

Selecting a serial interconnect for distributed computing

By Bill Davies

In next-generation VXS and VPX systems, VME designers have a wealth of serial choices, but only RapidIO brings low latency, scalable bandwidth, and an ecosystem anchored by open standards and an active trade association.

When designing a distributed computing system with more than one processor and shared memory, a high-speed connection is required. Choosing a technology for that high-speed data plane fabric requires a detailed understanding of the features and the useable bandwidth.

Serial interconnect is being used in all embedded market segments because it enables distributed computing by providing a fast connection within each card and across a backplane. Serial interconnect is specified in both VITA 41 (VXS) and VITA 46 (VPX) VME systems, but multiple choices have slowed adoption rates as companies wait for a clear winner to emerge. These serial fabric choices – ranging from Ethernet, InfiniBand, PCI Express, HyperTransport, and RapidIO – narrow significantly when key criteria are applied. Low latency, high bandwidth, and error management are required by VME customers to build distributed computing platforms with the reliability and scalability expected in these applications. Additionally, companies should look for ecosystem momentum, open standards, and an active trade association to ensure the serial interconnect thrives and evolves.

Why distributed computing

The fundamental problem of electronics is how to put more functions in the same space. More functionality results in greater heat generation despite shrinking silicon geometries. As any thermal engineer will attest, it is not the power but *power density* that creates the cooling challenge. So despite much hype about System-on-Chip (SoC) integration, cost-effective products distribute the power-hungry compute nodes throughout.

In addition to lowering system cooling costs, distributed computing enables the use of task-specific computing elements.

Network Processing Units (NPUs), ASICs, and DSPs exist because they are more cost-efficient at certain tasks than the CPU or SoC. Being able to hook them together with a serial fabric provides modularity and scalability through replacement of individual computing elements as needed. Rather than a major upgrade to install the latest SoC, performance can be incrementally improved by adding just some new DSPs, or a faster ASIC on a mezzanine card.

The challenge

The catch to this process is that distributed computing requires a very fast *data pipe* between computing elements or its performance suffers. Serial interconnect is providing that data pipe in multiple embedded applications. VITA specifications have laid out how to build systems using a variety of serial interconnect technologies. (See Figure 1 for a conceptual VXS single board computer showing multiple computing elements such as DSPs, CPUs, and FPGAs). Both VXS and VPX can also use this serial connection through the backplane to additional processing cards.

Serial switches like Tundra's RapidIO Tsi578 or Tsi568A are used to create the point-to-point connections required in serial communication. There is no multidrop bus at these gigahertz speeds. To meet reliability targets, the standard should support connection topologies like dual star or mesh so that a single connection failure does not bring the whole system to a halt.

The interconnect standard also plays a role in achieving high performance. It needs to be designed to ensure fast, efficient, and reliable data transmission. How packets are terminated is also important. Terminating packets includes dealing with things like addressing, error checking, and traffic priority levels and should take as little bandwidth and processing power as possible.

In addition to high performance and reliability, what else does a company need to consider when choosing an interconnect standard? A wise choice supports the required features, has a developed ecosystem of suppliers, is found as a

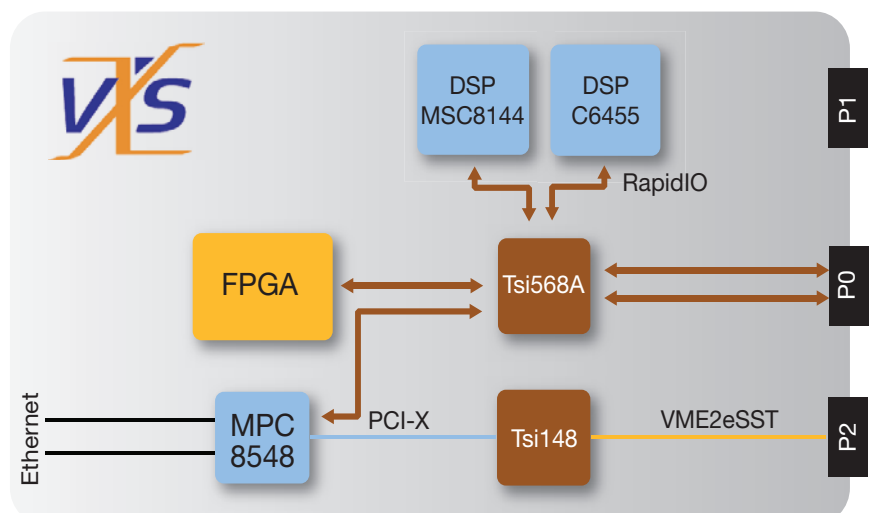


Figure 1

native port on the processing elements thereby eliminating the need for additional conversion or *bridge* chips, has an active road map, and is an open standard. At the top of the list for VME applications are longevity, data reliability, and enablement of distributed processing.

What makes for a high-performance switch?

Low latency and high bandwidth are critical in distributed processing. Simply put, switch latency is measured as the time data takes to traverse the switch. So a low latency value indicates a single fast *data pipe*. Latency varies widely between vendors' silicon and depends on both the design and the type of forwarding mode used. But to prevent traffic congestion and truly claim high performance, the switch must also be nonblocking. A switch can claim low latency on one path, but if it isn't nonblocking, other data paths will be waiting while that one transaction completes.

A company should also look for the lowest power so that the power budget can be spent on processing. Having ports that are independently configurable in both speeds and widths means that power can be minimized by not using ports or bandwidth you don't need. To sum it up, one needs to shop carefully for switches because they are not all the same and there are many variables.

What makes a high-performance serial interconnect protocol?

The interconnect standard affects performance and reliability. Efficiency in the structure of the protocol, efficiency in how errors are checked, the speed of the physical layer, and topologies supported determine a protocol's potential.

Bandwidth is a product of both efficiency and the physical transmission rate.

Efficiency is a ratio of the bytes available for data to the bytes required for overhead (addressing and error checking). The effective bandwidths presented in Table 1 are calculated using efficiency and the data rates available today. How packets are terminated is also important. Termination is essentially taking the header apart so as to read addresses and perform error checking. Gigabit Ethernet (GbE) using TCP/IP can take a significant amount of computing power as only up to Level 2 of the protocol stack is handled in silicon; the rest is dealt with in software. Schemes like UDP are employed in GbE to reduce the overhead. In PCIe and RapidIO, termination is done in the silicon up to Level 3. (See Table 1 for a comparison.)

Reliability in the context of an interconnect standard means that it should support multipoint connection topology – a fabric – and offer advanced error handling features. PCIe is really a host-centric architecture. A PCIe application has a single CPU *mastering* the traffic – called the *root complex*. This leaves the system vulnerable to root complex failure. It also means that mapping distributed memory so it can be globally shared by multiple processors is very difficult in PCIe. Advanced Switching Interconnect (ASI) was developed to address both these issues by creating a fabric using PCIe, which is really meant as a local interconnect. ASI is not compared here because – as stated in the *EE Times* article of July 17, 2006, "Plug Pulled on Comms Interconnect" – ASI has failed to gather enough support from silicon vendors.

RapidIO offers all of the features required to build a VME VXS or VPX system with the greatest bandwidth and reliability.

Items to note include:

- The effective bandwidth includes CRC encoding scheme
- GbE figures assume a UDP implementation (TCP/IP would be worse and Layer 2 slightly better)
- Protocol Data Unit (PDU) = header, data, CRC, and so forth

VME customers require long-term supply

A company cannot risk developing a system with an immature technology in the VME market. Being used as a *native port on the processing elements* indicates an interconnect's viability and also removes the need for bridge chips. Right now only RapidIO, PCIe, and GbE can claim to be native ports on processors and to have multiple vendors selling silicon. Figure 2 shows a maturing RapidIO ecosystem.

Open standards and trade associations also help. Apart from enabling multiple silicon suppliers, they also promote an active roadmap and specifications that will ensure the backwards compatibility needed in VME systems that have a long life. Interoperability is also fostered. Tundra has just launched an interoperability lab – RIOLAB – a state-of-the-art, independent testing facility that provides device interoperability and specification compliance reports.

RapidIO: Tackling today's challenges

Considering all the challenges above, the serial interconnect specification that best addresses these today is the RapidIO standard. RapidIO is found as a native port on many processing elements, has multiple suppliers with silicon available today, has features to support error management

	RapidIO	GbE	PCIe
VITA 46/VPX Spec. Number	VITA 46.3	VITA 46.6	VITA 46.4
VITA 41/VXS Spec. Number	VITA 41.2	VITA 41.3	VITA 41.4
Data rate (per lane)	3.125 Gbaud	1.0 Gbaud	2.5 Gbaud
Effective bandwidth per lane with a 2048 byte PDU	2.4 GBps	1.0 GBps	2.0 GBps
Effective bandwidth per lane with a 256 byte PDU	2.4 GBps	0.8 GBps	1.8 GBps
Effective bandwidth per lane with a 32 byte PDU	1.7 GBps	0.3 GBps	1.1 GBps
Protocol layer terminated in silicon	Up to layer 3	Up to layer 2	Up to layer 3
Supports globally shared memory	Yes	No	No
Messaging support	Up to 4K messages	No	Interrupt & Event Signaling

Note: These figures include coding scheme: SRIO uses 8b/10b

Table 1

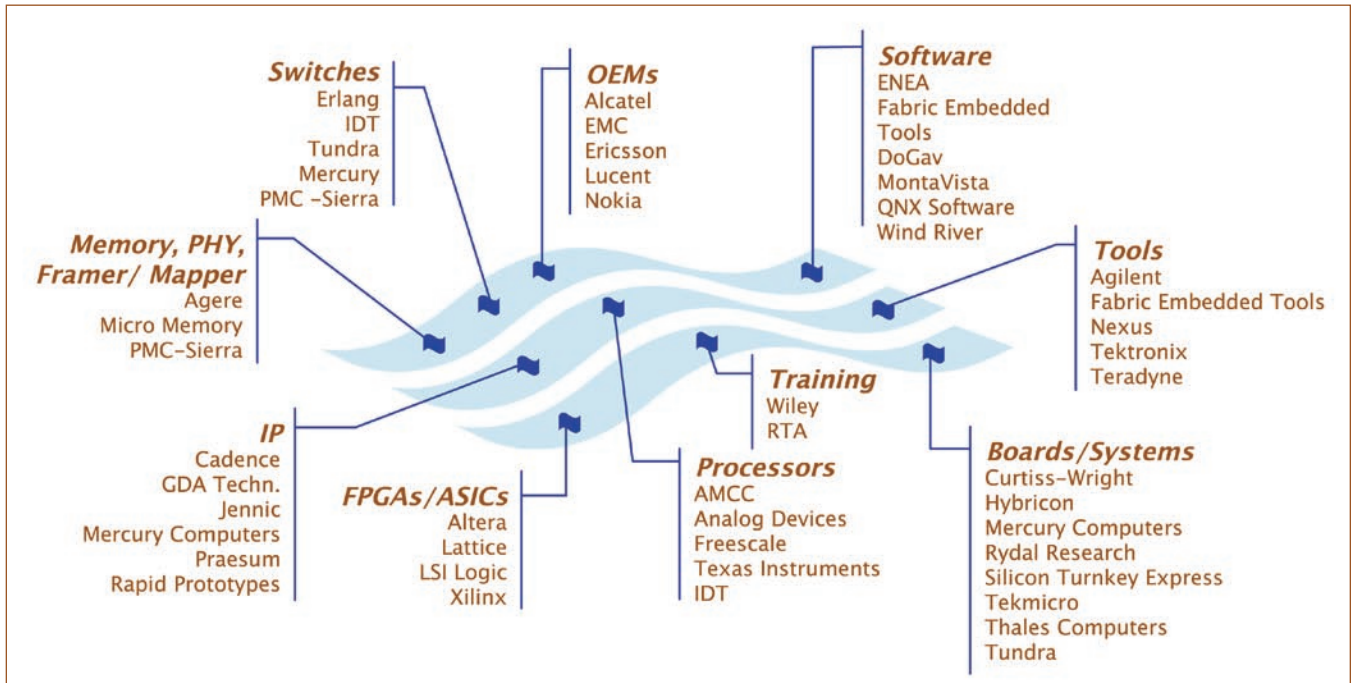


Figure 2

with small overhead, and has the highest bandwidth that enables distributed processing. With 25 years of products behind it, VME has succeeded not just because it is an open standard, but also because it promotes backwards compatibility, has a large supply base, and has continued to evolve over time. Any serial interconnect should offer the same potential. RapidIO technology does just that. Ω



Bill Davies is VME Product Manager at Tundra Semiconductor Corp. and has 20 years of semiconductor experience in engineering and business roles.

Prior to joining Tundra, he worked at Nortel Networks and General Dynamics. Bill has a BAsC in Engineering from the University of Waterloo.

To learn more, contact Bill at:

Tundra Semiconductor Corp.
 606 March Road
 Ottawa, Canada, K2K 2M5
 Tel: 613-592-0714
 E-mail: bill.davies@tundra.com
 Website: www.tundra.com