A major purpose of the Technical Information Center is to provide the broadest dissemination possible of information contained in DOE's Research and Development Reports to business, industry, the academic community, and federal, state and local governments.

Although a small portion of this report is not reproducible, it is being made available to expedite the availability of information on the research discussed herein.

1

LA-UR--89-2560

DE89 015283

Los Alamos National Laboratory is operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36

TITLE THE HIGH-SPEED CHANNEL (HSC) AND ULTRA-SPEED NETWORKS

AUTHOR(S) Don E. Tolmie, C-5

SUBMITTED TO BROADBAND (Foc/Lan) '89
San Francisco, California
October 30-November 3, 1989

Los Alamos

Los Alamos National Laboratory
Los Alamos, New Mexico 87545

THE HIGH-SPEED CHANNEL (HSC) AND ULTRA-SPEED NETWORKS

Don E. Tolmie

Los Alamos National Laboratory
C-5, Computer Network Engineering, MS-B255
Los Alamos, New Mexico 87545
(505) 667-5502

Summary

We have had excellent success, and user productivity increases, with a research project which provided computer-generated video-type output of physics calculations. Now we are investigating ways to provide this visualization capability to many users in a networked environment. We see the need for networking with channel speeds in the 800-Mbit/s range for data transfers to graphics displays, to storage systems, and between machines. A basic building block to achieve this capability is the High-Speed Channel (HSC) being worked on in American National Standards Institute (ANSI) Task Group X3T9.3. The HSC provides a point-to-point channel with speeds of 800 and 1600 Mbit/s (100 and 200 MByte/s). Industry interest is very high, and vendors are starting to produce hardware. Los Alamos is also working with some vendors on computer networks built around the HSC, crossbar switches, and intelligent network interfaces. A project is also underway to develop a fiber-optic transport media to support the higher layer protocols of HSC, the Intelligent Peripheral Interface (IPI) and the Small Computer System Interface (SCSI).

Introduction

High speed is a relative term. The HSC being defined in ANSI Task Group X3T9.3 is specified for peak data rates of 800 or 1600 Mbit/s (100 or 200 MByte/s). The HSC is a point-to-point channel for interconnecting computers and other data-processing equipment. In comparison to other physical interconnections, the 800-Mbit/s HSC is 80 times faster than Ethernet at 10 Mbit/s and 8 times faster than FDDI at 100 Mbit/s.

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.

HSC Characteristics

The HSC was designed to move data from memory to memory at such high rates that it has been likened to a fire hose. An early guideline was to keep the design simple. As such, it looks more like a communications channel than like many traditional computer channels.

Overview of the HSC Features

In relation to the Open Systems Interconnection (OSI) Basic Reference Model, the HSC covers the physical layer and a small portion of the data-link layer. The current version of the HSC uses a parallel data path with copper cable. The 800-Mbit/s HSC uses a 32-bit data bus, and the 1600-Mbit/s version of the HSC uses a 64-bit data bus. The major emphasis has been on the development of the 800-Mbit/s version.

The HSC is a simplex channel, capable of transferring data in one direction only. Two HSCs may be used to implement a full-duplex channel. The HSC is a point-to-point channel that does not support multi-drop. The point-to-point limitation considerably simplified the electrical and protocol aspects of the HSC. Crossbar switches and other networking methods are being considered to achieve the equivalent of multi-drop. An addressing mechanism is included to support these networking concepts.

The signal sequences provide look-ahead flow control to allow the average data rate to approach the peak data rate, even over distances of tens of kilometers. This is a benefit for future, fiber-optic versions of the HSC. Data transfers and flow control are performed in increments of bursts, with each burst nominally containing 256 words (1024 or 2048 bytes).

Error detection, but not error correction, is provided by the HSC. Byte parity is used on the data bus. In addition, each burst of data, where the burst is 256 words or less, is immediately followed by a length/longitudinal redundancy checkword (LLRC). We envision that error recovery would be done at a higher-layer protocol. Driving a video frame buffer, for example, is best served by ignoring errors that the next frame overwrites anyway.

The HSC provides support for low-latency, real-time, and variable-size packet transfers. The signal-line control sequences are simple and do not require any new silicon to implement. Prototype versions of the HSC are being built today with off-the-shelf commercial parts.

HSC Data Framing and Signalling Sequences

Figure 1 shows the basic organization of the information or data framing on the HSC. A connection is made in a fashion similar to the connection made when dialing the telephone. Once a connection is established a packet (or multiple packets) can be sent from the source to the destination. Each packet contains zero or more bursts, and each burst contains one to 256 words. Bursts that contain less than 256 words may only occur as the first or last burst of a packet. Words are composed of 32 or 64 bits. The amount of wait time between packets and bursts may vary. Maximum wait times depend on the data flow to or from the upper-layer protocols and on the data flow to or from the opposite end of the channel.

The interface signals are illustrated in Figure 2. The numbers in parentheses indicate the number of signal lines when using the 1600-Mbit/s option. The other numbers indicate the number of signal lines when using the 800-Mbit/s option. All signals, except for the INTERCONNECT signals, use differential emitter-coupled logic (ECL) drivers and receivers. The INTERCONNECT signals use single-ended ECL drivers and receivers.

Fifty-pair, twisted-pair cables are used for distances up to 25 meters. The 800-Mbit/s option uses one cable, and the 1600-Mbit/s option uses two cables. All of the signal lines in the HSC are unidirectional to accommodate future, fiber-optic implementations and crossbar switches. All of the control and data signals are timed in relation to the constant 25-MHz CLOCK signal with a period of 40 nanoseconds.

Typical HSC waveforms are shown in Figure 3 for a sequence that establishes a connection, sends a packet containing two bursts, sends a packet containing one burst, and then disconnects. A connection is made from the source to the destination much like a telephone connection. The source supplies the I-Field on the data bus (like a telephone number), and asserts the REQUEST signal. If the destination wants to accept the connection, it asserts the CONNECT signal. Although the contents of the I-Field are not specified in the HSC standard, the I-Field was intended for addressing or other control operations.

Once a connection is established, single or multiple packets may be transferred from the source to the destination. Packets are delimited by the PACKET signal being true. Packets are composed of zero or more bursts.

Bursts are delimited by the BURST signal being true. Bursts consist of a group of words sent on the data bus, one word per clock period, during contiguous clock periods. Bursts contain one to 256 words. Bursts that contain less than 256 words are called short bursts. There can be only one short burst per packet, and it must be either the first or last burst of the packet. Short bursts are for applications like packet headers, variable-length data transfers, and short messages that must be acted on quickly.

The LLRC checkword is sent from the source to the destination on the data bus during the first clock period following the burst. The destination controls the flow of data by issuing a READY pulse for each burst that it is prepared to accept from the source. If the READY pulses arrive at the source before the source is ready to send the next burst, there will be no time lost between bursts. Hence, the flow control is distance independent if the cable-length time is shorter than the time required to transmit the number of bursts. This requires about one burst buffer for each two kilometers of cable distance.

The INTERCONNECT signals indicate to both the source and destination that the cable(s) are connected and that the other end is powered up. They may also be used to indicate whether the 800-Mbit/s or 1600-Mbit/s HSC option is in use.

High-Speed Channel Uses

An immediate use for a channel of this speed is to enhance user productivity with computer visualization. For visualization, the digital data, usually in pixel or raster format, is output in a continuous stream or video form from a computer directly to a graphics screen. An image format of 1024 x 1024 pixels, with 8 bits of color information per pixel, requires 8 Mbit per image. With a refresh rate of 30 frames per second, this requires a sustained data rate of about 240 Mbit/s. These numbers are achievable with the 800-Mbit/s speed of the HSC. Experiments have shown that computer visualization successfully increases user productivity. Visualization will become commonplace as more commercial equipment becomes available.

Effective networking of supercomputers requires networking systems that operate at the highest speeds available on the supercomputers. By their very nature, supercomputers are traditionally peripheral poor, depending upon networks for their input data and connections to output devices. To balance the power and speed of the supercomputers, the networks that connect to them must be capable of handling data with minimal impact on the overall throughput of the supercomputers. Computer channels are now becoming available with bandwidths in the 800-Mbit/s range, well beyond the 50-Mbit/s rates of networking components available today.

Other identified uses for the channel include driving a paging memory from a mainframe, real-time data input and output, and closely coupling computers into a multi-processor system. The HSC can also be used in a tinker-toy fashion to assemble hardware and software systems for special applications, for example, interconnecting a Connection machine and a Cray computer.

New Opportunities

The much higher I/O rates are a new capability that requires us to rethink some old problems. For example, should simple graphics commands be expanded into the final image in a dedicated graphics workstation or in a shared supercomputer. When only low-speed communications links were available, the obvious answer was the workstation. Now that adequate communications are available, it may be advantageous to support a shared mainframe rather than many dedicated (and only partially utilized) graphics workstations. There are good arguments and applications for both cases.

Another concern related to visualization is how do we schedule the machines. Interactive visualization gives tremendous boosts in user productivity, but at a cost of central processing unit (CPU) horsepower. An environment that gives a researcher all of a supercomputer's resources for a five- to ten-minute interactive session may be advantageous over a timeshared environment in which only a few seconds continuous activity are allowed to provide essentially immediate response throughout the day. Timesharing is great for editing and problem setup but is not conducive to interactive visualization.

Los Alamos Networking Plans

The Los Alamos National Laboratory currently uses a local area network based on 50-Mbit/s channels for interconnecting a large set of supercomputers and other services. We plan to use the HSC to provide movie-mode visualization from the supercomputers. Further enhancements will be to use the HSC to interconnect the supercomputers and other services as equipment with these speed capabilities becomes available.

Since the HSC is only a point-to-point channel, we plan to develop and use crossbar switches to interconnect the HSCs. A feature of crossbar switches is that they allow multiple connections to pass data simultaneously and independently, a must for supporting movie-mode devices. In addition, we plan to experiment with intelligent interfaces in the interconnecting links to off-load the network protocol work from the hosts. The Los Alamos crossbar-switch system (16 x 16 or 32 x 32) with the intelligent network interfaces is called CP*. We are working with industry partners on the development of the crossbar switches and intelligent network interfaces.

The Fiber Channel

An original desire, in early 1987, of the ANSI task group working on HSC was that the HSC should be implemented with fiber optics to take advantage of the small connector footprints, longer distances, and improved EMI/RFI characteristics. The task group recognized an immediate need for a channel of the HSC's capability, that copper cabling is satisfactory for the immediate need, and that appropriate fiber-optic components at the HSC speeds were not fully developed at that time. Hence, the HSC physical layer was defined with copper cables, with the intent to transition to a serial fiber-optic implementation at a later date. The task group is now working on a project called Fiber Channel to provide the desired fiber-optic capabilities.

The Fiber Channel provides a transport vehicle for the upper layer IPI and SCSI command sets, and the HSC data link layer. The Fiber Channel is capable of replacing the SCSI, IPI, and HSC physical interfaces with a protocol-efficient alternative that provides performance improvements in distance and/or speed. IPI-3 commands, SCSI commands, and HSC data link operations, may all be intermixed on the Fiber Channel without impact on one another.

The Fiber Channel is optimized for predictable transfers of large blocks of data such as used for file transfers between processors (super, mainframe, super-mini, etc.) storage systems (disk and tape), communications, and to output only devices such as laser printers and raster scan graphics terminals.

The Fiber Channel standards are organized in four documents called FC-0 through FC-3.

FC-0 defines the physical portions of the Fiber Channel including the fiber, connectors, and optical parameters for a variety of data rates. A serial copper version is also included for limited distance low cost applications. This standard will cover speeds from about 50 Mbit/s to about 2000 Mbit/s (2 Gbit/s) using several fiber types and optical technologies.

FC-1 defines the transmission protocol, which includes the serial encoding and decoding, and the error control. Examples of coding methods include scrambling, run length codes such as 4B/5B and 8B/10B, and error correcting codes. The goal is to achieve DC balance, allow for special control codes, and simplify the clock recovery circuits.

FC-2 defines the signalling protocol which includes the frame structure and byte sequences. The framing sequence includes such things as source and destination addresses, packet sizes, and other control information.

FC-3 is highest layer in the Fiber Channel standards. It defines the channel protocol, or mapping, between the lower layer FC standards and the IPI-3 and SCSI command sets, and the HSC data link layer.

The Fiber Channel protocol is simple in order to minimize implementation cost and enhance throughput. The transmission medium is isolated from the control protocol so that implementation of point-to-point links, multi-drop bus, rings, crosspoint switches, or other special requirements may be made in a technology best suited to the environment of use.

The HSC as an ANSI Standard

Standards benefit the whole computer community. The user benefits by being able to purchase equipment from multiple vendors. The vendors benefit by having a wider set of equipment that can connect to their machines. System integrators and network vendors benefit by being able to concentrate on their value-added functions instead of developing special adapter boxes to interconnect proprietary channels. Everyone benefits from the economies of scale when common components are used.

Having standards available for interconnecting the supercomputer class of machine is something new. It is useful to have a standard like the VME bus, which is the generic backplane for workstations, where additional functions can easily be added. Until the HSC, interconnecting at channel speeds required special adapter boxes or proprietary equipment available mainly from Network Systems Corporation.

Status of the HSC Standard

The HSC physical layer was first proposed to ANSI in early 1987. The standard was drafted in a relatively short time because of the clear objectives, the force of the market, and the strong commitment of the active membership. The group not only met for the usual bimonthly meeting but added working meetings in between. The current HSC working draft is considerably different from the original straw-man proposal. The task group action has refined it and strengthened it in numerous areas. Design by committee really worked well in the case of the HSC standards.

The definition phase of the HSC standards effort is complete, and the approval phase is underway. The approval phase involves public review and letter ballots by various groups, which may result in suggestions for change. The task group will consider these suggestions and may change the draft to accommodate them. Implementations of the HSC can start at any time, but there is a small risk that the specification may change.

Based on past standards experience, the HSC physical layer (HSC Mechanical, Electrical, and Signalling Protocol) should become an approved standard the first or second quarter of 1990. The HSC Data Link should follow by about six months. The HSC will also be proposed as an international standard.

Follow-on Standards Work

The current HSC standard covers the mechanical, electrical, and signalling protocol requirements of a channel and corresponds to the OSI physical layer. The ANSI task group is now working on a companion data-link layer for the HSC. Existing data-link layers, for example IEEE 802.2, are not applicable to the HSC.

The Fiber Channel project mentioned earlier is also a follow-on effort to provide a fiber-optic version of the HSC. The Fiber Channel is in the definition phase.

Participants in the HSC ANSI Effort

The Los Alamos National Laboratory is leading the effort to standardize the HSC through the ANSI organization. The Laboratory wants the ability to purchase equipment from multiple manufacturers and, with minimal effort, interconnect all of the equipment into a cohesive network.

The list of vendors participating in the ANSI task group X3T9.3 is long and illustrious. Manufacturers of mainframes, superminis, graphics devices, workstations, fiber-optic components, integrated circuits, networking systems, storage systems, as well as other national laboratories and universities are all working together on the HSC standard. Some of the more well-known companies that are participating include IBM Corporation, Digital Equipment Corporation, Control Data Corporation, Cray Research, Inc., AT&T, Scientific Computer Systems, Ultra Network Technologies, Amdahl Corporation, Data General, Thinking Machines Corporation, and Network Systems Corporation. Without the commitment and dedication shown by the participating organizations and individuals, the HSC in its current form would not exist.

The ANSI meetings are open to all interested parties, and you are welcome to participate if you have the interest, time, and expertise in this area.

Conclusions

The HSC is moving quickly towards becoming an ANSI standard. It is based on proven technology and the principle of keeping the design simple. The HSC has excellent support from the vendors and potential users. Also, the HSC is getting support from, and will affect, the mini-supercomputer, workstation, and graphics industry segments. The HSC is expected to change the way supercomputer users connect their systems; namely, the HSC supports the "supercomputer glue" business of interconnecting machines in a tinker-toy fashion. Presently the HSC is based on limited-distance copper cables, and the ANSI task group is starting to specify a fiber-optic version for longer distances and the other fiber-optic benefits. The HSC is being proposed as an international standard.

Acknowledgements

The HSC owes a large debt of gratitude to all of the participants in the ANSI effort. Major Los Alamos contributors include Karl-Heinz Winkler who provided the initial impetus through his computer visualization work, Michael McGowen who conceived the HSC, Gene Dornhoff for electronic contributions, and Randy Hoebelheinrich and Richard Thomsen for their work on the higher layer protocols.

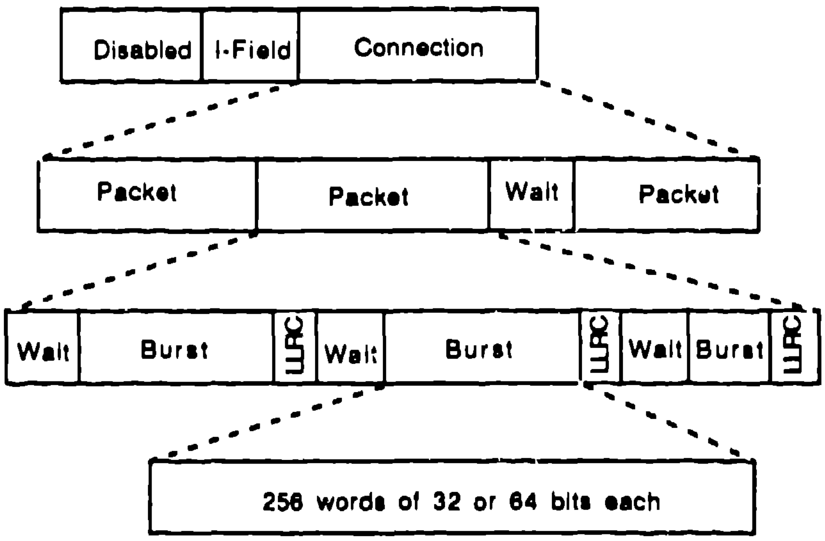

Figure 1. Framing hierarchy of the HSC.

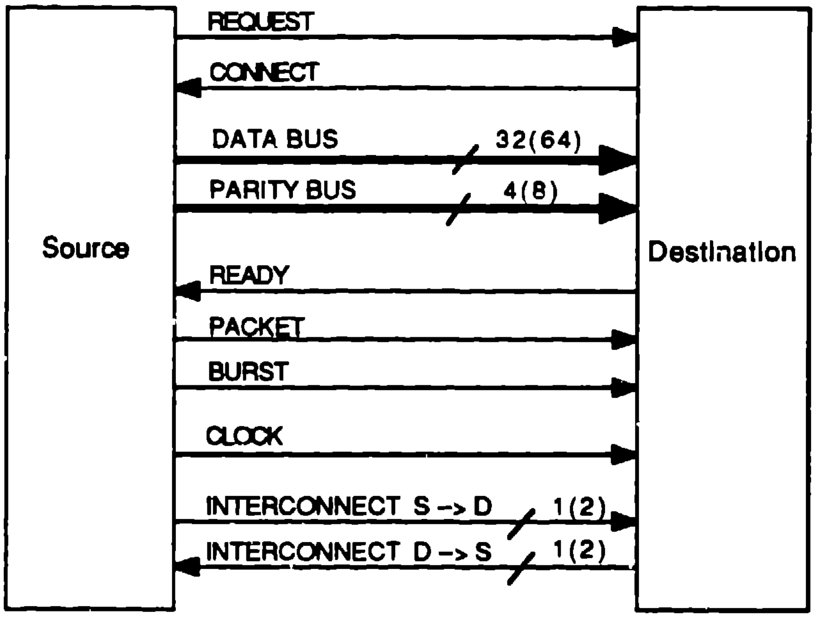

Figure 2. Summary of the HSC interface signals. The numbers indicate the number of signal lines when using the 800-Mbit/s (1600-Mbit/s) option.

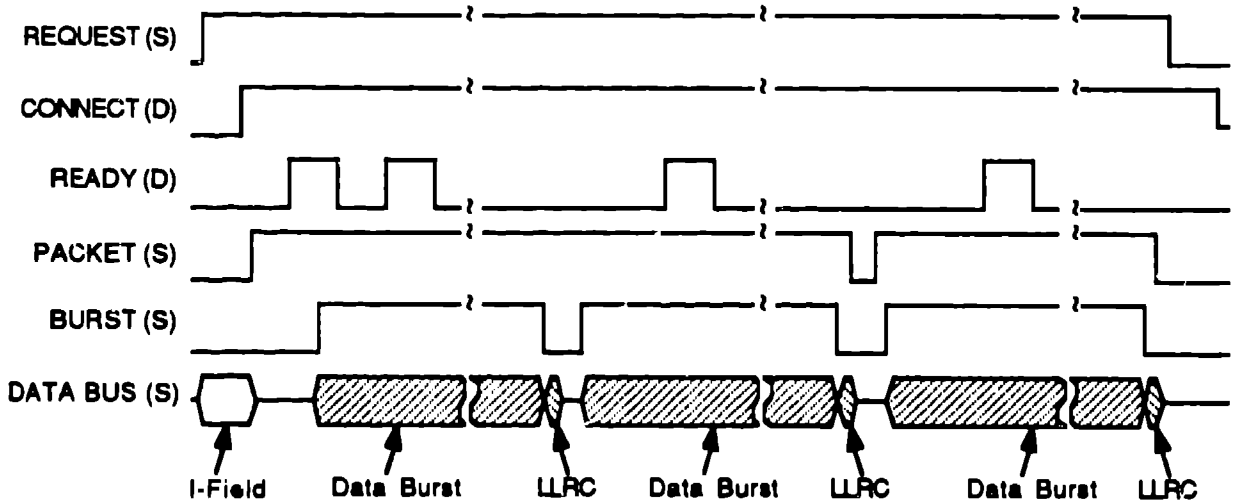

Figure 3. Typical HSC waveforms.