National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file   Votre référence

Our file   Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

Performance Improvement Schemes for the Transport of MPEG Video Streams over ATM Networks

by

Taeseop Han, B.A.Sc.

A thesis submitted to the
School of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

MASTER OF APPLIED SCIENCE

Ottawa-Carleton Institute of Electrical Engineering

Department of Electrical Engineering
Faculty of Engineering
University of Ottawa

August, 1995

©Taeseop Han

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-612-07820-5

Canadä

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

I hereby declare that I am the sole author of this thesis.

I authorize the University of Ottawa to lend this thesis to other institutions or individuals for the purpose of scholarly research.

Taeseop Han

I further authorize the University of Ottawa to reproduce the thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Taeseop Han

to my family .....

Contents

Chapter 4 PRIORITIZATION & DUMMY CELL INSERTION SCHEMES

Chapter 5 SLICE-SIZE CONTROL SCHEME

Chapter 6 CONCLUSIONS AND FUTURE RESEARCH

Abstract

In this thesis, the use of three different schemes are studied to overcome the loss of information when transferring MPEG video sequences through ATM networks. In the first scheme, the prioritization algorithm assigns different priorities to the video stream based on the different types of video frames. In the second scheme, the use of dummy cells to replace lost ATM cells is investigated in the ATM adaptation layer. For this second scheme, different patterns to be used for the dummy cells are investigated. In the third scheme, the data size of slices and the number of slices are controlled based on the different types of video frames and the order in a group of pictures. The experimental results show the effectiveness of these schemes to overcome the loss of information when transferring packetized video streams. Statistics of the signal to noise ratio (SNR) for different cell loss ratios (CLR) in each case are presented. Statistics gathered from the three types of schemes can be applied in several ways for network design and operation.

Acknowledgments

I would like to express my sincere appreciation to my thesis supervisor, Dr. Luis Orozco-Barbosa, for his guidance during my thesis work.

I also appreciate Dr. Georganas in University of Ottawa, Peter Ashton and Rashid in BNR for their help during my research.

I am truly grateful to my forever beloved wife ***Jinmi,*** for her consistent support, coping with all difficulty in a new environment. Without her, this work would not have been possible for me.

I am thankful to TRIO for its financial support.

List of Figures

List of Tables

Acronyms

AAL	ATM Adaptation Layer
AAL-IDU	AAL Interface Data Unit
AAL-PCI	AAL Protocol Control Information
AAL-SDU	AAL-Service Data Unit
ATM	Asynchronous Transfer Mode
ATM-SDU	ATM Service Data Unit
B-ISDN	Broadband ISDN
CAC	Connection Admission Control
CCITT	Consultative Committee on International Telegraphy & Telephony
CD	Compact Disk
CDV	Cell Delay Variation
CER	Cell Error Ratio
CLR	Cell Loss Ratio
CMR	Cell Misinsertion Ratio
CLP	Cell Loss Priority
CRC	Cyclic redundancy check
CS	Convergence sublayer
CS-PDU	CS protocol data unit
CSI	Convergence sublayer indication
DCT	Discrete Cosine Transform
FEC	Forward error correction
GFC	Generic Flow Control

GOP	Group of Pictures
HDTV	High Definition Television
HEC	Header Error Control
ISDN	Integrated Services Digital Network
ISO	International Standards Organization
LAPD	Link Access Protocol on D channel
LSB	Least Significant Bit
MPEG	Moving Picture Expert Group
MSB	Most Significant Bit
MUX	Multiplexer
NNI	Network Node Interface
NPC	Network Parameter Control
NT	Network Termination
PCM	Pulse Code Modulation
PDU	Protocol Data Unit
PTI	Payload Type Identifier
QOS	Quality of Service
SAP	Service Access Point
SAR	Segmentation and Reassembly Sublayer
SAR-PDU	SAR Protocol Data Unit
SAR-SDU	SAR Service Data Unit
SN	Sequence Number
SNP	SN Protection
SNR	Signal to Noise Ratio
STM	Synchronous Transfer Mode
TE	Terminal
UNI	User Network Interface

UPC	Usage Parameter Control
VC	Virtual Channel
VP	Virtual Path
VCC	VC Connection
VCI	VC Identifier
VLC	Variable Length Code
VPC	VP Connection
VPI	VP Identifier

Chapter 1

INTRODUCTION

1.1 Motivation

The demand for increased bandwidth to support new services which integrate diverse media such as video, voice, graphics and text has lead to the introduction of Broadband Integrated Services Digital Network (B-ISDN) [WIB92]. Asynchronous Transfer Mode (ATM), a packet switched approach based on fixed cell sizes, has been proposed as the switching and multiplexing scheme that can provide a unified transport structure for services in B-ISDN [DBA90]. This is mainly due to the high flexibility of the ATM networks in handling multimedia services. Among the variety of services that will be supported by B-ISDN, video is becoming an increasingly important communication medium, in addition to voice and data [GPV89, ACV92, RGI89, MPC93].

MPEG (Moving Picture Expert Group) is presently considered to be the prime coding scheme for video transmission. MPEG can combine streams from the video and audio into a single stream and enables the synchronized decoding at the destination [MCM93, MCP93]. The Experts Group for ATM Video Coding in CCITT SG XV was established to develop unified video-coding standards for ATM environments in the B-ISDN [CDS91]. This group is responsible for the development of a standard video-coding algorithm appropriate for the ATM environment. However, in spite of the many advantages offered by the ATM protocol, one potential problem in ATM networks, caused by the bursty nature of traffic and statistical multiplexing, is cell loss [WIA88, MBC89, YVB91]. When several sources transmit at their peak rates simultaneously, the buffers available at some switches may be inadequate causing overflow. The congestion at these switch buffers and the subsequent dropping of cells due to overflow will most likely be the major component of cell loss in B-ISDN. As well, cell loss occurs in a minor way due to bit errors or erroneously routed cells onto another virtual connection. These can lead to severe degradation in service quality, especially if cells are discarded indiscriminately at these switch buffers. Therefore, cell-loss-compensation is becoming an important issue for packet video [MCC93, TVH93, QCC93]. ATM-cell-loss -compensation schemes must be developed in both video-coding algorithms and ATM adaptation layers.

Figure 1.1.1 shows a reference ATM architecture for video services. The terminal equipment is connected via an ATM access multiplexer (MUX) which contains the ATM Adaptation Layer (AAL). The AAL maps the user protocol data unit (PDU) into the information field of the ATM cells and can perform among other functions forward error correction (FEC), dummy cell insertion and priority mechanisms which enhance the performance of the ATM layer [CBS92]. The performance objective could be defined at three different levels: End user level, B-ISDN level (i.e. adaptation layer), and ATM level. In order to support video services over ATM networks, an understanding of end-to-end performance requirements is necessary.

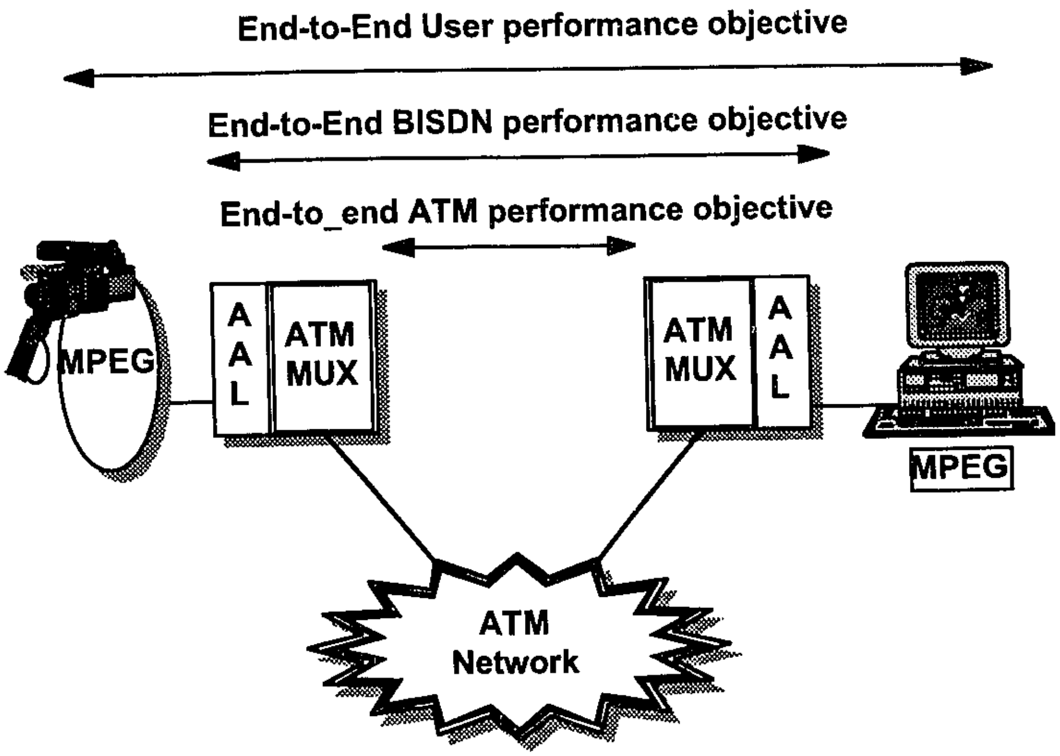


Figure 1.1.1 Reference ATM Architecture

1.2 Objectives

The objective of this study is to improve the performance which will satisfy the end user performance requirement. In this work, we investigate the use of the following three schemes to overcome the loss of cell information based on each layer (ATM, AAL, and Application).

- Prioritization Scheme at the ATM layer,
- Dummy Cell Insertion Scheme at the AAL layer,
- Slice-size Control Scheme at the Application layer.

The prioritization scheme at the ATM layer is accomplished by using the Cell Loss Priority(CLP) bit in a cell header, based on the type of frames in a group of pictures of an MPEG video stream. In ATM networks, a CLP bit is available for the identification of the priority of the ATM cell. Cells with the CLP = 1 may be discarded by the network. On the other hand, cells with the CLP bit = 0 have high priority, and an anticipated capacity has been allocated for them throughout the network. In MPEG video coding, in order to achieve high compression, a number of techniques is used. The algorithm of the temporal processing is, as an example, used to reduce the temporal redundancy so that it causes the different priorities of frames.

The determination of performance requirements of AAL layer can be done assuming that either there are no mitigation methods employed at the AAL, or there are mitigation methods (such as dummy cell insertion in place of lost cells) employed at the AAL. In a dummy cell insertion method, when a cell is determined to be missing from the incoming cell stream, the receiving AAL will insert a dummy cell containing a fixed payload to replace the lost cell. Insertion of dummy cells maintains the bit count integrity and thus helps preventing the impairments from being extended.

At the application layer, the coding characteristics of MPEG video can be exploited to improve performance. In MPEG-based video communication, due to the temporal processing as previously mentioned, the propagated amounts of impairments caused by cell losses are determined by the dependency of pictures in a group of pictures at the decoder. The objective of controlling the slice size is to explore the use of shorter slice size in the frames on which more pictures depend. This reduces the total impairments propagated by the cell loss of the frames. The macroblock relationship of frames in a group of pictures is analyzed to minimize the total impairments.

1.3 Thesis Outline

The thesis is organized as follows. In Chapter 2, a brief overview of the BISDN (Broadband Integrated Services Digital Network) architecture and protocols for integration of a wide variety of services is described. ATM techniques, QOS (Quality of Services) parameters and resource management in ATM networks are then briefly explained. Sources of cell loss, which is an important QOS parameter in this work, are also presented. Then, the basic principles of the ATM adaptation layer (AAL) are discussed and AAL 1 is described in detail.

In Chapter 3, the basic concepts of MPEG (Moving Picture Expert Group) 1 and MPEG2 proposed by the ISO (International Standards Organization) for video compression are briefly described. The MPEG coding algorithm is then presented, and basic terminology is introduced.

In Chapters 4 and 5, three different schemes for improving the performance of MPEG transport over ATM are proposed. The first two schemes, Prioritization and Dummy Cell Insertion, are presented in Chapter 4 and the third scheme, the Slice-size Control, is presented in Chapter 5.

Chapter 4 illustrates the motivation of the first two schemes and the effects on the MPEG video sequence due to the loss of information based on the frame types, and describes the specifics of the proposed schemes. Experimental results are provided including the statistics of SNR (Signal to Noise Ratio) for different CLR (Cell Loss Ratio) and the pictures of the video.

Chapter 5 also illustrates the motivation of the third scheme and the cell loss effect on the video sequence according to the dependency of pictures in GOP (Group of Pictures). The specifics of the slice-size control scheme are described, and the analytical and experimental results are presented.

Chapter 6 concludes the thesis and outlines the future directions of research.

Publications

1. Taeseop Han and Luis Orozco-Barbosa, "Performance Improvement for the Transport of MPEG Video Streams over ATM Networks," *IEEE International Conference on Communications*, Seattle, USA, June 1995

2. Taeseop Han and Luis Orozco-Barbosa, "Error Control Schemes for Improving the Performance of MPEG-based Video Communication over ATM Networks," International Conference on Computer Communication, Seoul, Korea, August 1995.

3. Taeseop Han and Luis Orozco-Barbosa, "Slice to Cell Mapping Mechanism for MPEG Video Communications over ATM Networks," Canadian Conference on Electrical and Computer Engineering, Montreal, Canada, September 1995

Chapter 2
BROADBAND ISDN

2.1 Introduction

The need for B-ISDN (Broadband Integrated Services Digital Network) has been recognized worldwide, and considerable effort is currently being focused on the development of the B-ISDN transfer technology, system architecture and network introduction strategy [WIB92]. In future B-ISDN a wide variety of services with different characteristics will be integrated. Different services have different network performance requirements. Voice traffic can tolerate a limited delay, but can accept a moderate loss ratio. On the other hand, data might be very sensitive to loss, but their delay requirements are not so strict. These performance requirements are negotiated during the call setup phase. The network will be responsible for keeping the values of these performance requirements within the negotiated range. The fundamental objective of future telecommunication networks is to be able to meet the requirements of all the above services.

Broadband ISDN Architecture

B-ISDN differs from narrowband ISDN in a number of ways. To meet the requirement for high-resolution video, an upper channel rate on the order of 150 Mbps is needed. To simultaneously support one or more interactive services and distributive services, a total subscriber line rate of about 600 Mbps is needed. In terms of today's installed telephone plant, this is a stupendous data rate to sustain. The only appropriate technology for widespread support of such data rates is optical fiber. Hence, the introduction of B-ISDN depends on the pace of introduction of fiber subscriber loops.

Internal to the network, there is the issue of the switching technique to be used. The switching facility has to be capable of handling a wide range of different bit rates and traffic parameters. Despite the increasing power of digital circuit-switching hardware and the increasing use of optical fiber trunking, it is difficult to handle the large and diverse requirements of B-ISDN with circuit-switching technology. For this reason, there is increasing interest in some type of fast packet-switching as the basic switching technique for B-ISDN. This form of switching readily supports a new user-network interface protocol known as ATM (Asynchronous Transfer Mode) [WDC94].

Broadband ISDN Protocols

The protocol architecture for B-ISDN introduces some new elements not found in the ISDN architecture, as depicted in Figure 2.1.1. For B-ISDN, it is assumed that the transfer of information across the user-network interface will use asynchronous transfer mode (ATM). One difference between X.25 and ATM is that X.25 includes control signaling on the same channel as data transfer, whereas ATM makes use of common-channel signaling. Another difference is that X.25 packets may be of various length, whereas ATM packets are of fixed size, referred to as cells.

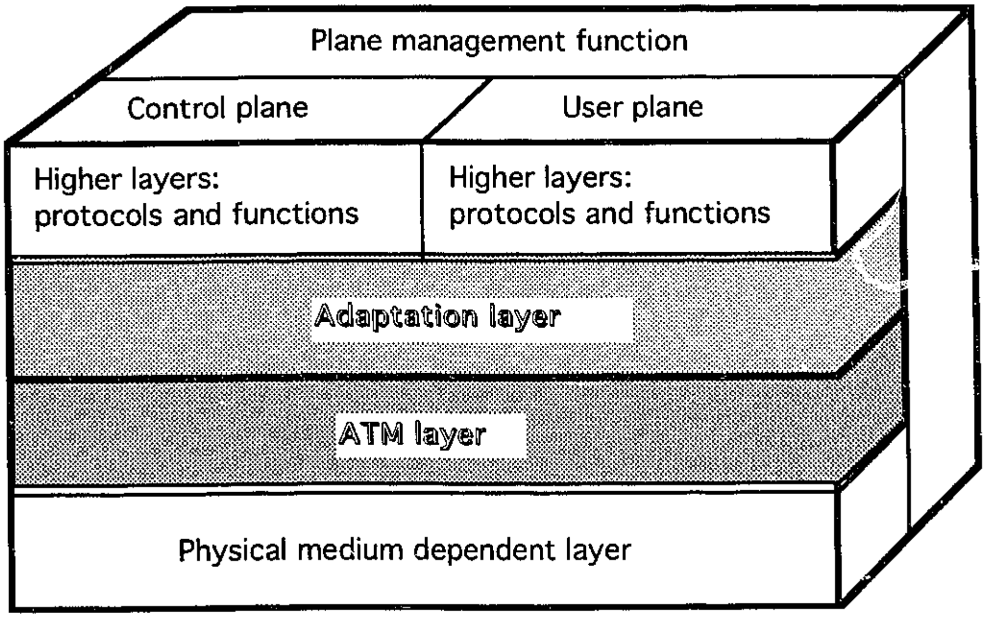


Figure 2.1.1 B-ISDN protocol model for ATM

The decision to use ATM for B-ISDN is a remarkable one. This implies that B-ISDN will be a packet-based network, certainly at the interface and almost certainly in terms of its internal switching. Although the recommendation also states that B-ISDN will support circuit-mode applications, this will be done over a packet-based transport mechanism.

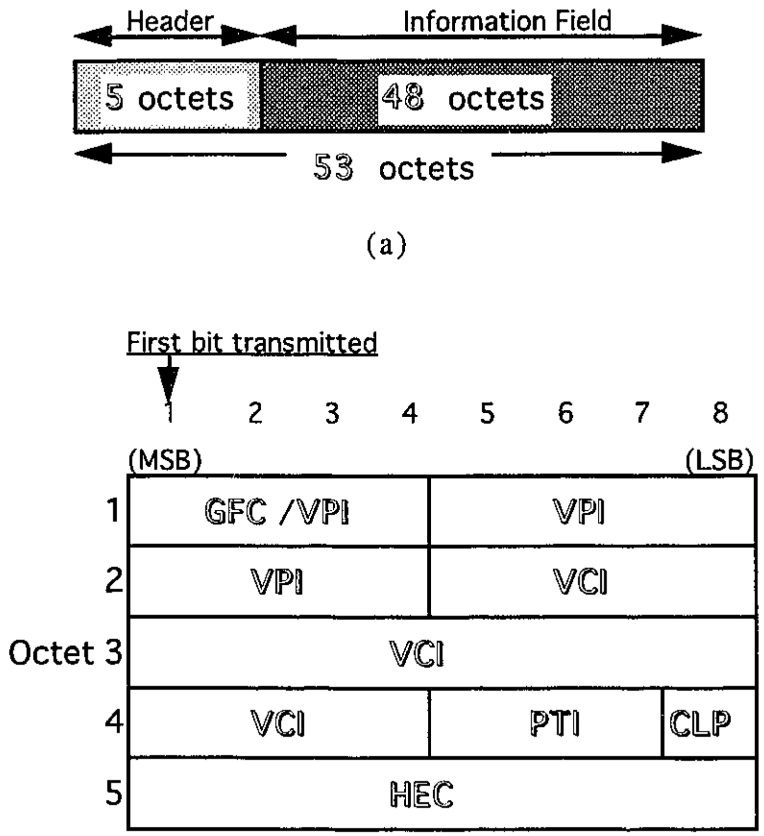
Two layers of the B-ISDN protocol architecture relate to ATM functions. There is an ATM layer common to all services that provides packet transfer capabilities and an ATM adaptation layer (AAL) that is service dependent. The AAL maps higher-layer information into cells to be transported over B-ISDN, then collects information from ATM cells for delivery to higher layers. The use of ATM creates the need for an adaptation layer to support information transfer protocols not based on ATM. Two examples listed in I.121 are PCM (Pulse Code Modulation) voice and LAPD (Link Access Protocol on D channel) [CBI90]. PCM voice is an application that produces a stream of bits. To employ this application over ATM, it is necessary to assemble PCM bits into packets for transmission and to read them out on reception in such a way as to produce a smooth, constant flow of bits to the receiver. For LAPD, it is necessary to map LAPD frames into ATM packets; this will probably mean segmenting one LAPD frame into a number of packets on transmission and reassembling the frame from packets on reception. By allowing the use of LAPD over ATM, all of the existing ISDN applications and control signaling protocols can be used on B-ISDN.

2.2 Overview of ATM Networks

ATM is an attractive packet-oriented switching and multiplexing technique for transfer of future broadband communication services having a broad spectrum of traffic statistics and Quality of Service (QOS) requirements [DBA90]. ATM is expected to offer higher bit rate flexibility and a greater degree of integration, in both transmission and switching functions, than the Synchronous Transfer Mode (STM) [DBA90, CBG90]. In the initial phases of introduction, and for the purposes of interworking, however, ATM must coexist with the STM. Current STM-based network systems are mainly dedicated to supporting specific services such as telephony, circuit or packet switching services for low speed data, and narrowband ISDN services. This has happened because the current communication technologies were designed around specific applications, and are very difficult to adapt to new services. In ATM networks, through statistical multiplexing, several individual sources may share a high transmission rate link of capacity less than the sum of their peak arrival rates. Through statistical bandwidth assignment, a significant multiplexing gain can be achieved, especially for bursty traffic sources such as video telephony, video retrieval and document retrieval. This mode of operation is particularly advantageous when compared with STM, since in STM a physical allocation of network resources (peak rate) to any connection is assumed for its duration. However, to achieve this higher efficiency, ATM requires effective traffic control strategies to guarantee a minimum QOS in all connections [GOQ89, JES90, JAT92, AQS91, CTC92].

2.2.1 ATM Technique

ATM is a packet-oriented switching and multiplexing technique based on the use of cells which are of a fixed length, having a payload field and a header. The technique is expected, at a high transmission rate, to offer full bandwidth flexibility and a higher bandwidth utilization than the STM through statistical multiplexing. The gain in bandwidth utilization, however, will require the provision of a more complicated connection acceptance mechanism. The cells are transmitted contiguously on a transmission link, and are not identified by their position in relation to a fixed time reference but by means of address information in the header defining a 'virtual channel'. The technique is asynchronous in the sense that the cells carrying a particular address (i.e. within a particular virtual channel) may appear at irregular intervals within the cell-stream. The technique is connection oriented in that a 'virtual circuit' is established at call set-up time, and this will associate the virtual channels used on a series of network links to form the end-to-end connection [CBI90]. The ATM cell is the basic unit of information transfer within the ATM network. It consists of a 5-octet header field and a 48-octet information field (see Figure 2.2.1a). The header field which contains mainly routing information (VCI (Virtual Channel Identifier), VPI (Virtual Path Identifier)) is transmitted first, followed by the information field. The structure of the ATM cell header is depicted in Figure 2.2.1b.



(a)

(b)

Figure 2.2.1 a)ATM cell structure b)ATM cell header structure at UNI

The size of the ATM cell header (5 octets), as well as the size of the information field of the ATM cell, remain constant at both the User Network Interface (UNI) and the Network Node Interface (NNI). The only difference of the ATM cell format between the UNI and the NNI is that the bits allocated to the GFC (Generic Flow Control) field of the ATM cell header at the UNI are allocated to extend the VPI field of the ATM cell header at the NNI. Thus, while the VPI field at the UNI consists of 8 bits, at the NNI the same field consists of 12 bits [CBL90]. A description of each field of the ATM cell header at the UNI is given below.

- *Generic Flow Control (GFC)*: this four bit field will be used by the generic flow control mechanism to support traffic flow control at the UNI in the direction towards the network to avoid overload situations; namely, the traffic flow control at the UNI is supported by the GFC field of the ATM cell header.
- *VPI / VCI* : these fields (8 or 12 bits for the VPI and 16 bits for the VCI) contain the routing information of the cell. The VPI and the VCI fields identify, respectively, one particular VP (Virtual Path) within the physical link on which the cell is transmitted, and one specific VC (Virtual Channel) inside the VP.
- *payload Type Identifier (PTI)*: three bits are available for the identification of the payload of the cell. Two different payload types have been considered so far: user information and network information. In user information cells, the payload contains user information and service adaptation function information. Where the payload type field of a cell indicates network information, additional information regarding the type of the network control is given in the information field of the cell.
- *Cell Loss Priority (CLP)*: one bit is available for the identification of the priority of the ATM cell. Cells with the CLP bit set may be discarded by the network. On the other hand, cells with the CLP bit not set have high priority, and an anticipated capacity has been allocated for them throughout the network [CBF90].
- *Header Error Control (HEC)*: this 8 bit field is used for cell header error protection, and provides recovery from single-bit header errors and a low probability of delivery of cells with errored headers.

2.2.2 Quality of Service Parameters in ATM Networks

The QOS of bearer services in an ATM network is characterized by the following parameters:

- *Call control parameters*: connection set-up delay, connection release delay and connection acceptance probability.
- *Information transfer parameters*: cell information field bit error ratio (CER), cell loss ratio (CLR), cell misinsertion ratio (CMR), end-to-end transfer delay and cell delay variation (CDV).

\- *Cell Loss Ratio* : The CLR is the ratio of lost cells to total cells in a population of interest. A cell is considered to be lost when it does not arrive at its intended destination within time T of being sent. The value for T is currently not defined in CCITT Rec. I.35b. However, since ATM is supposed to provide cell sequence integrity, anytime a cell with sequence number k+1 arrives before cell k, then cell k may be assumed to be lost[1]. Note that the CLR parameter includes only compliant cells which have been discarded by the network. Cells which are non-compliant and tagged by the UPC with CLP=1 are not counted in this parameter if they are discarded.

\- *Cell Misinsertion Rate*: The CMR is defined as the total number of misinserted cells observed during a specified time interval divided by the time interval duration. All duplicated and misdelivered cell events are included in this parameter. Note that the most likely cause of a misinserted cell on a particular connection is an undetected

---

[1] In this case, we are assumming the use of an AAL1 entity. At the time of writing, the CCITT has only released a draft standards of AAL 1 which defines a sequence number for every cell associated to a frame.

header error in a cell being transmitted on another connection. Since the mechanism that most often causes misinserted cells has nothing to do with the number of transmitted cells on the observed connection, this performance parameter is most correctly expressed on an end-to-end basis as a rate, rather than a ratio.

- *Cell Transfer Delay*: End-to-End transfer delay can be defined as the value of the elapsed time between the start of transfer and the successful completion of transfer of a cell. According to the CCITT, end-to-end transfer delay is defined as the one-way propagation time between two S/T interfaces. In ATM networks the main factors contributing to end-to-end transfer delay are delay in the transmission media and delay in the queues of the switching nodes and vary according to the instantaneous traffic load condition of the network. The contribution of the switching nodes to the delay is at most 10 ms per switching node.

- *Cell Delay Variation*: There are currently two CDV definitions specified in draft I.35b. Cell clumping measures the variability of cell arrivals at a single measurement point. Only cells which arrive before their expected arrival time count towards the cell clumping parameter. The other CDV parameter is called two-point CDV and measures the cell transfer performance between two measurement points. The cell parameter is used to engineer UPC (Usage Parameter Control) and NPC (Network Parameter Control) algorithms(Figure A.1) whereas the two point CDV parameter is of interest for determining cell transfer performance of a switch or network connection portion. Cell delay variation is caused mainly by variations in cell delay in the network nodes. As a compromise between switching centers and terminals, a maximum peak-to-peak cell delay variation of some 100 microseconds per ATM switching centers can be expected.(For the detail of resource management, refer to Appendix A)

The appropriate values for these parameters are influenced on one side by what is realistically achievable by an ATM network, and on the other side by what is required from particular services. The values of these QOS parameters are negotiated between the user and the network during the call set-up phase. The network will be responsible for keeping these values within the negotiated range. In ATM networks, the most important QOS parameters are those dealing with cell loss [KQM90], cell delay and cell delay variation [AQS91].

2.3 Sources of Cell Loss

Cell loss occurs whenever a cell is discarded due to bit errors, congestion control, or erroneously routed to another virtual connection (or nonexistent virtual connection). Detection of lost and misinserted cells by using cell sequence numbering and discarding misinserted cells are assumed as minimum adaptation layer functions for all services. Protection against cell loss events may be service dependent.

Errors on the Transmission Media

The transmission medium (e.g. fiber) is the lowest level source of errors in an ATM network. Bit errors occur across both the cell header and payload and may occur on each and every segment of fiber a cell traverses. Errors in the cell payload do not contribute directly to cell loss prior to the destination AAL, but header errors can cause problems at every point where an ATM layer receives cells (destination or intermediate switching node).

The cell header is protected by an 8-bit header error check (HEC) field (Fig2.2.1b). The HEC is capable of providing single-bit error correction and multi-bit error detection. A cell receiver operates in two modes: correction and detection. In detection mode, all cells with detectable header errors are discarded; in correction mode, single-bit errors are corrected while cells with multi-bit errors are discarded. Detecting an error while in correction mode causes the receiver to switch to detection mode after processing the current cell; reception of a valid cell while in detection mode causes the receiver to process the cell and switch back to correction mode [CBU92].

Cell loss occurs when a corrupt cell is detected and properly discarded. Cell loss also occurs when header corruption is not detected but has resulted in a new VPI or VCI. The cell then "disappears" from the virtual connection it should have been on. A further cause of cell "disappearance" is when a single-bit error is incorrectly reconstructed, again possibly leading to a new VPI or VCI. (These disappearances can affect another virtual connection, by injecting lost cells into it. This cell insertion has been shown to be much less significant than cell loss in most cases [MCI92])

Discarding Cells for Congestion Control

The broad effect of congestion control is due to the combination of policing units acting close to the cell source, and overload response mechanisms acting at the inputs to cell switching nodes. In general, each virtual connection will have an associated agreed traffic parameter. Policing units act to ensure the compliance of the network user with their agreed traffic parameters. A policing unit may simply discard cells sent in violation of the agreed parameters.

Most switching nodes, by design, discard cells when the switch fabrics and associated buffers are overflowing. The CLP bit in the header is used to indicate to switch nodes which cells they should discard first, if possible. A variation on the discarding policing unit is a "violation tagging" policing unit, where cells in violation of the traffic parameters are allowed through but tagged so that switch nodes know they may be discarded first [CCB91].

Processing Errors in Switching Nodes and End Points

A final area of cell loss occurs when header processing is incorrectly performed at a switching node or virtual connection end point, possibly leading to complete loss or misrouting of cells. Actually symptoms in this area depend on the physical and logical architecture of switching nodes along the path taken by any given virtual connection. A receiving node may incorrectly interpret a cell's VPI or VCI and fail to match it to an open virtual connection, thus losing the cell.

2.4 ATM Adaptation Layer

The ATM adaptation layer (AAL) enhances the services provided by the ATM layer to support the functions required by the next higher layer. The AAL performs functions required by the user, control and management planes and supports the mapping between the ATM layer and the next higher layer. The functions performed in the AAL depending upon the higher layer requirements. The AAL supports multiple protocols to fit the needs of the different AAL service users. The AAL is therefore service-dependent. Examples of services provided by the AAL include: handling of transmission errors; handling quantization effect due to cell information field size; handling of the lost and misinserted cell condition; flow control and timing control [IBF93].

2.4.1 Basic principles of the AAL

The AAL isolates the higher layers from the specific characteristics of the ATM layer by mapping the higher layer protocol data units (PDUs) into the information field of the

ATM cell and vice-versa. The AAL entities exchange information with the peer AAL entities to support the AAL functions.

To support services above the AAL, some interdependent functions must be performed in the AAL. These functions are organized in two logical sublayers, the convergence sublayer (CS) and the segmentation and reassembly sublayer (SAR). The prime functions of SAR are: segmentation of higher layer information into a size suitable for the information field of an ATM cell; reassembly of the contents of ATM cell information fields into higher layer information. The prime function of CS is to provide the AAL service at the AAL-SAP (Service Access Point). This sublayer is service-dependent.

In order to minimize the number of AAL protocols, a service classification is defined based on the following parameters: timing relation between source and destination (required or not required); bit rate (constant or variable); connection mode (connection-oriented or connectionless). Other parameters such as assurance of the communication are treated as quality of service parameters, and therefore do not lead to different service classes for the AAL. Since not all combinations of the above parameters are foreseen, four classes are distinguished, according to Figure 2.4.1.

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

Figure 2.4.1 Service classification for AAL

Examples of services in the classes A, B, C and D are as follows:

- Class A	Circuit emulation; constant bit rate video
- Class B	Variable bit rate video and audio
- Class C	Connection-oriented data transfer
- Class D	Connectionless data transfer

2.4.2 Specification of AAL type 1

The layer services provided by AAL type 1 to the AAL user are: transfer of service data units with a constant source bit rate and the delivery of them with the same bit rate [IBS93]; transfer of timing information between source and destination; transfer of structure information between source and destination; indication of lost or errored information which is not recovered by AAL type 1, if needed.

Function

The following functions may be performed in the AAL type 1 in order to enhance the ATM layer service: segmentation and reassembly of user information; handling of cell delay variation; handling of cell payload assembly delay; handling of lost and misinserted cells; source clock frequency recovery at the receiver; recovery of the source data structure at the receiver; monitoring of AAL-PCI (Protocol Control Information) for bit errors; handling of AAL-PCI bit errors; monitoring of user information field for bit errors and possible corrective action. Other functions are for further study.

Segmentation and Reassembly (SAR) sublayer

The SAR sublayer functions are performed on an ATM-SDU (Service Data Unit) basis.

- *Mapping between CS-PDU and SAR-PDU* : The SAR sublayer at the transmitting end accepts a 47 octet block of data from the convergence sublayer (CS), and then prepends a one octet SAR-PDU header to each block to form the SAR-PDU. The SAR sublayer at the receiving end receives the 48 octet block of data from the ATM layer, and then separates the SAR-PDU header. The 47 octet block of SAR-PDU payload is passed to the CS.
- *Existence of CS function* : The SAR sublayer has the capability to indicate the existence of a CS function. Associated with each 47 octet SAR-PDU payload, it receives this indication from the CS and conveys it to the peer CS entity. The use of this indication is optional.
- *Sequence numbering* : Associated with each SAR-PDU payload, the SAR sublayer receives a sequence number value from the CS. At the receiving end, it passes the sequence number value to the CS. The CS may use these sequence number values to detect lost or misinserted SAR-PDU payloads (corresponding it lost or misinserted ATM cells).
- *Error protection* : The SAR sublayer protects the sequence number value and the CS indication against bit errors. It informs the receiving CS when the sequence number value and the CS indication are errored and cannot be corrected.

In the format of SAR protocol, the SAR-PDU header together with the 47 octets of the SAR-PDU payload comprises the 48 octet ATM-SDU, cell information field. The size and positions of the fields in the SAR-PDU are given in Figure 2.4.2.

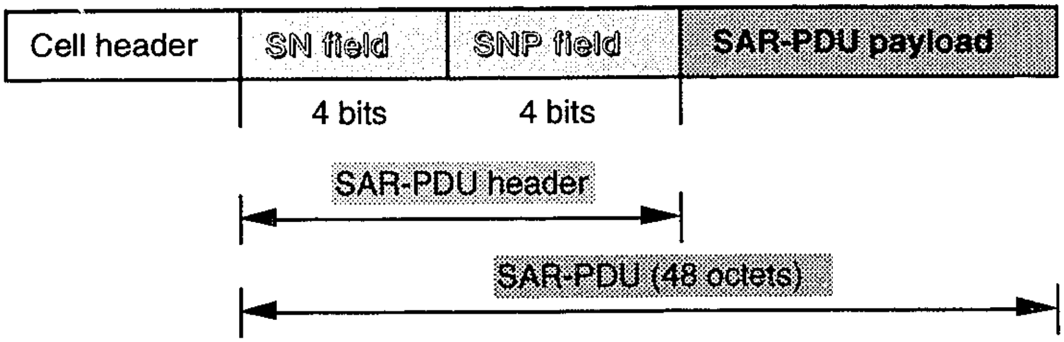


Figure 2.4.2 SAR-PDU format of AAL type 1

The SN (Sequence Number) field is divided into two subfields as shown in Figure 2.4.3. The sequence count field carries the sequence count value provided by the CS. The CSI (Convergence Sublayer Indication) bit carries the CS indication provided by the CS. The default value of the CSI bit is "0".

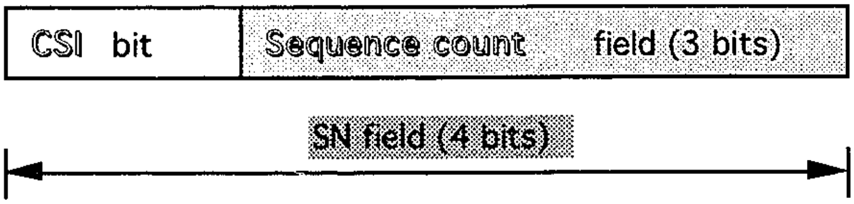


Figure 2.4.3 Sequence number (SN) field format

The SNP (Sequence Number Protection) field provides error detection and correction capabilities over the SAR-PDU header. The format of this field is given in

Figure 2.4.4. A two step approach is used for the protection: the SN field is protected by a 3 bit CRC code; The resulting 7 bit codeword is protected by an even parity check bit.

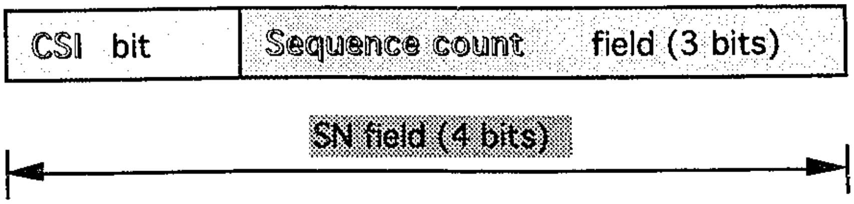


Figure 2.4.4 SNP field format

Convergence sublayer (CS)

The CS may include the following functions.

a) Handling of cell delay variation is performed at this sublayer for delivery of AAL-SDUs to an AAL user at a constant bit rate.

b) Processing of sequence count may be performed at this sublayer. The sequence count value and its error check status provided by the SAR sublayer can be used by the CS to detect cell loss and misinsertion. Further handing of lost and misinserted cells is also performed in this layer.

c) The CS can utilize the CS indication provided by the SAR sublayer to support CS functions for some AAL users.

d) For AAL users requiring recovery of source clock frequency at the destination end, the AAL can provide a mechanism for a timing information transfer.

e) For some AAL users, this sublayer provides the transfer of structure information between source and destination.

f) For video and high quality audio signal transport, forward error correction may be performed to protect against bit errors. This may be combined with interleaving of AAL user bits (e.g. octet interleaving) to give more secure protection against errors.

g) The CS may generate reports giving the status of end-to-end performance as deduced by the AAL. The performance measures in these reports could be based on: events of lost and misinserted cells; buffer underflow and overflow; bit error events.

Chapter 3
MPEG CODING

3.1 Introduction

The MPEG (Moving Picture Expert Group) coding algorithm was developed primarily for storage of compressed video on digital storage media [DMV91]. Provisions were therefore made in the algorithm to enable random access, fast forward/reverse searches and other features when decoding from any digital storage media. However, the coding standard is suitable for a much wider range of video applications. Recent applications of MPEG-like coding algorithms have appeared for a variety of video services from multimedia workstations to high definition television [PMC94].

Recently, the International Standards Organization (ISO) has proposed the MPEG 1 and MPEG 2 standards for video compression. MPEG has been developed by ISO/IEC/JTC1/SC 29/WG11 committee of the ISO. It covers motion video as well as audio coding according to the ISO/IEC standardization process. MPEG-1 specifies a coded representation that can be used for compressing video sequences up to bit rates around 1.5 Mbits/sec. It was developed in response to the growing need for a common format for representing compressed video on various digital storage media such as CD's (Compact Disk's), Winchester disks and optical drives [MCM93]. The use of MPEG-1 means that motion video can be manipulated as a form of computer data and can be transmitted and received over existing and future networks. Further developments in the area of video coding techniques are based on a target rate of up to 100 Mbits/s. This is known as MPEG-2 [MCP93]. It strives for a higher resolution similar to the digital video studio standard CCIR601 and leading to HDTV. The MPEG-2 video standard specifies the coded bit stream for high-quality digital video. As a compatible extension, MPEG-2 video builds upon the completed MPEG-1 standard by supporting interlaced video formats and a number of advanced features including those supporting HDTV (High Definition Television). The MPEG-2 Main Profile was defined to support digital video transmission in the range of about 2 to 80 Mbits/sec over cable, satellite and other broadcast channels as well as for digital storage media and other communication applications[MCP93].

3.2 Overview of the Algorithm

The coded representation defined in MPEG achieves a high compression ratio while preserving a good picture quality. The algorithm is not lossless as the exact pel values are not preserved during coding. The choice of the techniques is based on the need to balance a high picture quality and compression ratio with requirement to make random access to the coded bitstream. Obtaining good picture quality at the bitrates of interest demands very high compression, which is not achievable with intraframe coding alone.

The need for random access, however, is best satisfied with pure intraframe coding. This requires a careful balance between intra- and interframe coding and recursive and non-recursive temporal redundancy reduction.

A number of techniques can be used to achieve high compression. The first, which is almost independent from this International Standard, is to select an appropriate spatial resolution for the signal. The algorithm then uses block-based motion compensation to reduce the temporal redundancy. Motion compensation is used both for causal prediction of the current picture from a previous picture and for non-causal interpolative prediction from past and future pictures [PMC91]. Motion vectors are defined for each 16-pel by 16-line region of the image. The difference signal called the prediction error is further compressed using the discrete transform(DCT) - Important contributions can be found in [DPT90, FFS90, ND91, HFR87] - to remove spatial correlation before it is quantized in an irreversible process that discards the less important information. Finally, the motion vectors are combined with the residual DCT information, and transmitted using variable length codes.

3.2.1 Compression Processing Schemes

Temporal Processing

Because of the conflicting requirements of random access and highly efficient compression, three main picture types have been defined. Intra coded pictures (I-Pictures) are coded without reference to other pictures. They provide access points to the coded sequence where decoding can begin, but are coded with only moderate compression. Predictive coded pictures (P-Pictures) are coded more efficiently using motion compensated prediction from a past intra or predictive coded picture and are generally used as a reference for further prediction. Bidirectionally-predictive coded pictures (B-Pictures) provide the highest degree of compression but require both past and future reference pictures for motion compensation. Bidirectionally-predictive coded pictures are never used as references for prediction. The organization of the three picture types in a sequence is very flexible. The choice is left to the encoder and will depend on the requirements of the application. Figure 3.2.1 illustrates the relationship among the three different picture types.

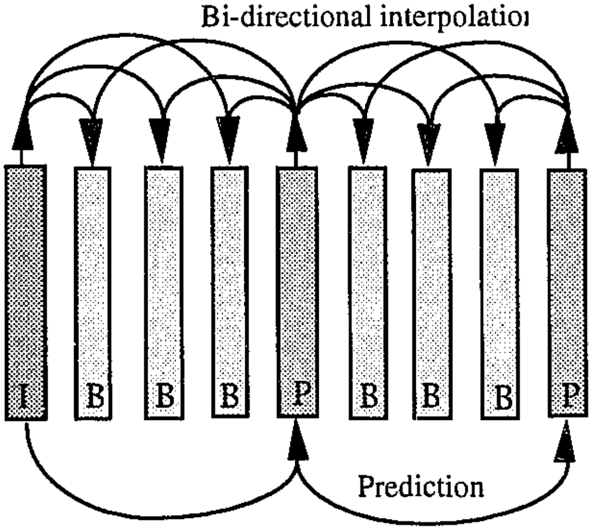


Figure 3.2.1 Example of a temporal picture structure

Motion representation

The choice of 16 by 16 macroblocks for the motion-compensation unit is a result of the trade-off between the coding gain provided by using motion information and the overhead needed to store it. Each macroblock can be one of a number of different types. For example, intra-coded, forward-predictive-coded, backward-predictive coded, and bidirectionally-predictive-coded macroblocks are permitted in B-pictures. Depending on the type of the macroblock, motion vector information and other side information are stored with the compressed prediction error signal in each macroblock. The motion vectors are encoded differentially with respect to the last transmitted motion vector, using variable length codes. The maximum length of the vectors that may be represented can be programmed, on a picture-by -picture basis, so that the most demanding applications can be met without compromising the performance of the system in normal situations.

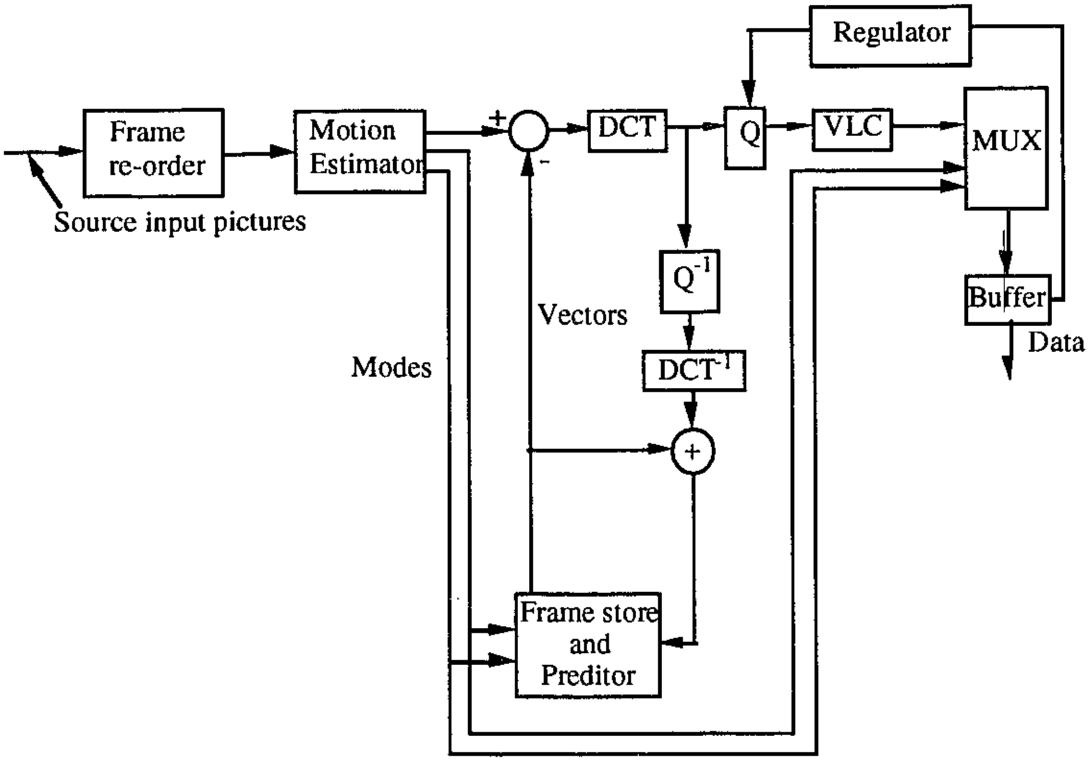
Spatial Redundancy Reduction

Both original pictures and prediction error signals have high spatial redundancy. MPEG uses a block-based DCT method with visually weighted quantization and run-length coding. After motion compensation prediction or interpolation, the residual picture is split into 8 by 8 blocks. These are transformed into the DCT domain where they are weighted before being quantized. After quantization many of the coefficients are zero and so two-dimensional run-length and variable length coding are used to encode the remaining coefficients efficiently.

3.2.2 Configuration of Coding

Encoding

MPEG does not specify an encoding process. It specifies the syntax and semantics of the bitstream and the signal processing in the decoder. As a result, many options are left open to encoders to trade-off cost and speed against picture quality and coding efficiency. Figure 3.2.2 shows the main functional blocks (This is illustrated based on MPEG1).



Where

DCT is the discrete cosine transform

DCT^{-1} is the inverse discrete cosine transform

Q is the quantization

Q^{-1} is the inverse quantization

VLC is the variable length coding

Figure 3.2.2 Simplified Video Encoder Block Diagram

The input video signal must be digitized and represented as a luminance and two color difference signals (Y, Cr, Cb). This may be followed by preprocessing and format conversion to select an appropriate window, resolution and input format. MPEG requires that the color difference signals (Cr and Cb) are subsampled with respect to the luminance by 2:1 in both the vertical and horizontal directions and are reformatted, if necessary, as a non-interlaced signal.

The encoder must choose which picture type to use for each picture. Having defined the picture types, the encoder estimates motion vectors for each 16 by 16 macroblock in the picture. Typically in P-Pictures, one vector is needed for each macroblock and in B-Pictures one or two vectors are needed.

If B-Pictures are used, some reordering of the picture sequence is necessary before encoding. Because B-Pictures are coded using bidirectional motion compensated prediction, they can be decoded after the subsequent reference (an I or P-Picture) has been decoded. Therefore the pictures are reordered by the encoder so that the pictures arrive at the decoder in the order for decoding. The correct display order is recovered by the decoder.

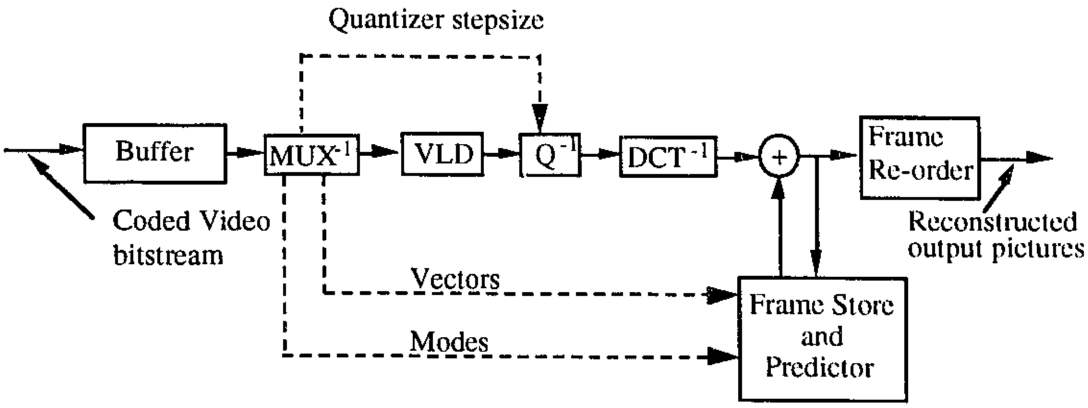
The basic unit of coding within a picture is the macroblock. Within each picture, macroblocks are encoded in sequence, left to right, top to bottom. Each macroblock consists of six component 8 by 8 blocks: four blocks of luminance, one block of Cb chrominance, and one block of Cr chrominance. The picture area covered by the four blocks of luminance is the same as the area covered by each of the chrominance blocks. This is due to subsampling of the chrominance information to match sensitivity of the human visual system.

Firstly, for a given macroblock, the coding mode is chosen. It depends on the picture type, the effectiveness of motion compensated prediction in that local region, and the nature of the signal within the block. Secondly, depending on the coding mode, a motion compensated prediction of the contents of the block based on past and/or future reference pictures is formed. This prediction is subtracted from the actual data in the current macroblock to form an error signal. Thirdly, this error signal is separated into 8 by 8 blocks (4 luminance and 2 chrominance blocks in each macroblock) and a discrete cosine transform is performed on each block. The resulting 8 by 8 block of DCT coefficients is quantized and the two-dimensional block is scanned in a zig-zag order to convert it into a one-dimensional string of quantized DCT coefficients. Fourthly, the side-information for the macroblock (type, vectors etc.) and the quantized coefficients data is encoded. For maximum efficiency, a number of variable length code tables are defined for the different data elements. Run-length coding is used for the quantized coefficient data.

The final step in the encoder is to regenerate I-Pictures and P-Pictures by decoding the data so that they can be used as reference pictures for subsequent encoding. The quantized coefficients are dequantized and an inverse 8 by 8 DCT is performed on each block. The error signal produced is then added back to the prediction signal and limited to the required range to give a decoded reference picture.

Decoding

Decoding is the inverse of the encoding operation. It is considerably simpler than encoding as there is no need to perform motion estimation and there are many fewer options. The decoding process is defined by MPEG. Figure 3.2.3 shows the main functional blocks.



Where

DCT^{-1} is inverse discrete cosine transform

Q^{-1} is inverse quantization

MUX^{-1} is demultiplexing and variable length decoding

VLD is the variable length decoder

Figure 3.2.3 Basic Video Decoder Block Diagram

The decoder reads the data elements in the bitstream according to the defined syntax. As the decoder reads the stream, it identifies the start of a coded picture and then the type of the picture. It decodes each macroblock in the picture in turn. The macroblock type and motion vectors, if present, are used to construct a prediction of the current macroblock based on past and future reference pictures that have been stored in the decoder. The coefficient data is decoded and inverse quantized. Each 8 by 8 block of coefficient data is transformed by an inverse DCT, and the result is added to the prediction signal and limited to the defined range.

After all the macroblocks in the picture have been processed, the picture has been reconstructed. If it is an I-picture or a P-picture it is a reference picture for subsequent pictures and is stored, replacing the oldest stored reference picture. Before the pictures are displayed they may need to be re-ordered from the coding order to their natural display order. After reordering the pictures are available, in digital form, for post-processing and display in any manner that the application chooses.

3.2.3 Coding Layers

The MPEG bit stream model is organized in six layers, each of which either supports a signal processing or system function (Figure 3.2.4)

Layers of the syntax
Sequence layer
Group of pictures layer
Picture layer
Slice layer
Macroblock layer
Block layer

Figure 3.2.4 Structure of the coded bit stream

- *Sequence layer* : A coded video sequence commences with a sequence header and is followed by one or more groups of pictures and is ended by a sequence_end_code.

Immediately before each of the groups of pictures there may be a sequence header. Within each sequence, pictures shall be decoded continuously.

- *Group of pictures layer* : A group of pictures (GOP) is a series of one or more consecutive pictures intended to assist random access into the sequence. In the coded bitstream, the first coded picture in a group of pictures is an I-picture. The order of the pictures in the coded bitstream is the order in which the decoder processes them. In display order, the last picture in a group of pictures is always an I-Picture or a P-Picture, and the first is either an I-Picture or the first B-Picture of the consecutive series of B-Pictures which immediately precedes the first I-Picture.

- At the encoder input : I B B P B B P B B I B B P B B P B B I B B P B B P B B I
- At the encoder output, in the coded bitstream, and at the decoder input :

  I P B B P B B I B B P B B P B B I B B P B B P B B I B B
- At the decoder output : The decoder should display pictures in the same order as the encoder input.

- *Picture layer* : A picture in MPEG terminology is the basic unit of display and corresponds to a single frame in a video sequence. The spatial dimensions of a frame are variable and are determined by the requirements of an application.
- *Slice layer* : A slice, which is a horizontal strip within a frame, is the basic processing unit in the MPEG coding scheme. A slice is a series of an arbitrary number of macroblocks with the order of macroblocks starting from the upper-left of the picture and proceeding by raster-scan order from left to right and top to bottom. Every slice shall contain at least one macroblock. Slices shall not overlap and there shall be no gaps between slices. The position of slices may change from picture to picture. The first slice shall start with the first macroblock in the picture and the last slice shall end with the last macroblock in the picture. A frame is divided into a

number of contiguous macroblock slices. A slice is an autonomous unit since coding a slice is done independently from its neighbors for synchronization under error environment.

- *Macroblock layer* : A macroblock is the basic coding unit in the MPEG algorithm. A macroblock is a 16x16 pixel segment in a frame. A macroblock contains a section of the luminance component and the spatially corresponding chrominance components.
- *Block layer* : A block is the smallest coding unit in the MPEG algorithm. The term "block" can refer either to source and reconstructed data or to the DCT coefficients or to the corresponding coded data elements. When "block" refers to source and reconstructed data it refers to an orthogonal section of a luminance or chrominance component with the same number of rows and columns. It is made up of 8x8 pixels and can be one of three types; luminance(Y), red chrominance(Cr), and blue chrominance(Cb). When "block" refers to the DCT coefficients there will be 64 coefficients. The block is the basic unit in intraframe DCT coded frames.

Chapter 4

PRIORITIZATION & DUMMY CELL INSERTION SCHEMES

4.1 Motivation

4.1.1 Previous Work

In order to secure the delivery of the continuous video information through an ATM network, different levels of priority have been proposed [QCC93, PMC94, MCC93]. The cells that can be lost without disastrous effects are transmitted with lower priority to those that must be received within a strictly specified time period.

One method for immunization of the video services to cell losses is the multilayer coding technique [GSC87, KVB89, PLT89, MTB88]. In this method, the video signal is partitioned into various classes, depending on the importance of the information. For example, synchronization and control signals belong to the most important class. In two-layer video coding, the picture information is divided into two categories, each carrying part of the information needed for the full representation of the original picture [MTC89]. Although various types of two-layer coding have been proposed over the years [GTV91, STV91, SUA90], the basic idea remains the same. In a two-layer video codec the input picture frame is coarsely coded by a first-layer or base-layer coder and the coded data are packetized into fixed-length cells. The base-layer cells are called guaranteed cells as they require guaranteed transmission with no losses [MTC89]. The second layer is formed by coding the difference between the input signal and the locally decoded information of the first layer. The resulting data are packetized into low-priority cells for transmission through the network. The second-layer cells are called enhancement cells because their reception improves the picture quality. At the decoder both the guaranteed and enhancement cells are decoded separately and added together to form the reconstructed picture. Multiresolution techniques that lead to layered coding schemes are particularly amenable to prioritized transmission. However, these coding schemes have not gained favor in practice [PMC94]. We should consider several issues in the prioritization of cells.

1) Can it be implemented using simple schemes ?
2) Can the output of a MPEG coder be prioritized satisfactorily ?
3) Are the characteristics of MPEG fully exploited ?
4) What is the cost of prioritization in terms of extra bandwidth required ?

The main disadvantages of the two-layer coding technique are: they require another layer in addition to the basic layer; some extra bandwidth is required to implement the two layers. Instead of using the additional layer, we can simply utilize the cell loss priority provided by ATM networks. A recent study has used block-based intraframe discrete cosine transform (DCT) coding [MCC91], but has not utilized any interframe coding which is the main property of temporal processing technique in MPEG coding. This property is the important factor of causing error propagation temporally in a video sequence.

4.1.2 Cell Loss and Video Quality

Due to the way the MPEG (Moving Picture Expert Group) sequences are encoded, the impairments caused by the cell losses very much depend on the contents of the information being conveyed by the missing cells. In the sequel, we illustrate the impairments caused when losing various kinds of pieces of information of an MPEG sequence.

Impairments by frame type

Figures 4.1.1, 4.1.2 and 4.1.3 show the impact over the different pictures of a group due to the loss of cells carrying information belonging to the I, P and B frames, respectively. In these figures, the signal to noise ratio for each one of the frames is shown. This signal to noise ratio (SNR) has been calculated as [AFD89]:

$$\mathrm{SNR[dB]} = 10 \log_{10}(\sigma^2 / \sigma_{ls}^2) \qquad -(4.1)$$

where σ_{ls}^2 is its average least squares error. Among the quantitative measures, a class of criteria used often is called the *mean square criterion*. It refers to the average or sum of squares of the error between two images. For $M \times N$ images of which the pixel levels of input and output sequences of a two-dimensional system are given by $u(m,n)$ and $u'(m, n)$, respectively, the quantity

$$\sigma_{ls}^2 \cong \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} |u(m,n) - u'(m,n)|^2 \qquad -(4.2)$$

is called the *average least squares error.* And σ^2 is the variance of the original frame i.e. the average value of u(m,n) is used instead of u'(m,n). In our case, we have set σ to 1. Therefore, a value of SNR = ∞ means that the frame has not been affected. In the figures, this fact is represented by the value SNR=0.

In the particular cases depicted in Figures 4.1.1 to 4.1.3, two cells have been lost. In Figure 4.1.1, one cell from the first I frame and another cell from the second I frame were lost. The loss of these cells affects both the P and B frames in the same group of pictures. Figure 4.1.2 shows the result of losing two cells from two different P frames. As expected, in this case only the P and B frames are affected as the decoding of the I frames does not depend on the P frames. Finally, Figure 4.1.3 shows that the loss of cells pertaining to a B-frame does not affect all other frames.

Visual Impact

In order for the decoder to be able to interpret the different parts of information in a video stream, the coder requires to include in the video stream some control information regarding, among other information, the beginning of some layers of the video stream. The loss of this control information may cause that the decoder, upon receiving the video stream, be unable to identify the various components of the stream. Among the various types of control information, the proper identification of the start code pertaining to a new picture and new slice is essential for the proper operation of the decoder. Even though the design of the MPEG coding scheme has been done by taking into account some of the problems which may arise when the information is lost, the loss of packetized video units (cells in an ATM environment) may induce a large number of impairments.

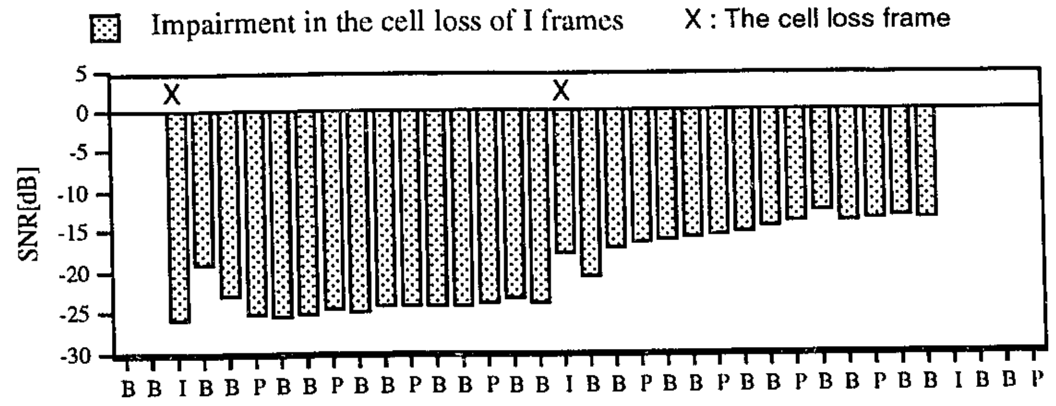


Figure 4.1.1 Impairment due to the loss of I frames cell.

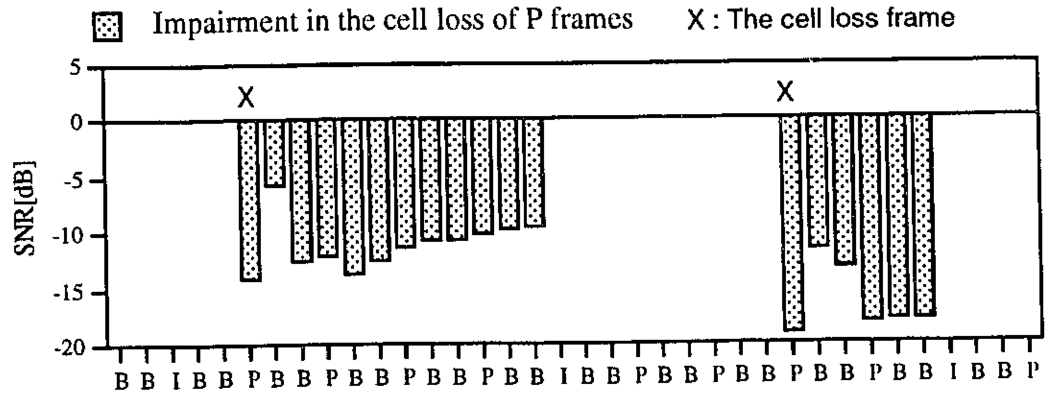


Figure 4.1.2 Impairment due to the loss of P frames cells

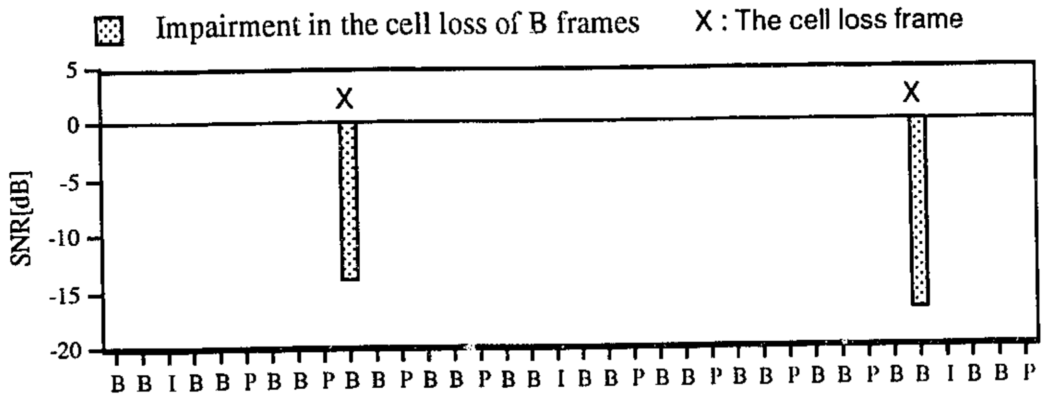


Figure 4.1.3 Impairment due to the loss of B frames cells

The loss of different parts of the MPEG bit stream, and the fact of propagating the impairments of losing information can have the following visual effects:

- *Blockiness* caused by non-detected error in a block or macroblock.
- *Block Displacement* caused by an error in the motion vectors,
- *Block Bluriness and Artifact* caused by an error in the quantization scale or DCT coefficients,
- *Jerky Motion and Hesitation* from the failure of the error concealment (replacement of corrupt blocks by blocks in a previously displayed frame).

4.2 Cell loss and Impairments

In section 4.1.1, we illustrate the problems that may arise when losing the information pertaining to the different types of frames. In this section, we look into the kind of problems that the decoder may face in order to properly identify the different pieces of information of the video stream under the presence of cell losses.

1) Mismatched Macroblocks

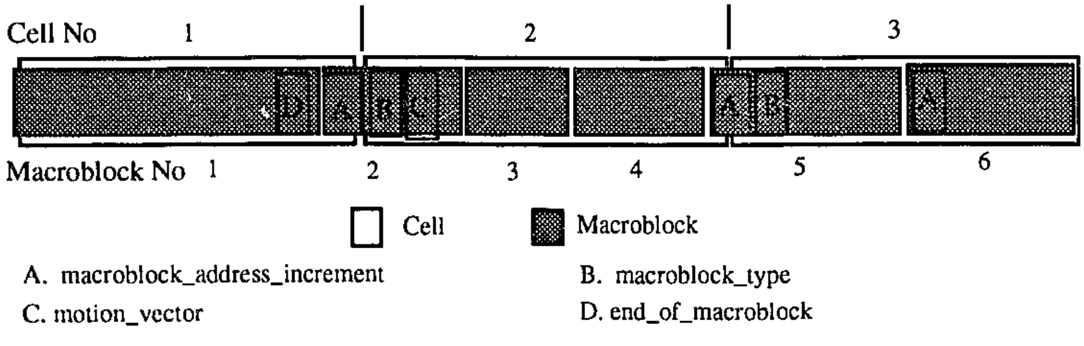


Figure 4.2.1 The case of mismatched macroblocks

This example illustrates the case when the loss of information may cause an MPEG decoder to misinterpret information at the macroblock layer. From Figure 4.2.1, let us suppose that cell 2 is lost. In this case, the decoder receives the macroblock address increment, A, for the macroblock 2 followed by the third ATM cell carrying information of macroblock 5. In this case, the decoder can hardly detect the loss of information and interprets the contents of macroblock 5 as being the sequel of the macroblock 2. This overrun is caused by the fact that, according to the standard[MCM93], a decoder processing a macroblock will continue to do so until detecting the end of the block. However, if part of the macroblock is lost, the decoder will be unable to detect the beginning of the next macroblock (for further reference, the reader may consult [MCM93] pp. 79-85) . This situation will cause the decoder misinterpret from there on the information, therefore propagating the error until the end of the current slice. As we shall see, the use of a dummy cell to be inserted when an ATM cell loss occurs aims to provide the decoder with the means of detecting the loss of information upon receiving the dummy cell. This kind of situation may arise when losing macroblock information such as the macroblock type and motion vector in addition to the macroblock address increment. This phenomenon will result in visual effects, such as, blockiness, block displacement, block bluriness and artifact, jerky motion and hesitation.

2) Mismatched Slices

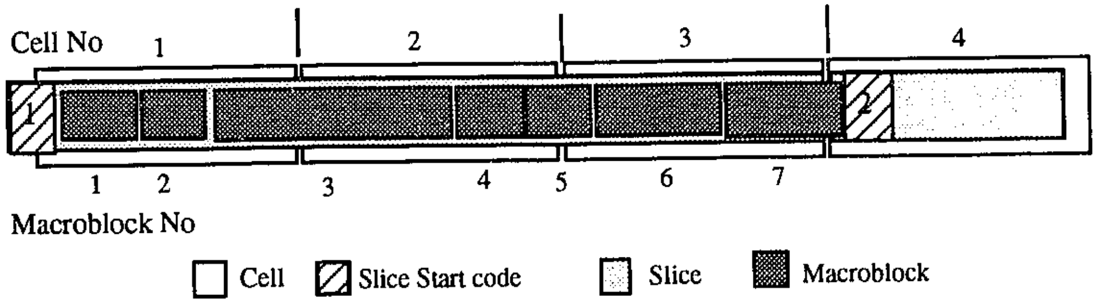


Figure 4.2.2 The case of mismatched slices

In this case, the loss of information may result in the loss of several macroblocks within the corrupted slice as well as all macroblocks of the slice following the corrupted slice. From Figure 4.2.2, let us suppose that cell 3 is lost. In this case, an MPEG decoder can hardly detect the loss of information. Upon receiving the cell 4, the decoder will interpret the information in this cell as belonging to macroblock 5. Even though the decoder will eventually recognize that this slice start code does not make part of macroblock 5, the decoder can not recover this slice start code without having to go back and decode the slice start code. This operation is unsuitable for a decoder aiming to work in a high-speed network environment. This means that all the macroblocks in this second slice will have to be lost (352 x 16 pixels)

3) Mismatched Pictures

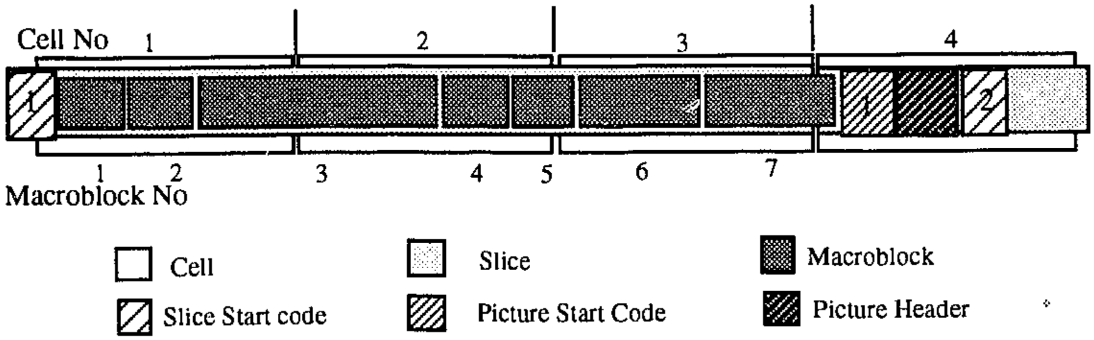


Figure 4.2.3 The case of mismatched pictures

This case is very similar to the previous one, but the extend of its impact may propagate much more. From Figure 4.2.3, let us suppose that cell 3 is lost. In this case, the decoder receives the second part of data of the macroblock 5 in cell 2 and comes to lose the part of data in cell 3. An MPEG decoder can hardly detect the loss of information and will interpret the picture start code in cell 4 as being part of the macroblock 5, thus missing the picture start code. This will result in the loss of all the macroblocks in this picture (352 x 240 pixels). Even though the decoder will eventually recognize that this picture start code does not make part of the fifth macroblock, the decoder can not recover this picture start code because it is already late to do so.

4.3 Prioritization and Dummy Cell Insertion

4.3.1 The Priority Scheme

In a cell loss priority determination method, the cell priority is used to provide guidance to the network in the event of congestion. A value of 0 indicates a cell of relatively higher priority, which should not be discarded unless not any other alternative is available. A value of 1 indicates that this cell is subject to discard within the network. We propose to make use of this facility when transmitting an MPEG video sequence.

As shown in the Figure 4.3.1, we propose to assign high priority to all the headers from the sequence layer to the picture layer. The amount of this information represents a very small portion of the overall information associated with the video stream. However, this information is essential for the decoding process and display of the incoming video bit stream. All the other information, slice headers and macroblocks can be transmitted using the high or low priority bits based on frame types.

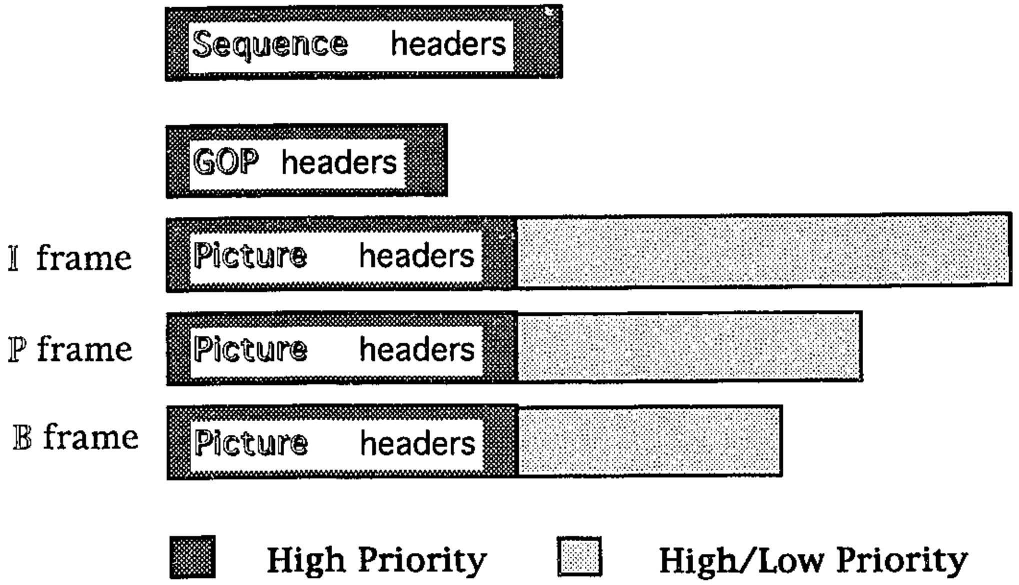


Figure 4.3.1 The structure of prioritization

In section 5, we explore the following three scenarios:

1) Use of high priority to I and P frames.
2) Use of high priority to I frames.
3) No priority is used.

According to the assignment of priorities given above, we assume that only frames which have been assigned low priority may be lost. For instance, in case 2 above, only B and P frames may be lost.

4.3.2 The Dummy Cell Insertion Scheme

In a dummy cell insertion method, when a cell is determined to be missing from the incoming cell stream, the receiving AAL (ATM Adaptation Layer), specifically AAL1 which comprises a sequence number field, will insert a dummy cell containing a fixed payload to replace the lost cell. Insertion of dummy cells maintains the bit count integrity and thus helps to prevent the impairments from being extended.

Characteristic of decoder

- *Searching next_start_code*: The next_start_codes are used to identify the start of each layers from the sequence layer to the slice layer. In a frame, if the coded data are corrupted, and the decoder detects this, then the decoder can search for the next slice start code and resume decoding the picture from that point. If the data are corrupted, and the decoder fails to detect this, then the decoder may encounter an unexpected slice start code. It should then start decoding from the position indicated in the slice header.

- *Definition of next_start_code function*: The next_start_code function removes any zero bit and zero byte stuffing and locates the next_start_code. This function checks whether the current position is byte aligned. If it is not, zero stuffing bits are present. After that any number of zero bytes may be present before the start-code. Therefore start-codes are always byte aligned and may be preceded by any number of zero stuffing bits.

```
next_start_code(){                                        No. of bits
    while(!bytealigned())
        zero_bit                                              1
    while(nextbits() != '0000 0000 0000 0000 0000 0001')
        zero_byte                                             8
}
```

- *VLC (Variable Length Code) pattern used in the MPEG decode*r: For more compression of the number of bits, MPEG is using the Huffman coding as a statistical coding technique that assigns codewords to values to be encoded. Values with high frequency of occurrence are assigned short codewords, and those of infrequent occurrence are assigned long codewords. On average, the more frequent shorter codewords dominate, such that the coding string is shorter than the original data. This coding method generates variable length codes.

- *Reserved next_start_codes*: Start-codes are reserved bit patterns that do not otherwise occur in the video stream. All start codes are byte aligned. Some example start codes are shown in Table 4.3.1.

name	**hexadecimal value**
picture_start_code	00000100
slice_start_codes	00000101 through 000001AF
sequence_header_code	000001B3
group_start_code	000001B8

Table 4.3.1 The example of start codes

Dummy Cell Bit Pattern

We saw that when the corrupted data happened it is very important for a decoder to be able to detect the loss of information to reduce the propagation of impairments. Our objective is therefore to define a proper code for the dummy cell to be inserted to allow the decoder to quickly detect that information has been lost.

According to the reserved starting codes used, the VLC cannot have the pattern of more than twenty three '0' bits. Therefore, in order to make the decoder detect the corrupted data so that it maintains the bit count integrity, we consider using more than twenty three '0' bits as dummy cell pattern. Furthermore, in order to avoid a mismatching of this pattern with the code "next_start_code", the bit value '1' as next bit after the twenty three '0' bits should be avoided.

Finally, the dummy cell should not be filled out with all '0' bits in the last byte in order to avoid the possibility of creating a start code when concatenating the information of the dummy cell with the contents of the following cell. These cases can make the decoder encounter a next_start_code. This will result in the same problem as the ones previously mentioned. The principle is simple. We should prevent a decoder from generating more impairments by misinterpreting information. Misinterpreting will result in worse results as compared to the case when no dummy cell is used.

As a result, we can consider the following structure for defining the composition of the dummy cell.

1) at least, the first twenty three bits should be '0' to detect the corrupted data as soon as possible.

2) the last byte should not be 00(H)

3) the code next_start_code pattern should be avoided in any place.

When we consider 1) and 3), the first three bytes in a dummy cell should be zero and when we consider 2) and 3) the last byte should be at least 02(H). The examples shown in Table 4.3.2 are possible patterns (hexadecimal value).

Byte No.	1	2	3	4	5	.....	.....	.....	n-3	n-2	n-1	n
Pattern	**00**	**00**	**00**	00	00				00	**00**	**00**	**02**(H)
:												
Pattern	**00**	**00**	**00**	02	00				00	**00**	**00**	**02**(H)
:												
Pattern	**00**	**00**	**00**	ff	ff				ff	**ff**	**ff**	**ff**(H)

Table 4.3.2 The examples of possible dummy cell patterns

Figure 4.3.2 shows the simple dummy cell pattern and how to detect the corrupted data and search a next_start_code.

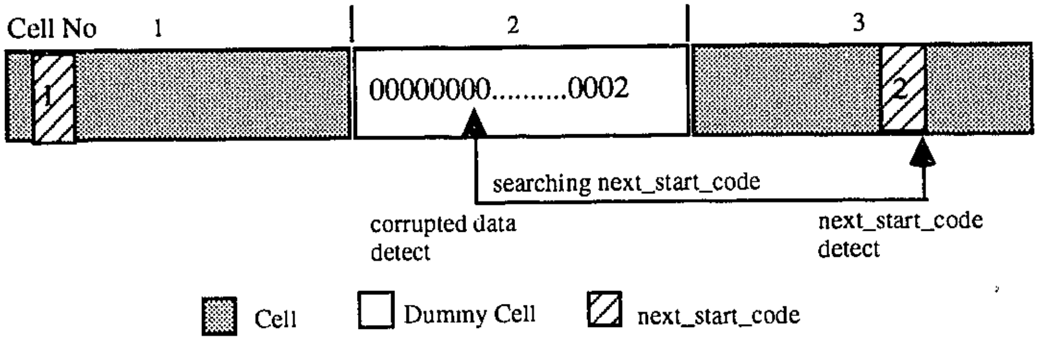


Figure 4.3.2 An example of the use of a dummy cell

4.4 Experimental Results

For this study, we used the "flower" bit stream (MPEG1 video bit stream) and the decoder programmed by the University of Berkeley and available through the Internet [MUB93]. The frame sequence, frame composition and cell distribution per slice are the major factors affecting the performance of the suggested schemes. The results herein are representative of the problems that may arise when information is lost.

4.4.1 Experimental Data

Figure 4.4.1 shows the frame composition of the MPEG flower video stream per frame type I, P, B frames. The data shown were calculated only by considering the bit stream rates of macroblocks which are the basic coding units. That is, the bit stream of headers above the picture header was not considered in our calculations.

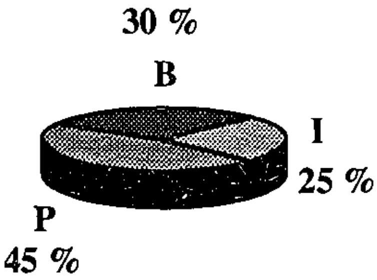


Figure 4.4.1 Frame Composition of the MPEG flower video stream

Figure 4.4.2 shows the cell distribution per slice in different frames. As previously discussed, if the decoder is unable to detect the lost information this may result in the loss of a large number of the macroblocks belonging to a slice, and even macroblocks to subsequent slices. According to the coding scheme, the slice size is variable. Therefore, in the case of lost information, the slice size is an important parameter determining the amount of information being lost. In this sequence, the average cell numbers of a slice in I, P, B frame are 3, 11, 27 respectively. This means that a lost cell of a slice in one I frame can affect the impairment of frame more than a lost cell of a slice in any of the other frames.

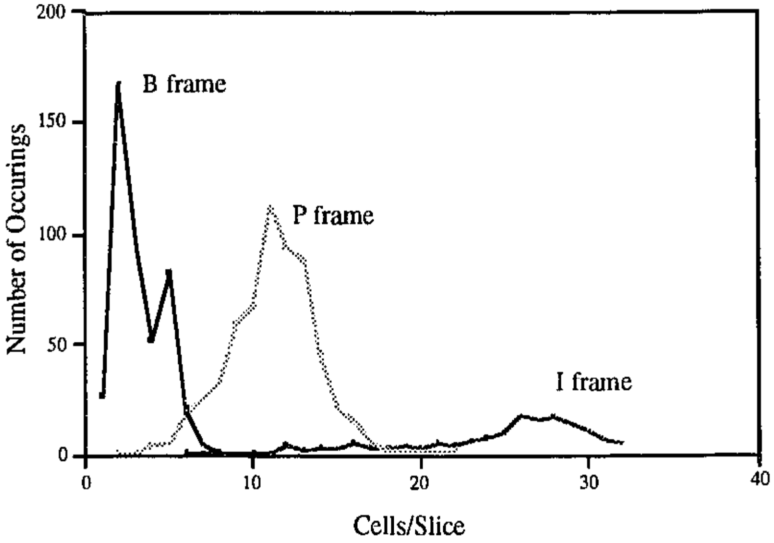


Figure 4.4.2 Cell distribution per slice in different frames

4.4.2 Performance of the Priority Scheme

Figure 4.4.3 shows the image quality for the three different priority assignments. In our simulation, cell losses followed a uniform distribution. As expected, when the high priority is assigned to the I and P frames, the degradation of the image quality resulting from the cell loss of low priority data was much smaller than when the high priority was only assigned to the I frames. As seen from this figure, the use of priority greatly helps to protect the quality of the image. The total number of cells used in this video stream is about 15000, and the numbers of lost cells tested are from 1 (10^{-2} %) to 150 (1 %). These results were obtained by running the simulation approximately one hundred times. These results provide with a baseline for the results when using the dummy cell scheme with and without the priority scheme.

Figures 4.4.4 and 4.4.5 show the sequential images in both cases of no priority and high priority to I & P frames. As we evaluated in section 4.1.2 (Cell Loss and Video Quality), when we did not assign priority to frame types, the impairments due to the cell loss of I or P frames have been propagated to P and B frames so that we can easily and apparently see the visual impact at each frame. But, since the impairments due to the cell loss of B frames occur only on B frames themselves, we can hardly see the visual impact in this case.

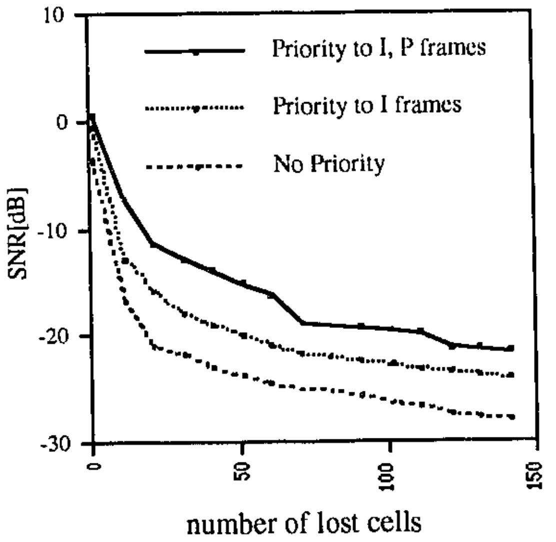


Figure 4.4.3 SNR vs CLR on Prioritization without dummy cell insertion

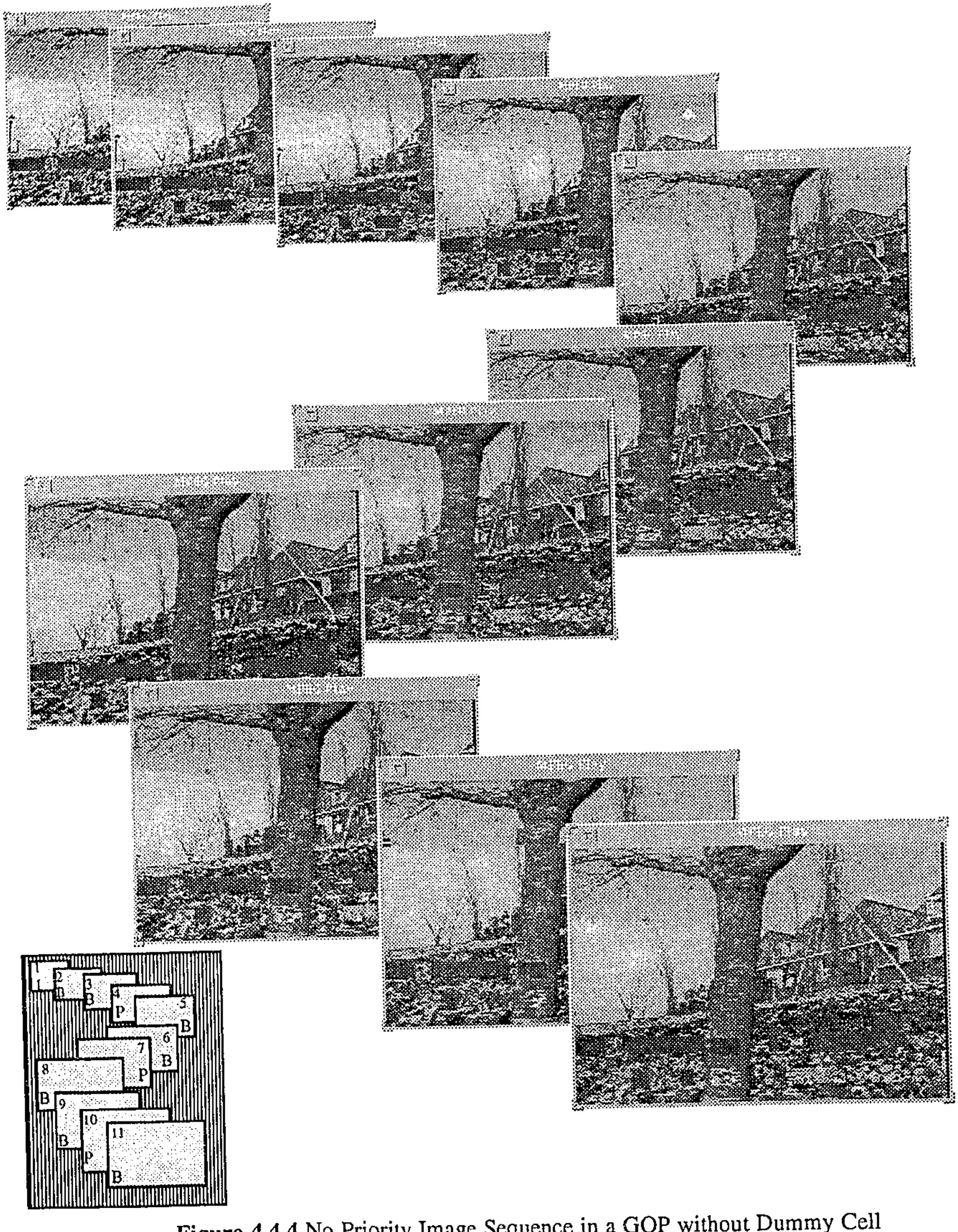


Figure 4.4.4 No Priority Image Sequence in a GOP without Dummy Cell

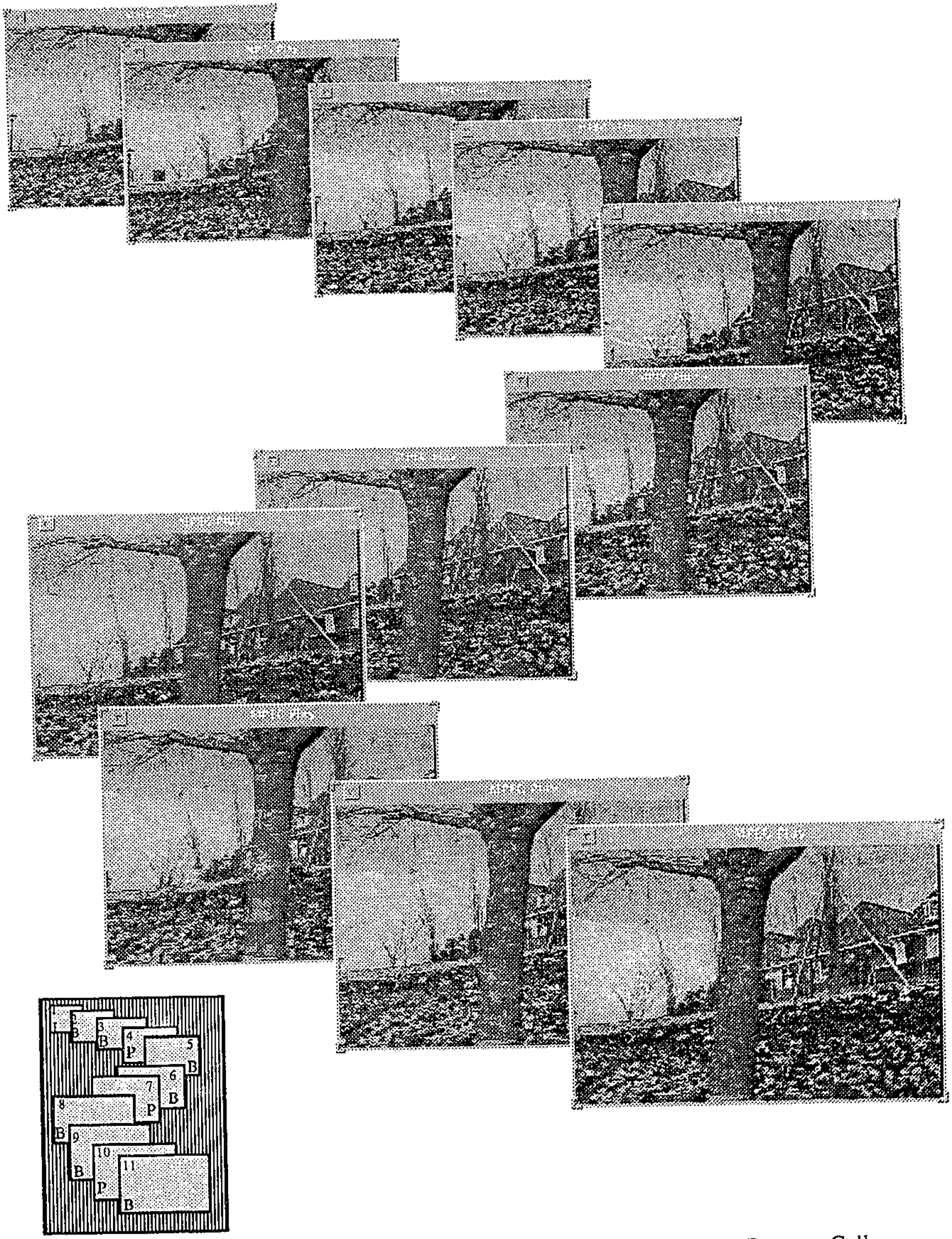


Figure 4.4.5 Prioritized Image Sequence in a GOP without Dummy Cell

4.4.3 Performance of the Dummy Cell Insertion Schen

Dummy Cell Bit Pattern

Before we tested the effectiveness of inserting a dummy cell to replace lost cells, we had explored the performance of various dummy cell patterns. As we have already mentioned, it is important to choose the right dummy cell pattern providing the decoder with the means of detecting the loss of information. In comparing different dummy cell patterns, we fixed the number of cell losses to 0.7 % jointly with the different prioritization schemes previously mentioned. The examples shown in Table 4.4.1 are the dummy cell patterns which were used in this test. Among these patterns, the pattern A corresponds to the case when no dummy cell is used.

Byte No.	1	2	3	4	5	.....	.....	n-3	n-2	n-1	n
Pattern A											
Pattern B	00	00	00	00	00			00	00	00	00(H)
Pattern C	00	00	00	00	00			00	00	00	01(H)
Pattern D	00	00	00	00	00			00	00	00	02(H)
Pattern E	00	00	00	ff	ff			ff	ff	ff	ff(H)
Pattern F	ff	ff	ff	ff	ff			ff	ff	ff	ff(H)

Table 4.4.1 The examples of the dummy cell patterns used

As shown in Figure 4.4.4, we obtained the best results when using the dummy cell patterns D, E and the worst results when using the dummy cell pattern C. In pattern B and C, the decoder was often unable to continue to decode the bit stream. This was caused by the wrong generation of the code indicating the beginning of a new slice. This problem was more often encountered when using the pattern C than the pattern B. This problem was even more common in case A, i.e., when no dummy cell was used.

From these results, we conclude that patterns B and C must not be used in MPEG standard because they may generate with high probability bit patterns corresponding to reserved bit streams. For patterns D and E, we can see the same results, which correspond to the characteristics explained in section 4.2.1. Finally, pattern F does not show good results. This is because pattern F can result in the creation of a reserved code used in the macroblock structure, such as '111111111..'. This code is used by MPEG for coding information fields, such as the Macroblock Addressing, Macroblock Type, Macroblock Pattern, Motion Vectors and DCT Coefficients.

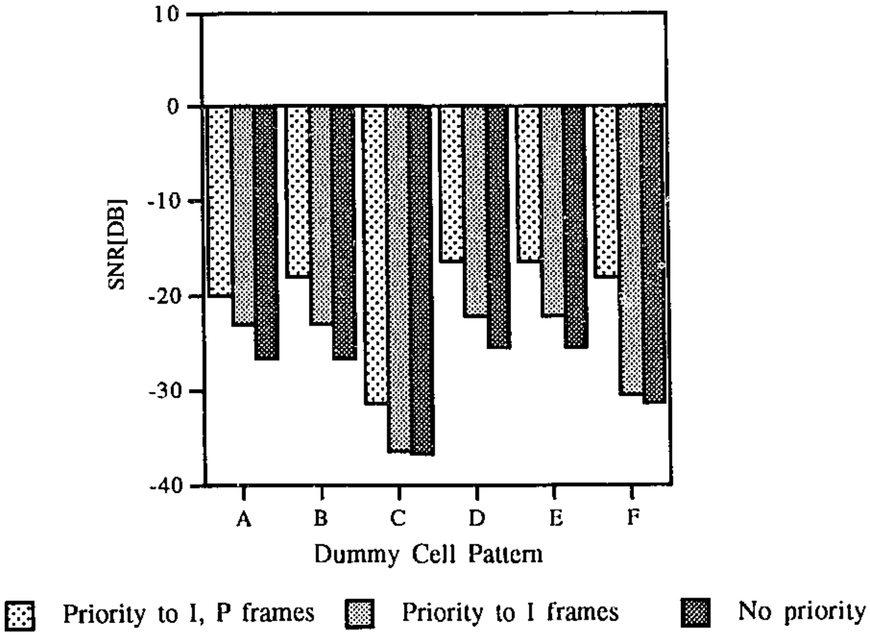


Figure 4.4.4 Comparison on dummy cell patterns

Priority and Dummy Cell Insertion Scheme

For this study, we used the dummy cell pattern D defined in the previous section. The proper dummy cell patterns have been tested and this pattern, among them, is the simplest one of many patterns which are effective in increasing SNR at the same CLR (Cell Loss Ratio).

Figures 4.4.5, 4.4.6 and 4.4.7 show SNR vs CLR when we insert a dummy cell for the lost cell compared to the no dummy cell insertion in each case of prioritization. These figures illustrate better quality of video image by using the dummy cell which has the appropriate dummy cell pattern. In Figure 4.4.8, we can see that even though we use the dummy cell, the effect of prioritization is still remained. It is important to notice that the results depicted in Figures 4.4.5 to 4.4.7 have been obtained by running our experiments several times. It is also important to point that in some of these cases, except for the cases where the bit patterns of the dummy cells used were E and F, the decoder stopped displaying the sequence. This fact confirms that by using these dummy cells bit patterns, we have been able to do the system more robust. Figures 4.4.9 and 4.4.10 show a mismatched slice and a mismatched macroblock under a cell loss respectively. In figure 4.4.9(b), when the dummy cell insertion scheme has not been applied, the impairments have appeared at the point '1' , '2', '3' due to 1) the lost information of a cell carrying macroblocks, 2) the mismatched slice, 3) and the error propagation from the previous frame respectively. Whereas, when the dummy cell insertion scheme has been applied, the impairments of points '2' and '3' have disappeared. Similarly as in figure 4.4.9, the impairment of a mismatched macroblock has appeared at the point '2' in figure 4.4.10(b) while it has disappeared in figure 4.4.10(c).

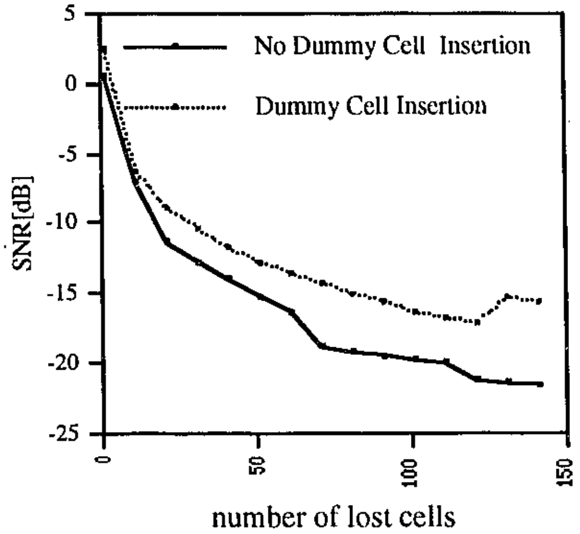


Figure 4.4.5 SNR vs CLR with High Priority to I, P frames

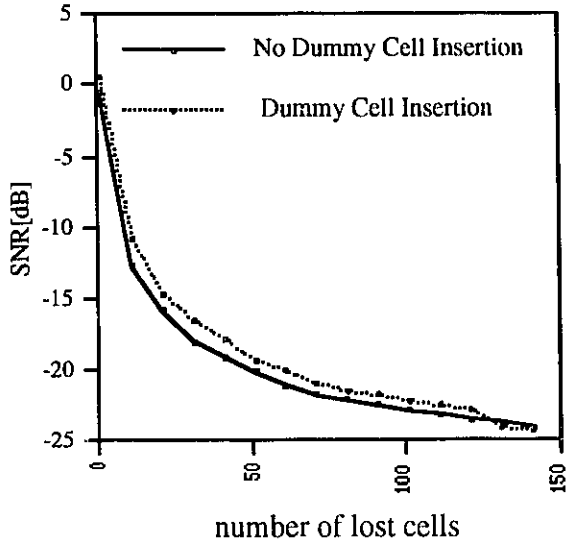


Figure 4.4.6 SNR vs CLR with High Priority to I frames

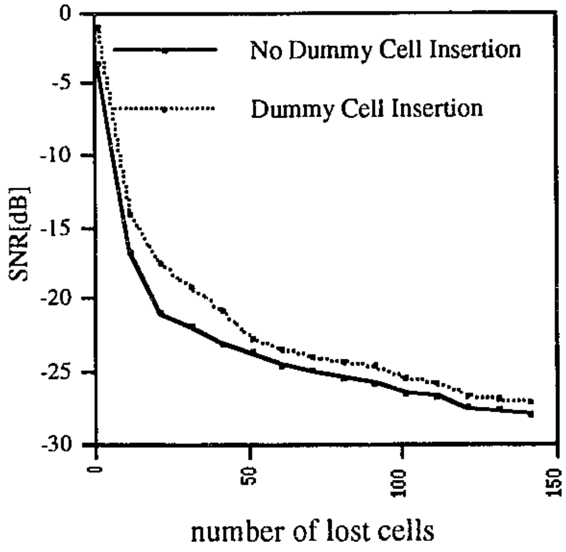


Figure 4.4.7 SNR vs CLR with No Priority

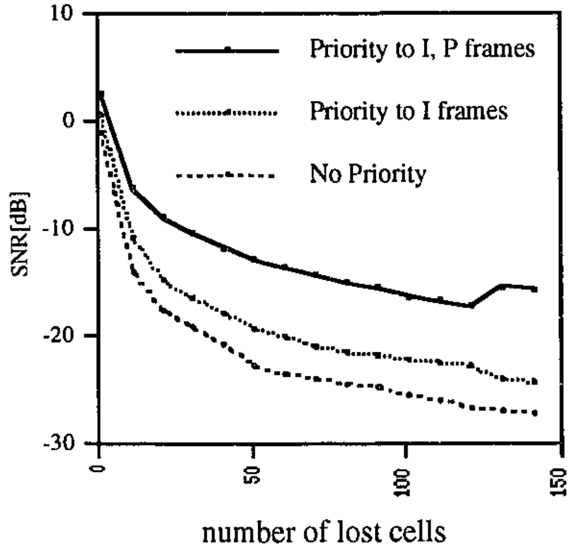
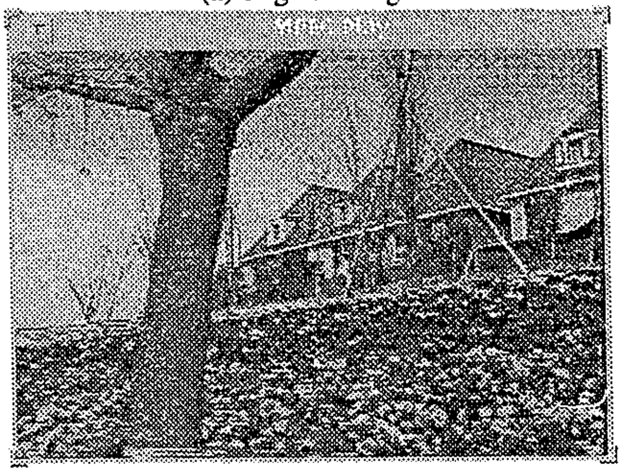
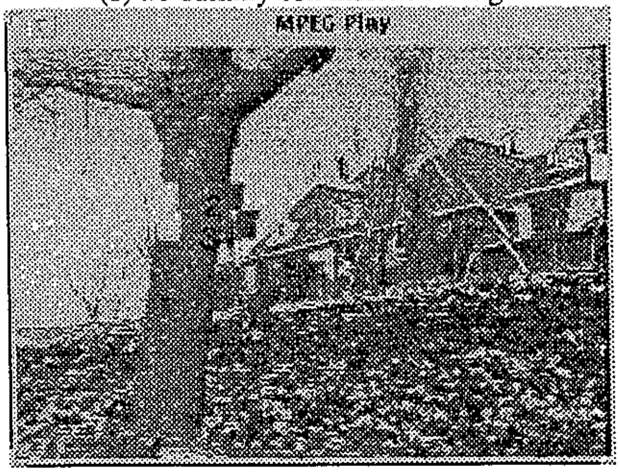


Figure 4.4.8 SNR vs CLR with Prioritization and dummy cell insertion

(a) original image



(b) no dummy cell insertion image



(c) dummy cell insertion image

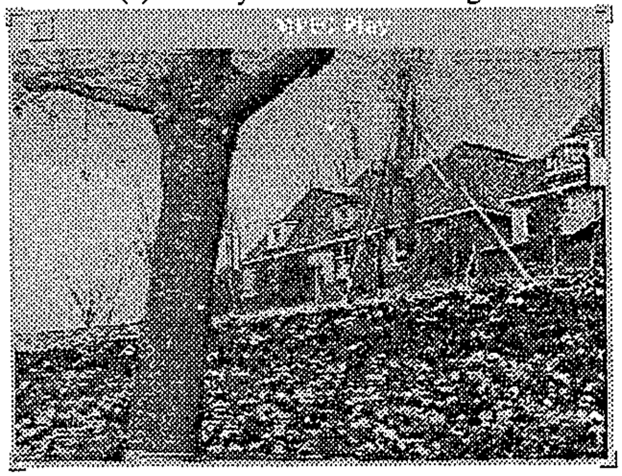
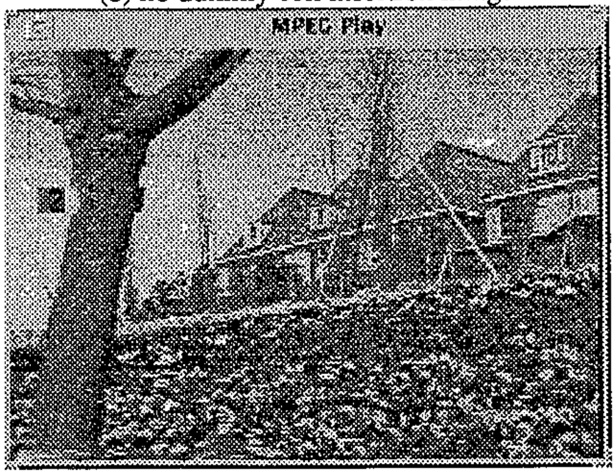


Figure 4.4.9 Comparison between No Dummy Cell Insertion Image and Dummy Cell Insertion Image for Mismatched Slices in ***flower.mpg***

(a) original image

(b) no dummy cell insertion image



(c) dummy cell insertion image

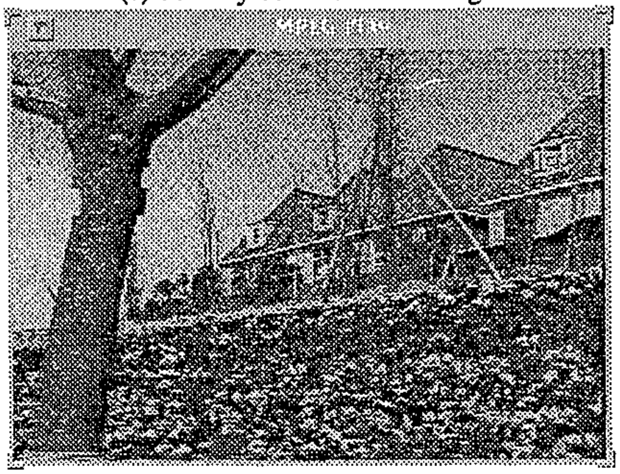
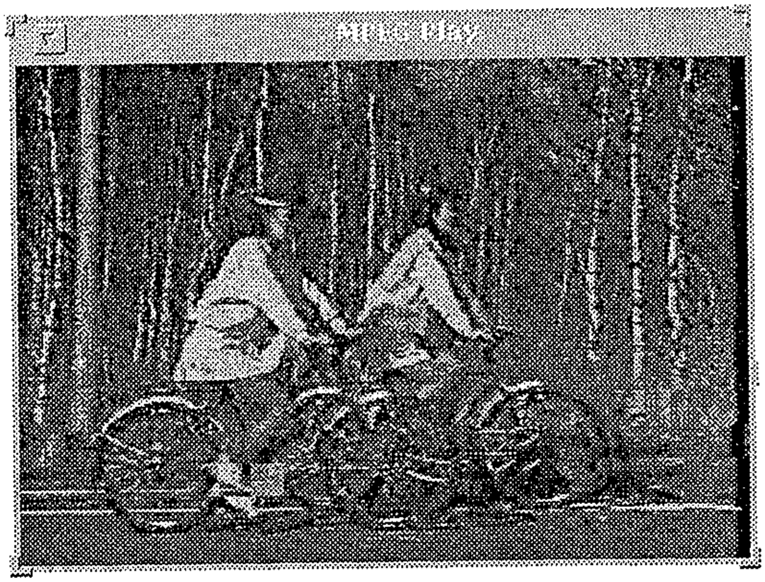


Figure 4.4.10 Comparison between No Dummy Cell Insertion Image and Dummy Cell Insertion Image for Mismatched Macroblocks in ***flower.mpg***

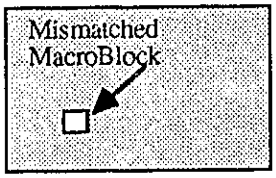
The results from the dummy cell insertion, for several kinds of video sequences - bicycle.mpg, stoelendans.mpg, waterski.mpg and zoom.mpg, are shown in the Figures 4.4.11 - 4.4.17. For each video sequence, the average number of cells per slice, the frame composition ratio, the configuration of a GOP and motion direction are informed. Some example impairment of the mismatched macroblocks and slices are shown for the first three video sequence, and the propagated impairment due to the mismatched slice are shown for zoom.mpg.

4.5 Conclusion

The multilayer coding technique has been proposed to secure the delivery of the continuous video information over an ATM network. But it has some drawbacks such as the increasing amount of information and the complexity of implementation. In the prioritization scheme, exploiting the benefit of the concept of multilayer coding technique and recovering the drawbacks, the video signal has been partitioned based on the frame types. In this scheme, the algorithm was simply implemented and the SNR effectively improved. In the dummy cell insertion scheme, the system (MPEG video) can be made more robust by using the dummy cells of specific patterns determined in this work, as well as the SNR more improved in addition to the improvement of the prioritization scheme. The scheme has proved to be effective in a large variety of video sequences, varying from video sequences containing high level of activity to low level of activity.



(a) UP : no dummy cell insertion image



(b) Down : dummy cell insertion image

Video Sequence	Frame Size	Motion
bicycle.mpg	362X270	High → ↻
Average cells/slice	I : P : B	
14: 9: 4	15:39:46	
GOP		
IBBPBBPBBPBBPBB		

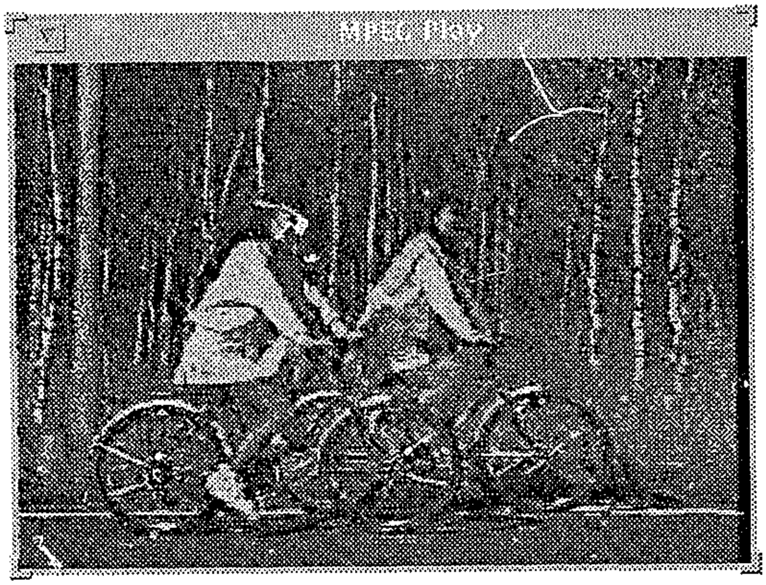
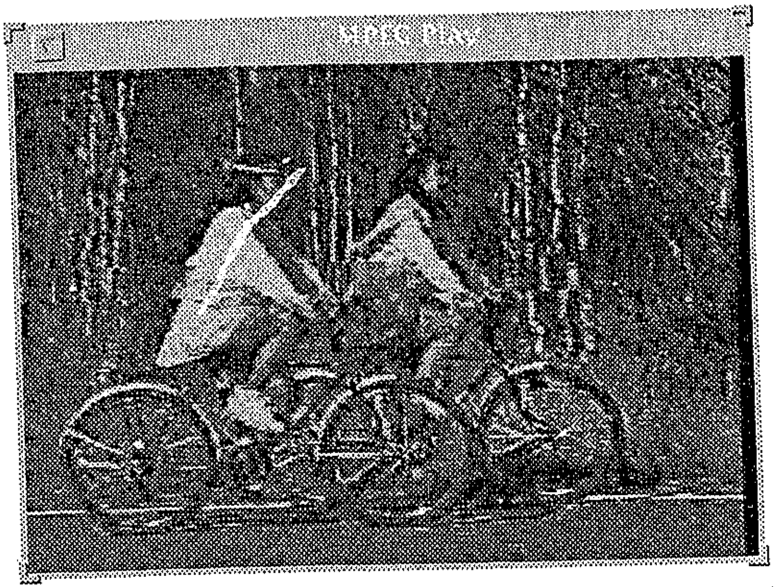
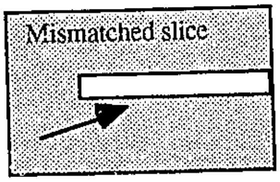


Figure 4.4.11 Mismatched Macroblocks in ***bicycle.mpg***



(a) UP : no dummy cell insertion image



(b) Down : dummy cell insertion image

Video Sequence	Frame Size	Motion
bicycle.mpg	362X270	High → ↻
Average cells/slice	I : P : B	
14: 9: 4	15:39:46	
GOP		
IBBPBBPBBPBBPBB		

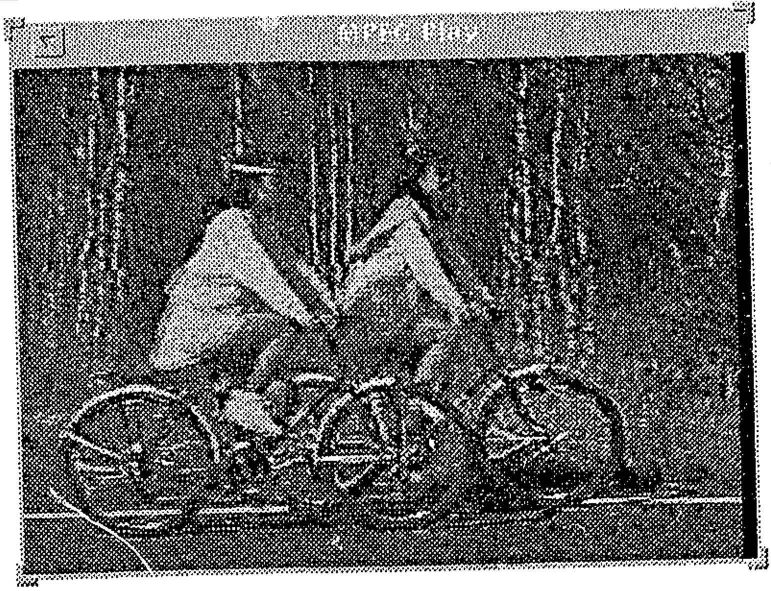
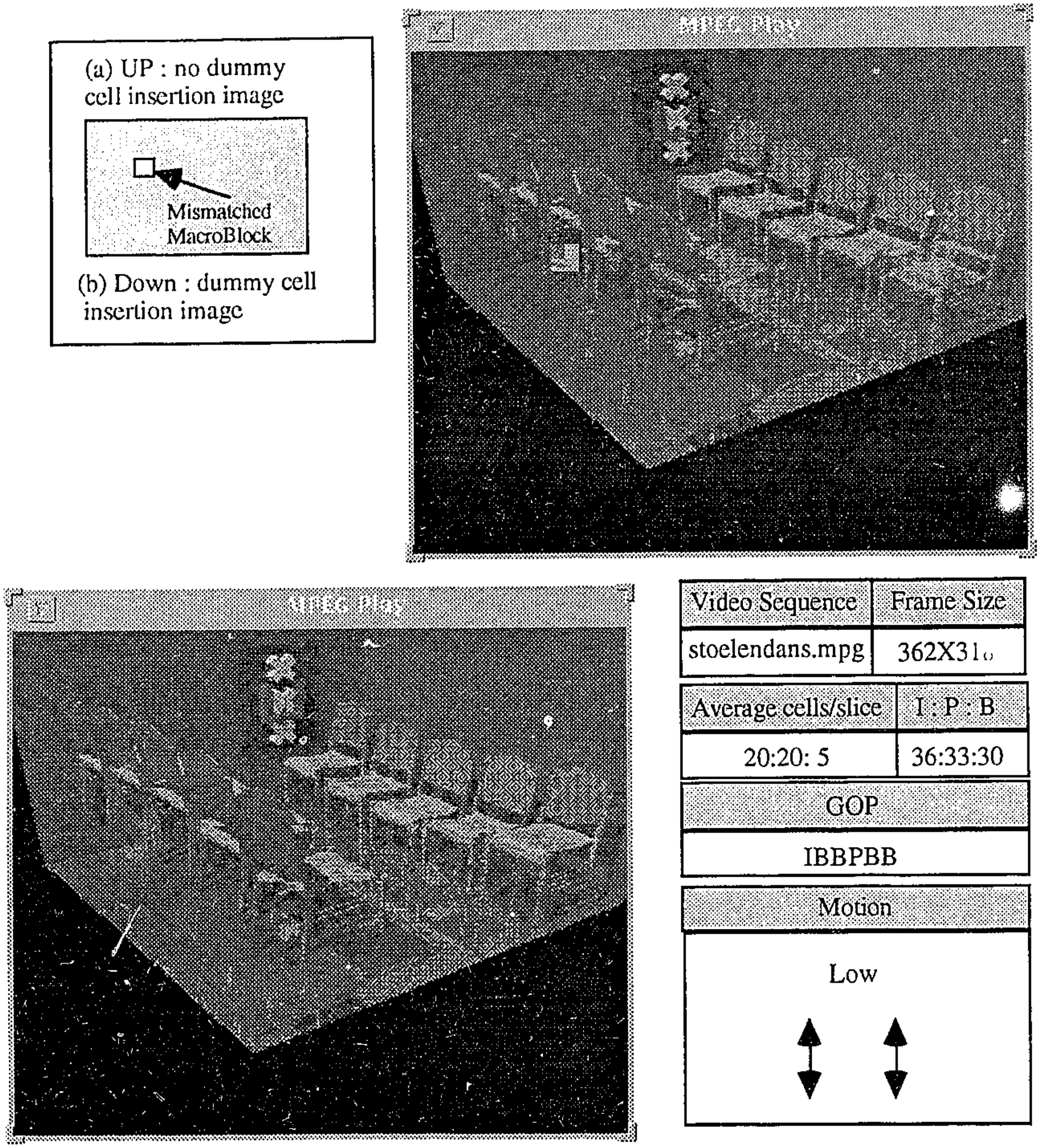


Figure 4.4.12 Mismatched Slices in ***bicycle.mpg***



(a) UP : no dummy cell insertion image

(b) Down : dummy cell insertion image

Video Sequence	Frame Size
stoelendans.mpg	362X316

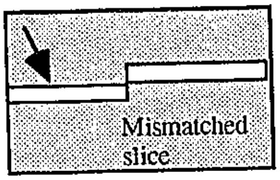
Average cells/slice	I : P : B
20:20: 5	36:33:30

GOP
IBBPBB

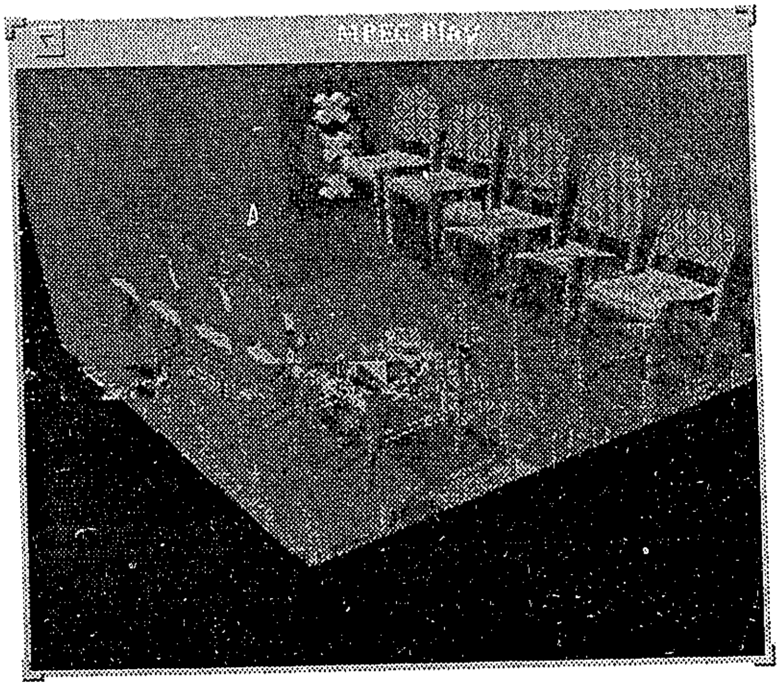
Motion
Low

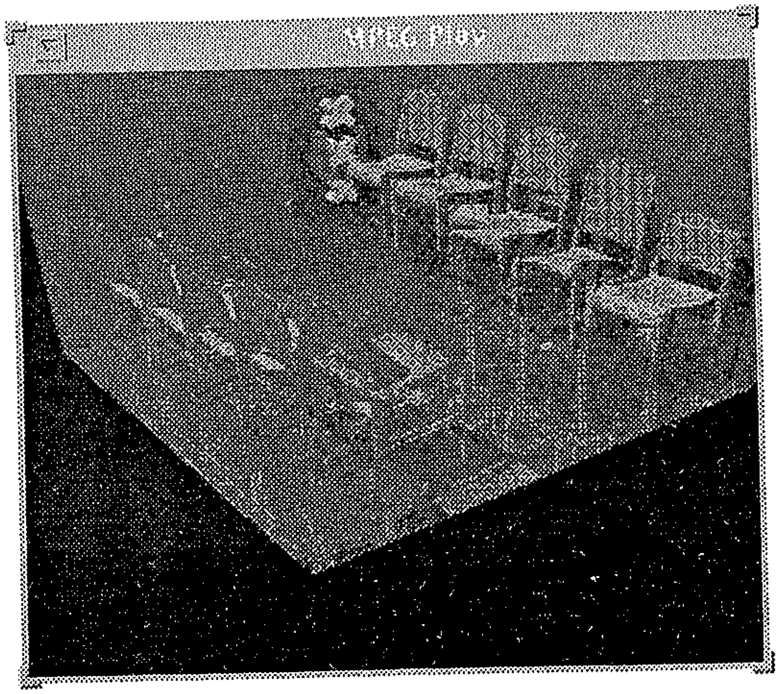
Figure 4.4.13 Mismatched Macroblocks in ***stoelendans.mpg***

(a) UP : no dummy cell insertion image



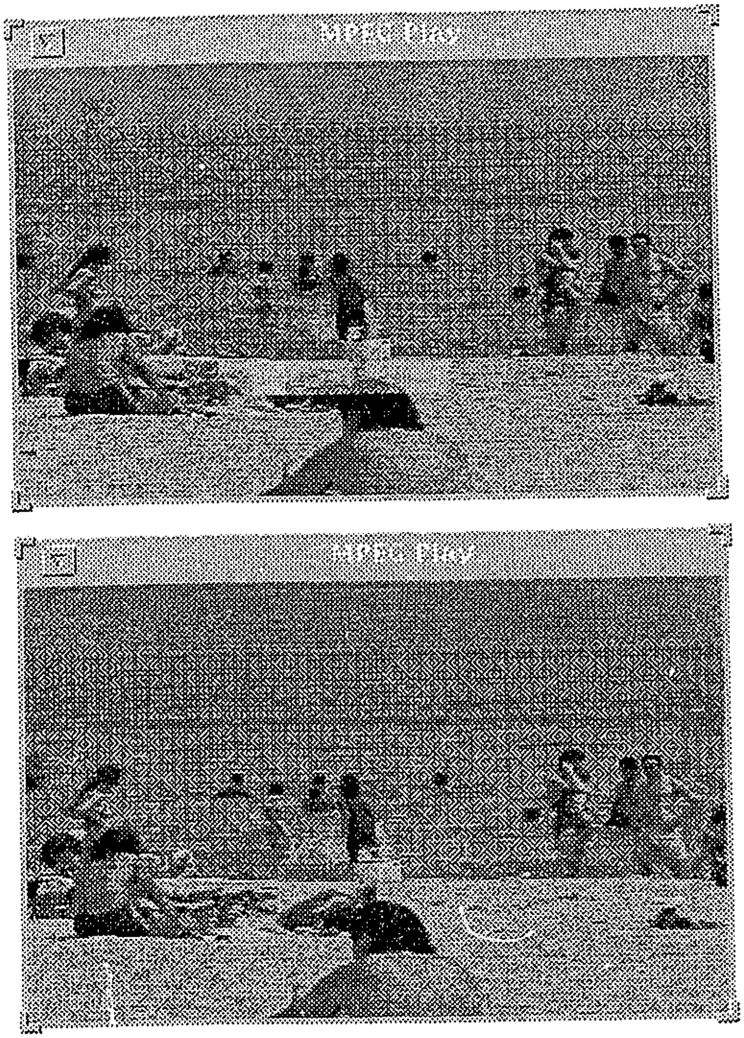
(b) Down : dummy cell insertion image



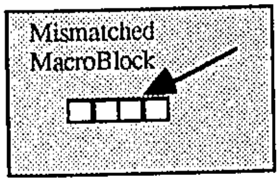


Video Sequence	Frame Size
stoelendans.mpg	362X318
Average cells/slice	I : P : B
20:20: 5	36:33:30
GOP	
IBBPBB	
Motion	
Low ↕ ↕	

Figure 4.4.14 Mismatched Slices in *stoelendans.mpg*



(a) UP : no dummy cell insertion image

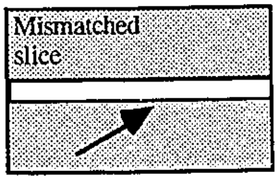


(b) Down : dummy cell insertion image

Video Sequence	Frame Size	Motion
waterski.mpg	346X238	Middle
Average cells/slice	I : P : B	
26:10: 5	26:43:31	
GOP		
IBBPBBPBBPBBPBB		

Figure 4.4.15 Mismatched Macroblocks in ***waterski.mpg***

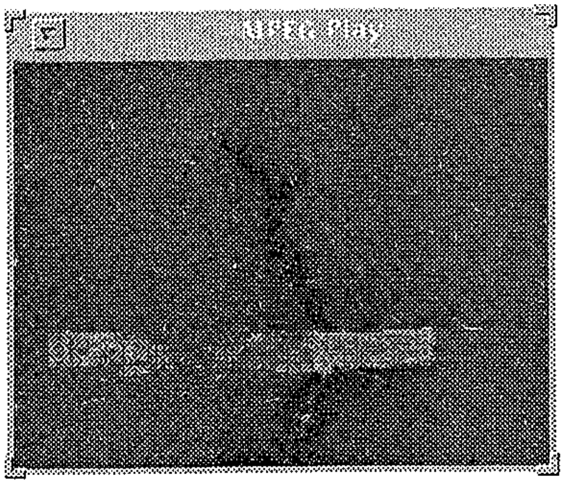
(a) UP : no dummy cell insertion image



(b) Down : dummy cell insertion image

Video Sequence	Frame Size	Motion
waterski.mpg	346X238	Middle
Average cells/slice	I : P : B	
26:10: 5	26:43:31	
GOP		
IBBPBBPBBPBBPBB		

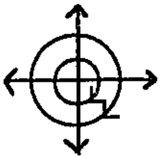
Figure 4.4.16 Mismatched Slices in ***waterski.mpg***



Video Sequence	Frame Size
zoom.mpg	266X222

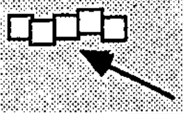
Average cells/slice	I : P : B
10:10:4	31:31:38

GOP
IBBPBB

Motion
High 

(a) UP : no dummy cell insertion image

Propagated impairements by a mismatched slice



(b) Down : dummy cell insertion image

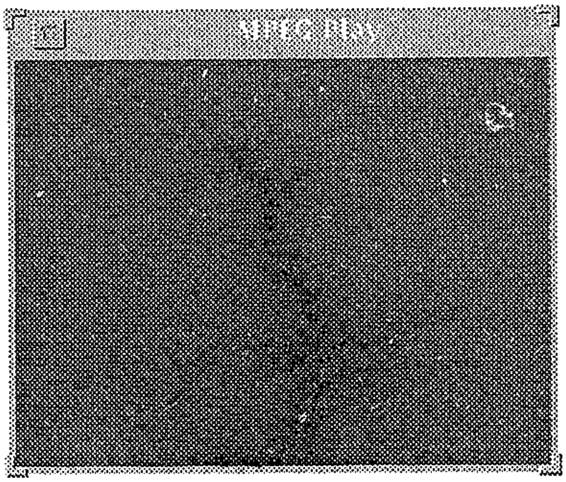


Figure 4.4.17 Mismatched Slices in *zoom.mpg*

Chapter 5

SLICE-SIZE CONTROL SCHEME

5.1 Motivation

5.1.1 The dependency of pictures in GOP

A group of pictures (GOP) is set of pictures which are contiguous in display order. A GOP must contain at least one I-picture. This required picture may be followed by any number of I and P-pictures. Any number of B-pictures may be interspersed between each pair of I or P-pictures, and may also precede the first I-picture. According to the properties of structure of a GOP, some examples of GOP are given in Figure 5.1.1.

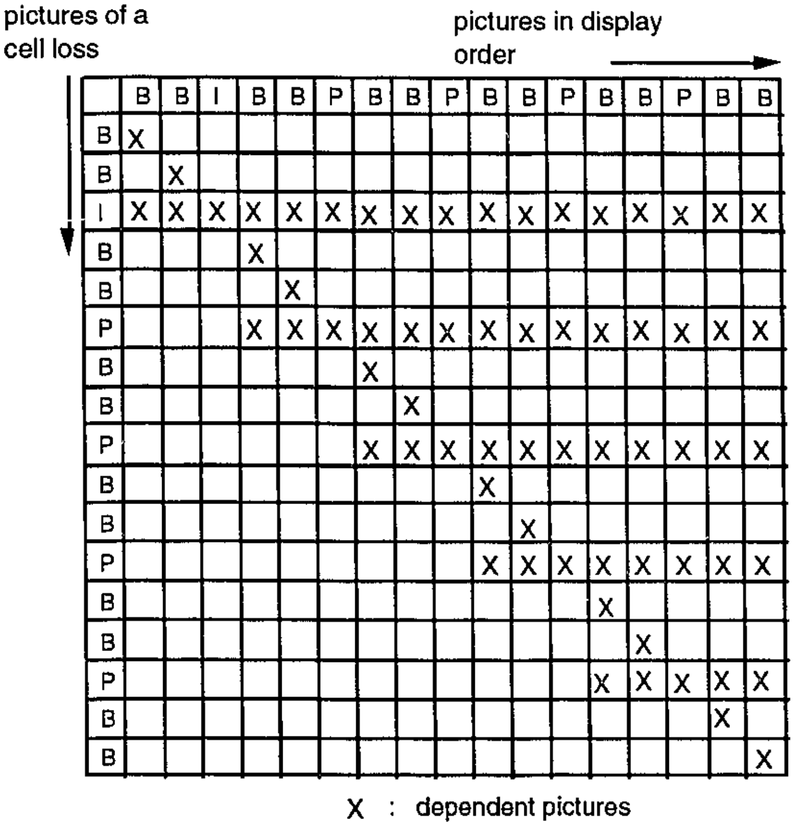
The original concept of a GOP was a set of pictures that could be coded and displayed independently of any other group. In the final version of the International Standard this is not always true, and any B-pictures preceding (in display order) the first

I-picture in a group may require the last picture in the previous group in order to be decoded. Even though encoders can still construct groups of pictures which are independent of one another, we have used the video stream which contains B frames using as a reference the first I-picture in the next group. A group of pictures used for our study contains 15 pictures including 1 I picture, 4 P pictures and 10 B pictures as (6) in figure 5.1.1.

(1) I
(2) I P P
(3) I B P B P
(4) B B I B P B P
(5) B I B B B B P B I B B I I
(6) B B I B B P B B P B B P B B P

Figure 5.1.1 Examples of groups of pictures in display order

Figure 5.1.2 shows the dependency of pictures in a GOP in display order at the decoder. The impairments caused by cell losses are determined by the property of dependency, which is the main factor of propagation of impairments to other pictures in a GOP, according to the picture types and the order of display. In this figure, as a example, the impairments of 17 pictures depend on the cell loss of an I-picture and the impairments of 14, 11, 8 and 5 pictures are respectively affected by the cell loss of P_i-pictures in a GOP (where i = 1, 2, 3, 4). Only B-pictures have no dependent picture which is affected by the cell loss of the picture.



	B	B	I	B	B	P	B	B	P	B	B	P	B	B	P	B	B
B	X																
B		X															
I	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
B				X													
B					X												
P				X	X	X	X	X	X	X	X	X	X	X	X	X	X
B							X										
B								X									
P							X	X	X	X	X	X	X	X	X	X	X
B										X							
B											X						
P										X	X	X	X	X	X	X	X
B													X				
B														X			
P													X	X	X	X	X
B																X	
B																	X

Figure 5.1.2 The dependency of pictures in display order

5.1.2 Coding schemes on each type of frames

• *Coding I-picture* : Each I-picture is divided up into one or more slices, which are, in turn, divided into macroblocks. I-pictures use intra coding. The encoder has two main decisions. These are: how to divide the picture up into slices, and how to set the quantizer scale.

- *Coding P-picture* : As in I-pictures, each P-picture is divided up into one or more slices, which are, in turn, divided into macroblocks. Coding is more complex than for I-pictures, since motion-compensated macroblocks may be constructed. The difference between the motion compensated macroblock and the current macroblock is transformed with a 2-dimensional DCT giving an array of 8 by 8 transform coefficients. The coefficients are quantized to produce a set of quantized coefficients. The quantized coefficients are then encoded using a run-length value technique. In coding P-pictures, the encoder has more decisions to make than in the case of I-pictures. These decisions are: how to divide the picture up into slices, determine the best motion vectors to use, decide whether to code each macroblock as intra or predicted, and how to set the quantizer scale.

Slices are divided into macroblocks in the same way as for I-pictures. The major difference is the complexity introduced by motion compensation. The position of the macroblock is determined by the macroblock address. Whereas the macroblock address increment within a slice for I-pictures is restricted to one, it may be larger for P-pictures. Any macroblocks thus skipped over are called "skipped macroblocks". The decoder copies them from the previous picture into the current picture. Skipped macroblocks are as predicted macroblocks with a zero motion vector for which no additional correction is available. Skipped macroblocks have no VLC code. Instead they are coded by having the macroblock address increment code skip over them.

- *Coding B-picture* : As in I and P-pictures, each B-picture is divided up into one or more slices, which are, in turn, divided into macroblocks. Coding is more complex than for P-pictures, since several types of motion compensated macroblocks may be constructed: forward, backward, and interpolated. The difference between the motion compensated macroblock and the current macroblock is transformed with a 2-dimensional DCT giving an array of 8 by 8 transform coefficients. The coefficients are quantized to produce a set of quantized coefficients. The quantized coefficients are then encoded using a run-length value technique. In coding B-pictures, the encoder has more decisions to make than in the case of P-pictures. These decisions are: how to divide the picture up into slices, determine the best motion vectors to use, decide whether to use forward or backward or interpolated or to code as intra, and how to set the quantizer scale.

Slices are divided into macroblocks in the same way as for I-pictures. The position of the macroblock is determined by the macroblock address. Whereas the macroblock address increment within a slice for I-pictures is restricted to one, it may be larger for B-pictures. Any macroblocks thus skipped over are called "skipped macroblocks". Skipped macroblocks in B-pictures differ from skipped macroblocks in P-pictures. Whereas in P-pictures skipped macroblocks have a motion vector equal to zero, in B-pictures skipped macroblocks have the same motion vector and the same macroblock type as the previous macroblock, which cannot be intra coded.

5.1.3. Cell Loss and Video Quality

As mentioned in the prioritization scheme, the impairments caused by the cell losses very much depend on the contents of the information being conveyed by the missing cells. Therefore, in order to present the quality of video affected in the presence of cell loss, the impairments caused when losing various kinds of pieces of information of an MPEG sequence are illustrated.

Impairments by frame type and order

In order to illustrate the impairments, we have chosen the video sequence which has a GOP, in display order, as follows:

I B B P_1 B B P_2 B B P_3 B B P_4 B B

Figure 5.1.3 shows the impact over the different pictures of a GOP due to the loss of cells carrying information belonging to the I, P1, P2, P3, P4 and B frames, respectively. In this figure, the signal to noise ratio for each one of the frames is shown. This signal to noise ratio has been calculated as done in section 4.1.1. In the particular cases depicted in the figure, a cell has been lost pertaining to I, P1, P2, P3, P4 and B frames respectively. The loss of a cell from I frame affects the I, the consecutive 4 P (P1, P2, P3, P4) and the 10 B frames in the same GOP and the 2 B frames in the previous GOP.

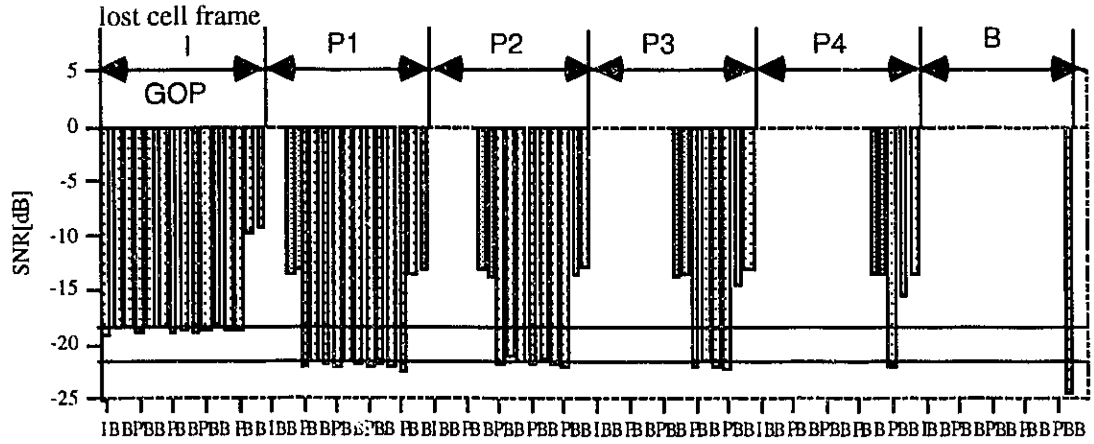


Figure 5.1.3 The propagated impairments caused by a cell loss of frames based on types and order

As expected, in the case of the cell loss from P1 frame the 1 P (P_1), 3 P (P_2, P_3, P_4) frames and 10 B frames are affected but I frame is not as the decoding of the I frames do not depend on the P frames. The cell loss from P_2, P_3 or P_4 frame affects its own frame and the rest depending frames as the cell loss from P1 does. Finally the loss of cell pertaining to a B-frame does not affect any other frame.

5.2 The Slice Data Size Control Scheme

In this control scheme, we control the number of bytes per slice based on the type and order of frames (I, P_1, P_2, P_3, P_4 and B frames). As we have seen from the dummy cell insertion scheme, the loss of a cell may reduce the chances of the decoder to detect the beginning of a new slice causing a propagation of errors. By introducing a dummy cell pattern, we have attempted to make more robust the communication process. The decoder may easily detect the loss of information and resynchronize by finding the next slice header. The objective of controlling the slice size is to explore the use of shorter slice size to the frames on which more pictures depend so that reduce the total impairments propagated by the cell loss of the frame.

5.2.1 Sample MPEG Video Sequences

The MPEG standard provides the bottom line to control the slice size. According to the MPEG standard, the coding structure permits great flexibility in dividing a picture up into slices depending on the environment. If the data are to be used in an error free environment, then one slice per picture may be appropriate. If the environment is noisy, then more slices may be more desirable. Since each slice header requires 40 bits, the use of a smaller slice size may increase the amount of information to be transferred which implies a larger number of cells.

Figures 5.2.1 shows the average data size (number of cells) per slice on each frame in video bit streams which are used as standard sample video sequences. In these video sequences, the average number of cells per slice on I frame is much larger than on P or B frame, and P frame is much larger than B frame. As seen from Figure 5.1.3, the types and order of these frames may affect the quality of the images in the presence of cell losses. In order to reduce this effect, the length of slice on I frames should be shorter than one on P and B frames, and the first coming P frame should be shorter than the later coming P frame, and all P frames should be shorter than B frames.

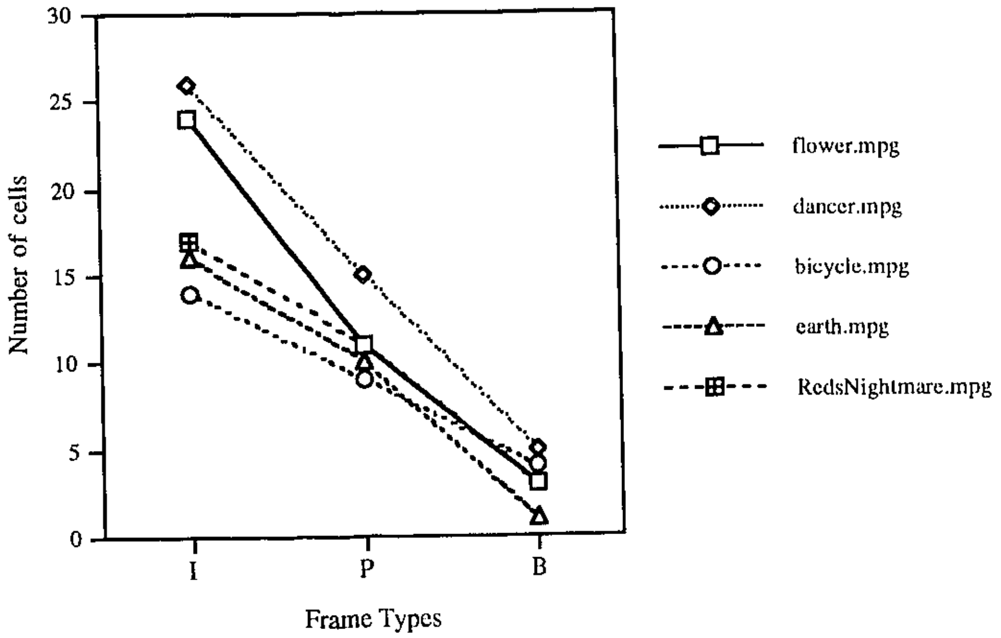


Figure 5.2.1 The average number of cells per slice

Furthermore, in order to prevent an increase in the number of cells, the slice data size of frames such as I or P1 frame on which other frames more depend should be decreased while the slice data size of frames such as P4 or B frame on which other frames less depend are increased. However, when we control the slice size in this method we should notice that the choice on the slice size based on each frame has to take into account the variation of the total amount of impairments occuring in a group of pictures in order to minimize it. This factor will be considered in section 5.3.

5.2.2 Impairments per frame based on a macroblock

In order to investigate the impairments relationship among frames based on a macroblock, we have used a 16 by 16 macroblock as a basic unit since a macroblock is used as a unit for motion representation. By using the GOP mentioned in the section 2.3, Figure 5.2.2 shows the general configuration of the three types of macroblocks (Intra-coded, Inter-coded and Skipped macroblocks) and motion vectors in each type of frame. An I frame comprises only Intra-coded macroblocks and P & B frames comprise all three types of macroblocks.

Let N be the total number of macroblocks in a frame and $N_1(\text{pic})$, $N_2(\text{pic})$, $N_3(\text{pic})$, for pic = I, P and B frames be the number of intra-coded, inter-coded and skipped macroblocks in each type of frame. Here, we assume that each frame of the same type has the same number of macroblocks of each kind, i.e., $N_1(\text{Pi}) = N_1(\text{Pi+1})$, $N_1(\text{Bi}) = N_1(\text{Bi+1})$ and so on, where i = 1, 2, 3 ... Then, the total numbers of macroblocks in I, P and B frames are $N_1(\text{I})$, $N_1(\text{P}) + N_2(\text{P}) + N_3(\text{P})$ and $N_1(\text{B}) + N_2(\text{B}) + N_3(\text{B})$ respectively.

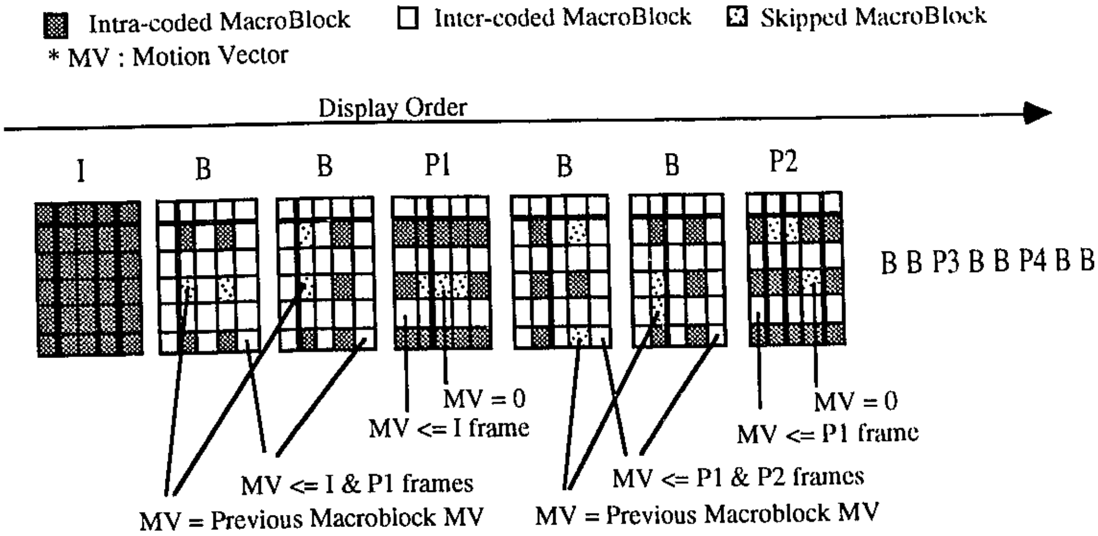


Figure 5.2.2 Configuration of Macroblocks in each type of frames

By considering these compositions, we can evaluate the total number of impaired macroblocks which comprises the macroblock of the lost frame plus the macroblocks affected by the loss of a macroblock. First, for the lost macroblock, we should consider three cases depending on the frame type:

- I frame macroblock loss: Since the macroblocks of an I frame are all intra-coded, any macroblock loss results in a macroblock corrupted. Therefore, when a macroblock is lost, the number of macroblocks corrupted is 1.
- P frame macroblock loss: For skipped macroblocks in P frames, the decoder copies them from the previous I or P frame into the current P frame as predicted macroblocks with a zero motion vector for which no additional correction is available. Thus, as far as it is not corrupted from the previous I and P frame, the skipped macroblock cannot

be lost. In the case of losing an intra-coded or inter-coded macroblock, the number of macroblocks corrupted is 1.

- B frame macroblock loss: In the case of skipped macroblocks in B frames, they differ from skipped macroblocks in P frames. Whereas in P frames skipped macroblocks have a motion vector equal to zero, in B-pictures skipped macroblocks have the same motion vector and the same macroblock type as the previous macroblock so that the skipped macroblock can also be lost if the previous macroblock is lost. Therefore, when a macroblock is lost, the number of macroblocks corrupted is more than 1. That is,

$$1+\frac{N_3(B)}{N-N_3(B)}$$

Note : Even though this ratio in the strict sense should give an integer number, we will consider that it may take value in the real set number as we are interested in evaluating the mean number of macroblocks lost.

Second, we should consider the impairments propagated to other frames. Here, we suppose that in B frames the number of forward inter-coded macroblocks is equal to the number of backward inter-coded macroblocks. This is particular true for long video sequences. Four cases are considered as follows:

1) I to P propagation
2) I to B propagation
3) P to P propagation
4) P to B propagation

The number of macroblocks corrupted in each case are expressed as follows:

1) $1-\frac{N_1(P)}{N}$ 2) $\frac{1}{2}\left(1-\frac{N_1(B)}{N}\right)$ 3) $1-\frac{N_1(P)}{N}$ 4) $\frac{1}{2}\left(1-\frac{N_1(B)}{N}\right)$

These expressions allow us to obtain the total number of macroblocks being corrupted, *Ep*(pic) where pic = I, P and B, which are caused by a macroblock loss at each type of frame. The total impairments are as follows:

- When a macroblock is lost at a B frame,

$$\mathrm{E}\,p(\mathrm{B}) = 1 + \frac{N3(B)}{N - N3(B)} \qquad \text{-(5.1)}$$

- When a macroblock is lost at a P frame,

$$E_p(P_i) = 1 + \left[\frac{1}{2}\left(1 - \frac{N_1(B)}{N}\right) \times B_f(P)\right] + E_p(P_{i+1})$$

$$E_p(P_{i+1}) = \left[1 - \frac{N_1(P)}{N}\right] + \left[\frac{1}{2}\left(1 - \frac{(N_1(B) + N_1(P))}{N}\right) \times B_f(P)\right] + E_p(P_{i+2})$$

$$E_p(P_{i+2}) = \left[1 - \frac{2N_1(P)}{N}\right] + \left[\frac{1}{2}\left(1 - \frac{(N_1(B) + 2N_1(P))}{N}\right) \times B_f(P)\right] + E_p(P_{i+3})$$

o

o

$$E_p(P_k) = \left[1 - \frac{(k-i)N_1(P)}{N}\right] + \left[\frac{1}{2}\left(1 - \frac{(N_1(B) + (k-i)N_1(P))}{N}\right) \times B_f(P)\right]$$

As a result,

$$E_p(P_i) = \sum_{i=l}^{k}\left(1 - \frac{(i-l)N_1(P)}{N} + \frac{1}{2}\left(1 - \frac{N_1(B) + (i-l)N_1(P)}{N}\right) \times B_f(P)\right) \qquad \text{-(5.2)}$$

where k is the total number of P frames in a GOP, 1 is the first P picture number which contains the lost information of a macroblock, and $Bf(P)$ is the number of B frames which use a P frame as a reference.

- When a macroblocks is lost at an I frame,

$$E_p(I) = 1 + \left[\frac{1}{2}\left(1 - \frac{N_1(B)}{N}\right) \times B_f(I)\right] + E_p(P_1)$$

$$E_p(P_1) = \left[1 - \frac{N_1(P)}{N}\right] + \left[\frac{1}{2}\left(1 - \frac{N_1(B) + N_1(P)}{N}\right) \times B_f(P)\right] + E_p(P_2)$$

$$E_p(P_2) = \left[1 - \frac{2N_1(P)}{N}\right] + \left[\frac{1}{2}\left(1 - \frac{N_1(B) + 2N_1(P)}{N}\right) \times B_f(P)\right] + E_p(P_3)$$

o

o

$$E_p(P_k) = \left[1 - \frac{kN_1(P)}{N}\right] + \left[\frac{1}{2}\left(1 - \frac{N_1(B) + kN_1(P)}{N}\right) \times B_f(P)\right]$$

That is,

$$E_p(I) = 1 + \frac{1}{2}\left(1 - \frac{N_1(B)}{N}\right) \times B_f(I) + \sum_{i=1}^{k}\left(1 - \frac{iN_1(P)}{N} + \frac{1}{2}\left(1 - \frac{N_1(B) + iN_1(P)}{N}\right) \times B_f(P)\right)$$

-(5.3)

where k is the total number of P frames in a GOP and *Bf(I)* is the number of B frames which use an I frame as a reference.

5.2.3 Impairments per frame as a result of a cell loss

Let *Mc(*pic*)*, *Cs(*pic) for pic = I, P and B be the number of macroblocks per cell and the number of cells per slice both for each type of frame, respectively. Then, the number of macroblocks lost from a cell loss of each type of frame, *Im*(pic), is given by:

$$Im(pic) = Ep(pic)Mc(pic)$$

If a cell loss occurs in a slice, the total average number of lost cells is

$$Cs(pic)/2$$

Similarly, the number of corrupted macroblocks on a slice loss are

$$Im(pic)Cs(pic)/2$$

Finally, the total number of corrupted macroblocks in the presence of cell losses in a video sequence is expressed as

$$T_c = \mathrm{Im}(I)\frac{Cs(I)}{2} + \mathrm{Im}(P)\frac{Cs(P)}{2} + \mathrm{Im}(B)\frac{Cs(B)}{2} \qquad \text{- (5.4)}$$

Now, let us decide the proper slice data size of each type of frames with the formulas derived above such as to minimize the number of corrupted macroblocks for a fixed total number of cells, i.e., the fixed number of slices, m. If we define the total number of slices in each type of frame as $Sn($pic) and the total number of cells in each type of frame as $Cf($pic), then

$$Cs(pic) = Cf(pic)/Sn(pic) \qquad \text{[cells/slice]} \qquad \text{-(5.5)}$$

From (5.4) and (5.5), we obtain

$$T_C = \mathrm{Im}(I)\frac{Cf(I)}{2S_n(I)} + \mathrm{Im}(P)\frac{Cf(P)}{2S_n(P)} + \mathrm{Im}(B)\frac{Cf(B)}{2S_n(B)} \qquad \text{-(5.6)}$$

and,

$$S_n(I) + S_n(P) + S_n(B) = m \qquad \text{-(5.7)}$$

In order to simplify the Eqn. (5.6) and (5.7), let's denote that:

$$a = Im(I)\frac{Cf(I)}{2}, b = Im(P)\frac{Cf(P)}{2}, c = Im(B)\frac{Cf(B)}{2}$$

$$x = Sn(I), y = Sn(P), z = Sn(B)$$

then Eqn (5.6) and (5.7) can be expressed as,

$$T_C = \frac{a}{x} + \frac{b}{y} + \frac{c}{z} \qquad \text{-(5.8)}$$

$$x + y + z = m \qquad \text{-(5.9)}$$

From (5.8), (5.9) we can get the x, y, z values minimizing $T_C(x,y,z)$. That is,

$$x = \frac{m\sqrt{a}}{\sqrt{a}+\sqrt{b}+\sqrt{c}}, \quad y = \frac{m\sqrt{b}}{\sqrt{a}+\sqrt{b}+\sqrt{c}}, \quad z = \frac{m\sqrt{c}}{\sqrt{a}+\sqrt{b}+\sqrt{c}} \qquad \text{-(5.10)}$$

Proof)

If we suppose x+y equal to k, then

$$Tc(x) = \frac{a}{x} + \frac{b}{k-x} + \frac{c}{m-k} \qquad \text{-(5.11)}$$

The necessary conditions for minimization of *Tc(x)* are obtained by differentiating it with respect to x and equating the results to zero. This gives

$$Tc'(x) = -\frac{a}{x^2} + \frac{b}{(k-x)^2} = 0 \qquad \text{-(5.12)}$$

$$x = \frac{\sqrt{a}}{\sqrt{a}+\sqrt{b}}k,\ \ y = \frac{\sqrt{b}}{\sqrt{a}+\sqrt{b}}k,\ \ z = m-k \qquad \text{-(5.13)}$$

In order to get k, we can substitute 5.13 in 5.11, then

$$Tc(k) = \frac{a\left(\sqrt{a}+\sqrt{b}\right)}{\sqrt{a}k} + \frac{b\left(\sqrt{a}+\sqrt{b}\right)}{\sqrt{b}k} + \frac{c}{m-k}$$

To minimize Tc(*k*), if we differentiate

$$Tc'(k) = -\frac{a\left(\sqrt{a}+\sqrt{b}\right)}{\sqrt{a}k^2} - \frac{b\left(\sqrt{a}+\sqrt{b}\right)}{\sqrt{b}k^2} + \frac{c}{(m-k)^2} = 0$$

$$k = \frac{m\left(\sqrt{a}+\sqrt{b}\right)}{\sqrt{a}+\sqrt{b}+\sqrt{c}}$$

As a result,

$$x = \frac{m\sqrt{a}}{\sqrt{a}+\sqrt{b}+\sqrt{c}},\ \ y = \frac{m\sqrt{b}}{\sqrt{a}+\sqrt{b}+\sqrt{c}},\ \ z = \frac{m\sqrt{c}}{\sqrt{a}+\sqrt{b}+\sqrt{c}}$$

5.3 Experimental Results

5.3.1 Analytical Performance

Table 5.3.1 shows the configuration of the video sequence used in our experiments and the results obtained when the proposed scheme is used. The video sequence used, flower.mpg, contains 330 macroblocks, i.e. $N = 330$. The I frames are only composed of intra-coded macroblocks and the P (P1, P2, P3, P4) and B frames consist almost exclusively of inter-coded macroblocks. The number of macroblocks corrupted by a macroblock information loss, Ep, are gradually decreasing from the I frames to the B frames by steps of approximately 2. The number of macroblocks per cell, Mc, for the B frames are about 7 and 3 times of the ones in the I and P frames respectively. For P frames this number is almost twice the number of an I frame. The number of corrupted macroblocks from a cell loss for I frames are smaller than the one of P1 and P2 frames. This is due to the smaller number of macroblocks of the I frames in a cell, even though the values of Ep are larger. For the same reason, the P4 frames have smaller values of Im than B frames. The number of cells per frame type, Cf, is also important factor to determine the slice size. The number of cells for one I frame is about 2 and 8 times the number of cells of P and B frames, respectively.

frame types	**I**	**P_1**	**P_2**	**P_3**	**P_4**	**B**
no. of frames	10	10	10	10	10	98
Intra MB/frame	330	28.4	28.4	28.4	28.4	4.2
Inter MB/frame	0	293.6	293.6	293.6	293.6	285.4
skipped MB/frame	0	8	8	8	8	40.4
Ep	12.42	10.42	8.2	5.7	3	1.1
Mc	0.92	1.94	1.94	1.94	1.94	6.57
Im	11.43	20.21	15.91	11.06	5.82	7.23
Cf	3591	1655	1655	1655	1655	4321
Sn/frame*	15	15	15	15	15	15
Sn/frame	33.9	30.6	27.1	22.7	16.4	9.4
*Cs**	24	11	11	11	11	3
Cs	10.6	5.4	6.1	7.3	10.1	4.7
Bs	508	259	293	350	485	226

Ep : Error propagated impairments per one macroblock loss.

Mc : number of Macroblocks per cell.

Cf: number of Cells per frame type in a video sequence.

Im : error propagated Impairments per one cell loss.

Sn : number of Slices per frame type in a video sequence.

Cs : number of Cells per slice.

Bs : number of Bytes per slice

* the standard video sequence.

Table 5.3.1 Configuration of the video sequence and results

Figure 5.3.1 shows the number of slices, *Sn*, based on each type of frame in both the original and the slice size controlled video sequences. In the controlled video sequence, in order to keep unchanged the number of cells, the numbers of slices for the B frames are decreased while they are increased for the I and P frames. In this figure, the reason for which only B frames have a decreased number of slices is due to the large number of B frames contained in a GOP, which is 10 and 2.5 times to I and P frames, respectively so that the numbers of slices are enough to increase the number of slices of the other frames.

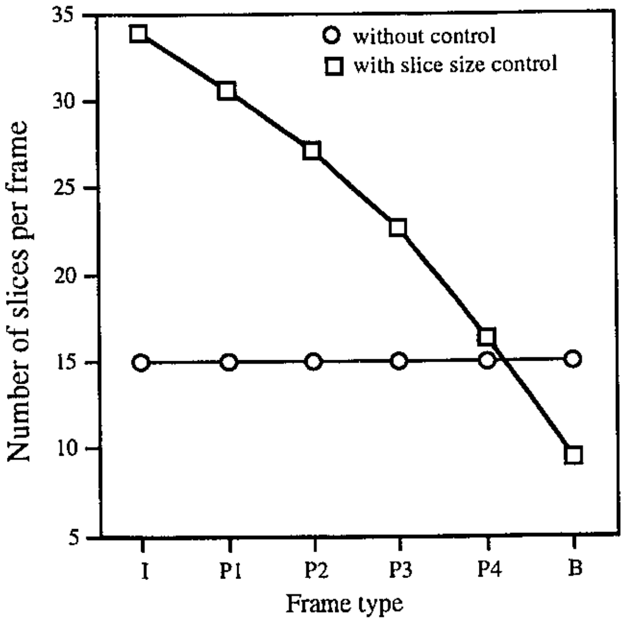


Figure 5.3.1 Number of slices for each type of frame

According to the result, we can evaluate that in order to minimize the impairments in presence of a cell loss of a bitstream exploiting the temporal processing algorithm, an I frame should contain a greater number of slices than the first P frame to be displayed, and that the first P frame should have a greater number of slices than the second P frame to be displayed, and so on.

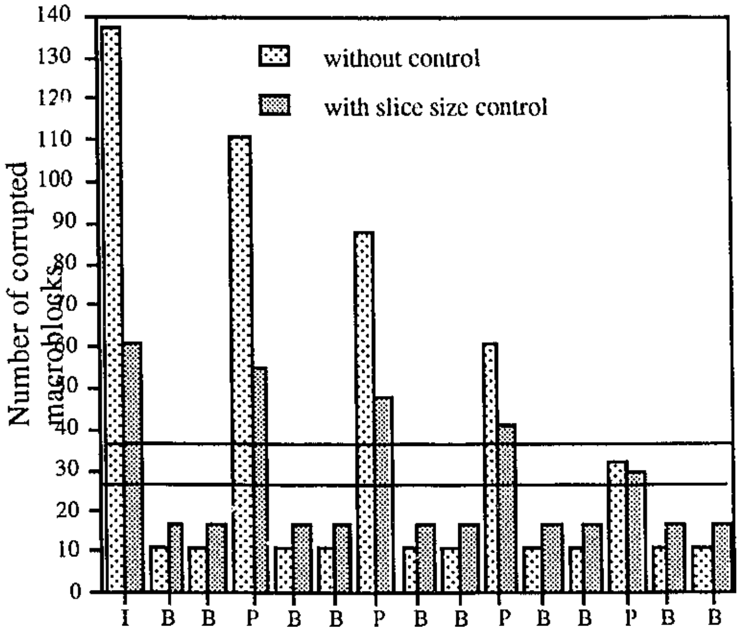


Figure 5.3.2 The number of corrupted macroblocks

In Figure 5.3.2, the numbers of corrupted macroblocks based on a cell loss, i.e. based on the loss of half of a slice, are compared in both the original and the slice size controlled video sequences. The average number of corrupted macroblocks for the controlled sequence is 27, whereas it is 36 for the standard sequence. This implies that the impairment has been decreased by 9 macroblocks (25 %) by exploiting this scheme. In addition, the impairments induced from a cell loss of the I, P1 and P2 frames have been apparently decreased while they have been increased a little for the B frames on which the differences are not easily distinguishable.

5.3.2 Experimental Performance

For the experimental simulation of this study, the slice data size of the I and P frames is decreased while the slice size of the B frames has been increased. One of our objectives is to keep the total number of slices unchanged, i.e., the total number of ATM cells should also be left unchanged. Figure 5.3.3 depicts the impact over the quality of the different frames as a function of the maximum slice data size and in the presence of losses in the different frames. N represents the maximum data size of a slice of I and P frames while M is the maximum data size of B frames, all given in bytes. As shown in this figure, the impairments of the loss of a cell on I, P are largely reduced when a shorter slice size is used, while the quality of the B frames is not heavily affected.

Figure 5.3.4 shows SNR as a function of the CLR for different slice data sizes. This figure illustrates that a better quality video can be obtained by controlling the slice data size and the number of slices on each type of frames. It is also important to notice that the total number of ATM cells needed to carry the video stream has been kept constant.

Figures 5.3.5 and 5.3.6 show an example when two different slice sizes, N=400 (M=40) and N=100 (M=130), are used. Figure 5.3.5 illustrates the case when an ATM cell carrying the 400 bytes length of slices of an I frame is lost. Figure 5.3.5a shows that most of the upper part of the image is lost and both the P and B frames are also much affected (Figure 5.3.5b and c). Figure 5.3.6a shows the case when the ATM cell carrying the 100 bytes length of slices of an I frame is lost. In this second case, the loss of information belonging to the I frame is limited which in turn also reduces the effect over the I an P frames (Figures 5.3.6b and c).

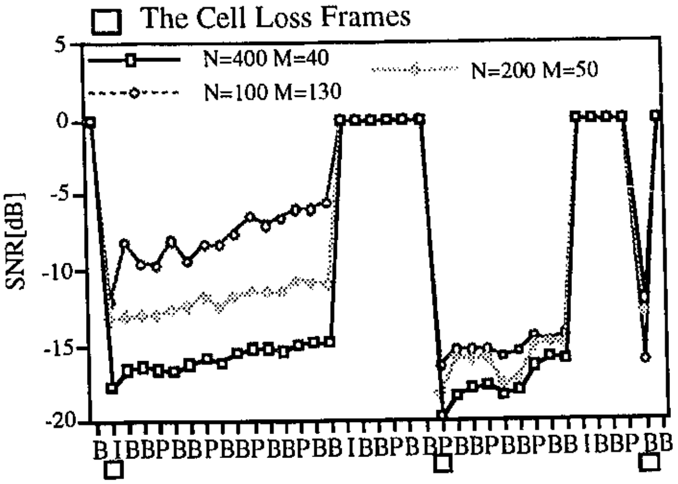


Figure 5.3.3 Impairment due to the loss of each frames cell with variable slice size

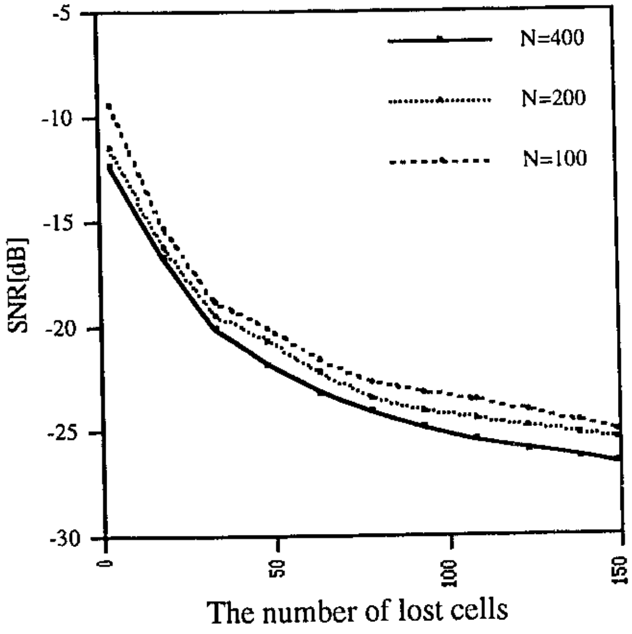
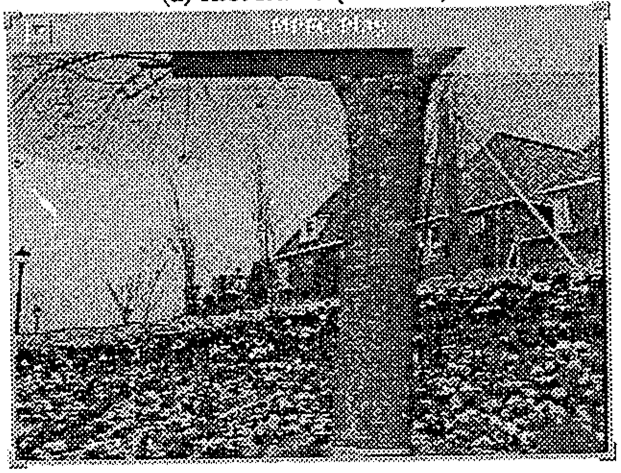
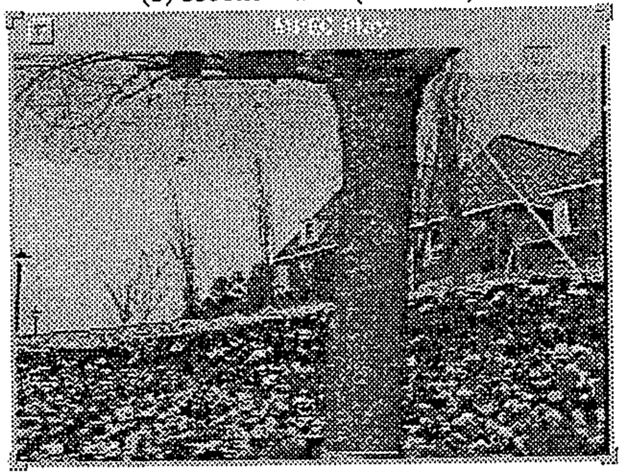


Figure 5.3.4 SNR vs CLR on Maximum Slice Data Size

(a) first frame (I frame)



(b) second frame (P frame)



(c) third frame (B frame)

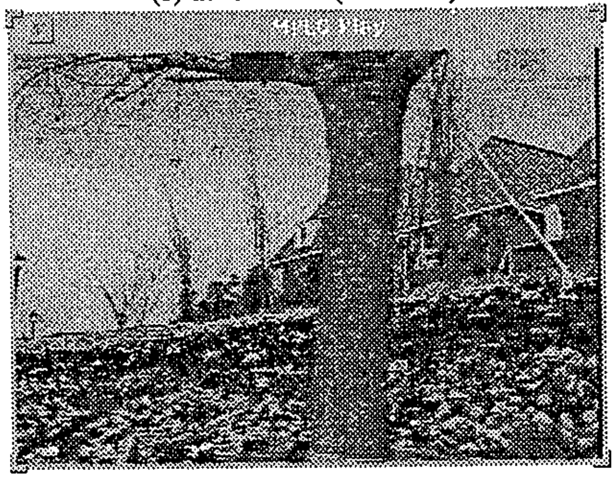
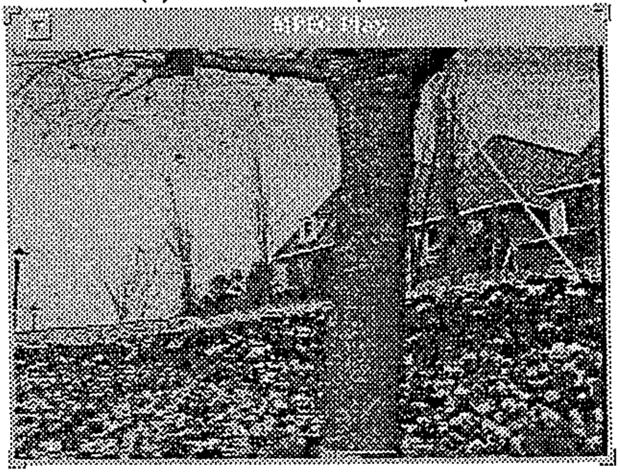


Figure 5.3.5 Impairments Propagation in Sequential Images when N=400

(a) first frame (I frame)

(b) second frame (P frame)



(c) third frame (B frame)

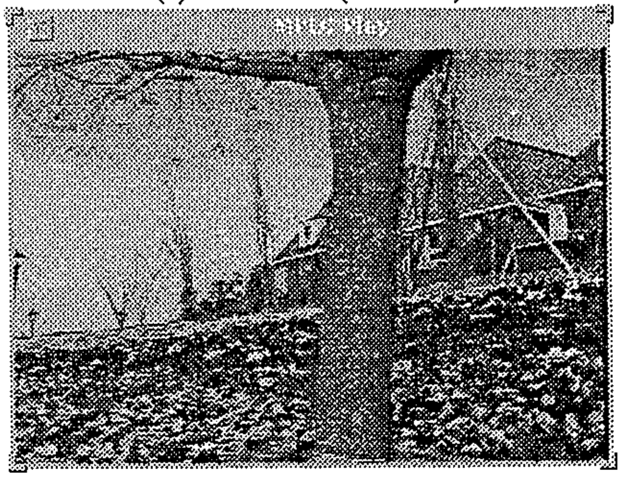


Figure 5.3.6 Impairments Propagation in Sequential Images when N=100

5.4 Conclusion

The MPEG standard utilizes a slice layer for synchronization and provides the bottom line to control the slice size. This feature is particular useful in noisy environments when video sequences may get corrupted. However, smaller slice size will introduce more overhead(slice headers), which in turn increases the amount of data to be transferred. If we consider that a cell loss occurs mostly from congestion from various resources using the same virtual path, it should be prevented to use more slices which can invoke more congestion. In this slice size control scheme, therefore, the study has been focused on improving the performance while maintaining to a minimum the number of cells to be sent. By applying this scheme to the original sample video sequence, we have got a 25 % better video quality, based on the unit of corrupted macroblocks, at the same cell loss error rate without increasing the amount of information to be sent.

Chapter 6

CONCLUSIONS AND FUTURE RESEARCH

6.1 Conclusions

In this thesis, we have investigated the use of three schemes to overcome the problems of losing cells when transferring a MPEG video sequence through an ATM network. From these experimental results we have also seen that if no proper action is taken, the loss of information often causes the decoding and display of the video sequence to stop.

The first scheme, the prioritization scheme, aims to direct the network on the actions to take when congestion occurs. We should notice that this scheme has been implemented by the simple algorithm of prioritization and without causing an increase in the number of cells. For prioritization, we explored the property of the temporal processing of MPEG coding. The results show that the SNR at the same CLR can much improve when the high priority is given to the frames used as references to other frames.

The second scheme, the insertion of dummy cells, aims to make the system more robust in the presence of information loss. Our results show that the use of dummy cells is very useful in providing the decoder with the means of detecting the loss of information. Furthermore, our results have also shown that the bit pattern of the dummy cell has to be chosen carefully. We have been able to determine the bit pattern providing the best results. The dummy cell scheme can be implemented by using the AAL1 which includes the sequence number.

In the third scheme, the slice-size control scheme, the slice-sizes have been controlled based on the frame types and the order of frames in GOP in order to limit the error propagation. The MPEG standard provides the bottom line to control the slice-size but it results in increasing the amount of information (the number of cells) to be transferred if the environment is noisy. This scheme can minimize the total impairments occurred under the same environment. The results have shown the effectiveness of this scheme, i.e., less error propagation for the same percentage ATM cell losses.

6.2 Future Research

We need to compare the prioritization scheme suggested in this work and the multilayer coding scheme already suggested on the basis of the increment of the amount of information to be transferred and the SNR improvement. The way of implementation in a real ATM system and the performance requirement should be studied.

To specify the AAL type for MPEG video is also necessary. Even though AAL 1 contains the sequence number which can be used for the dummy cell insertion scheme, it has been suggested for constant bit rate video. The study of a new AAL type containing the idea of the dummy cell is suggested as a future work.

In the slice-size control scheme, we have experimented only with MPEG 1 video sequence, CBR (Constant Bit Rate). In the case of variable bit rate video, we need to statistically analyze the MPEG video sequence to predict the characteristics of next GOP for implementation.

Appendix A

Resource Management in ATM Networks

The concepts of a VP and VC let us remove such physical concepts as the digital path hierarchy and the fixed-bit-rate channel from the path and channel management, and they provide us flexible network management. The change of the VP/VC bandwidth does not require complicated controls in physical equipment and can be easily performed. This situation differs from the digital path/channel control in a STM network. However, the flexibility attained in ATM networks raises challenging new issues on resource management [CTC92]. Preventing the network from becoming congested, achieving network performance objectives, and optimizing the use of network resources requires the following traffic controls: CAC (connection admission control), UPC, NPC, cell-level quality control (= priority control), and congestion control (Figure A.1).

CAC is defined as the set of actions, executed at connection set-up, taken to decide whether or not the VCC (Virtual Channel Connection) or VPC (Virtual Path Connection) requesting establishment can be accepted. This decision is based on the connection's anticipated traffic characteristics, the requested quality of service, and the current network load. The connection's anticipated traffic characteristics are described by a source traffic descriptor, and a user notifies source traffic descriptor values to the network when the connection is set up. If the request is accepted, network resources (such as bandwidth) are implicitly allocated to the connection.

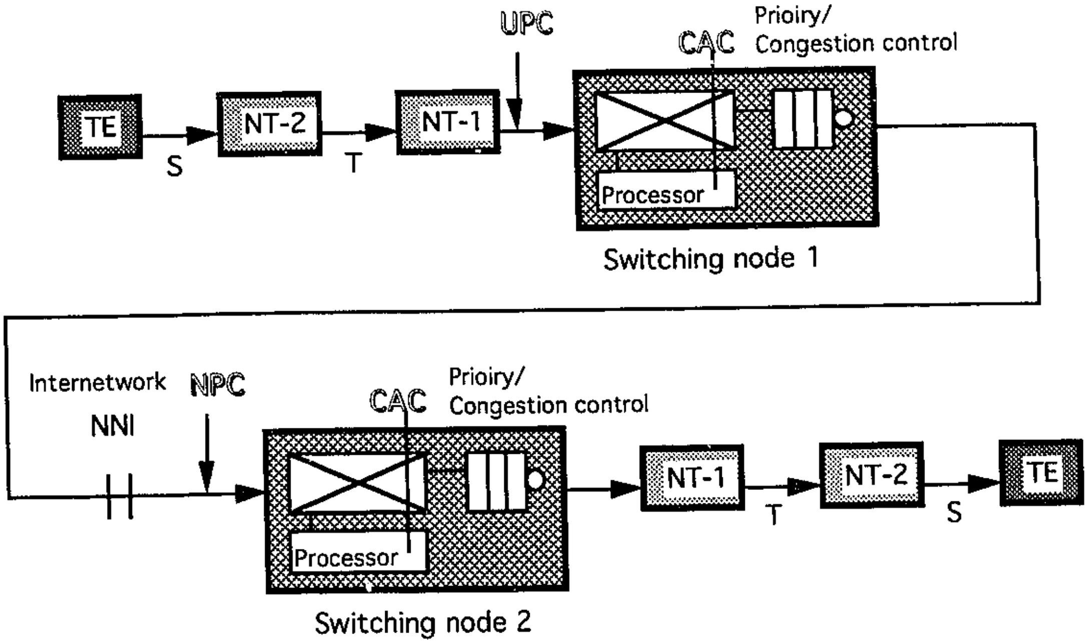


Figure A.1 Reference model of traffic control

The network decides admission under the assumption that source traffic descriptor values are correct; if they are not, the network may become congested. To prevent this, UPC and NPC are provided. These controls monitor the traffic at a UNI (UPC) and at an NNI (NPC), and to assure conformity between the monitored traffic and the anticipated traffic, they adjust the actual traffic by discarding or tagging cells. These parameter controls can exclude malicious cell streams from fair cell streams, thereby protecting the network. Ordinarily, UPC is performed for each traffic parameter in a source traffic descriptor. For example, if the source traffic descriptor consists of the peak cell rate and the average cell rate, UPC is necessary for both of them. Priority control in the ATM layer refers to the time and space priority assigned to a cell implicitly and explicitly and to the actions taken according to the priority. Congestion control, the set of actions reducing the spread and duration of congestion, includes selectively discarding cells. The functions of these controls and such related topics as the traffic parameters are described.

A.1 Traffic source characterization parameters

An ATM network offers services that have diverse traffic characteristics, not all of which require the same QOS. The network must support these services economically and with assurance that QOS objectives are satisfied. The traffic characteristics of an ATM connection are described by such parameters like peak cell rate (**P**), average cell rate (**m**), burstness (β), peak duration (**ton**), and source type.

P : the peak arrival rate of the cells when the source is at the active state (peak rate), or the maximum amount of network resource requested by the source. Alternatively, this parameter may be defined as the reciprocal of the minimum interarrival time between two consecutive cells belonging to the same connection. The above definition is called *Instantaneous Peak Cell Rate* by the CCITT [CTP90]. In the same reference, *Integrated Peak Cell Rate* is defined as the number of cells belonging to the same connection measured during a predefined short time interval T, divided by T. In the case where an ATM source consists of two types of cells with respect to their priority, then the peak arrival rate of the high priority cells is another traffic characterization parameter of that particular ATM source.

m : average cell arrival rate, or the average amount of network resource requested by the source [CDT90]. Two parameters are discussed by the CCITT that can characterize the average cell rate: the *True Average Cell Rate* - the number of cells measured during the duration of a connection, divided by this duration; the *Estimated*

Average Cell Rate - the number of cells measured during a long time interval T, divided by T. In the case where an ATM source generates two types of cells with respect to their priority, the average cell arrival rate of the high priority cells is another traffic characterization parameter of the ATM source.

β : burstiness, defined as the ratio between the peak cell rate and the average cell rate (β = p/m), and can be viewed as a measure of the duration of the activity period of a connection.

ton : the average duration of the active state (peak duration).

The above traffic characterization parameters, or a subset of them, are used in the operation of important network functions such as the admission control, usage parameter control and resource allocation.

A source traffic descriptor is the set of traffic parameters specifying the intrinsic traffic characteristics of the connection at the time a connection is set up. A traffic parameter contributing to a source traffic descriptor should be understandable by the user, of significant use in resource allocation, and enforceable by the network provider through UPC and NPC [CBG90].

A.2 Usage Parameter Control and Network Parameter Control

UPC and NPC have similar functions but at different interfaces: UPC is done at the user-network interface; and NPC is done at the internetwork NNI. These parameter controls are performed for each VC or VP, and their functions include monitoring cell streams, checking the conformity between the actual cell stream and the nominal cell stream (specified by the traffic descriptor values) and taking necessary action when disconformity is detected.

Violation cells may be either discarded immediately or tagged for discard at a congestion point when the network is congested. The CLP bit may be used for this tagging. In this case, the stream of cells with CLP= 0 is checked first and the violation cells are tagged by setting the CLP bit to 1. Then the stream of all the cells (irrespective of CLP value) is checked and the violation cells are discarded at UPC [CTC92]. The cells with CLP =1 are discarded when congestion occurs, although we cannot distinguish between tagged cells and cells whose CLP bit was originally set to 1.

UPC and NPC may make detection errors mainly because of the CDV at NT-2 prior to the UPC point and because of CDV in the network prior to the NPC point. The error that occurs when UPC and NPC detect violation for a cell stream conforming to the traffic descriptor is part of the edge-to-edge network QOS degradation [CTC92].

A.3 Priority Control

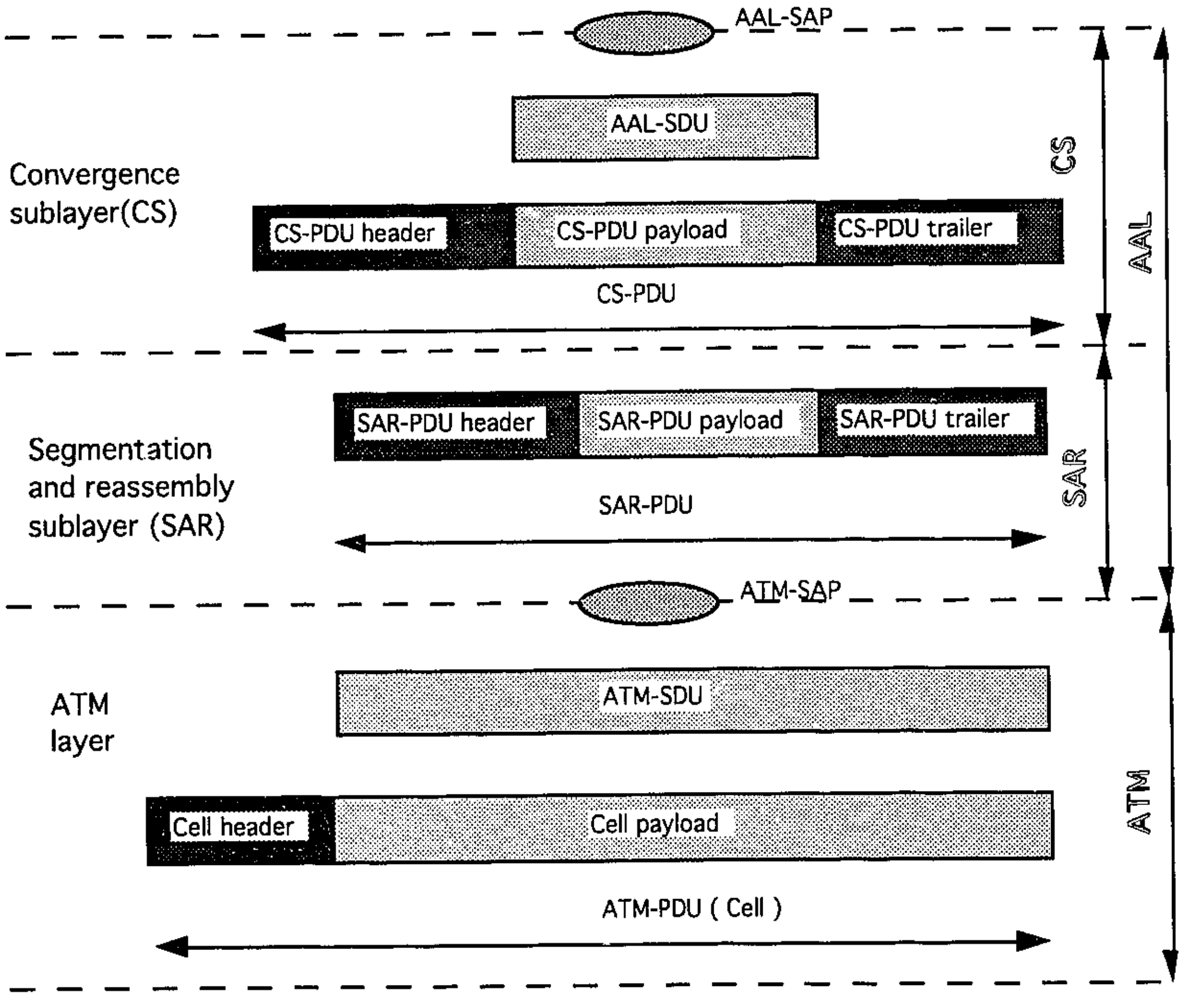
ATM networks support heterogeneous QOS requirements and may introduce priorities to cope with such requirements effectively. There are two ways to express the priorities of a cell. One is to express priorities implicitly by VP/VCI. At the connection setup, the user specifies a QOS class (or QOS requirement) as a part of traffic contract [CTC92]. Thus the VPI/VCI values given at the connection setup implicitly refer to QOS class of the connection. If the network introduces higher priorities for the stringent QOS class, the VPI and VCI values implicitly show these priorities. This implicit designation of priority can be applied to the time priority (that is, the priority for cell queuing delay) as well as the space priority (the priority for cell loss because of buffer overflow). With this implicit designation, all the cells in a connection have the same priority.

Another way to express cell priority is by using the CLP bit: the space priority can be given explicitly by this bit. This explicit designation enables different cells in the same VC connection to have different priorities. The compatibility of the explicit and implicit designations is controversial: it has not been concluded whether the priority designation using the CLP bit is still valid when the individual QOS class is specified by VPI and VCI. Although a cell loss ratio objective is given to the stream of cells with CLP = 0, it is controversial whether a cell loss ratio objective should also be given to the stream of cells with CLP = 1 or to all cells irrespective of CLP values. And of course no conclusion has been reached on such related topics as whether the cell loss ratio objective for cells with CLP = 0 should be the same as the objective for high-priority VC connections.

In any case, some buffers may be reserved for cells given a high-priority cell loss ratio, or these cells may be given the right to push low-priority cells out of buffers and enter those buffers. The high-priority cells are thus relatively unlikely to be lost because of buffer overflow.

Appendix B

Details of the data unit naming conversion



NOTES

The figure is to indicate the naming of the AAL data unit only.

Figure B.1 General data unit naming conversions

Bibliography

[ACV92] Amy R. Reibman, Barry G. Haskell, "Constraints on Variable Bit-Rate Video for ATM Networks," IEEE Transaction on Circuits and Systems for Video Technology, vol. 2, No. 4, December 1992.

[AFD89] A. K. Jain, Fundamentals of Digital Image Processing, 1989, Prentice-Hall, Inc. New Jersey.

[AQS91] Anagnostou, M E, Theologou, M E, Vlakos, K M, Tournis, D and Protonotarios, E N "Quality of Service requirements in ATM based B-ISDN," Comput. Commun. Vol 14 no 4 (May 1991) pp 197-204.

[CBF90] CCITT Recommendation I.150, B-ISDN ATM Functional Characteristics, CCITT, Geneva, Switzerland (May 1990)

[CBG90] CCITT Recommendation I.311, B-ISDN General Network Aspects, CCITT, Geneva, Switzerland (May 1990)

[CBI90] CCITT Recommendation I.121, Broadband Aspects of ISDN, CCITT, Geneva, Switzerland (May 1990).

[CBL90] CCITT Recommendation I.361, B-ISDN ATM Layer Specification, CCITT, Geneva, Switzerland (May 1990)

[CBP90] CCITT Recommendation I.321, B-ISDN Protocol Reference Model and its Application, CCITT, Geneva, 1990 (May 1990)

[CBS92] CCITT Draft Recommendation I.363, B-ISDN ATM adaptation layer (AAL) specification, Study Group XVIII, Geneva, July 1992.

[CBU92] CCITT SGX VIII - Report R119, Draft Recommendation I.432, "B-ISDN User-Network Interface - Physical Layer Specification," June 1992.

[CCB91] C. J. O'Neill, "Comparison of Broadband ISDN Congestion Control Methods," Proc. Australian Broadband Switching and Services Symposium '91, July 1991, Sydney, Australia, vol. 1.

[CDS91] CCITT SG XV, Working Party XV/1, Experts Group for ATM Video Coding, "Draft status report on ATM video coding standardization," Temp. Document No. 3, CCITT SG XV(Paris), May 23, 1991.

[CDT90] CCITT Study Group XVIII Contribution D.941 Definition of Traffic Parameters, CCITT, Geneva, Switzerland, Nov. 1990.

[CTC92] CCITT Draft Recommendation I.371, "Traffic Control and Congestion Control in B-ISDN," Geneva, 1992.

[CTP90] CCITT Study Group XVIII Temporary Document 43, Traffic parameters and Descriptors, CCITT, Geneva, Switzerland, Nov - Dec. 1990.

[DBA90] Draft Recommendation I.121: Broadband Aspects of ISDN, CCITT, Working Party XVIII, SWP8, Jan. 1990.

[DBF90] Draft Recommendation I.150, B-SIDN ATM Functional Characteristics, CCITT SG XVIII, Report R34, June 1990.

[DMV91] D. Le Gall, "MPEG: A video Compression Standard for Multimedia Applications," Communications of the ACM, vol. 34, no. 4, April 1991, pp. 305-313

[DPT90] Duhamel P, Guillemot C, "Polynomial transform computation of the 2-D DCT," IEEE International Conference on Acoustic, Speech and Signal Processing, Albuquerque, N.M., pp 1515-1518, 1990.

[FFS90] Feig E, "A fast scaled DCT algorithm," In: Pennington KS, Moorhead RJ II (des) Image Processing Algorithms and Techniques, Proc. SPIE 1244:2-13, Santa Clara, Calif, 1990.

[GPV89] G. Karlsson, M. Vetterli, "Packet Video and Its Integration into the Network Architecture," IEEE Journal on Selected Area in Communications, vol. 7. No. 5, June 1989

[GSC87] G. Karlsson and M. Vetterli, "Sub-band coding of video signals for packet switched networks," in Proc. SPIE Visual Commun. Image Processing II, vol. 845, Oct. 1987, pp. 446-456.

{GSQ89] G. Hebuterne and A. Gravey, "A space-priority queuing mechanism for multiplexing ATM channels," ITC Specialist Seminar, Adelaide, September 1989.

[GTV91] G. Morrison and D. Beaumont, "Two-layer video coding for ATM networks," Signal Processing: Image Commun., vol. 3, no. 2-3, pp. 129-141, June 1991.

[HFR87] Hou HS, "A fast recursive algorithm for computing the discrete cosine transform," IEEE Trans Acoustics Speech Signal Processing 35:1455-1461, 1987.

[IBF93] ITU-T Recommendation I.362, "B-ISDN ATM Adaptation Layer (AAL) Functional Description," March 1993.

[IBS93] ITU-T Recommendation I.363, "B-ISDN ATM Adaptation Layer (AAL) Specification," March 1993.

[JAT92] J. Le Boudec, "The Asynchronous Transfer Mode: a tutorial," Computer Networks and ISDN Systems 24, pp. 279-309, 1992.

[JES90] J. Le Boudec, "An efficient solution method for Markov models of ATM links with loss priorities," Research Report RZ 2001, IBM, July 1990.

[KQM90] Kroner, H K, Theimer, T H and Briem, U "Queuing models for ATM systems: a-comparison," 7th ITC Seminar, New Jersey, USA (October 1990)

[KVB89] K. Shimamura, Y. Hayashi, and F. Kishino, "Variable bit-rate coding capable of compensating for packet loss," Proc. SPIE Visual Commun. Image Processing, vol. 1001, Nov. 1989, pp. 991-998.

[LND91] Linzer EN, Feig E, "New DCT and scaled DCT algorithms for fused multiply/add architectures," Proc. IEEE International Conference on Acoustic, Speech and Signal Processing, Toronto, Canada, pp 1515-1518, 1991.

[MBC89] M. Nomura, T. Fuji, and N. Ohta, "Basic characteristics of variable rate video coding in ATM environment," IEEE J. Select. Areas Commun., vol. 7, pp. 752-760, June 1989.

[MCC91] M. Garrett and M. Vetterli, Congestion control strategies for packet video. In Proc. Fourth

[MCC93] M. Ghanbari and V. Seferidis, "Cell-Loss Concealment in ATM Video Codecs," IEEE Trans. Circuits and Systems For Video Technology, vol. 3, no 3, June 1993, pp. 238-247.

[MCI92] M. E. Anagnostou, M. E. Theologou, and E. N. Protonotarios, " Cell Insertion Ratio Analysis in Asynchronous Transfer Mode Networks," Computer Networks and ISDN Systems, vol. 24, p.335. 1992

[MCM93] MPEG ISO IEC JTC 1, Information Technology - Coding of moving pictures and associated audio for digital storage media up to about 1.5Mbit/s. ISO/IEC IS 11172, 1993

[MCP93] MPEG ISO IEC JTC 1, Information Technology - Coding of moving pictures and associated audio for digital storage media, test model 4. Draft, MPEG 93/255b, 1993

[MPC93] Mohammad Ghanbari and Charles J. Hughes, "Packing Coded Video Signals into ATM Cells," IEEE ACM Transactions on Networking, vol 1. No 5, October 1993.

[MTB88] M. Tanimoto, N. Chiba, H. Yasui, and M. Murakami, "TAT (time-axis transform) bandwidth compression system of picture signals," IEEE Trans. Commun., vol. 36, pp.347-354, Mar. 1988.

[MTC89] M. Ghanbari, "Two-layer coding of video signals for VBR networks," IEEE J. Select. Areas Commun., vol 7, no. 5, pp. 771-781, June 1989.

[MUB93] MPEG Decoder, University of Berkeley, Internet.

[PLT89] P. Haskel, K. H. Tzou, and T. R. Hsing, "A lapped-orthogonal transform based variable bit rate video coder for packet networks," in Proc. ICASSP'89, Paper M10.2, May 1989, pp. 1905-1908.

[PMC91] Puri A, Aravind R, "Motion compensated video coding with adaptive perceptual quantization," IEEE Trans Circuits Syst Video Technol 1:351, 1991.

[PMC94] P. Pancha and M. El Zarki, "MPEG Coding For Variable Bit Rate Video Transmission," IEEE Communications Magazine, vol. 32, no. 5, May 1994, pp. 54-66.

[QCC93] Q. Zhu, Y. Wang, and L. Shaw, "Coding and Cell-Loss Recovery in DCT-Based Packet Video," IEEE Trans. Circuits and Systems For Video Technology, vol. 3, no 3, June 1993, pp. 248-258.

[RGI89] R. Kishimono, Y. Ogata, and F. Inumaru, "Generation Interval Distribution Characteristics of Packetized Variable Rate Video Coding Data Streams in an ATM Network," IEEE Journal on Selected Area in Communications. vol. 7, No. 5, June 1989.

[STV91] S. Tubaro, "A two-layer video coding scheme for ATM networks," Signal Processing: Image Commun., vol. 3, no. 2-3, pp. 179-195, June 1991.

[SUA90] S. D'Agostino, J. David, R. Heib, and W. Verbiest, "Universal ATM video coding architecture," Electrical Commun., vol. 64, no. 2/3, pp. 186-193, 1990.

[TVH93] T. Kinoshita, T. Nakahashi, and M. Maruyama, "Variable-Bit-Rate HDTV CODEC with ATM-Cell-Loss Compensation," IEEE Trans. Circuits and Systems For Video Technology, vol. 3, no 3, June 1993, pp. 230-237.

[WDC94] W. Stallings, Data and Computer Communications, Macmillan, 4th Edition, 1994, New York.

[WIA88] W. Verbiest, L. Pinnoo, and B. Voeten, "The impact of the ATM concepts on video coding," IEEE J. Select. Areas Commun., vol 6, pp 1623-1632, Dec. 1988.

[WIB92] W. Stallings, ISDN and Broadband ISDN, MacMillan, 2nd. Edition, 1992, New York.

[YVB91] Y. Zhang, W. W. Wu, K.S. Kim, R. L. Pickholtz and J. Ramasastry, "Variable bit-rate video in the broadband ISDN environment,' IEEE, vol. 79, pp. 214-222, Feb. 1991.