NCS TIB 99-2

NATIONAL COMMUNICATIONS SYSTEM

TECHNICAL INFORMATION BULLETIN 99-2

RELATIONSHIPS OF POPULAR TRANSMISSION CHARATERISTICS TO PERCEIVED QUALITY FOR DIGITAL VIDEO OVER ATM

JANUARY 1999

OFFICE OF THE MANAGER
NATIONAL COMMUNICATIONS SYSTEM
701 SOUTH COURT HOUSE ROAD
ARLINGTON, VA 22204-2198

This document was prepared under contract to the

Office of the Manager
National Communications System

Contract #F41621-93-D50060060

by:

Gary L. Ragsdale and Gerard P. Lynch
Southwest Research Institute
P.O. Drawer 28510
San Antonio, TX 78228
210-522-5378 (VOICE)
210-522-5499 (FAX)

Executive Summary

Managing video quality and its delivery, especially during NS/EP events, requires a coordinated effort between network design, video codec design, and network operating practice. Four sets of standards impact the ability of ATM networks to deliver quality video and provide for NS/EP preferences: MPEG-2 video communication standards, ITU-T transport protocol standards, ATM standards, and network carrier standard operating practices.

MPEG-2 has video encoding features that provide for the segregation of video traffic, variable traffic coding rate, managed connection establishment, and conditional information access. They constitute the basic elements for managing network traffic and reducing congestion at its entry point, i.e., the video encoder. Additions of error concealment to the video decoder will further improve quality by reducing the perceived impact of residual errors and cell losses. Carriage of video traffic within the United States will be complicated by the presence of the ATSC standard and its departures from MPEG-2. ATSC lacks important traffic layering features found in MPEG-2. Consequently, ATSC does not offer good opportunities for controlling traffic at the source. Steps should be taken to avoid incompatibilities between MPEG-2 and ATSC networks.

ITU-T transport protocol and ATM standards hold the key for managing the ATM interface and the proper handling of traffic within the network. The transport protocol and ATM interface should share flow control, admission control, and traffic carriage priority information. These standards should promote sharing of available bandwidth, traffic classification, priority, and flow control information from the MPEG-2 video encoder to the transport codec and the ATM interface and vice versa.

There is a need for an overarching standard or set of standards that define the relationship between MPEG-2 traffic generation, traffic transport, and ATM switching. Creation of this overarching standard would highlight deficiencies within the underlying standards, i.e., MPEG-2, ITU-T, and ATM, thereby leading to their eventual enhancement. In so doing, traffic management for commercial and NS/EP-related traffic would improve with an overall benefit to video quality.

Finally, there will be a need for network quality standards and standard network operating procedures within the network operators. The number of switches and the quality of the bearer channels have significant impact upon video quality. Therefore, network organization has significant implications to congestion management. Network quality standards would define the expected behavior of these networks. Quality standards would lead to uniform network design and operating practice in much the same way as AT&T's "Notes on Direct Distance Dialing" and Bell Telephone's technical publications defined high-quality practices within the telephone network years ago.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

AAC	Advanced Audio Coding
AAL	ATM Adaptation Layer
ARQ	Automatic Repeat-Request
ATM	Asynchronous Transfer Mode
ATSC	Advanced Television Standards Committee
BCH	Bose-Chadhuri-Hocquenghem
BER	Bit Error Rates
CATV	Cable Television
CBR	Constant Bit Rate
CCIR	International Radio Consultative Committee
CD	Compact Disk
CD-ROM	Compact Disk-Read Only Memory
CLP	Cell-Loss Priority
DC	Direct Current
DCTs	Discrete Cosine Transforms
DSM-CC	Digital Storage Media Command and Control
DTS	Digital Television Standard
FCC	Federal Communications Commission
FLC	Fixed Length Compression
fps	Frames Per Second
GFC	Generic Flow Control
GOP	Group of Pictures
GSTN	General Switched Telephone Network
HDTV	High Definition Television
HEC	Header Error Control
ISDN	Integrated Services Digital Network
ISO/IEC	International Standards Organization/International Electrotechnical Commission
ITU	International Telecommunications Union
ITU-T	International Telecommunications Union–Telecommunications Standardization Sector
JPEG	Joint Photographic Expert Group
LAN	Local Area Network
MPEG	Moving Picture Experts Group
N6	Technology and Standards Division
NCS	National Communications Systems
NS/EP	National Security and Emergency Preparedness
NTSC	National Television Standards Committee
OMNCS	Office of the Manager, National Communications System
PAL	Phase Alternating Line
PESs	Packetized Elementary Streams
PIP	Picture in a Picture
PSNR	Peak Signal to Noise Ratio

PSTN	Public Switched Telephone Network
PTI	Payload Type Identifier
PVCs	Permanent Virtual Connections
QoS	Quality of Service
R-S	Reed-Solomon
RTI	Real-Time Interface
SDH	Synchronous Digital Hierarchy
SDTV	Standard Definition Television
SNA	System Network Architecture
SNR	Signal-to-Noise Ratio
SONET	Synchronous Optical Network
SVCs	Switched Virtual Connections
TV	Television
VBR	Variable Bit Rate
VCI	Virtual Channel Identifier
VCR	Videocassette Recorder
VCs	Virtual Connections
VLC	Variable-Length Coding
VLSI	Very Large Scale Integration
VPI	Virtual Path Identifier

1.0 INTRODUCTION

The mission of the National Communications System (NCS) is to maintain a state of readiness and ensure the effective use of technology during National Security and Emergency Preparedness (NS/EP) conditions. To meet this challenge, the Office of the Manager, National Communications System (OMNCS), and in particular, the Technology and Standards Division (N6), is assigned the responsibility of ensuring maximum network interoperability and the availability of the latest applicable technology. In general, the N6 Division stays abreast of the latest technology through its participation in national and international standards committees and its association with both government and industry organizations. As a result, the OMNCS is able to apply the latest available technology to meet the critical mission objectives.

One technology of importance to the OMNCS is asynchronous transfer mode (ATM). ATM is a layered communications transport technology that is based on the transfer of information using a 53-byte packet and statistical process mechanisms to guarantee quality of service (QoS) as a part of the basic messaging functions. Furthermore, ATM is considered the next generation integrated services digital network (ISDN) technology, thereby adding great improvements to the ability of networks to process multimedia traffic.

Video communications is playing an ever-increasing role in the way we are entertained, how we work, how we stay informed, and how we manage crises. In the past, video's emphasis was entertainment and the dissemination of information. Growth in desktop and mobile teleconferencing, on-the-scene video broadcasting, and a broad array of organizationally-oriented telecasts demonstrates the importance of visual information in dealing with complex activities including those related to national security and emergency preparedness. Video communications will therefore be a part of information gathering, situation assessment, response coordination, and organizational management in the days to come.

2.0 BACKGROUND

There are many reasons to distribute video information. Sample applications are live video broadcast, delayed video replay, video teleconferencing, and pay-per-view video entertainment. There are well-established, analog distribution channels including TV broadcast stations, cable TV systems, and receive-only satellite systems. Video networks distribute programs across combinations of distribution channels using either analog or digital switching systems. The distribution channels and switching systems currently conform to the National Television Standards Committee (NTSC) format for analog transmission of video information.

NTSC video distribution is giving way to new methods as regulatory, technological, and market forces bring about change in the world. Most notable of these forces are recent Federal Communication Commission (FCC) directives requiring the conversion of NTSC broadcasts to the new high definition TV (HDTV) standard. FCC mandates are leading the video industry away from analog video distribution and into a new digital distribution era.

An industrial group called the “Moving Picture Experts Group” (MPEG) formulated new standards called "MPEG-2" that define the all-digital HDTV format for video information [1]. Video compression algorithms defined within the MPEG-2 standard reduce massive amounts of video data to a quantity compatible with available line-of-sight radio, satellite, and fiber optic transmission media. ATM is the popular candidate for switching MPEG-2 video within the new video distribution networks.

High definition TV proponents and the viewing public expect good quality video reception across a wide variety of networks. Under normal conditions, network behavior varies, creating differences in transmission channel quality and ATM switch performance. Times of war, natural disasters, acts of terrorism, and other times of national security or emergency preparedness are also times when networks experience crises of their own. Surges in volume may congest the network. Natural or manmade transmission interference may inject errors into otherwise good quality systems and cause essential video information to be lost. Left to themselves, error-prone transmission channels and ATM network congestion will reduce digital video quality to unacceptable levels. It is during these times that good quality communications are most important for reasons of command, control, support, and maintaining an informed populace. Therefore, steps should be taken within the HDTV network to compensate for errors and congestion. This paper discusses the factors affecting video quality and measures for maintaining digital video quality within an ATM distribution network, especially when events inject errors and losses within the network.

3.0 DIGITAL VIDEO

Conversion of video imagery to a digital format has long been an objective of research and product development. MPEG led the way in formalizing the research into a well-defined standard called "MPEG-2." The Advanced Television Standards Committee (ATSC) adopted a modified subset of MPEG-2 as its video format and compression for HDTV. This paper focuses on the distribution of MPEG-2 and similar digital video information over ATM networks. It is important to understand the nature and requirements of MPEG-2 as background to designing an appropriate distribution system.

3.1 Video Compression: Evolution and Revolution

Video compression is an evolving technology with a history spanning nearly a decade. The emergence of new multimedia electronics, new transmission networks, and a burgeoning market demand rallied industrial support for interoperable multimedia devices. A need arose for reliable multimedia standards with video compression as a core requirement. A series of video compression standards were developed to address two applications: multimedia teleconferencing and digital TV program distribution.

Communication carriers and video conferencing users have driven, and continue to drive, the development of teleconferencing standards. Rapid developments in digital signal processing, high-speed local area networks (LANs), ISDN, and ATM networks made desktop teleconferencing feasible. The International Telecommunications Union – Telecommunications Standardization Sector (ITU-T) spearheaded video conferencing standards in 1990 by creating the video compression standard H.261 for ISDN and LAN multimedia teleconferencing [2]. H.263 followed in 1995 by defining standard protocols for low-bit rate teleconferencing over the general switched telephone network (GSTN). ITU-T and International Standards Organization/International Electrotechnical Commission (ISO/IEC) collaboration created H.262 for video conferencing over broadband networks, especially ATM networks. With these standards in place, the ITU specified the H.32x series of standards for multimedia communications systems. H.322, H.323, and H.324 define the system standards for isoEthernet, Ethernet, and low bit-rate dial-up multimedia communications systems. The most common standard, H.320, addresses the system standards for narrow-band telephony networks such as ISDN. H.320 allows for using H.262 or H.263 for video compression. H.321 addresses the system standards for broadband ISDN by extending the H.320 specification to address ATM specific issues relative to the ATM Adaptation Layers (specifically AAL-1 and AAL-5). Thus, the teleconferencing standards developed from the need to communicate compressed video over a variety of networks.

A different set of imperatives drove the evolution of MPEG standards. The name "MPEG" comes from the joint ISO/IEC Motion Picture Expert Group standards committee formed for the purpose of codifying multimedia compression [3], [4]. New developments in digital storage and a move to digital TV broadcasts focused interest on new video compression technologies. The desire for video program distribution over new digital media motivated the formation of MPEG, driven largely by the entertainment industry and government regulators. This creative standards committee is leading developments in MPEG-1, 2, and 4 standards.

Video compression encoders produce compressed video information at a continuous coding rate or at a variable-coding rate [2]. Continuous-rate encoders maintain a constant output rate by varying the compression ratio and the image quality. If the video material is highly uncorrelated, the continuous-rate codec increases the compression ratio and may discard some image detail, i.e., it performs a "lossy" compression. Conversely, it adds unnecessary filler bits to the output when the video material produces too little information to maintain the desired output data rate. Continuous-rate codecs operate over synchronous point-to-point channels or synchronous switched networks, such as ISDN and DS-3 circuit switched networks. It is clear that the fixed rate codec must either waste transmission bandwidth or sacrifice picture quality in order to operate over a synchronous network and maintain a constant coding rate.

A variable-rate codec adapts to the video material and varies its coding rate according the spatial and motion content of the video material. A variable rate codec is less complex to design, makes better use of transmission services, and reproduces image quality consistently across a broader range of materials [2]. The network supporting a variable rate codec must also vary with the "asynchronous" video coding rate. Variable rate, asynchronous transmission is one of the operating modes of ATM and is well suited for operation of variable coding rate video transmissions over an ATM network.

The rest of this paper devotes its attention to variable-rate codecs operating over ATM networks. Any reference to a codec, a transmission channel, or a network will implicitly mean variable rate codec, ATM bearer channel, and ATM network, respectively.

3.2 MPEG-1 — Genesis Of Video Compression Standards

ISO/IEC approved MPEG-1 in 1992 for video and audio storage and retrieval at 1.5 Mbps. MPEG-1 codifies distribution of multimedia programming on CD-ROM and digital tape storage media. It has videocassette recorder (VCR) start, stop, rewind, fast-forward, fast-reverse, and other retrieval functions.

MPEG-1 is a landmark standard. It provides the basis for, and is a subset of, MPEG-2. MPEG-2 expands upon MPEG-1's video compression techniques for spatial and motion compression. MPEG-1 audio compression is a subset of the MPEG-2 audio standard. The ITU-T incorporated MPEG-2 video compression (a.k.a. ITU-T H.262) and MPEG-1 audio compression into its H.310 multimedia conferencing standard during 1996 [2]. This means that MPEG-1 is upward compatible with MPEG-2 and ITU-T H.262. In other words, a MPEG-2 or a H.262 compliant decoder can receive a valid MPEG-1 transmission.

MPEG standards only prescribe the decoder structures and the bit stream formats. Designers are free to innovate on the development of encoder algorithms and structures. MPEG-1, 2, and 4 are lossy encoding standards. All use Discrete Cosine Transforms (DCTs) as the means to describe the information content of an image. Video image quality derived from lossy compression is a function of the image complexity and the compression technique's sophistication. Many different encoding techniques may develop as new image processing and higher performance electronics appear in the market place. Standards ensure MPEG decoders will continue to receive intelligible transmissions as technology evolves.

3.3 MPEG-2 — The High Definition Standard

MPEG established MPEG-2 (a.k.a. ISO/IEC Standard 13818) as a framework for superior video quality over high bandwidth networks. MPEG-2 provides a range of video quality beginning with NTSC standard quality and ranging upwards to International Radio Consultative Committee (CCIR) standard 601 quality. Video quality and content varies the encoding rates from 4 Mbps to 80 Mbps. MPEG-2's intended applications are digital cable television (CATV) distribution, digital video tape recording, digital satellite, and terrestrial broadcast.

MPEG-2 conventions cover much more than video compression. The standard prescribes methods for manipulating multimedia information. MPEG-2 maintains network independence by not prescribing the underlying information transmission network. The standard has ten parts individually referred to as ISO/IEC 13818-1 through 13818-10 [5], [6]. Parts 1, 2, and 3 are the central parts of MPEG-2. These parts were ISO/IEC approved in November 1994. Parts 4, 5, 6, 7, and 9 became standards during subsequent years. They support or extend the central parts of the standard. As of this writing, Part 10 is under development. Part 8 may never achieve approval for lack of industrial interest. The first five Parts of the MPEG-2 standard define the same functions as the corresponding Parts of MPEG-1. The remaining five Parts define new multimedia functions not found in MPEG-1.

ITU-T collaborated with MPEG during the development of MPEG-2's System and Video Parts (Parts 1 and 2). ITU-T subsequently ratified MPEG-2 within the H.262 standard. Thus, ISO/IEC 13818-1, ISO/IEC 13818-2, and ITU-T H.262 are common standards enjoying recognition within the world's leading standards bodies.

3.4 A Brief Overview of the MPEG-2 Standard

MPEG's greatest innovation lies within the video coding area. Therefore, the bulk of this study will focus on MPEG-2's video compression methods. Table 1 provides a high level comparison of MPEG-2 and its precursor MPEG-1 as an illustration of their relationship. Readers wanting a complete understanding of MPEG-2 should begin their study with the Joint Photographic Expert Group (JPEG) and MPEG-1 standards followed by a complete study of all parts of the MPEG-2 standard. The following sections provide a brief overview of the MPEG-2 standard.

Table 1. Typical MPEG-1 and MPEG-2 Coding Parameters [4], [7].

Parameters	MPEG-1	MPEG-2	
Standardized	1992	1994	
Main Application	Digital video on CD-ROM	Digital TV Profile	High Definition TV Profile
Spatial Resolution	CIF Format (1/4 TV) at 288 x 360 pels	576 x 720 pels	1152 x 1440 pels
Temporal Resolution	25 - 30 frames/s	50-60 fields/s	100-120 fields/s
Bit Rate	1.5 Mbit/s	approx. 4 Mbit/s	approx. 20 Mbit/s
Quality	comparable to VHS	comparable to NTSC/PAL for TV	near studio quality TV
Compression Ratio over PCM	approx. 20 - 30	approx. 30-40	approx. 30-40

3.4.1 Part 1: Systems

Part 1 (officially known as ISO/IEC 13818-1) covers MPEG-2 system standards and parallels the MPEG-1 Systems Multiplex. Part 1 defines multiplexing techniques for combining one or more elementary bit streams of compressed video, compressed audio, and other information into a composite bit stream. MPEG-2 segments elementary video, audio, and data bit streams into Packetized Elementary Streams (PESs). Packet headers specify the type of elementary stream contained in the packet payload area. The composite stream is suitable for storage or transmission. Two multiplexing formats are allowable under the MPEG-2 standard: the Program Stream and the Transport Stream. The two stream options have different applications in mind.

Program Stream multiplex supports video distribution and storage within an error-free environment (e.g., video production studio, computer storage). Packets within the Program Stream are variable in length and may be much longer than Transport Stream packets. MPEG-2 Program Stream is similar to MPEG-1 PES.

Transport Stream multiplex combines one or more PESs and their associated time bases into a single bit stream suitable for transmission over error-prone networks. A Transport Stream may carry a large number of programs. Programs consist of video, audio, or data elementary streams sharing a common timebase. Tables carried within the stream point to the location of the individual programs. MPEG-2 supports two special streams called ECM and EMM. These special streams carry information for deciphering encrypted programs. Copyright identification fields carried with each program enable monitoring of intellectual property as it flows through a network. MPEG-2 segments Transport Stream information into fixed length, 188 byte packets, a length chosen with ATM networks in mind.

Support for NS/EP video traffic should begin with a realization for traffic prioritization and information security features found within ISO/IEC 13818-1. As will be shown later, MPEG-2 has features for subdividing video information into an essential information (base) layer and optional (enhancement) information layers. ATM video encoders should incorporate prioritized cell discard at the ATM network interface. ATM flow control mechanisms should signal the encoder to discard enhancement layers, thereby reducing unnecessary traffic at its point of entry. Security is not, strictly speaking, an ATM network issue. However, encryption and conditional access mechanisms should be present in the ATM video transceivers in a manner that limits access to authorized receivers. For example, work is underway within the "pay-per-view" and "pay-TV" industry to implement these features in the next generation of cable TV receivers.

3.4.2 Part 2: Video

In principle, MPEG-2 standard, Part 2, is a superset of MPEG-1 video compression. It expands upon MPEG-1 techniques to provide a wider variety of coding tools. MPEG-2 organizes these tools and capabilities into application-relevant levels and profiles. See Table 2 and Table 3 for a description of MPEG-2's application structure. Applications can subscribe to a given level and profile without embracing the entire MPEG-2 standard. Profile compliance ensures compatibility within a given industrial application without the excessive cost of full standard compliance. A later section will discuss how MPEG-2 levels and profiles apply to the North American HDTV industry.

Table 2. MPEG-2 Profiles [4], [7].

Profile	Algorithms
HIGH	Supports all functionality provided by the Spatial Scalable Profile plus the provision to support three layers within the SNR and Spatial scalable coding modes 4:2:2 YUV-representation for improved quality requirement (See note below.)
SPATIAL Scalable	Supports all functionality provided by the SNR Scalable Profile plus an algorithm for Spatial scalable coding (two layers allowed) 4:0:0 YUV-representation (See note below.)
SNR Scalable	Supports all functionality provided by the MAIN Profile plus an algorithm for SNR scalable coding (two layers allowed) 4:2:0 YUV-representation (See note below.)
MAIN	Non-scalable coding algorithm supporting functionality for: ▪ coding interlaced video ▪ random access ▪ B-picture prediction modes 4:2:0 YUV-representation (See note below.)
SIMPLE	Includes all functionality provided by the MAIN Profile but does not support B-picture prediction modes 4:2:0 YUV-representation (See note below.)

Notes: "YUV" describes the ratio of luminance and chrominance information samples within video representation (e.g., "4:2:2" is a representation with two luminance samples for every chrominance sample).

3.4.3 Part 3: Audio

MPEG-2, Part 3, defines MPEG-2 audio as an extension of MPEG-1 standards. Its extensions support multichannel audio while maintaining backward compatibility with MPEG-1 audio elementary streams. An MPEG-1 decoder can decode up to two channels of an MPEG-2 stream. Conversely, an MPEG-2 audio decoder can decode an MPEG-1 stream in full compliance with the earlier audio standard.

Expanding upon MPEG-1, MPEG-2 has a maximum of 7.1 channels for digital surround sound [7]. There are five front channels (LL, LC, CC, RC, and RR from left to right) and two rear channels (LS and RS). The "0.1" channel drives a low-frequency sub-woofer for heavy bass.

Table 3. MPEG-2 Profile Parameters Upper Bound At Each Level [4], [7].

Level	Parameters
HIGH	1920 samples/line 1152 lines/frame 60 frames/s 80 Mbit/s
HIGH 1440	1440 samples/line 1152 lines/frame 60 frames/s 60 Mbit/s
MAIN	720 samples/line 576 lines/frame 30 frames/s 15 Mbit/s
LOW	352 samples/line 288 lines/frame 30 frames/s 4 Mbit/s

3.4.4 Part 4. Compliance Testing

Part 4 establishes conventional methods for standards compliance testing within new multimedia systems. MPEG-2, Part 4, corresponds directly to Part 4 of the MPEG-1 standard.

3.4.5 Part 5. Software Simulation

Part 5 prescribes MPEG-2 software simulation methods. MPEG-1, Part 5, is its counterpart.

3.4.6 Part 6. Digital Storage Media Command and Control

Digital Storage Media Command and Control (DSM-CC) became a standard in July 1996 as a specification for bit stream management protocol. DSM-CC defines a server and client model for the exchange of program material over heterogeneous networks. It creates a central supervision entity for managing the creation of client-server sessions and allocating media resources.

NCS should encourage modifications to ISO/IEC 13818-6 that support prioritization of video sessions across an ATM network. Admission control mechanisms and session priority controls could screen new video sessions at the ATM-video terminal interface with preference shown toward NS/EP traffic.

3.4.7 Part 7. Non-Backwards Compatible Audio Coding

Backward compatibility with MPEG-1 exacts a quality penalty. There arose a need for a higher quality audio standard for applications using MPEG-2, but not tied to MPEG-1.

Part 7 reached international standard status in April 1997 as the "Advanced Audio Coding" (AAC) standard.

Part 7 is a more advanced form of audio compression. It achieves higher audio quality at an encoding rate comparable to MPEG-1 audio compression, or it can achieve quality equivalent to MPEG-1 at 50% of MPEG-1's encoding rate.

3.4.8 Part 8. Untitled

MPEG-2, Part 2, quantizes the video sample at eight bits per sample. Part 8 would have been a 10-bit quantizing standard for use in the professional video industry. Work on Part 8 ceased when industry lost interest in its development.

3.4.9 Part 9. Real-Time Interface

Real-Time Interface (RTI) became an international standard in July 1996. Part 9 defines a real-time network adaptation layer between Transport Stream decoders and communication networks. It guarantees required network performance and, at the same time, enables the design of MPEG-2 decoders with appropriate buffer and timing recovery mechanisms. Consumer electronics, computers, games, and other multimedia devices may use Part 9 to ensure interoperability between devices and their networks.

3.4.10 Part 10. Conformance Testing of DSM-CC.

Part 10 is under development as of this writing.

3.5 Compression Requirements and Coding Rate Constraints

It is instructive to view MPEG-2 from the standpoint of the video encoder. Keep in mind that MPEG-2 sets no standards on the encoder. Rather MPEG-2 defines an information exchange interface between the video encoder and decoder. Within the context of the interface, MPEG only prescribes video encoder output and leaves designers free to innovate on means of achieving the prescribed result.

The goal of video compression is data reduction for storage and transmission [4]. Normally one thinks of information transmission as the spatial displacement of information from one geographic point to another. A transmission "network" may deliver the information in real-time (e.g., live TV broadcast) or delayed in time (e.g., pay-per-view movies). Notice the introduction of time displacement as a factor in video transmission. Temporal displacement implies program storage and retrieval. In some cases there may be no transmission network involved in the information delivery. For instance, image transmission may be the storage, sale, and viewing of a video compact disc (CD).

MPEG-2 draws no distinction between storage, retrieval, and transmission systems. The standard applies equally well to all types of transmission systems using any combination of spatial or temporal delivery. All of the transmission systems have a common constraint. They can only convey a finite amount of information. A video CD can retain only a certain amount of information encoded upon its surface. A TV channel or ATM network can convey only a certain amount of information per second. The transmission system sets upper limits upon the information transmitted across time and space.

Video compression is a compromise between the natural information content of an image, delivered image quality, and available transmission rate for its distribution. Video encoders strike a balance between these independent constraints by eliminating redundancy within the video image, varying the reproduced image quality, and, when possible, varying the instantaneous video encoding rate. The latter is only possible in transmission systems that support variable bit rate (VBR) transmission (e.g., ATM networks). Even VBR systems set constraints on peak and average transmission rate, so a transmission rate constraint exists albeit much less severe than in a constant bit rate (CBR) transmission system. The trick in all cases is to convey only as much information as required by the viewer's subjective requirement for image quality.

MPEG-2 video compression deals with three types of video information redundancy: subjective, spatial, and temporal. The amount of redundancy depends upon human perception and the video image. Compression reduces the encoded video bit stream by reducing all three types of redundancy under the triple constraint of perceived reproduction quality, video material content, and the transmission system capacity.

3.6 Satisfying the Human Eye — Image Sampling and Subjective Redundancy

Some conventions will help with the understanding of a complex environment. Think of an "image" as a scene with a background and one or more objects in motion as seen within a camera's field of view. For example, consider the image in Figure 1. A camera views the image as continuous three-dimensional information along a vertical axis, a horizontal axis, and a temporal axis. A "snapshot" holds the temporal dimension constant and captures the image's instantaneous two-dimensional or spatial information within the snapshot "frame." The term "picture" and "frame" are interchangeable references to a snapshot representation of an image's spatial information. Along any given axis, there are variations in the frame's luminance and chrominance. Variations along the spatial axes describe the image's form, color, and texture. Temporal axis variations are objects in motion, changes in scenery, changes in ambient lighting, or even transitions to a new image, such as panning the camera. Perfect rendition of the image requires the encoder to send a continuous stream of spatial and temporal information. This is not practical within most image delivery systems. Instead, the video encoder seeks compromises compatible with the viewer's requirement for perceived image quality.

3.6.1 Reductions in Subjective Temporal Redundancy

The human eye has a certain "persistence." That is to say, the eye cannot distinguish between continuous image change and a rapid succession of discrete frames. Experiments show that the eye perceives a continuous image if the discrete "frame rate" is greater than twenty frames per second (fps). The eye perceives the frame transitions as image "flicker" at lower rates.

MPEG-2 video encoders take snapshots of a camera's field of view at frame rates ranging from 24 to 60 frames per second. Higher frame rates imply greater image reproduction quality. Higher rates also imply greater amounts of information to be transmitted to the video decoder. Transmission systems set upper bounds on the information rate produced by the encoder. Therefore, encoder compression techniques must strive to reduce the transmitted information to an amount compatible with the transmission system capacity.

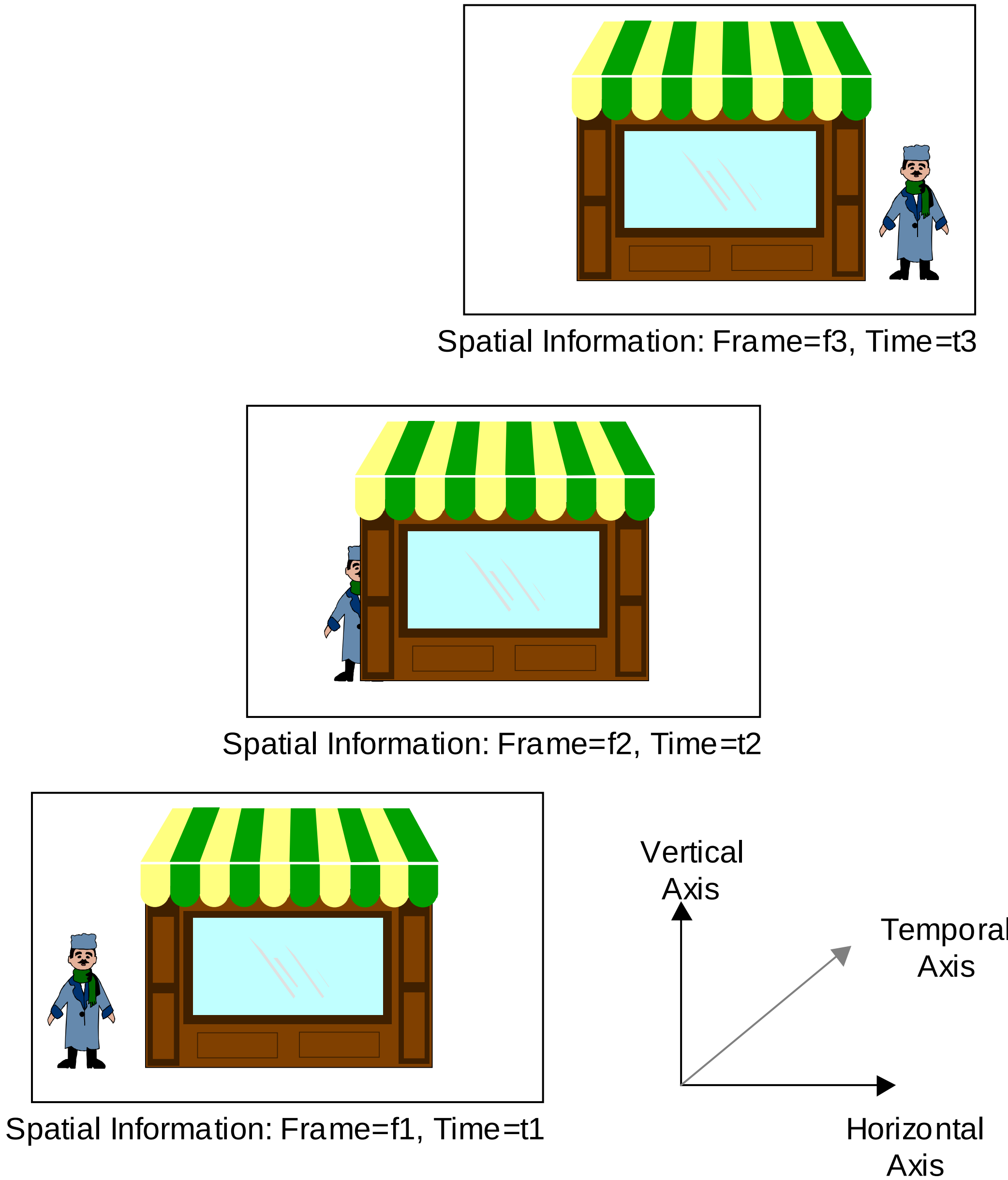
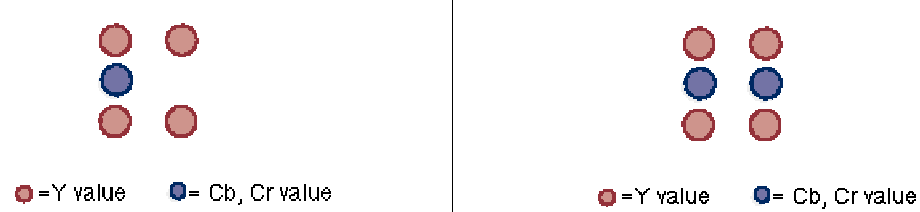


Figure 1. A Video Image and Its Interpretation [3], [5].

3.6.2 Reductions in Subjective Spatial Redundancy

Each frame has continuous variation in luminance and chrominance along both spatial axes. An MPEG-2 encoder samples the image in a left to right, top to bottom order in a process often called "frame scanning." Each sample is called a "pel." Some pels convey light intensity, i.e. luminance. Convention labels the luminance pels with the letter Y. Others pels convey one of two sampled colors. These are chrominance pels, and they are labeled either Cr or Cb. It can be shown that placing combinations of Y, Cb, and Cr pels in close proximity convey a full range of image intensities, colors, and textures to the human eye.

The human eye is much more sensitive to variations in luminance than variations in chrominance. MPEG-2 encoders cater to this difference in sensitivity and reduce transmitted information volume by sampling luminance and chrominance pels at different rates. As it scans from left to right, encoders produce clusters of pels in different formats. For example, an encoder may produce a pel cluster in the format shown in Figure 2a. This is the popular ATSC configuration known as "Y:Cr:Cb = 4:2:0." It contains four luminance pels, one Cr pel, and one Cb pel. The chrominance pels reside one on top of the other. A higher quality image results in Figure 2b's configuration known as "Y:Cr:Cb = 4:2:2," (a format prescribed by the CCIR Standard 601). CCIR's 4:2:2 format carries with it twice as much chrominance information as the 4:2:0 format. Both formats are MPEG-2 formats and represent different levels of picture quality within the standard.



a. 4:2:0 pel format.

b. 4:2:2 pel format.

Figure 2. Popular MPEG-2 Pel Formats [5].

MPEG-2 terminology defines a horizontal row of luminance and chrominance pels to be a "frame line." The human eye cannot distinguish pels and frame lines in close proximity. Instead, the human vision process blends the discrete image samples into an estimate of the original image. Creating a subjectively accurate estimate of the image requires pels and lines to appear at a regular frequency along the horizontal and vertical axes. The frequency of pels and lines is known as the "resolution" of the sampled frame. An adequately high resolution places combinations of pels in close proximity and satisfies the eye's perception of continuous color, texture, and detail. Table 2 and Table 3 describe the pel formats and resolutions supported by MPEG-2.

It is not possible to discuss information compression further without first discussing information structure. In fact, it is not possible to quantify redundancy without describing the organization of information and its interrelationship, known as its "correlation." It is possible to organize video information in different ways as evidenced by the NTSC and PAL formats for analog television broadcast. Digital video must also adopt an organization and then use that organization as a device for managing the flow of information from the encoder to the decoder.

3.7 Elementary Stream Structure

MPEG-2 defines an organizational structure for the video packetized elementary stream. At the lowest level of the structure, the standard organizes pel clusters into "blocks." A block is an 8x8 array of luminance or chrominance pels. Now consider Figure 3 composed of four consecutive pel clusters taken from four consecutive frame lines. The array defines a square area within the frame containing a complete block of Y pels, one quadrant of a Cr block, and one quadrant of a Cb block. If the figure contained the CCIR 4:2:2 format, then it would contain a complete luminance block, two adjoining Cb quadrants, and two adjoining Cr quadrants. A block is an important unit in MPEG-2 compression, but more about that later.

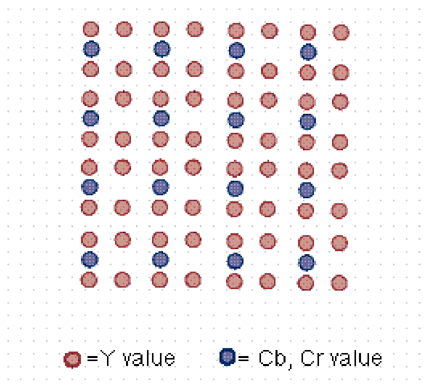


Figure 3. Block in 4:2:0 Format [5].

A block is a relatively small unit in the video elementary stream, so MPEG-2 collects four blocks into a square array called a "macroblock." A 4:2:0 macroblock contains a 2x2 array of luminance blocks, one complete Cr block and one complete Cb block, as illustrated in Figure 4 [5]. A set of consecutive macroblocks spanning the frame horizontally is a "frame row." It turns out that the macroblock is the smallest addressable unit within the video packetized elementary stream.

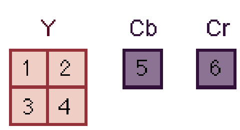


Figure 4. 4:2:0 Macroblock [5].

An MPEG-2 "slice" contains one or more consecutive macroblocks. Each slice begins with a uniquely formatted slice start code. Decoders use the start code to resynchronize with the bit stream if the preceding slice is in error. In this way, slices limit error propagation to no more than one frame row. There may be more than one slice per row. For example, the beginning of every ATSC row is also the beginning of a slice. This is allowable but not required under MPEG-2 standards. More slices mean quicker error recovery but the additional slice overhead may be at the expense of video coding rate or quality.

Figure 5 illustrates the relationship between blocks, macroblocks, slices, and a picture or frame. As mentioned earlier, frame and picture are interchangeable terms within MPEG-2 and both consist of all macroblocks within the active camera's viewing area. Slices within the picture subdivide the frame for quicker decoder synchronization and error recovery.

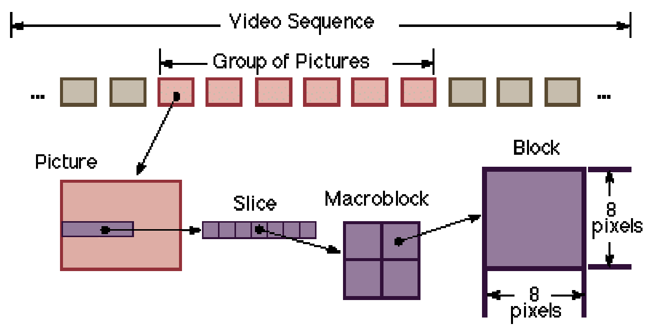


Figure 5. Video Elementary Stream [5].

MPEG-2 organizes one or more consecutive pictures into a group of pictures, or "GOP." GOPs act as boundaries, much as a slice, for inter-picture compression coding and as a way of maintaining time registration. GOPs are optional both in MPEG-2 and ATSC.

A video sequence consists of a collection of one or more consecutive pictures. It has a sequence header and end-of-sequence code. A video sequence may contain multiple headers. Each header acts as an entry point where the decoder may begin decoding the sequence. Picture (frame) sequences maintain a relationship important to interpicture (interframe) video compression. Upcoming discussion will elaborate on this relationship.

3.8 Video Compression and Information Redundancy

Statistically speaking, there is a great deal of similarity or "correlation" among adjoining pels. This similarity exists along the spatial axes and persists over consecutive frames. MPEG-2 relies upon inter-pel correlation as a means of reducing information redundancy. It assumes that the magnitude of a given pel can be predicted from nearby pels within the same frame using intraframe-coding techniques or from pels of nearby frames using interframe coding techniques.

3.8.1 Intraframe Correlation

Consider for a moment the simple statistical model provided by Thomas Sikora in Figure 6 [4], [7]. It illustrates high correlation for a given pel and its adjoining pels, a concept referred to as "spatial correlation." In Sikora's model, there is high correlation between adjacent pels and monotonically decaying correlation as the distance between pels increases.

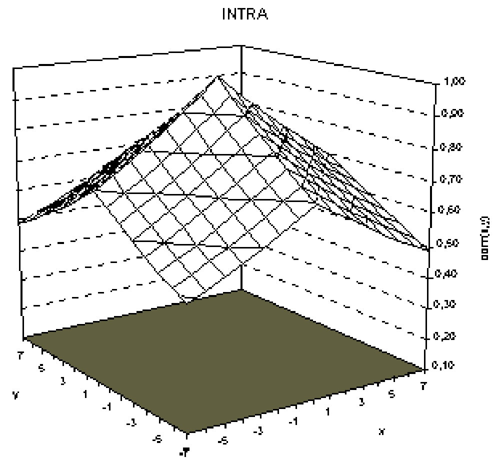


Figure 6. Block Spatial Correlation Model [7].

Consider the implications of an image with high local correlation and monotonically decreasing correlation with distance. The image's cross section is a set of discrete, non-negative luminance or chrominance intensities. The values have smooth transitions and few discontinuities. The Discrete Fourier series of this cross section has predominantly low-frequency coefficients and a zero-frequency, "DC" component. The majority of the higher frequency Fourier coefficients would be relatively small or approach zero. The usefulness of the Fourier series ends with its ability to only handle the one-dimension cross section. In its place, MPEG-2 makes use of another transform, the Discrete Cosine Transform, suitable to two-dimensional information.

3.8.2 Discrete Cosine Transforms

MPEG-1, MPEG-2, and the lossy form of JPEG analyze images using the DCT. The DCT converts pel intensities into a set of coefficients using the following equation [8].

$$F(u,v) = \frac{1}{4} C(u) C(v) \sum_{x=0}^{7} \sum_{y=0}^{7} f(x,y) \cos\left[\frac{(2x+1)u\pi}{16}\right] \cos\left[\frac{(2y+1)v\pi}{16}\right]$$

where x and y are pel indices within an 8x8 block, u and v are DCT coefficient indices within an 8x8 coefficient block, and:

$$C(w) = \frac{1}{\sqrt{2}} \quad for\ w = 0$$

$$C(w) = 1 \qquad for\ w = 1, 2, \ldots, 7$$

Thus an 8x8 array of pel intensities f(x, y) is the input to a mathematical formula and the output is an 8x8 array of frequency coefficients, F(u, v).

If the MPEG-2 encoder uses the DCT to express and transmit the block coefficients, then the decoder can recover the original pel intensities by applying the inverse transform to the coefficients. The Inverse Discrete Cosine Transform is given by [8]:

$$f(x,y) = \frac{1}{4}\sum_{u=0}^{7}\sum_{v=0}^{7} C(u)C(v)F(u,v)\cos\left[\frac{(2x+1)u\pi}{16}\right]\cos\left[\frac{(2y+1)v\pi}{16}\right]$$

Applying a DCT to the block in Figure 6 produces an 8x8-coefficient block similar to the one shown in Figure 7. MPEG-2 encoders apply the DCT to luminance and chrominance blocks within each frame. DCTs produce blocks of luminance and chrominance coefficients, many with the distribution shown in Figure 7.

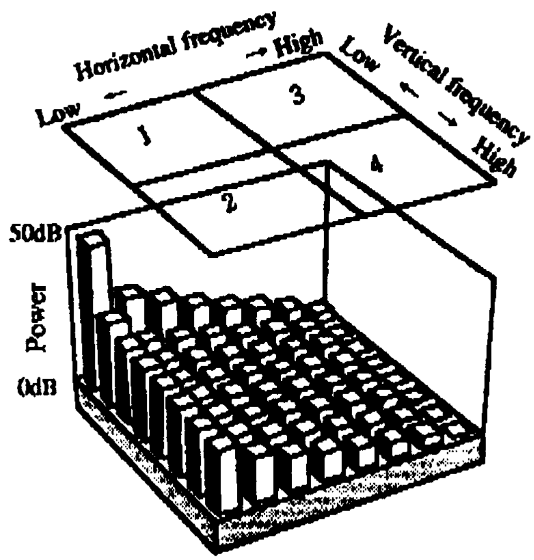


Figure 7. Block DCT Coefficients [9].

3.8.3 Quantization and Spatial Redundancy

At first glance, the transformation seems to be an even trade, i.e. 64 continuous pel intensity values for 64 continuous DCT coefficients, resulting in no reduction in the amount of information to be transmitted. Now suppose circumstances permitted discarding many of the coefficients without reducing the perceived image quality. An encoder could send the essential coefficients with a net reduction in transmission. As before, the solution lies within the human eye.

Studies of the human eye lead to ways to reduce information without materially affecting perceived video quality. Earlier, knowledge of the eye's preference to luminance permitted a reduction in subjectively redundant chrominance information. It happens that the eye is much more sensitive to the average intensity of the frame and to low frequency, gradual changes in the image composition. The eye is less sensitive to fine, high-frequency detail and sharp transitions. In terms of DCT coefficients, the eye is more sensitive to error in the DC and low-frequency coefficients. Missing information in the high-frequency coefficients is not as noticeable and will, with precautions, be overlooked.

MPEG-2 reduces spatial redundancy by taking advantage of the eye's spatial preferences. Notice in Figure 7 that the DCT coefficients tend to decrease in value with increasing spatial frequency. This is a natural outcome of a DCT applied to an image with high local correlation and monotonically decreasing spatial correlation, properties found to be statistically common in video images. Recall that the DCT produces a continuous set of values for all 64 block coefficients. MPEG-2 reduces these values to a binary form by quantizing each coefficient into eight bits. If the encoder quantizer were linear, then it would give equal accuracy to all of the block coefficients. Many of the higher frequency coefficients would take small, nonzero values. The eye tends to overlook this high-frequency information, so sending it to the decoder is not as valuable as sending detailed low-frequency information.

An MPEG-2 quantizer is nonlinear. That is to say, it assigns values to the coefficients in a nonlinear fashion. The quantizing steps are small and assign greater accuracy to DC and low-frequency coefficients with deference to the eye's sensitivity to these terms. On the other hand, MPEG-2 quantizer steps are larger for high-frequency coefficients causing many coefficients to round down to zero. Typically, a quantized block will have a few nonzero coefficients at the low frequencies and the balance of the block coefficients will be zero. See Figure 8 as an example of a nonlinear quantizing result.

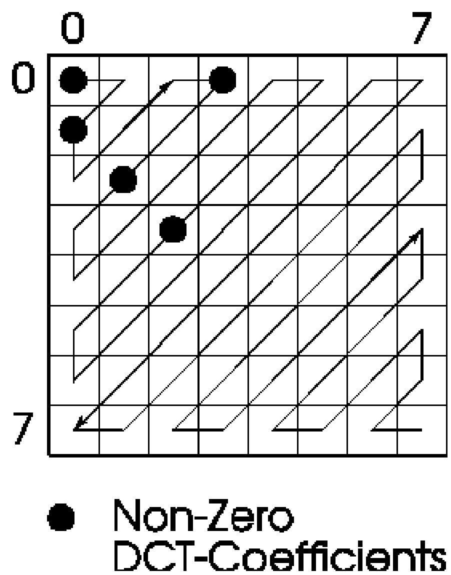


Figure 8. Quantized Coefficients and Zigzag Run-Length Encoding [7].

3.8.4 Huffman Variable-Length Coding

MPEG-2 applies Huffman run-length coding, more generally referred to as "variable-length coding (VLC)," to a quantized block in a zigzag fashion as shown in Figure 8. Reading in zigzag order maximizes the number of consecutive zeroes passed to the run-length encoder and shortens the binary description of the block. For a given image, the run-length-encoding rate becomes smaller as the quantizer increases the number of zero coefficients in the macroblock. Figure 9 summarizes the entire intraframe-encoded process. The process reduces the spatial redundancy within a single frame and produces an intraframe-encoded frame, called an "I-frame."

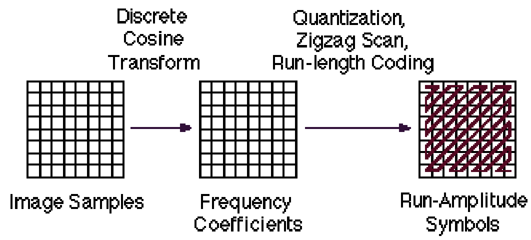


Figure 9. Intraframe (I-frame) Encoding [5].

Quantizing the block coefficients leads to a lossy image. In other words, the reproduction is not an exact recreation of the original image. Almost all of the MPEG-2 information "loss" takes place within the quantizer. The amount of loss depends upon the amount of quantizing nonlinearity applied to the macroblock coefficients.

Encoders vary quantizer accuracy and the corresponding encoding rate by assigning a scaling factor to each macroblock. The scaling factor specifies the amount of nonlinearity applied to the macroblock's coefficients. When transmission capacity permits, encoders will reduce the scaling factor and, in so doing, increase the received image accuracy and the amount of information transmitted to the decoder. Conversely, when transmission system capacity is overtaxed, MPEG-2 encoders increase the scaling factor driving more coefficients to zero with corresponding reductions in coding rate and delivered image quality. The combination of variable nonlinear quantizing and run-length encoding reduces the spatial redundancy within a given frame.

ATM networks and video encoders should interact in manners that set priorities for the creation of video connection, allocation of network bandwidth, and regulation of traffic flow. As will be discussed in greater detail in later sections, video encoders should show preference toward NS/EP priority traffic when setting their coding rate. Interactions between the encoder and ATM network interface should adjust the video coding rate to accommodate bandwidth made available to NS/EP and non-NS/EP traffic. Prioritized coding rate variation and discarding nonessential video information at the encoder would limit the entry of lower priority traffic into the ATM network, thereby reducing congestion especially during times of crisis.

3.8.5 Motion Compensated Prediction

Pel correlation, as seen in Figure 6, holds across multiple frames so long as there is no motion. Lacking any change, there would be perfect correlation between consecutive frames and all of the frames, but one would be redundant. Subtracting corresponding pel values within a given block from the same block in a neighboring frame would yield an all-zero "residue" block. The DCT of the residue would be an 8-by-8 array of zero coefficients. MPEG-2 run-length encoders require very few bits to describe a zero residue block. Thus, any steps taken that make nearby frames correlate result in a reduction in the transmitted coding rate.

Figure 10 shows the correlation of a block from the Nth frame within a search area of the preceding frame. MPEG-2's "motion compensation" algorithm finds the block in the prior frame exhibiting the highest correlation with the given block. It "compensates" for inter-frame by creating a motion vector describing the displacement of the current block relative to the matching block. The algorithm subtracts the prior frame from the compensated frame yielding a "residue," i.e., the difference between the frames after motion compensation. MPEG-2 performs a DCT on the residue and quantizes the result, forming a "prediction" relative to a prior frame. MPEG-2 convention refers to the prior frame as the "anchor frame." The encoder sends the anchor frame, the quantized residue coefficients, and the motion vector to the decoder.

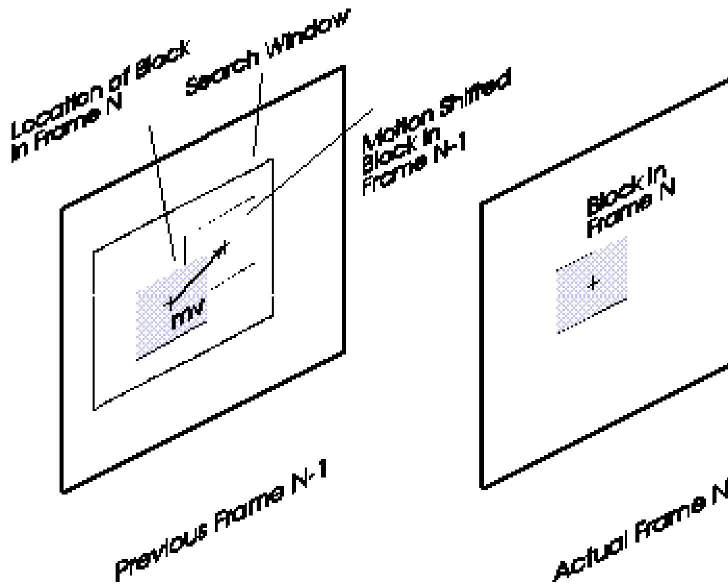


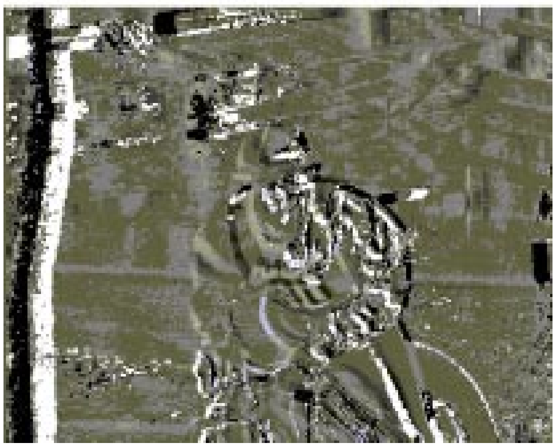
Figure 10. Motion Compensation Using Block Matching [4], [7].

Figure 11 uses two consecutive frames to illustrate the motion-compensated prediction process. Figure 11a represents the Nth frame in a video sequence. Motion exhibited by the jockey in Figure 11a spatially relocates objects appearing in the prior frame (N-1 frame). Object geometry and shading may vary between frames as well. Even so, there is great similarity between the macroblocks depicting the objects in consecutive frames if allowances are made for motion. Therefore, blocks in Figure 11a have high correlation with a displaced set of blocks in a prior frame. MPEG-2 finds block correlation through a process called "motion compensated prediction." In so doing, MPEG-2 seeks to reduce the encoder's coding rate by reducing duplication between frames, referred to as "temporal redundancy."

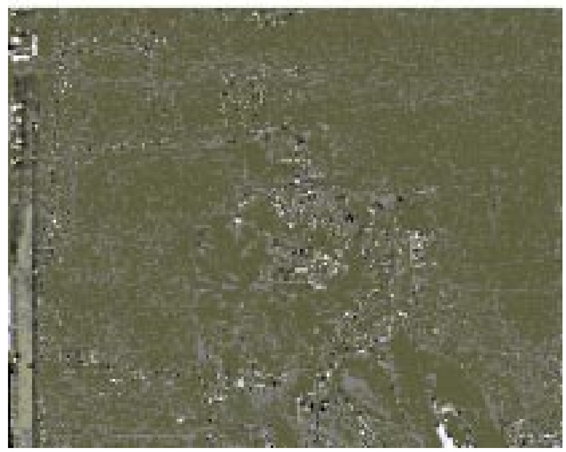
Refer again to the jockey in Figure 11. Figure 11b is the prior frame with white marks representing MPEG-2 motion vectors. Subtracting Figure 11a from Figure 11b without motion compensation yields Figure 11c. Notice the large amount of residual detail representing frame decorrelation caused by motion. Compare this result with the motion compensated version in Figure 11d. Very little residual detail remains. This series of figures demonstrates why inter-frame compression using motion compensated prediction yields the largest portion of MPEG-2's coding rate reduction.

a. Frame N to be encoded.

b. Anchor frame N-1 with Frame N motion vectors overlaid.



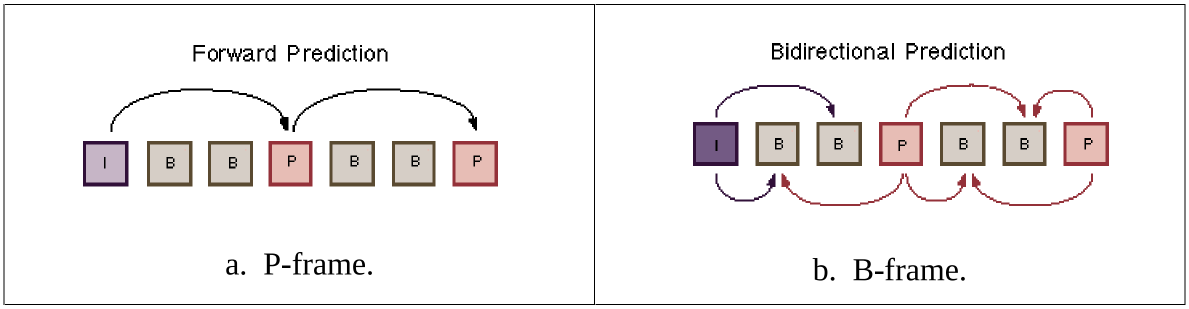
c. Prediction error, i.e., Residue = (Frame N – Frame N-1), with motion vectors set to zero.



d. Prediction error, i.e., Residue = (N – (N-1)), with motion vectors from (b).

Figure 11. Motion Compensation [4], [7].

MPEG-2 defines three types of compressed frames. MPEG encoders create I-frames solely from intraframe encoding as discussed earlier. I-frames constitute the lowest degree of video compression. Interframe encoding relies on anchor frames as the basis for motion compensated prediction. Frames that rely exclusively on preceding anchor frames are progressively encoded frames, called "P-frames." P-frames may reference earlier I-frames or P-frames as anchor frames. So-called "B-frames" may reference prior anchor frames, subsequent anchor frame, or both. Figure 12 illustrates the differences between P- and B-frame prediction.



a. P-frame.
b. B-frame.

Figure 12. Motion Prediction Frames [5].

Certain rules apply to the interframe encoding process. I-frames require no anchor frames as references and may be decoded immediately upon arrival at the MPEG-2 receiver. Anchor frames must arrive at the decoder prior to any P-frames and B-frames that reference them. P-frames may act as anchor frames for B-frames and other P-frames. B-frames may not act as anchor frames.

These rules place requirements on the encoder and decoder. Encoders send B-frames out of sequence. MPEG-2 decoders must hold anchor frames in memory while subsequent P- and B-frames are decoded. Decoders must reorder the frame sequence prior to displaying the decompressed video image. Figure 13 illustrates the differences between the frame encoding order and the frame transmission order.

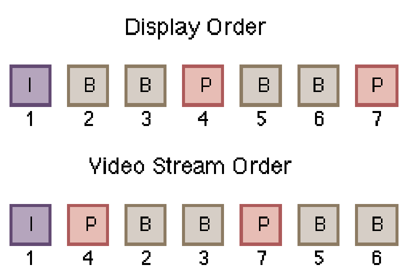


Figure 13. Frame Encoding Versus Transmission Order [5].

I-frames must appear at least once in every 132-frame sequence. I-frames refresh the decoding process by mitigating DCT drift at the decoder. They also act as entry points into the decoding process (for example, during a TV channel change or when entering a recording mid-track). I-frames may appear more frequently in applications where random access is important [5]. Figure 14 illustrates a sequence with I-frames appearing two times per second in the transmitted elementary stream.

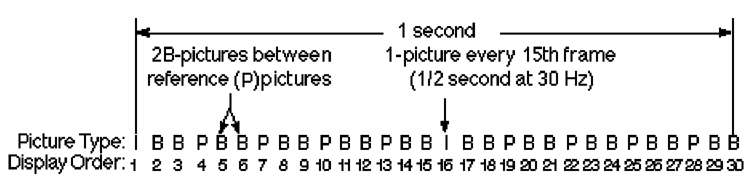


Figure 14. Periodic I-Frame Refresh [5].

3.8.6 Frames and Fields

Many video cameras treat each frame as two interlaced fields. The odd lines may appear in the first field and the even lines may appear in the second. MPEG-2 permits the encoder to either combine the fields into a composite frame or compress the fields separately. Encoders electing the first choice create I-, P-, and B-frames as discussed earlier. In the latter case, encoders create I-fields, P-fields, and B-fields in the same manner as described for frame encoders. A decoder receives, reorders, and displays the fields in the correct sequence in the same manner as a frame decoder.

3.8.7 Huffman Encoding

MPEG-2 merges the frames (or fields) and their associated motion vectors into a video elementary stream. Bit patterns may frequently repeat within the resulting bit stream and represents bit pattern redundancy. MPEG-2 achieves a further reduction in the encoding rate by passing the stream through a variable-length Huffman encoder. Huffman encoders assign a shorter representation to recurrent bit patterns according to the frequency of their occurrence.

3.9 MPEG-2 Profiles

MPEG-2 addresses a broad range of possible applications including a digital version of conventional TV, high definition TV, and broadband video conferencing. The standard offers an equally broad range of parameters from which applications may choose. Many applications require only a subset for which full implementation of all MPEG-2 options represents unnecessary cost. To avoid unnecessary complexity, MPEG grouped MPEG-2 parameters into subsets called "profiles," as shown in Table 2. Each profile represents a different degree of encoder-decoder sophistication. Within the profiles, applications may choose from a set of image quality "levels," i.e., image resolution, frame rates, and resulting coding rate (See Table 3). Manufacturers may choose to support a given application profile up to a certain level of image quality according to the application's requirements. Support for a common profile and level assures equipment compatibility across equivalent brands of encoders and decoders. Lower profiles and levels are upward compatible with higher profiles and levels, thereby assuring portability and larger market interest.

3.10 Scalability

The profiles above the Main Profile support scalability tools. Scalable video encoding encourages interoperability between services (e.g., standard definition TV (SDTV) and HDTV) and a mix of receiver capabilities. Scalable encoders create layered bit streams containing subsets of the full resolution video as illustrated in Figure 15. MPEG-2 supports up to three scalable layers within a video elementary bit stream.

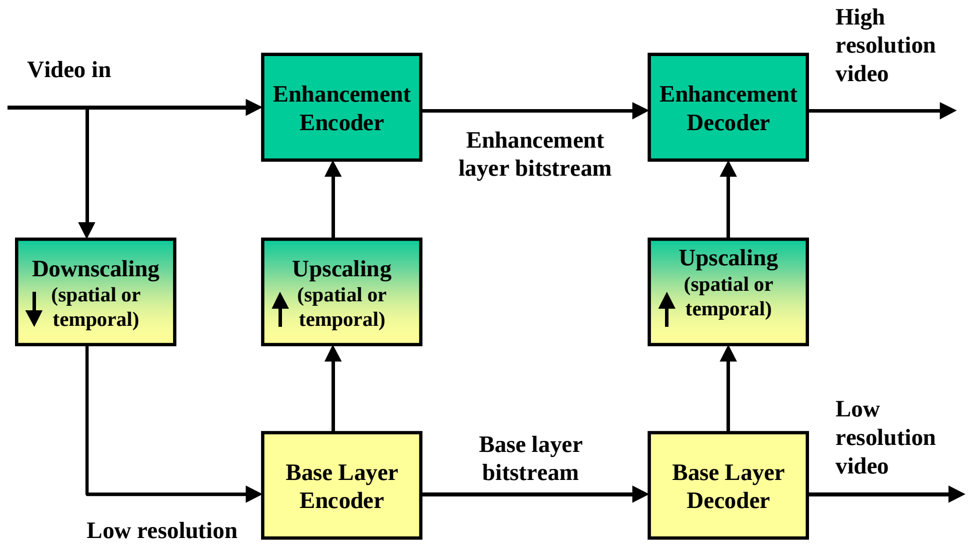


Figure 15. Two-Layer Scalable Video Elementary Stream [4], [7].

Scalable receivers may decode part or all of a high-resolution stream depending on receiver display capabilities or network conditions. For example, an SDTV receiver might decode and display a scalable HDTV bit stream at an SDTV-compatible resolution and frame rate. Layered bit streams permit flexibility within transmission networks as well. For example, an ATM-based video network may assign higher priority and extra error protection to low resolution layers, thereby assuring basic video quality during periods of congestion or low signal-to-noise ratio (SNR). All together, there are three scalability options.

SNR scalability, mentioned in the preceding example, subdivides high-resolution bit streams into resolution substreams. Low-resolution bit streams receive higher levels of error protection and higher transmission priority. Decoders may reduce video quality gracefully by decoding lower resolution layers when higher resolution layers are unavailable. The higher layers may be absent either because the video encoder did not send at high resolution or because the higher resolution layers were lost to corruption or congestion.

Spatial scalability facilitates interoperability between video systems by separating image detail into spatial layers with each layer representing a different display resolution. Lower resolution video receivers may decode lower resolution layers without the waste of receiving a separate lower resolution transmission. Higher resolution receivers may decode some or all of the layers depending on the viewer's quality requirement. For example, a scalable television receiver with "picture in a picture (PIP)" features might display the primary program at high resolution while displaying the smaller caption at low resolution. A scalable decoder could easily interchange the program and PIP displays with a corresponding change in their resolutions.

Temporal scalability provides for interoperability between stereoscopic and conventional receivers. Stereoscopic encoders send left and right view images as a base layer and an enhancement layer, respectively. Receivers may display a single view for a conventional video display or, if supported, both views for three-dimensional effects. These three-dimensional enhancements are optional and could be discarded during times of national crisis to conserve available bandwidth for NS/EP traffic without serious impact to conventional, two-dimensional video quality.

MPEG-2's scalability options offer good opportunities for robust and interoperable video distribution systems provided steps are taken within the video codecs and ATM networks. Video codecs should support more than one form of MPEG-2 scalability, thereby affording the maximum flexibility when dealing with codecs from different manufacturers. Interaction between a video encoder and its ATM network should negotiate the creation and delivery of video layers according to the capabilities of the receiving decoder and available network bandwidth. During times of NS/EP crisis, scalability and interoperability standards would show preference to NS/EP traffic, reduce nonessential video layers at their source, and foster rapid deployment of available resources through codec interoperability.

3.11 ATSC and MPEG-2

ATSC's Digital Television Standard (DTS) is an interesting example of MPEG-2 modified to address regional television market [8]. DTS complies with the Main Profile of MPEG-2. It supports random access and the conventional 4:2:0 pel format. The North American standard supports most, but not all, aspects of the High Level. Thus, a Main Profile, High Level, receiver can fully decode an ATSC DTS-encoded video stream, but the converse will not be true in all cases. DTS does not support scalable bit streams. Consequently, North American TV stations must send DTS quality and NTSC quality programs as separate broadcasts, which leads to duplication of network traffic and a lack of interoperability.

DTS departs completely from MPEG-2 audio by adopting the Digital Dolby AC-3 audio compression standard. Dolby Digital is a stereo, "surround sound" audio system designed for digital videodisks. It is a "4.1" channel format in which the "4" specifies the presence of four full fidelity, surround sound channels, and the "0.1" specifies a low-frequency, narrow band channel for a heavy bass subwoofer [8], [10], [11]. In contrast, high definition PAL follows the ITU H.262 standard and uses the MPEG-2 audio compression described in Section 3.4.3. Thus, MPEG-2 Main Profile, High Level, and PAL receivers can decode DTS video bit streams, but not DTS audio bit streams.

ATSC designed DTS primarily for the North American open-air broadcast market. As designed, DTS produces an average unencoded bit rate of 1 Gbps [12]. Application of MPEG-2 techniques encodes DTS at a 50:1 compression ratio and produces a 19 Mbps constant bit rate transmission. The transmission is specifically designed for the existing 6 MHz broadcast channel common to North American open-air and CATV systems.

ATSC's departures from the MPEG-2 standards create at least two concerns for NS/EP traffic: priorities for nonessential video enhancement layers and codec interoperability. ATSC's lack of video scalability (layering) features prevents priority discard of optional video enhancement layers at the ATM interface. Regrettably, admitting ATSC video into an ATM network is an "all or nothing" decision.

ATSC is a North American standard and is unlikely to appear in markets outside the region. Consequently, there will be interoperability issues related to audio reception between MPEG-2 compliant and ATSC compliant codecs. Strictly speaking, this is not an ATM network issue. After all, both codecs could be fully ATM compliant. However, incompatibilities between the two standards make the codecs incompatible and will impede the exchange of NS/EP traffic. These incompatibilities will likely be felt during multi-national activities such as international diplomacy, joint military operations, and multi-national disaster relief efforts. NCS should encourage options in the ATSC standards for MPEG-2 audio reception and should encourage codec manufacturers and video broadcasters to accommodate both standards within their systems. Their systems should incorporate both ATSC and MPEG-2 compatibility in much the same way as public switched telephone network (PSTN) modems currently support multiple standards, (e.g., ITU-T V.34, ITU-T V.90).

4.0 ASYNCHRONOUS TRANSFER MODE

The International Telecommunications Union (ITU) formulated the ATM standard in response to a need for a high speed switching technology. Earlier standards proved inadequate for transmission speeds above a few hundred kilobits per second. Demands for high-speed data and video switching brought about ATM as a multi-gigabit per second switching technology capable of addressing applications like HDTV, multimedia Internet, complex imagery, and video conferencing [2]. There are capabilities and limitations within ATM that affect the overall design of a video distribution network.

4.1 ATM Switching Mechanisms

ATM defines the exchange of information between two communicating entities in terms of "virtual connections (VCs)." ATM transfers information over either permanent virtual connections (PVCs) or switched, i.e., temporary, virtual connections (SVCs). In either case, ATM subdivides the information exchange into fixed length, 48-byte cells and adds a 5-byte cell header as shown in Figure 16.

ATM supports both continuous bit rate (CBR) and VBR applications as appropriate to the codec design. The primary interest of this paper is the proper operation of variable coding rate codecs operating over variable bit rate virtual channels. The remaining discussion focuses on system considerations for the reliable delivery of video over variable rate codec and virtual channel combinations.

Bits

8 7 6 5	4 3	2 1	Bytes ↓
Generic flow control (GFC)	Virtual path identifier (VPI)		1
Virtual path identifier (continued)	Virtual channel identifier (VCI)		2
Virtual channel identifier (continued)			3
Virtual channel identifier (continued)	Payload type identifier (PTI)	Cell-loss priority (CLP)	4
Header error control (HEC)			5
Cell payload			6 • • • 53

Figure 16. ATM Cell Format [5].

ATM combines, relays, and distributes cells over a selected set of switches and bearer services using virtual path and channel identifiers. This simple but effective cell routing algorithm lends itself to hardware realization for switching speeds of several gigabits per second. Hardware switching speeds make ATM uniquely suited for distribution of high definition TV and other high-volume, video transmissions.

Switches servicing variable coding rate codecs can encounter more cell traffic than available bearer services will handle. ATM manages traffic congestion by using information stored in the generic flow control and cell loss priority fields. ATM responds to impending congestion by regulating the flow of new traffic into the network. It invokes admission control and flow control mechanisms that reject new SVCs and reduces cell submission rate of established virtual connections. Periods of heavy cell traffic may lead to switch buffer overflow and cell loss. Encoders may specify cell importance using the cell loss priority (CLP) field. ATM discards low priority cells during severe switch congestion.

MPEG-2 session establishment and ATM's admission control methods should interact at the network interface in a way that sets priorities with preference for NS/EP traffic. Video codecs should employ the CLP field as a means of setting higher priority for base layer bit streams essential to basic NS/EP video transmissions. ATM would discard lower priority enhancement layer cells, thereby improving the likelihood of good NS/EP communication during periods of congestion. As will be seen later, the video codec has several ways of concealing distortion introduced by cell loss, especially for enhancement layer cells.

ATM provides error recovery primarily as a means to avoid misrouted cells. The header error control field corrects one error and detects up to two errors in the header [13]. ATM discards cells with uncorrectable errors and may misroute cells with undetectable header errors. ATM provides no error correction for the payload area. Mechanisms external to ATM are responsible for the reliable delivery of data and the definition of acceptable levels of performance. It will be shown later in the document that HEC error correction and detection is not, by itself, adequate for many types of digital video transmission.

4.2 A Separation of Switching, Error Correction, and Data Integrity

ATM departs from the philosophies of earlier packet switching technologies. IBM's System Network Architecture (SNA), the ITU's X.25, and X.75 standards all include error correction at the data link layer and automatic repeat-request (ARQ) packet recovery at the virtual connection layer. Protocols like SNA, X.25, and X.75 compensate for channel impairments and switch congestion, thereby ensuring the reliable delivery of data packets.

Cell loss either through misrouting, error discard, or switch buffer overflow affect end-to-end data integrity. ATM's HEC error correction mechanism reduces, but does not eliminate, the probability of discarded or misrouted cells. Likewise, ATM's flow control and admission control methods reduce, but do not eliminate, the possibility of switch congestion, cell loss to buffer overflow, or low-priority cell discard. ATM provides no mechanism for recovering lost cells.

In comparison, ATM takes a much narrower approach to information transmission with high-speed switching as its primary objective [14]. ATM only provides error correction for the cell header for the purpose of reliable cell switching. Payload errors go undetected. There is no mechanism built into ATM for the purpose of mitigating channel impairments. ATM does not concern itself with payload bit errors, cell losses, or end-to-end data integrity. These matters are left to the designers of bearer services and network terminal devices.

ATM assumes other components within the communication network will compensate for errors and lost cells according to the needs of the individual application. For example, Figure 17 illustrates a possible network arrangement for high-quality video transmission over ATM. In the figure, video propagates from the video source downward through five layers of protocol and upward through the same five layers to the video receiver. The MPEG-2 encoder layer accepts video from the source and compresses the video for efficient transmission. Beneath the entropy encoder lies a transport encoder. The transport encoder organizes the compressed video, adds transport control information, and submits the resulting video message to the ATM network interface. Taken together, the MPEG-2 and transport encoders comprise what is often called a "video encoder [15]." Figure 17 places an ATM switch between the bearer channel, which may be one of several different types, and the video encoder or decoder. The bearer channel formats the ATM cells into a media-compatible form and transmits the result across a physical transmission media. The layered protocol process reverses itself at the receiver as it passes upward through the bearer channel decoder, ATM switch, transport decoder, and entropy decoder. All of the encoding, switching, transmission, and decoding steps make up an ATM-based, video "virtual connection" between the source and receiver. Notice that all of the application and media dependent processes take place outside the ATM switching layer including those matters related to error concealment.

ATM standards delegate payload error concealment to the bearer channel or the video codecs. Individual applications must tailor a combination of error compensation mechanisms in the layers surrounding ATM sufficient to meet its particular needs. It is desirable that this tailoring should take place on a connection-by-connection basis so varying combinations of terminal devices and bearer channels operate together effectively, especially during national emergencies where tailoring should show preference to NS/EP traffic, but more about this later. First, it is necessary to discuss the characteristics of the layers making up the ATM virtual connection.

5.0 BEARER CHANNEL CHARACTERIZATION

Three types of video bearer channels frequently appear in ATM network research. Fiber optic channels provide high speed and high transmission quality. Satellite channels offer high speeds and geographic coverage. Line-of-sight radio channels promise tetherless mobility. Transmitting digital video over these bearer channels presents different challenges to the ATM network designer.

5.1 Fiber Optic Channel

Fiber optic transmission systems offer the most reliable and economical form of transmission available today. Network designers use fiber optic cables in building LANs, constructing campus area networks, and spanning continents. It seems natural that high-quality video transmissions would travel over fiber channels.

Fiber channels provide exceptionally high data rates (up to several Gbps) at relatively low error rates. Typical fiber channel bit error rates (BERs) are 10^{-9} errors per bit (epb). Such low error rates could lull the inattentive into a false sense of security. Consider a variable bit rate, HDTV signal operating at an average coding rate of 19 Mbps over a fiber bearer channel. Standard HDTV transmissions display thirty frames per second containing, on average, 1,494 ATM cells per frame. Two or more errors in the cell header cause the cell to be lost and a frame distortion. Assuming all bit errors are independent random events, the probability of a discarded or misrouted cell is a binomial distribution over the forty header bits. Figure 18 illustrates the probability of cell loss as a function of the bit error probability. The probability of erroneous header cell loss for a single bearer channel operating at a 10^{-9} BER is negligible at $2 \bullet 10^{-15}$. It is unlikely that error-induced cell loss would cause frame distortion over a single fiber channel.

Now consider how the payload error rate impacts a video distribution network. Define the number of ATM bearer channels connected in tandem in terms of the variable N. Random payload errors are also a binomial process operating upon the 384 payload bits. The graph in Figure 19 shows a point-to-point fiber channel (N=1) experiencing 10^{-9} epb. A single fiber channel has an aberrant payload probability of $3.8 \bullet 10^{-7}$. The combined effect of payload errors and error-induced cell losses has a combined frame error probability of $5.7 \bullet 10^{-3}$, as shown in Figure 20. At 30 frames per second, a point-to-point ATM channel will average one frame error every minute. So, for the trivial case of N=1, there is little need for error correction; and complacency seems justified when using fiber bearer channels.

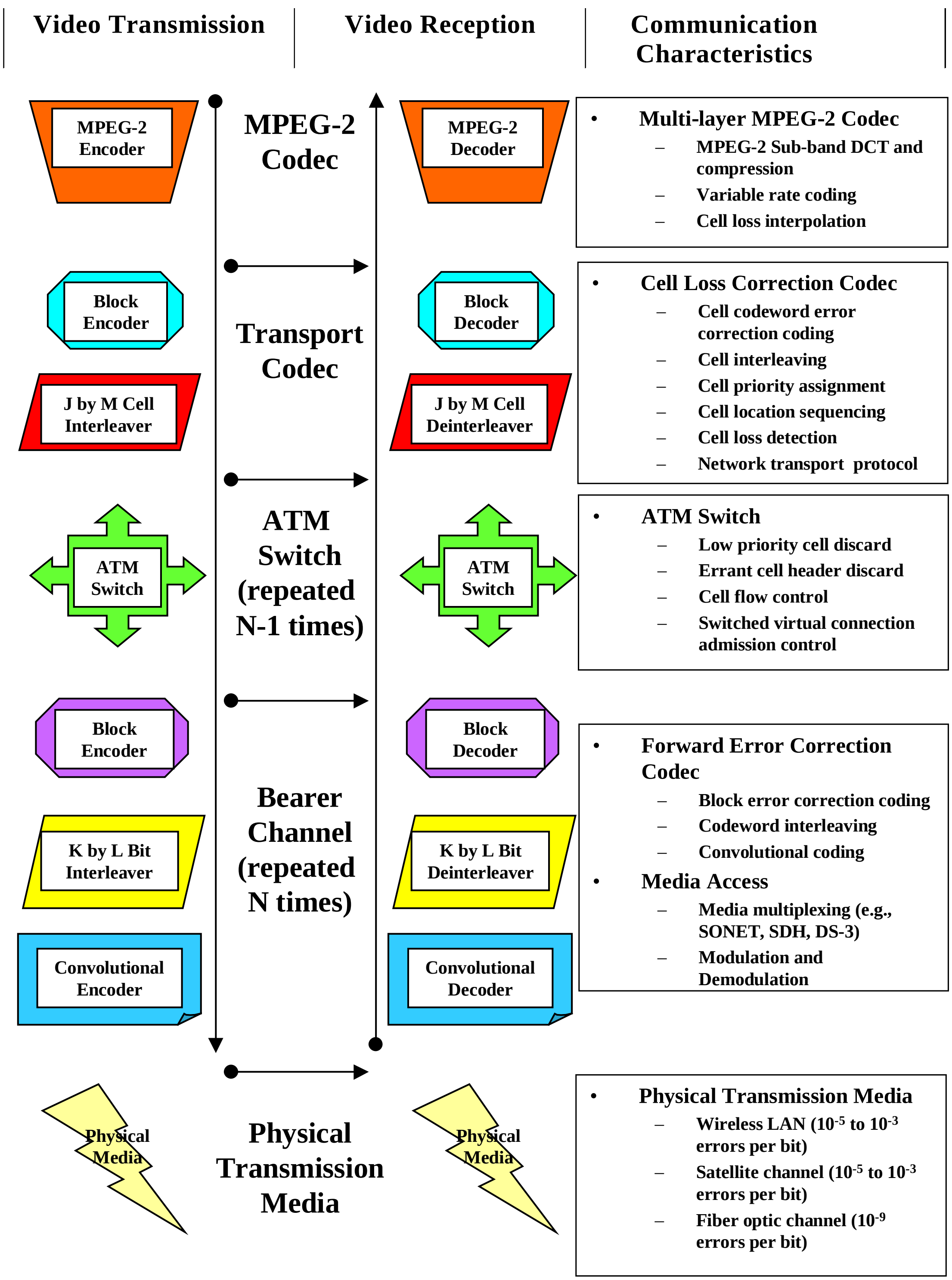


Figure 17. Digital Video Over ATM Architecture (four layers of protocol).

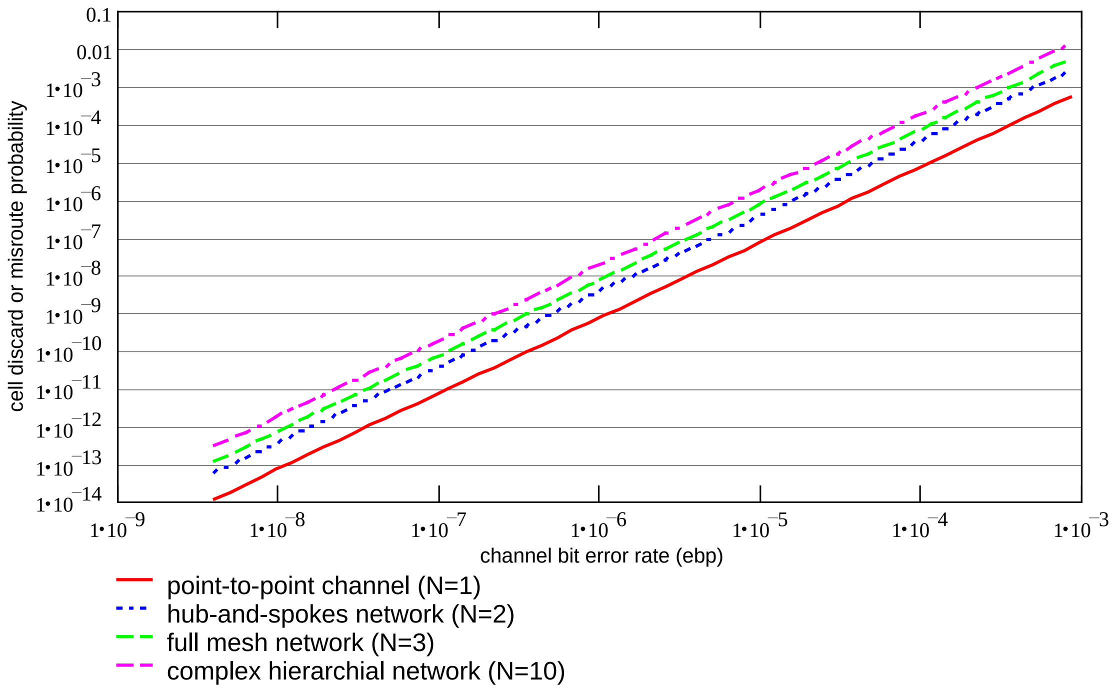


Figure 18. Cell loss Probability for N Tandem Channels.

A more realistic video distribution network consists of some number of fiber bearer channels connected in tandem by one or more ATM switches. For instance, consider a hub-and-spokes network in which a video transmission traverses one switch and exactly two fiber channels. From Figure 20, the error rate is $1 \bullet 10^{-3}$ errors per frame, or two frame errors per minute. Next, consider a large distribution network like the Internet or a pay-per-view cable system. The number of switches increases greatly and so does the number of fiber channels connected in tandem. For example, Figure 20 illustrates the case for ten fiber channels traversing nine ATM switches. The frame error rate increases to ten frames per minute, which is unacceptable for high-quality video distribution. Clearly, video quality decreases as the number of tandem channels increases until some form of error correction is necessary.

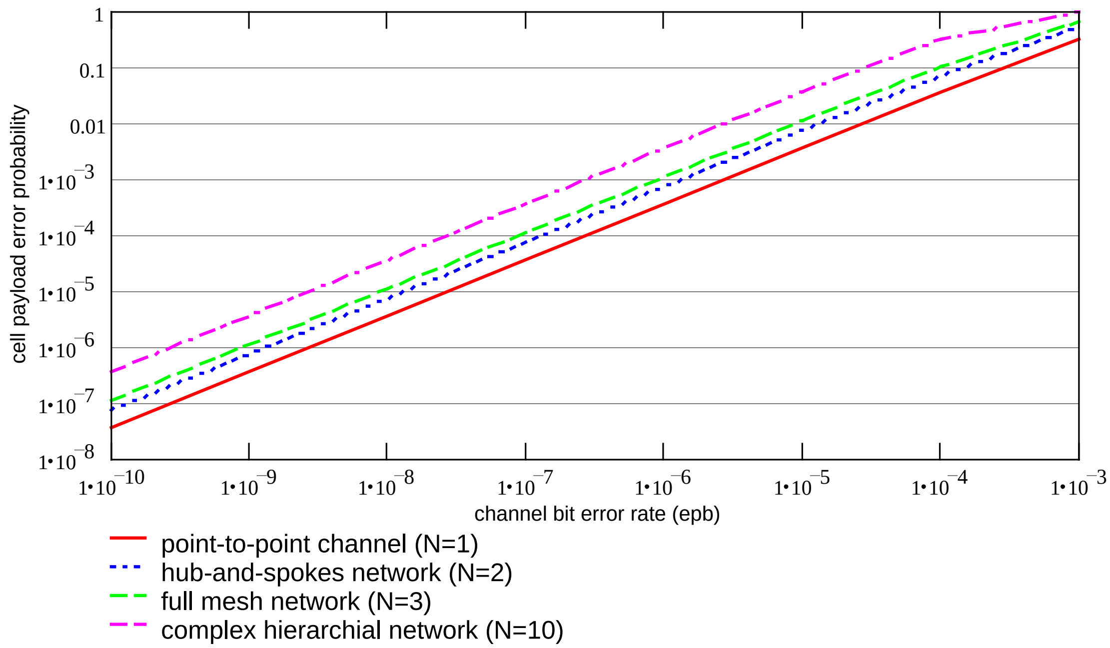


Figure 19. Aberrant Payload Probability for N tandem channels.

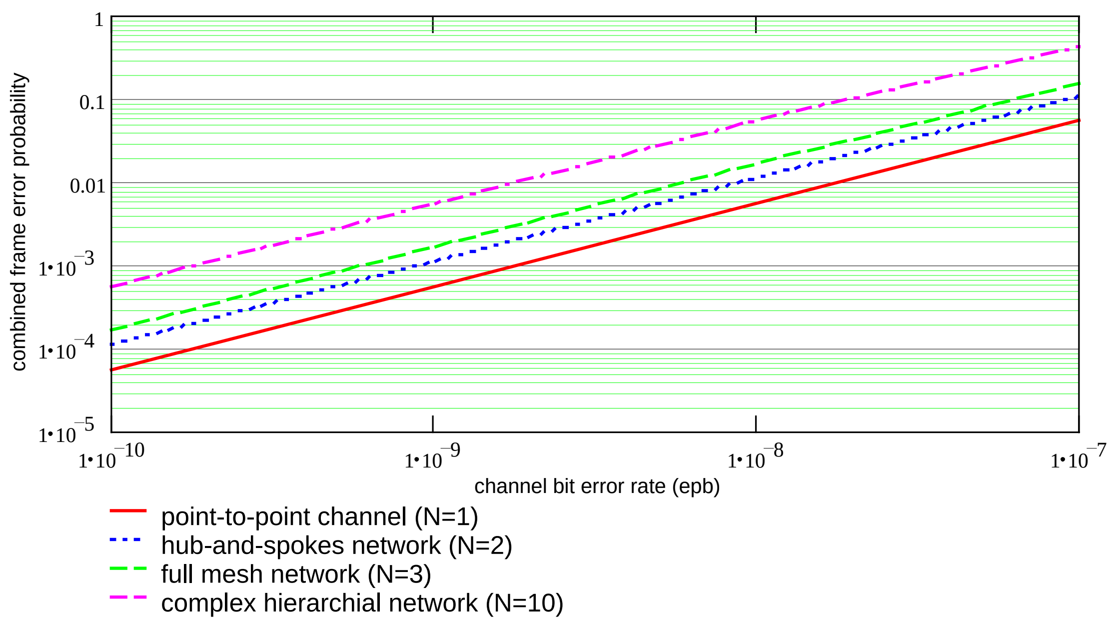


Figure 20. Combined Frame Error Probability for N Tandem Channels.

5.2 Geostationary Satellite Channel

Geostationary satellites have long been popular for distributing video over large areas. Satellite channels have a round-trip propagation delay of 512 milliseconds. Consequently, network designers seldom connect satellite channels in tandem. Instead, it is much more common to treat satellite networks as hub-and-spoke systems with video distribution centered at the hub. ARQ error correction protocols are generally unsuitable given the video transmission rate, propagation delay, and ARQ memory buffer requirements.

Satellite channels suffer from noise, rain fading, and severe transmit power constraints. Consequently, uncorrected bit error rates are on the order of 10^{-5} to 10^{-3} epb [16]. Burst errors may occur during heavy rainfall at the transmitter or receiver. Figure 20 shows that a satellite channel will deliver every frame with one or more errors, even if there is but one channel involved in the virtual connection. Clearly, error correction is essential for satellite video networks.

5.3 Wideband Line-of-Sight Radio Channel

Wideband line-of-sight radio channels are popular media for ATM wireless LANs [17]. Wireless LAN topology is typically a hub-and-spoke configuration over, at most, two wireless channels. Propagation delay is negligible. Noise, transmit power limitations, and multi-path fading are common line-of-sight channel impairments [17]. Consequently, the wireless channels exhibit much higher bit error rates than fiber, on the order of 10^{-5} to 10^{-3} epb [18]. The errors do not always occur in an independent fashion [13], [17]. Burst errors from multi-path fading are common problems. Figure 20 assumes only independent Gaussian noise and ignores burst errors. As such, the figure predicts a lower bound of one or more errors per frame over hub-and-spokes wireless network. Here too, error correction is essential.

6.0 ROBUST APPLICATION OF VIDEO ERROR CONCEALMENT

The following sections show that error concealment should treat a virtual connection as a composite channel comprised of bearer channels connected in tandem by one or more ATM switches. This "virtual connection" model encompasses the simplest case of a single point-to-point connection and complex cases comprised of multi-switch virtual connections. It is necessary to view the virtual connection and its components as a whole. The discussion that follows explores appropriate combinations of error compensation for virtual connections over tandem channels.

6.1 Codec Architectures

Transmission of video over a VBR ATM connection involves several reversible steps within the video encoder. The encoder must digitize (quantize) the video waveform coming from a video source (e.g., camera, VCR), compress the resulting video bit stream, and packetize the stream adding the MPEG-2 bit stream control packets. Finally, it must add transport control packets to coordinate stream transfer from the video encoder, across the ATM network interface, and to the decoder. The resulting bit stream passes from the encoder output to the ATM network interface where interface parses the stream into cells, sends the cells across the ATM connection, and delivers the reconstructed stream with errors or cell losses to the video decoder. Application of rigorous error concealment embedded within the codec offers good opportunities to mitigate disruptions introduced by an ATM connection.

The MPEG-2 encoder in Figure 17 consists of two subcomponents. Within the entropy encoder, there is a waveform encoder that transforms the image into DCT coefficients, digitizes the coefficients, and applies lossy compression to the binary result using variable-scale quantization. An entropy encoder applies lossless compression to the waveform encoder's output by mapping repeating bit patterns into variable length code words according to their statistical properties (e.g., Huffman run-length coding). Entropy encoders further compress the waveform by computing differences between consecutive frames and producing the difference as a compressed estimate of the original image's motion, i.e., motion-compensated image prediction coding.

As seen earlier, MPEG-2, Part 2, uses DCT transforms, quantization, motion-compensated prediction coding, and Huffman run-length encoding. Therefore, MPEG-2 combines waveform and entropy encoding within the scope of one video compression standard.

In Part 1, MPEG-2 transport encoding adds codec synchronization words, multiplexing control packets, and other protocols to the video bit stream to maintain encoder-decoder synchronization. However, MPEG-2 does not concern itself with network-related stream transportation. Instead MPEG-2 delegates network issues to the network interface portion of the transport encoder. Transport encoders add control packets comprising a "transport protocol." A transport protocol provides for the real-time delivery of compressed information over a particular type of network (e.g., video transport protocols are H.221 in H.320, H.223 in H.324, and H.225 in H.323). Figure 17 illustrates the relationships and roles of the MPEG-2 codec, transport codec, ATM switches, and bearer channel. Video transport protocols are important avenues for improving the robustness of NS/EP traffic. This point will become clearer as later sections discuss error concealment methods, many of which apply to the video transport codec.

6.2 Error Effects on Visual Quality

Bit errors are more serious in variable length coding (VLC) (e.g., Huffman run-length coding) than in fixed length compression coding (FLC). Both compression methods may loose video information (referred to as an information "erasure") when there are packet losses, burst errors, or system failures. Fortunately, these types of errors occur relatively infrequently in many networks. On the other hand, the most common type of errors, random bit errors, cause erasures in VLC compression because a bit error causes subsequent bits in the compressed video bit stream to be undecodable and for all purposes lost.

Compressed digital video is a highly interdependent information set. Codecs perform intra- and inter-frame compression to achieve very high compression ratios on the order of 50:1. Compressed temporal information spans two or more frames and is highly dependent upon the accuracy of spatial information. Compressed video information describes many pixels along the horizontal, vertical, and time axes.

An errant bit in uncompressed video information affects only a single pixel in the display area. Compression creates a close relationship between spatial and temporal information. The relationship exacerbates the effects of even one bit error across a much larger pixel area.

Errors have a multiplicative effect on picture quality. Bit errors propagate across multiple video raster lines distorting the picture. For example, consider the effect of bit errors as shown in Figure 21. In the figure, bit errors create distortions spanning many nearby pixels. Cell losses represent the absence (erasure) of 384 consecutive bits of compressed information and have even greater impact on video quality. A single cell loss propagates image distortion over as many as eight raster lines [19]. In the worst case, high incidents of bit errors and lost cells may force the video codec out of macroblock synchronization. The resulting distortion spans an area of 16 raster lines high by 352 pixels wide, i.e., 5632 pixels. Thus, picture distortion grows worse when high bit error rates corrupt more pixels in the cell payload and when congestion or errors cause compressed video cells to be discarded.

(a) Original image

(b) Image with errors

Figure 21. Multiplication of Image Errors [20].

6.3 Consideration for Virtual Connection Characteristics

The characteristics of a video distribution network play an important role in the quality of received video information. Delay, error rates, error distribution, and switch congestion affect the end-to-end quality of the video image as it passes through an ATM network. The goal of error concealment is to mitigate network impairments and raise the perceived quality of the video presentation.

6.3.1 Connection Delay

Transmission delay is an important consideration in any error concealment scheme. ATM virtual connections are subject to bearer channel propagation and switching delay. Some error concealment methods also add delay. Traditional ARQ error correction protocols require large memory buffers for high transmission speeds and significant propagation delays. This is especially true for virtual connections involving several ATM switches or satellite bearer services. Generally speaking, propagation delays, ARQ buffer memory requirements, and the time-definite nature of video make ARQ protocols impractical for many video applications [13], [17]. In some cases, judicious use of ARQ protocols improves video quality especially where latency is acceptable (e.g., video streaming), but misuse can lead to varying video quality, lost codec synchronization, and added network congestion. For these reasons, ARQ-based error concealment is not appropriate for most NS/EP traffic. Instead, NS/EP-relevant error concealment standards should incorporate some or all of the mechanisms described in the following sections.

Standard HDTV video decoders must receive compressed video information at a rate sufficient to maintain the display frame rate (e.g., thirty frames per second). A fixed network propagation delay is acceptable in most viewing situations, (high-fidelity video conferencing being a notable exception). Video codec malfunctions develop when delay variations starve the codec of compressed information.

Therefore, virtual connections must maintain a predictable propagation delay. ATM connections must deliver compressed information to the video decoder at the rate matching the average coding rate. Error correction mechanisms may contribute to fixed delays within the acceptable limits of the applications. Delay variation cannot exceed the video decoder’s ability to buffer and decompress frames at a constant display rate. Optimally, the selected video error correction method introduces an acceptably small, highly predictable processing delay.

6.3.2 Bearer Service Errors

Applications of digital video over ATM range across fiber optic, satellite, and wireless bearer channels with hybrid combinations likely. All of the bearer channels suffer from additive white Gaussian noise in varying degrees. Error rates range from 10^{-9} for fiber channels to 10^{-3} for satellite and wireless channels as discussed earlier. Figure 19 shows that connecting ATM bearer services in tandem has a multiplicative effect on payload error rate.

Satellite and wireless bearer services suffer from burst errors as an additional impairment. Burst errors occur during periods of heavy rainfall, solar, or multi-path interference. The errors are closely grouped and have more serious effects than the equivalent number of random errors. Channel error correction must correct random and burst errors. Correction "strength" must be sufficient to offset tandem connection error effects and maintain good end-to-end video quality.

6.3.3 ATM Switch Behavior

Switch congestion will force an ATM network into flow control, cell loss, and, consequently, leads to video distortion. When ATM invokes flow control, the ATM switches effectively reduce virtual connection bandwidth. If designed to do so, variable-rate codecs will respond to bandwidth reduction by reducing the video coding rate and, consequently, lower the video quality. In severe cases, flow control should force the video codec to discard cells at the source. Reducing nonessential information at its source is an important feature for managing network congestion and should be implemented with preferences shown toward NS/EP traffic. Specifically, ATM flow controls and video encoding methods should reduce bandwidth and discard nonessential information, thereby increasing the network capability to deliver NS/EP traffic.

Congestion may persist despite the ATM's efforts to constrain traffic. If so, ATM will discard low priority cells in an effort to shed excess traffic. If flow control and traffic priorities do not suffice, then ATM switches will loose cells as switch buffers overflow.

Clearly, ATM will discard cells with corrupted cell headers or when congestion threatens the switch's ability to effectively distribute cells. Losses may involve a single cell or a burst of cells. Either way, cell losses are large erasures of highly interdependent video information.

6.4 Video Error Concealment

A good video codec design formulates a combination of waveform, entropy, and transport codecs that maximizes perceived video quality for a given video source model, channel bandwidth, and connection error characteristic. Codec design is a difficult problem for the following reasons:

Video sources have time-varying statistics (nonstationary) and, therefore, are hard to model across a broad range of sources.

ATM connections have time-varying statistics during periods of congestion and present another difficult modeling problem.

Satellite and other wireless channels have time-varying characteristics related to weather, geography, and, in some cases, motion.

There are three types of error concealment codecs [15]:

Forward error concealment uses source encoding provided for error correction at the decoder.

Error concealment by postprocessing attempts to recover lost information at the decoder by estimation and interpolation of the remaining video information.

Interactive error concealment relies upon cooperation between the encoder and decoder to perform ARQ and selective predictive coding based on feedback from the decoder.

Criteria for judging their relative worth are image quality, required delay, bit rate overhead added by redundancy, and processing complexity. The relative importance of each criterion depends upon the video application.

In the final analysis, there may only be limited opportunities to improve the quality of ATM bearer circuits or reduce lossy behavior within the switches, especially during times of national crisis or natural disaster. However, much can be done at the video codec to conceal the network errors and reduce network congestion. The following sections describe a set of video codec features that are highly desirable for good video quality over error-prone, ATM networks.

6.4.1 Congestion Avoidance

Consider a video codec designed with an intermittently congestive, error prone network in mind. First, the video codec would interact with the network's flow control mechanisms to regulate the video-coding rate in keeping with the available network bandwidth. For example, the video encoder should discard nonessential enhancement layers at the source. Discarding enhancement layers at the source has a graceful effect of momentarily reducing video resolution at the receiver. The lowering of video resolution would likely go unnoticed especially if it were brief. Now, compare the encoder's proactive approach to that of delivering the cells to a congested network for subsequent discard. In the later case, ATM networks randomly discard cells in a burst. Unpredictable discarding of video information makes the video decoder's job of maintaining codec synchronization and delivering good quality video much more difficult.

Therefore, steps taken to mitigate congestion at the source are far more effective than remedial steps taken within the network and receiver. NCS should encourage standards and collaboration that yield interaction between the ATM network interface and the video codec for the purpose of regulating video traffic at the encoder. The interactions should favor NS/EP traffic by preferentially assigning bandwidth to NS/EP-designated video encoders.

6.4.2 Bearer Channel Error Correction

High-quality digital video depends on low error rates on the order of 10^{-9} errors per bit. These rates are commonly found on individual SONET or SDH channels. Using SONET as the standard, link-layer error compensation should improve ATM bearer channel performance to 10^{-9} errors per bit as a minimum standard for reliable ATM switching [21].

ATM video distribution networks will employ multiple bearer services and switches to distribute video programs within a city, a nation, or several nations. The number of switches and types of channels will vary according to the geography and economic considerations. It is easy to envision scenarios where the combined effects of several tandem channels will lower cell payload quality to unacceptable levels. Therefore, video error concealment is an appropriate measure, especially when some or all of the ATM bearer channels are prone to error.

6.4.3 Convolutional Encoding

Modems commonly use forward error correction consisting of convolutional or Viterbi encoders [13]. The amount or "strength" of the error correction is variable according to the bearer channel performance goals and the uncorrected channel quality. To illustrate, consider Figure 22. A convolutional codec with two-error correction ability is sufficient to bring a 10^{-3} BER connection up to acceptable quality standards. Such a convolutional codec lends itself to very large scale integration (VLSI) realization as described in [22]. It is, therefore, practical to implement convolutional codecs compatible with video bearer service behavior and speed.

6.4.4 Block Codes

Block encoding is similar to convolutional coding in that it adds information to the bit stream for the purpose of error concealment. Block codes can deal with larger, fixed length blocks of information and can be tailored for a broader range of correction power. As its name suggests, block encoders organize a block of information consisting of K bits into a codeword N bits long. The extra bits add parity information to the block used by the transport decoder for correcting some number of errors and detecting a larger number of errors. Block encoders frequently use Hamming, Reed-Solomon (R-S), or Bose-Chadhuri-Hocquenghem (BCH) encoding algorithms.

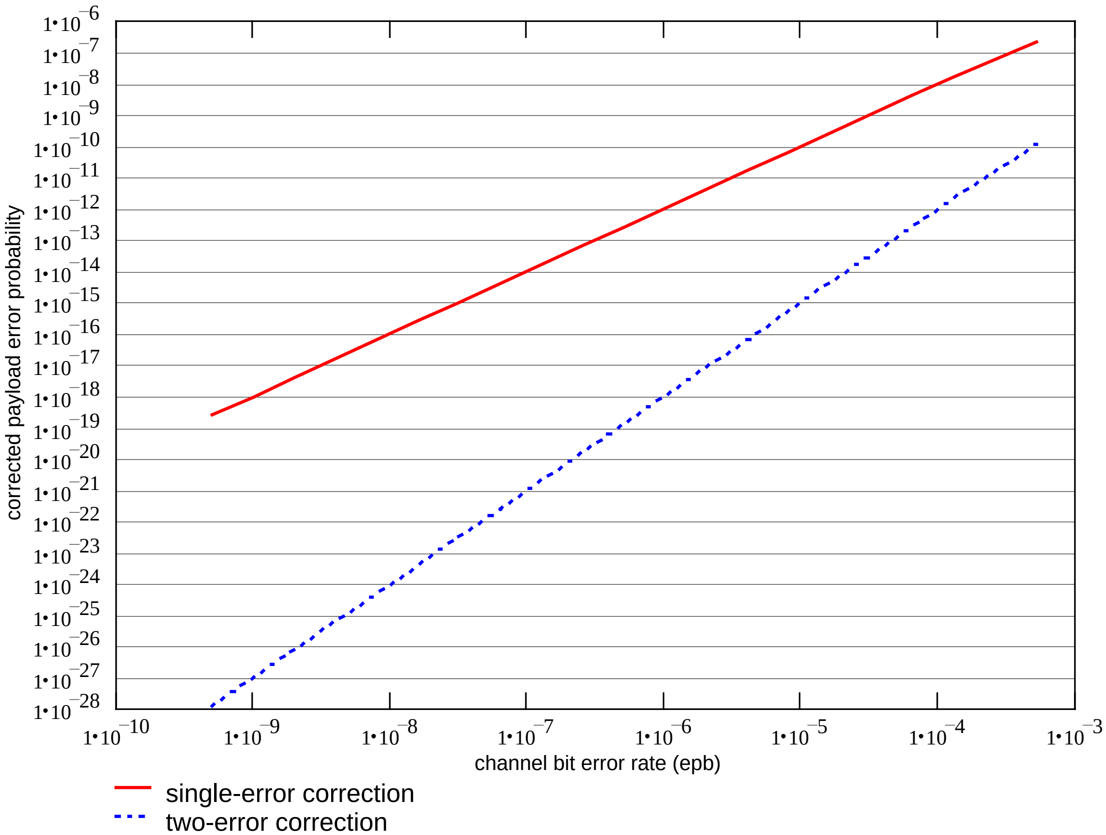


Figure 22. Convolutional Correction of Cell Bit Errors.

6.4.5 Concatenated Codes

There is a downside to convolutional coding. A decoder produces a burst of errors at its output when the number of correctable errors is exceeded at its input 6.0, [16], [23]. In so doing, the convolutional codec alters the behavior of an additive Gaussian noise channel creating a burst error behavior in its place. Convolutional burst errors add to burst error behavior exhibited by satellite or wireless channels. Burst errors can be highly disruptive causing cell discards and noticeable distortions in the displayed image.

A common solution to burst errors is the creation of a concatenated code. A concatenated code combines a block correction encoder (a Reed-Solomon encoder in the case of Figure 17), code word interleaver, and a convolutional encoder operating in tandem [13]. An R-S correction encoder subdivides the compressed video data into blocks comprised of multiple bits. The R-S encoder adds parity bits to form codewords [24], [25], [26], [27]. Each codeword is J bits long consisting of video bits and parity bits. An interleaver organizes L codewords into a matrix containing L rows of J codeword bits. The interleaver transposes the matrix passing the result in row order to a convolutional encoder.

The doubly encoded and interleaved bit stream travels over the ATM bearer channel and arrives at the input of the receiver’s convolutional decoder. The convolutional decoder corrects some number of bit errors within the limits of its design. Having corrected as many errors as possible, the convolutional decoder passes the corrected bits and possibly some burst errors to a deinterleaver. The deinterleaver transposes the K by L matrix returning the RS codewords to their original order. Note that the interleaver-deinterleaver pair does not detect or correct errors. By transposing the matrix, the interleaver spreads or "randomizes" any burst errors across the entire matrix making error concealment easier at the decoder. The matrix passes to a block decoder where some residual errors are corrected and a greater number are detected. A concatenated error concealment strategy chooses an interleaving length sufficient to mitigate the majority of burst errors. Therefore, knowledge of the channel and convolutional decoder behavior is important for optimum concatenated code performance. The result can be a highly robust encoding method that is resilient to random and burst errors introduced by an ATM bearer channel.

NCS should encourage development of ATM QoS plans by network carriers that ensure an acceptable error rate across bearer channels and virtual connections. ATM carriers should apply these plans to their networks in much the same manner as long distance carriers apply "via-net loss" quality plans across the telephone networks. Guided by these QoS plans, the carriers should design each bearer channel to provide statistically high quality across a broad range of virtual connections. The designs would apply error concealment methods at the bearer channel level in a well-planned and consistently executed manner. Published standard operating practices and even FCC mandated standards for carrier service quality seem appropriate for public networks carrying teleconferencing, public broadcast, and subscription video streams.

Interoperability between bearer channel codecs is important for maintaining open and equal access between competing carriers especially if national security or natural disaster call for rapid reconfiguration of a public ATM network. Therefore, standards for bear channel codec operation are an important part of implementing an ATM network QoS plan.

Concatenated coding is a promising way to mitigate random bit errors introduced by an ATM bearer channel. Improving the error performance of the ATM connection reduces the multiplicative effects of tandem channel connections. Concatenated codes can greatly reduce the burst errors from convolutional decoders, channel fading, and multi-path fading. Connection error correction does much to improve the cell transmission quality. Notwithstanding, there remains the problems of cell losses brought about by ATM switch congestion.

6.5 Cell Loss Correction

Cell losses, whether due to intentional traffic constraint, buffer overflow, or uncorrectable header errors, constitute a substantial loss of compressed video information. As shown earlier, random loss of even one cell can affect a large display area. Cell loss due to switch congestion may be more damaging than random cell loss.

Episodes of switch congestion may occur randomly, but the resulting cell loss may be prolonged, spanning many cells. Consequently, congestion often leads to the loss of multiple cells, or "burst cell loss," an effect analogous to burst bit errors. Loss of multiple cells can be highly disruptive to the display. Large cell losses may force the video decoder out of synchronization causing massive display disruption. Congestive loss suggests the need for error concealment encompassing the ATM virtual connection [28].

Consider applying error concealment to a block of cells within the video transport encoder. Using Figure 17 as a guide, group a set of cells together into a "cell block." Add a sequence number to each cell identifying its place in the block. Apply a block encoder to the entire cellblock, thereby, producing codeword having a length of J cells and consisting of video cells plus parity cells [2]. These parity cells hold information sufficient to completely regenerate one or more missing cells and the ability to detect the erasure of some larger number of cells. Organize the cell codewords into a J by M matrix where every row is one codeword. Transpose the codeword matrix using a cell interleaver and send the M by J matrix in row order through the ATM interface and across the virtual connection [18].

At the video transport decoder, a deinterleaver transposes the cell matrix back to its J by M order of the cells and identifies missing cells using the cell sequence. A block decoder treats missing cells as cell erasures and regenerates the missing cells using the parity cells created at the encoder. Locations of unrecoverable cells pass to the video compression decoder for reasons to be explained in a moment.

6.6 Video Interpolation

Often, considerations for error compensation stop with error correction. This is certainly true for the transmission of essential data elements, such as those used in data processing. After all, one cannot gracefully lose elements of a financial record. On the other hand, the human eye is tolerant to low levels of visual distortion, especially if they are short lived. This visual latitude offers an opportunity to further mitigate the effects of virtual connection errors.

These opportunities are especially important when available ATM services cannot carry sporadically heavy volumes of video traffic or when impairments momentarily degrade channel performance. Problems of congestive traffic loads and low transmission quality may plague NS/EP traffic. Under these circumstances, it may be better to gracefully degrade picture quality using our knowledge of the human eye and error location information supplied by the video transport decoder. A prioritized allocation of network capacity would further public interests by degrading nonessential video traffic in favor of preserving NS/EP traffic.

6.6.1 Error Concealment

Video decoders can have the ability to "conceal" errors. That is to say, decoders can be made to reduce the noticeable effects of received errors. Given the location of an errant or missing cell, a video decoder can interpolate information from prior frames and create an estimate of the original information. Interpolation makes modest improvements in the peak signal-to-noise ratio (PSNR), an objective measure of video impairment. However, the result is more pleasing to the human eye than the uninterpolated display [29].

Error concealment in video often begins by partitioning the video information into layers. The base layer contains essential video information for acceptable video quality. Higher layers add enhancements to the base level video quality. Layered error concealment assigns the greatest error protection to the base layer and lesser protection to the higher layers. Layered informational encoding can be by spatial resolution refinement, temporal resolution refinement, amplitude resolution refinement, or frequency domain partitioning [15]. MPEG-2 terminology calls these four methods "temporal scalability," "spatial scalability," "signal-to-noise (SNR) scalability," and "data partitioning," respectively. MPEG-2 deals with scalability as a means of varying coding rate and maintaining interoperability across a broad range of receivers.

In Figure 17, a block decoder provides a video compression decoder with the locations of missing cells. A post-processing video decoder uses the cell locations and information from prior frames and adjoining cells to estimate the information in the missing cells. Thus, the decoder has the ability to conceal missing information. The degree of improvement is largely related to the accuracy of the interpolation and the eye's sensitivity to the residual distortion.

Keeping the eye's sensitivities in mind, consider the benefits of only losing cells of low importance to subjective visual quality. Such is the scheme when a multi-layer video codec, Reed-Solomon cell correction codec and ATM's cell priority mechanism work in unison to deal with network congestion [17]. In this scheme, a multi-layer video codec divides the spatial information into DCT sub-bands as described earlier. Video compression occurs independently at each layer. Low-frequency DCT coefficients are especially important to subjective visual quality and receive the highest cell priority. Higher frequency coefficients receive lower cell priorities. The video transport encoder and ATM network discard low priority cells during periods of network congestion, thereby improving chances that essential base-band information arrives at the video decoder. A Reed-Solomon cell decoder restores missing cells up to the limits of its correction ability and passes uncorrectable cell locations to the video compression decoder for error concealment. Decoder error concealment interpolates available information masking the missing, predominantly high-frequency information. The result is subjectively better because the eye is less sensitive to high-frequency information distortion.

Thus, MPEG-2 scalability layers video information and, in so doing, adds opportunities for improved error concealment. Spatial scalability provides the greatest error robustness and has the greatest overhead. Data partitioning provides the least error robustness but exacts less overhead. SNR scalability falls between the two alternatives for both robustness and overhead. It is important to note again that ATSC standards do not support MPEG-2 scalability options. Therefore, ATSC is not the best standard for NS/EP traffic distribution.

From a NS/EP standpoint, MPEG-2 encoding with two or more video encoding layers is a better standard for sending video information across the ATM network. The video transport encoder and ATM network could cooperatively drop nonessential image enhancement cells during moments of congestion or errant transmission without seriously impacting video quality. Moreover, network management systems could first apply flow control and cell discard policies to non-NS/EP video traffic and then to NS/EP traffic. Error concealment mechanisms within both traffic classes would use error concealment to mask the worst effects of the cell loss. During normal, non-crisis operations, the same prioritized discard and concealment mechanisms would elevate the overall quality and consistency of digital video transmission.

7.0 CONCLUSIONS

This paper began by describing video "quality" as visual perception of a projected image but, through its development, expanded that definition to include the management of video information in times of normal operation and in crisis. Managing video quality and its delivery, especially during NS/EP events, requires a coordinated effort between network design, video codec design, and network operating practice. Government and industry should work together along all three avenues to create video communication systems that provide good video quality on a priority basis according to the public interests.

Four sets of standards impact the ability of ATM networks to deliver quality video and provide for NS/EP preferences: MPEG-2 video communication standards, ITU-T transport protocol standards, ATM standards, and network carrier standard operating practices. MPEG-2 has video encoding features that provide for the segregation of video traffic, variable traffic coding rate, managed connection establishment, and conditional information access. They constitute the basic elements for managing network traffic and reducing congestion at its entry point, i.e., the video encoder. Additions of error concealment to the video decoder will further improve quality by reducing the perceived impact of residual errors and cell losses.

ITU-T transport protocol and ATM standards hold the key for managing the ATM interface and the proper handling of traffic within the network. The transport protocol and ATM interface should share flow control, admission control, and traffic carriage priority information. Clearly, this information relates to the MPEG-2 traffic encoding features as well. Therefore, the transport protocol must share this information with the video encoder. These standards should promote sharing of available bandwidth, traffic classification, priority, and flow control information from the MPEG-2 video encoder to the transport codec and the ATM interface and vice versa.

There is a need for an overarching standard or set of standards that define the relationship between MPEG-2 traffic generation, traffic transport, and ATM switching. Creation of this overarching standard would highlight deficiencies within the underlying standards, i.e., MPEG-2, ITU-T, and ATM, thereby leading to their eventual enhancement. In so doing, traffic management for commercial and NS/EP-related traffic would improve with an overall benefit to video quality.

Carriage of video traffic within the United States will be complicated by the presence of the ATSC standard and its departures from MPEG-2. The creators of ATSC dealt primarily with concerns regarding open-air and conventional CATV broadcast over dedicated channels. Switched networks require greater flexibility and sophistication within the video codec for good network performance than ATSC currently offers. Unfortunately, ATSC carries its designation as the North American standard for high definition TV, which will be delivered over many different types of networks including ATM. There will be difficulties, incompatibilities, and compromises when MPEG-2 and ATSC codecs attempt to communicate.

ATSC lacks important traffic layering features found in MPEG-2. Consequently, ATSC does not offer good opportunities for controlling traffic at the source. There should be modifications to ATSC supporting upward compatibility into MPEG-2 codecs. Receiver manufacturers will be building systems to both standards and should be encouraged to combine both standards within a common system. These and other steps should be taken to avoid incompatibilities between MPEG-2 and ATSC networks.

Finally, there will be a need for network quality standards and standard network operating procedures within the network operators. As shown earlier, the number of switches and the quality of the bearer channels have significant impact upon video quality. Therefore, network organization has significant implications to congestion management. Network quality standards would define the expected behavior of these networks. Quality standards would lead to uniform network design and operating practice in much the same way as AT&T's "Notes on Direct Distance Dialing" and Bell Telephone's technical publications defined high-quality practices within the telephone network years ago.

8.0 LIST OF REFERENCES

[1] B. Bhatt, D. Birks, D. Hermreck, "Digital Television: Making It Work," *IEEE Spectrum*, October 1997.

[2] Dave Lindbergh, "The H.324 Multimedia Communication Standard," *IEEE Communications Magazine*, December 1996.

[3] Rob Koenen, "Overview of the MPEG-4 Version 1 Standard," International Organization for Standards, MPEG98, February 1998, San Jose, CA, ISO/IEC JTC/SC29/WG11 N2078, http://drogo.cselt.stet.it/mpeg/standards/mpeg-4.htm, 3/10/98.

[4] Thomas Sikora, "MPEG Digital Video-Coding Standards," *IEEE Signal Processing Magazine*, September 1997.

[5] Roberto Grandi and Rene' Prince, "Course of Visual Communication: MPEG 4 Compared to Other Standards," Swiss Federal Institute of Technology Zurich, http://www.stud.ee.ethz.ch/~rrggrandi/cover.html, last modified 03 July 1996.

[6] Leonardo Chiariglione, "MPEG and Multimedia Communications," http://www. cselt.stet.it/ufv/leonardo/paper/isce96.htm.

[7] Thomas Sikora, "MPEG-1 and MPEG-2 Digital Video Coding Standard," Heinrich-Hertz-Insitut Berlin - Image Processing Department, <sikora@hhi.de> http://wwwam. hhi.de/mpeg-video/papers/sikora/mpeg1_2/mpeg1_2.htm, 3/10/98.

[8] "Guide to the Use of the ATSC Digital Television Standard," Advanced Television Standards Committee, Doc. A/54, 12 Apr. 95, revised 4 Oct. 95.

[9] P. Nasiopoulos and R. Ward, "Improving the HDTV Picture Performance Under Noisy Transmission Conditions," *Image and Multidimensional Signal Processing*, Volume V of V, 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing, April 27-30, 1993, Minneapolis, MN.

[10] "Superb Surround Sound Just Like at the Movies NTSC," TOSHIBA Internet Community Plaza, http://eiplaza.toshiba.co.jp/dvd/e/wats/0001/35.htm, 3/26/98.

[11] C. Basile, A. P. Cavallerano, M. S. Deiss, R. Keeler, J. S. Lim, W. C. Luplow, W. H. Paik, E. Petajan, R. Rast, G. Reitmeier, T. R. Smith, and C. Todd, "The US HDTV Standard - The Grand Alliance," *IEEE Spectrum*, April 1995.

[12] B. Bhatt, D. Birks, and D. Hermreck, "Digital Television: Making It Work," IEEE Spectrum, October 1997.

[13] J. B. Cain and D. N. McGregor,"A Recommended Error Control Architecture for ATM Networks with Wireless Links," *IEEE Journal on Selected Areas in Communications*, Vol. 15, No. 1, January 1997.

[14] Wolfgang Fischer, Eugen Wallmeier, Thomas Worster, Simon P. Davis, and Andrew Hayter, "Data Communications Using ATM: Architectures, Protocols,

and Resource Management," *IEEE Communications Magazine*, pp. 24-33, August 1994.

[15] Yao Wang and Qin-Fan Zhu, "Error Control and Concealment for Video Communication: A Review," *Proceedings of the IEEE*, Vol. 86, No. 5, May 1998.

[16] James W. Marshall and Anh Q. Le, "Approaches and Requirements for Tactical SATCOM ATM Integration," *MILCOM '95 Conference Record*, Volume 1 of 3, IEEE Communications Society, 6 November 1995, ISBN 0-7803-2489-7/95.

[17] E. Ayanoglu, P. Pancha, A. R. Reibman and S. Talwar, "Forward Error Control for MPEG-2 Video Transport in a Wireless ATM LAN," *IEEE Proceedings of the International Conference n Image Processing*, September 16-19, 1996, Lausanne, Switzerland.

[18] George Thomas, "Impact of Cell Recovery Coding on MPEG Video Over Wireless ATM," IEEE International Symposium on Personal, Indoor, and Mobile Radio, Taipei, Taiwan, ROC, October 15-18, 1996.

[19] W. Verbiest, L Pinnoo, and B. Voeten, "The Impact of the ATM Concept on Video Coding," *IEEE Journal On Selected Areas In Communications*, Vol. 6, No. 9, December 1988.

[20] P. Nasiopoulos and R. Ward, "Improving the HDTV Picture Performance Under Noisy Transmission Conditions," *Image and Multidimensional Signal Processing*, Volume V of V, 1993 IEEE International Conference on Acoustics, Speech, and Signal Processing, April 27-30, 1993, Minneapolis, MN.

[21] Kent Springer, "A Forward Error Correcting Code for Gigabit Fiber Optic Links," *Proceedings of SPIE (Society of Photo-Optical Instrumentation Engineers)* — The International Society of Optical Engineers, Vol. 1577 High-Speed Fiber Networks and Channels, 4-6 September 1991.

[22] D. Oh, Y. Kim, S. Hwang, "A VLSI Architecture Of The Trellis Decoder Block for the Digital HDTV Grand Alliance System," *IEEE Transactions on Consumer Electronics*, Vol. 42, No. 3, August 1996.

[23] Ron R. Mankarious, "A Full Mesh Asynchronous Transfer Mode (ATM) Satellite Communications Network," *MILCOM '95 Conference Record*, Volume 1 of 3, Sponsored by IEEE Communications Society. 6 November 1995.

[24] Bernard Sklar, *Digital Communications Fundamentals and Applications*, Prentice Hall, 1988.

[25] Charles T. Retter, "Research on Error-Correcting Codes," http://homepage. arl.mil/~retter/coding.html, 2/15/98.

[26] Shaw-Min Lei, "Forward Error Correction Codes for MPEG2 over ATM," *IEEE Transaction on Circuits and Systems for Video Technology*, Vol. 4, No. 2, April 1994.

[27] J. K. Wolf, "A Survey of Coding Theory 1967-1972," *IEEE Transactions Information Theory*, Vol. IT-19, July 1973, pp. 381-389.

[28] Bernd X. Weis, "Impact Of Error-Detecting/Error-Correcting Codes on Reliable Data Transmission Over Noisy Channels in ATM Systems," *IEEE Transactions On Communications*, Vol. 39, No. 4, April 1991.

[29] V. Papdakis, T. Le-Ngoc, W. Lynch, "Combined Error Detection and Error Concealment for MPEG2 Over ATM," 1997 Canadian Conference on Electrical and Computer Engineering, May 25-28, 1997, St. John's, Newfoundland, Canada.

[30] Donald R. Marks, "ATM from A to Z: A Definitive Glossary for Enterprise Network Managers," *Data Communications*, pp. 113-122, December 1994.

[31] R. Kishimoto and K. Irie, "HDTV Transmission System in an ATM-Based Network," *Signal Processing: Image Communication 3 (1991)*, Elsevier Science Publishers B.V., pp. 111-122.

[32] "Superb Surround Sound Just Like at the Movies PAL," TOSHIBA Internet Community Plaza, http://eiplaza.toshiba.co.jp/dvd/e/wats/0001/p35.htm, 3/26/98.