LA-UR- 96-1246

CONF-9606166--2

Title: HIPPI - What Is It, Where Did It Come From, and Where Is It Going?

Author(s): Don Tolmie

Submitted to: 10th Annual International Conference on High Performance Computers, June 5-7, 1996, Ottawa, Canada

Los Alamos
NATIONAL LABORATORY

HIPPI - What Is It, Where Did It Come From, and Where Is It Going

By: Don Tolmie

Los Alamos National Laboratory

Presented at:

10th Annual International Conference on High Performance Computers (HPCS 96)

5 June 1996, Ottawa, Canada

ABSTRACT

The first commercial High-Performance Parallel Interface (HIPPI) equipment was delivered in 1988, and HIPPI is the current interface of choice for supercomputers and other high end equipment. HIPPI currently supports data rates of 800 and 1600 Mbit/s. The evolution of HIPPI will be described, including some of the design choices made. Comparisons to the Fibre Channel, ATM, and SCI competing technologies will be made, describing where the author feels each came from, and the strengths and weaknesses of each. The HIPPI standards committee is now working on a new physical layer interface supporting data transfer rates of 6.4 Gbit/s, an increase of eight times the current common HIPPI rate. The design objectives will be discussed, and the new interface will be described.

1 - INTRODUCTION

As computers get faster and more powerful, the networks that interconnect them must also increase in speed to keep a computer from going idle while data is being transferred to or from the computer. Not only must the networks be fast, but there may also be requirements for low latency from the sender to the receiver, for small as well as large messages, for long distances, or for continuous rate transmission such as voice or video. This paper examines interconnection technologies that operate at or near the gigabit/s rate, and that are proposed for use in local area networks (LANs).

The trend is for communications bandwidths to increase as computer power increases. It wasn't many years ago that 9600 baud (approximately 8 Kbit/s of data) was considered high speed. Today Ethernet runs at 10 and 100 Mbit/s, Fibre Distributed Data Interface (FDDI) at 100 Mbit/s, the High-Performance Parallel Interface (HIPPI) at 800 Mbit/s, and the Scalable Coherent Interface (SCI) at 8000 Mbit/s.

Often a simple name on an interface does not give a very good idea of what application area the interface was intended for. One way is to look at the background of the people who designed the interface--what sorts of interfaces have they developed in the past is a good clue of what the new interface would be good at. The interface world can be divided into five separate cultures. These are listed below, followed by examples.

- Backplane

    VME, Futurebus +, SCI

- Peripheral I/O Channel

    SCSI, IBM Block Mux, IPI, Fibre Channel

- Local Area Network (LAN)

Ethernet, FDDI, HIPPI

- Metropolitan Area Network (MAN)

FDDI, IEEE 802.6 (DQDB)

- Wide Area Network (WAN)

Telecommunications, X.25, Internet, SONET, ATM

The cultures can be differentiated by the problems they are trying to solve and the goals used to approach the problem. They can also be categorized by "Control" and "Trust". For example, in a backplane environment, a single user has complete control, and gives complete trust since usually the card being installed can read or write anywhere in the system. Compare this to the wide area telecommunications network--here a single user has very little control (does your cross-country phone conversation go through St. Louis, Chicago, or Dallas? And do you care?), and shows the lack of trust by using protocols with lots of checking and firewalls.

Another example is charging and data reliability. On a backplane, Peripheral I/O channel, or LAN, once you buy the equipment the vendor does not care if you use it at 1% or 90% of the available bandwidth. But in these environments the data must be delivered correctly, or if an error occurs, this should be a very rare exception and the error should be flagged. In contrast, the end user does not directly buy telecommunications network central office equipment, but instead rents bandwidth. Hence, the goal for the telecommunications providers is to keep the channels as full of data as possible since this is where they get their revenue. To achieve the full telecommunications bandwidth, the providers may oversubscribe the links, depending upon statistics over a large user population to avoid dropping data. The telecommunications WANs were developed primarily for voice traffic where dropping small amounts of voice data is not catastrophic; but dropping packet data is catastrophic.

If an interface is developed for a particular "culture", it is not unrealistic to also use the interface for a different culture, but it may not be optimized for that culture. It is highly unlikely that an interface developed for a backplane would be appropriate for a WAN, and vise versa. If someone touts a particular interface as the "best for all applications," you should be very cautious--it is probably wishful thinking.

2 - HIGH-PERFORMANCE PARALLEL INTERFACE (HIPPI)

The High-Performance Parallel Interface (HIPPI) was developed in American National Standards Institute (ANSI) Task Group X3T9.3 (recently re-named Technical Committee X3T11). Basically HIPPI was developed as a "fire hose" for moving data, with an emphasis on simplicity. Also in keeping with the simplicity goal, no new technology development or new silicon was required. Because of the simplicity of the design and the economics of using existing components, products incorporating HIPPI interfaces were quickly developed, marketed, and installed starting in the late 1980s. [1,2] These products and their second-generation successors--for computers, peripherals, and networking equipment--are now available from some 65 vendors.

The HIPPI physical layer specifies an 800 or 1600 Mbit/s (100 or 200 MByte/s) simplex interface for distances of up to 25 meters using copper twisted-pair cables. The 800 and 1600 Mbit/s variants are called HIPPI-800 and HIPPI-1600 respectively. There is much more HIPPI-800 hardware in use than HIPPI-1600 hardware.

Other HIPPI documents define the packet format, mappings for upper-layer protocols, control of physical layer switches, fiber-optic extenders, a mapping to ATM, definition of IP and ARP on HIPPI, and some management protocols. [3]

HIPPI came from the high-speed local network culture. Attempts to include features to better support direct disk and tape I/O were expressly omitted. HIPPI is intended as a memory-to-memory interface, and is used in this mode even with high-end stripped disk and tape systems. Most HIPPI installations use crossbar switches to interconnect end systems using point-to-point HIPPI links. A good description of the HIPPI signaling, features, formats, and headers is contained in reference. [4]

Figure 1 shows the HIPPI framing hierarchy. It starts with a connection--the Source puts an address (called the I-Field) on the data bus and raises the Request line. If the destination can accept the call, it raises the Connect line. This looks a lot like dialing a telephone. Note that the connections are simplex, that is, you can transfer data from the Source to Destination only. If you need full-duplex, then separate links can be installed in each direction.

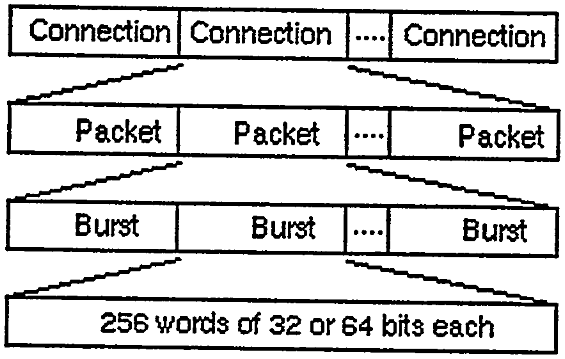

Figure 1 - HIPPI Logical Framing Hierarchy

Once a connection has been made, then packets can be sent. A packet is made up of bursts, with each burst made up of 1 to 256 words of 32 or 64 bits each. An 800 Mbit/s HIPPI uses 32-bit words; a 1600 Mbit/s HIPPI uses 64-bit words. Within a burst the words are sent every 40 ns. There can be idle time between bursts or packets, for example if the Source cannot supply a continuous stream at 800 Mbit/s.

The HIPPI physical layer uses copper twisted-pair cables for simplicity and low cost. A 32-bit parallel data path provides the 800 Mbit/s speed, and a 64-bit parallel data path provides the 1600 Mbit/s speed (using two cables instead of one for the 800 Mbit/s speed). Byte parity, and a longitudinal burst checksum provide physical level error checking. The control signals use individual wires and simple signaling sequences. A synchronous 25 MHz (40 ns) clock is used to strobe the data and control signals. The committee would have liked to use fiber optics rather than parallel copper cables, but at that time (1988) felt that the optical parts were not mature enough, needed an immediate solution, and hence settled for copper cables.

The documents defining HIPPI, and the nomenclature of the standards document, include (1) the physical layer (HIPPI-PH, ANSI X3.183-1991), (2) a framing protocol that defines the packet format (HIPPI-FP, ANSI X3.210-1992), (3) a mapping to IEEE 802.2 to support communications protocols such as TCP/IP over HIPPI (HIPPI-LE, ANSI X3.218-1993), (4) a definition for physical layer crossbar switches (HIPPI-SC, ANSI X3.222-1993), a fiber-optic extender for distances up to 10 kilometer (HIPPI-Serial, X3.xxx-199x), and a mapping to ATM (HIPPI-ATM, X3.xxx.199x). The ANSI documents are also being processed as international standards.

Additional documents being developed in ANSI X3T11 include (1) mappings to the IPI-3 generic command sets for magnetic disks and tapes, (2) a mapping allowing upper-layer HIPPI protocols to use a Fibre Channel physical interface, (3) a mapping between HIPPI and ATM, (4) striping an upper-layer-protocol's large file over multiple HIPPI physical interfaces, and (5) converting Serial-HIPPI for use as a native workstation interface and processing it as an ANSI standard.

2.1 - HIPPI Advantages

HIPPI is the current interface of choice, largely because it was the first standard at close to the gigabit speed. It came to fruition quickly because of a "keep it simple" goal, and a well-focused direction in the standards committee that avoided adding lots of bells and whistles. Some of the advantages of HIPPI include:

- It is simple, elegant, and easy to understand.

- It has a good physical level flow control. The flow control even works with very long links by the addition of extra buffering at the receivers (approximately 1 kilobyte per kilometer of distance).

- A good tester was developed early on which allowed vendors to test implementations in-house so that interconnection with other vendor's equipment was usually plug-and-play.

- A variety of products with HIPPI interfaces from a number of vendors currently exist. Many are second or third generation designs, incorporating improvements from earlier designs.

- HIPPI crossbar switches are available from several vendors.

- HIPPI-specific integrated circuits are available. Even so, some vendors find that small scale integration parts are more suitable due to the simplicity of the physical interface and limitations of the HIPPI-specific ICs.

- HIPPI to SONET adapters are available for very long distance links using telephone network facilities.

2.2 - HIPPI Shortcomings

HIPPI is not without limitations and shortcomings. Perceived shortcomings include:

- It is not a mass-market item, the number of applications that require the bandwidths are not that numerous. Hence the price is higher. It is questionable whether competing gigabit/s technologies, for example, Fibre Channel or ATM would be any cheaper.

- HIPPI does not support speeds slower than 800 Mbit/s. Slower speeds might help make it more of a mass market item.

- HIPPI does not support multiplexing. Since addressing information is only available at the connection level, it is not possible to intermix packets or bursts from different streams. Hence, if you transfer a megabyte over a HIPPI channel, or through a HIPPI switch, as a single entity then it will take at least 10 milliseconds. During this time the channel cannot be used for any other communications unless the large single entity is broken into smaller pieces which are sent as separate connections.

- HIPPI does not support time-dependent or isochronous data.

- The HIPPI specification limits the distance to 25 meters (82 feet) with copper twisted-pair cable. HIPPI-Serial defines a fiber-optic version usable for distances up to 10 kilometers.

- The copper-cable connector is large and somewhat fragile.

3 - FIBRE CHANNEL

Fibre Channel is an emerging standard, also from the ANSI X3T11 Technical Committee. [5] The Fibre Channel work started in 1988, one year after HIPPI started. The first Fibre Channel documents are just now being completed, and the first products being delivered. Fibre Channel supports burst data rates of 100, 200, 400, and 800 Mbit/s. As the name implies, it is based on serial transmission over fiber optics, whereas HIPPI was based on parallel transmission over copper wires.

You may see Fibre Channel referred to with different rates, for example, 133, 266, 531 or 1062.5 Mbit/s. These rates are the serial stream signaling rates that include the 8B/10B encoding and other overhead. The corresponding rates for the user data portion of the serial stream are 100, 200, 400 and 800 Mbit/s respectively.

The Fibre Channel Physical and Signaling Interface (FC-PH) document is an approved standard, and work on follow-on enhancement standards is underway now. The FC-PH standard is available from ANSI or Global Engineering as ANSI X3.230-1994. [3] Included in FC-PH are:

- FC-0, Physical media dependent

- FC-1, 8B/10B code assignments

- FC-2, Frame format and signaling protocol

The physical media (FC-0) portion of FC-PH includes:

- Media speeds of 133, 266, 531 or 1062.5 Mbaud,

- Physical media including single-mode fiber, multimode fiber, coaxial cable, and twisted-pair cables,

- Distance and speed specifications for the different media, and

- Open fiber safety circuits for short-wave lasers.

There are 10 optical variants, and four copper variants specified. The short-wave lasers are interesting. They are the same lasers that are used in compact disk players, and hence have large production volumes and low prices. A disadvantage of early short-wave CD lasers was that they emitted at 780 nm where the human eye is quite susceptible. Hence, extra safety circuitry was needed to shut off the laser if the cable path is opened. Newer short-wave laser components operate at longer wavelengths, for example, 980 nm, and with power levels that do not require the extra safety circuitry--hence at lower costs.

The 8B/10B (FC-1) portion of FC-PH specifies a very clever code that uses 10 serial bits to represent 8 bits of data. The 8B/10B code provides dc balance, that is, an equal number of 1's and 0's, and some error checking to detect bit errors in the serial stream. Since only 256 combinations, out of the 1024 possible combinations, are needed for data, the other combinations are available for control functions. Some additional error checking is provided since receiving an unused combination signifies a transmission error. In addition, the Fibre Channel transmission code is organized so that byte and word synchronization can occur very quickly. All of the control flags are four bytes long to aid in acquiring and keeping word synchronization. Integrated circuits implementing the 8B/10B portion of Fibre Channel are available now. Even though the 8B/10B code set is only a small part of Fibre Channel FC-PH, some people using the 8B/10B code set chips by themselves are claiming conformance to Fibre Channel. This is somewhat true, but can be misleading--it is too bad that the 8B/10B code set portion of the Fibre Channel FC-PH document was not broken out as a separate standard.

3.1 - Fibre Channel FC-2 Building Blocks

The Fibre Channel FC-2 building blocks are:

- Frame - the indivisible unit transferred between end points. The Frame format is shown in figure 2.

- Sequence - one or more related data frames transmitted unidirectionally between end points, with corresponding control frames, if applicable, in response.

- Exchange - one or more non-concurrent Sequences, either unidirectional or bi-directional, between the same end points.

On a single physical link, the Frames from multiple Sequences, or multiple Exchanges, can be multiplexed, that is, the Frames of Exchange 1 can be intermixed with the frames of Exchange 2.

The Frame Header contains 24-bit source and destination addresses, frame counts, control bits, and identifiers. Identifiers include a Sequence Identifier, and identifiers for the Exchange as seen by the transmitter and as seen by the receiver. These identifiers, plus an optional Association Header, may be used to direct the data to a specific process or user memory area--helping to avoid multiple data copies and increase efficiency.

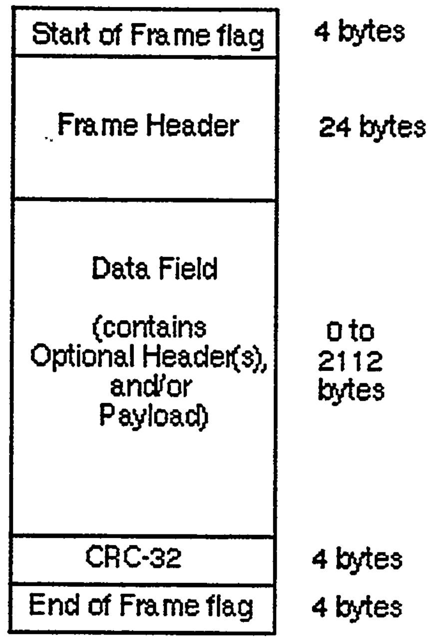

Figure 2 - Fibre Channel, FC-PH, Frame Format

The Data Field contains up to 2112 bytes, consisting of optional headers and/or user data. The ordering is optional headers followed by user data. Optional headers are specified for:

- Expiration timer and security field

- Network header containing other addresses, such as 48-bit MAC addresses.

- Association identifiers

- Device specific header

Other Fibre Channel documents under development include:

- FC-EP, enhanced physical layer with support for isochronous, stripped physical layers (for example, running three FC-PH physical layers in parallel for three times the bandwidth on a single transfer), and other things left out of FC-PH.

- FC-IG, implementation guide with state diagrams for FC-PH and a collection of folklore and helpful hints

- FC-SB, mapping to single-byte command code sets, that is, IBM Block Mux command sets

- FC-FP, mapping to HIPPI upper-layer protocols

- FC-LE, mapping to IEEE 802.2 for support of communications protocols such as TCP/IP

- SCSI-FCP, mapping for SCSI protocols to use a Fibre Channel physical layer for higher speed SCSI devices

- SCSI-GPP, mapping for SCSI generic packetized protocols

- IPI-3 Disk and Tape, revisions to the existing IPI-3 standards to include running over the Fibre Channel physical layer

- FC-FG, fabric generic requirements

- FC-XS, crosspoint switch fabric

- FC-AL, arbitrated loop

- FC-DF, distributed fabric

3.2 - Fibre Channel Advantages

Fibre Channel came from the "mainframe I/O channel" culture, and it should provide an excellent solution for that application. Fibre Channel's success at penetrating the LAN environment remains to be seen. Fibre Channel is considerably more complex than HIPPI, but it also includes many more features. It will be interesting to see if this extensive set of Fibre Channel features turns out to be a boon or a bane. Advantages of Fibre Channel include:

- Very versatile; can do almost anything.

- Supports multiplexing of 2 kilobyte frames of different information transfers since addressing is included in each frame.

- Supports dedicated switched circuits.

- Supports datagram service, that is, best-effort transfers without acknowledgments.

- Supports a broad range of speeds with common integrated circuits.

- Defines a variety of interconnection fabrics, including (1) a crossbar for highest throughput, and highest cost, (2) distributed switching elements for maximum flexibility and ease of growth, or (3) arbitrated loop (a ring architecture with no extra fabric elements required) for lowest cost and lowest performance. Several commercial storage products are currently using the arbitrated loop (FC-AL), with data transfer rates of up to 100 MByte/s. Since the distances spanned in these cabinet-based systems are small, most of the interconnections are copper-based, further reducing costs.

- The switch definitions allow for easy mixing of speeds in a single system. For example, a switch can simultaneously interconnect 200 and 800 Mbit/s end nodes. As you may expect, speed mixing is not supported on the arbitrated loop.

- Fibre Channel may support time-dependent isochronous data, for example, voice or video, in future releases. Striping across multiple physical channels is also planned as a future enhancement.

3.3 - Fibre Channel Shortcomings

In trying to be "all-things-for-all-people" Fibre Channel included an extensive set of options, which some people label as "bells and whistles". Only the future will tell if this large option set made Fibre Channel stronger by being useful for a large set of applications. The large set of options could also be a detriment, resulting in vendors having difficulty making interoperable products, that is, the set of options used by one vendor are not compatible with another vendor's equipment. Fibre Channel may have tried to do too much in one interface. Other perceived shortcomings of Fibre Channel include:

- The development process has taken a long time; Fibre Channel may miss its window of opportunity.

- Integrated circuits supporting Fibre Channel are just becoming available, and they may be made obsolete by later changes in the specification.

- Fibre Channel does not currently support time-dependent, that is, isochronous, data. There are plans within the ANSI committee to add this support, but when it will be developed, and when it will be available in integrated circuits, are open questions.

- All of the options and capabilities resulted in a specification that is quite difficult to read and understand. The complexity will also make interfaces difficult to implement, check out, and verify against other vendor's interfaces.

- Fibre Channel is in a race with ATM, and if Fibre Channel products are not forthcoming soon, then ATM may win by default.

4 - ASYNCHRONOUS TRANSFER MODE (ATM)

In this context, ATM stands for Asynchronous Transfer Mode, not automatic teller machine. ATM came from the telecommunications community, and defines a protocol for sending information in 53-byte cells. [6,7]

Note that ATM is not a physical level interface. SONET (Synchronous Optical NETwork) is the physical layer interface most often mentioned with ATM. SONET is a point-to-point interface supporting data rates from 51 Mbit/s to gigabits per second. [8] SONET does not support switching by the end users, ATM will provide this function. Hence, SONET is comparable to a leased line, and ATM to a dial-up connection. The SONET speeds on optical interfaces are designated as OC-n, where the serial speed on the link is about 51.8 Mbit/s times n, that is, OC-3 is about 155 Mbit/s, OC-48 is about 2.4 Gbit/s. As with Fibre Channel, these rates include overhead; the actual user data rates are about 75% to 85% of the signaling rates.

The 53-byte ATM cell was designed for carrying many separate voice traffic connections over a single physical media. The ATM cell is composed of a 5-byte header with routing, control, and checking information, and a 48-byte payload. Adaptation layers, called AAL1 through AAL5, define the nature of the information in the payload. For example, AAL1 is intended for constant bit rate data, for example, voice or video, while AAL5 is intended for packet data that has no specific timing requirements.

ATM standards are being developed in ANSI, in ISO, and in the ATM Forum.

4.1 - ATM Advantages

ATM was designed for wide area telecommunications networks, but there is also a lot of interest in using ATM technology in LANs. There does not seem to be much activity on using ATM as a native storage system interface. Some of the advantages of ATM-based LANs would include:

- ATM already has good support for mixing time-dependent data, for example, voice and video, with packet data.

- ATM is independent of the underlying physical media, but is most often mentioned in conjunction with SONET. This combination supports a wide range of speeds from megabits per second to gigabits per second.

- It is easy to mix equipment with differing speed interfaces, for example, OC-3 and OC-12, in the same system.

- By using the same technology as the telecommunications industry, larger volumes of common components should result in lower prices.

- Bridging between LANs using ATM and wide area networks (which also use ATM) should be simpler than converting between two dissimilar standards.

- WANs presently have extensive network management tools and these tools may be available in an ATM based LAN.

- There is a lot of interest and momentum behind the ATM work--if it fails it will not be due to the lack of talented people working on it, or for lack of effort.

4.2 - ATM Shortcomings

ATM was not specifically designed for LAN usage, and hence has some shortcomings when used in that environment. Perceived shortcomings include:

- Vendors are building and delivering products before the standards and problems have been solved. This is largely a result of "over hyping". If too many troubles or delays occur, then there may be a backlash against ATM.

- The ATM flow control allows cells to be dropped when congestion occurs. Which this may not be a serious problem in wide area network with voice traffic, it can cause serious lost data problems in LANs.

- The loss or corruption of a single 53-byte ATM cell may result in the re-transmission of a much larger entity, for example, a 16 KByte packet.

- The early ATM equipment for LAN usage supported only Permanent Virtual Circuits (PVCs). This required circuits to be dedicated on the basis of "might be needed sometime." Support for Switched Virtual Circuits (SVCs) is being developed, and is starting to appear in commercial equipment.

- Setting up an SVC may take a fairly long time, for example, milliseconds.

- The most common speeds supported today with ATM seems to be 155 Mbit/s. There is some equipment supporting 622 Mbit/s starting to appear, but higher speeds have yet to be tapped. The next higher speed used in the wide area networks will be 2.4 Gbit/s; it remains to be seen if this speed or an intermediate speed is used in LANs. Hence, true gigabit speeds from ATM are not available now, and probably will not be available in the near future.

5 - SCALABLE COHERENT INTERFACE (SCI)

The Scalable Coherent Interface (SCI) came from the backplane culture, and was developed in the IEEE. [9]

The basic document is IEEE Std P1596-1992; other companion standards are under development as IEEE P1596.n. In the name, scalable means that it is intended to interconnect several to many processors. Cache coherence is a major thrust of SCI, and is something that is very useful in parallel processor systems.

The SCI physical layer consists of unidirectional links--often interconnected in a ring architecture. The links may also be used with switches for higher-performance systems. The links use a 16-bit-wide parallel path, transfer at 8000 Mbit/s (1 GByte/s), and use differential ECL signals. A serial fiber extension operates at one-eighth the speed, using the same serial chipset as is used in HIPPI-Serial. The logical packet interface has a 16-byte header on each packet, and each packet can also carry 0, 16, 64, or 256 bytes of user data payload.

5.1 - SCI Advantages

SCI is starting to appear as a multiprocessor interconnection. Perceived SCI advantages include:

- Very high speed data transfers, that is, 4 to 7.5 GByte/s when the 16-byte header overhead is taken into consideration.

- Cache coherence built in at the hardware level.

- Chipsets are starting to become available.

5.2 - SCI SHORTCOMINGS

In the author's mind the following are disadvantages of SCI:

- The vendor's who have used SCI have not strictly adhered to the standard, but implemented a proprietary variant. This makes sense when you consider that the interconnect mechanism is a large part of the value added for a vendor, and is often what differentiates one vendor's product from anothers.

- Since the vendors are using proprietary SCI variants, the products from different vendors cannot be easily interconnected.

- Trying to achieve cache coherence across multiple vendor's products does not seem likely.

- The cache coherence mechanism used in SCI may not be optimum for all applications.

- Ring-based architectures suffer from the bandwidth limitation of a single data path, and SCI switches are not presently in use.

6 - HIGHER SPEED HIPPI DEVELOPMENT

HIPPI has been a work-horse interface at 800 Mbit/s since 1988, but the speed of the end-devices has not stood still, and today higher speeds are needed. Work is currently underway on development of a new physical layer standard called HIPPI-6400-PH with the following characteristics:

- A data transfer rate of 6400 Mbit/s (800 MByte/s). This is eight times the HIPPI-800 rate in common use today. The 6400 Mbit/s is fully available to the user--control information is carried in a parallel sideband and is not included in the bandwidth number.

- A full-duplex link capable of independent full-bandwidth transfers in both directions simultaneously. HIPPI-800 defined a simplex link, but most systems used two simplex links back-to-back to achieve full-duplex. The message-based control mechanisms of HIPPI-6400 requires full-duplex on all of the links.

- Four virtual circuits, which provide a multiplexing capability. On HIPPI-800, a packet transfer tied up a link until the transfer was complete, allowing a large block transfer to inhibit delivery of a small control message until the large transfer was complete. The four virtual circuits in HIPPI-6400 alleviate this limitation.

- A 32-byte fixed-size transfer unit (micro-packet) for hardware efficiency. HIPPI-6400 has some relationship to ATM in that both use a small fixed-size transfer unit that can be efficiently processed in hardware.

- The small transfer unit results in low latency for short messages, and a building block for large messages. Efficiency is enhanced by the 32-byte granularity.

- Credit-based flow control that prevents buffer overflow. HIPPI-800 used a credit-based flow control to prevent loss of data, and HIPPI-6400 continues the tradition. The HIPPI-6400 flow control is done on a virtual channel basis, preventing a stall on one virtual channel from blocking the other virtual channels.

- End-to-end, as well as link-to-link, check sums. Two 16-bit cyclic redundancy checksums (CRCs) are used, and checked at each node. A link-to-link CRC covers both the data and control portions of a 32-byte micro-packet. An end-to-end CRC covers only the data portion, and is cumulative over all of the micro-packets in a message. The two separate 16-bit CRCs give the equivalent checking capability of a 32-bit CRC without the difficulty of implementing a 32-bit CRC.

- Automatic retransmission of errored data providing guaranteed, in-order, error free, data delivery. The CRC on each 32-byte micro-packet is checked at each node, and if an error is detected the micro-packet is automatically retransmitted by the hardware without software intervention. Hardware checking, and retransmission on error, help keep the overall bandwidth high, even in the presence of transient errors.

- As the name implies, HIPPI uses parallel signal lines to transfer the data. Two variants are specified in HIPPI-6400-PH, using 8 and 16 bit-wide data paths. On the 8-bit-wide system the signal lines switch at 1 Gbit/s (1 ns) and the clock runs at 500 MHz (data is strobed on both edges of the clock). On the 16-bit-wide system the signal lines switch at 500 Mbit/s (2 ns) and the clock runs at 250 MHz. Dynamic skew adjustment is used on the signal lines to compensate for unequal signal length paths, circuitry differences, and temperature effects.

- An ac coupled parallel electrical interface for limited distance applications, and a parallel fiber-optic interface for longer distances. A 16-bit interface uses 20 signal lines: 16 data bits, 4 control bits, 1 Frame bit, and Clock. An 8-bit interface uses 12 signal lines: 8 data bits, 2 control bits, 1 Frame bit, and Clock. The signals are 4b/5b encoded to provide dc balance.

- A mapping for carrying legacy HIPPI-800 and HIPPI-1600 traffic, either on a HIPPI-6400 backbone, or between HIPPI-800, or HIPPI-1600, devices and HIPPI-6400 devices. Some people are planning to use up to eight HIPPI-800 interfaces to couple to a HIPPI-6400 environment and transfer at the full 6400 Mbit/s rate. This allows an easy transition from legacy interfaces and systems to the next generation systems.

- Support for existing HIPPI upper-layer protocols. Software costs are a major part of any system, and we have designed HIPPI-6400 so that the existing upper-layer protocols will run over this new higher-speed physical layer.

As with HIPPI-800, the emphasis in HIPPI-6400 is for simplicity, making it easier to specify, and to implement. Much of the HIPPI-6400 specification is based on previous work done by Silicon Graphics Inc. First HIPPI-6400 implementations are expected in 1997.

7 - SUMMARY

Local area network speeds are increasing to keep up with new generation computing equipment, and gigabit per second speeds are becoming a reality. Switches are replacing shared media at the higher speeds, and fiber optics are changing the error characteristics. Standards are becoming more important with few customers willing to invest in proprietary solutions.

HIPPI, as defined in the approved ANSI standards, supports data rates of 800 Mbit/s and 1.6 Gbit/s. Equipment based on the HIPPI standards has been available since 1988, and HIPPI is currently the interface of choice at the gigabit-per-second rate. Simplicity was a key concept in defining HIPPI, and resulted in an interface that was easy to understand and implement. Simplicity also helped HIPPI travel relatively quickly through the definition and standardization process. Interfaces between HIPPI and SONET, ATM, and Fibre Channel provide long distance interconnections and bridges to the future. Work on a HIPPI physical interface carrying user data at 6400 Mbit/s is underway, and first products should be available in 1997.

Fibre Channel, ATM, and SCI are all excellent interfaces, with advantages in some environments--there is no one interface that is best for all environments or applications. For high-speed local networking of heterogeneous machines, that is, from different vendors, HIPPI is seen as the dominant interface in use today. The higher-speed versions of HIPPI will provide a growth path for use with higher-speed computers and devices.

Acknowledgments

The Los Alamos National Laboratory is operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36. This work was performed under auspices of the U.S. Department of Energy.

References

[1] D. Tolmie, J. Renwick, "HIPPI: Simplicity Yields Success," IEEE Network - The Magazine of Computer Communications, January 1993, pp 28-32.

[2] D. Tolmie, D. Flanagan, "HIPPI: It's Not Just for Supercomputers Anymore," Data Communications Magazine, May 8, 1995.

[3] Approved ANSI standards are available from the American National Standards Institute, Attn: Customer Service, 11 West 42nd Street, New York, NY 10036, (212) 642-4900, or from Global Engineering, 15 Inverness Way East, Englewood, CO 80112-5704, (800) 854-7179. The HIPPI documents are also available via anonymous FTP from the X3T11/hippi directory of ftp.network.com.

[4] D. Tolmie, "High Performance Parallel Interface (HIPPI)," in High Performance Networks - Technology and Protocols, A.. Tantawy, ed., Kluwer Academic Publishers, Norwell, Massachusetts, 1994, pp. 131-156.

[5] M. Sachs, "Fibre Channel," in High Performance Networks - Technology and Protocols, A.. Tantawy, ed., Kluwer Academic Publishers, Norwell, Massachusetts, 1994, pp. 109-129

[6] R.Handel and M.Huber, Integrated Broadband Networks - An Introduction to ATM-Based Networks, Addison-Wesley, Wokingham, England, 1991.

[7] M. Zeug, "Broadband Integrated Services Digital Network (B-ISDN) Standards," in High Performance Networks - Technology and Protocols, A.. Tantawy, ed., Kluwer Academic Publishers, Norwell, Massachusetts, 1994, pp. 183-204.

[8] G. Shenoda, "Synchronous Optical Network - SONET," in High Performance Networks - Technology and Protocols, A.. Tantawy, ed., Kluwer Academic Publishers, Norwell, Massachusetts, 1994, pp. 205-228.

[9] David B. Gustavson, "The Scalable Coherent Interface and Related Standards Projects" , IEEE Micro, February 1992, pp. 10-22.

Biography

Don Tolmie joined the Los Alamos National Laboratory in 1959 as a Technical Staff Member, and has been involved with networking of supercomputers since 1970. His current task is defining the next generation computer network to support higher speeds and visualization, and working with vendors to provide the appropriate products. He has been involved in the computer interface standards activities for over 13 years, and for five years chaired ANSI Task Group X3T9.3 (since renamed to X3T11), responsible for the High-Performance Parallel Interface (HIPPI), Intelligent Peripheral Interface (IPI), and Fibre Channel (FC). Tolmie received a BSEE degree from New Mexico State University in 1959 and an MSEE degree from University of California, Berkeley, in 1961. He can be reached at the Los Alamos National Laboratory, MS-B255, Los Alamos, New Mexico, 87545; Internet, det@lanl.gov.