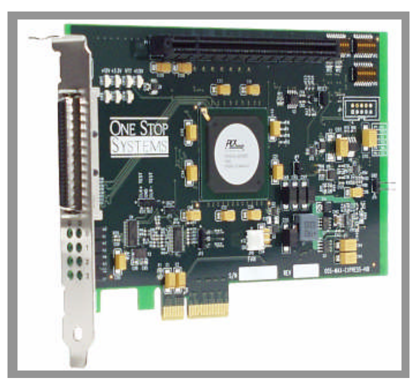

Figure 4: Host Interface Board

of such a single host system to allow for high-traffic I/O board to I/O board transfers through a single switch leads to efficient system design with respect to use of available bandwidth. By placing I/O boards that require intensive communication on the same switch in the system, traffic can be isolated from the remainder of the system communication as shown in Figure 5.

High-density, high-availability single-host computing applications also benefit from the cabled PCI Express bus. These applications typically require a high-density of I/O, ability to hot-swap I/O boards and the latest technology processors. In some instances the horsepower of a dual processor or even 4 or 8-way processor is required in this application. With the limits of the physical form factor of these high density systems such as PXI, CompactPCI or CompactPCI Express (CPCIe), it may not be possible to incorporate these larger or higher powered CPUs into the design without using multiple slots or causing excessive thermal issues. PCI Express cabling can allow a system to take advantage of the latest high-performance server systems on the market and the hot-swap, front loaded I/O boards of the PXI, CompactPCI or CompactPCI Express form factor while only minimally increasing required used rack space. In the case of a CompactPCI Express system a 1U host system with a PCI Express HIB can be cabled to a CPCIe ELB inserted in the CPCIe host slot, allowing the 1U host to control the high-density CPCIe I/O boards. In the case of PXI or CompactPCI the 1U host system with the same HIB will be cabled to an Expansion Bridge Board (EBB) inserted into the PXI or CompactPCI host slot. The EBB consists of a x4 PCI Express to PCI-X bridge allowing the 1U PCI Express host to control the PCI based I/O subsystem. Since all of the communications are PCI with only a serial to parallel translation handled by the bridge board, this configuration is transparent to the operating system and application. There are also no special software drivers required.

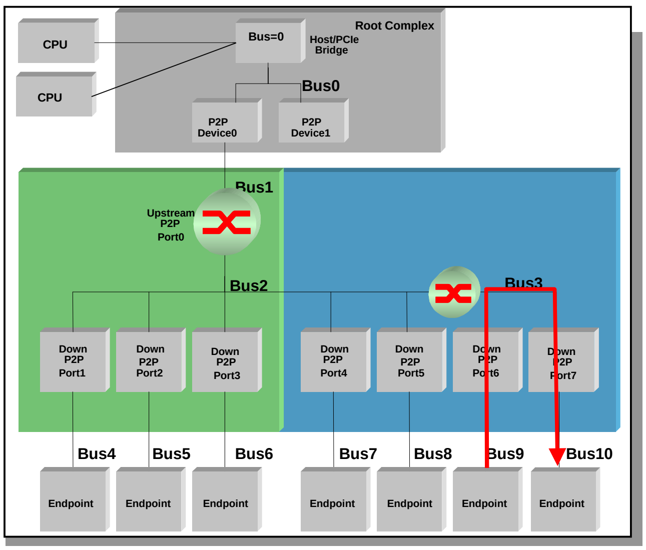

Figure 5: Switches as traffic isolation

The software compatibility of PCI Express to legacy PCI cannot be understated. In fact, this is the primary differentiator between PCI Express and other serial busses such as Infiniband and Rapid I/O. PCI Express cable expansion systems maintain the software compatibility and plug-and-play features we expect over the cabled link allowing add-in cards to be added to an expanded system as if they were being added to the base add-in card slots of a motherboard. For example, a PCI Express Ethernet Network Interface Card (NIC) inserted into a motherboard PCI Express add-in slot will auto detect and install any software drivers specific to that card using the plug-and-play method we are used to with PCI-based systems. When that same Ethernet NIC is inserted into the expansion chassis, which is cabled to an HIB inserted into the motherboard add-in card slot, the plug-and-play approach operates in exactly the same manner. While the other serial busses mentioned above have their market niches, PCI Express is the bus of choice for the commercial PC add-in bus due to the software compatibility and speed, solidifying it as the mainstream serial bus of choice for at least the next 10 years.

Advanced Switching Expansion: The "network" approach

With the adoption of Advanced Switching PCI Express in the second half of 2005 expanding the possibilities of networked systems with ultra-high bandwidth interconnects of 80 to 160 Gigabits per second becomes possible. In comparison, Ethernet networks are typically 1 Gb/s and lack high availability features. Other networked approaches that incorporate the high-availability features that Ethernet lacks, include Myrinet, Infiniband, Rapid I/O and Star Fabric. Myrinet has been widely used in high-availability (HA) computer clusters for the last 8 years. Currently its development lags other fabrics on the market and its speed is limited to 4 Gb/s. Infiniband, Rapid I/O and Star Fabric reach higher speed interconnects and incorporate most of the HA features required for clustering and fail over. What they lack is the lower cost that the large economy of scale PCI Express will bring from the desktop market to the HA market and the low latencies associated with the use of the native host bus of the CPU root complex. Single systems with multiple masters with an ultra-high speed interconnect or a network of several systems each with a host CPU controlling large numbers of I/O cards paves the way for larger clustered systems attacking a single application in an almost unlimited fashion. With PCI Express Advanced Switching, the addition of HA features inherent in other interconnects, such as heartbeat continuity monitoring and alternative redundant communications paths, will offer a high speed interconnect at a reasonable price point.

Advanced switching allows a virtual unlimited networked cluster of systems with multiple I/O functions per system to communicate over a cabled PCI Express network. Communications core equipment or the most demanding processor intensive applications can benefit from this clustering ability of advanced switching. In an I/O intensive application, advanced switching can be combined with the basic switched expansion system. This hybrid "network-tree" topology allows an unlimited number of hosts with up to 150 add-in boards each to communicate over redundant links at high speeds using the same PCI compatible communications throughout the network as shown in Figure 6.

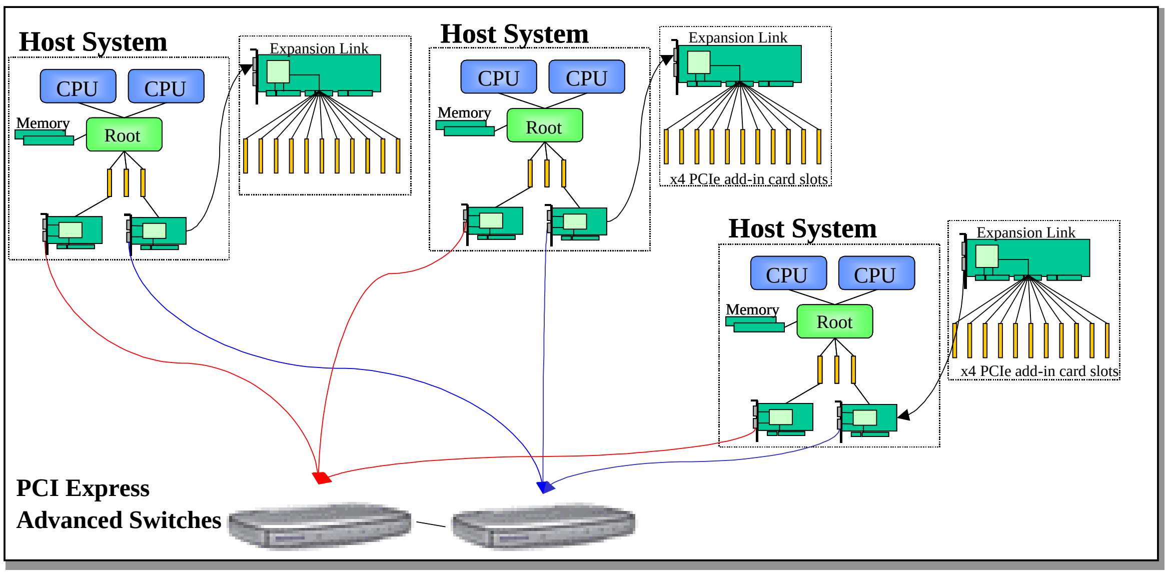

Figure 6: Compound "network-tree" clustered expansion system

PCI Express is changing the landscape of computer technology from the consumer market to the largest telecommunications infrastructure OEM. The "serialization" of high-speed data connectivity is affecting all aspects of the traditional computer from the add-in card bus to the chip-to-chip connections to the data storage functions. The last great bastion of parallel technology, system memory, is also likely to fall to the serialization movement in the coming years. PCI Express not only increases bus speed but also increases system flexibility through the use of bus cabling allowing virtually unlimited system expansion via networking. With over 54 different PCI Express based form factors under definition in 2004-2005, connection between systems of similar or different PCI Express and PCI based form factors can be accomplished with expansion products. Expansion and interconnection will occur over seven meters with standard cables or kilometers with the adoption of high-speed fiber optics.

Jim Ison is the Product Marketing Manager for One Stop Systems and has more than 11 years experience in the bus-board marketplace. Prior to One Stop Systems Jim has held various field sales and management positions centered on industrial and converged communications accounts for Ziatech Corporation and Rittal Corporation. More recently, he has held the global positions of CompactPCI Product Manager and Director of OEM Business development with I-Bus. One Stop Systems is the Secretary of the PICMG 1.3 SHB Express specification and working member of the PICMG EXP.0 and XMC subcommittees. One Stop Systems is also a member of the PCI-SIG PCI Express Cable committee. Jim holds a BS in Aeronautical Engineering from California State Polytechnic University at San Luis Obispo.

Ihr kompetenter Ansprechpartner:

BRESSNER Technology GmbH

Breslauer Str. 34

D - 82194 Gröbenzell

Tel. 08142 / 47284-0

Fax 08142 / 47284-77

E-Mail: info@bressner.de