

CHASSIS DESIGNS ARE SHIFTING FROM PROPRIETARY TO STANDARD BACKPLANES.

Evolving to AdvancedTCA-compliant devices

DESIGNING A NEW SYSTEM BACKPLANE involves many disciplines, including the application-specific functions and generic blocks that are part of any telecom or data system. These generic blocks, such as mechanical design, EMI integrity, backplanes, power supplies, and shelf management, are common to all designs. They do not contribute to a design's differentiation and from an R&D perspective can be considered overhead. On the other hand, system-related features are application- and implementation-specific and enable system vendors to introduce new and unique features that differentiate their products from the competition's.

Designing the common, generic blocks is an expensive undertaking in design resources and specialization. This expense can greatly impact smaller design projects, because it forces companies to allocate significant engineering resources and budget to design subsystems that do not directly increase business success. Designing these generic blocks creates a significant market-entry barrier, because the investment to create a new chassis requires backplanes to support a capacity sufficient for several years. If your company lacks high-volume products and a variety of products to amortize the engineering effort for these generic blocks, you might not need a new backplane design for several years and will have to address what to do with the backplane design team in the interim.

The plethora of proprietary backplane standards is a hindrance to creating interoperable line cards and switches for data-networking and -communications infrastructure equipment. The ATCA (Advanced Telecom Computing Architecture) defines a standard chassis, shelf-management mechanisms, and card form factors, and it offers designers a choice of standard protocols to run over the backplane; it promises to bring COTS (commercial off-the-shelf) interoperability within a developer's reach.

The ATCA standard defines the chassis itself and provides a choice of backplane protocols and guidelines for card format and shelf management. ATCA enables large companies to more easily leverage designs across multiple product families and small companies to compete with large companies, because they can enter the market with a sin-

gle specialized ATCA line card instead of creating an accompanying proprietary backplane and switch fabric. For example, ATCA lets companies move up the food chain, perhaps shifting from pizza-box devices to scalable and higher capacity modular equipment, and avoid the headaches and NRE (nonrecurring-engineering) costs of chassis design.

By standardizing the chassis, ATCA allows design teams to first focus on key features that add value to a design and get to market faster with a concept product. The design team can develop a customized version of the concept product that improves the cost and performance. For smaller companies, the economies of scale for a widely deployed off-the-shelf chassis may result in less expensive boxes than the company could develop for itself. Similarly, when using common hardware architectures, software vendors can leverage IP (intellectual property) across multiple vendors and product lines with the same base software. The low-level drivers can accommodate differentiation in the hardware architectures, and software developers need not wait for

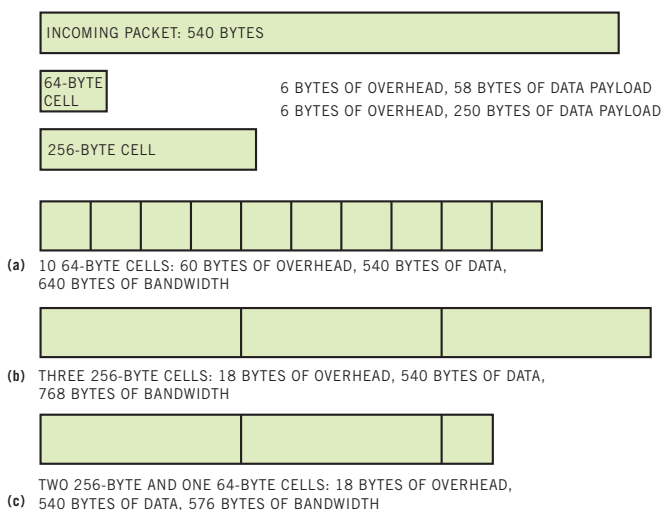


Figure 1 A small, fixed cell fabric consumes additional bandwidth because of cell overhead (a). A large cell size keeps overhead low but potentially leaves more unused bandwidth (b). Multiple cell sizes enable you to break large packets into a combination of cells that reduce overhead as well as limit unused bandwidth (c).

a completed proprietary chassis before beginning development.

MIGRATING TO ATCA

OEMs are migrating from proprietary implementations at a logical level to standard protocols. ATCA proposes standardizing the chassis to allow designers to select the logical protocol that best meets their application's needs. Provisions within the specification enable cards to detect whether the backplane protocol is compatible; if it is not, the card will not connect.

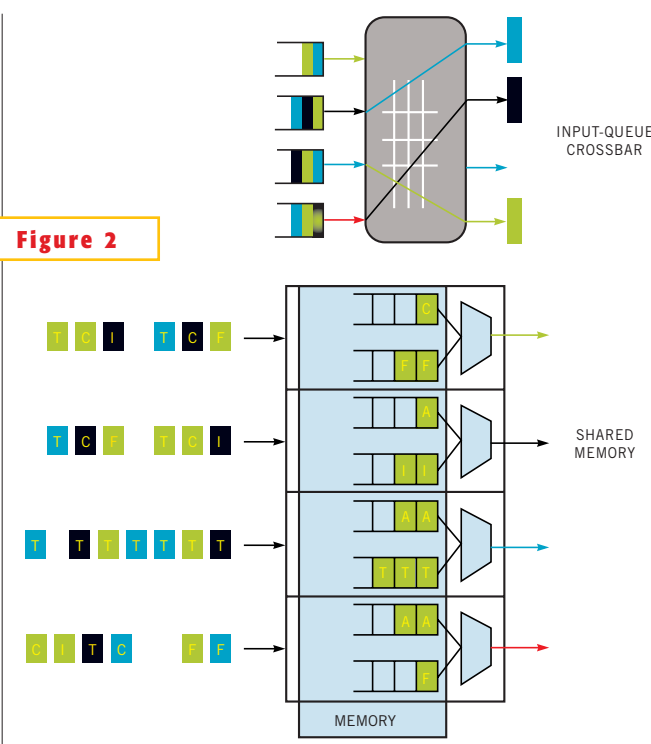
ATCA consists of a base specification for the implementation of a plug-in chassis architecture, in which plug-in cards interconnect via point-to-point differential links each operating at 2.5 Gbps. The base specification supports star and mesh networks based on as much as 10 Gbps between any two cards and provides a detailed specification of all mechanical, thermal,

backplane, and shelf-management characteristics of the chassis. For switch fabrics to scale from 2.5 to 10 Gbps, they must support four SERDES (serializer/deserializer) links for each line card.

The ATCA chassis specification includes profile specifications to support a variety of protocols across its backplane. Protocols such as 802.3ap Ethernet BP (Ethernet over the backplane), Serial RapidIO, Advanced Switching, and InfiniBand are vying for market position, even though some of these standards are new and not yet stable. At this time, it is unclear which of these standards will establish sufficient market share to survive. Therefore, the ATCA standard allows the definition of OEM-specific profiles, or proprietary protocols, and their support in a standard chassis. Because a new ATCA chassis must be able to interoperate with existing equipment, OEMs need to provide ATCA-compatible systems that can support deployed proprietary fabrics.

Although the ATCA spec defines the basic characteristics of backplane traces, differences exist in the functions of each trace, depending upon the protocol in

Figure 2



Managing quality of service and priority can be complex for crossbar fabrics using dynamic cells; a shared-memory architecture offers a less complex approach and improves the delay guarantee, fairness between queues, delay variation, and lossless guarantees.

use. Not all of the protocols are fully specified for the ATCA backplane yet, so designers need to leave headroom for adjustments. The limited availability of fully compatible ATCA devices is another important migration issue that drives the need for testing early ATCA products for interoperability with specific products until the compatibility issues are resolved. Designers of these early devices need to include a level of flexibility, both to adapt to de facto interpretations of ambiguities in the specification and to accommodate the intermediate implementations of necessary but loosely defined functions.

A key advantage of the new ATCA spec is that there are no legacy systems to compromise interoperability or performance. However, developers of the ATCA standard are already discussing the next generation of lane speed (5 to 6.25 Gbps). This change, which is expected to see adoption within the next 12 months if testing verifies that the ATCA chassis can successfully handle such rates, would double chassis capacity.

Implementing the backplane interface through a programmable SERDES allows

a designer to scale designs to higher lane speeds. The shelf manager can test an inserted board for compliance with the system. For example, it can determine whether the two partners on a point-to-point port connection can interoperate using a mechanism called E-Keying. The SERDES programming can initially support a lower data rate, such as 2.5 Gbps, for compatibility with the rest of the system, and when a higher data rate, such as 3.125 Gbps, is required, the system can dynamically reprogram and reconnect the SERDES.

Architectures that employ mezzanine cards can more easily maintain backward compatibility with proprietary systems already deployed in the field. Mezzanine cards can support scaling the I/O that is feeding data to the board by allowing a designer to connect different PHYs at

different data rates to the same base line-card design. By using a modular approach, a designer avoids having to completely redesign a card to support multiple protocols. In some cases, it may be impossible to design a single line card that supports say, both Ethernet and InfiniBand, but with ATCA as a base, all of the wiring stays the same for each card.

THE NEED FOR SPEED

Although ATCA claims to support speeds as high as OC-768 (40 Gbps), there is currently a mismatch between the number of traces to each board and their speeds; driving four traces at 2.5 Gbps each does not provide enough overspeed to achieve 10 Gbps, let alone 40 Gbps. Because the specification supports a mesh-network architecture, a designer can implement 40 Gbps connected to the board as long as the payload is split before connecting to the backplane.

The amount of overspeed that a chassis requires is a function of the architecture and the application. For example, shared-memory architectures are non-blocking when you run noncorrelated (Bernoulli) traffic through them; how-

ever, traffic in the real world is often correlated, resulting in congestion. Features such as multicast and bursting further exacerbate the congestion. Typical fabric overspeed is 1.5 to 2 times, although more efficient fabrics can drop the requirement to a 1.3-times increase.

Overspeed addresses physical-increase link encoding and protocol-driven overhead; it is not a mechanism to solve congestion due to bursts of traffic heading

through the same egress port. Because fabrics are cell-based, support for packet-based protocols entails packet SAR (segmentation and reassembly), which introduces static and dynamic (or fragmentation) overhead. Static overhead comprises the additional overhead appended to each cell in addition to the payload. Dynamic overhead occurs when the packet length is not an integer number of cell payloads, so that the last cell is partially empty and padded with non-data bits.

Protocol-driven overhead occurs because crossbar-switch fabrics depend on a request-grant protocol between the FIDs (fabric-interface devices) and the scheduler. Because crossbar fabrics are input-based, the scheduler must match requests from the various input queues to the availability of the output queues and determine which input queues will receive the right to send across the fabric. Granting requests requires transactions over the backplane, which introduces the need for an additional overspeed, usually on the order of 1.4 times.

Shared-memory fabrics, which are output-based, require no direct communication between input and output queues and therefore do not require protocol-driven overspeeding. The output-buffer queue directs the data flow that reduces port contention. When the output buffer fills, back-pressure mechanisms slow the flow of data to the port without further loading the backplane with requests or grants.

Another source of congestion results from supporting multiple protocols. Multiprotocol support stresses a system with encapsulation, fixed-cell inefficiencies, and the addition of SAR functions. Encapsulation, such as advanced switch-

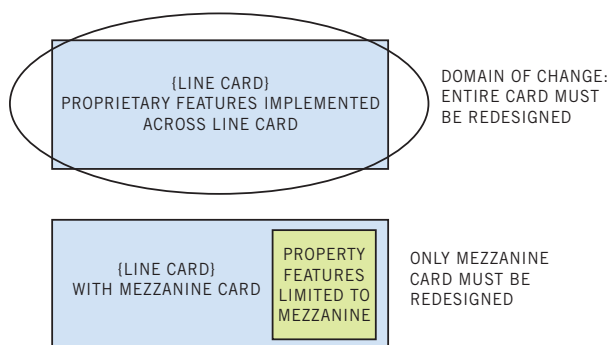


Figure 3

Implementing all proprietary features on a mezzanine card limits the domain of what you need to redesign when you replace proprietary features with standard technology.

ing proposes, frames multiple protocols in a common format, adding a fixed overhead. Using a fixed cell size simplifies fabric architecture and scheduling but results in unused bandwidth, because not all packets are an integer number of cells (Figure 1a). By using a large cell, a designer can minimize static overhead; however, more unused bits may increase the fragmentation overhead (Figure 1b). This situation represents a trade-off between static and dynamic overheads. Small cells introduce large static overhead but incur a low dynamic overhead. Large cells do the opposite.

The most efficient architecture employs a number of cell sizes, enabling the system to break large packets into a combination of cells that reduce overhead as well as the amount of wasted bandwidth (Figure 1c). As a consequence, the FIDs must be able to accommodate the different cell sizes. Finally, adding SAR functions adds latency to a system as well as the cost for the various buffers each port requires.

Dynamic cell size balances between maintaining the flexibility of a general-purpose fabric and optimizing the throughput of a protocol-specific implementation. However, managing QOS (quality of service) and priority becomes prohibitively complex for crossbar fabrics using dynamic cells. Managing QOS at the output queue, such as with a dynamic shared-memory architecture, offers a less complex approach, because whichever port has priority fills the queue. Because the overhead is lower, it also enables a more efficient backplane when using an ATCA platform, facilitating the design of 10-Gbps interfaces for edge and router switches that more closely achieve a true 10-Gbps throughput

(Figure 2). However, the implementation of a shared-memory switch is complex, because a 16×10-Gbps FDX (full-duplex-transmission) switch requires support for a challenging memory with a bandwidth capacity of 320 Gbps.

COMPLIANCE

For high-performance applications, the protocol standards that ATCA supports may be too inefficient. To achieve the required level of performance,

a designer may need to employ proprietary mechanisms that are either always on (potentially breaking interoperability with other ATCA-compliant devices) or can switch from standard to enhanced operating mode. During initial configuration, devices can attempt to negotiate working in enhanced mode with other devices. In this way, a designer can still focus on high-end applications that require performance beyond the current ATCA standard's offering and still take advantage of ATCA cost reductions, such as from COTS management. This ability is another step on the migration path from proprietary to standard protocols and eases the design challenge. The key is balancing the enhancements so that you don't break interoperability in standard mode.

Designers should consider enhancing base cards that differ from mezzanine cards. Conceptually, when a designer limits proprietary technology to a mezzanine card, it becomes easier to manage (Figure 3). As new standards come into play, replacing proprietary technology with standard technology has less impact on the overall system.

ATCA boxes accommodate as much as 200W per card with as much as 3200W per box; this power budget is generous for applications such as those in telecom environments, in which an entire chassis might have an aggregate budget of 1200W. In this respect, the ATCA standard may offer too much tolerance. For ATCA to be successful, vendors need to understand that the 200W limit is not a license to consume power. If you need that much power, you can design a box that requires a power waiver, allowing a line card to consume more than its "fair" share. Doing so may not affect the oper-

ation of a chassis, because some of the cards may draw less power. In any case, it's nice to know that the box will not melt if you power it up to 3200W.

The amount of power a line card consumes may also affect its interoperability. Although line cards might be logically interoperable, the application in which you deploy the chassis may impose its own power limitations. Regulatory issues or other requests that customers impose might constrain an ATCA platform to less than the 200W the standard specifies. The shelf-management system collects the total power consumption per card and calculates whether the combination of cards in the system complies with the configuration. Highly reliable systems employ power-based E-Keying, in which the chassis provides just enough current to a newly plugged-in card to identify the card and how much power it is rated to consume. The line card is activated based on the chassis' available power budget. If the line card will cause the chassis to exceed this budget, the chassis does not activate the card.

One mechanism for managing the power consumption of a card is to use a modular architecture. By implementing I/O and processing resources on mezzanine cards, a designer can gracefully upgrade a card, introducing and scaling resources as needed, as long as the system remains within the 200W constraint. However, a line card that consumes 200W today offers no scalability, because as many as four mezzanine cards contribute to the 200W budget. AMCs (advanced mezzanine cards) are limited to 60W each, and the shelf manager has an important responsibility to guarantee compliance with the total power consumption.

MORE TO COME

As with any emerging standard, the ATCA specification has room to grow, providing mechanisms for additional functions as the market demands them. For example, for line cards to be truly interoperable, they each need to support a standard shelf-management mechanism. However, when a standard is too well-defined, interoperability comes at the expense of product differentiation and high-performance features. Having room to stretch is essential for leading-

edge applications that, by definition, exceed the capabilities of what is currently standard.

Where the interoperability line will fall—and whether it will take the form of guidelines or requirements—depends on each application market and what is actually built. Interoperability can be dictated by a standard or self-regulated; for example, you could build an implementation of a PCI-Express chassis on the ATCA foundation, placing additional restrictions to define shelf-management mechanisms. Only the line cards that support those outlined restrictions interoperate within this chassis.

Currently, network administrators need to buy switches and FIDs from the same vendor to ensure interoperability. With ATCA-standard shelf management, any card will be able to identify itself, its interfaces, and its capabilities to the rest of system. This ability allows multiple cards serving different applications to reside in the same chassis. Therefore, someone who has already invested in proprietary equipment will, without a fork-lift upgrade, be able to implement a new standard application in an ATCA chassis with line cards and a switch fabric that coexists side by side with the proprietary system. ATCA ensures investment protection and simple migration to future technologies in the same chassis.

Silicon vendors will play a key role in enabling interoperability with reference designs. Instead of exchanging proprietary design details early in the design cycle, reference designs let you provide a neutral base target that competing companies can use to facilitate interoperability. For example, line-card vendors can achieve initial I/O interoperability on the network side by working with the same transceiver reference design, implemented in a mezzanine card to match the standard interfaces the carrier uses, and on the fabric side by testing with the same switch reference design.

One of the most expensive aspects of shifting from a proprietary to a standard design is getting designers and implementers up to speed on new protocols. The SPI 4.2 interface specification, for example, runs more than 400 pages. The goal behind ATCA is to provide basic functions, such as interfaces and shelf management, in off-the-shelf products. In

this way, designers need not become experts in standards they are not familiar with. For example, it is much easier to design in a simple memory card than it is to develop a memory interface from scratch.

Simplification is one of the key factors that make ATCA a potential appliance platform for networking systems. Today, available line cards are CPU-based, communicating across the 1-Gbps Ethernet-management bus, and many vendors offer switch cards that interoperate on an ATCA chassis. Plugging multiple cards into the same chassis supports the development of appliances for networking with strong scalability due to their modular design.

In the future, NPU (network-processing-unit) vendors working with capable fabric vendors will provide line cards with a higher-order-of-magnitude throughput, driving fabrics with 10 Gbps per line card. These cards will connect with an off-the-shelf switch card, which in turn will connect to the ATCA standard backplane. Expect such interoperable and ATCA-compliant devices from companies such as EZchip, which uses NP1c NPU, and TeraChip with its TCF switch fabric.

High-speed line cards based on complex NPU architectures, of course, will involve a much steeper learning curve. The degree of complexity depends upon the design. However, ATCA means that understanding the intricacies of interfaces is no longer a key issue. Developers need no longer worry about how to build data pipes; this situation is akin to buying a Pentium CPU board and developing just the software instead of developing the entire CPU board from scratch. □

AUTHOR'S BIOGRAPHY

Micha Zeiger is chief executive officer of TeraChip. Before founding the company, he designed an ATM switch, encryption machines, FFT-based modems, a CDMA wideband wireless-local-loop base station, radar DSP ASICs, and navigation systems. Zeiger has a bachelor's degree in electrical engineering from the Technion, Israel Institute of Technology (Haifa) and a master's in business from Tel Aviv University.

TALK TO US

Post comments via TalkBack at the online version of this article at www.edn.com.