The Ins and Outs of New Local I/O Trends

Linda Dailey Paulson

As processor performance, memory, networking, storage, and other technologies have improved, vendors have had to develop higher-speed local interconnects to ensure that overall system performance keeps pace.

For example, today's faster processors need to receive instructions and data in a timely manner to process information efficiently. Likewise, today's faster PCs, servers, voice- and data-networking products such as routers and switches, signal processing devices, and other technologies need faster, more functional I/O.

This is particularly the case for machines handling data arriving via today's fast technologies such as multi-Gigabit Ethernet, Serial Attached small computer system interface, Internet SCSI, and Fibre Channel.

New interconnect technologies are especially important for computers processing today's streaming media, games, security processing (including virus checking), video editing, voice recognition, advanced encryption, 3D animation and rendering, and other applications that require high bandwidth and dependable, smoothly flowing data transmissions.

Previous standard local I/O technologies, including PCI (peripheral component interconnect), cannot provide the necessary performance to offer the instant data availability that many

systems need. Users have tried working with multiple PCI buses, but this adds cost, power consumption, and space usage.

Therefore, vendors have upgraded the PCI standard and developed several new approaches for internal expansion buses, and chip-to-chip, chip-to-I/O, chipset component coupling, and chipset-to-I/O connectivity.

NEW APPROACHES

Like PCI, PCI-X (Extended) and PCI-Express provide interconnections within the box, via slots or internal expansion buses.

Rapid I/O and HyperTransport are mezzanine connectors from a system's main processor or memory controller to another I/O, often a PCI bus, within the system, perhaps for an additional chip or component. They also can connect outside the box, such as within a data center.

Meanwhile, InfiniBand, which industry observers at one time thought would replace PCI, is now seen as best for external interconnects between systems.

The new interconnect technologies offer more bandwidth and less latency. All are point-to-point, packet-based approaches and reduce power consumption—particularly important for handheld devices—via lower pin counts.

Lower pin counts also help maintain or even reduce chip size, said John Beaton, an interconnect program director for Intel.

PCI-Express and HyperTransport have serial architectures. PCI-X has a parallel architecture. Rapid I/O has both serial and parallel architectures.

Thus there are new interconnects for working with the traditional parallel technology or the increasingly popular serial approach.

The older parallel-bus technology provides high bandwidth and simplicity and is familiar to many developers. However, it operates over shorter distances, uses more power, and requires more I/O pins.

Serial I/O approaches operate over longer distances and use fewer wires and pins, as well as less power. However, they create more latency and require high-performance silicon and additional serializer/deserializer hardware.

PCI-X

PCI, a shared-memory technology, has been an I/O standard since 1992. Since then, said Karl Walker, chief technology officer of Hewlett-Packard's Industry Standard Server Division, "PCI and its subvariants have been the dominant standard in the industry."

PCI-X, currently on version 2.0, is the dominant I/O peripheral standard in servers. It was approved in September 1999 by the PCI-SIG, a special interest group that administers PCI standards.

Compaq Computer, HP (which has since purchased Compaq), and IBM jointly developed PCI-X to increase I/O performance for high-bandwidth technologies such as Gigabit Ethernet. They designed the technology largely

for server applications, unlike prior PCI versions that were developed for general I/O uses.

Vendors implement PCI-X in various ways, including in chipsets and network cards. Compared to PCI, PCI-X offers better memory read-write performance, faster clocking, and a wider bus, which can move more bits at a time.

PCI-X differs from PCI in that it uses register-to-register signaling, which more accurately resolves signals. This increases the speed at which the I/O reads data and lets a bus have higher clock and data rates.

In newer versions of PCI-X, as well as PCI-Express, the clock is embedded with the signal. This makes the process more efficient and synchronization more accurate than clocking that comes from an outside source.

Currently, PCI lets one 64-bit bus run at 66 MHz and then lets either additional 32-bit buses run at 66 MHz or additional 64-bit buses run at 33 MHz, yielding a maximum data rate of 532 Mbytes per second.

With PCI-X-133, one 64-bit bus runs at 133 MHz and the rest run at 66 MHz, allowing for a maximum data rate of 1.06 Gbytes per second.

For PCI-X 2.0, the PCI-SIG is developing PCI-X-266 and PCI-X-533, yielding maximum data rates of 2.1 and 4.3 Gbytes per second, respectively.

PCI-X also offers advantages other than speed over earlier technologies. For example, the technology is more fault tolerant than PCI. PCI-X can reinitialize a faulty line card or take it offline before it fails.

PCI-X is backward compatible with older PCI versions, although it functions only at the older technologies' slower rates, said Alan Goodrum, an HP staff fellow. Backward compatibility is one of PCI-X's strong points, he said. "It's investment protection."

PCI-EXPRESS

The PCI-SIG designed PCI-Express as a serial, switched PCI technology. Its layered architecture enables connections with copper, optical, or other media.

The PCI-SIG approved PCI-Express in July 2002. The organization and key supporters such as Intel designed PCI-Express, once known as Third Generation I/O (3GIO), for use with serial technologies in various markets, including computing and communications servers, as well as handheld and embedded devices.

Intel's Beaton said, "PCI-Express is taking PCI-X and moving it from the bus to a high-speed switch interconnect." Proponents say the switched approach offers better performance than PCI's traditional shared-bus approach.

The new interconnect technologies offer more bandwidth and less latency.

Unlike PCI, which uses transistor-transistor logic signaling, PCI-Express uses low-voltage differential signaling. LVDS communicates using the voltage difference between two wires rather than the voltage in a single wire, as is the case with TTL. LVDS incurs less electrical interference, so the system can more accurately identify signals.

PCI-Express currently moves data up to 250 Mbytes per second in each direction per lane (a lane is a pair of two wires, one for transmitting and one for receiving signals), yielding 16 Gbytes per second in a typical 32-lane configuration. Users can add up to 32 more lanes to provide more bandwidth.

PCI-Express buses initially will run at 2.5 GHz. Later implementations could reach 6.25 GHz, with the potential to scale beyond 10 GHz in conjunction with fiber-based technologies.

PCI-SIG chair Tony Pierce, technical evangelist for Microsoft's Hardware Strategy Group, said PCI-Express also offers aggressive power management and hot-plug capabilities, which enhance system availability by allowing the insertion of PCI adapter cards without rebooting.

In addition, Beaton noted, PCI-Express has a quality-of-service option that lets users assign different transmissions to various virtual channels, giving priority to some of the channels.

Intel plans to release the first PCI-Express chips and chipsets later this year.

HYPERTRANSPORT

The HyperTransport Technology Consortium manages the HyperTransport standard, initially developed by AMD and the now-defunct API Networks, both semiconductor companies.

HyperTransport, shown in Figure 1, sends packet-based data and command information over fast unidirectional links. Proponents say this has two main advantages over shared-bus technologies such as PCI and PCI-X: Unidirectional links permit better signal integrity at high speeds, and they enable faster data transfers with low-power signals.

HyperTransport offers clock speeds of up to 800 MHz, double-data-rate signaling, and 32-bit data-transfer technology. The technology is thus quite fast, with a theoretical maximum aggregate data rate of 12.8 Gbytes per second.

Because it is optimized for high-speed data transfer, HyperTransport is particularly good at connecting high-speed components such as processors and closely coupled chipset elements.

Also, HyperTransport links of different widths can connect. For example, a 2-bit-wide link can connect to an 8-bit-wide link, which lets users daisy-chain parts of a system or application. This allows companies to mix and match communications and embedded products for use in applications.

In addition, HyperTransport provides scalable bandwidth where needed.

HyperTransport 1.10, slated for release in the near future, will offer a series of features for networking applications, said Brian Holden, chair of the HyperTransport Technology Consor-

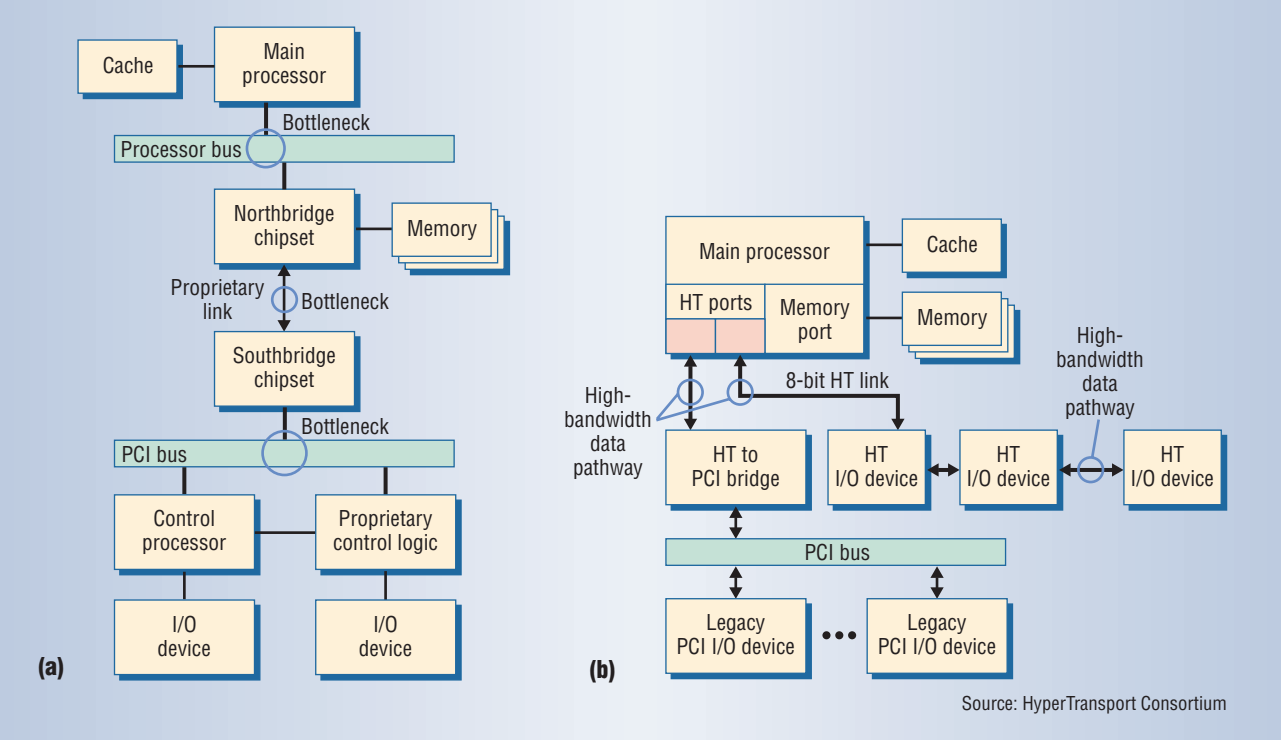

Figure 1. (a) Traditional PC architecture has multiple layers of older PCI bus technology that can create data-transfer bottlenecks, particularly as processor performance, memory, and other technologies have improved and as computers run more-demanding applications. (b) HyperTransport (HT) targets bottlenecks by streamlining the interconnect structure and providing high-speed links. The technology also works with legacy PCI buses.

tium's Technical Working Group. These include the capacity to handle packets natively and network extensions that let HyperTransport bridge to other I/O types.

RAPID I/O

Motorola and Mercury Computer Systems jointly developed Rapid I/O between 1998 and 2000 for embedded systems, primarily for the networking and communications markets. The technology is typically implemented in processors, controllers, switches, bridges, field-programmable gate arrays, and application-specific integrated circuits.

Several companies formed the RapidIO Trade Association in early 2001 to handle the standard's development, maintenance, and evolution, according to the association's marketing chair Kalpesh Gala, IBM's PowerPC strategic marketing manager.

Explained association president Sam Fuller, "Rapid I/O is a chip-to-chip and board-to-board interconnect."

Rapid I/O offers a switched architecture that increases data rates and reduces latency. The technology also uses LVDS.

With Rapid I/O, all processing is done in a CPU or some other type of hardware. This eliminates the need to write I/O software, run the software on a CPU (which slows down the processor), or spend extra money on a dedicated processor to run it. This reduces latency.

And because Rapid I/O runs only in hardware, the technology, unlike the other new local I/Os, operates with no impact on the OS, is transparent to applications, and thus doesn't need special device drivers. This makes the system simpler and more efficient.

Rapid I/O can bundle multiple differential links into a single link for use in a task. This lets the technology support bandwidths of up to 60 Gbits per second for each direction of a bundled bidirectional link. "Currently available systems based on Rapid I/O already offer backplanes with [aggregate] bandwidth of over 480 Gbits per second," Fuller said.

A parallel version of Rapid I/O provides the speed and low latency necessary for high-performance chip and system connectivity. The parallel approach is best for interconnectivity over short distances, such as between modules and carrier boards.

There is also a serial version for such purposes as serial backplane communications and connectivity in digital-signal-processor farms, explained Dan Bouvier, architecture manager for the Motorola Somerset Design Center. Serial Rapid I/O is designed for applications that require longer transmission distances, he said.

INFINIBAND

InfiniBand, initially called System I/O, was born when two projects—Future I/O and Next Generation I/O—merged in 1999.

Industry observers initially predicted InfiniBand would replace PCI, particularly when PCI was seen as a major system bottleneck with few alternatives, said Ramon Acosta, chief technical officer for InfiniSwitch, a networking hardware and software vendor.

However, said HP's Goodrum, InfiniBand was optimized for use in networks and is unnecessarily complex for a local I/O technology. Proponents say it is better used as an external system interconnect, particularly in data centers or between external networking devices such as those used for storage.

Each of the local I/O technologies may survive by cultivating a market niche. "I see them all being in the marketplace," said analyst Jonathan Eunice with Illuminata, a market research firm. "As long as they have a different value proposition, they will survive."

For example, he said, HyperTransport's niche could be as the mezzanine bus for AMD systems. PCI-X, on the other hand, is popular in servers but hasn't been widely adopted for client PCs, which don't need the higher bandwidth, according to PCI-SIG Chair Pierce.

PCI-Express could be used in many scenarios but may not be supported in servers, at least initially, because of PCI-X's popularity. PCI-Express might replace accelerated graphics port technology in graphics chips.

Meanwhile, proponents are improving the new local I/O technologies.

In the process, users are trying to figure out whether they can afford the new technologies right away, said Bert McComas, founder and principal analyst for InQuest Market Research. "A lot of people are going to have to work hard to make this cost-effective." ■

Linda Dailey Paulson is a freelance writer based in Ventura, California. Contact her at ldpaulson@yahoo.com.

Editor: Lee Garber, *Computer*, 10662 Los Vaqueros Circle, PO Box 3014, Los Alamitos, CA 90720-1314; l.garber@computer.org