But high-performance embedded applications cannot live on InfiniBand technology alone. Data-intensive applications place demanding requirements on platform hardware, particularly the I/O subsystems. As such, a third-generation (3G) I/O bus is also needed to provide high-bandwidth, low latency connectivity for chip-to-chip, adapter cards, and other interconnects such as InfiniBand technology. There are other switched fabric architectures on the market – Rapid I/O and StarFabric among them – but over time, a de facto or established standard will emerge, driven by compatibility with popular I/O bus architectures.

This article will focus on the InfiniBand switched fabric network architecture, citing its features/benefits for embedded applications, and detailing compatible I/O bus architectures. These I/O buses include PCI Express with a higher bandwidth, and scalable I/O technology that is compatible with the current PCI software environment; VITA-41, for connectivity to legacy VMEbus-based applications; and the Advanced Telecom Computing Architecture (AdvancedTCA) for the next generation of carrier grade communications equipment.

In addition, because InfiniBand technology has gained market acceptance in the HPC (High Performance Computing) arena, it is likely that continuing investments in the core technology will make the technology more robust over time. And as InfiniBand technology gains critical mass, price/performance ratios should continue to improve. In short, this article will show developers why InfiniBand technology is well poised to become the de facto standard switched fabric network architecture for high-performance embedded applications.

Local bus connectivity
For inside-the-box communications, the most commonly used buses have been VMEbus (for embedded applications), and PCI bus (for commercial and embedded applications). The PCI bus is a multi-drop, parallel bus that is close to its practical limits of performance, and as such it cannot be easily scaled up in frequency, or down in voltage. Its synchronously clocked data transfer is signal skew limited, and the signal routing rules are at the limit.

However, recent advances in high-speed, low pin count, point-to-point technologies
offer an attractive alternative for major bandwidth improvements for I/O communications. Originally called Third Generation I/O (3GIO), PCI Express is designed for use in chip-to-chip applications, and is likely to replace the PCI/PCI-X buses over time.

PCI Express is defined as a serial I/O point-to-point interconnect that uses dual simplex differential pairs to communicate. The intent of this serial interconnect is to establish very high bandwidth communication over a few pins, versus low bandwidth communication over many pins. It also preserves customer investments in the PCI/PCI-X bus to facilitate migration.

Celebrating its 20-year anniversary last year, the VMEbus is still going strong, especially in the military, aerospace, industrial control, and instrumentation markets. However, the need for higher performance has shaken up the VMEbus market somewhat. While military/aerospace applications are not screaming for higher bandwidth at the moment, communications and other applications are. The new VITA-41 Specifications for the VX5S backplane allows fabric architectures such as InfiniBand technology to run across the P0 (Port Zero) portion of a new VME64x-type backplane, thereby extending legacy VMEbus-based systems dramatically.

AdvancedTCA is a new bus geared for the telecom and communications market that offers Gigabit/Terabit performance. AdvancedTCA, which is primarily a packet-based, switched-blade architecture, also interfaces with InfiniBand technology and other switch fabric architectures for outside-the-box connections.

**Switch fabric connectivity**

PCI Express, the VMEbus, and AdvancedTCA were designed for internal connections, whereas InfiniBand technology provides true fabric architecture for internal and external networks.

Originally known as System I/O, InfiniBand supports circuit trace, copper wire, and optical fibers. InfiniBand architecture is a combination of Intel’s Next Generation I/O (NGIO), and Future I/O from IBM, HP, and Compaq.

InfiniBand technology uses switched, point-to-point channels similar to those used in mainframes. InfiniBand technology provides a data path from 500 MBps to 6 Gbps between each pair of nodes, and is designed as a true fabric architecture that can extend connections via internal and external networks.

InfiniBand technology is implemented differently with each local bus. For example, a device called the InfiniHost® II Ex from Mellanox can connect directly to a PCI Express 8x (20 Gbps) link, and dual 10 Gbps InfiniBand links. As an additional example, SBS Technologies offers an InfiniBand 4x Dual-Port PMC Host Channel Adapter (HCA) that is ideal for a connection to a SBC in an embedded system. As shown in Figure 2, SBS Technologies IB4X-PMC-2 HCA is based on Mellanox’s InfiniHost silicon and is capable of full wire speed transmissions over InfiniBand links.

With both sides matched in speed, latencies are not expected as the PCI Express-to-InfiniBand interface enables processor-level bandwidth throughout the system. In short, InfiniBand extends compute nodes through the use of PCI Express into an intelligent fabric-based cluster.

InfiniBand is also suited to inside-the-box connectivity. A PCI Express-to-InfiniBand bridge can be implemented on nodes within a chassis such as a single board computer (SBC) with VITA-41 standard connectors. This is useful in applications where two cards need to communicate quickly without interrupting other cards in the same chassis. For example, an application where data acquisition, processing, and display cards in three chassis – one for data acquisition, another for data processing, and a third to display the results – communicate with a traditional system. With InfiniBand on the backplane, the HCA/InfiniBand silicon can handle the networking overhead so the acquisition and processing nodes can concentrate on their own specific tasks. InfiniBand can send the results out-of-the box directly to a user’s display computer.

**InfiniBand features/benefits**

InfiniBand has many features and benefits. With 10 Giga1bits per second full-duplex bandwidth for 4X, 30 Gigabits per second for 12X, and the announcement of future bandwidths up to 120 Gigabits per second, the InfiniBand Architecture is a high-performance interconnect technology that will meet the bandwidth needs of today and well into the future. InfiniBand technology not only provides high speed, but its intelligent fabric offloads the computer’s processing system from complex overhead and housekeeping tasks.
InfiniBand technology accesses memory directly with its Remote Direct Memory Access (RDMA) feature which bypasses the CPU for information transfer, and this is essential when space and computing power are critical.

As such, InfiniBand technology has caught the attention of the HPC market, which uses tens to hundreds of servers organized in clusters that often run on open source operating systems such as Linux. In fact, many big companies have replaced mainframes with HPC clusters linked by InfiniBand technology, thereby achieving a huge improvement in price/performance. The success of InfiniBand technology in a commercial market such as HPC almost guarantees that the price of silicon will continue to come down due to economies of scale, and the embedded market will realize these benefits as well.

The primary component of InfiniBand architecture is the HCA that provides transport services over the switch fabric. The HCA, or more fundamentally, the HCA silicon chip allows the bridge between PCI-X/PCI Express and InfiniBand architecture. Fabric end nodes are either host nodes with HCAs, target nodes with Target Channel Adapters (TCAs), or as SBCs/nodes with integrated InfiniBand ports. As shown in Figure 3, the primary component of InfiniBand is the HCA that provides transport services over the switch fabric. Fabric end nodes are either host nodes with HCAs, or target nodes with Target Channel Adapters (TCAs).

Extensive reliability and data integrity features are built into InfiniBand. The hardware automatically checks for data integrity at every step in the transfer of a packet which guarantees that a destination will reject corrupted data, and that a corrupted operation won’t corrupt adjacent data. In addition, each device in the fabric maintains its own separate clock and power domain, ensuring that a failed device only affects its own attached hardware link and associated queue pairs, and that communications between other devices in the system are unaffected.

At the physical level, the InfiniBand architecture is implemented as a point-to-point packet switched fabric, which provides a basic scaling advantage. The addressing capabilities of the InfiniBand protocol allow as many as 48,000 nodes to be supported in a single subnet, and there are no theoretical limits to the size of a system. It scales up linearly, and performance doesn’t degrade as components are added. InfiniBand architecture provides fault tolerance through redundant components.

The HCAs and TCAs are dual-ported to provide redundant connections if needed for failover. Most of the components are hot pluggable, so if a component fails, it can be swapped out without powering down the chassis. InfiniBand architecture offers security through partitioning, where only members of a partition can see each other while others outside the membership cannot. The partitions provide security transparently while sharing common network components. This security scheme provides the end-to-end access control and authentication services.

For real-time determinism, the switch fabric technology of InfiniBand architecture goes a long way to alleviate resource contention, which guarantees an upper threshold on message latency and jitter. InfiniBand architecture is flexible enough to handle isochronous traffic such as streaming video, while at the same time it accommodates more stringent levels of determinism.

And surprisingly, InfiniBand architecture is more affordable than the other network technologies. It is expected to come in at roughly $300/port for Gigabit network performance at network/telecom market prices.

**Conclusion**

While InfiniBand architecture has caught on with the HPC market in a big way, it remains to be seen if it will become the switch fabric of choice for embedded applications. InfiniBand architecture is compatible with PCI Express at the silicon level (which is likely to be deployed everywhere), not to mention the VXS backplane via the VITA 41.1 Specification (which couples it with legacy embedded applications), as well as AdvancedTCA (the new bus for high-speed communication in the telecom market). As such, InfiniBand technology stands a very good chance to succeed as the standard switched fabric network architecture for high-performance embedded applications.

InfiniHost is a registered trademark of Mellanox. Other brand or product names are registered trademarks or trademarks of the respective holders.

For further information, contact Lori and Steve at:

**SBS Technologies**

Corporate Headquarters
2400 Louisiana Blvd. NE Suite 5-600
Albuquerque, NM 87110
Tel: 800-727-1553 • Fax: 505-875-0400
E-mail: ldunbar@sbs.com
E-mail: scook@sbs.com
Website: www.sbs.com