The Development of ATM Standards and Technology: A Retrospective

The telecommunications and information technology industries currently see asynchronous transfer mode as the next major infrastructure technology. With roots in experimental switching technologies, ATM evolved along lines suggested by standards organizations. Because its services and technology coalesced in such organizations, it has achieved global acceptance. Though now finding use in private and public networks, ATM must still overcome outstanding technical, economic, and regulatory issues before it becomes a major commercial, and therefore standards, success.

Richard Vickers

Northern Telecom

The telecommunications and information technology industries currently see asynchronous transfer mode for broadband telecommunications support of multimedia services as the next major infrastructure technology. Although it had its roots in both technology and standards development, public standards bodies provided the focus for technological development of ATM. Though claims for ATM's commercial success are premature, the current industry momentum behind the technology is immense.

It is interesting to recount how ATM evolved, what challenges developers of its technology and standards faced, and how they addressed these challenges. A number of hurdles remain, including regulatory, tariff, economic, and technical challenges. Paradoxically, ATM might fail in the short term because of its current success; expectations may be too high, and the short-term demands may be greater than the technology can deliver. However, no other technologies appear to match the potential of ATM for supporting the same variety of applications in local-area, wide-area, private, and public networks.

ATM development has passed through three main stages. The first stage started with a service vision, a technology vision, and a standards base. This first stage culminated in the issue of the first framework CCITT (Comité Consultatif International de Téléphonique et Télégraphique) Recommendations in 1988. The core ATM development occurred in industrial and university research laboratories, and in two or three public network standards groups that provided the catalyst for the convergence of protocols. This second stage involved development of a set of implementable base standards. Since completion of the base standards, the technology has moved into the implementation, or third, stage. A much wider community of interest is now addressing both standards and technology: customer-premises equipment (CPE) and terminal equipment manufacturers have joined the original public network carriers, equipment manufacturers, and researchers.

As public network standards bodies carried out the majority of front-end ATM development, our discussion here should start with a survey of the key organizations and their operating methods. Then we can move onto the development of ATM itself. We will look at early activities in terms of standards and technology developments. The coalescence of these activities into a single stream occurred exclusively in the formal standards development process. For the first time, a single global standard for a transport technology arose. Subsequent to the formal standards-setting process, the technology moved into different working environments, and into differ-

0272-1732/93/1200-0062$03.00 © 1993 IEEE

ent applications. End users are just now becoming involved, but here we must carefully divorce the technology from its application. Technology is of little interest to end users; applications butter their bread.

The standards-setting process

The first stage of the standards cycle is the development of the standard. A standard is generally not sufficient to define equipment or an interface. Specifications use standards as key components, and will define equipment and interfaces to a level that product developers can implement them. The specification phase translates as the application of the standard. The final responsibility of the standard is to ensure interoperability between different implementations of the same standard.

Standards development. In the United States, Committee T1, an American National Standards Institute-accredited body, generates public network telecommunications standards. In Europe, the European Telecommunications Standards Institute develops related standards. Other countries also have national standards bodies. Internationally, the standards take shape as Recommendations by the Telecommunications Standardization Sector of the International Telecommunications Union (ITU-T—formerly CCITT), the ITU being a treaty organization between the national governments of the member countries.

Each of these standards bodies works on the development of standards based on written contributions against Project Proposals (T1) or Study Questions (ITU-T), reaching agreement on a consensus basis. A vote of the members tests the formal consensus. The consensus process does not necessarily mean unanimity in the agreement of the standard; it does ensure that any dissenters have their objections heard and discussed in open forum. In practice, most organizations work diligently to clear comments and objections before issuing the standard. In Committee T1, for example, external ANSI review follows two stages of balloting by members and comment resolution.

Although in the past each of the national or regional organizations has tended to produce independent national standards, a trend to place the standards development effort into the international forum has now developed. National standards refer directly to the international recommendations together with any national enhancements and option selections. Using the ITU-T Recommendations ensures that a designer does not face ambiguity because two different texts describe the same requirement, and also avoids unnecessary duplication of editorial effort in the development of the text.

The national or regional standards bodies perform most of the base technical work before passing it to the ITU by means of written contributions. The transfer is not generally direct, because the initiating standards body is usually a private-sector organization, but delegations to the ITU are responsible to their national governments or telecommunications administrations. Before they can be considered as representing the national position, generated contributions to the ITU must be endorsed by the appropriate committee, with the endorsement process generally performed in national ITU bodies.

The final responsibility of the standard is to ensure interoperability between different implementations of the same standard.

For example, in the United States, contributions generated by Committee T1 go to the appropriate US Study Group. The US Study Groups have the delegated authority of the Department of State to establish US positions. Individual organizations can have membership in the ITU either in the Recognized Operating Administration category if they are a carrier, or Scientific and Industrial Organization if the organization is research oriented or a manufacturer. Although the process seems unwieldly, it is mitigated by the fact that most delegates attending the ITU also work in national standards development.

At the beginning of the standardization activity on ATM, single working groups in both Committee T1, ETSI, and CCITT undertook most of the work. With diversification of the technical work, the formal standards development process has expanded from the original CCITT Task Group and Committee T1 Subworking Group until it now spans a number of study groups and T1 working groups. This diversification, however, introduced a communications problem between the various activities. Both the ITU and Committee T1 have a formal process of liaisons between the various groups. The formal communication proceeds with Liaison Letters between the separate groups. A Liaison Rapporteur, who generally attends the meetings of the two or more groups involved in the liaison, usually supplements these letters.

Standards application. The individual standards developed by the standards bodies are insufficient in themselves for a designer to build equipment. The published standards provide the essential components for use in developing formal specifications. To build equipment, the designer must select an appropriate suite of standards and tailor it to the application. The application of the standard to an equipment specification may require some optimization, such as elimi-

The core standards activities are diversifying from the original development of the ATM protocol, to cover items such as signaling, network management, and equipment specification.

nation of options, and perhaps some customization to provide features not envisaged when the standard was originally developed. The formal standards development bodies do perform some applications work. For example, the protocols are generally developed as a coherent stack. Much of the work of specification, though, takes place outside the formal standards development process.

To move the standards to the implementation and specification stage, a new body entered the arena of ATM—the ATM Forum. Founded in October 1991 by Northern Telecom, Sprint, Cisco, and Adaptive, it is developing implementation agreements, which provide the selection of standards and options within standards to form a particular interface. The ATM Forum also has a stronger customer-premises equipment vendor community than the formal standards bodies, and is extending the formal standards to encompass the requirements of that community.

The ATM Forum reaches decisions by a majority vote, rather than consensus. This procedure allows a more rapid closure of issues, but it does risk alienating parts of its membership. Other than the ATM Forum, specifications tend to be customer-specific. For Regional Bell Operating Companies, Bellcore provides the development of specifications in the form of Technical Requirements and Technical Advisories. Although they have been developed for the RBOCs, these documents will probably form the basis of requirements for most US carriers.

Conformance and interoperability. We achieve interoperability when different implementations of the same specification correctly provide the intended service or application. This requires that individual protocols will work with the protocols above and below in the protocol stack. In reality, interoperability comes in two phases. First, conformance testing measures an implemented protocol against the standard. Second, interoperability testing checks that when two implementations of the protocol interwork, the system as a whole meets its functional requirements. Currently, the interoperability aspects of standardization comprise a relatively minor portion of the total standards effort. However, conformance and interoperability will become increasingly important issues as more vendors enter the arenas of ATM and B-ISDN (Broadband Integrated Services Digital Network).

Many standards under development now contain a protocol implementation conformance statement (PICS) pro forma. The PICS consists of a checklist of requirements (both mandatory and optional) contained within the standard. A manufacturer can indicate the level of conformance of an implementation by using the PICS pro forma and checking off requirements on an item-by-item basis. The result is the PICS for a particular implementation.

Complementing the PICS pro forma is an abstract test suite, consisting of a set of test cases that rigorously check the behavior of a protocol implementation through all possible state transitions of the protocol. The abstract test suite has traditionally been generated outside the formal standards development bodies, but recently the International Standards Organization/International Electrotechnical Committee (ISO/IEC) and the ITU-T have been developing abstract test suites for some protocols. Interoperability testing tends to be less rigorous that conformance testing. Typically, it involves observing the behavior of two or more interconnected implementations. Generally in interoperability testing, we test the normal operating conditions of the protocol, but not all of the error legs. Conformance testing and interoperability testing are complementary steps in ensuring interoperability. The ATM Forum has recently established a Testing Subworking Group to address interoperability as part of its mandate, and is currently defining its work plan.

Other ATM activities. There are other bodies with an interest in ATM. For example, the Internet Engineering Task Force is working on the use of transmission control protocol/Internet protocol over ATM. There is also work in the International Standards Organization on ATM rings. ATM has now moved beyond the formal standards-setting process to the applications over ATM. The core standards activities are diversifying from the original development of the ATM protocol, to cover items such as signaling, network management, and equipment specification.

"Fast" switching—the technology roots

In the early 1980s, two basic digital switching technologies existed: 64-Kbps circuit switching and X.25 packet switching. The circuit switches were actually voiceband switches that used digital crosspoints; the idea of end-to-end digital circuit switched service was under development, but not yet a reality. Only a very limited number of such switches were in service in the 1980s. X.25 was a packet switching technology for low-speed data applications. It used heavyweight, link-by-link flow control and retransmission protocols to avoid loss from congestion and to recover from errors induced by the underlying transmission systems of the time, primarily copper-based with some digital radio systems. The flow control procedures allowed operation at high network efficiencies on the basis that the cost of transmission was high. X.25-based networks provided very high quality and reliability of data transfer. While they still fill a vital need in the business of reliable data transfer, their application is limited to low-speed data.

The telecommunications industry had traditionally introduced new services to operate over a voiceband channel, and had been quite successful in doing so. X.25 required an overlay network approach; dedicated transmission, switching, and operations support had to be provided. The network components were expensive, and the traffic volumes small, with the result that X.25 networks struggled for a long time to become profitable.

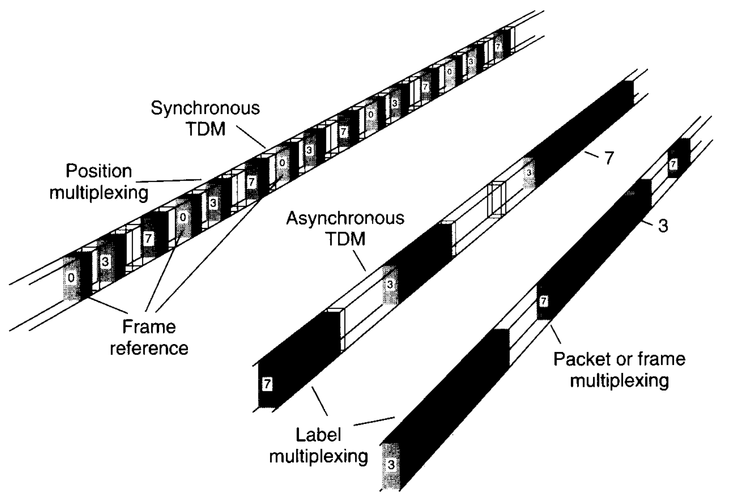


Figure 1. Multiplexing methods. Numbers indicate channel or virtual channel.

The poor economics associated with providing overlay networks for each new service offering clearly pointed to the requirement for a different approach. The approach now taken is to provide a single switching technique and network capable of supporting a wide variety of services. The commercial success of the network would not then hinge on the commercial success of a single service. The research of the early 1980s therefore aimed at developing much more versatile switching techniques than either the conventional circuit switch, or the X.25 packet switch, since its ultimate objective was a single, or integrated, network for all services.

Three different techniques were under development at the time: fast circuit switching, fast packet switching, and asynchronous TDM. These techniques generally relied on some form of preprocessing for routing and in-band labels associated with the data for switching. The three systems differed in their method of multiplexing. Fast circuit switching used conventional TDM position multiplexing. Fast packet switching used variable length frames carried on virtual channels. The header of each frame carried virtual channel identifiers. Asynchronous TDM was a hybrid of the two previous schemes. It used fixed-length frames on virtual channels, and carried labels in the header of each frame. Figure 1 illustrates the different multiplexing methods.

All three of these techniques minimized the processing required for switching. Fast circuit switching retained the characteristics of conventional synchronous TDM switching systems in that the bandwidth of a channel was fixed, resulting in a lack of flexibility to accommodate different source bit rates. For this reason the technology was not pursued, and especially because most sources (including voice and video) do not intrinsically generate constant bit rates.

Researchers developed the fast packet technology, but for the transfer of medium-speed data only, rather than for the mix of voice and data. This development produced frame relay, now deployed by a number of carriers. Frame relay will likely serve as an intermediate technology to support applications such as LAN-to-LAN traffic before full ATM service becomes available. Somewhat ironically, current plans include using ATM to support frame relay.

As asynchronous TDM development emerged, it became known as ATM. The early versions were somewhat primitive, and required considerable development, but all the basic principles were in place. The label multiplexing and the fixed-length packets, now called cells, were all components of the original asynchronous TDM technology.

These experiments set the technology base for what was to be developed in the standards. They established the viability of using hardware and in-band headers for controlling switching. The technologies themselves were experimental, either in the form of laboratory prototypes, or limited field trials, but they did provide one of the cornerstones for what was to follow in the development of ATM.

Early standardization efforts

The major standardization effort of the early part of the

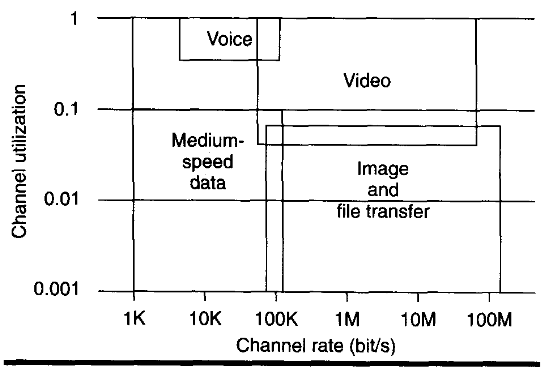


Figure 2. Bandwidth/channel utilization characteristics.

1980s was on the ISDN. Intended to provide an end-to-end digital service, both circuit switched (64 Kbps) and X.25 packet switched, ISDN concentrated on integrated access from the transport perspective, but based on existing digital switching technologies. It made no attempt to coalesce the underlying switching structures.

In 1985, industry recognized that new services would emerge that would require transport capabilities beyond those supported by the ISDN then under development. The term B-ISDN then came into existence, meaning broadband aspects of the ISDN. Note that B-ISDN was considered a part of the ISDN.

B-ISDN split from the mainstream of ISDN development in 1986; both CCITT and ANSI created separate subgroups. In the longer term, this separation of the work allowed for a less constrained development of the standards.

During the early phases of the standards development, most of the work concentrated on the services that would use B-ISDN, but with some discussion of the supporting transport technologies. The issue of the first CCITT B-ISDN recommendation in 1988[1] marked the completion of the early phase of development. This recommendation set the framework for subsequent developments.

B-ISDN is an entire network concept. It encompasses transport, multiplexing, and signaling, as well as operations, administration, and maintenance.

Transport. By the early 1980s, we had optimized the transmission hierarchy around the capabilities of copper and radio-based systems with limited bandwidth capabilities. We generally timed transmission systems from local free-running clocks. To accommodate timing differences, we performed multiplexing above the primary rate (DS1—1.544 Mbps) asynchronously by bit stuffing techniques. At the same time, optical fiber systems were starting their deployment in volume. A single optical fiber has enormous intrinsic bandwidth, but the transmission hierarchy of the time was designed to handle relatively low bandwidth systems, and was really not suited to efficiently manage the large bandwidths available with optical fiber systems.

To extract the most from fiber transmission, researchers developed a new synchronous transmission format. In North America it is known as Sonet (synchronous optical network) and in Europe as SDH (synchronous digital hierarchy). A detailed technical discussion of Sonet/SDH is beyond our scope here, but it possesses two extremely important properties for B-ISDN. First, bandwidth can be easily concatenated to form very large channels. Second, it provides a transport capability and common management platform that is independent of the format of the payload. Both these properties are important for B-ISDN and provide a strong transport base.

Transfer modes and multiplexing. Sonet provided the transport platform suitable for carrying broadband services. Work in standards now concentrated on the development of suitable multiplexing methods for carrying a number of services with different bandwidth and other characteristics (see Figure 2). Three modes of transport were under discussion: synchronous transfer mode, ATM, and packet transfer mode. STM is the traditional digital voice transfer mode, where the transfer takes place with fixed blocks of information at fixed intervals. Packet transfer mode proceeds using variable-size blocks of information at varying intervals. ATM uses fixed blocks of information at variable intervals.

Complementary to the discussions on transfer mode were the talks on multiplexing. Because of the difficulties of carrying constant bit rate information and the asynchronous nature of label processing in such systems, pure packet multiplexing eliminated itself fairly quickly. Hybrid circuit/packet structures and asynchronous TDM remained for considerations. The proponents of the hybrid system based their arguments on compatibility and interworking with existing systems, whereas the proponents of asynchronous TDM cited the flexibility of the system and its ability to automatically adapt to variations in traffic mix.

After much debate during 1986, the standards bodies agreed to focus on studies and standardization of ATM, and the standardization of asynchronous TDM as the multiplexing technique followed almost automatically.[1] The reasons for the decision were complex, but one major advantage was that we could now employ a single multiplexing layer for all services on top of the common transport layer of Sonet/SDH, thus simplifying layer management functions. Note also that the overall ATM system efficiency will generally be greater for a mix of services than a circuit-switched system despite the added overhead because of ATM's ability to rate adapt the transport to the requirements of the source. The ability to statistically multiplex the resultant variable bit rate sources may further enhance ATM's efficiency.

Other standardization activities. Two other major standardization activities occurred through the 1980s: fiber distributed data interface LAN activities (in ANSI X3T9) and the metropolitan area network activities (in IEEE P802.6).

The Fiber Digital Data Interface project was initiated in the early 1980s to provide an optical fiber-based LAN at 100 Mbps. Envisioned at its inception as the interconnection of mainframe and peripherals, the standard was fully developed and products are now available. FDDI now tends to serve as a backbone LAN rather than the original, envisaged application. Developed purely for data applications, a subgroup of FDDI spun off early in the development cycle to look at the carriage of isochronous services. This subgroup defined a new signal structure (FDDI-II), where we could divide the 100-Mbps bandwidth between data and isochronous applications. This hybrid packet/circuit structure had obvious application in multimedia LANs, but in retrospect the timing was wrong, predating the market requirements for multimedia LANs by some years.

FDDI is deployed in LANs today. In addition, some of the technology developed for FDDI is finding application in other areas. Of interest to ATM, the ATM Forum has defined an ATM user-to-network interface for customer premises applications using the physical layer of FDDI.

Activity started for the MAN standard in the early 1980s, but languished for some time, until in 1987 a protocol (distributed queue dual bus—DQDB) was proposed that had adequate bandwidth/distance characteristics for MAN applications. Also, its format closely resembled the format under development for B-ISDN, in that it was based on fixed-length packets, or cells. Because the format was so similar to the ATM cell format, investigators expended considerable effort to maintain commonality between the cell sizes in ATM and IEEE P802.6. Work on the basic IEEE P802.6 standard concluded in 1990.[2]

IEEE P802.6, like FDDI-II, has an isochronous capability. In IEEE P802.6, cells placed at fixed intervals provide the isochronous capability, giving it built-in multimedia capabilities. IEEE P802.6 has major importance in the transition of B-ISDN services. We use it as the access technology to support the switched multimegabit data service defined by Bellcore. SMDS is a broadband connectionless data service with access rates initially at the DS1 (1.544 Mbps) and DS3 (44.736 Mbps) rates. Initially, we thought that the service would be supported on an overlay IEEE P802.6 network; more recently the network support for the service has moved to an ATM base.

Development of the ATM standard

Once the decision was made to use ATM during 1986, the parameters of the ATM cell had to be standardized, which led to protracted and heated discussions in the standards forums. The North American position centered on a 64-octet cell payload; the European position centered on a 32-octet payload.

Once the decision was made to use ATM during 1986, the parameters of the ATM cell had to be standardized, which led to some protracted and heated discussions in the standards forums.

In North America, the IEEE P802.6 committee had also adopted a cell-based format for MANs, as well as a particular format of a 64-octet payload and a 5-octet header. There was also SMDS to consider. SMDS is a datagram service that routes on an E.164 address[3] embedded in the header. The header itself could fit into one cell (32 or 64 octets), but the routing decision had to be made very quickly. Processing the header in one cell time would eliminate queuing of headers. With a 64-octet payload, observers felt that there was sufficient time to perform the processing, but 32 octets left insufficient time. In addition, a 32-octet payload with a 5-octet header would result in a maximum efficiency of 86 percent, considered too low, especially when combined with the efficiency of Sonet/SDH, and detraction of the 32-octet payload by adaptation layer functions.

Europeans made a strong push for a total integration of services, including voice. For performance reasons the delay requirements on voice are quite stringent. At 64 Kbps it would take 8 ms to fill a 64 octet cell, and 8 ms from the European perspective was unacceptable, as it would mean major investments in the deployment of echo control devices to counter the additional delay. Europe was therefore proposing a cell size of 32 octets.

After much debate, the various parties agreed upon a cell consisting of a 48-octet payload and a 5-byte header in June 1989. The selection of a common cell was vital to the development of ATM. The adoption of different cell sizes in different regions of the world would have prevented interworking at the ATM layer. Instead, the signals would need reconstruction in their original form (64 Kbps, data packets) and redivision into the different cell structure of the receiving country. This form of interworking would have to be performed on a service-by-service basis, and in the long term would have severely constrained the development of international ATM networks. The compromise cell size is

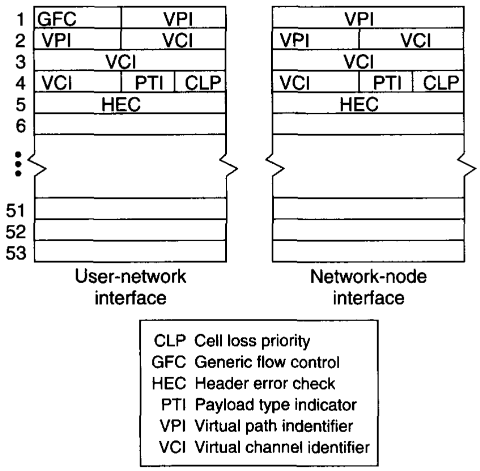


Figure 3. ATM cell formats.

probably not ideal, but the need for international compatibility outweighed the technical arguments put forward for different cell sizes. The IEEE P802.6 Committee subsequently modified its cell size to align with the 48-octet payload.

The parties also agreed on mapping of ATM into the 155-Mbps Sonet/SDH payloads. The combination of these mappings, and a common cell format meant that for the first time we had a common transport format that we could apply in all regions of the globe.

Conferees in subsequent meetings also accepted and developed other functions and fields in the header, until the final header and cell formats (Figure 3) were incorporated into the 1992 versions of the CCITT ATM layer recommendations.[4,5] The formats for use at the user-to-network interface and for use internal to the network (network node interface) differ in one significant detail. At the UNI the header dedicates four of its bits to a function called generic flow control. At the NNI it allocates these bits to the virtual path identifier.

It is interesting in looking at these header structures to examine the developments from the primitive two-octet header implementations of some early ATM applications. These developments have all arisen because of the public technical debates that occur in the standards forums, and in every sense make ATM a more viable and flexible technology.

The essence of the header is in the routing field. The routing field is the label, and provides the association of the cell with a particular channel or circuit. It is simply a number that has significance only on the local link. As the cell traverses a switching point the number will generally change, just as in a synchronous TDM system the timeslot number will generally change as the channel traverses a switch.

The routing field is divided into two portions: the VPI and the virtual channel identifier. The VPI allows us to group VCIs and switch the group as an entity. Grouped switching, which operates as the ATM equivalent of trunk grouping, was initally proposed by Australia and Japan. It also gives users access to blocks of channels, for example, to interconnect ATM PBXs. The VCI is the single-channel identifier, and thus identifies the basic switched channel.

Agreement on the size of the routing field proved elusive. The size of the routing field dictates the maximum number of virtual channels that the same multiplexed interface can support simultaneously. Again, the sizes finally selected represented a compromise. On one hand, some emerging applications of virtual path would require large numbers of virtual paths both within the network and at the UNI. On the other hand, the routing field would need to support a large VCI field for the general user of B-ISDN. The compromise probably fully meets the needs of both groups in practice. The maximum number of virtual channels that the interface can support is 16 million at the UNI and 256 million at the NNI.

The second field is the payload type indicator. Early developmental stages showed the ATM protocol would need additional capabilities to provide for operational functions not related to the user information. Unlike statistical time-division multiplexing, a virtual channel may have a very intermittent flow of user information. To check whether the channel is still viable during a period of inactivity, we could insert a dummy cell to check that it is transported over the entire virtual connection. Although it would carry the same VPI/VCI as a cell carrying user data, *it is not actually carrying* user data; the payload consists of information used by the network in checking the viability of the channel. The PTI allows that cell to be extracted by the monitoring function from the user data stream; if it were interpreted as user data it would cause an error.

We have found other uses for the PTI, which has been extended from the two bits originally allocated to three bits. The functions now include an in-channel multiplexing identifier and a congestion indication (both for user data), two types of operations administration and maintainence cell, and a resource management cell (for network use).

The last functional bit in the header is the cell loss priority bit. This bit provides the network with a selective discard capability. The user may mark cells high or low priority. In some cases, however, the network may change the value of the CLP bit.

The final field in the header is the header error check field. Added to prevent errors in the header that cause a misrouting of the cell into another user's data stream, this field contains the result of the CRC-8 calculation across the previous four octets. When a switch or terminal terminates the header, errors will be detected with a high probability. If the error is a single-bit one, it may be corrected. If the error is not correctable, the whole cell must be discarded. Considerable work on the theoretical characteristics of the error check algorithms led to agreement on the detection and correction capabilities, with the results being mapped onto the measured performance of optical fiber transmission systems in the field. The ability to correct single bit errors improves system performance, but with a slight degradation of the error detection capability. Note, though, that certain types of transmission systems use block coding schemes. Such schemes introduce error multiplication. If we use ATM over such transmission systems, we should use the header error check only in detection mode.

The GFC field at the UNI is the only field where we have not fully defined the functions and protocol. The prime purpose is to provide arbitration between contending terminals on the same interface. Standardization of GFC should be completed sometime in 1994. Note that for a simple uncontrolled interface, the GFC field is not used. Therefore, absence of complete encoding does not impede the deployment of ATM.

That completes a brief description of the ATM layer fields and functions. With the exception of the GFC, the ATM layer standards were completed in 1992. ATM is a compromise technology, and as such it retains some of the features of the parent technologies and loses others. As a result, to carry specific services over ATM, we must restore certain functions to the overall protocol stack before the service can be carried. The adaptation layers carry these additional functions.

Adaptation layers

By design, the transmission and multiplexing functions are largely service generic; that is, these layers are not customized for service-specific functions. The adaptation layers provide the functions that enable specific services to be carried. While a full description on the adaptation layers is beyond our scope here, a basic understanding of the type of functions they provide will be helpful. For a full description, see the CCITT recommendations I.362[6] and the protocols described in I.363.[7]

The early work on services established classifications based on various service characteristics (Figure 4). These service characteristics identified requirements that the ATM layer did not necessarily provide. To provide these characteristics to the services, we identified four adaptation layer types, labeled Types 1 to 4:

	Class A	Class B	Class C	Class D
Timing between source and destination	Required		Not required	
Bit rate	Constant	Variable		
Connection mode	Connection-oriented			Connection-less

Figure 4. Classification of services.

- Type 1 adaptation for constant bit rate services (Class A). requires provision for reconstructing the service clock at the destination. We have developed a protocol for Type 1 adaptation. This protocol will serve initially in the carriage of services that emulate the performance of an existing constant bit rate service, such as DS1. Providing the adaptation layer functions requires one of the 48 payload octets in each cell.
- Type 2 adaptation is for variable bit rate services where the timing relationship between the cells is significant to the service (Class B). One service that will require this type of adaptation is variable bit rate video. However, we have not yet defined VBR video services, and the adaptation layer protocol is not yet required.
- Type 3 adaptation is for VBR connection-oriented data, where the timing relationship between cells is not important (Class C). It provides for the basic segmentation of variable length frames into cells at the sending end, and the reassembly of the frame at the receiving end. It also has error checking and a multiplexing capability built into each cell. The adaptation layer functions consume four of the 48 payload octets.
- Type 4 adaptation is for connectionless data (Class D), and is almost identical to Type 3. The only difference is in the use of the multiplexing identifier. Type 3 assigns a value for the duration of the connection whereas Type 4 assigns the value only for the duration of the frame.

Subsequent to the identification of these four adaptation layer types, a new simpler adaptation layer emerged for connection-oriented VBR data service. The Type 3/4 adaptation layer requires that four octets of information be detracted from the cell payload to carry necessary information from transmitter to receiver. The Type 5 adaptation layer does not use any of the 48-octet payload for adaptation, but instead requires an identifier in the ATM header, and uses the in-channel multiplexing identifier function of the payload type field. Type 5 provides only segmentation and reassembly at the cell level. It provides all other functions, such as length checking and error detection, on a frame basis.

The introduction of Type 5 has virtually made Type 3 obsolete. Services such as signaling and frame relay have moved from Type 3 support to Type 5. Type 5 represented a challenge to the standards community in two ways. Type 5 originated in the customer-premises equipment industry, a group

The thrust of standards development has moved away from ATM itself to other standards required for full B–ISDN capabilities.

that works on much shorter time scales than the public network industry. The challenge was to accept a proposal from a different segment of the industry and complete the standardization much more quickly than is normal for such standards. The standards community met most of these challenges, and Type 5 AAL achieved international agreement within eighteen months of its first appearing in Committee T1.

The work on adaptation layers will continue for some time. As new applications emerge, we will likely define new adaptation layers. However, the development of new adaptation layers will have minimum impact on public network facilities. Since adaptation does not form part of the base transport, we implement it in the CPE or in network servers. The core ATM transport will not be affected.

Developments since 1992

With completion of the base standards in 1992, ATM entered the implementation phase. Until then public network operators and equipment had almost exclusively developed the standards. They had also been developed to a fairly relaxed timetable, as there was no real commitment (apart from trials) to the implementation of networks based on the standard.

During 1991, some organizations in the business of producing high-end workstations and other CPE equipment realized that ATM technology could fill some of their needs. Workstation I/O capabilities were reaching 40 Mbps. At those rates the then current state-of-the-art LAN (FDDI) could handle only a total bandwidth of 100 Mbps, a capacity that would not be suitable for networking such high-speed applications. Multimedia was also emerging, with trials of multimedia editing workstations operating on voice, video, graphics, and text files. The development environment for ATM standards had very quickly become much more pragmatic.

Most immediately, the CPE industry needed Implementation Agreements for interfaces to terminals and ATM switches. Such agreements would provide a complete definition of an interface, including operations, management, and signaling. Standards provided the building blocks and some guidelines on protocol stacks to support certain applications.

Announced in October 1991 with the specific charter of producing such Implementation Agreements, the ATM Forum has issued such agreements on the UNI.[8] Besides using formal standards, in certain instances it has also filled specific needs by developing extensions to the standards (such as for multipoint connection capabilities) or new technologies where no standards exist (such as ATM over unshielded twisted-pair cable). Eventually, some of these developments should be fed back into the standards process.

Formal standards development since 1992 has also become more pragmatic. Recognizing that 150 Mbps could not be economically deployed to all customer premises in the near term, the standards community is developing interfaces at 44 Mbps and 1.5 Mbps, based on the old, but widely available asynchronous transmission formats. The thrust of standards development has moved away from ATM itself to other standards required for full B-ISDN capabilities: signaling, OAM, and interworking with other networks.

The standards-setting process

The development of public network standards has traditionally been slow. The ITU used to work on a four-year cycle for issuing new recommendations. In periods of slow technological change this four-year cycle was adequate, but the current environment, with topics such as ATM and related standards, requires a much faster development cycle. The standards community is establishing a number of processes to aid this more rapid development.

The ITU has decoupled the approval of Recommendations from the four-year study period, and now has a process to generate new Study Questions during the study period. Its formal approval process is still slow, taking about a year from the completion of the text to formal approval of the standard, while publishing the Recommendations adds more time. The ITU is now developing electronic publishing and document handling capabilities, which should speed the publishing process.

The need to publish the formal documents in three languages is another expensive and time-consuming processing issue within the ITU. Currently, people perform the translations, with technical experts reviewing the translated text for consistency with the original. Before long, computers may be able to perform the initial translation, although the technical review of accuracy will still have to be manual. Again, this will speed the standards development cycle.

In the early days of ATM standards, progress was slow, due largely to misunderstandings between the technical experts of different countries that resulted in entrenched positions and dogma. The debating environment was formal and was not conducive to informal technical discussions between participants. To provide a more conducive environment, representatives of Committee T1, ETSI, and TTC (Telecommunications Technology Committee) of Japan agreed that an annual B-ISDN Technical Workshop should be established. Held in 1991, the first workshop discussed specific issues in an informal atmosphere to help eliminate conflict resulting from technical misunderstanding. In this objective, the workshop has been extremely successful.

Standards, implementation, and interoperability. The underlying purpose of standards is to ensure that when various standardized components are brought together that they interoperate. Standards also should define boundary conditions, but are not intended to constrain implementation. However, as protocols become more complex, interoperability becomes more difficult to write into the standard. There are a number of reasons why this is occurring.

1. *The use of prose text in standards.* Using prose text makes it very difficult to define a function precisely in such a way that it can be interpreted unambiguously, or such that two different implementations of the same function will interwork.
2. *The definition of options within standards.* Options may be included in standards for a number of reasons, some technical, some political. For whatever reasons they exist, options can lead to confusion and incorrect implementations.
3. *Errors of omission.* Prior to implementation, verifying the completeness of a standard is difficult. When the standard is implemented, different designers may choose to implement the missing parts in different ways, leading to interoperability problems.

The use of prose text in standards definition is traditional. However, the inexact meanings of prose text cause problems, even before standards are issued. Standards developers spend much time debating whether the wording reflects the intent of the standard, and whether it is unambiguous. But even agreement may not eliminate all problems. For a direct comparison, look at the law, where even more precise meanings to language do not entirely eliminate problems of ambiguity and interpretation. Therefore, case law must supplement legislation to provide judicial interpretation of the legislation.

In the field of standards, replacing prose description with formal description techniques has helped. FDTs provide a method of specifying protocols or behavior unambiguously and precisely. Development of FDTs has proceeded for a number of years, and in the case of SDL (functional specification and description language) the technique has been used in standards for some time. However, the SDL portion of the standards is only informative, one reason being that SDL is not sufficiently abstract to be independent of implementation. Also, from an accuracy viewpoint, we have not had the tools to debug the SDL as written.

SDL is just like a computer language; specifications written in SDL must be verified for semantics and syntax, then run to check that the protocol as defined meets its original requirements. Only recently have tools to perform these tasks become available on the open market. SDL is only one of several FDTs under development, but it is certainly the most mature and widely used. Other FDTs may be more independent of implementation, but are not as easy to use, nor do the tool sets exist to assist the designer.

If FDTs are to form the normative part of a standard, the standards community will have to reconsider the sanctity of the principle that standards should not constrain implementation. One possibility is the use of the tool in reference implementations. Even though a designer would have freedom of implementation, the validity of the design would always be judged against the reference implementation.

The standards development bodies are just starting to work on conformance and interoperability. As systems become more complex, conformance and interoperability will have to form a much more important role in the standards development process. The standards bodies cannot claim to be doing a complete job until they recognize the importance of these topics.

The future of ATM

All segments of the telecommunications industry now believe that ATM will be the transport technology for cost-effectively supporting multimedia services. However, before it fulfills its promise as the ubiquitous transport for broadband services, it must still face a number of challenges, some technical, but some legal, regulatory, and business related. These factors will determine where the technology has its first applications, and how it will develop.

Technical issues. The standards completed to date are sufficient for the implementation of basic services on ATM. For certain types of service, though, the network efficiency could be quite low, and these are the highly bursty data services. Within the network, using large trunk cross sections (600-2,400 Mbps) will allow the law of large numbers to come to bear for statistical multiplexing of services with peak bit rates up to a few tens of Mbps. The local-area and private networks, however, generally have an insufficient volume of users or traffic to use the law of large numbers; such networks are frequently, indeed routinely, congested. In local networks, therefore, media access control protocols form the basis of the LANs such as Ethernet (CSMA/CD), Token Bus, and Token Ring. Such protocols work in LANs because the geographical distance is limited. More important, the LAN effectively forms a single-stage switch, making contention an easier problem to solve than in a large, wide-area network. Such protocols effectively share the available bandwidth among the users during times of congestion, rather than discarding excess traffic at the expense of degraded throughput performance. Currently defined

ATM protocols would be forced to discard should congestion occur.

The challenge is to extend the same service provided by a LAN in the local area across the wide area. Today, we interconnect LANs across a wide area by using routers attached to the LANs and interconnected by fixed low-bandwidth facilities. The same method applies to an ATM network, although the ATM network can provide higher bandwidth pipes. The evolution of the network will see provision of LAN-like capabilities provided on a wide area, ubiquitously available basis. Standards bodies are currently discussing a number of mechanisms, including Fast Reservation, link-by-link flow control, and backward congestion notification.

Other major technical areas under debate include signaling and network management. ITU-T Study Group 11 is developing signaling protocol. Signaling work focuses on the capabilities that B-ISDN will provide to meet marketplace requirements. The group has scheduled two releases:

- Capability Set 1, scheduled for completion by the end of 1993, provides for basic call establishment for point-to-point connections; and
- Capability Set 2, scheduled for completion at the end of 1994, will provide for limited multipoint and multimedia capabilities.

The study group has identified some capabilities for Capability Set 3, but has not yet defined its completion. Capability Set 3 will provide for more sophisticated multimedia and multipoint capabilities.

Development of network management is progressing along two streams. For private networks, the customer network management in North America is evolving from the Simple Network Management Protocol defined by the Internet Engineering Task Force. In the public network, the Network Management Protocols build on OSI, the ISO/ITU-T protocols stack. In Europe, many administrations favor the OSI approach for customer network management.

The standards for deploying a basic ATM network are complete. Such a basic network would have only virtual private line capability, and would not be capable of a high degree of statistical multiplexing. The ATM Forum has developed some extensions to the formal standards to allow for switched service, and a limited amount of statistical multiplexing that will assist in getting started with some more sophisticated techniques.

Signaling developments will enrich the variety of services provided by B-ISDN, and improvements in traffic management techniques will improve the efficiency of the network. Still, network operators will need a period of adjustment as they learn how to manage ATM networks and how to tailor them to increase their efficiency and utilization. We should not overextend the ATM technology in the early days.

Economics and tariff. ATM technology operates better as the bit rate of the services gets faster, and as the number of users increases. Some issues associated with traffic management manifest themselves in start-up costs for ATM networks, particularly in the public network environment.

The cost of the equipment needed to support the service dominates the cost of ATM in customer-premises applications, and the cost and performance benefits of ATM over competing technologies will determine its success there. ATM has a start-up cost disadvantage, as even in the smaller sizes, the switches tend to have switching capacities in the gigabit-per-second region, compared with the 100 Mbps of FDDI. Even if the cost per bit per second of ATM may be a fraction of the cost of FDDI, initially only a few users networking top-end workstations could justify the ATM capacities. Consequently, ATM is likely to start as a backbone technology in applications where FDDI does not provide sufficient capacity.

For connection to the desktop, ATM has a lot of competition. Its success will depend on the development of applications that really require the tens or hundreds of megabits per second that ATM can provide. The 10 Mbps provided by a dedicated or switched Ethernet connection could provide sufficient capacity for many applications, including video, and at much lower cost than ATM. However, another model states that applications will always grow to consume the available bandwidth—that is how the personal computer industry has made new demands on processing speed and storage requirements.

In public network applications the economics of ATM are more complex. The capacity of public network ATM switches in trunking applications will start at about 10 Gbps. Compared to the total data traffic carried by the network, this capacity is enormous. The interconnecting trunks will also be very large, 150 and 600 Mbps being the internationally agreed network bit rates. As with the customer premise applications, public network applications must also address a start-up issue.

The cost of public network service to the end-user manifests itself in the tariff. A full discussion of tariff strategy is beyond the scope of this article, but certain aspects of tariffs can encourage or discourage certain behavior from the user of the network. For example, within the wide area network, it is more efficient now to shape bursty traffic to a fixed peak bandwidth than to pass it through the network unshaped. For items that do not depend on network delay, we can structure the tariff to encourage the end user to shape traffic for the network to handle it in the most efficient way.

In both public and private network applications of ATM, as with any technology, the economic factors will determine the ultimate market position. The cost of the technology must be such that sufficient volume deployment arises to justify manufacturers making major investments in cost reduction. Under these circumstances, ATM could see the levels of growth and demand exhibited by the personal computer. With the current level of industry momentum behind it, ATM is unlikely to fail completely. However, if the cost is too high, it could remain a niche technology, much as X.25 has remained a niche network technology.

Regulation, legislation, and market trends. Historically, the local exchange carriers have operated as a monopoly under Public Utility Commission regulation and federal legislation. Under the law they cannot generally provide protocol conversion, and they are restricted in the types of service they can provide. However, their monopoly now faces challenges on a number of fronts: alternate access carriers for business services, CATV operators for residential services, and cellular radio carriers for mobile services. The regulation and legislation are now becoming major constraints to competition to those companies that operate under its umbrella. To counter the threat, the LECs have achieved some success in challenging the legislation prohibiting them from providing entertainment video services in their own operating regions. They are also obtaining more and more waivers to the protocol translation prohibitions. The whole industry is moving to a competitive basis from a regulated basis, just as the interexchange business started to transition in the early 1980s.

ATM can only benefit from such competition. As a technology suitable for supporting the richness of options that service providers will want to offer to retain their competitive edge, ATM use will certainly facilitate the generation of competition in residential applications. Its application to the public network-owned local loop and CATV delivery will provide a common platform on which to compete for services.

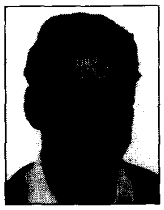
ATM IS IN THE EARLY IMPLEMENTATION PHASE. As we gain experience from their implementation, standards should be extended and perhaps amended. The long-term commercial success of the technology will depend not only on its continued technical development, but also on economic, regulatory, and market situations. Optimistically, ATM could trigger an appetite for bandwidth growth analogous to the appetite for computing power on the desktop. Pessimistically, ATM will remain a niche technology for high-end computing and multimedia communications. With the momentum that exists behind the development of ATM and ATM products, it is unlikely to fail completely.

References

1. *CCITT Recommendation I.121, Broadband Aspects of ISDN,* Comité Consulatif International de Téléphonique et Télégraphique, Geneva, 1988.
2. *P802.6 Distributed Queue Dual Bus (DQDB) Subnetwork of a Metropolitan Area Network (MAN),* IEEE, Piscataway, N.J., 1991.
3. *CCITT Recommendation E.164, Numbering Plan for the ISDN Era,* CCITT, 1988.
4. *CCITT Recommendation I.150, B-ISDN Asynchronous Transfer Mode Functional Characteristics,* CCITT, 1992.
5. *CCITT Recommendation I.361, B-ISDN ATM Layer Specification,* CCITT, 1992.
6. *CCITT Recommendation I.362, B-ISDN ATM Adaptation Layer (AAL) Functional Characteristics,* CCITT, 1992.
7. *CCITT Recommendation I.363, B-ISDN ATM Adaptation Layer (AAL) Specification,* CCITT, 1992.
8. *ATM User to Network Interface,* V. 3.0, ATM Forum, 1993.

Acknowledgments

I acknowledge assistance given by my colleagues and management in the preparation of this article. In particular, thanks go to Roy Mills for many useful comments and to Rob Williams for his assistance with conformance and interoperability.



Richard Vickers has represented Northern Telecom in the development of B-ISDN standards since 1987, both in Committee T1 and CCITT (now ITU-T). Currently, he serves as chair of T1S1.5, the standards group responsible for the development for ATM and B-ISDN standards in the US, and for the development for the ITU-T. Earlier, he served as chair of the ATM Subworking Group in T1S1.5. At Marconi in Chelmsford, England, he developed digital transmission and switching equipment. For BNR in Ottawa, Canada, he developed a variety of technologies including fast circuit switching and ATM multiplexing.

Vickers graduated from the University of Wales, Cardiff, with a BSc in electrical engineering.

Direct questions concerning this article to the author at NTI, Department 6LMS242, PO Box 3511, Station C, Ottawa, Ontario K1Y 2H7, Canada; vickers@bnr.ca.

Reader Interest Survey

Indicate your interest in this article by circling the appropriate number on the Reader Service Card.

Low 168 Medium 169 High 170