Gone are the days when communications equipment manufacturers could employ sophisticated, proprietary backplane designs with custom ASICs, take two or more years to develop a product, and then charge a premium for it.

On the other hand, manufacturers that might have chosen standards such as PCI, H.110, or CompactPCI to keep costs down and speed product development are discovering the outer limits of those technologies in next-generation equipment. The scalability, performance, and quality of service requirements from customers are going up, and the standards today don’t have the horsepower to meet them.

As a result, equipment manufacturers are now adopting the best of both worlds: an open switch fabric architecture that combines the compatibility, low cost, and ease of design of standards-based bus architectures like PCI, H.110, and CompactPCI, with the scalability, physical form-factor options, high-availability features, and multiple traffic classes of custom backplane designs.

PCI has enjoyed great success for the control plane of mid-range and low-end communication equipment like DSLAMs, voice-over-network gateways, wireless base stations, and multiservice access platforms. Manufacturers had ample access to components that were affordable and offered performance, ease of design, multi-vendor interoperability, and software compatibility.

However, even for mid-range and low-end products, PCI presents several limitations in terms of reliability and scalability. For example, as a bus-based architecture, any single device connected to the bus can bring down the entire system, and all devices connected to it are forced to share its 4-Gbits/sec of bandwidth (at 64-bit/66 MHz). Furthermore, PCI limits the total number of devices that can use a single bus, typically five to eight depending on the speed of the bus. Also limiting scalability is PCI’s physical design limit of about 1m. Therefore, it can only support single chassis solutions.

H.110, as a telecommunications standard, presents very similar challenges to product designers.

Enhancing the standards: The move to switch fabrics

Because of these issues, the industry has looked toward open switch fabric designs to provide the power and flexibility needed to build next-generation communication equipment while meeting cost, compatibility, ease-of-design, and time-to-market requirements.

Switch fabrics provide the scalability and reliability not found in bus-based interconnects. For example, each end point in an open switch fabric is connected to every other end point through one or a series of switches. End points can be considered bridges to existing standard buses or components. Consequently, open switch fabrics allow many devices the ability to transmit and receive simultaneously. By building a complex mesh with a series of end points and switches, many different topologies with ever-scalable bandwidth can be supported.

Maintaining multiple routes between the same two end points addresses another key requirement in next-generation systems – reliability. If one route fails, traffic takes an alternative route. This additional reliability not only ensures no single point of failure, but also allows devices to be added or removed without affecting the overall system.

To address the need for an elegant evolution from current bus-based architectures to an open switch fabric, the PCI Industrial Computer Manufacturing Group (PICMG) has embarked on the CompactPCI StarFabric Specification, the PICMG 2.17 initiative. The specification, based on StarGen’s StarFabric technology, will enable system and board developers to realize the many benefits of a multi-Gigabit switched interconnect within the existing PICMG 2.x CompactPCI framework.

Sponsored by Agere Systems, Bustronic/Elma, PCI, Hybricon, Motorola, Natural MicroSystems, Pigeon Point, StarGen, and Sun Microsystems, the PICMG 2.17 subcommittee currently has 65 participating companies.

PICMG 2.17, when approved, will support a variety of system topologies, including:

- Bus-compatible and bus-less system designs
- Centralized and distributed switching configurations
- PCI mezzanine modules
- Multi-chassis systems

Because PICMG 2.17 is based on the StarFabric switched interconnect technology, it will address a number of issues currently facing system designers, including system scalability, high availability, and support for multiple traffic types and communication protocols with a single interconnect fabric. With an emphasis on providing an evolutionary path with strong investment protection, the specification provides for the reuse of current CompactPCI line cards and processor cards, as well as the ability to get started with no software changes using existing CompactPCI software, including BIOS, drivers, and operating system support.

Copyright 2002 Applied Computing
A blueprint for open switch-fabric applications

Although open switch fabrics are relatively new and the PICMG 2.17 standard is still in the final stages of approval by leading manufacturers, many real-world applications are currently under development that attest to the breadth and depth of the technology’s applicability for communications device design.

A few such applications that are employing StarFabric technology today are described below.

Access concentrator

Successful telecommunication access systems such as Class 5 switches and digital loop carriers have always been based on a large number of dumb line cards and an intelligent centralized switch fabric. Current generation DSLAMs, on the other hand, require that each line card has some intelligence to perform ATM-level traffic management. This architecture is too expensive, too difficult to manage, and requires too much power.

At the device level, the architecture that broadband access requires is a centralized computing architecture much like that of the POTS-based access infrastructure. The line cards act like devices, and the communication model is a processor-to-device model. If the uplinks have processors, the network processing unit (NPU)-to-Uplink communication model is a processor-to-processor unit.

StarFabric supports both types of communication models over a unified switch fabric. Figure 1 illustrates the generic StarFabric architecture. Here, the line cards are dumb. This architecture requires only the PHYs and Utopia-to-StarFabric Bridge devices. These cards are less expensive as well, since a single lower-cost, low-power Utopia-to-StarFabric Bridge replaces the multiple Layer 2 processors. Line cards are the cost multipliers in an access concentrator where many low-speed links are multiplexed onto a few high-speed uplinks. Thus, a cheaper line card drastically reduces the overall system cost.

The NPU cards are made up of an array of network processors. Each Utopia stream from a PHY is provisioned to an NPU port for processing. In the event that an NPU port failed, the controller would simply redirect the PHY to a known good port.

The WAN cards may or may not require additional traffic shaping (Layer 2) processors. The function could be in the NPU cards. In either case, the NPU cards would send the processed ATM flows to the uplink card via the fabric. The switch fabric would be implemented in two cards for redundancy, made of a redundant SG1010-based StarFabric switch topology.

With this architecture, the service provider could easily upgrade the DSLAMs software, because each line card, of which there could be hundreds, would not require a software upgrade. Only the uplink cards, if intelligent, and the NPU cards would require an upgrade. With this approach, the disruption of a system upgrade is minimal.

In addition, the service provider has the ability to upgrade the NPU cards as required to support new services. In the original architecture, each line card would need to be replaced if the Layer 2 processors could not support the new feature or protocol.

The system will ultimately grow to a point where the uplink is too oversubscribed for the service provider’s business model. Here again, the WAN cards can be upgraded independently of the other cards.

As shelves are added in the rack (see Figure 2), the second shelf contains two switch fabric cards and high-density line cards. The results are higher port densities and lower costs at the system level.

As the system scales to multiple shelves and to multiple racks, the service provider can simply add more NPU cards to process the additional line cards. As noted before, the uplink cards can be upgraded to meet the over-subscription ratio of the carrier’s business model.

A large access system will have a lot of traffic movement from card-to-card and from shelf-to-shelf. The original architect is overburdened by having to process protocol at every interconnect. With StarFabric, the protocol overhead is only incurred at the

Applied Computing / Spring 2002
ingress and egress to the system. StarFabric moves the data throughout the system in a memory-to-memory fashion.

**Media gateway**

Media gateways are those devices that bridge the existing TDM-based PSTN with an IP-based network. Current media gateway architectures create a number of challenges to the system designer, including trying to support TDM traffic, packet traffic, and control traffic concurrently. To move TDM, control, and packet traffic through a multi-shelf system requires three independent backplanes. If redundancy is also required, the system must support six separate buses on the backplane and support them from chassis-to-chassis.

The media gateway presents an interesting communications system. The communication between the line cards and the DSP card is a processor-to-device model. The communication between the DSP card and the packet WAN link is a processor-to-processor model. The challenge is to find an efficient interconnect system that supports both. StarFabric unifies the multiple traffic planes into a single, high-availability switched interconnect fabric (see Figure 3).

StarFabric supports seven classes of service including the isochronous, asynchronous, and control traffic required in a media gateway. Using a credit-based flow control algorithm, each traffic class can be treated as required. The weights of the credits can be modified to best suit the specific application’s traffic patterns.

To support both processor-to-processor and processor-to-device communications in an efficient manner requires moving to an open switch fabric. However, switch fabrics optimized for packets do not support TDM traffic well. Similarly, those optimized for TDM are inflexible when sending asynchronous packets. Further, interconnect systems optimized for processor-to-processor communications are burdensome in processor-to-device communication applications. The reverse is true as well. The StarFabric architecture was designed from the outset to support both types of communication models. It’s a mezzanine-level switched interconnect technology ideally suited for these systems.

Carriers demand increased levels of scalability. As the number of ports increases by adding additional shelves of line cards, the totality of the shelves must look like a single system to the management system. When a second shelf is added, it will only require redundant switch cards and high-density line cards. As the number of ports exceeds the capacity of the uplink, the car-
rier will only upgrade the WAN card.

Storage
One limitation of IDE drives has been the 18-inch distance limit from the drives to the disk controller due to the ATA interface. StarFabric overcomes this limitation by bringing the CPU’s PCI bus to the disk controller (Figure 4). In fact, the PCI bus can be “remoted” by up to 10 meters. Now, innovated 1U implementations and rack-mount, chassis-based storage arrays can be realized. Any disk drive technology could be used. The only requirement is that the disk controller has a PCI interface.

StarFabric enables innovative system implementations by eliminating the distance limitations of the PCI bus and overcoming the distance limitation of ATA interfaces. With this approach, new system architectures are possible. Figure 5 shows how this topology could be implemented in 1U systems, and Figure 6 shows how this topology could be implemented using a chassis-based system. The benefits of either implementation are that the storage and CPUs can scale independently from each other.

If additional CPU resources are required, single-board computers could be added to any slot in any system. The StarFabric control software would interconnect the storage resources as required.

Conclusion
Switch fabrics provide the basis for developing open-platform-based communications equipment, such as VoIP and media gateways, wireless base stations, and broadband loop carrier equipment. The flexibility and power of open switch-fabric technology, including StarFabric, is also applicable in a wide range of other embedded applications such as storage, distributed and cluster computing, and other applications that require a large amount of data to be moved at gigabit per second speeds. By offering key features such as scalability, high availability, and quality of service in an open standard, while maintaining 100 percent backwards compatibility with existing technologies, StarFabric and the PICMG 2.17 effort enable a logical progression path from today’s bus-based systems to higher performance, switched interconnect architectures.

Greg Whelan is director of product marketing at StarGen. He has 15+ years of technical marketing experience in the telecommunication industry with a focus on broadband access. Greg has driven broadband product strategies at Cayman System, Aware, Cascade Communications, and Analog Devices. At Analog he was a founding member and vice president of the DSL Forum. Greg has a BSEE from Cornell and an MBA from Northeastern.