RC 20279 (11/17/95)
Engineering and Technology

IBM Research Report

Fibre Channel

Martin W. Sachs

IBM Research Division
T.J. Watson Research Center
Yorktown Heights, New York

Anujan Varma

Computer Engineering Department
University of California
Santa Cruz, California

IBM Research Division
Almaden · T.J. Watson · Tokyo · Zurich

Fibre Channel

Martin W. Sachs[*]   Anujan Varma[†]

IBM Research Report RC-xxxxxx

Nov. 7, 1995

Abstract

Fibre Channel is a new ANSI standard for attachment of high-performance I/O devices to processors and for clustering processors. We provide an overview of Fibre Channel, a summary of a performance study, and a description of current industry activity.

{**Keywords: Fibre Channel, interconnection networks**}

[*]IBM T. J. Watson Research Center, P.O. Box 704, Yorktown Heights, NY 10598, sachs@watson.ibm.com, (1-914-784-7287)

[†]Computer Engineering Department, University of California, Santa Cruz, CA 95064, varma@ce.ucsc.edu, (1-408-459-3505)

Introduction

In the late 1980s, Accredited Standards Committee X3's working group responsible for developing the ANSI Intelligent Peripheral Interface (IPI), Small Computer Systems Interface (SCSI), and High Performance Parallel Interface (HIPPI) began to design Fibre Channel, a common serial optical interconnect to replace the parallel electrical interconnects on which these standards were based. As the work progressed, the design evolved to encompass a broader range of applications than those served by the earlier standards, and continuing work will result in enhanced standards that encompass an even broader range of applications. In addition, to satisfy the most cost-sensitive applications, provision for serial electrical interconnect was added.

Fibre Channel consists of a suite of standards of which the first, the base standard, *Fibre Channel Physical and Signalling Interface* (FC-PH), became an ANSI standard in 1994. Additional members of the suite are now being completed by the X3T11 Fibre Channel working group and are expected to be issued during 1995 and 1996.

Initially, Fibre Channel is targeted at attachment of storage devices to systems and system clustering at distances up to several km for applications involving both the large data sets typical of scientific and technical computing and medical imaging, and the smaller but higher-volume data transfers typical of on-line transaction processing. Initial uses of Fibre Channel include clustering of workstations, connection of workstations to supercomputers, and attachment of SCSI storage subsystems to workstations or supercomputers.

FC-PH defines a point-to-point or switched point-to-point link which includes definitions of the physical layer, transmission code, and higher-level functions such as flow control and multiplexing. Associated standards define generic requirements for the interconnection fabric

(FC-FG) and more specific interconnection details including a combined circuit- and frame-switched fabric (FC-SW) and a low-cost arbitrated loop (FC-AL). Also nearing completion is a fabric-services standard (FC-GS) which includes definitions of directory, management, and time services.

What has been described above is a generic data-transfer interconnection system. Layered on top of FC-PH is a set of mapping definitions which enable specific higher-level protocols such as the SCSI device command set to be carried by FC-PH. The higher-level protocols are discussed in a subsequent section.

Overview of FC-PH

FC-PH defines a system consisting of a collection of nodes interconnected by a fabric. A node is, in general, a computer run by an instance of an operating system and may be a general purpose application processor, a server, or an I/O subsystem. The entity within a node which connects the node to another node or to the fabric is called an N_Port ("node port"). A node may contain one or more N_Ports. An N_Port is connected to another N_Port or to the fabric by means of a link (the physical transmission medium). An N_Port contains the link transmission and reception hardware, an interface to the node memory, and the function which moves data between the memory interface and the link hardware. The equivalent entity in the fabric, to which the link from an N_Port is connected, is called an F_Port ("fabric port"). Figure 1 illustrates a set of interconnected nodes.

The FC-PH standard includes three levels of function: the physical level, the transmission protocol, and the signalling protocol. Figure 2 shows this level structure along with two higher levels which will be discussed later. Architecturally, the physical, transmission, and signalling

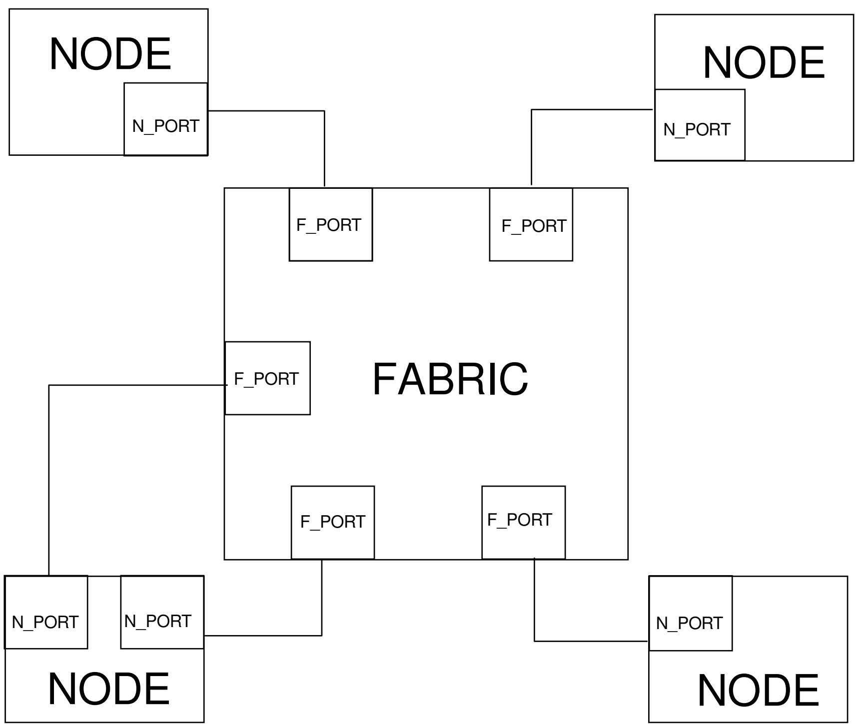

Figure 1: Several nodes interconnected by a fabric

levels reside in the N_Port and F_Port although the specific signalling-protocol functions are somewhat different in these two places. Figure 3 illustrates the structure of an N_Port.

Physical level: The physical level includes definitions of bit rates, transmission media and their parameters, and connectors. It includes a menu of technologies suitable to different cost and performance design points, illustrated in Figure 4. Transmission rates of 1062.5, 531.5, 265.625, and 131.8125 Mbaud are specified, corresponding to payload rates of 100, 50, 25, and 12.25 Mbytes/s, respectively. A future enhancement of the standard will add 4250 and 2125 Mbaud. Transmission media include optical fiber, copper coaxial cable, and copper shielded twisted pair cable. Both single-mode (9 μm) and multimode (50 μm and 62.5 μm) fiber are provided. Optical emitters are long-wave LED, short-wave laser, and long-wave laser. Optical wavelengths are 1300 nm and 780 nm. The maximum distances are 500 m to 10 km for optical and 10-100 m for electrical media, depending on the specific combination of parameters chosen.

Transmission protocol: The transmission protocol includes the definition of the transmission code and related functions. The transmission code is an adaptive 8-10 code[1] which provides sufficient transition density and DC balance to meet the requirements of an optical receiver. It also provides a large number of control codes in the 10-bit domain which are in addition to the 256 code points of the data alphabet. The transmission code intrinsically provides for detection of all odd-bit errors and a large number of other error patterns. This is reinforced by a 4-byte cyclic redundancy check code (CRC) on each frame.

Information on the link is organized in groups of four 10-bit codewords. This permits efficient implementation of 4-byte interfaces and data flows in the N_Port hardware. Control information is represented by groups of four 10-bit codewords known as ordered sets. An

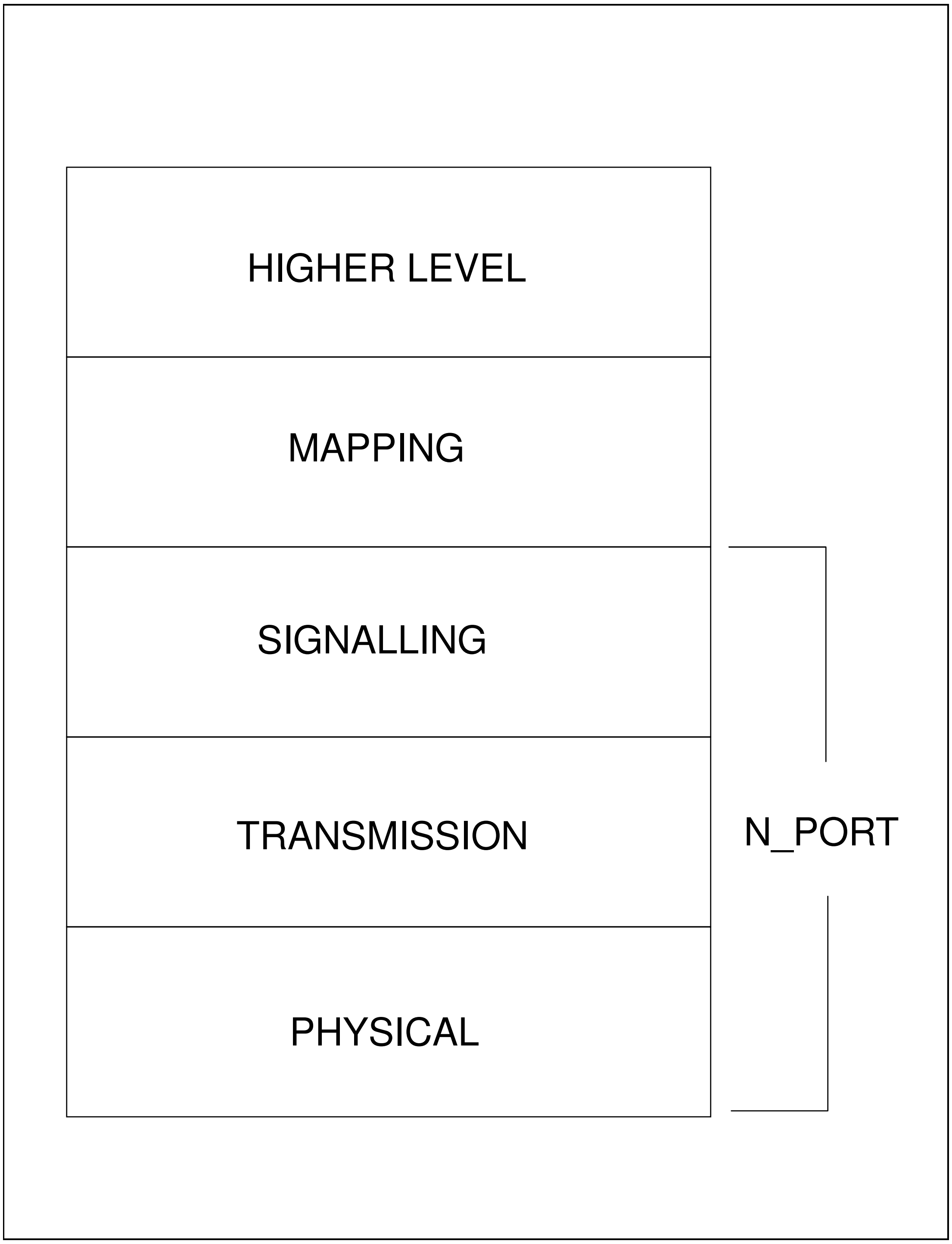

Figure 2: Level structure of FC-PH

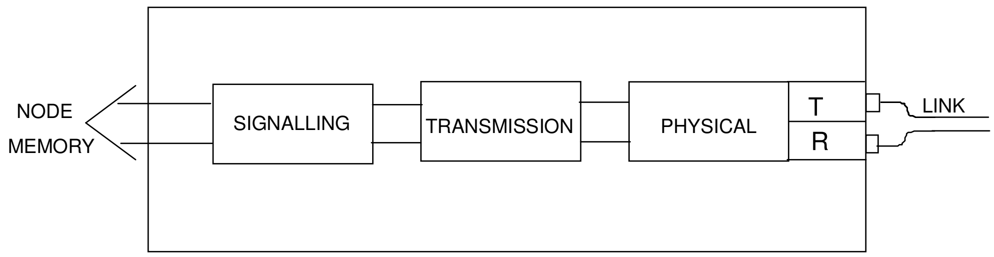

Figure 3: Structure of an N_Port

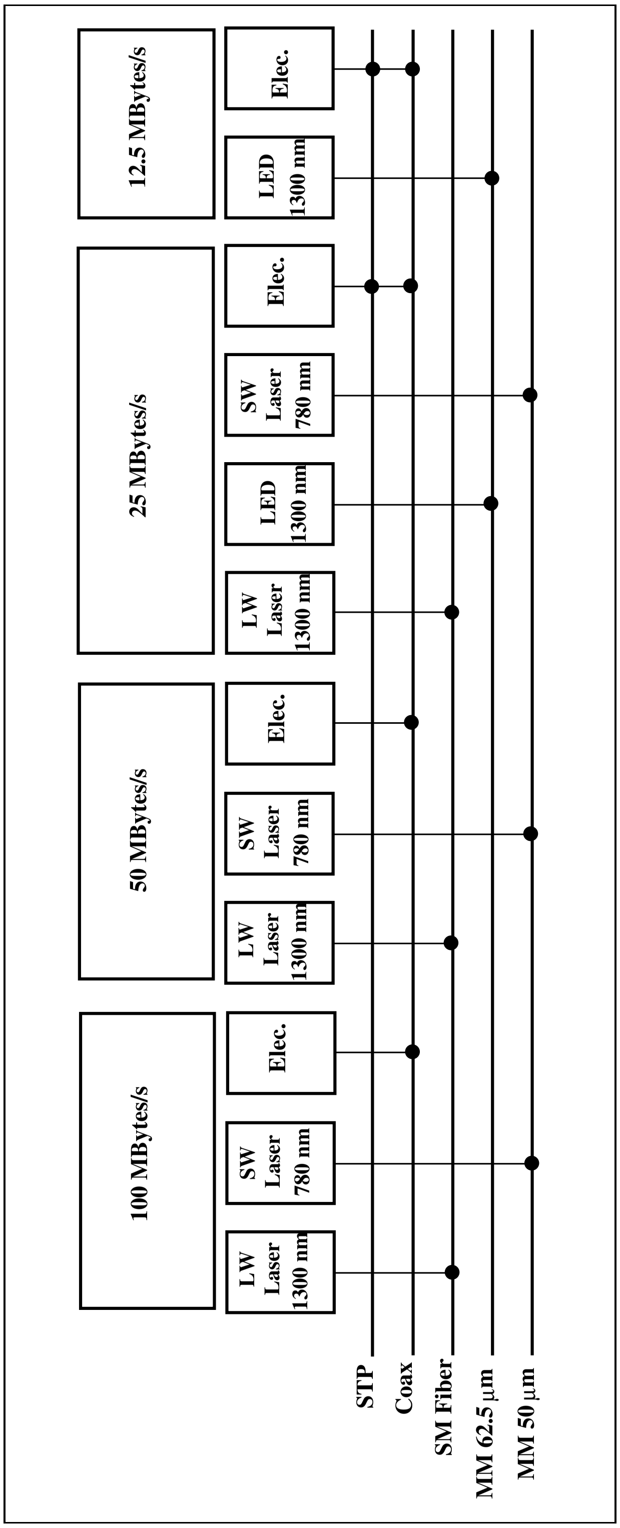

Figure 4: FC-0 technologies. Abbreviations: LW = long wave; SW = short wave; STP = shielded twisted pair; SM = single-mode fiber; MM = multimode fiber.

ordered set consists of a control codeword followed by three data codewords. The control codeword used in all currently defined ordered sets is the comma character denoted K28.5. This codeword has the special property that it cannot appear as a result of the juxtaposition of two codewords in an error-free data stream. It is thus used to enable a receiver to align itself to the boundaries of the 4-codeword groups. Among the functions performed by ordered sets are frame delimiters, idles (transmitted between frames), and certain control operations performed with continuous streams of particular ordered sets (primitive sequences).

Signalling protocol: Functions provided by the signalling protocol include the definition of the frame and its header, initialization (login), routing through the fabric, data flow control, multiplexing management, and connection management. Three classes of service are provided. Class 1 provides full-duplex circuit-switched connections through the fabric. Class 2 is a frame-switched service, with acknowledgements and flow control. Class 3 provides unacknowledged datagrams. The basic entity transferred from sender to recipient is a sequence, which is a collection of bytes provided by the higher level function in the node and transferred to its counterpart in the recipient node. A sequence is transferred as the payload of one or more data frames. Error recovery is on a sequence basis. In class 1 and class 2, it is guaranteed that a sequence is either delivered or an error is indicated. In class 3, the higher level must handle error detection and recovery.

Each frame contains up to 2112 bytes of payload plus a 24-byte header and ends with a 4-byte CRC field. It is preceded by a start-of-frame delimiter and followed by an end-of-frame delimiter (1 ordered set each).

Initialization is provided by the login function, by which N_Ports identify themselves to each other and to the fabric and exchange information about optional features and operating parameters such as number and sizes of buffers. During login with the fabric, the fabric may assign an N_Port its address.

Routing of a frame through the fabric is based on a 24-bit address in the frame header which identifies an N_Port which is the source or destination of the frame. It should be understood that in class 1, routing according to the destination address is used mainly for the connection request frame. One a physical circuit is established through the fabric, frames flow on that circuit without involving routing function. From the viewpoint of an N_Port, the address is a single-level address. However a fabric may internally structure the address to facilitate routing. A recommended address structure is given in the FC-FG document.

Flow control is credit-based. In classes 1 and 2, there is end-to-end flow control between communicating N_Ports, using acknowledgments to data frames. In class 2, there is an additional level of flow control, called buffer-to-buffer flow control, which operates between each N_Port and the fabric and is used to control fabric congestion. In class 3, there is only buffer-to-buffer flow control. Flow control is illustrated in figure 5.

Multiplexing is based on the concept of an exchange. Each exchange consists of a unidirectional or bidirectional data stream which flows between logical entities, termed the originator and responder, at the two communicating N_Ports. Multiplexing is performed by interleaving frames belonging to different exchanges. Each data stream is labelled with a pair of identifiers called exchange identifiers; one member of the pair is supplied by each of the communicating N_Ports when the exchange is initialized. In class 1, there can be multiple exchanges between two connected N_Ports. In classes 2 and 3, exchanges between a source and multiple destinations, and between a destination and multiple sources, can be multiplexed. An "intermix" facility permits one or both N_Ports in a class-1 connection to also communicate with other N_Ports using class-2 frames, which have lower priority than the class-1 data, multiplexed

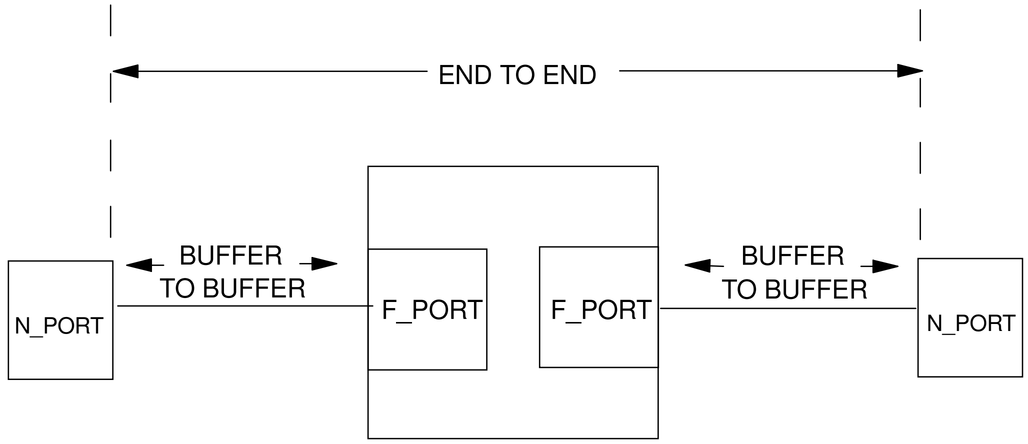

Figure 5: Flow control domains. Two N_Ports are shown communicating through a fabric.

with the class-1 exchanges.

Fibre Channel Fabric

Instead of defining the fabric as a specific interconnection topology, the Fibre Channel defines a functional model for the interconnection network in terms of the functions it must support, rather than its specific topology and implementation. A wide variety of fabric models is possible.

For high-performance I/O systems, a switched network consisting of one or more switches is the preferred choice of fabric. A Fibre Channel switch can provide non-blocking access to the N_ports, thus avoiding the problem of contention in the fabric. Significant flexibility exists in the design of Fibre Channel switches, such as the switch architecture (crossbar, multistage), classes of service supported, and other options provided. Techniques such as camp-on and stacking may be used to improve the throughput of the switch [4, 5]. In the future, Fibre Channel switches will incorporate mechanisms for providing quality-of-service (QoS) guarantees for real-time data.

For I/O systems where a switch-based fabric is too expensive, a low-cost interconnect based on a ring topology is defined in the Fibre Channel Arbitrated Loop (FC-AL) standard. An important use of FC-AL is to serve as the fabric for interconnection of SCSI-FCP disk drives within a disk subsystem. In FC-AL, the N_ports are connected via point-to-point links to form a closed loop. A loop may accommodate up to 127 N_ports, in addition to an F_port for interconnection to a larger Fibre Channel fabric. Access to the ring is controlled by an arbitration algorithm. An N_port requiring access to the medium transmits a request around the loop when the medium is idle; the request is granted when it returns, unmodified, to the

N_port that transmitted it. A higher-priority N_port in the loop may modify the request; thus, the highest-priority node among the requesting N_ports wins the arbitration. A fairness algorithm is defined to provide fair access to all ports in the loop. The fairness algorithm allows each requesting N_port to access the loop during an access window.

Future Enhancements to FC-PH

The Fibre Channel working group is developing enhancements to the original FC-PH standard. In addition to the 4250 and 2125 Mbaud link rates previously mentioned, following are some of the enhancements which will be submitted to the formal standards process during 1995.

Fractional Bandwidth: A new class of service, class 4, will provide bidirectional virtual circuits between N_Ports. A virtual circuit has an associated quality-of-service description which includes guaranteed bandwidth and latency. Each virtual circuit provides a fractional allocation of the total bandwidth available between the communicating N_Ports.

Data Compression: Adaptive lossless data compression will be provided as part of the signalling level. The higher-level protocols will be able to request compression separately for each sequence.

Multicast: Unacknowledged class-3 multicast messages will be provided. A multicast message is sent to a specific group of N_Ports identified by a multicast alias.

Hunt Group: A hunt group is a group of N_Ports identified by a hunt-group alias. A message sent to a hunt-group alias will be delivered to any one available N_Port in the hunt group. Typically, the N_Ports in a hunt group are at the same node. The hunt group is similar to the Asynchronous Transfer Mode (ATM) concept of "anycasting".

Alias Services: The alias services provide for defining address aliases to represent multicast groups and hunt groups and registering individual N_Ports as members of groups denoted by specific aliases.

Process Login: Process login provides initialization of communications between higher-level processes or groups of processes.

State-Change Notification: If an event occurs at an N_Port which affects resources involved in logins with other N_Ports, state-change notification will notify the other involved N_Ports that recovery actions may be needed.

Buffered Class 1: In the base definition of class 1, two N_Ports communicating over a circuit-switched connection must use the same link baud rate. Buffered class 1 will permit N_Ports which have different link baud rates to communicate using class 1.

Dedicated Simplex: In the base definition of class 1, full-duplex circuit connections are provided. Dedicated simplex provides simplex circuit connections. An N_Port using dedicated simplex will be able to send over a circuit to one destination while receiving over a circuit from a different source. ACKs will be sent using class 2.

Camp-On: In the base definition of class 1, if the destination of a class-1 connection request is busy, a busy indication is returned to the connection initiator. Camp-on will enable the connection request to be queued until the destination is available to accept the connection request.

In addition, an avionics working group, denoted FC-AE, is developing enhancements to address applications of Fibre Channel which require high reliability, fault tolerance, and deterministic behavior (for real-time applications), especially for the commercial and military aerospace industries. Among the subjects being addressed are static switching, real-time loop or mesh, the requirements of real-time embedded systems, trusted computer base, time distribution and clock synchronization, error process policy, and distributed interactive simulation.

Mapping Layer

As mentioned earlier, FC-PH provides a point-to-point, loop, or switched data transmission service which can be utilized by a variety of specific higher-level protocols. The initial use of FC-PH will be to provide serial data transmission for several existing standards which were originally designed with parallel electrical busses.

As viewed by the higher-level protocol, FC-PH provides a set of primitive services. For each of the higher-level protocols, a mapping standard (known as an FC-4 standard) must be designed to specify how the higher-level protocol uses the FC-PH primitives. Figure 2 shows the relationship of the mapping and higher-level protocol to the other FC levels.

The following FC-4 mapping standards are nearing completion or under development at this time. The document name is shown in parentheses.

- SCSI Fibre Channel Protocol (SCSI-FCP): This standard defines the rules for mapping the SCSI device command set protocol onto FC-PH. Effectively, FC-PH replaces the SCSI parallel bus.

- IPI-3 on Fibre Channel: This standard defines the rules for mapping the IPI-3 device generic protocols onto FC-PH, thus replacing the IPI-1 parallel bus by FC-PH. This mapping standard is published as a clause in the IPI-3 standards.

- Link Encapsulation (FC-LE): This standard defines the rules for mapping the ISO 8802.2 Logical Link Control onto FC-PH. Its definition allows existing communications software to treat FC-PH as an additional member of the suite of local area network protocols supported by ISO 8802.2.

- IP and ARP on Fibre Channel: This document, which is a proposed Internet standard is a subset of the FC-LE definition which is sufficient to define how Internet Protocol (IP), with ISO 8802.2 SNAP, is carried by Fibre Channel.

- Mapping of HIPPI-FP (FC-FP): This standard defines the rules for mapping the HIPPI framing protocol onto FC-PH, thus allowing Fibre Channel to be substituted for the HIPPI physical layer and HIPPI switch.

- Mapping of ATM (FC-ATM): This standard defines the rules for mapping ATM/AAL5 onto FC-PH. This standard will permit building fabrics which can carry both Fibre Channel and ATM data streams.

- Mapping of Single-Byte Command Code Sets (FC-SB): This standard defines the rules for mapping the IBM ESCON[1] I/O Interface[3] protocol onto FC-PH. It will enable migrating systems and devices which were originally designed to the ESCON architecture to open Fibre Channel systems on which multiple of the protocols in this list can be used concurrently.

- Simple Network Management Protocol: This standard defines the rules for using SNMP

[1]ESCON is a trademark of the IBM Corporation.

to manage Fibre Channel systems. Accompanying it are management information block (MIB) definitions for the N_Port and fabric.

The FC-PH services are expressed in terms of a set of service-interface primitives which are published as an annex (appendix) in the FC-PH standard document. The annex defines the services which are provided but not the manner of implementation. In order to leave product designers the maximum freedom, the set of primitives does not constitute a mandatory interface. Since the conceptual interface defined by the primitives is generally viewed as existing entirely internally to each product, there is no interoperability issue. An implementer may choose to implement the higher-level protocol, the FC-4, and FC-PH as separate or combined software, hardware, and microcode elements in any matter which satisfies cost, performance, modularity, or other constraints.

Following are some examples of the service-interface primitives in FC-PH.

FC_PH_SEQUENCE.request: When the FC-4 level issues this primitive, it causes a data sequence to be transmitted to a destination FC-4 level. Parameters include the data to be transferred, an identifier representing the destination address, etc. The FC-4 definition corresponding to a given higher-level protocol specifies things such as the rules for sending commands, data, and completion information as sequences, maximum and minimum sequence lengths, rules for sending sequences between the communicating FC-4 levels at different N_Ports, and other things related to constructs in the higher-level protocol.

FC_PH_SEQUENCE.indication: This primitive signals the receiving FC-4 level that a sequence has arrived. Associated information includes the identification of the source of the sequence.

N_PORT_LOGIN.request: This primitive causes a login sequence to be sent to a destination N_Port. The sequence contains operating parameters supplied by the FC-4 level.

I/O Operations on Fibre Channel

The mapping layer described in the previous section allows all the common I/O device protocols, such as SCSI, IPI, and the IBM ESCON I/O Interface, to be used in Fibre Channel. The mapping layer provides the services necessary to perform the I/O operations in the device command set using the services available from FC-PH. The operation of a device command can be illustrated with respect to the SCSI Fibre Channel Protocol (FCP) as follows:

Data transfers in Fibre Channel are performed using *sequences*. A sequence represents one or more Fibre Channel frames transmitted in one direction. Error control is performed at sequence boundaries; that is, in the event of a frame loss, the entire sequence containing the frame is retransmitted. I/O operations are mapped into exchanges by the mapping layer. An I/O operation is performed by transfer of sequences between the exchange originator and responder according to the rules defined by the particular mapping protocol (FC-4).

An I/O operation initiated by FCP may consist of a single request or a list of chained requests. Each request specifies the parameters associated with the corresponding SCSI command, including the location of the buffers to be used for data transfer. The N_port that acts as the SCSI initiator sets up an exchange with the SCSI target. This exchange is then used to transfer commands, responses, and data associated with the I/O operations. Commands, responses, and data are distinguished by the use of distinct sequence names within the FCP information units.

Command execution starts by the initiator sending the SCSI command and its accompanying parameters to the SCSI target over the exchange set up over the Fibre Channel. On interpreting the command, the target transmits a data-transfer request to the initiator, specifying the size of the data block to be transferred. If the operation is a read, the target may transfer data immediately following the data transfer request. If the operation is a write, the initiator starts the data transfer on receipt of the request from the responder. Data transfers during the execution of the I/O command may be controlled either by means of explicit delivery requests between the two endpoints, or other system-specific mechanisms. On completion of the transfer, the responder transmits status information to the initiator, which marks the end of the I/O operation.

FCP allows multiple I/O operations to be in progress at the same time via multiple concurrent exchanges. An exchange may utilize either class-1 or class-2 service of FC-PH. Using class 1 allows data from multiple I/O operations between the same endpoints to be multiplexed. Class 2 allows I/O operations to be multiplexed between multiple end points. Class 3 cannot be used with FCP as class 3 does not provide confirmation of delivery.

Device addressing in FCP is performed using an 8-byte address that allows hierarchical addressing. While the actual choice of the address structure is left to the vendor, the address format is intended to provide flexibility in addressing devices. For example, the collection of disks in a disk array can be addressed as a single logical unit during normal operation, but the individual disks in the array can be addressed independently for configuration and diagnostics.

Performance of a Fibre Channel Fabric

In the context of the High-Performance Switching System (HPSS) project at Lawrence Livermore National Laboratory, a detailed simulation model was developed for a Fibre Channel fabric based on a three-stage switching network. The HPSS is a distributed switching system designed to operate with the Fibre Channel protocols. The largest configuration of the switch has 4096 full-duplex ports. The system is constructed from interconnected crossbars, based on a folded version of the Clos three-stage network. The system uses distributed control, unlike most of the previous designs based on Clos networks. Each fabric port supports a data rate of 1.0625 Gbits/s, corresponding to the highest speed currently supported by Fibre Channel. The simulation software consists of more than 10,000 lines of CSIM code.

The HPSS switch in Figure 6 has three stages, referred to as the *I/O*, *transfer*, and *cross-connect*(CC) stages, respectively. The building blocks of the switch are called *racks*, each of which provides up to 64 internal ports. The racks have somewhat different functionality than single crossbars. A crossbar in the first stage of the Clos network extends from an I/O-stage rack to the corresponding transfer-stage rack; a crossbar in the second stage of the Clos network is completely contained within a cross-connect stage rack; and a crossbar in the third stage of the Clos network extends from a transfer-stage rack to the corresponding I/O-stage rack. Each rack is equipped with a set of microcontrollers to perform connection management and signalling.

A connection is set up in the HPSS switch starting at the source port and proceeding through the I/O and transfer racks into a CC-stage rack, subsequently to a transfer-stage rack, and finally to the I/O-stage rack corresponding to the destination port. This path setup is performed in a distributed fashion, with each microcontroller choosing an available path

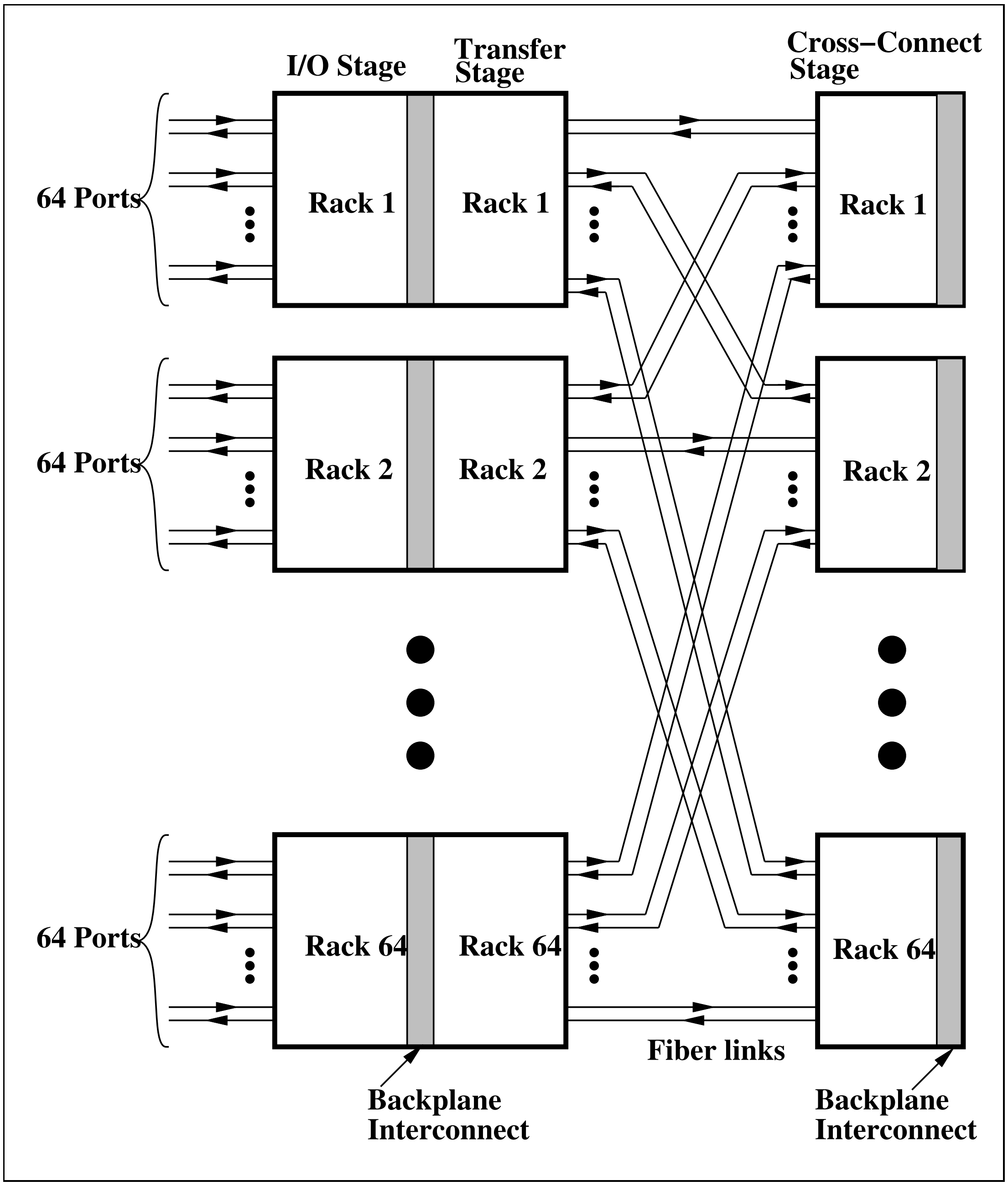

Figure 6: An HPSS configuration with 4096 ports.

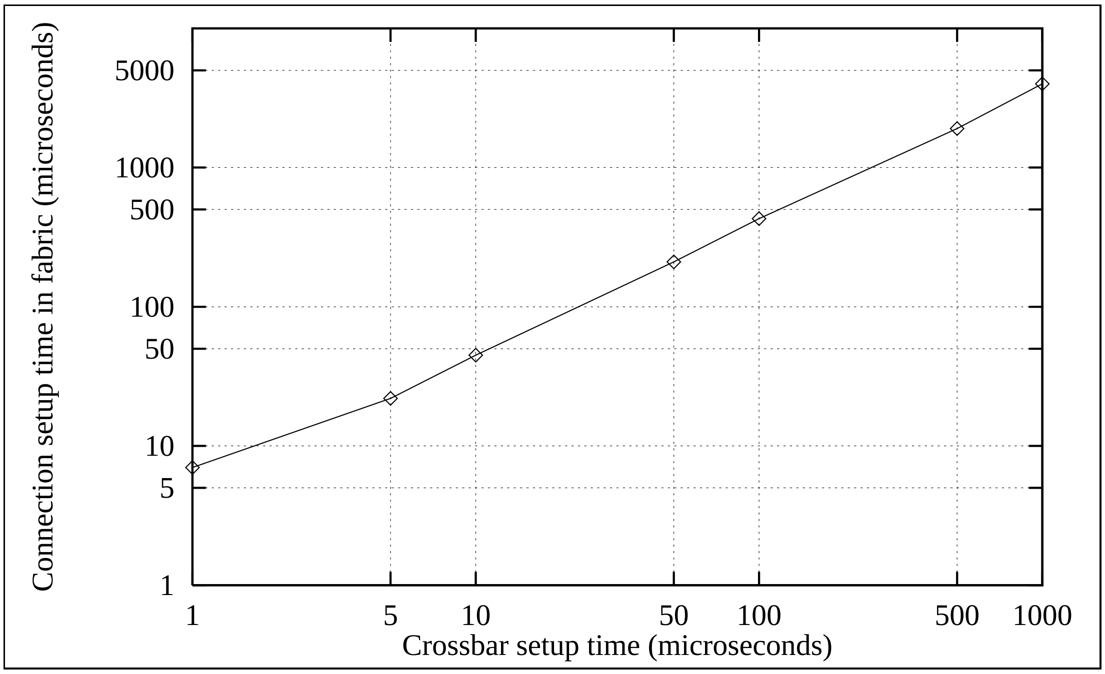

Figure 7: Connection setup time in a Fibre Channel fabric based on a three-stage switching network with 512 F-ports as a function of the setup time of the individual crossbar switches.

to the next stage. If no path is found, the request backs off to the previous stage. To aid in the path selection, each microcontroller maintains a list of available paths to the next stage. The HPSS uses a status update scheme to broadcast state changes from each microcontroller to others.

Figure 7 provides a sampling of the simulation results for a 512-port HPSS configuration. The figure plots the connection setup delay in the three-stage fabric as a function of the setup delay of the individual crossbars. The configuration simulated consists of eight I/O-stage racks, each providing 64 F-ports, and the same number of transfer-stage racks. The inter-stage fiber links are assumed to be 100 meters long. A traffic distribution was used in which a fraction α of the attempted connections are short (one Fibre-Channel frame with 2140 bytes) and $(1 - \alpha)$ are long connections (a total of 3 Mbytes per connection). The results shown in Figure 7 are for $\alpha = 0.95$. Connections of each type were generated from an Erlang-2 distribution and the destinations of connections were chosen from a uniform distribution. With the chosen set of parameters, The points in Figure 7 correspond to the saturation throughput for each value of switch latency on the x-axis. For more detailed results on the performance of HPSS, the reader is referred to [2].

Industry Activity

ANSI publishes standards but deals with neither interoperability issues nor with commercialization. In a world of open standards, these issues are best dealt with by industry-wide associations of which perhaps the best known is the ATM Forum.

Fibre Channel, because of its goal of addressing a broad range of applications and being attractive to the entire industry, has a large number of optional features. This causes a concern that interoperability difficulties might cause problems for both vendors and customers. These problems are being addressed by an industry association and by Fibre Channel interoperability testing facilities created at Lawrence Livermore Laboratory and proposed (as of this time of writing) at the University of New Hampshire InterOperability Laboratory.

In early 1993, Hewlett-Packard, IBM, and Sun Microsystems announced a joint effort, called the Fibre Channel Systems Initiative (FCSI). FCSI stated goals of advancing Fibre Channel as a high speed interconnection for workstations and systems, promoting open systems for distributed computing, and addressing interoperability issues. To address interoperability, FCSI developed profiles which define sets of Fibre Channel options to address specific applications such as workstation clustering and attachment of storage subsystems to workstations. FCSI expected to complete its work and go out of business in mid-1995, as this article was being written.

Shortly after FCSI began its work, a broader-based group, called the Fibre Channel Association (FCA) was founded, open to the entire industry. It was started by 19 companies and grew to 74 by April, 1995. Its goals are to promote industry awareness, acceptance, and advancement of Fibre Channel products. It encourages publications of articles, conducts seminars and conferences, and provides a Fibre Channel presence at trade shows and other events. It will also continue and extend FCSI's work on interoperability profiles and will work to foster interoperability testing. FCA also intends to maintain technical liaisons with other standards bodies and consortia.

According to information provided by the Fibre Channel Association, as of May, 1995, 28 companies had announced Fibre Channel products as available or in development. Announced products include physical layer chips and transducers, signalling protocol chips, adapters for several different system busses, switches, testers, and management software.

Summary

Fibre Channel is an ANSI standard which was originally proposed to provide a serial fiber-optic interconnect to replace the parallel electrical interconnects of IPI, SCSI, and HIPPI. Its design has evolved to encompass a much broader range of applications including both performance-sensitive and cost-sensitive applications.

The functional layers in Fibre Channel include a physical interface specification, a word-oriented transmission protocol, and a signalling protocol which includes multiplexing, end-to-end flow control, and error detection. Fibre Channel includes a specification of a generalized multipoint fabric in place of the previous-generation parallel busses. Mapping standards define how various upper-level protocol standards are supported by Fibre Channel.

References

[1] A. X. Widmer and P. A. Franaszek, "A DC-Balanced, Partitioned-Block, 8B/10B Transmission Code," *IBM J. Res. Develop.* Vol. 27, no. 5, pp. 440–451, Sept., 1983.

[2] A. Varma, V. Sahai, and R. Bryant, "Performance Evaluation of a High-Speed Switching System Based on the Fibre-Channel Standard," *Proceedings of the Second IEEE International Symposium on High-Performance Distributed Computing*, Spokane, Washington, July 1993, pp. 144–151.

[3] J. C. Elliott and M. W. Sachs "The IBM Enterprise Systems Connection (ESCON) Architecture," *IBM J. Res. Develop.*, Vol. 36, No. 4, pp. 577–591, July, 1992.

[4] A. Varma, S. Murthy, and R. Bryant, "Using Camp-on to Improve the Performance of a Fibre Channel Switch," *Proceeding of the 18th IEEE conference on Local Computer Networks*, Minneapolis, Minnesota, September 1993, pp. 247–255

[5] W. Zhang, A. Varma, and P. Rupert, "Performance Evaluation of Stacked Connection Scheme in the High-Performance Switching System (HPSS)," Research Report, Computer Engineering, Univerity of California, Santa Cruz, September 1994.

Side Bar: Fibre Channel Standards

The Fibre Channel working group has a number of projects covering different aspects of Fibre Channel. As of September 22, 1995, the results of two of these projects had been issued as ANSI standards and work was in progress on 13 other standards. In addition, several Internet-related projects are in various stages of processing towards becoming Internet standards. Following are the projects.

Completed Standards

FC-PH Fibre Channel Physical and Signalling Interface, American National Standard X3.230-1994

FC-FP Fibre Channel Mapping to HIPPI-FP, American National Standard X3-254-1994

Awaiting Final Approval as ANSI Standards

FC-SB Fibre Channel Single-Byte Command Code Sets Mapping Protocol, Draft Proposed American National Standard X3.271-19xx

FC-AL Fibre Channel Arbitrated Loop, Draft Proposed American National Standard X3.272-19xx

HIPPI-FC High-Performance Parallel Interface Mapping to Fibre Channel Upper Layer Protocols, Draft Proposed American National Standard X3.283-19xx

Work in Progress

FC-PH-2 Enhancements to FC-PH

FC-PH-3 Further enhancements to FC-PH

FC-LE Fibre Channel - Link Encapsulation

FC-FG Fibre Channel Fabric Generic Requirements

FC-SW Fibre Channel Switch Topology and Switch Control

FC-ATM Fibre Channel Mapping to Asynchronous Transfer Mode

FC-GS Fibre Channel Generic Services

FC-GS-2 Enhancements to FC-GS

FC-IG Fibre Channel Implementation Guide

FCP SCSI Fibre Channel Protocol (development of this standard is by the SCSI X3T10 working group)

FC-AE Fibre Channel Avionics Environment

FC-AL-2 Enhancements to FC-AL

Work in Progress on Internet Standards

- IP and ARP on Fibre Channel
- Definitions of Managed Objects for the Fabric Element in Fibre Channel Standard using SMIv2 (Fabric Element MIB)
- Definitions of Managed Objects for the Node in Fibre Channel Standard using SMIv2 (Fibre Channel node MIB)

Side Bar: Related Standards

A naive view of standards would suggest that there should be only one standard to do any given thing. Yet communications-related standards sometimes appear to be proliferating faster than the previous generations of proprietary architectures. In addition, some of these standards, including Fibre Channel, include a sufficient number of options to produce an interoperability nightmare. Further thought suggests that is no more surprising that there are multiple overlapping communications standards than that there are multiple kinds of standard fastener. The richness of the Fibre Channel physical layer is no more unreasonable than is the existence of a large number of sizes and types of standard screws and nuts. Different cost and performance requirements lead to different design points. Yet another reason for multiple overlapping standards is the need to migrate previous generations of products to open standards even if it is not possible to migrate them to the same standard. Here, we discuss several related communications and I/O standards some of which are mentioned elsewhere in this article: High Performance Parallel Interface (HIPPI), Asynchronous Transfer Mode (ATM), Serial Storage Architecture (SSA), and Single-Byte Channel Command Connection Architecture (SBCON), the newest of them.

HIPPI was designed several years ago as a high-speed (100 or 200 Mbytes/s) wide (32 or 64 bits) short-distance (25 m) parallel electrical interface to replace a proprietary supercomputer interface. It includes switching capability. It is in wide use for both I/O and computer-to-computer interconnect for applications involving high-bandwidth transfer of large sets of data, such as found in scientific and technical computing, engineering graphics, and medical imaging. It was eventually augmented with fiber-optic distance extension. The FC-FP standard (mapping of the HIPPI framing protocol onto FC-PH) provides for eventual migration of existing HIPPI applications and, to some extent, system designs to Fibre Channel.

ATM is designed to provide wide-area (indeed, global) transport of voice, video, and traditional computer data (perhaps a good generic term is “encoded data”). However its design is optimized for voice and video, both of which require fairly rigidly isochronous data transfer, and both of which can tolerate some cell loss due to switch congestion. From time to time, consideration is given to using ATM switches for computer I/O interconnect. This is certainly conceivable, though the small (53-byte) cell size makes ATM somewhat less efficient than Fibre Channel in its use of link bandwidth. However, I/O interconnects are usually designed for transmission media with extremely low error rates. While the fiber links used in most ATM networks meet this requirement, the end to end error rate includes ATM cell losses in the switches between the source and destination of a data stream. ATM switches with cell loss rates comparable to the error rates in fiber-optic links a few km long are possible; however it is unclear whether the costs of such switches will be competitive with Fibre Channel or HIPPI switches. Another solution to cell loss which has been proposed is to incorporate credit-based flow control into ATM for optional use with encoded data. In any case, as of the time this article was written, we are not aware of any effort to define mapping of the important I/O protocols onto ATM.

SSA is another serial I/O interconnection architecture currently undergoing standardization. The SSA architecture was first developed by IBM for interconnection of disks to controllers, but was later adopted for standardization by an industry group, and eventually transferred to ANSI X3T10.1. SSA currently provides data rates of 20 and 40 Mbytes/s in a given direction, with higher speeds expected in the future. The architecture defines a physical layer based on copper or fiber links, a transport layer, and a mapping layer for SCSI commands. The physical topology may consist of full-duplex strings, loops, and switched networks. An SSA loop is topologically similar to an FC-AL configuration, but there are significant differences in the protocols used to control access to the loop. SSA employs an elaborate token-based protocol that allows multiple concurrent transmissions in the loop, providing potential throughput improvements over conventional token-based protocols. SSA is primarily targeted at short-distance interconnection of disks and controllers, and the protocols are optimized for short distances.

The SBCON standards activity is both the newest and the oldest of these standards. It was begun by the X3T11 working group late in 1994 in order to address the need, felt by both vendors and customers of the IBM ESCON I/O Interface, to have an open standard based on the ESCON definition and compatible with existing ESCON products. It is the oldest because work on the ESCON I/O Interface was begun within IBM well before any of the other architectures being discussed here. The ESCON I/O Interface (and therefore SBCON) was designed specifically to provide enhancement of both distance and data rate for products originally designed for the prior IBM parallel I/O interface while requiring minimal design change, above the physical layer, to existing products and their software. While it is conceptually possible to use the Fibre Channel FC-SB mapping definition instead of SBCON, this would not meet the need for compatibility with the existing ESCON products since FC-PH is not compatible with the corresponding function in the ESCON I/O Interface. The SBCON working group intends that when major functional enhancement is needed, the vehicle for such enhancement be Fibre Channel and FC-SB.