Multimedia and Multimedia Communication: A Tutorial

Chwan-Hwa Wu, *Senior Member, IEEE*, and J. David Irwin, *Fellow, IEEE*

***Abstract*— This paper presents, in a tutorial fashion, the important features of multimedia technology. The specific areas addressed are multimedia and compression standards, computer networks, multimedia transport, and some specific applications employed by industry to date. Multimedia and the effective and efficient communication of multimedia using compression and networks are fused together in this tutorial in an attempt to demonstrate the tight coupling which exists between these two interrelated technologies. First, the techniques and properties inherent in both multimedia and compression standards are presented. Then, the important characteristics of the major local and wide area networks are summarized. Next, the communication techniques for the transport of video and video conferencing are discussed. The new strategies employed to connect homes through cable TV and the telephone companies, as well as the new Ethernet technologies, are also described. Finally, some modern applications of multimedia communication derived from the automotive industry are used to describe the use of this technology in design, manufacturing, and sales.**

***Index Terms*— Compression standards, computer networks, multimedia communication, multimedia standards.**

I. Introduction

THE assimilation of multimedia and multimedia communication by industry for applications in design, manufacturing, and training marks a significant turning point. This important and constantly evolving area comprises a number of technologies, including multimedia compression, computer networks, and the transport of multimedia over these networks. The standards and technology for multimedia and multimedia communication are evolving quickly and, therefore, it is challenging to keep pace with the wide spectrum of this rapidly advancing technology. This tutorial is designed to provide an overview of the fundamental and critical aspects of this area, in order to foster a better understanding of this technology, which, in turn, will result in the efficient and effective design of multimedia communication networks for numerous applications.

Multimedia and multimedia communication can be globally viewed as a hierarchical system. The multimedia software and applications provide a direct interactive environment for users. When a computer requires information from remote computers or servers, multimedia information must travel through computer networks. Since the amount of information involved in the transmission of video and audio can be substantial, the multimedia information must be compressed before it can be sent through the network, in order to reduce the communication delay. Constraints, such as limited delay and jitter, are used to ensure a reasonable video and audio effect at the receiving end. Therefore, communication networks are undergoing constant improvements, in order to provide for multimedia communication capabilities. Local area networks are used to connect local computers and other equipment, and wide area networks and the Internet connect the local area networks together. Better standards are constantly being developed, in order to provide a global information superhighway over which multimedia information will travel.

One of the goals of this paper is to describe the hierarchical system in a structural manner, so that readers can gain a comprehensive understanding of this area.

The format for this tutorial is as follows. First, the multimedia and compression standards are presented and, then, the major computer networks, including local area networks and wide area networks, are analyzed. This material is followed by a discussion of methods for transmitting multimedia over networks, including compressed video and video conferencing, new developments from the cable TV and telephone industries for multimedia and Internet access, multimedia over the Internet, and the new trend in improving the Ethernet for multimedia communication. The automotive industry is highlighted as an application of multimedia communication for design, manufacturing, and sales. Finally, concluding remarks are given for comparing standards and technologies, as well as providing a vision for future development.

II. Multimedia and Compression Standards

A. Multimedia Technologies and Standards

There are a number of widely accepted multimedia technologies or products [such as QuickTime, Video for Windows, and Indeo (Intel Video)] [1], and there is a new International Standard Organization (ISO) standard entitled Multimedia and Hypermedia Information Coding Expert Group (MHEG). The following is a brief description of the popular multimedia technologies and the appropriate standards.

1) QuickTime: This technology was developed by Apple Computer for Apple computers, Microsoft Windows (3.1, 95, and NT), and UNIX platforms. It is a proprietary technology and is the most widely accepted format for multimedia storage and communication. It supports Motion Picture Expert Group (MPEG), an ISO compression standard, Indeo (Intel Video), Kodak's photo CD format, and Musical Instrument Digital

Manuscript received September 7, 1996; revised January 17, 1997.

The authors are with the Department of Electrical Engineering, Auburn University, Auburn, AL 36849-5201 USA.

Publisher Item Identifier S 0278-0046(98)00416-X.

0278–0046/98$10.00

Interface (MIDI). To play a QuickTime movie, a software player can be downloaded from the Internet without charge.

2) Video for Windows: Video for Windows (VFW) was developed by Microsoft for a Windows environment (3.1, 95, and NT), and ActiveMovie will soon replace VFW and be backward compatible with VFW. The Audio Video Interleave (AVI) file format interleaves the audio and video. It is designed to playback videos in a small window using software; the size of small windows is 320 $\times$ 240 pixels (quarter VGA screen) in an Intel 486 computer or 640 $\times$ 480 pixels (VGA screen) in an Intel Pentium computer equipped with a graphics accelerator. It uses the run-length encoding method to compress information, and it supports other video compression algorithms, such as QuickTime and Indeo.

3) Indeo (Intel Video): This technology was originally developed by the David Sarnoff Research Center for converting a National Television System Committee (NTSC) analog signal into digital video, in order to store and play back video in a personal computer (PC). Software-only Indeo is a choice of compression schemes in QuickTime and Video for Windows. One major difference between the above two technologies (QuickTime and Video for Windows) and the hardware Indeo is that Indeo uses hardware for compression and playback to achieve higher frame rate and resolution, whereas the other two technologies only support software compression. Indeo requires a PC-size card using digital video interface (DVI) technology that is based on the Intel i750 chipset. This proprietary compression method, using vector quantization and run-length coding, can achieve 160 : 1 compression ratios and store 70 min of video on a CD-ROM. A major application of Indeo is videoconferencing. The Personal Conferencing Work Group (PCWG) led by Intel is promoting a product based on Indeo for video conferencing. To make the price of the product acceptable, software compression is used, instead of hardware. (Intel has other videoconferencing products based on standards.) Since Indeo uses the vector quantization technique, it is much faster than the MPEG and H.261 (an International Telecommunication Union—Telecommunication (ITU-T) video conferencing standard) based on the discrete cosine transform (DCT).

4) MHEG: MHEG is a standard (ISO/IEC JTC1/SC29 WG12) set by the International Standard Organization for a set of object classes used to control the presentation of multimedia and hypermedia information. MHEG-5 is the standard developed to support the distribution of interactive multimedia applications in a multivendor client/server environment [28]. Application developers who use this standard need only develop an application once, since information representation is interchangeable in a multivendor computer and network environment [29]. In comparison to the document description capabilities of Hyper Text Markup Language (HTML), the MHEG provides the additional multimedia handling capabilities. The ISO's Abstract Syntax Notation (ASN.1) is used for representation between different computer platforms, operating systems, and languages. It uses space (three spatial coordinates) and time to synchronize the presentation. A product prototype, the ARMIDA Service, was developed for video on demand in Italy.

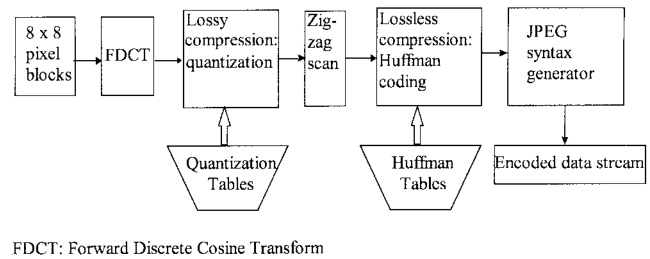

Fig. 1. The block diagram of the JPEG compression scheme.

B. Compression Standards

There are many compression techniques, for example, the proprietary graphics interchange format (GIF) is the most widely used format in the Internet and Bulletin Board Service (BBS). This discussion will focus on two major compression standards developed by the ISO. It is believed that these two standards will be widely accepted as the high-quality compression techniques for multimedia communication.

1) Joint Photographic Expert Group (JPEG): JPEG is the standard for compressing and decompressing continuous-tone (color or grayscale) images [2]. This standard was set by ISO/IEC JTC1/SC29 WG10. JPEG can be used for compressing still pictures for audiographical conferencing, color FAX, image archiving, desktop publishing, multimedia, and medical imaging.

An overview of the JPEG compression process is shown in Fig. 1. An image is divided into blocks, and each block contains 8 $\times$ 8 pixels. Blocks are transformed to frequency domain using the forward discrete cosine transform (FDCT). The two-dimensional (2-D) DCT coefficients are quantized using quantization matrices (8 $\times$ 8 matrix). This process is a lossy one, since the original values cannot be recovered after the quantization process is complete. The high-frequency DCT coefficients have smaller values in comparison to the low-frequency DCT coefficients. Furthermore, the high-frequency DCT coefficients are not so sensitive to the human vision system as the low-frequency coefficients. Therefore, the quantization process can quantize the high-frequency coefficients much coarser than the low-frequency coefficients; this quantization process produces numerous zeros in the high-frequency part of the 2-D DCT coefficients. Then, the quantized 2-D DCT coefficients are converted to a one-dimensional (1-D) matrix using a zigzag scan. The 1-D matrix is coded using predefined Huffman tables. As a result of the zigzag scan, numerous zeros in the high-frequency portion are next to one another, and the Huffman coding produces a better compression ratio. The Huffman coding process is a lossless compression process that provides a compression ratio of about two, whereas the lossy quantization process provides a much higher compression ratio. The last compression step is to generate the JPEG syntax that can be understood by the JPEG decompression processes. JPEG decompression reverses the compression process. The first step is to decode the JPEG syntax; then, Huffman decoding is performed using the Huffman tables, and a 1-D matrix is obtained for each block. The reversed zigzag scan converts the 1-D matrix to the 2-D matrix. The dequantization process uses the quantization table to regenerate the 2-D DCT coefficients that are transformed back to the spatial domain by the inverse DCT (IDCT) process to reconstruct the block.

For a colored image, we can encode red, green, and blue (RGB) data as three independent inputs or use luminance (Y) and subsampled chrominance (C_r and C_b, or U and V). A linear transformation can be used to convert the RGB values to YUV values and vice versa. A YUV image has values of Y, C_r, and C_b for every pixel and is labeled as $Y{:}C_r{:}C_b =$ 4:4:4. If the chrominance values are subsampled every other pixel along the x coordinate, then $Y{:}C_r{:}C_b =$ 4:2:2 and, if the chrominance values are subsampled every other pixel along both the x and y coordinates, then $Y{:}C_r{:}C_b =$ 4:2:0.

The JPEG standard defines the following four coding processes: 1) baseline system; 2) extended system; 3) hierarchical encoding mode; and 4) lossless method. They are defined for the following different applications.

1) The baseline system is a simple and efficient system that is adequate for most applications. It uses a single pass through the data (a sequential encoding) to encode the image and has 8-b resolution for each input.
2) The extended system enhances the baseline system to satisfy applications requiring higher resolution (12 b), progressive buildup, and arithmetic encoding. It also supports sequential coding (similar to the baseline system) and 8-b resolution. The progressive encoding mode supports spectral selection (from low-frequency blurry image to high-frequency sharp image buildup) and successive approximation (from the most significant bits to the least significant bits buildup). Note that the progressive buildup is suitable for applications sensitive to communication time, such as sending images for real estate selection. Arithmetic encoding can compress 5%–10% better than the Huffman coding.
3) The hierarchical encoding mode supports a sequence of frames from low- to high-resolution buildup, like a pyramid, from a low number of pixels to a high number of pixels. The difference between the current and the subsequent frame is encoded using the DCT. This process supports either 8 or 12 b.
4) The lossless method is used in cases that require an exact reconstruction of the original image, such as medical applications. It does not use the DCT. The digital values of each pixel can be from 2 to 16 b. This process supports sequential encoding, and a user can choose either Huffman or arithmetic encoding.

2) MPEG: The MPEG standard was developed by ISO/IEC JTC1/SC29 WG11 for compressing motion pictures, in contrast to JPEG, which is used for still images [3], [4]. The MPEG standards only specify the decoding process and the bitstream syntax and leave room for research on the encoding process. The MPEG has two defined standards, MPEG-1 [3] and MPEG-2 [4], and a standard currently under development, MPEG-4 [5].

1) MPEG-1 is used for compressing video on digital storage media (CD-ROM) at 1 to 1.5 Mb/s or through communication networks at various rates [6]–[9]. The compression ratio is approximately 130 : 1.

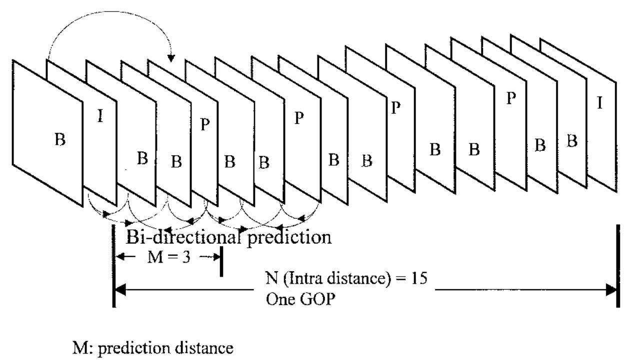

Fig. 2. The GOP structure in the MPEG.

2) MPEG-2 is used for compressing video for the following: a) interlaced digital video, cable, and satellite TV; b) high definition TV (HDTV); c) high-quality digital storage media; and d) video service over communication networks [6]–[9]. The compression ratio is 30–100 : 1.
3) MPEG-4 will be set for a wide range of bit rates and applications [5], [30]. The activity started in 1993, with an expected completion date of 1998. The near-term project is to refine H.261 at a rate from 8 to 30 kb/s. The long-term project is to define an advanced coding method. In addition to the compression capabilities, MPEG-4 offers the following: 1) universal access capabilities for robustness in error-prone environments and content-based scalability and 2) content-based interactivity for content-based multimedia data access, content-based manipulation and bitstream editing, hybrid natural and synthetic data coding, as well as improved temporal random access. MPEG-4 Syntactic Description Language (MSDL) was proposed for flexibly describing the bitstream structure, configuring, and programming decoders, as well as selecting, describing, and downloading tools, algorithms, and profiles. MSDL includes the following: 1) profile—algorithm or combination of algorithms for a particular application; 2) algorithm—collection of tools to provide one or more functionality; and 3) tool—basic technique. MSDL's syntax structure consists of three layers: 1) FLEX 0, for switching between predefined algorithms (MPEG-1, MPEG-2, JPEG, and so forth); 2) FLEX 1, for configuring decoders using standard tools (e.g., DCT and Huffman decoding); and 2) FLEX 2, for tool updating and downloading. Notice that the MSDL makes MPEG-4 significantly different from previous standards.

In contrast to the compression ratios of MPEG, the Motion JPEG (using JPEG to compress motion pictures) has a compression ratio of 7–27 : 1 and the H.261/H.320 (ITU-T video conferencing standards) have compression ratios around 24 : 1.

The MPEG defines a group of pictures (GOP) structure. Each GOP starts from an I (Intra) frame; the number of frames in a GOP is N (including the first I frame), and there are a few P (forward prediction) frames in a GOP. The frames between the I and P frames, P and P frames, and P and I (belonging to the next GOP) are B (bidirectionally predicted) frames, and the number of B frames between I and P is defined as $M - 1$. This format is illustrated in Fig. 2.

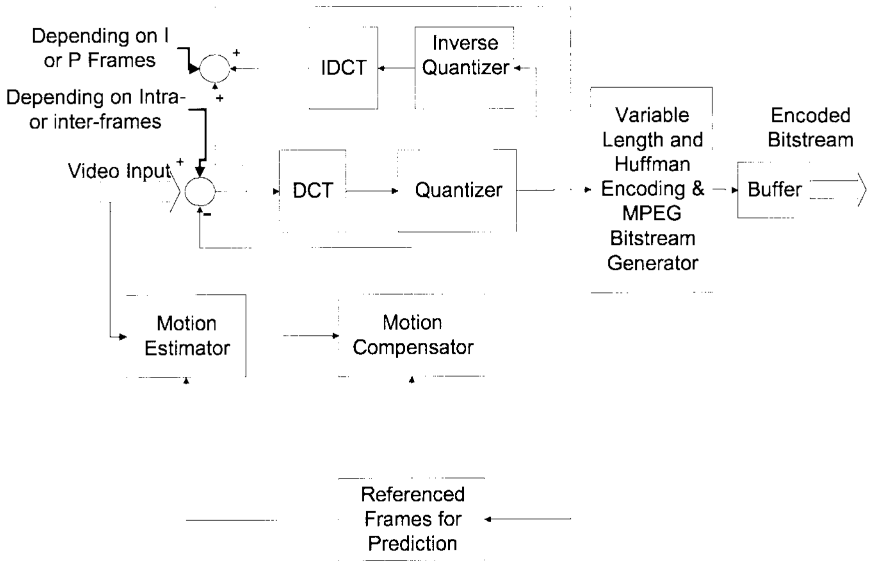

Fig. 3. The block diagram of the MPEG encoding scheme.

Typically, $N = 15$ and $M = 3$. The I frame is compressed in a manner similar to that employed in the JPEG, without using the redundancy between frames. The P frame is compressed by using the nearest reconstructed (from a compressed frame) I or P frame as the referenced frame to estimate the motion vector for each macro block (MB), containing four 8×8 blocks). Based upon the motion vectors, a predicted frame (a motion-compensated frame) can be constructed using the referenced I or P frame and, then, the difference between the predicted frame and the actual frame is compressed. The B frame is compressed by using the nearest reconstructed I and P, P and P, or P and I (belonging to the next GOP) as referenced frames to estimate the motion vector for each MB bidirectionally. Then, a bidirectionally predicted frame can be generated using the referenced frames and motion vectors, and the difference between the predicted frame and the actual frame is encoded. The P and B frames are called inter frames in contrast to I frames.

The encoding process is illustrated in Fig. 3. The encoding technique for I, P, and B frames is as follows.

1) The I frame is compressed in a manner similar to that used in the JPEG by taking the DCT, then the quantization process, variable length, and Huffman coding, and the frame is put in the MPEG bitstream. The quantized DCT coefficients are inverse quantized and the IDCT is used to generate the referenced frame which is stored in the memory. This referenced frame is used for motion estimation for the P and B frames.
2) The motion vectors of the MB's for the P frame are estimated by the motion estimator based on the referenced I or P frames; then, the motion-compensated P (or the forward predicted) frame is generated using the motion vectors and the referenced frame. The DCT coefficients for the difference between the input P frame and the predicted frame are quantized; variable length and Huffman coding are performed, and the result is coded in an MPEG bitstream. The referenced P frame is generated by the inverse quantization process, the IDCT of the difference between the predicted frame and input frame and the addition of this difference and the motion-compensated frame.
3) The motion estimator is used to bidirectionally estimate the motion vectors based on the nearest referenced I and P, P and P, or P and I frames. The motion-compensated frame is generated using a pair of nearest referenced frames and the bidirectionally estimated motion vectors. The DCT coefficients for the difference between the input B frame and the bidirectionally predicted frame are quantized; variable length and Huffman coding are performed, and the result is coded in an MPEG bitstream. The referenced B frame is generated by the inverse quantization process, the IDCT for the difference between the predicted frame and input frame and the addition of the difference and the motion-compensated frame.

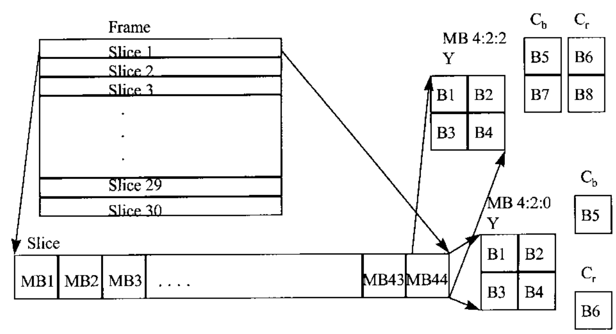

Fig. 4. The slice and MB structure of the MPEG-2.

A frame is divided into a sequence of slices and a slice is further divided into MB's. An example of an MPEG-2 frame is shown in Fig. 4. An MB may contain two blocks of C_b and C_r (4:2:2) or one block of C_b and C_r (4:2:0). MPEG-2 can also support 4:4:4 format; however, the MPEG-1 only supports 4:2:0.

TABLE I
MPEG-2 PROFILE AND LEVELS (S: SCALABLE; NS: NONSCALABLE; F/s: FRAMES/SECOND; MS: MEGASAMPLES/SECOND)

Profile / Level	Simple NS 4:2:0 (No B)	Main NS 4:2:0 (with B)	SNR S 4:2:0 (2 layers)	Spatial S 4:2:0 (3 layers)	High (3 layers) NS 4:2:2, S 4:2:0/4:2:
Low		352 x 288, 30 F/s, 3.04 MS/s, up to 4 Mbps	352 x 288, 30 F/s, 3.04 MS/s, up to 4 Mbps		
Main	720 x 576, 30 F/s, 10.4 MS/s, up to 15 Mbps	720 x 576, 30 F/s, 10.4 MS/s, up to 15 Mbps	720 x 576, 30 F/s, 10.4 MS/s, up to 15 Mbps		720 x 576, 30 F/s, 11.06/14.75 MS/s, 20 Mbps
High-1440		1440 x 1152, 60 F/s, 47 MS/s, up to 60 Mbps		1440 x 1152, 60 F/s, 47 MS/s, up to 60 Mbps	1440 x 1152, 60 F/s, 47/62.7 MS/s, up to 80 Mbps
High		1920 x 1152, 60 F/s, 62.7 MS/s, up to 80 Mbps			1920 x 1152, 60 F/s, 62.7/83.5 MS/s, up to 100 Mbps

The most computationally intensive part of MPEG is the motion estimation. A number of algorithms have been developed [5], and many investigations continue. The basic idea is that of using a full search by matching an MB in the input frame with every MB in the referenced frame(s). If there is a good match, then the movement of the MB in the input frame is the motion vector corresponding to that MB. The accuracy of the motion vectors is one-half pixel. For high motion scenes in MPEG-2, the input MB contains two field blocks, each containing 16×8 pixels, corresponding to the interlaced format, whereas, the normal motion scene uses a 16×16 frame block. Note, however, that MPEG-1 does not support interlaced format.

The quantization process is similar to that used in the JPEG. In order to control the communication rate, a quantization scale factor (mquant, an integer between 1 and 31) is used to multiply the quantization matrices. The Simulation Test Model 5 (STM5) is defined for testing the MPEG-2, and it checks the fullness of the buffer and uses the mquant scale factor to adjust the quantization level to obtain the desired bit rate [10], [11]. The quantization process in the STM5 may take more than one iteration.

The intra frame is handled differently from the inter frames in MPEG. The zigzag scan is used for inter frames to convert the 2-D matrix to a 1-D matrix (including ac and dc coefficients), whereas a newly defined alternate scan is used to scan the ac coefficients for intra frames [4]. The 1-D matrices of the ac coefficients are coded using variable length (or run length) and Huffman coding. In contrast, it is the difference between the dc coefficients of consecutive MB's (forms a 1-D matrix) that is coded by variable length and Huffman coding.

The MPEG-2 supports different profiles and levels, as shown in Table I, to satisfy constraints imposed by the communication network, user equipment (TV or monitor), and variable window size in a computer [12], [13]. The scalability to satisfy these constraints includes signal-to-noise ratio (SNR), spatial, temporal and hybrid scalability by multiple layers. The SNR, spatial, and temporal scalabilities have two layers. In SNR, both layers have the same spatial resolution, but different qualities (quantization and chroma), and the lower layer of spatial scalability has a lower spatial resolution in comparison to the high layer. The lower layer of the temporal scalability has a lower frame rate. The hybrid has three layers that use any two of the scalabilities. The high profile uses the hybrid scalability.

III. COMPUTER NETWORKS

The computer network is the means by which the multimedia are transported. Thus, in this section, the major characteristics of computer networks [14]–[16] will be described.

A. 802.X Local Area Networks (LAN's)

The 802.X standard defined by IEEE contains several LAN standards:

1) 802.1—including the relationship between 802.X and the open system (OSI) interconnect reference model, interconnecting 802.X by bridges and network management;
2) 802.2—logical link control for addressing multiple service points to multiple simultaneous applications and for connection-oriented, connectionless, and acknowledged operations;
3) 802.3—Ethernet standard;
4) 802.4—token bus standard manufacturing automation protocol (MAP);
5) 802.5—token ring standard;
6) 802.12—100VG-AnyLAN standard.

The Ethernet is the most widely deployed LAN and includes the newly developed fast Ethernet standard. The token bus has very limited deployment for manufacturing. Token ring connects stations as a ring and uses a token to resolve contention. 100VG-AnyLAN is mainly for multimedia communication.

B. Fast Ethernet and 100VG-AnyLAN

The Ethernet was originally designed to connect stations by a coaxial cable (10BASE5, thick cable and 10BASE2, thin cable), and later, it evolved to use twisted pairs to connect stations to a hub (10BASE-T). The transport rate is 10 Mb/s. To improve the rate, the fast Ethernet (100BASE-T) was developed to accommodate 100 Mb/s and must use a hub to connect stations. The 100BASE-T can use two pairs of category 5 UTP (unshielded twisted pair) or four pairs of category 3 UTP cables.

The 100VG-AnyLAN (802.12) was designed to handle multimedia communication using a priority demand scheme.

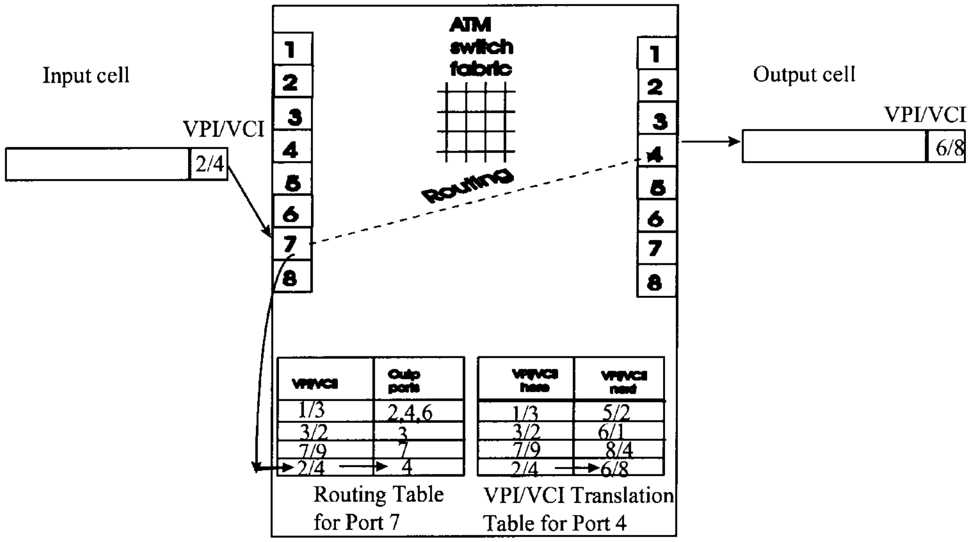

Fig. 5. An example of cell switching by using the routing table and translation table.

It assigns high priority to multimedia, in contrast to the Ethernet (802.3), which uses the carrier sense multiple access with collision detection (CSMA/CD) and cannot assign high priority to time-critical traffic, such as isochronous (video and audio) data. The hubs of a 100VG-AnyLAN can be connected in a hierarchical star geometry with three layers maximum, and 10BASE-T or token ring LAN's (but not both) can be connected as part of the network. The same kind of cables used in the fast Ethernet can be used in the 100VG-AnyLAN.

C. Switched LAN's

The shared medium in the LAN's presented above has a scalability problem. As the number of stations increases, the effective bandwidth for each station decreases. This problem can be alleviated by using a switched LAN that interconnects stations through a switching hub. The communication cannot be blocked if there is no contention for the same destination; hence, the performance of the network is scalable. The layer 2 switch performs at Layer 2 of the OSI reference model and it behaves like a bridge; hence, it can connect multiple LAN's and stations. The layer 2 switch is much faster than the bridge because it uses application specific integrated circuits (ASIC's) to switch traffic. The switch can learn the location of a LAN station using the medium access control (MAC) address in a LAN station in a manner similar to that used by a bridge. The layer 3 switch is designed to improve the performance and lower the cost of a router that connects the network at Layer 3 of the OSI reference model. A router must perform two functions: 1) generating the routing table for the next hop by collecting the route information from neighboring routers and 2) switching the frame to the appropriate output port based on the routing table. The layer 3 switch is designed to perform the second function using ASIC's, and the routing table can be obtained from a route server that is designed to perform the first function. Several switches can share one route server to reduce the cost and the ASIC's can switch faster than a reduced instruction set computer (RISC)-based router.

D. Asynchronous Transfer Mode (ATM)

The ATM performs the following functions: 1) asynchronously multiplexes small packets (called cells, each consisting of 5 octets of headers and 48 octets of payload) going from a number of information sources to various destinations, into seats in a constantly flowing train, if there is a seat for the specified destination with the required quality [quality of service (QoS)]; 2) switches cells during transport if necessary (just like changing trains in a train station); and 3) lets cells jump off the train at the destination. The ATM uses hardware switching to achieve much faster switching speeds in comparison to other communication switches. The QoS that can be negotiated between applications and the ATM provides the required bandwidth for video and audio (isochronous) communications. ATM can also transfer bursty or asynchronous data and, therefore, is designed to accommodate all kinds of communication.

The ATM uses a connection-oriented operation. It establishes a sequence of switches, so that a connection is made from the source to the destination. Such a connection is called a virtual circuit connection (VCC). The switches can be established to perform simplex, duplex, multicast, and broadcast communications. A virtual connection (VC) is a connection between a switching node and the next node; hence, a VCC consists of a series of VC's. There are two kinds of VCs: 1) a permanent VC (PVC) for a leased line and 2) a switched VC (SVC) for a dynamically established connection. To simplify the management of VC's, a number of VC's with the same starting and ending node is grouped together as a virtual path (VP). To identify a VP or a VC, a number is used as the identifier and is labeled VPI/VCI (VP identifier/VC identifier).

An example of cell switching is shown in Fig. 5, which illustrates how two tables are used to perform the switching. These two tables can be established by operators for PVC or signaling for SVC. In this example, an input cell with a VPI/VCI = 2/4 in the cell header appears at input port 7. The routing table for port 7 indicates that the cell should be routed to output port 4. The VPI/VCI translation table for output port 4 modifies the VPI/VCI to 6/8 in the cell header for next switch. If the input VPI/VCI value is 1/3, then a multicast is performed by sending the cell to output ports 2, 4, and 6.

To adapt to different characteristics of the traffic, ATM provides five types of adaptation. Type 1 is for circuit emulation at constant bit rate (CBR) service for isochronous data, type 2 is for variable bit rate (VBR), connection-oriented service for isochronous data (no standard yet), type 3 is for connection-oriented data service, type 4 is for connectionless data service, and type 5 is for LAN emulation and all other possible traffic. The ATM Forum also defines available bit rate (ABR) (guarantees a minimum rate, but delay may vary) and the unspecified bit rate (UBR) (similar to ABR, but does not guarantee a minimum rate and cells may be lost due to congestion).

LAN emulation (LANE) was developed by the ATM Forum for interconnecting LAN's and ATM networks. LANE works at Layer 2 of the OSI reference model. LANE uses LAN emulation server (LES), LANE configuration server (LECS), and broadcast and unknown server (BUS) for configuration, address resolution, broadcast, and the resolution of unknown addresses. A more complicated multiprotocol over ATM (MPOA) is under development by the ATM Forum to perform Layer 3 operations for multiple protocols such as transmission control protocol/Internet protocol (TCP/IP) and Internet packet exchange/sequenced packet exchange (IPX/SPX).

The ATM is suitable for multimedia communication because it provides a guaranteed QoS. Although the adoption of ATM is slow now, it is expected to increase, because the cost of ATM is decreasing and there is an expanding demand for multimedia communication. The ATM is currently being used for a collapsed backbone of LAN's [to replace fiber distribute data interface (FDDI)] and for power user groups. The migration to ATM is steady, since the LANE and MPOA make the ATM easier to manage.

E. Synchronous Optical Network (SONET) and Broad-band Integrated Service Digital Network (BISDN)

The SONET was developed by Bellcore for wide area networking and, it is an American National Standard Institute (ANSI) standard used in the U.S. and Canada. The European counterpart, synchronous digital hierarchy (SDH) is an ITU-T standard. The SONET and SDH are physically compatible at synchronous transport signal (STS)-3C (155.52 Mb/s) and synchronous transport module (STM)-1, and the speeds above. The SONET uses a frame to carry lower-speed information (tributaries: video, audio, and data). The bandwidth for each tributary is guaranteed. The SONET uses synchronous multiplexing (SMUX) and add/drop multiplexers (ADM) to insert a signal from a source and to extract a signal at the destination.

The BISDN uses ATM over the SONET to multiplex signals. The cells are packed in the SONET frame's payload. BISDN provides LAN and metropolitan area network (MAN) connections, including 802.X, FDDI, frame relay, and switched multimegabit data service (SMDS), supports integrated service digital network (ISDN) and isochronous communication. It is expected that the information superhighway will use the BISDN for the next generation of Internet to accommodate the multimedia communication.

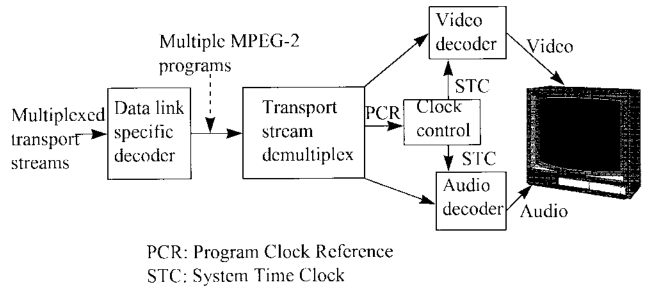

Fig. 6. The processes of decoding transport streams.

IV. Multimedia Communication

A. MPEG-2 Over ATM

The MPEG-2 defines the system part [17] for combining one or more program elementary streams (PES's) of video and audio for transmission or storage, and it includes a transport stream (TS) for an erroneous environment and a program stream (PS) for an error-free environment. A TS packet is composed of 188 octets, including a packet header and a payload (PES) packet. A PES packet contains the packet start code prefix, stream I.D., packet length, priority, and data. The video and audio are encoded based on the system time clock (STC) time base; then, they are packetized and multiplexed into TS's. The TS's are converted to the appropriate cells (for ATM) or frames [(LAN's and wide area networks (WAN's)] and transmitted through a communication network to the destination. The receiver converts the cells and frames back to multiplexed MPEG-2 TS's using a data link specific decoder. The transport stream demultiplexer demultiplexes the multiple programs. The program clock reference (PCR) is used by clock control to regenerate the STC for decoding the synchronized video and audio. The block diagram for the TS decoder is shown in Fig. 6.

The MPEG-2 can be sent over the ATM for multimedia services, such as video on demand. There are two types of ATM adaptation that can be used for transferring MPEG-2.

1) Type 1 (constant bit rate) for guaranteed bandwidth and latency is jitter free, but still suffers from cell loss and bit errors that may cause error propagation through a period of one GOP [18].
2) Type 5 (AAL5, ATM Adaptation Layer 5) standard is phase one for video on demand and was approved by the ATM Forum in January 1996 [19]. There are several products available based on this standard. The later phases will include video conferencing (part of the H.32X and broadcast video). The MPEG-2 over AAL5 is believed to be the standard widely supported by many vendors. Hence, mapping MPEG-2 to AAL5 will now be discussed.

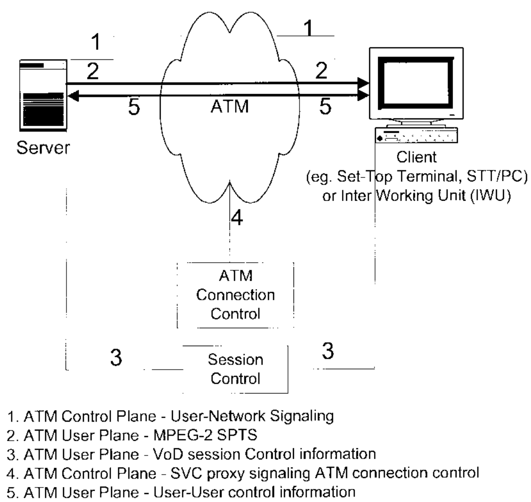

Fig. 7. MPEG-2 over AAL5: the video on demand reference configuration.

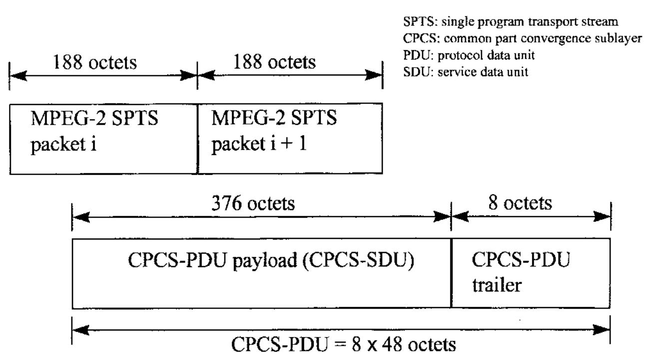

Fig. 8. Mapping the TS's of MPEG-2 into AAL 5PDU's.

The video on demand reference configuration for ATM over AAL5 is shown in Fig. 7 [19]. There are five interfaces for the five kinds of information flow that are mapped to separate VC's using the user–network interfaces (UNI's). Interface 1, which belongs to the control plane, is for signaling between the client and the network, as well as between the server and the network. Interface 2, which belongs to the user plane, transmits the single program TS (SPTS) from the server to the client. Interface 3, which belongs to the user plane, controls the video on demand (VoD) session. Interface 4, which belongs to the control plane, performs the signaling to set up the connection between the ATM switches and routers for Interfaces 2, 3, and 5. Interface 5, which belongs to the user plane, transmits the control information between the VoD server and the client. MPEG-2 over ATM can be also used over the SONET, asymmetrical digital subscriber loop (ADSL), hybrid fiber coax (HFC), fiber to the curb (FTTC), and fiber to the home (FTTH) [20], [21].

One example which illustrates mapping the SPTS of MPEG-2 into protocol data units (PDU's) of AAL5 is shown in Fig. 8 [19]. N packets of the SPTS form a unit, which maps to a number of cells. All of the equipment must support the $N = 2$; two SPTS packets (376 octets) are mapped to eight cells. The common part of convergence sublayer (CPCS) handles the conversion between the MPEG-2 and ATM, including the assembly of headers, payload, and trailers (for error detection and correction). Of the eight cells, 376 octets are used for payload as the service data unit (SDU) and eight octets are used for the trailer. The PDU's are assembled at the transmitter end and disassembled at the receiver end. It is necessary to use buffers at both ends, and jitter may occur due to communication delays and the simultaneous processing of disassembled SPTS's from an SDU. The control of jitter is still under study.

B. Video Conferencing

Video conferencing is widely used for effective communication in industry. A number of standards have been set by the ITU-T for video conferencing.

1) H.320 for ISDN includes the following.
   a) H.261 [22] is for compression and decompression. It is also known as $p \times 64$ $(p = 1, \cdots, 30)$ since multiple 64 kb/s of ISDN B channels can be used. Two formats, common intermediate format (CIF, 288 $\times$ 360 pixels) and quarter CIF (QCIF, 144 $\times$ 180 pixels), can be chosen based on the communication bandwidth available.
   b) H.200 is for audio and still graphics conferencing.
   c) H.230 is for multiplexing audio, video, and data into a digital channel.
   d) H.231 is for the multiple control unit (MCU) for a group video conferencing.
2) H.263 is for very low bit rate visual telephone [part of H.324 is for plain old telephone system (POTS], including:
   a) H.263 for video coding narrowband signals below 64 kb/s;
   b) H.223 for a multiplexing protocol for very low bit rate multimedia systems.
3) H.323 is for shared nonguaranteed bandwidth networks, such as LAN's and the Internet. Internet video phone and cable TV video phone are examples.

The most widely accepted compression standard for video conferencing is H.261. . MPEG is downward compatible to H.261. The system uses both intra frames and inter frames. The intra frame is only sent for the first picture or after a change of scene; there is no motion estimation for the intra frame. DCT, quantization, zigzag scan, and variable length and Huffman coding are used for each MB. The inverse quantization and the IDCT for the quantized frame form a referenced frame. The inter frame uses motion estimation by comparing the input frame to the referenced neighbor frame (the reconstructed neighbor picture) to find the motion vectors. If the difference between the input block and the referenced block is below the threshold, no information need be sent; otherwise, the difference is transformed by DCT, quantized, zigzag scanned, and coded using variable length and Huffman coding. A referenced frame can be generated by reconstructing the frame using inverse quantization and the IDCT on the quantized difference (if it is larger than the threshold, otherwise, the difference is zero) and adding the motion-compensated picture to this difference. A loop filter that removes the high-frequency noise can be used to improve the visual effect.

Increasing p up to $p = 6$ (rate = 384 kb/s) in H.261 can improve the quality and frame rate of video conferencing. When the rate = 384 kb/s, the CIF format can be used; however, it is difficult to see any visual quality improvement even if a higher rate is used. The QCIF is usually used when $p < 6$.

C. High Definition TV (HDTV)

The Federal Communications Commission (FCC) solicited and received proposals in 1988 for HDTV. The Grand Alliance, consisting of MIT, General Instruments, Philips, David Sarnoff Research Center, Thomson Consumer Products, AT&T and Zenith, was formed in 1993 [23]. The Grand Alliance selected the MPEG-2 main profile and high level for video encoding and transport and the AC-3 developed by Dolby Labs for audio encoding. MPEG-2 TS packetization and multiplexing will be used, as was described in Section IV-A. The bandwidth of the transmission channel is 6 MHz, the payload rate is 19.4 Mb/s, and the MPEG packets are scrambled and error correction information is added.

D. Cable TV (CATV)

The new services from CATV include telephone, fax, video on demand, and computer access to the Internet. A cable modem, which must be used in order to connect to the Internet, is connected between the cable feed and the Ethernet network interface card (NIC) inside a computer. Note that the cable's bandwidth is shared by users connected to the same cable from CATV, as is done with the Ethernet. Depending upon the cable modem vendor, the downstream (from headend to home) bandwidth is from 4 to 27 Mb/s, and the upstream (from home to headend) bandwidth is from 96 kb/s to 27 Mb/s. The price range for cable modems is from $400 to $600. The major problem in the cable modem deployment is the noise in upstream communication. The noise problem will be solved in the next few years. A large-scale deployment of cable modems (from 350 000 to 800 000 modems) in the Bay area of California will use the ISDN for upstream until the upstream noise problem is solved. In addition, CATV is also planning to implement telephone services by using SONET to connect headend offices to the public switching telephone network (PSTN).

E. Digital Subscriber Loop (DSL)

In order to compete with CATV for VoD and Internet access services, the telephone companies are deploying DSL, developed by Bellcore [24]. The DSL includes ISDN, asymmetrical digital subscriber loop (ADSL), high data rate DSL (HDSL), single line DSL (SDSL), and very high data rate DSL (VDSL). The downstream and upstream bandwidth, as well as the distance limit to the central office, are tabulated in Table II. The DSL supports analog telephone as well as ISDN basic rate interface (BRI) (two 64-kb/s B channels and one D channel for control) and H0 (384 kb/s for high-quality video conferencing). However, the ADSL-1 does not support ISDN and H0. 98% of North America's subscriber loops are less than 18 kft and 60% of them are less than 12 kft. Hence, it is expected that DSL can be deployed to many users. Special electronic circuits are used for modulation and demodulation of DSL. The monthly fee for ADSL from US WEST is about $60, and it will no doubt eventually drop to $35. The DSL technology will make multimedia communication affordable. Note that the DSL bandwidth is dedicated to a single user, whereas the bandwidth of cable modems is shared.

TABLE II
DSL CHARACTERISTICS: RATE AND DISTANCE LIMITATION

	Downstream	Upstream	ISDN BRI (2B+D) / H0 (384 kbps)	Distance
ADSL-1	1.544 Mbps	16 kbps	No	18 kft
ADSL-2	3.152 Mbps	640 kbps	Yes	12 kft
ADSL-3	6.312 Mbps or 8.448 Mbps	640 kbps	Yes	12 kft or 9 kft
HDSL	1.544 Mbps (2 pairs) / 2.048 Mbps (3 pairs)	1.544 Mbps (2 pairs) / 2.048 Mbps (3 pairs)	Yes	12 kft
SDSL	1.544 Mbps (1 pair) / 2.048 Mbps (1 pair)	1.544 Mbps (1 pair) / 2.048 Mbps (1 pair)	Yes	10 kft
VDSL	12.96, 25.82, or 52 Mbps	1.5 to 2.3 Mbps	Yes	4.5, 3 or 1 kft

F. Internet

The multicast backbone (MBONE) was developed by Steve Deering at Xerox and adopted by the Internet Engineering Task Force (IETF) for time-critical communications. The MBONE requires facilities such as an Internet protocol (IP) multicast router and a T1 (1.544 Mb/s) line. It is being used for video conferencing, CNN News, London-based World Radio Network short wave broadcast, and so forth. The resource reservation protocol (RSVP) set by IETF provides QoS for the IP; hence, isochronous communication over the Internet is enabled. The new IP version 6 (IPv6), providing 128 b of IP address (compared with the current 32-b address of IPv4) [25], [26], will have a flow label to identify the application, and this label can be used to provide the QoS, as is done with VPI/VCI in ATM. The real-time protocol (RTP) is for real-time applications in the Internet and can be run on RSVP or ATM.

G. New Trends

The most widely used LAN, the Ethernet, has low throughput and scalability problems and is not suitable for multimedia communication. It is important to overcome these problems without a significant amount of new investment. There are several new developments on the horizon.

1) Priority access control enabled workgroup switch (PACE) gives multimedia communication a high priority and improves the Ethernet efficiency to 98%. There is no need to replace the NIC card in a computer, but software upgrades for the computer and workgroup switch are required.
2) Cell in frames (CIF) allows conversion between the ATM cells and the Ethernet frame format, so the ATM cells can reach an Ethernet station without the need to replace the NIC in a computer. This technique provides

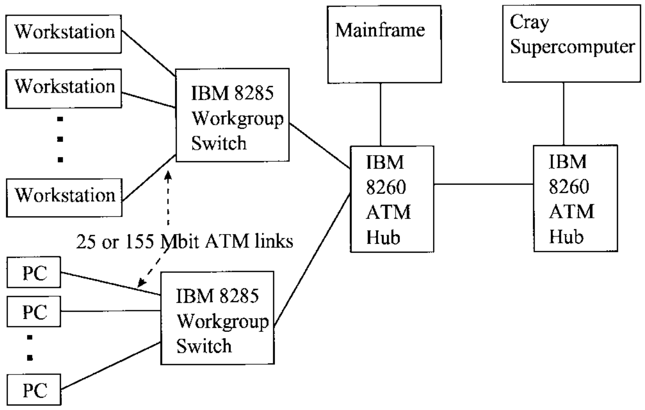

Fig. 9. Chrysler's ATM network for the MPEG video communication, training, CAE, and CAD communication.

an Ethernet physical layer for the ATM and is suitable for multimedia communication.

3) Gigabit Ethernet will provide gigabit connections among hubs and servers, so there is more bandwidth for multimedia communication.

V. Applications of Multimedia Communication

A. Automotive Industry: Engineering Design

Chrysler is using ATM networks for computer-aided design (CAD) and computer-aided engineering (CAE) [27]. The CAE applications for car crash and wind flow simulations are running on Silicon Graphics Inc. (SGI) workstations and a Cray Supercomputer. Engineers watch the MPEG-compressed video that is generated from CAE simulations. The video is viewed in a PC equipped with an ATM card in a network file system (NFS) environment developed by Sun Microsystems. The compressed video is replayed in an SGI Challenge box and sent to the PC using NFS and ATM. Chrysler also has multimedia training material stored in video servers that can be viewed by users anytime. The CAD software, CATIA, is from Dassault Systems in Europe and is running on IBM mainframes and RISC/6000 workstations. In addition, there is a need to transfer a significant volume of data in the networks, since data reside in the mainframe for shared access and must be exchanged among the mainframe and workstations. Furthermore, executable CAD macros need to be loaded to workstations for execution. The Chrysler ATM network for connecting the mainframe, the Cray Supercomputer, workstations, and PC's by ATM workgroup switches and ATM hubs is shown in Fig. 9. This ATM network provides a viable solution for video and data communication.

B. Automotive Industry: Sales

Ford Motor Company also uses one-way broadcast video from its headquarters to 1000 dealerships in Europe through Very Small Aperture Terminal (VSAT), a satellite communication system. Dealers can provide back feedback and receive real-time responses from headquarters using PSTN.

C. Automotive Industry: Manufacturing

Video conferencing is used to keep the automotive manufacturing managers in Japan and the U.S. in touch on a weekly basis. Hard-to-describe mechanical problems can be visually demonstrated. Top experts can be brought into a discussion at a moment's notice. As a result, video conferencing systems are installed on the assembly line. This process permits the manufacturers to resolve from two to three times the number of production issues in a week that could be resolved otherwise. Hence, this approach saves money and time and provides better quality products in a safer manufacturing environment.

VI. Concluding Remarks

This paper has presented multimedia and compression standards, computer networks, multimedia transport, and some industrial applications. A brief comparison/summary of existing standards and technologies follows.

1) QuickTime and Video for Windows provides software-only compression methods in which Indeo is one of the compression methods. QuickTime is used for PC and Macintosh platforms, whereas Video for Windows is designed specifically for PC.
2) MHEG is a standard for distributing interactive multimedia applications in a multivendor client/server environment.
3) JPEG is the standard for compression/decompression of continuous-tone still images; MPEG is for motion pictures.
4) Ethernet (802.3) provides the most widely used LAN standard and fast Ethernet provides a 100 Mb/s rate. Token ring (802.5) provides both 4 and 16 Mb/s rates. 100VG-AnyLAN (802.12) provides 100 Mb/s for multimedia communication.
5) Switched LAN's provide better efficiency, whereas the shared media LAN's experienced problems in scaling up. ATM provides QoS that is critical for multimedia communication. SONET is the standard for a high-speed WAN.
6) MPEG-2 over ATM is the standard for communicating high-quality video over networks. DSL and CATV represent two choices for broad-band communication to homes.
7) H.320 is a videoconferencing standard over ISDN, H.323 is used for shared computer networks (such as LAN's and the Internet), and H.324 is used for POTS.

The future of this advancing technology will be characterized as follows. MPEG-2 over ATM will be the standard for delivering high-quality video over network. DSL/CATV will deliver broad-band information to users directly. Multimedia applications with high-quality video will significantly enhance the information exchange, which, in turn, will stimulate many industrial applications that have been considered impossible.

References

[1] S. Heath, *Multimedia and Communication Technology*. Stoneham, MA: Focal, 1996.

[2] *Information Technoloty—Digital Compression and Coding of Continuous-Tone Still Images: Requirements and Guidelines*, ISO/IEC Standard 10918-1, Feb. 1994.
[3] *Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1.5 Mbits/s—Part 2: Video*, ISO/IEC Standard 11172-2, Aug. 1993.
[4] *Generic Coding of Moving Pictures and Associated Audio Information—Part 2: Video*, ISO/IEC Standard DIS 13818-2, May 1994.
[5] B. Sheu, M. Ismail, E. Sanchez-Sinencio, and T. H. Wu, *Microsystems Technology for Multimedia Applications: An Introduction*. Piscataway, NJ: IEEE Press, 1995.
[6] L. Chiariglione, "The development of an integrated audiovisual coding standard: MPEG," *Proc. IEEE*, vol. 83, pp. 151–157, Feb. 1995.
[7] R. Aravind, G. L. Cash, D. L. Duttweiler, H. W. Hang, B. G. Haskell, and A. Puri, "Image and video coding standards," *AT&T Tech. J.*, pp. 67–88, Jan./Feb. 1993.
[8] D. L. Gall, "MPEG: A video compression standard for multimedia applications," *Commun. ACM*, vol. 34, no. 4, pp. 47-58, Apr. 1991.
[9] FAQ with Answers about MPEG, MPEG1, and MPEG2. Available http://www.crs4.it/HTML/LUIGI/MPEG/.
[10] "Test model 5," ISO/MPEGII, International Standards Organization, Doc. AVC-491, Apr. 1993.
[11] S. Okubo, "Reference model methodology—A tool for the collaborative creation of video coding standards," *Proc. IEEE*, vol. 83, pp. 139–150, Feb. 1995.
[12] J. Morris, "MPEG2 the main profile," in *Proc. IEEE ISCAS'94 Tutorials*, May 1994, pp. 130–138.
[13] N. D. Wells and P. N. Tudor, "Standardization of scaleable coding schemes," in *Proc. IEEE ISCAS'92 Tutorials*, May 1992.
[14] W. A. Shay, *Understanding Data Communications and Networks*. Boston, MA: PWS Publishing, 1995.
[15] F. Halsall, *Data Communications, Computer Networks and Open Systems*, 4th ed. Reading, MA: Addison-Wesley, 1995.
[16] U. Black, *ATM: Foundation for Broadband Networks*. Englewood Cliffs, NJ: Prentice-Hall, 1995
[17] *Generic Coding of Moving Pictures and Associated Audio: Systems*, ISO/IEC Standard DIS 13818-1, Apr. 1995.
[18] M.-T. Sun, "MPEG2 systems and the transport over ATM," in *Proc. IEEE ISCAS'94 Tutorials*, May 1994, pp. 138–146.
[19] ATM Forum Technical Committee, "Audiovisual multimedia services—Video on demand specification 1.0, AF-SAA-0049.000," *ATM Forum*, Jan. 1996.
[20] D. Minoli, *Video Dialtone Technology*. New York: McGraw-Hill, 1995.
[21] F. Fluckiger, *Understanding Networked Multimedia Applications and Technology*. Englewood Cliffs, NJ: Prentice Hall, 1995.
[22] M. Liu, "Overview of the $P * 64$ kbits/s video coding standard," *Commun. ACM*, vol. 34, no. 4, pp. 60–63, Apr. 1991.
[23] K. Challapali, X. Lebegue, J. S. Lim, W. H. Paik, R. S. Girons, E. Petajan, V. Sathe, P. A. Snopko, and J. Zdepski, "The grand alliance system for US HDTV," *Proc. IEEE*, vol. 83, pp. 158–174, Feb. 1995.
[24] D. W. Lin, C.-T. Chen, and T. R. Hsing, "Video on phone lines: Technology and applications," *Proc. IEEE*, vol. 83, pp. 175–193, Feb. 1995.
[25] D. E. Comer, *Internetworking with TCP/IP*, vols. 1–3. Englewood Cliffs, NJ: Prentice-Hall, 1995.
[26] S. A. Thomas, *IPNG and the TCP/IP Protocols*. New York: Wiley, 1996.
[27] S. Schatt, "Asynchronous transfer mode," Computer Intelligence Info-Corp, Los Angeles, CA, 1995.
[28] *MHEG-5*, ISO/IEC Standard DIS 13522-5, Dec. 1995.
[29] MHEG home page for standards and general information. Available http://www.demon.co.uk/tcasey/wg12.html
[30] Y.-Q. Zhang, Guest Ed., Special Issue on MPEG-4, *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, pp. 5–233, Feb. 1997.

Chwan-Hwa "John" Wu (M'88–SM'93), for a photograph and biography, see this issue, p. 3.

J. David Irwin (S'60–M'63–SM'71–F'82), for a photograph and biography, see this issue, p. 3.