

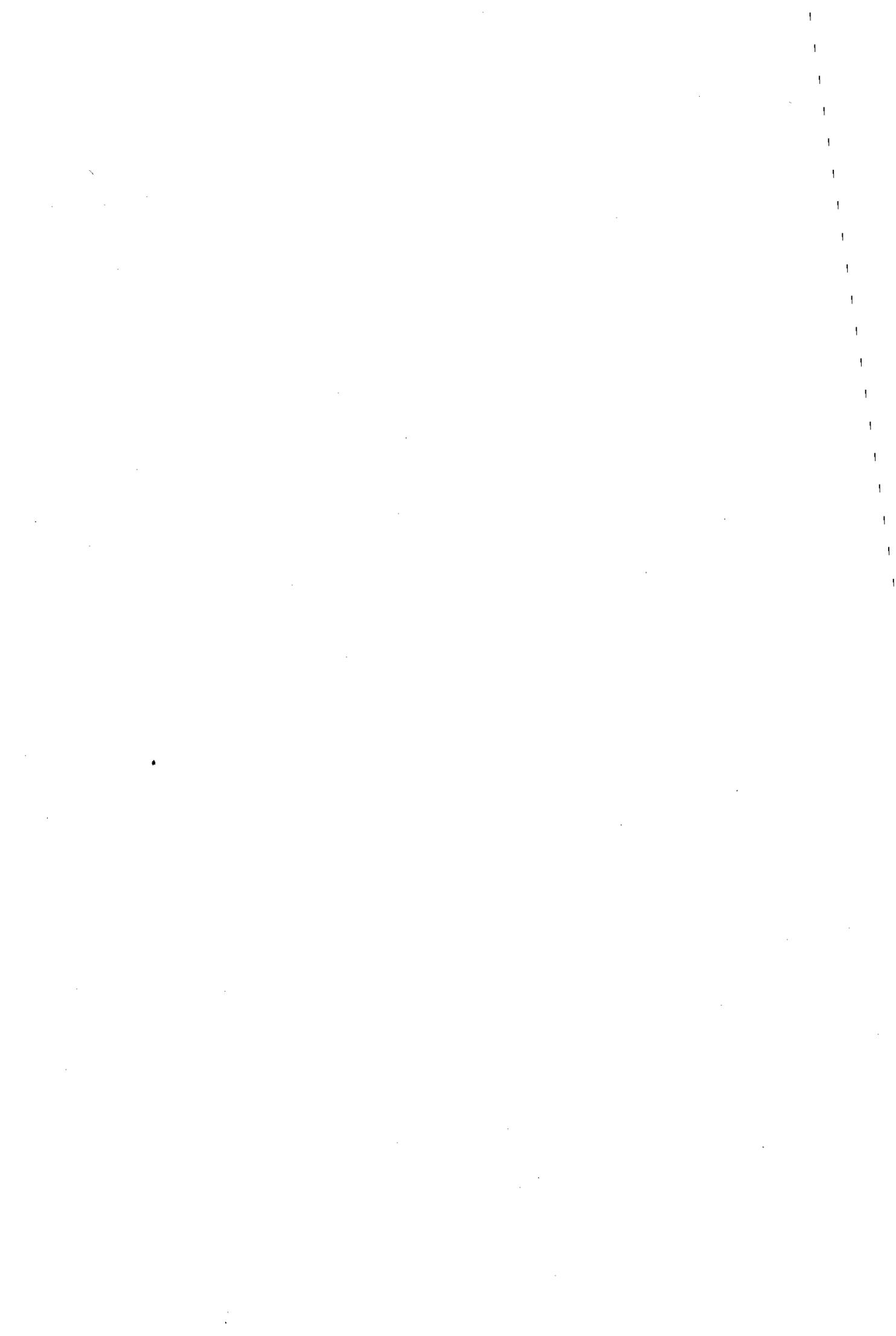
BLDSC no :- DX 88961

LOUGHBOROUGH
UNIVERSITY OF TECHNOLOGY
LIBRARY

| | |
|-----------------------|------------------------------------|
| AUTHOR/FILING TITLE | |
| CHIN, H S | |
| ACCESSION/COPY NO. | |
| 040015699 | |
| VOL. NO. | CLASS MARK |
| 6 JUL 1990 | Loan copy |
| | 3 OCT 1997 |
| 5 JUL 1991 | date due :- |
| 3 JUL 1992 | 23 MAR 1998 |
| 2 JUL 1993 | LOAN 3 WKS. + 3 UNLESS RECALLED |
| | TM 82660 |
| 27 JUN 1997 | |

040015699 7





TRANSMISSION OF VARIABLE BIT RATE VIDEO OVER AN ORWELL RING

by

Hin Soon Chin, B.Sc.

A Doctoral Thesis

*Submitted in partial fulfilment of the requirements for the award of the degree of
Doctor of Philosophy of the University of Technology, Loughborough.*

July, 1989

Supervisors : Professor J.W.R.Griffiths and Dr. D.J.Parish

Department of Electronic and Electrical Engineering,
Loughborough University,
England.

© by Hin Soon Chin, 1989

| | |
|--|-----------|
| Loughborough University of Technology Library | |
| DATE | Feb 20 |
| CLASS | |
| ACC. NO. | 040015699 |

59317997

In Memory of ...

My Beloved Father

&

With Love to ...

My Beloved Mother

ABSTRACT

Asynchronous Transfer Mode (ATM) is fast emerging as the preferred information transfer technique for future Broadband Integrated Services Digital Networks (BISDN), offering the advantages of both the simplicity of time division circuit switched techniques and the flexibility of packet switched techniques. ATM networks with their inherent rate flexibility offer new opportunities for the efficient transmission of real time Variable Bit Rate (VBR) services over such networks. Since most services are VBR in nature when efficiently coded, this could in turn lead to a more efficient utilisation of network resources through statistical multiplexing. Video communication is typical of such a service and could benefit significantly if supported with VBR video over ATM networks.

The primary objective of the research described in this thesis is to address the various issues associated with the transmission of VBR video over an ATM network called the Orwell Ring. There are two major issues concerned in this work. One relates to the relative merits of VBR video, and the other, to the efficient transmission and control of VBR video traffic over the ring.

VBR video signal statistics were collected from a COST211 video-conferencing codec which was modified to operate as a VBR video source. The picture material investigated were mainly head and shoulders scenes corresponding to videophone and video-conferencing type pictures. The statistics collected enabled an assessment of the relative merits of the type of VBR video sources concerned. The results indicated a bandwidth saving of about 2-3 over constant bit rate video, as well as other advantages such as improved picture quality and simpler codec design. The effects of statistical multiplexing the outputs of uncorrelated VBR video sources were also investigated.

The signal statistics collected also provided information on the characteristics of the VBR video signals, and enabled a simple VBR video source model to be developed. The model exhibited most of the characteristics of the video sources upon which it was modelled, but in view of the constraints in computing resources, some of the simplifications made suggested that the scope of the model was somewhat limited. It is nevertheless considered to be adequate for this study.

Simulation studies of the transmission of VBR video over the Orwell Ring have shown that with correct dimensioning of the ring, proper exercise of the traffic control mechanisms provided by the protocol, and a reasonable amount of buffering, a good Quality of Service (QOS) can be obtained for all services in a multi-service environment. Cell loss rate and call

blocking rate which satisfy the required performance along with a mean cell delay in the order of 10 μ s, can be achieved without sacrificing much of the gain of VBR video. The protocol has also been demonstrated to be effective, even under overload conditions.

Finally, some experimental work in interfacing the modified COST211 codec as a VBR video source to a prototype Orwell Ring was carried out, in order to demonstrate the inter-working of the two systems and the effectiveness of the protocol. Considerations have also been given to the design of an addressing scheme for the video data suitable for the ATM environment.

ACKNOWLEDGEMENTS

The author would like to acknowledge the financial assistance received from Loughborough University and British Telecom Research Laboratories, which made this research possible. Acknowledgement also goes to BTRL for the loan of equipments and software which were essential for this project.

The author would also like to thank Professor J.W.R. Griffiths and Dr. D.J. Parish for their guidance throughout the research, for their help in many ways, in particular, for securing the financial support.

Special thanks must also extend to Dr. J.L. Adams and Mr. M.D. Carr of BTRL for their invaluable advice and discussion during the course of this research; to Mr. J.W. Goodge for the hardware support, the many useful discussion, and for reading and commenting on the draft of this thesis; to Mr. H.L. Lim, T. Rodger and S.G. Jayasinghe for reading the draft thesis; and to all my friends in the University and BTRL for their help and useful discussion throughout my research period, and for making this an enjoyable and memorable time.

My heartiest gratitude goes to my mother and my family for their patience and understanding during my absence from home, and for their constant encouragement; and to Ms. Z. Hashim for her encouragement and moral support.

H.S.Chin,
July, 1989

LIST OF SYMBOLS AND ACRONYMS

| | |
|--------------------|--|
| AR | Auto-Reset |
| BISDN | Broadband Integrated Service Digital Network |
| BTRL | British Telecom Research Laboratories |
| CBO | Constant Bitstream Oriented |
| CCITT | Consultative Committee on International Telephony and Telegraphy |
| COST | Co-operation in Scientific and Technological Research |
| CLR | Cell Loss Rate |
| EBU | European Broadcasting Union |
| FD | Frame Differences |
| LC | Layered Coding |
| MAC | Medium Access Control |
| MCL | Mean Cluster Length |
| PCM, DPCM | Pulse Code Modulation, Differential PCM |
| PMR | Peak-to-Mean Ratio |
| PSN | Packet Switched Network |
| QOS | Quality of Service |
| RI | Reset Interval |
| RR | Reset Rate |
| T_a, T_c, T_m | Call Acceptance, Ceiling and Masking Thresholds |
| VBR | Variable Bit Rate |
| η | Mean |
| σ, σ^2 | Standard Deviation, Variance |
| E | Erlang |
| kb/s, Mb/s | Kilo-bits and Mega-bits per Second |
| μ | Micro |

Symbols and Acronyms

| | |
|------------|---|
| ATM | Asynchronous Transfer Mode |
| 'd' | Unit in which bandwidth is allocated on the Orwell Ring |
| D | Pixel or sample delay |
| PDF | Probability Density Function |
| VLC | Variable Length Code |
| α | Shape parameter (Gamma Distribution); Correlation coefficient (Auto-correlation) |

DEFINITIONS AND TERMINOLOGY

Addressing Scheme

A scheme whereby unique codes and addressing information are inserted into the video data stream, to allow the decoder to reconstruct the picture and to enable the decoder to recover synchronisation from error conditions.

Asynchronous Transfer Mode

A packet and connection oriented transfer mode using time division multiplexing technique where the information flow is organised in fixed size data transport units called cells.

Auto-Reset

A mechanism which resets the 'd' allocation of the CBO services if no reset is received within a pre-defined interval.

Bridging

The forming of longer pixel clusters by joining groups of changed pixels which are separated by less than a pre-specified distance.

Call Acceptance Threshold

The reset rate below which call blocking commences.

Call Blocking

Refers to the situation when a new call request is rejected because of insufficient bandwidth on the ring to support the new call, without degrading the service performance of the existing connections.

Ceiling Threshold

The reset rate below which no increase in 'd' allocation for VBR video will be allowed.

Cell

Is the data transport unit in ATM technique. It consists of a header field, which contains routing and control information, and an information field for carrying user data, both fields are of fixed sizes.

Cell Delay

Refers to the time elapsed between a cell generation and its launching onto the ring, i.e. time spent waiting in the node buffer.

Cell Loss

Refers to cells discarded from a node either because of excessive cell delay (for synchronous service) or buffer overflow.

Cluster

A groups of adjoining changed pixels and pixels included by the bridging process to increase cluster length.

Data Rate Smoothing

Refers to the mechanisms (usually with feedback control) used to smooth a time varying data rate to a constant data stream.

'd' Allocation

Synonymous with bandwidth allocation; 'd' is the unit in which bandwidth allocation is measured. A 'd' of one is equivalent to a minimum allocation of 1 Mb/s for CBO services, but can be less for VBR services because of masked resets.

Frame Differences

The proportion of pixels in a video frame which are deemed to have changed significantly from the previous frame, as well as pixels included by the bridging process.

Layered Coding

A video coding concept whereby video signals are coded into layers of hierarchical information of different picture resolution. Layers carrying video related data e.g. voice, text, can also be included.

Load Control

The mechanisms which decide whether a new call request or a request for more bandwidth allocation can be accepted or not. The decision is based upon whether the increased load will cause the ring to overload. e.g. call blocking, dynamic 'd' allocation.

Masked Resets

An overload control mechanism provided by the Orwell protocol where some resets are masked from the VBR services at a rate proportional to the level of overload. This effectively removes bandwidth gradually from these services and protects the ring against overload.

Masking Threshold

The reset rate below which the masking mechanism is activated.

Monitoring Interval (of the dynamic 'd' allocation scheme)

The interval over which cell arrival rate is monitored.

Movement Detector

A mechanism used to decide whether a pixel has changed significantly from the corresponding pixel in the previous frame.

Observation Interval

The interval over which reset rate is monitored and computed.

Orwell Torus

A switch where rings are stacked in parallel. The Orwell protocol operates across the whole Torus making it appear as a single high capacity ring.

Overload Control

The mechanisms which protect the ring against overload conditions due to the statistical behaviour of the VBR traffic. e.g. masked resets, auto-resets.

Packet Video

The concept where digital video signals are segmented into blocks of data for transmission over packet switched networks.

Peak-to-Mean Ratio

The peak-to-mean bit rate ratio of a VBR coded video signal. It provides some measure of the burstiness of the signal and the VBR video gain over CBO video.

Quality of Service

Service performance in terms of call blocking rate and cell loss rate.

Reset Interval, Maximum

Time elapsed between two consecutive ring resets. Maximum reset interval is the pre-defined interval which should not be exceeded, e.g. 125 μ s.

Reset Rate

The number of resets observed on the ring within a given time interval, e.g. 2 ms. This provides a measure of the amount of unused bandwidth on the ring.

Ring Dimensioning

Estimating the ring capacity required to support a given level of offered load with a pre-specified QOS.

Traffic Control

This consists of load control and overload control.

VBR Video

Video signals which are coded such that the data rate is proportional to the spatial and/or temporal complexity of the picture at any given moment, thus maintaining a constant picture quality

Video Coding

The application of digital image processing technique to remove redundancy in the video signals, in order to reduce the signal bandwidth for transmission economy.

Logical Addressing

Indicates the desired destination without indicating the route to be taken. The route taken by the cell may be dynamically altered without any header manipulation.

TABLE OF CONTENTS

| | Page |
|--|------------|
| Abstract | i |
| Acknowledgements | iii |
| List of Symbols and Acronyms | iv |
| Definitions and Terminology | v |
| Table of Contents | ix |
| Chapter 1 : Introduction | 1 |
| 1.1 Broadband ISDN and Video Services | 1 |
| 1.2 Variable Bit Rate Video | 2 |
| 1.3 The Orwell Ring | 4 |
| 1.4 Research Objective | 5 |
| 1.5 Thesis Organisation | 6 |
| Chapter 2 : VBR Video : Coding and Transmission | 8 |
| 2.1 The Need for Efficient Video Coding | 8 |
| 2.2 Effects of Coding on Output Bit Rate | 9 |
| 2.3 Transmission of Packet Video over ATM Networks | 14 |
| 2.3.1 The Advantages and Problems | 14 |
| 2.3.2 Layered Coding | 17 |
| 2.3.3 Traffic Control Strategies for VBR Video | 19 |
| 2.3.3.1 Call Control Level | 19 |
| 2.3.3.2 Cell Transfer Level | 20 |
| 2.4 Summary | 21 |
| Chapter 3 : VBR Video Signal Statistics | 23 |
| 3.1 Introduction | 23 |
| 3.2 Experimental Arrangement | 24 |
| 3.2.1 The Codec Used in Experiments | 24 |
| 3.2.2 Threshold Selection | 27 |
| 3.2.3 Picture Source | 28 |

| | |
|---|-----------|
| 3.3 Experimental Results | 28 |
| 3.3.1 Statistics of Frame Differences | 29 |
| 3.3.1.1 PCM Mode | 29 |
| 3.3.1.2 Encoded Mode | 34 |
| 3.3.2 Statistics of the Output Bit Rate | 34 |
| 3.3.3 Probability Density Function (PDF) of Frame Differences and Output Bit Rate | 39 |
| 3.3.3.1 PDF of Frame Differences | 39 |
| 3.3.3.2 PDF of Multiplexed Frame Differences | 42 |
| 3.3.3.3 PDF of Output Bit Rate | 44 |
| 3.3.4 Cluster Length and Addressing Overhead | 44 |
| 3.4 Conclusions | 48 |
| Chapter 4 : A Simple VBR Video Source Model | 50 |
| 4.1 Introduction | 50 |
| 4.2 Auto-Correlation of the Frame Differences | 51 |
| 4.3 Distribution of Changed Pixels Within a Frame | 51 |
| 4.4 Modelling of the Source Behaviour | 53 |
| 4.4.1 Probability Density Function of Frame Differences | 55 |
| 4.4.2 Auto-Correlation of the Frame Differences | 56 |
| 4.4.3 Distribution of Changed Pixels Within a Frame | 57 |
| 4.4.4 Coding Performance and Addressing Overhead | 59 |
| 4.4.5 Field Synchronisation | 60 |
| 4.5 A Simple Source Model | 60 |
| 4.6 The Model in the Ring Simulation Environment | 66 |
| 4.7 Summary | 67 |
| Chapter 5 : Simulation of VBR Video on an Orwell Ring | 68 |
| 5.1 Introduction | 68 |
| 5.2 The Orwell Ring - A Brief Description | 69 |
| 5.2.1 Service Queues With Priority | 70 |
| 5.2.2 Masked Resets | 70 |
| 5.2.3 Auto-Resets | 71 |

| | |
|---|------------|
| 5.3 The Orwell Ring Simulation | 72 |
| 5.4 Aspects of Interest | 73 |
| 5.5 Dimensioning of the Ring for VBR Video | 75 |
| 5.6 The Simulation Study | 78 |
| 5.6.1 Ring Bandwidth Estimation | 78 |
| 5.6.2 The Reset Rate - Ring Load Characteristic | 83 |
| 5.6.3 Traffic Control and Thresholds Considerations | 83 |
| 5.6.3.1 Call Acceptance Threshold and Call Blocking | 85 |
| 5.6.3.2 The 'd' Allocation for VBR Video | 88 |
| 5.6.3.3 Masking Threshold | 92 |
| 5.6.4 Buffer Size | 95 |
| 5.6.5 Service Performance on the Ring | 97 |
| 5.6.5.1 VBR Video Traffic | 97 |
| 5.6.5.2 Mixed Traffic | 102 |
| 5.6.6 Reset Rate Observation Interval | 106 |
| 5.6.7 Cell Delay and Cell Loss | 107 |
| 5.6.8 Auto-Resets | 108 |
| 5.7 Discussion and Conclusions | 110 |
| Chapter 6 : Experimental Work on a Prototype Orwell Ring | 113 |
| 6.1 Introduction | 113 |
| 6.2 The Experimental Orwell Ring | 113 |
| 6.3 The VBR Video Codec | 115 |
| 6.4 An Addressing Scheme for the Video Data | 116 |
| 6.5 Codec to Ring Interfacing | 119 |
| 6.6 The Experimental Work | 119 |
| 6.7 Discussion | 121 |
| Chapter 7 : Discussion and Conclusions | 123 |
| References | 134 |
| Appendix I : Snapshots of Picture Sequences | 138 |
| Appendix II : Addressing Schemes | 142 |
| Appendix III : Tables Used in Ring Dimensioning | 147 |

Chapter 1:

Introduction

1.1 Broadband ISDN and Video Services

Broadband integrated services digital networks (BISDN) are emerging in response to the anticipated demand for new broadband services and the rapid advances in technologies - particularly in optical transmission, high speed switching and processing. The falling costs of implementing these technologies further expedite the process of network evolution towards the BISDN concept.

BISDN will be developed on the basis of the existing narrowband ISDN, and will support all existing services as well as any new services that may arise. Besides offering user access to a broader bandwidth well in excess of 64 kb/s, BISDN should also provide flexibility and be service independent. These requirements are essential for the evolution towards a single layer, multi-service network. The advantages of such an integrated communications network are widely acknowledged^[23,29,33,38].

Dynamic rate flexibility has been singled out as a very desirable feature for any future network. Such flexibility is required to cope with the uncertainties in future demand for network services, most of which will exhibit bursty traffic characteristics and have widely differing requirements^[38]. In order to provide these services economically and efficiently, BISDN must be able to adapt dynamically to the requirements of the individual services at any instant, and thereby provide the basis for the statistical multiplexing of all services, leading to a more efficient utilisation of network resources.

A very promising proposal for the implementation of BISDN is the use of the Asynchronous Transfer Mode (ATM) technique, where all information is transported and switched in packet form using fixed size data transport units called cells, as defined in CCITT Recommendation I.121. The use of cells as the basic transport unit in ATM based networks provides a high degree of flexibility, where the rate adaptation capability and the absence of

physical channel structures allow for the complete sharing of network resources among all the different connections; this enables full service integration and statistical multiplexing to be achieved easily.

Although the exact nature of the broadband services that are to be offered has yet to be considered by the CCITT, video-communication is expected to be one of the major services. The market potential for services such as videophone, video-conferencing and video distribution is widely recognised. Videophone for example, has a potential customer base equivalent to that of the present telephony service.

Video signals are characterised by their high bandwidth requirements and high transmission costs. However, the advanced signal processing technologies available today enable this visual information to be coded efficiently, thereby optimising the bandwidth requirements. The optimised bandwidth requirement may differ from one video source to another and may very often fluctuate with time. With an ATM based BISDN, this 'variable rate' information can be transported efficiently, thereby making more efficient use of the network resources for these services, rendering them more cost effective.

The transport of these variable information rate services over ATM networks is now a subject of worldwide research, and the work presented in this thesis addresses the problems associated with the transmission of variable bit rate video over one such network - the Orwell Ring.

1.2 Variable Bit Rate Video

Video related visual services place large demands on network resources. It is therefore important, for a given picture quality, to keep these demands to a minimum in order to reduce the transmission costs for such services. This objective can be achieved through the use of efficient video coding techniques. Since a video signal contains a varying amount of information which is proportional to the instantaneous complexity of the pictures, proper coding of the signal for a fixed picture quality will result in a variable bit rate output; hence the term Variable bit rate (VBR) video.

The concept of VBR video has not received much attention in the past primarily because there was no broadband VBR carrier available. Instead, the fluctuating bit rate has always been smoothed, using a buffer and some feedback mechanism, to yield a constant bit rate signal for transmission over a fixed bit rate channel. This is known as constant bitstream oriented (CBO) video. This technique inevitably leads to the inefficient use of bandwidth when the actual data rate is lower than the allocated bandwidth itself, and conversely, picture quality would be degraded when the data rate has to be reduced, in order that no loss of data

is incurred. In general, a bandwidth much higher than the expected mean data rate would have to be allocated in order to avoid excessive or prolonged degradation to the picture quality.

The advent of ATM based BISDN will offer the potential for VBR video to be supported in the future. With VBR video, video services can be provided more efficiently for a given picture quality, since only the exact amount of bandwidth will be required from the network at any given time. On average, the bandwidth utilised is equal to the mean bit rate of VBR video, which is much lower than the peak bit rate. See references [27,37] and Chapter 3 for some experimental measurements.

However, this bandwidth saving can only be realised if a large number of uncorrelated VBR video sources are available for multiplexing in a pool of shared resources. Since simultaneous demand for large bandwidth from several sources rarely occurs, the fluctuating bit rates are averaged out through bandwidth sharing. The resultant combined bit rate would have a mean equal to the sum of the means of the individual sources, and would have smaller peaks with less fluctuations. Therefore, on average, the bandwidth required for each connection is only slightly higher than the mean bit rate of a VBR video source. High bandwidth demand from a source can be met by the allocated but unused bandwidth from other sources which have low bandwidth demand, or from the unused resources on the network. Research carried out to investigate the statistical multiplexing of VBR video by Haskell^[27] and Koga^[37] suggests that a bandwidth saving of the order of 2 with respect to CBO video can be achieved. This represents a significant amount of saving, and hence cost reduction in the provision of these services.

In summary, VBR video offers some very attractive features for supporting video services over an ATM network, namely:

- constant picture quality;
- less complex codec design since no feedback mechanism is required;
- bandwidth saving and hence lower transmission charges through statistical multiplexing.

The last-mentioned feature (lower transmission charges), is of particular importance insofar as the viability of video services on a widely accessible scale is concerned.

1.3 The Orwell Ring

The Orwell Ring is, in essence, a high speed slotted ring implementing the Orwell medium access control protocol. The protocol was developed at British Telecom Research Laboratories (BTRL) as a flexible protocol for carrying mixed services such as voice, data, and possibly low bit rate VBR video on a single multi-service network. The ring supports a connection-oriented, cell-based transport mechanism which complies with the definition of ATM.

The ring, or more precisely, the Orwell protocol, is unique when compared with other slotted ring protocols in that slots are released at the destination nodes rather than the source nodes. This feature allows the ring capacity to be increased, or alternatively, the delay to be reduced, for the same load. A novel load control mechanism overcomes the potential hogging problem arising from the destination release policy, and guarantees an upper bound for cell delay.

The load control mechanism is fully distributed and is at the heart of the Orwell protocol, guaranteeing synchronous services small cell delay and high cell security, while allowing the remaining bandwidth to be used in a flexible and efficient manner by delay tolerant asynchronous services and other low priority services such as VBR video. The load control mechanism operates a bandwidth reservation scheme and provides an estimate of the unused bandwidth on the ring. By controlling access to the slots from the nodes according to their bandwidth reservation, and rejecting call requests that might cause the ring to overload, cell delay can be bounded. The protocol further provides overload control on non-priority services in order to protect the synchronous services.

The Orwell Ring has several advantages over other non-ring or non-bus based ATM switches. For instance, logical addressing can be used for fast signalling, which then allows for a greater distribution of the network processing resources. The ring also provides a more elegant way of broadcasting and multicasting information. However, the throughput of the ring is ultimately limited by the single path. While a ring speed of 140 Mb/s, achievable with today's technology (and possibly 1 Gb/s in the next decade), may be sufficient for private network needs, it is inadequate for public network use.

To meet the requirements of a public switch, rings can be stacked in parallel and their separate capacities made to appear as a single high capacity Orwell Ring as shown in figure 1.1. This multiple ring system is called an Orwell Torus, and besides offering increased capacity, the Torus also provides the reliability and the ease of maintenance and expansion that is required for a public switch. The throughput of the Torus will still be limited by the maximum number of rings that can be stacked together.

For more details of the Orwell Ring and Torus, see references [1,2,3,22].

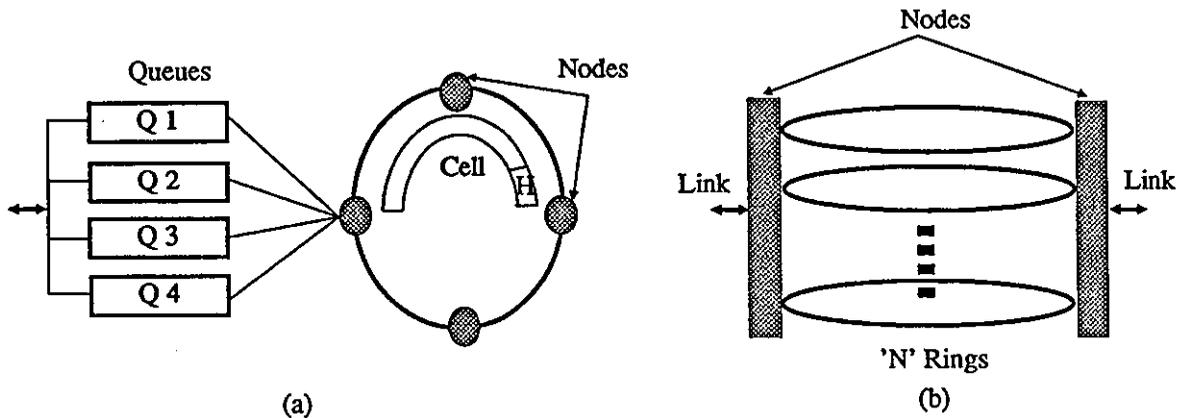


Fig. 1.1 (a) Orwell Ring and (b) Orwell Torus.

1.4 Research Objective

The primary objective of this research project is to address the various issues associated with the transmission of VBR video over an Orwell Ring. The two main issues which this work is concerned with, are:

1. the relative merits of VBR video, and
2. the efficient transmission and control of VBR video traffic over the ring.

(1) Earlier work^(27,37) has suggested that some gains in bandwidth saving and improved picture quality can be achieved with VBR video. The work in this thesis attempts to verify this claim, and to assess these gains by carrying out measurements on the signal statistics using a practical codec. This also provides a better understanding of the characteristics of VBR video, which will be important when addressing the second issue.

(2) Although the Orwell Ring is capable of supporting VBR video connections, the actual transmission and control of VBR video presents a major traffic control problem, especially in the presence of CBO services. Control must be exercised so that VBR video would only suffer minimal cell loss, with no resultant loss on the CBO connections, and at the same time achieve high network utilisation. Cell delay is another factor which needs to be controlled. Consideration must also be given to the recovery from, and the concealment of, cell loss related errors.

The research was conducted along the following lines to achieve the objective set out above:

- I. To study the statistical behaviour of VBR video signals in order to gain a better understanding of the signal characteristics, and to assess the relative merits of VBR video.
- II. To develop a simple computer model for a VBR video source using the results from (I).
- III. To incorporate the VBR video source model into an Orwell Ring simulation, and to study the control and performance of VBR video over the ring.
- IV. To produce and interface a variable bit rate video codec to an experimental Orwell Ring to facilitate practical assessment.

1.5 Thesis Organisation

Chapter 2 discusses the need for efficient coding of digital video signals for transmission over a switched network, and the effects that coding has on the output bit rate. The advantages and problems associated with the transmission of packetised VBR video over ATM networks and some proposed traffic control strategies are also examined in this chapter.

Chapter 3 describes the experiments carried out on a modified COST211 video-conferencing codec and a computer model of the codec, to study the statistics of VBR video which was coded using some simple compression techniques. Results on the statistics of frame differences and output bit rate, and the effect of source multiplexing, are presented. The relative advantages of VBR video are assessed.

Chapter 4 describes the development of a simple VBR video source model, based on the statistics obtained in Chapter 3 and other relevant aspects of the VBR video characteristics. The simplifications assumed in the model, and its limitations, are discussed.

Chapter 5 provides more information on the Orwell protocol, and describes the simulation study of VBR video on an Orwell Ring. Specific aspects of interest include ring dimensioning, resource allocation, call acceptance thresholds and call blocking, cell loss, cell delay, reset rate measurement, buffer size, effects of masked resets, and impacts on other CBO services. Simulation results are presented and discussed.

Chapter 6 describes some of the work carried out on the transmission of VBR video over an experimental Orwell Ring: the interfacing between a modified codec and the ring, an addressing scheme for video data and the performance of the system.

Chapter 7 contains a discussion of the work and the results, and presents the conclusions drawn from the research, within the framework of the stated objective.

Chapter 2 :

VBR Video: Coding and Transmission

2.1 The Need for Efficient Video Coding

If video signals are transmitted with their full bandwidth, they would require channel capacity of an enormous magnitude. For instance, a PAL composite video signal requires a bandwidth of 5.5 MHz or a digital equivalent of at least 70 Mb/s for its transmission^[1], compared to the 64 kb/s required for voice transmission. The transmission costs for signals with such large bandwidth requirements would be prohibitive for all except a few highly specific applications such as cable television. In the public network the problem is even more acute where network resources are becoming a scarcity with the ever increasing demand for bandwidth from many new and existing telecommunication applications.

This problem has long been realised and much research has been directed at looking for ways of reducing the bandwidth of video signals. As a consequence, many video coding algorithms, almost all employing digital signal processing techniques, have been developed to exploit the large amount of redundancy that exists within the signals in order to achieve bandwidth reduction or compression. Advances in microelectronics have already enabled some very complex and efficient coding algorithms to be implemented on coder-decoder (codec) of practical physical size, while the falling costs of the hardware will eventually bring the price of these codecs within the reach of everyone.

Table 2.1 gives a comparison of the bandwidth required for the direct digitisation of component video signals and the bandwidth compression that can be achieved for the different video quality standards^[46] with efficient video coding, using some of the currently available algorithms.

| Video Standard | Un-compressed | Compressed |
|-----------------------------------|---------------|--------------------|
| Videophone/ Video-conferencing | 20 Mb/s* | 64 kb/s- 2 Mb/s |
| Broadcast T.V. | 216 Mb/s | 17- 34 Mb/s |
| High Definition T.V. | 1.2 Gb/s | 70- 140 Mb/s |

*COST211 frame data format

Table 2.1

As can be seen, the reduction in signal bandwidth that can be attained is very significant and a much lower transmission cost would follow. Note however, that the figures quoted for the compressed signals are those for CBO video, and are oriented to the European transmission hierarchy. With future ATM based BISDN, such restrictions may be waived and the more efficient VBR video coding techniques can be used to obtain a lower mean bandwidth than those figures quoted above.

The aim of efficient video coding is, therefore, to reduce to a minimum the required transmission bandwidth for a given picture quality, with a reduction in the transmission costs. This is imperative if video services, particularly videophone and video-conferencing, are to make an impact on future telecommunications demand.

2.2 Effects of Coding on Output Bit Rate

Research into bandwidth or bit rate compression for video signals has emerged with a multitude of digital video coding techniques. The subject of video coding has been comprehensively covered elsewhere^[41,43,48]. The objective here is therefore not to review the various coding techniques or their performance, but merely to discuss briefly their effects on the output bit rate of the coded signal, i.e. whether a constant or a varying output would result.

Most of the coding techniques being used today operate on component video signals and follow two basic approaches: Predictive Coding and Transform Coding. More often, the two approaches are used together in what is known as a Hybrid Coding approach. As the

number of coding techniques that fall within the scope of these two approaches are too numerous to be examined individually, only those more commonly used will be considered, and generalisations will be made wherever possible.

2.2.1 Predictive Coding

In predictive coding, the digital signal samples, or picture elements (pixels), to be encoded are predicted from previously transmitted pixels either from the same video frame or from the previous frame, or from both. The aim is to make accurate predictions of the pixels to be encoded using the high correlation that exists among pixels in the same neighbourhood, so as to minimise the prediction errors and their dynamic range. As a result, fewer bits are required to code these errors for transmission. The statistical property of these errors also lend themselves to variable word length coding, leading to further bit reduction. As such, variable word length coding has almost become an integral part of predictive coding techniques.

Several more sophisticated variants to the basic approach described above have been developed. The better known are the Conditional Replenishment and Motion Compensation techniques, and all these will be examined.

In the case of the basic predictive coding (more commonly known as differential pulse code modulation (DPCM) coding), if the prediction errors are quantised and coded with fixed length codes, the output bit rate would remain constant since every pixel would still be represented by a fixed number of bits as in the original PCM signals. It is, however, more efficient to code the quantised prediction errors with variable length codes (VLC). The use of VLC implies that the resultant output bit rate would now depend on the local image complexity. For instance, the prediction errors in areas of high spatial detail or in frames with large amount of motion would be larger than in flat areas or in relatively still frames, and hence more bits would be assigned.

For intra-frame DPCM coding with VLC, the instantaneous bit rate thus varies within a frame as the spatial complexity varies from one part of the frame to the next, assuming of course that the quantiser remains constant for a given picture quality. The average bit rate per frame is unlikely to vary greatly from frame to frame unless it involves moving scenes, camera zooming or scene changes, whereby the image complexity changes significantly between frames. Adaptive prediction, which adapts to the local image complexity, can be used to ensure that the prediction errors are minimised at all times, and so the resultant bit rate will be less varying. Overall, the effect of the application of intra-frame DPCM coding with VLC to a video signal is to yield a variable bit rate signal. However, it has been suggested that it is possible to smooth this bit rate variation using a buffer of a reasonable

size and a simple feedback control mechanism^[5]. On that premise, even though the underlying data flow is time-varying in nature, it can be buffered to give a constant bit rate signal.

The use of intra-frame DPCM coding alone is, however, not an efficient coding method as it fails to exploit the high correlation that exists between pixels in adjacent frames, especially for scenes containing little motion. A simple inter-frame DPCM coder which makes use of the corresponding pixels from the previous frame for prediction would be more efficient in this case. The number of pixels with large prediction error would clearly depend on the amount of motion in the scene, and subsequent coding with VLC would result in a very bursty output bit rate which varies within a frame, as well as from frame to frame. Buffering alone in this case would not be sufficient to smooth the output^[7,21]. Other strategies are required and these usually involve varying the picture quality.

Intra-frame and inter-frame DPCM codings can also be used together to achieve higher coding efficiency; the former performs well with scenes containing large motions while the latter is excellent with relatively still scenes^[22]. The coder switches between the two modes of prediction depending on the amount of motion in the scene to minimise prediction errors. This combined approach along with the use of VLC would still result in a variable bit rate output comparable to the simple inter-frame DPCM case, but the peak bit rate would be upper bounded by the performance of the intra-frame DPCM. Smoothing of this output may not be easy.

The simple DPCM techniques described above can be made more efficient by incorporating more sophisticated adaptive functions, for instance, by coding and transmitting only those pixels that have prediction errors greater than a given threshold. This is the principle of conditional replenishment coding, and normally, inter-frame predictions are used. This technique, being basically an inter-frame DPCM technique, would again produce a bursty output data rate which is proportional to the amount of motion in the scene: the more motion, the more pixels would need to be coded and transmitted. Some very complex buffer feedback control mechanisms have been devised to level this bit rate variation. A codec employing this coding technique is described in Chapter 3.

The conditional replenishment technique is a very simple and efficient technique, achieving high compression and good picture quality. However, it can be further improved by taking into account the motion of objects in its prediction, assuming most motion to be translational and therefore estimable - this technique, and the most efficient predictive coding technique by far, is known as motion compensation. Predictions are made from the appropriately displaced pixels in the previous frame based on some estimated motion vectors. The pixels are classified into non-moving and moving pixels as in conditional replenishment. The former require no information to be sent, while the latter can be further sub-divided into compensable and non-compensable pixels depending on a prediction error criterion^[42]. The

first type requires only the motion vector information to be sent, whereas the second type requires the prediction errors to be transmitted as well. The last distinction is however not always made, in which case, both motion vectors and prediction errors are transmitted for all moving pixels.

Clearly, the resultant output data rate of a motion compensation coder would be motion dependent and therefore irregular. Results presented by many workers^[20,42] have served to illustrate this. Although the burstiness of the bit rate variation could be smaller than in the conditional replenishment case, it could still prove to be difficult to smooth, again without adjusting some of the coding parameters^[49].

2.2.2 Transform Coding

Transform coding techniques exploit pixel correlation in a different way to the predictive coding approach. In transform coding, the statistically dependent pixels are linearly transformed into a set of less correlated coefficients which are then quantised and coded for transmission. Compression is achieved by coarsely quantising the high order coefficients and not sending those that are small. The subject of transform coding has been discussed in great detail^[6]. Its effect on the output data rate will be discussed as a general case here without referring to any specific transform methods.

The picture to be encoded is usually segmented into blocks of sub-picture (blocks of 8 x 8 pixels for example). Transformation is carried out on the blocks in order to take advantage of the two dimensional correlation property of pictures. The transform operation results in the compaction of image energy into a few low order coefficients and the spread increases with the spatial detail of the blocks (less pixel correlation). Having carried out the transformation, the significant coefficients need to be selected for quantisation and transmission. Two basic schemes are used for this purpose - Zonal Sampling and Threshold Sampling.

Zonal sampling makes use of the statistical averages of the coefficients of a set of test pictures and chooses only those coefficients that have variances greater than a prescribed value; bits are allocated accordingly for coding each of the coefficients. In the non-adaptive case, this zone would be fixed, leading to a fixed bit rate system. This approach is highly inefficient since blocks with statistics deviating from the average would be poorly coded. This scheme can however be made adaptive, for instance, by using different zones for blocks with different high order coefficient activities as suggested by Chen^[9], where improvement in coding performance was reported. These zones may have a different number of selected coefficients and a different number of bits allocated, the result being a varying output data rate as the spatial detail varies from region to region within the frame.

Threshold sampling is an adaptive scheme where all coefficients that are greater than a preset threshold are selected for transmission. This scheme therefore adapts itself to the local image statistics but does require a high addressing overhead. The effect of this scheme is also to produce a variable output bit rate as the local image statistics vary.

With adaptive zonal sampling or threshold sampling (sometimes used in combination), the effects on the resultant bit rate could be quite similar to the intra-frame DPCM case, where the mean data rate per frame will only vary greatly from frame to frame if there is a large change in the spatial image complexity. It is not clear whether this variation can be smoothed easily, but being mainly intra-frame variation, it may be possible to do so using a buffer and a simple feedback mechanism; although bit allocation adjustment and threshold adjustment have been used for this purpose.

Block transform coding, although very efficient, fails to consider the inter-frame pixel correlation. Three dimensional coding has been studied^[50], but being extremely computation intensive and requiring immense storage space, is not considered practical for the time being despite better performance as reported. However, as inter-frame pixel correlation varies greatly with motion in the scene, adaptive three dimensional coding will undoubtedly yield large variation in the output bit rate as in the case of inter-frame predictive coding.

2.2.3 Inter-frame Hybrid Coding

In view of the difficulties in implementing three dimensional transform coding and the desire to extend the two dimensional transform to exploit temporal redundancy, a different approach has been adopted in which predictive coding is used for inter-frame coding to compensate for the deficiency of the two dimensional transform technique. Such an approach is known as Inter-Frame Hybrid Coding. It has very good coding performance and is probably the most widely used approach today.

Although intra-frame hybrid coding is possible, it is not as efficient, and recent work in hybrid coding are almost exclusively inter-frame based. This approach is covered in detail elsewhere^[51] and will not be discussed here. It is necessary only to point out that studies into adaptive intra-frame hybrid coding have revealed the variable nature of the resultant data rate. These results are not at all surprising since the underlying coding techniques themselves yield variable data flow as discussed in the previous sections.

Hybrid coding schemes can be put together in a variety of ways. For instance, the predictive step could precede the transform operation or vice versa; further, there is a variety of techniques under the two approaches to choose from. A scheme which is very popular in recent video coding work (for example CCITT Rec. H261, Esprit 925) is the use of block motion compensation techniques in the pixel domain, followed by transform coding on the prediction errors of the blocks which require transmission. This approach is highly efficient

since both component techniques are very efficient. Their effects on the output data flow have already been discussed separately; their combined effect can be expected to be of a similar nature, i.e. to yield a bursty data rate at the output.

The above examination of some of the more common coding techniques, albeit brief, indicates that the more efficient coding schemes are those which employ a hybrid approach with adaptive inter-frame predictive and transform coding, but these schemes also result in a bursty output bit rate if picture quality is to be maintained approximately constant. This result is hardly surprising since the information content in video signals varies with both spatial and temporal image complexity.

The unavailability of a broadband VBR carrier hitherto has meant that the bursty data rates must be smoothed before transmission. The feedback mechanisms used in some of the codecs for this purpose⁽⁴⁴⁾ are very complex and require continuous adjustments to the coding parameters or coding schemes. These adjustments cause picture quality to vary and, in general, the output must be smoothed to a bit rate significantly higher than the mean bit rate to reduce the probability of buffer overflow and to maintain a minimum picture quality at all times. Furthermore, the bit rate must be adjusted to fit in one level of the transmission hierarchy. With the proposed ATM based BISDN, the VBR output could be supported in its natural form without the aforementioned constraints, and this could potentially lead to a more efficient video transmission while maintaining a constant picture quality.

2.3 Transmission of Packet Video Over ATM Networks

In the ATM network environment where fixed size cells are the basic transport unit, digital video signals are packetised at the transmitter into packets of fixed length. These packets are then transported independently in the cells, routed by the cell headers, to the appropriate receiver where they are de-packetised and the video signals are retrieved for subsequent image reconstruction. Video signals transmitted by this mechanism is known as Packet Video, although the term does not necessarily refer to fixed length packets in packet switched networks (PSNs) other than ATM. The discussions that follow will be conducted in the context of ATM networks but they are equally applicable to other PSNs.

2.3.1 The Advantages and Problems

Packet video is attracting a lot of interest, in the light of the recent developments in high speed PSNs, as a probable means of providing video communication services over such networks. The advantages of transmitting packet video over ATM networks or any other PSNs arise from the characteristics of the networks and some of these have already been mentioned earlier, for instance, rate flexibility enables VBR video to be supported and allows

for source multiplexing, as well as offering a continuous range of transmission rates to which the outputs of CBO services can be optimised. All these could potentially lead to more efficient video transmissions.

The concept of packet video on ATM networks also eases the provision of multicasting capability. This feature can be implemented easily either by duplicating the packets and launching them on separate cells or by using the broadcast facility available on ring or bus type networks. For broadcast video applications, the layered coding concept (see section 2.3.2) can be implemented more efficiently in this framework; the technique offers an attractive method for introducing HDTV with downward compatibility with the existing television standards, and not least, the integration of video services with other media can be achieved more easily.

However, the cell-based transport mechanism of the ATM networks presents problems which are not encountered in the conventional circuit switched networks (CSNs), and these must be addressed carefully before packet video can be used to support video services on these networks. The problems that need to be considered, in addition to the well known bit error, include packetisation delay, cell delay and jitter, cell loss and cell sequence error. These will be discussed.

2.3.1.1 Packetisation Delay

Packetisation delay arises out of the need to complete the fixed length cell before launching, and its magnitude depends on the source output data rate and the cell size. This delay contributes to the overall end-to-end delay, and if it becomes excessive for a particular service, cells must be transmitted incomplete to avoid violation of the service delay constraint, and this leads to inefficiency. Most video services, however, have high output data rates, and with the proposed 32 or 64 byte wide cells, the packetisation delay will be very small, for example 125 or 250 μ s respectively for a 2Mb/s video signal. With VBR video, this delay is not constant, being dependent on the instantaneous bit rate of the signals; packetisation delay can be large during a low bit rate period.

However, the delay constraints for video services may not be as critical as for services like voice, where echo is a major problem. Since video services are normally accompanied by voice, video delay should therefore be specified in relation to the delay requirements of voice service, which are covered in several CCITT Recommendations. If the two services are transmitted on separate channels, the relative delay between the two services should subscribe to EBU Recommendation R37⁽²¹⁾, which specifies a maximum video lead of 40 ms to a maximum video lag of 120 ms. Delay constraint of such magnitude should not be difficult to meet.

2.3.1.2 Cell Delay and Jitter

Cell delay is caused mainly by cells having to wait in the transmit queues for access onto the network. Given the asynchronous nature of the ATM network and the variation in network load, this waiting time is not constant, and the varying cell delay results in cell jitter. In order to make the cell transport mechanism transparent to the services, particularly the CBO services, this jitter must be removed and a buffer is usually used at the receiver to restore synchronisation of the bit stream. This is a function of the adaptation sub-layer of the ATM model^[17]. The size of the buffer required will depend on the magnitude of the jitter; it will also affect the end-to-end cell delay.

With VBR video, cell jitter may be less of a problem since decoding of the compressed video can be carried out asynchronously and a frame store can be updated as soon as the information becomes available. A buffer may still be required at the receiver to absorb sudden bursts of cell arrivals, considering that the speed of the network is much higher than the peak bit rate that a VBR codec can handle.

Cell delay constraint has already been discussed previously. The total end-to-end delay of a cell will comprise of packetisation delay, cell delay over several switches and the buffering delay at the receiver. The total delay should conform to the delay constraint suggested.

2.3.1.3 Cell Loss

Cell loss is perhaps the biggest problem confronting the feasibility of packet video since it may have serious implication on the quality of the video services. Cell loss can be caused by either bit error in the cell header or buffer overflow due to network overload. The latter is of particular concern in the presence of VBR video, and indeed may cause burst cell loss in the VBR video connections.

ATM networks should have a design goal of a cell loss rate of no greater than the bit error rate,⁵⁴ and effective load control should enable the CBO services to experience cell loss rate of the same order. It is unclear what this loss rate should be although figures in the order of 10^{-6} - 10^{-9} have been quoted^[19,52]. With VBR video, this figure could be significantly higher since the networks do not provide for their peak bandwidth. The relationship between cell loss rate and picture quality is also not known, but will obviously depend on the coding technique used and the amount of compression applied - the higher the compression, the less tolerance to cell loss. If the cell loss rate is deemed to cause unacceptable degradation to the picture quality, then counter measures must be employed to alleviate this problem. To this end, new coding approaches, such as layered coding, which are more tolerant to cell loss should be used.

Cell loss has a more severe impact on picture quality than bit error because it introduces discontinuity into the data stream and causes loss of synchronisation. This could in turn lead to a massive service disruption if no counter measures are taken. The most basic requirements are to detect the occurrence of cell loss and to suspend decoding until synchronisation is recovered. These measures are necessary to prevent error propagation. Cell loss detection may require cells to be numbered in sequence, or other auxiliary information to be forwarded to the receiver at regular intervals; while rapid re-synchronisation necessitates the insertion of synchronisation information at some known point in the data stream. Error propagation can also be confined by packing cells or groups of cells into independent information units so that the loss of any cell will not result in error propagation beyond the boundary of the cell or group of cells. There are however two problems. First, this method may not be always possible. Second, it inevitably leads to inefficiency from the use of extra synchronisation information and incomplete cells.

The above measures, however, only prevent the propagation of errors and do not deal with the errors themselves. Consequently, the errors will persist and can be visually irritating. Some means of error recovery or error concealment are therefore required. Systematic unconditional update transmitted over a more reliable path can be used to clear up errors due to cell loss, but it can be rather slow. Bit scrambling with error correction codes^[39] may offer a solution against a small number of cell losses at the expense of more complexity in the codec and more packetisation delay, but it is unlikely to be adequate for burst cell loss which can be common with VBR video. Layered coding may offer a satisfactory solution to error concealment and allow for the graceful degradation of picture quality.

2.3.1.4 Cell Sequence Error

This refers to the situation where cells arrive at the receiver in the wrong order. This problem should not occur in connection-oriented services where cells are transferred over fixed path virtual channels. Furthermore, draft recommendation I.121 requires that cell sequence integrity on a virtual channel be preserved by the ATM sub-layer of the ATM model. However, if connectionless services are supported, a cell sequence number must be inserted in the cells to enable corrective action to be taken at the receiver.

2.3.2 Layered Coding

Layered coding (LC) is a coding concept whereby a video signal is coded into a number of layers of hierarchical information and combined with layers of audio and perhaps text information^[54]. Two approaches to LC are being actively pursued and they differ only in the ways of partitioning the signals into the different layers: one involves the direct coding of the signals into the different layers^[24], while the other decomposes the signals into different

frequency bands using digital filtering techniques before coding the separate sub-bands. The latter approach is appropriately called sub-band coding^[34]. The common objective is to code the signals into layers such that lower layers contain low resolution or basic information, while higher layers carry high detail or enhancement information. The two approaches are illustrated in figure 2.1.

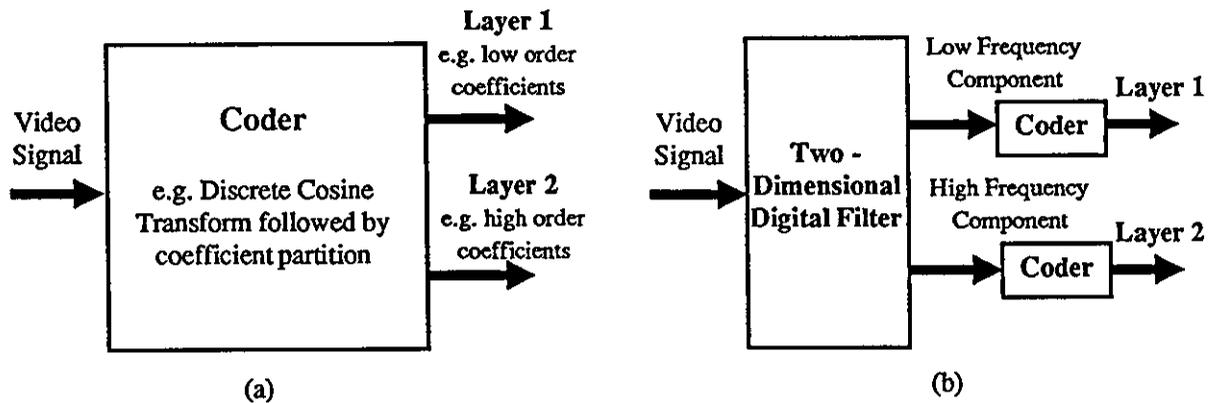


Fig. 2.1 Two Approaches of Layered Coding - (a) Hierarchical Coding and (b) Sub-Band Coding.

The advantages of LC have been briefly mentioned earlier and more details can be found elsewhere^[53]. In this work, only its error tolerant capability is relevant, and this will be dealt with in detail.

LC can be made more error resilient to cell loss error by providing better protection for the lower layers which contain the basic video information. The correct reception of this information will allow for the reconstruction of a picture of basic quality. With the basic image quality almost guaranteed, the loss of cells carrying higher layer information should only result in reduced image resolution without causing objectionable impairment to the picture. Propagation of errors in the higher layers can be reduced using information in the more secure lower layers. LC hence provides good error concealment capability and allows the graceful degradation of picture quality in case of cell loss.

Nomura^[45] and Ghanbari^[24] have shown that LC can maintain good picture quality even at cell loss rate of 10^{-1} in the upper layer of a two-layer video signal, while errors become objectionable at cell loss rate of 10^{-3} with conditional replenishment coding. However, it must be stressed that cell loss was assumed to be random rather than bursty in both cases; burst cell loss may result in more impairment to picture quality especially with non-layered coding. Hence, LC is probably more suitable for implementing VBR video where burst cell loss is likely to be encountered.

Still, there are some drawbacks to this coding concept: Coding efficiency may be reduced and there may be a higher transmission overhead. These could lead to a higher combined output data rate compared to the non-layered coding approach, although this may to some extent be compensated by being able to load the network to a higher loading level because of the reduced sensitivity to cell loss. A greater codec complexity would also result, and there may be difficulties in synchronising the different layers as they may have different packetisation delay and cell delay. The above arguments suggest that the number of layers in a LC model should be kept to a minimum to avoid inefficiency and to achieve practicality.

2.3.3 Traffic Control Strategies for VBR Video

VBR video, while offering some very attractive features for supporting video services over ATM networks, also gives rise to serious traffic control problems particularly in a multi-service environment, because of the magnitude and the bursty nature of their bandwidth requirements. Traffic control is essential not only to maintain a given performance requirement to the VBR video both in terms of cell loss and cell delay, but more so, to guarantee CBO services a very high service performance - near negligible cell loss and minimum cell delay - due to the sensitivity of these services to the two parameters. Classical traffic control strategies are inadequate in view of the bursty and non-deterministic nature of the load when VBR video and data traffic are present. The most important objective in devising traffic control mechanisms must be to prevent large bursts of VBR traffic from disrupting the CBO services, without inflicting a large cell loss on the VBR services, and at the same time maintaining a high network utilisation factor.

Several control strategies for VBR video have been proposed^[17,52,53] and research is continuing in this area. The suitability of a strategy, to a large extent, depends on the architecture of the ATM network being considered. For instance, the Orwell Ring has some control mechanisms embedded in the protocol which enable some sophisticated controls to be implemented.

Traffic control for VBR video can be implemented at two levels: call control level and cell transfer level.

2.3.3.1 Call Control Level

A call control mechanism accepts or rejects a new call request based on the status of the switch (or link) capacity and the bandwidth requirement of the new call. A request would be rejected if the acceptance of that call would degrade the performance of any existing connections below the specified requirements. The principle of this control mechanism may be simple, but it is difficult to apply in the presence of bursty traffic such as that generated by VBR video. This is because the non-deterministic characteristics of the VBR video sources make their bandwidth requirements difficult to estimate.

A call request should be accompanied by some indications of the characteristics of the call. In the case of VBR video, these may include an estimate of the mean, variance, and perhaps some measure of the burstiness, of the source data rate. With these parameters, the call control mechanism will have to determine the bandwidth required to support the new call for a specific service performance, taking into account the effects of statistical multiplexing. However, it will be difficult for a source to specify in advance the anticipated signal characteristics that the network must cater for, since these will be scene dependent. Incorrect specifications of the call characteristics can result in incorrect call control, and this can lead to low network utilisation or overloading. In the latter case, traffic control at the cell transfer level is imperative.

In switches where call control is performed on a link-by-link basis and where bandwidth allocation is static, the effectiveness of the call control mechanism will rely heavily on the accurate estimate of the total bandwidth requirement of the calls in progress on each link. Since the total bandwidth allocated on each link cannot be verified with the actual load carried, mismatches will result in incorrect call control with the consequences mentioned earlier. Call control is more adaptive with the mechanism used on the Orwell Ring, in that the unused bandwidth is estimated from some measurements of the ring load, thus alleviating the dependency on bandwidth requirement estimates. This mechanism, however, suffers from the difficulty in estimating the unused bandwidth on the ring accurately because of the bursty nature of the load and the dynamics of call establishments and terminations.

It is obvious from the above discussion that call control alone will not be sufficient to maintain the required service performance for the different services. Consequently, control on the cell transfer level is required to complement the call control mechanism.

2.3.3.2 Cell Transfer Level

Control at this level is very important if the performance criteria for some of the more critical services are to be met. It was mentioned earlier that CBO services and the lower layers of the layered coding model must be guaranteed negligible cell loss; for delay critical services, minimum cell delay must also be guaranteed. In a multi-service network, this can be achieved partially by allowing the cells from these services to take precedence in network access. A priority control mechanism is required where the different services would be assigned different priority queues according to their specific requirements. This mechanism must be coupled with an efficient priority access control algorithm to prevent low priority services from experiencing excessive cell delay and cell loss. With this mechanism, high priority services can be guaranteed to some extent, the required service performance. However, overloading caused by incorrect estimation during the call set-up stage, or by a large burst in the VBR video traffic, could still cause problems.

A second control mechanism, known as policing, can be used to alleviate the problem of a VBR video connection exceeding its bandwidth allocation as a result of the call not conforming to the pre-specified call characteristics^[53]. This mechanism compares the actual cell generating rate of the source against the pre-specified value, and if the latter is exceeded, the appropriate corrective action can be taken: the excess cell generated can either be simply discarded, or marked with a lower priority and be discarded only when the buffer is full. The first method is less efficient as cells will be discarded unnecessarily even when the network is not heavily loaded. The second method is more efficient but is more difficult to implement. Alternatively, the video source could be forced to reduce its output bit rate to conform to the pre-specified characteristics, thus resulting in a reduced picture quality. The problem with the policing mechanism lies in the conformance verification of the call characteristics.

A different mechanism can be used in place of policing, where the video sources are allocated bandwidth to cater for the needs as they arise, so long as they do not cause the network to overload. This mechanism requires the resource allocation controller to keep track of the state of the network and the output cell rate of the sources instant by instant. It can be very effective in preventing overload by denying VBR video connections further allocation if overload is eminent, while allowing network resources to be utilised efficiently at all times. If overload occurs due, for instance, to the loss of some of the network capacity because of partial system failure, the resource allocation controllers can remove bandwidth allocation from the VBR video, if necessary, in order to protect the high priority services. This mechanism is however more complex, and difficult to implement in most networks, but it can be implemented quite easily on an Orwell Ring since load monitoring and dynamic resource allocation functions are integral parts of the protocol.

In view of the difficulties in controlling VBR video traffic, particularly in a multi-service network, it has been suggested that a hybrid approach be adopted where part of the resources are dedicated to CBO services and the rest to VBR services. This simplifies the problem as the two classes of services can be controlled separately, but this approach also diminishes some of the advantages of ATM networks.

2.4 Summary

In this chapter, a number of aspects related to the transmission of video over ATM networks have been addressed. The need for the efficient coding of video signals was first established as a measure for transmission costs saving, which is crucial for the widespread acceptance of video-communication services. The effects of applying coding to video signals were then examined in the context of some commonly used coding algorithms, with the conclusion that the efficient coding of video signals generally leads to variable bit rate outputs, as a consequence of the fact that information rate varies with the spatial and temporal detail of the picture; hence VBR video.

The advantages and problems of supporting packetised VBR video in an ATM network environment were then discussed. Layered coding, which codes video signals into hierarchical information, was examined as a likely solution to error concealment in the event of cell loss. Error concealment is particularly important for VBR video where cell loss may occur in bursts. Finally some of the proposed traffic control strategies for VBR video were outlined.

Chapter 3:

VBR Video Signal Statistics

3.1 Introduction

VBR video promises a potential for more efficient network resource utilisation with a constant picture quality and a less complex codec design. One video-communication service that can benefit most from these advantages of VBR video is videophone, where the subject material is mainly head and shoulders type pictures, and where all the video sources are uncorrelated. This chapter describes the study of some aspects of the signal statistics of such a video source. Although source statistics have been looked at by some workers in the past^[7,27,28,37,51], the results were neither in a suitable form nor adequate for the purpose of this work. The more recent studies in this area have been conducted with the prospect of a broadband VBR carrier and the potential gain that can be achieved with VBR video in mind.

The main objective of this study is to acquire a good understanding of the video source behaviour, especially the videophone and video-conferencing type sources; with this, the relative merits of VBR video can be assessed and a source model developed to aid in the study of the transmission aspects of VBR video. A sound knowledge of the source behaviour is also required for the practical implementation of VBR video, for instance, in dimensioning network capacity and in formulating load control strategies.

In this chapter, the experimental arrangement and the codec with its modifications will first be described along with the picture source used. The results on the statistics of frame differences (FD), cluster length and the output data rate of the codec will be presented in section 3.3, followed by a conclusion on this study. Most of this work has been published in two papers^[11,12] and presented at a workshop^[10] recently.

3.2 Experimental Arrangement

Figure 3.1 shows the experimental arrangement used in this study. A commercial CBO codec was modified to work as a VBR video codec, and a video disc player was used as the picture source input. The codec was connected via an interfacing board to a BBC Master computer with an ACORN RISC second processor (ARM), and Winchester and floppy disc drives for data storage. The whole system was linked to a remote mainframe for data processing.

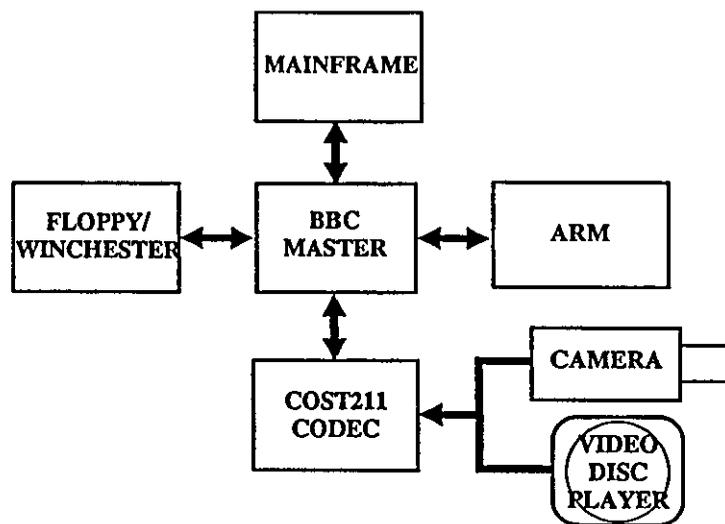


Fig. 3.1 Experimental Arrangement.

3.2.1 The Codec Used in Experiments

The experimental codec was based on a prototype BT/GEC COST211 video-conferencing codec^[8,44], which has a resolution of 286 lines in two interlaced fields and 255 pixels horizontally. Chrominance resolution is reduced in the ratio of 2:1 in the vertical direction and 5:1 in the horizontal direction. The full frame rate of 25 frames/s is used in the codec and pixels are represented using 8 bits (but only 224 levels are used for coding); without compression, these will generate a data rate of 17.5 Mb/s. Although the codec is a colour system, only the luminance data was used in this study as it was considered that the inclusion of the chrominance data would not alter the results significantly^[39], and preliminary investigations confirmed this assumption.

Figure 3.2 shows the functional block diagram of the modified codec. Normally the codec operates as a CBO system with a fixed output of 2 Mb/s. A VBR source was obtained from this codec by opening the feedback control loop from the buffer, which normally

constrains the output to a constant bit stream by varying the coding parameters, performing element or field subsampling and even stopping further coding, thus constantly varying the picture quality.

A conditional replenishment coding algorithm is used in the codec to detect moving, or 'changed', pixels in the incoming video frame with respect to a reference frame. The reference frame keeps track of the received frame at the remote codec assuming no loss of data. The output of the conditional replenishment coder is then DPCM coded using non-adaptive two dimensional intra-frame prediction; the prediction errors are quantised and Huffman coded with a maximum word length of 10 bits. In this study, it is not only the encoded output that is of interest but also the raw PCM statistics, and hence a further modification was made to enable the codec to operate in PCM mode as shown in figure 3.2.

Before being fed to the movement detector, the input signal is passed through a non-linear temporal filter which has the effect of cleaning up very small changes due to noise, thus improving the performance of the movement detector as well as reducing source entropy⁽⁴⁴⁾.

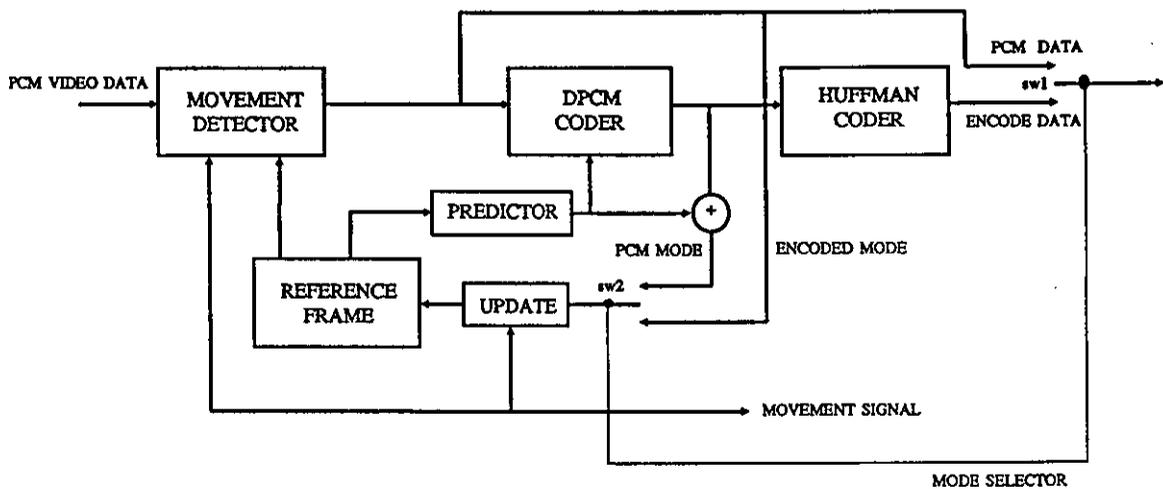


Fig. 3.2 Block Diagram of the Modified Codec Functions.

A block diagram of the movement detector is shown in figure 3.3. The movement detector subtracts the output of the temporal filter from the corresponding pixel in the reference frame and makes use of the sum of 5 consecutive absolute pixel differences, weighted according to the expression:

$$S_i = w_2 \cdot D_{i-2} + w_1 \cdot D_{i-1} + w_0 \cdot D_i + w_1 \cdot D_{i+1} + w_2 \cdot D_{i+2} \quad \dots(3.1)$$

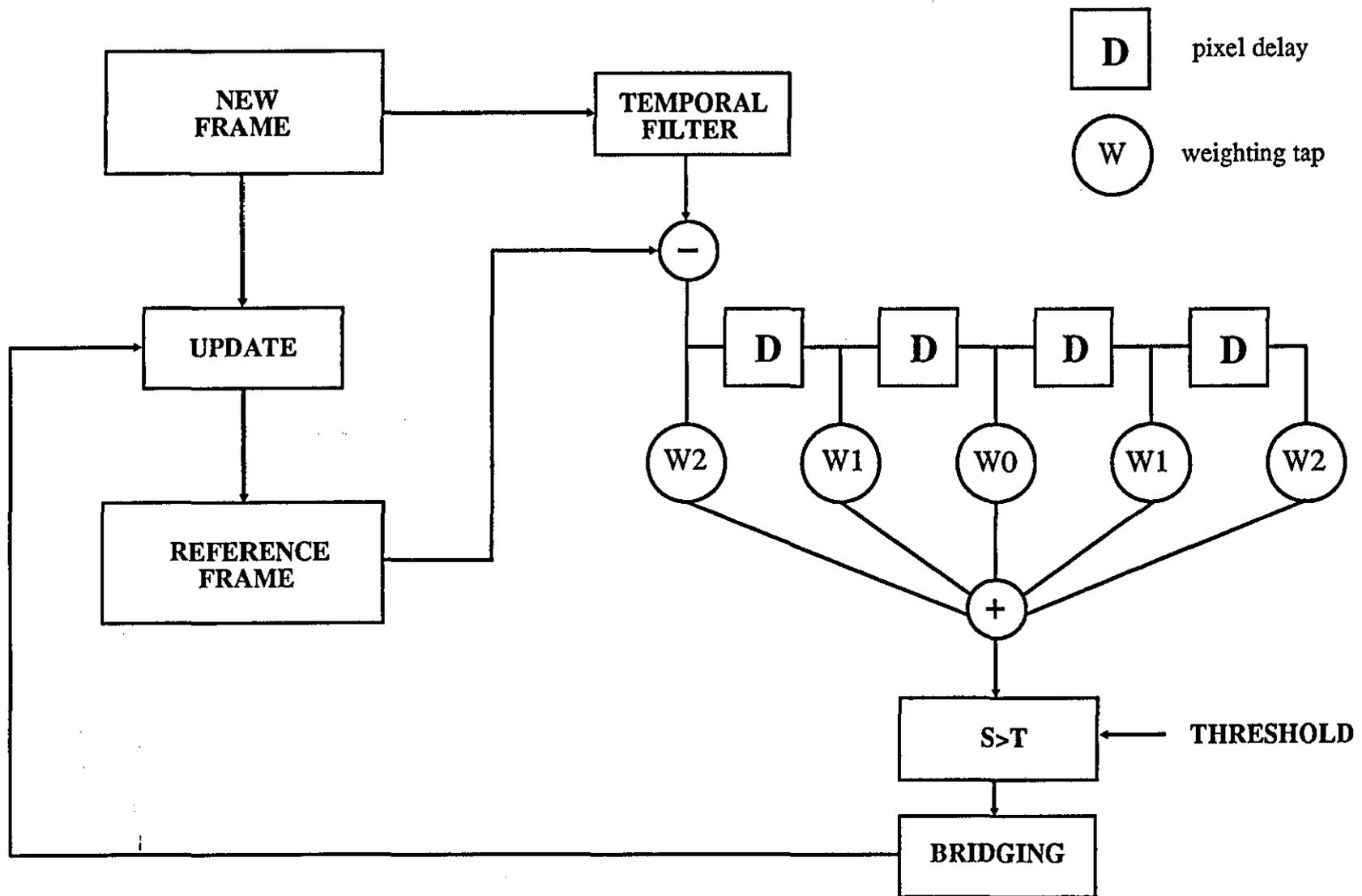


Fig. 3.3 A schematic representation of the COST211 codec movement detector.

The sum S_i is then compared against a threshold to determine if the current pixel has changed. The use of the 5-tap movement detector helps to reduce the detection of false movement. The changed pixel is bridged into an adjacent cluster of changed pixels if it is separated from the previous changed pixel by an amount less than or equal to a pre-specified value (in this study the value used is normally 3), otherwise it forms a new cluster. A cluster length of 1 pixel is allowed, provided of course that it is separated sufficiently from its neighbours. These clusters constitute moving areas and cause the reference frame to be updated accordingly, either directly with PCM data in the PCM mode, or with the decoded DPCM data if subsequent coding is employed.

3.2.2 Threshold Selection

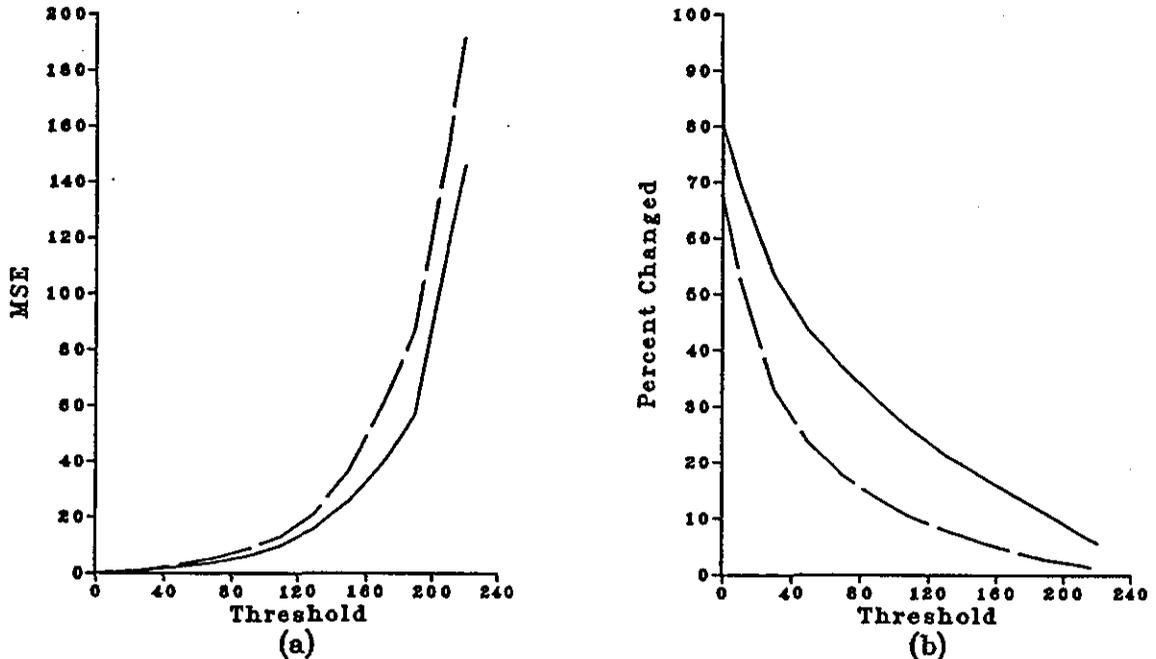
With the feedback control loop disabled, the threshold value must be held constant, and it must be chosen to give a reasonable picture quality. The picture quality will then be fixed but the output data will vary according to the amount of movement in the picture.

Experiments were carried out to investigate the effects of varying the threshold using a computer model of the codec developed on the ARM second processor and picture sequences captured from the actual codec. The model only operates in PCM mode to reduce processing time. Two non-visual tests, namely the mean square error (MSE) between the actual pictures and the reconstructed pictures, and the amount of change generated, were carried out on a head and shoulders sequence. The results are shown in figures 3.4(a)-(b). Only two results were plotted on each graph representing the extreme cases. It can be seen that a good compromise in the choice of the threshold would be in the range 80-150. Unfortunately, MSE is not a particularly suitable criterion for measuring the quality of pictures, and is used here only to measure the amount of error in the reconstructed pictures arising from the movement detection process.

The final choice of the threshold in this study was based on visual inspection of the performance of the actual codec while the threshold value was varied over the range determined earlier, particularly with scenes involving camera zooming and panning. The threshold chosen was the maximum value used, above which edges in the pictures began to break up and leave 'dirty windows'. A lower threshold would yield a better picture but also more change. The chosen threshold of 110 is a compromise between minimum change and acceptable picture quality, and no significant degradation in picture quality was observed.

However, it should be noted that the choice of 110 is slightly on the high side. In a separate experiment, the movement detector was modified to use only a single unweighted pixel difference in its movement detection process and a threshold of 7, which is about 3% of

the maximum pixel value. It was found to generate roughly the same amount of change as the 5-tap movement detector with a threshold of 110. In most other work^[27,28,37], a much lower threshold of 1.5% of the maximum pixel value was used to obtain a high quality picture.



Figs.3.4 Variation of (a) the mean square error (MSE) of the reconstructed frame and (b) the amount of frame-to-frame differences generated, with different thresholds.

3.2.3 Picture Source

The picture source used was a Pioneer Laser Disc Player which outputs a 625 lines PAL signal. The discs used were recordings of documentary programmes produced by the BBC^[V1,V2,V3]. The picture material contained high background detail, and some degree of involuntary camera movement was noticed.

3.3 Experimental Results

Frame differences and output data rate can be obtained from the codec via an interfacing board which taps into the appropriate signal lines on the codec and performs basic statistic collection functions; it also enables statistics on cluster characteristics to be obtained. The length of the sequences were only limited by the length of suitable picture material. A variety of picture material was used, ranging from head and shoulders to waist-up scenes, and a combination of these with scene changes and multiple characters. Snapshots for some of the sequences are shown in Appendix I.

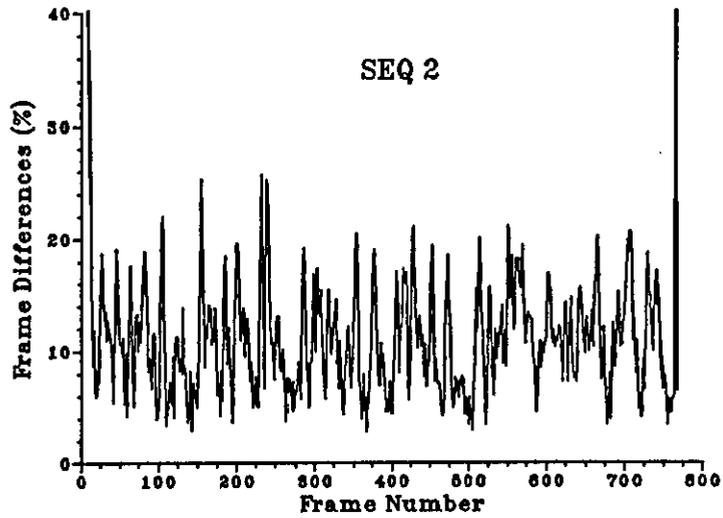
3.3.1 Statistics of Frame Differences

The term 'frame differences (FD)', strictly speaking, refers to the proportion of pixels in a frame that are deemed to have changed since the previous frame. However, in this experiment, it also includes those pixels that have been bridged into the clusters and which otherwise would not have changed. This is not expected to increase the actual FD by much. In fact, investigations carried out using the computer model of the codec indicated that the increase in the amount of change was no more than 1.5%. Furthermore, in most practical inter-frame codecs not using block coding, changed pixels are bridged into clusters to improve subsequent coding and transmission efficiencies. It was therefore considered more appropriate to examine the bridged FD rather than the actual FD. (Note that in block coding, not all pixels within a block must change before the block is deemed changed; this is to some extent similar to bridging).

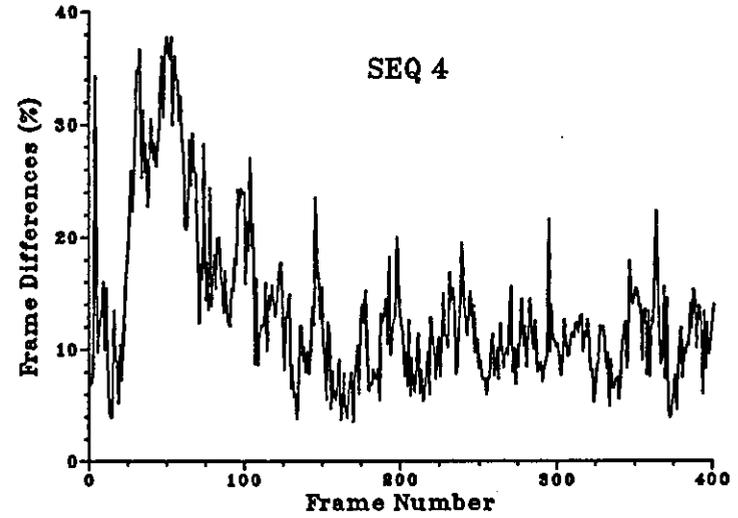
Frame differences statistics were studied under two different conditions. Firstly, without the influence of subsequent coding, i.e. in PCM mode, where the output of the movement detector was fed directly back into the reference frame; and secondly, when the output of the movement detector was DPCM coded and decoded before being fed back into the reference frame as in figure 3.2 (for convenience, this will be referred to as the encoded mode). The purpose of studying FD statistics under PCM mode is to obtain some general results which are not dependent on any specific coding algorithms other than conditional replenishment. The FD obtained under this condition represent the amount of raw information that needs to be transmitted; the actual coded bit rate would obviously depend on the subsequent coding algorithms used. If some knowledge about the characteristics of the coding algorithms used is available, then the coded bit rate can be approximated from the FD. However, subsequent coding will introduce more errors into the reference frame and result in a higher FD than in the PCM case. The magnitude of this increase will depend on the coding algorithms used and the amount of degradation permitted, but the profile of the variation of the FD with time is not expected to be affected significantly.

3.3.1.1 PCM Mode

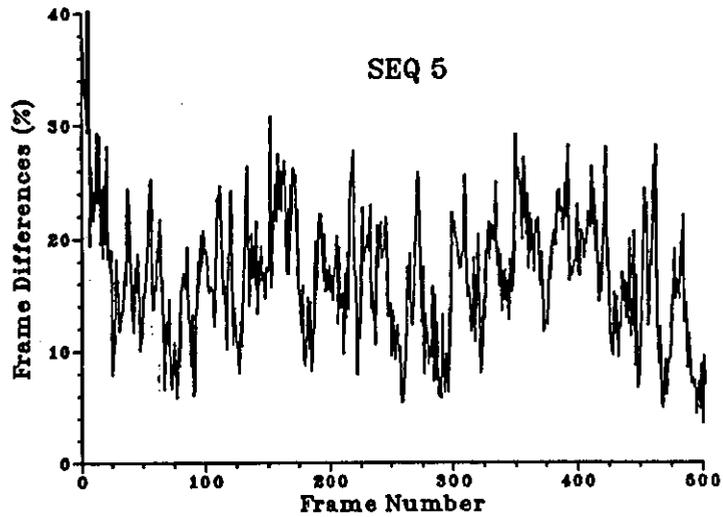
The graphs in figure 3.5 show how, for four different sequences, the frame differences vary with time (note that the peaks have been cut off at 40% change to give a better scaling). It is observed that the variation in FD is very significant, revealing the fundamental VBR nature of video in terms of transmission requirement. It also highlights the difficulties and inefficiency in trying to fit this data into a CBO channel. Large changes in FD are usually associated with sudden scene changes and generally last for no more than a few frames before settling down; those generated by large bursts of motion may have longer durations as can be seen in figure 3.5(b).



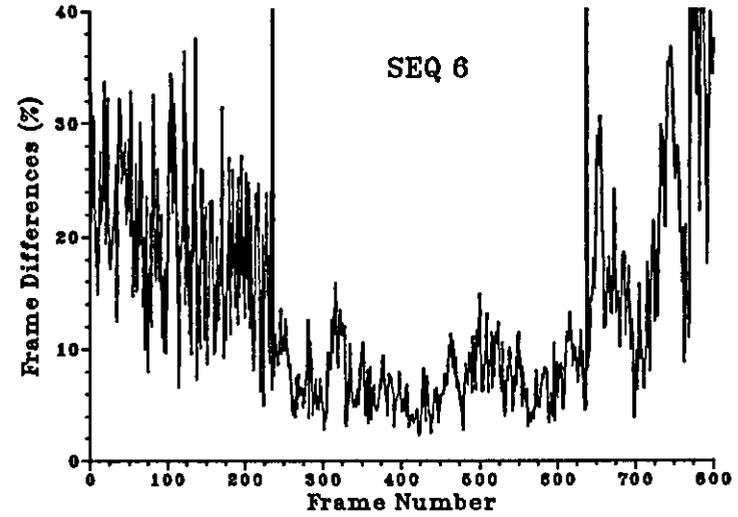
(a)



(b)



(c)



(d)

Fig(s). 3.5 (a)-(d) Variation of Frame Differences with time for 4 different sequences (25 frames/s).

Table 3.1 summarises the measurements on the FD for 7 sequences. If scene changes occur in the sequence, as in sequences 2 and 6, the peak-to-mean ratio (PMR) is about 5, but if there are no major scene changes as in the other sequences, the ratio is about 2-3. Since scene changes would have to be allowed for in normal operation, a PMR of about 5 would be a more probable figure. This would represent, in a simplistic view, the possible bandwidth gain in VBR video over CBO video, assuming that the peak demand must be catered for by the CBO codec. However, in practice, a CBO codec would make use of the reduced sensitivity of human eyes to spatial detail during scene changes to reduce the peak drastically, without causing any observable degradation in the picture quality^[28]. Under such circumstances, the gain inferred from the PMR with scene changes would not be valid. A gain of 2-3 is then probably more realistic for the sequences used. This gain could be further eroded if the network has to be lightly loaded in order to achieve some given performance requirements.

PMR is however not necessarily a good measure, nor should it be the only criterion for comparing VBR and CBO video, since it fails to take into account such factors as the duration of the peaks and the amount of FD variation. It also fails to consider the limitations of CBO video that lead to inefficiency and less superior picture quality.

An examination of the results shows that peaks caused by large bursts of motion in the sequences could last for an appreciable part of a second or even longer, and unless the channel can cope with the expected peak output of the codec, such video bursts would cause data loss, or picture quality would have to be reduced to constrain the output data rate. The duration of the bursts makes the use of a smoothing buffer impractical as this would need to be excessively large and could result in unacceptable delay. The large variances of the FD illustrate the difficulties in attempting to smooth out the FD variation. While the large FD variation suggests the need for a complex feedback control mechanism to adjust the coding parameters (and therefore the picture quality) continuously in order to maintain a stable output. The different characteristics of the individual video sources also make it difficult and inefficient to allocate a fixed bit rate channel to the video sources, since the optimum bit rate requirement for each source will be different. As a result, the source characteristics and the CBO channel allocated, which normally corresponds to a level of the transmission hierarchy, are usually mismatched, and the consequence is gross inefficiency.

The last column in Table 3.1 and the graph in figure 3.6 show the result of multiplexing the 7 sequences used earlier. This was achieved by adding up the FD of all the sequences on a frame by frame basis, and expressing the result on a *per source* basis. The interesting aspects of this graph are that the variance of the FD is very much smaller and the PMR is considerably reduced. The peak at the beginning of the multiplexed sequence results from the fact that all the individual sequences start with a peak, and this would be the case if large

Table 3.1 Statistical Measurements on Picture Sequences

Number of samples / sequence : 395

Frame Differences are quoted in percentages

Type : *WP* - waist up *HS* - head and shoulders *VR* - variety

| Sequence Number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | * <i>COM</i> |
|---------------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|--------------|
| Type of Picture Material | <i>WP</i> | <i>HS</i> | <i>HS</i> | <i>WP</i> | <i>HS</i> | <i>VR</i> | <i>HS</i> | <i>VR</i> |
| Maximum Frame Differences | 28 | 54 | 23 | 38 | 31 | 72 | 16 | 21 |
| Mean Frame Differences | 8.67 | 11.04 | 8.72 | 13.43 | 16.83 | 14.60 | 6.89 | 11.45 |
| Variance | 11.33 | 36.29 | 15.35 | 51.52 | 27.06 | 73.13 | 9.65 | 6.09 |
| Standard Deviation | 3.36 | 6.02 | 3.92 | 7.18 | 5.20 | 8.55 | 3.11 | 2.47 |
| Peak-to-mean Ratio | 3.23 | 4.89 | 2.64 | 2.83 | 1.84 | 4.93 | 2.25 | 1.84 |

**COM* - a combination of all sequences

peaks coincide with one another. This situation is, however, highly unlikely if the sources are uncorrelated. The smoother profile of the FD variation of the multiplexed sequence is advantageous as this would make traffic control easier on an ATM network.

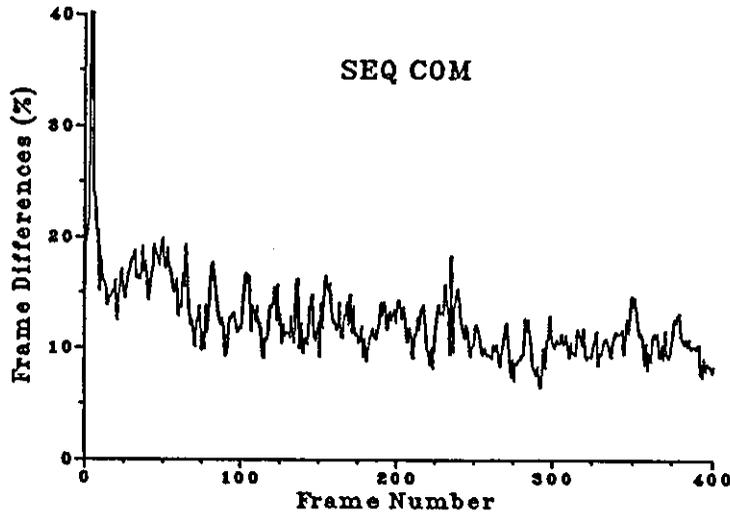


Fig.3.6 Variation of the frame differences with time for the multiplexed sequence (25 frames/s).

It is also worth noting that the mean FD of the multiplexed sequence is simply the average of the mean FD of the individual sequences. The resultant variance σ_m^2 is

$$\sigma_m^2 = \frac{1}{N} \sum_{i=1}^N \sigma_i^2 \quad \dots(3.2)$$

where σ_i^2 are the variances of the FD of the component sources.

This implies that the resultant variance is reduced by a factor of N. Assuming that the mean of the variances of the component sequences stays approximately constant, the resultant variance will diminish quickly with an increase in the number of component sources.

These explain the smoother sequence in figure 3.6. Variances calculated using equation 3.2 with different combinations of the sequences have been found to approximate those obtained experimentally.

Statistically multiplexing the outputs of several uncorrelated VBR video sources results in an overall smaller bandwidth per source being required for a given performance criterion, since the probability of all sources having their maximum output at the same time is extremely remote. On the other hand, the bandwidth required per source for a CBO system

does not alter as the number of sources are increased. The ability to share the available bandwidth is one of the main advantages of VBR video, and the gain would be better if the outputs of more uncorrelated sources were multiplexed^[27,54].

3.3.1.2 Encoded Mode

As a result of the DPCM coding process used in this mode, the reference frame of the encoder contains not only errors from the movement detection process but also those from the coding. However, these additional differences are small, thus the FD generated show almost identical patterns of variation to those generated in PCM mode - a difference of no more than 1%-1.5% on average. The plots of the variation of FD with time are not shown since they are virtually the same as those shown in figure 3.5.

Table 3.2 compares the FD statistics for the PCM mode with those for the encoded mode, and as can be seen, there is not much difference between the two. The slight decrease in the PMR in the encoded mode when scene changes were included was due to the increase in the sequence means, while the peak FD remained almost unchanged - a consequence of the fact that when a peak occurs, almost all the pixels in a frame have changed, and the small errors introduced by DPCM coding become insignificant.

3.3.2 Statistics of the Output Bit Rate

In this section, the statistics of the data rate at the output of the coding pipeline which comprises conditional replenishment, DPCM and Huffman coding algorithms were examined. The statistics of the data rate after addressing overhead had been added were also obtained. These results will be specific to the COST211 codec and codecs employing a similar coding strategy; their implications on codecs using other coding strategies, especially those with very different characteristics, need to be interpreted with great care. Data rate was measured in bits per frame and was assumed to be constant within a frame.

The statistics of the output data rate are presented in Table 3.3. It is clear from the table that the PMR for the encoded bit rate without addressing information is reduced compared to the PMR for the frame differences. This is a direct result of the cluster length characteristics. Since the first pixel of every cluster is always transmitted as a PCM pixel, it makes the coding of short clusters very inefficient as these PCM overheads form a significant portion of the compressed cluster data. The coding efficiency is thus not constant but increases with cluster length, and therefore with the amount of change in the frame as can be deduced from figure 3.13 (section 3.3.4). This characteristic results in a proportionally higher mean and lower peak, and thus a lower PMR; it is also expected to produce a smaller variance than the FD. This is indeed the case if a constant coding performance equivalent to the mean coding performance of the sequences is assumed, and this can be deduced from Tables 3.2 and 3.3 to

| | SEQ 2 | | SEQ 4 | |
|-----------|---------|-------|---------|-------|
| | Encoded | PCM | Encoded | PCM |
| Mean | 11.78 | 10.91 | 15.50 | 14.11 |
| Std. Dev. | 5.27 | 5.20 | 10.63 | 10.62 |
| Peak | 95.29 | 95.25 | 92.79 | 92.99 |
| PMR | 8.09 | 8.73 | 5.99 | 6.59 |

Table 3.2 : Statistical measurements of frame differences for PCM and encoded data. (figures quoted in percentage of frame changed)

| | SEQ 2 | | SEQ 4 | |
|-----------|-----------|---------|-----------|---------|
| | Data only | + Addr. | Data only | + Addr. |
| Mean | 30.95 | 53.04 | 42.34 | 68.24 |
| Std. Dev. | 10.95 | 12.88 | 19.97 | 23.01 |
| Peak | 202.40 | 214.30 | 163.79 | 177.57 |
| PMR | 6.54 | 4.04 | 3.87 | 2.60 |

Table 3.3 : Statistical measurements of data generated per frame for picture data only and with addressing added. (figures quoted in kbits/frame)

| | SEQ 2 | | SEQ 4 | |
|-----------|---------|-------|---------|-------|
| | Encoded | PCM | Encoded | PCM |
| Mean | 11.64 | 10.83 | 14.58 | 13.16 |
| Std. Dev. | 4.33 | 4.24 | 7.21 | 6.97 |
| Peak | 25.79 | 24.37 | 48.73 | 46.14 |
| PMR | 2.21 | 2.25 | 3.34 | 3.50 |

Table 3.4 : Statistical measurements of frame differences for PCM and encoded data (without scene change). (figures quoted in percentage of frame changed)

| | SEQ 2 | | SEQ 4 | |
|-----------|-----------|---------|-----------|---------|
| | Data only | + Addr. | Data only | + Addr. |
| Mean | 30.45 | 52.47 | 41.12 | 67.16 |
| Std. Dev. | 8.78 | 11.13 | 16.30 | 20.36 |
| Peak | 58.17 | 99.70 | 98.36 | 129.66 |
| PMR | 1.91 | 1.90 | 2.39 | 1.93 |

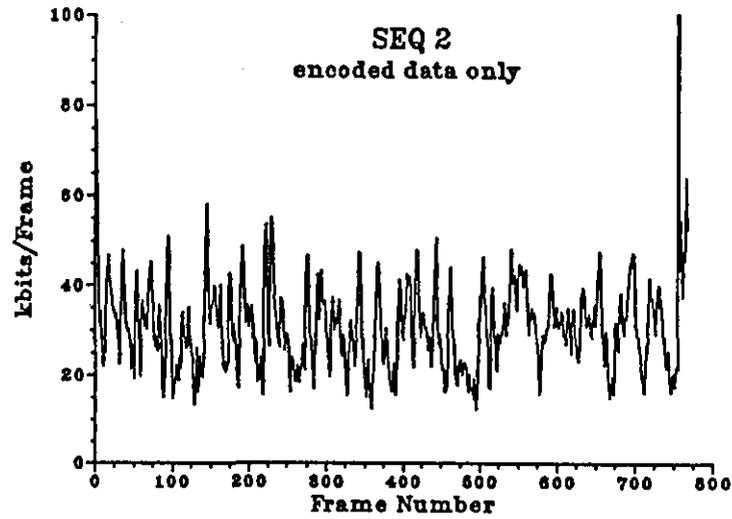
Table 3.5 : Statistical measurements of data generated per frame for picture data only and with addressing added (without scene change). (figures quoted in kbits/frame)

be in the range of 3.6-3.75 bits/pixel; the estimated standard deviations for the FD with scene changes included were 13.84 and 29.07 kbits respectively for the two sequences, as compared to 10.95 and 19.97 kbits of the measured values for the data streams. These results also have some implications on the underlying probability distribution of the output bit rate when compared with the FD distribution (see section 3.3.3).

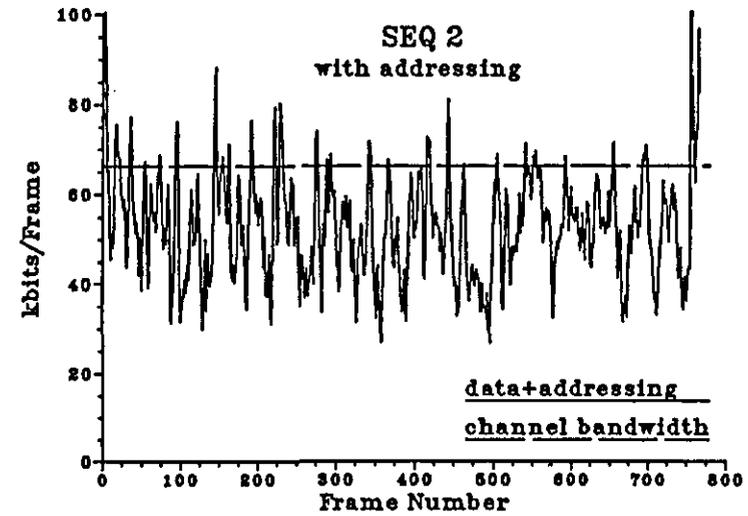
When the addressing information used by the COST211 codec (see Appendix II for its description) is included in the output data rate, the PMR is further reduced (as shown in Table 3.3) for a similar reason as above, but with addressing rather than the PCM pixel as the overhead. Since the size of the addressing overhead is fixed for every cluster irrespective of its length, the addressing efficiency, as with coding efficiency, is not constant but increases with the cluster length and hence the amount of FD. The PMR, as can be seen in Tables 3.2 and 3.3, is halved by the introduction of coding and addressing, but is nevertheless, still significant. The standard deviation indicates a fluctuation of about 300-500 kb/s in the instantaneous bit rate, which could still cause problems in buffering in a CBO system.

For completeness, results similar to those obtained above but excluding any scene changes are presented in Tables 3.4 and 3.5. These results account for cases where full resolution is not considered necessary for scene changes. Similar observations are derived from these sets of results, i.e. a continuous reduction in the measured PMR, but less drastic in this case. The reason being the relatively smaller peaks resulted in only a small increase in the cluster length, and hence only in a marginal improvement in the coding and addressing efficiencies. Consequently, the peaks and the means are affected almost equally by the coding process and the addition of the addressing information, the PMR therefore remains fairly constant. Overall, a gain factor of about 2 can be achieved with this system. However, PMR, as has been stressed earlier, is inadequate and should not be the only basis for assessing the gain of VBR video.

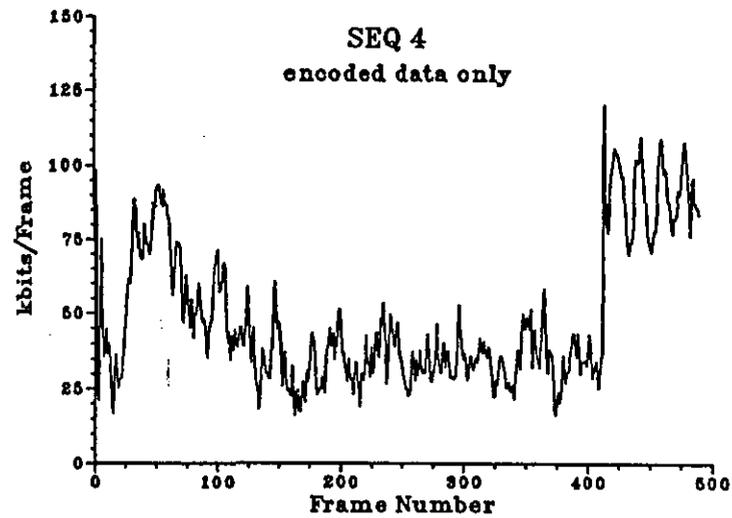
Figures 3.7(a)-(d) show how the number of bits per frame for the coded sequences, with and without addressing overhead, varies with time. As can be expected, the variation of bits per frame follows an almost identical pattern to that of the FD. It is worth noting that the burst of data at the end of sequence 4 has lower peaks relative to the rest of the sequence when compared to those encountered in its FD counterpart, and this illustrates the high coding and addressing efficiencies associated with these large changes. The horizontal line in figures 3.7(b) and (d) corresponds approximately to the bandwidth allocated for the transmission of luminance data in the COST211 codec, which is about 1.65 Mb/s. It is obvious that the allocated bandwidth and the bandwidth required by the two sequences are mismatched in different ways; the allocated bandwidth is largely under-utilised in the case of sequence 2, which has a smaller mean bit rate than the channel (~ 0.8 utilisation factor), and yet there are occasions when the channel is unable to cope. Therefore, despite the excess capacity, some data reduction mechanisms may be required to constrain these peaks.



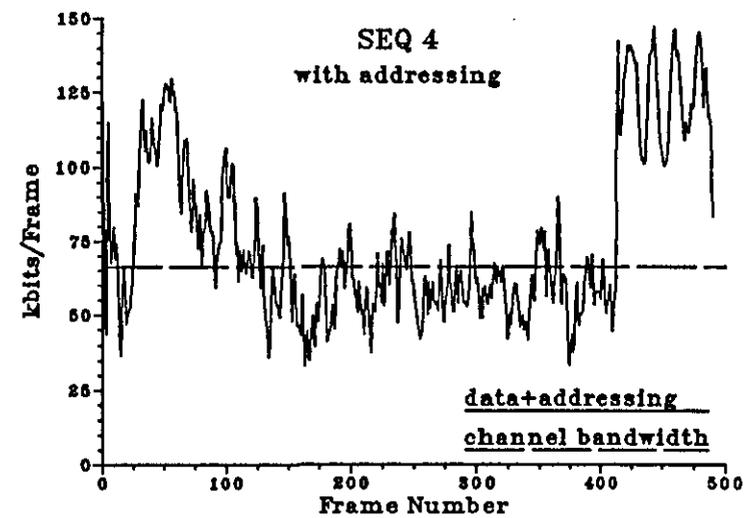
(a)



(b)



(c)



(d)

Fig(s). 3.7 (a)-(d) Variation of the codec output bit rate with time - with and without addressing overhead (25 frames/s).

Figure 3.8 indicates that the number of continuous frames exceeding the channel bandwidth, or video-spurts, could last up to 7 frames. Simple buffering techniques may be able to reduce these video-spurts in this case. On the other hand, the channel is hardly adequate for supporting sequence 4, with the channel failing to meet demand for 40% of the time, and video-spurts of up to 67 frames are present as shown in figure 3.8 (the large data burst at the end of the sequence has been excluded since it is from unsuitable picture material). This will necessitate the use of a higher movement detector threshold or subsampling in order to comply with the channel bandwidth at the expense of picture quality.

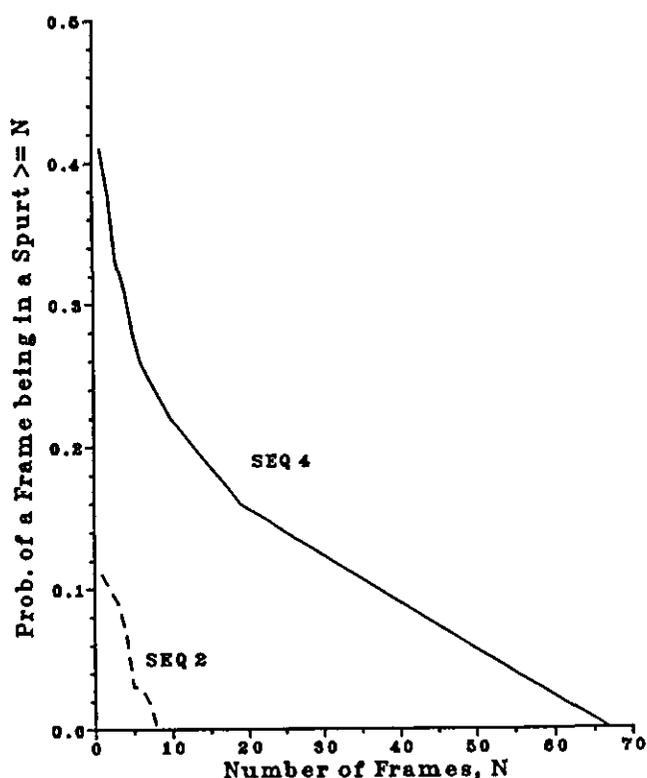


Fig.3.8 Video-Spurts (with spurt threshold = 66000 bits/frame).

These results reveal a very serious weakness of the CBO system. Clearly it is not possible to find a single fixed bandwidth channel that could satisfy the widely differing requirements of the different video sources shown in Table 3.1. This mismatch between source requirement and channel capacity translates into inefficiency both in terms of resource utilisation and service provision as exemplified by the two cases above. A higher efficiency can be achieved with VBR video; this, together with the redundancy of a complex feedback control mechanism, make VBR video an attractive way of transmitting video information.

It is also interesting to note that the addressing information on average occupies 35%-40% of the total data, but it represents only 6%-8% at the peaks. When the amount of change in the frame is very small, the addressing overhead could rise to 50%-60% as a result of the presence of a large number of short clusters. Similar observations can be expected on the effects of the PCM pixel at the beginning of a cluster. This problem of poor coding and addressing performance will be examined later in section 3.3.4 from the perspective of the cluster length.

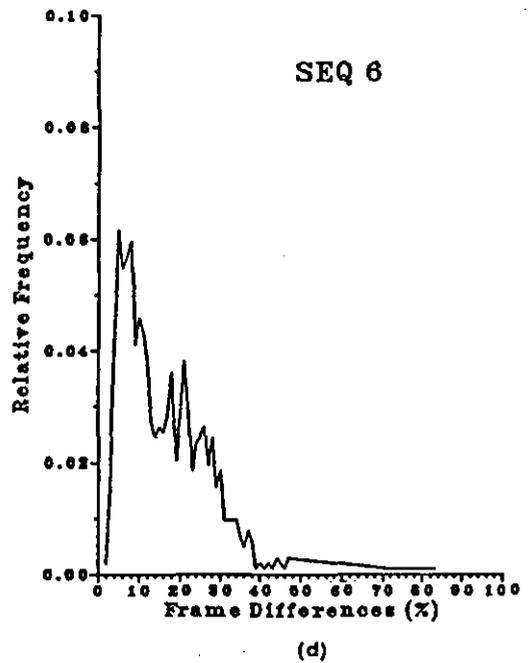
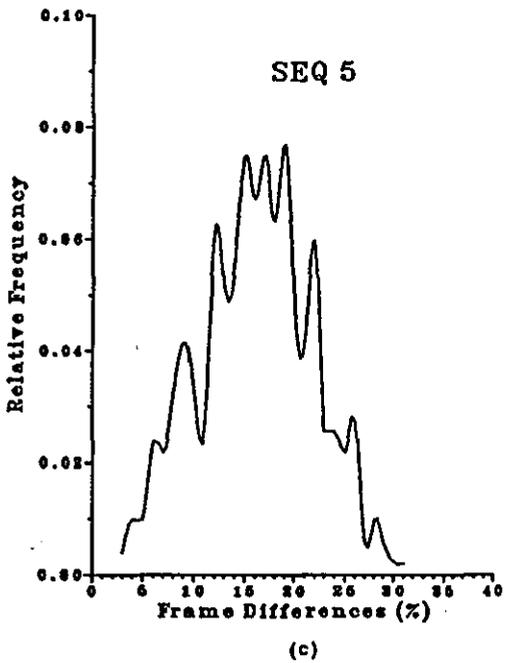
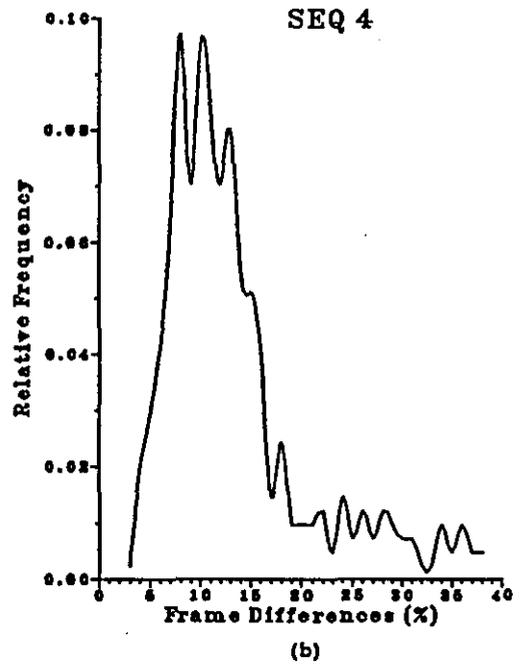
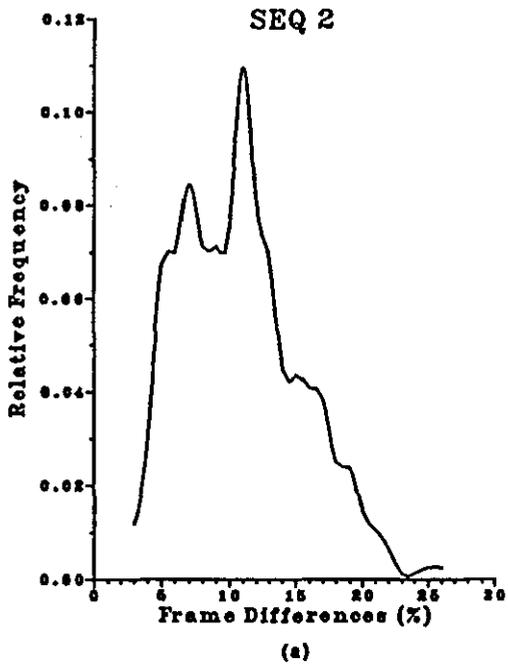
3.3.3 Probability Density Function (PDF) of Frame Differences and Output Bit Rate

The data collected on the frame differences and the output bit rate were further analysed to study their underlying PDFs. The effects which statistical multiplexing, DPCM and Huffman coding as well as addressing have on the PDFs, were also investigated.

3.3.3.1 PDF of Frame Differences

Frame differences, as seen in the last section, were almost unaffected by DPCM coding, thus only data collected in the PCM mode was studied here. However, the results are equally applicable to the FD in encoded mode. Since the PDF of FD had been studied in detail by Seyler⁽⁵¹⁾, the exact nature of the PDF for each of the sequences used was not determined. Instead they will be discussed in the context of the results obtained in the reference.

The PDFs of the FD were computed for all the sequences used, four of which are shown in figures 3.9(a)-(d). These results resemble closely those PDFs presented in the reference in exhibiting the same general shapes. On inspection, these PDFs appear to possess a mixed feature of both the asymmetric Gamma distribution and the symmetric Normal distribution: they are generally skewed to the right to a varying extent, which is a characteristic of the Gamma distribution, while the sharp peaks present in some sequences are a distinct feature of the Normal distribution. Multi-modality is clearly evident in some cases. Sometimes a clear distinction between the two distributions for a given sequence can be difficult to ascertain since the Normal distribution is a special case of the Gamma distribution.



Fig(s). 3.9(a)-(d) The PDF of Frame Differences for 4 different sequences.

Assuming that the PDF of FD can be approximated by the Gamma distribution as suggested by Seyler, then the PDF of the frame differences fd is given by:

$$h(fd) = \frac{fd^{\alpha-1} \exp(-fd/\beta)}{\Gamma(\alpha)\beta^\alpha} \quad \dots(3.3)$$

where α and β are given by the computed mean (η) and standard deviation (σ) of fd such that

$$\alpha = (\eta/\sigma)^2 \quad \text{and} \quad \beta = \sigma^2/\eta$$

and for $\Gamma(\alpha)$ see Seyler⁽⁵¹⁾.

As α (the shape parameter) increases, the PDF becomes increasingly symmetrical i.e. tending to a Normal distribution⁽⁵²⁾. The observed FD PDF for a particular sequence will therefore depend on the nature of the sequence, which is characterised by the mean to standard deviation ratio, α , of the FD. For a given η , the FD PDF of a sequence will become more symmetrical as σ decreases; this corresponds roughly to a sequence with small and uniform motion.

For a videophone type picture where the motion is bursty, the FD PDF can be expected to exhibit a strongly asymmetric Gamma characteristic. But with sequences which contain little motion, the PDF may tend to be more symmetrical, thus resembling more of a Normal distribution. The PDFs of the FD are however more often multi-modal - the occurrence of a superimposing Normal characteristic, in the form of a sharp peak, on a Gamma distribution in most sequences is probably due to the motion of the objects in the sequences remaining fairly constant over some interval of time; the short-term variances are thus correspondingly small and hence a higher α . With these points established, the results presented in figures 3.9(a)-(d) will be examined more closely.

Sequences 2, 4 and 5 are all head and shoulders or waist-up type pictures, with little to moderate motion against a stationary background. The resultant PDFs, as can be expected, lie between the Gamma and the Normal distributions. Sequence 2 exhibits a strong multi-modal characteristic with a distinct peak near the sequence mean (11%), which resembles a Normal distribution on an otherwise typical Gamma distributed FD. Sequence 4 also shows similar characteristics but multi-modality is less obvious as the two peaks, more or less, overlap. The values of α for the three sequences are 1.83, 1.72 and 3.24 respectively, and sequence 5, with the largest ratio, does appear to resemble more of a Normal distribution. When examined closer, the variance of the FD for sequence 5 is found to be relatively small for a sequence with a large mean FD as can be seen in Table 3.1. Inspection of its FD variation further reveals the small burstiness of the sequence.

Sequence 6 consists of four different short sequences concatenated, each with a different characteristic, and two of which contain scenes with multiple characters and non-stationary background. These large variations in picture content resulted in a much variable FD and thus a lower α of 1.6. The resultant PDF therefore has a more pronounced Gamma characteristic with its highly skewed shape.

Curve fitting for the FD PDF of sequence 2 with a Gamma curve of approximately the same mean and variance as those measured for the sequence is shown in figure 3.10. A good fit is obtained if the multi-modality is ignored.

The observations made in this work are in accord with those made by Seyler, who suggested that the FD PDF converges to Gamma^{Distribution} when the sequence contains a large variety of scenes and motion. This is also consistent with the suggestion that the PDF becomes more asymmetrical as α decreases (or σ increases). The observations of multi-modality for head and shoulders scenes are also consistent.

3.3.3.2 PDF of Multiplexed Frame Differences

The data computed for the multiplexed sequence *per source* (shown in figure 3.6) was subjected to the same analysis. Figure 3.11 shows the PDF of the FD for this sequence. The PDF is only slightly skewed and tending towards a Normal distribution.

This observation is well in accordance with the central limit theorem (CLT)^[35], which suggests that as the outputs of a large number of uncorrelated VBR video sources are multiplexed together, the resultant output will tend to have a Normal distribution irrespective of the underlying distributions of the component sources. There may however be some deviation near the tail of the distribution. Verbiest^[54] found that with 16 sources, the agreement with the Normal distribution is within 10%, while Haskell^[27] suggested a similar number of sources to be the minimum for this effect to be observed. The result in figure 3.11 was obtained with only 7 sequences; it therefore exhibits considerable deviation from the Normal distribution, especially at the tail section.

From a different perspective, the resultant multiplexed sequence in figure 3.6 has a much smaller FD variation, and hence a smaller variance (see Table 3.1). The value of α measured for this sequence is 4.63, which is significantly higher than the four individual sequences above, and from previous discussion, it is clear that the FD PDF for such a sequence will tend to be Normal in nature.

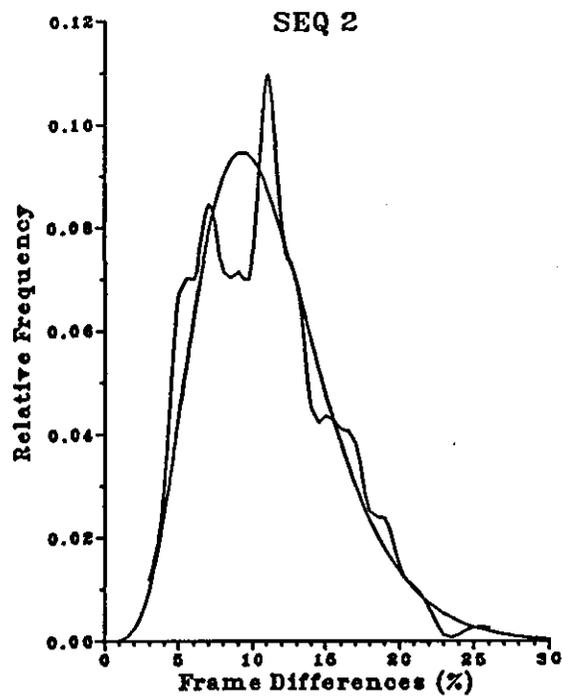


Fig. 3.10 Curve fitting for the PDF of Frame Differences for SEQ 2 using the Gamma Distribution.

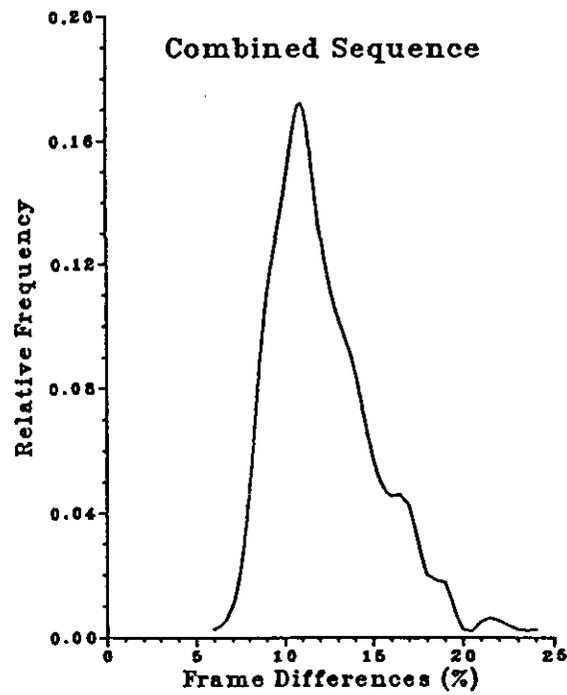


Fig. 3.11 The PDF of Frame Differences for the multiplexed sequence.

3.3.3.3 PDF of Output Bit Rate

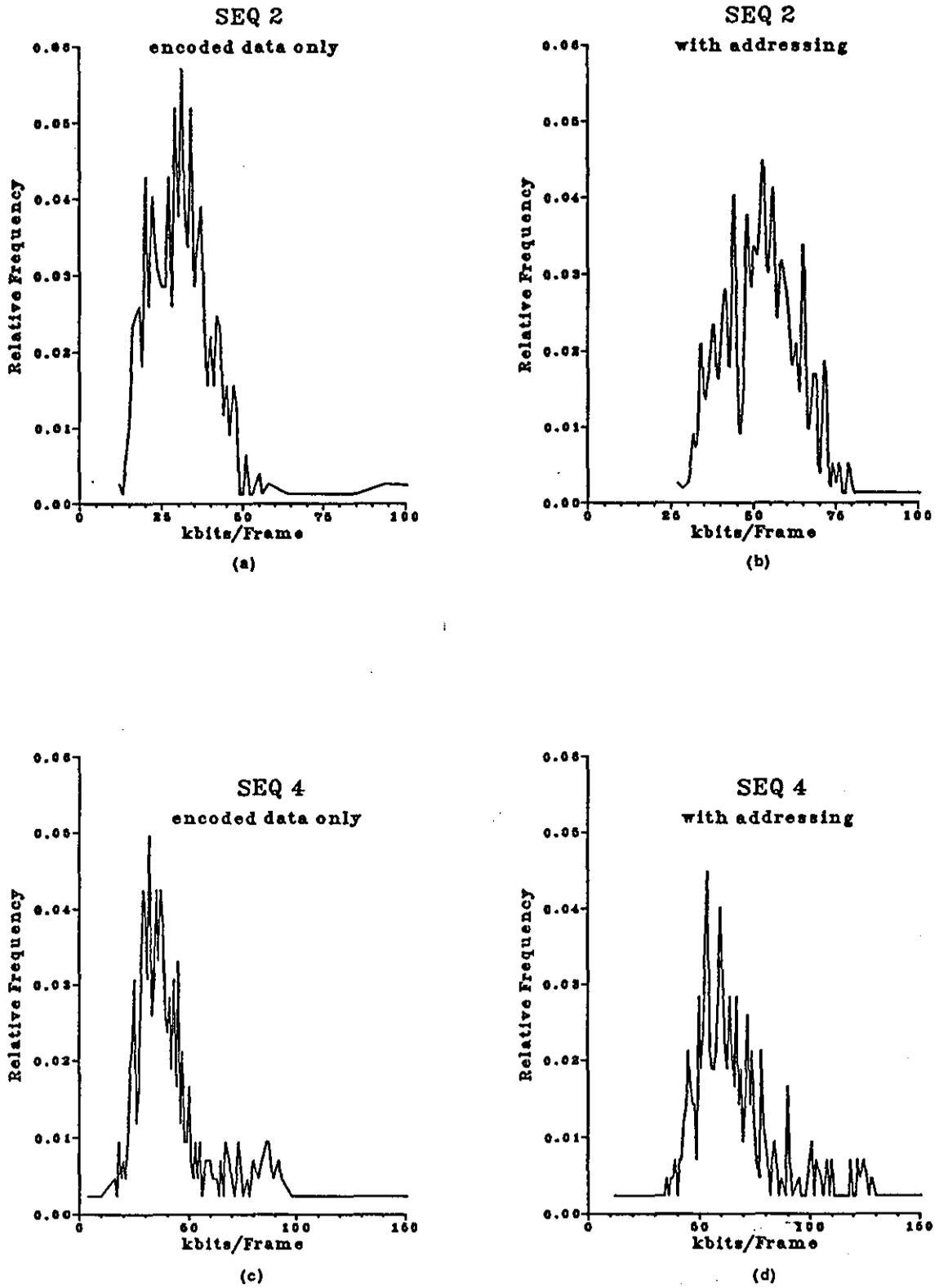
The output bit rate PDFs for sequence 2 and sequence 4 were computed from the results presented in section 3.3.2. Two PDFs, representing the encoded data rate before and after addressing overhead had been added, for each of the two sequences are shown in figures 3.12(a)-(d).

The PDFs are less skew than their FD counterparts, and α computed for the encoded data rate before and after the addition of addressing information were 2.81 and 4.0 respectively for sequence 2, and 2.1 and 3.0 for sequence 4. These figures are higher than their FD counterparts which were computed at 1.83 and 1.72 for the two sequences respectively, and as suggested earlier, a larger α tends to imply a more symmetrical distribution. Therefore the effect of DPCM and Huffman coding was to make the PDF of the output bit rate more symmetrical, or to become more 'Normal' like. Ghanbari^[25] made a similar observation with the same coding strategy. The addition of addressing information to the encoded data had a similar effect on the PDF.

These effects were due to the uneven coding and addressing efficiencies for clusters of different length, which in turn relates to the amount of FD in a frame. It was established earlier that these efficiencies improved with an increase in FD, the effect of which was therefore to reduce the degree of skewness of the PDFs by compressing the data harder and adding proportionally less addressing overhead when the FD was large, thus shortening the tails of the PDFs. These deductions are generally true for cluster coding, but they may not hold for block-based coding because the block size is fixed irrespective of the FD and thus a fairly constant coding and addressing performance could be expected.

3.3.4 Cluster Length and Addressing Overhead

One of the more important aspects of picture statistics with cluster-based coding strategy is the cluster length, because every cluster will need to be addressed before transmission. This overhead usually constitutes a significant proportion of the total data generated as shown in section 3.3.2, and will ultimately determine the efficiency of the system. This is particularly important for VBR video since it may affect the performance of the system when there is little movement present in the frames and where short clusters prevail. The net effect is to lower the bandwidth gain of VBR video. This problem is less of a concern in CBO video since a fixed amount of data must be transmitted regardless of the FD. Figure 3.13 shows the mean cluster length (MCL) per frame plotted against the FD, and as can be seen, the relationship is approximately linear for the range of FD plotted.



Fig(s). 3.12 The PDF of the codec output bit rate - with and without addressing overhead - for 2 different sequences.

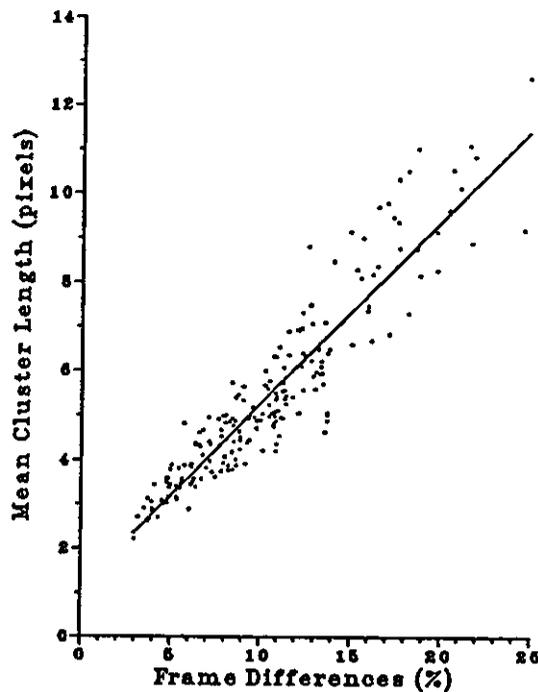


Fig.3.13 Variation of the mean cluster length with FD for SEQ 2.

The problem of addressing is even more acute for VBR video in an ATM network environment^(22,54), where a large demand by the source may not be met by the network and cells could be lost. Although this should occur only very infrequently, it results in a discontinuity in the data stream at the decoder. This type of error could be difficult to detect and correct, and thus it requires more redundancy to be introduced into the addressing to aid in the detection of cell loss and a rapid recovery without too much degradation to the picture quality. This could further offset the gain of VBR video.

Some possible ways of increasing the MCL have been considered, but it was difficult to study these proposals on the codec as it would entail undesirable modifications to the hardware. As a result, this study was conducted using the computer model of the codec and it allowed for greater flexibility.

One possibility of increasing the MCL considered was to reduce the threshold of the movement detector slightly. The threshold was subsequently reduced from 110 to 100, and it became apparent that this method would not produce the increase in efficiency required. This is partly due to the generation of more short clusters; more importantly, the marginal gain in the MCL was easily offset by the large increase in extra pixels requiring transmission. It is therefore not a method to be preferred.

A more promising method would be to increase the bridging gap. At present, a bridging gap of 3 is used. Simulation studies using bridging gaps of 5 and 7 were carried out producing the desired results of slightly longer clusters (on average, about 1 and 2 pixels longer respectively) and an overall smaller number of clusters. The increase in pixels requiring transmission is small and can be partly or completely offset by the reduction in addressing. However, the gain when there was little change was only marginal as the clusters were scattered too far apart for bridging to work. A further increase in the bridging gap would bring diminishing gain and would be offset by the increase in the extra pixels needed to be transmitted. The exact trade-off between reduced addressing overhead and an increased number of pixels would depend on the size of the overhead and the compression on the pixel data.

Another possible way to improve efficiency is to increase the minimum cluster length to 2. This is not expected to produce any significant visual degradation to the picture quality^[28], as very fine detail is probably not so important in videophone pictures. Furthermore, with a large bridging gap, a moving pixel would require a large horizontal velocity to escape bridging, and this would only happen for drastic motion. Under such circumstances, single isolated changes are few and lie mainly outside the active region as can be seen in figure 3.14, which shows the moving areas for a frame with a FD of 20%. Most single pixel clusters can therefore be safely regarded as noise, and genuine changes can be left to be updated by the background refresh process.

The methods described here for improving the efficiency of addressing are not specific to the ATM networks only, but apply equally to the circuit switched networks. However, with the bridging method, the trade-off between reduced addressing overhead and increased pixel data may differ in these two types of network since a smaller addressing overhead could be used in the circuit switched case. It should be noted that an improvement in the MCL not only increases the addressing efficiency, but also increases the coding efficiency.

With the above two strategies, a small gain in the MCL can be obtained for frames with a small amount of change, for instance, an increase of about 2 pixels on average with a bridging gap of 7. The characteristic of video which produces isolated short clusters when the amount of change is small restricts any further improvement.



Fig.3.14 Moving areas for a frame with a FD of 20%.

3.4 Conclusions

Several aspects of a video source with videophone type picture material have been investigated, both for PCM data and encoded data. The results show the inherent VBR nature of video and highlight the problems and inefficiency thereof in trying to constrain the data within a CBO channel. Using a CBO channel invariably means sending unnecessary information when the amount of movement in a frame is small, and reducing picture quality in order to constrain the output data rate when it is high. The use of a buffer alone to smooth out the data rate is not sufficient, as video-spurts could still present problems. A VBR network will allow the picture source to use as much bandwidth as it needs instant by instant and thus maintain picture quality, provided the peak bandwidth requirement can be met.

When the outputs of a number of uncorrelated VBR video sources are multiplexed in a network, the peak bandwidth requirement of the combined sources is lower than the sum of

the peaks of the individual sources. The lower bandwidth requirement results from bandwidth sharing among the sources, and leads to a more efficient utilisation of network resources and eases the traffic control problems.

The results also reveal the problems of using peak-to-mean ratio of the data rate variation as a measure of VBR gain. Not only does it vary with different types of picture material, but it also fails to account for factors such as the duration of the data bursts and the amount of data rate variation. Other factors which require consideration are the widely differing characteristics of the different sources, codec complexity, traffic control on the network, as well as the human visual response.

The probability PDFs of the frame differences have been found to exhibit mixed characteristics of both the Gamma and Normal distributions, and very often, they are multi-modal. Their inclination to either distribution depends on the amount of variation in picture content in the sequences. When several sequences are multiplexed together, the resultant frame differences PDF tends towards a Normal distribution as suggested by the central limit theorem. The coding of the frame differences with DPCM and Huffman algorithms, and the addition of addressing overhead, also have the effect of making the PDF of the output data rate more symmetrical.

The amount of addressing overhead in a VBR video could be unacceptably large when the amount of change in a frame is small, or when the pixel data are compressed. Some techniques have been described which increase the mean cluster length and hence improve the efficiency of addressing. Longer clusters will also lead to improved coding efficiency.

Chapter 4 :

A Simple VBR Video Source Model

4.1 Introduction

Having attained some understanding of the VBR video source characteristics, the next phase in this project was to use this understanding to develop a simple source model to be incorporated into an Orwell Ring simulation. The source was modelled using a discrete-event simulation technique, and the model itself was developed using Simula^[47] with Demos^[5] facilities.

Detailed modelling of the source was not attempted as source behaviour is very diverse and complex: it depends on, among other things, the type of picture material concerned and the coding strategies employed. It was also deemed unnecessary for the purpose of this research. Moreover, the model was developed with the constraint of limited computing resources, considering that it would have to be incorporated into the ring simulation, which in itself is extremely compute-intensive. For instance, a single run of a 34 Mb/s ring simulation normally requires about 3 hours of processor time on an ICL 3900 Mainframe! Consequently, only a simple source with restricted scope was modelled, based on a videophone type source with characteristics similar to those obtained in the preceding chapter, and simplifications were made wherever possible.

The source characteristics studied in the preceding chapter may be sufficient for the purpose of assessing the merits of VBR video, but are somewhat lacking in detail for the development of a model. Two additional pieces of information were identified as important for a basic model, namely, the auto-correlation of the frame differences and the distribution of changed pixels within a frame. The former was required to account for the video-spurt characteristic and the latter was necessary to account for the instantaneous data burst within a frame. Although intra-frame data burst has hitherto been ignored, it is nevertheless an important feature since no data smoothing is carried out at any stage in a VBR video source. These two aspects of the source behaviour will be considered next.

4.2 Auto-Correlation of the Frame Differences

The successive FD in a picture sequence is highly correlated since the amount of motion in successive frames is very similar⁽⁷⁾. This property of the FD is evident from figure 3.5, as the variation of the FD would be more haphazard and the long video-spurts less probable if this correlation did not exist.

The FD auto-correlation coefficient functions were computed for the different sequences, with different starting points. The results of the two extreme cases are presented in figure 4.1. As can be seen, the correlation coefficient functions approximate a first order Markov process, with the correlation coefficient falling within the range of 0.7-0.9. These results are consistent with those obtained by Verbiest⁽⁵³⁾ for videophone and video-conferencing type pictures.

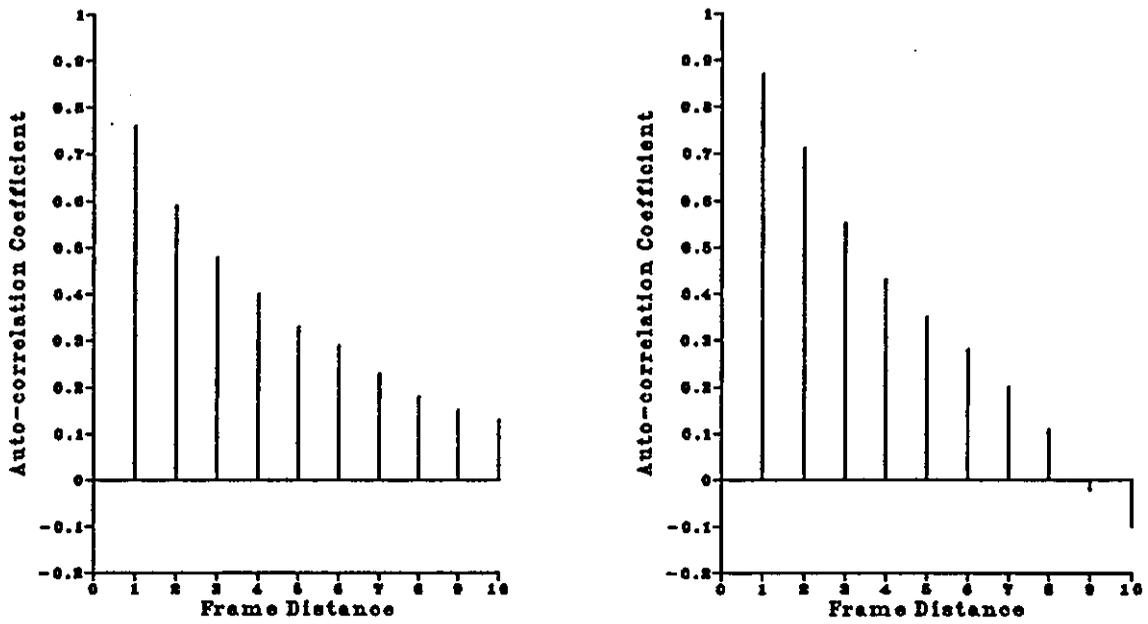
It must be stressed that these results represent the long-term auto-correlations which were computed over the entire length of the sequences (in this work, about 400-1000 frames). The short-term auto-correlations are not expected to deviate significantly from these results (except, perhaps over the short periods in which scene changes occur), and they are not likely to vary significantly with time.

4.3 Distribution of Changed Pixels Within a Frame

In almost all reported work on video signal statistics, the data rate within a video frame (or field) has always been assumed to be smooth. However, in a truly VBR video codec, data are output almost as soon as they have been coded (a small amount of buffering would always be present, for example, for packetisation purposes), and as such, the instantaneous output data rate would fluctuate according to the local image complexity, which varies across the picture.

In codecs employing inter-frame coding techniques (such as the COST211), the variation of the instantaneous output data rate is related to the concentration of the changed pixels and their locations. It must be noted that the term 'instantaneous output data rate' does not refer to the actual speed of the bit stream, which is a function of the clock rate of the codec and thus representing the upper bound for the output data rate. Instead, the term relates to the short-term average data rate, for example, over a video line. This instantaneous data rate is an important aspect of VBR video, especially when the interaction between a VBR video source and a network is to be considered. The instantaneous data rate can be much higher than the frame data rate, and consequently the model must incorporate this feature.

Experiments were carried out on the codec to collect the statistics on the distribution of changed pixels on a line-by-line basis over the video frames. Figure 4.2 shows the distributions for the two video fields of a head and shoulders scene (sequence 2). The results



Fig(s). 4.1 The auto-correlation coefficient functions of 2 different sequences.

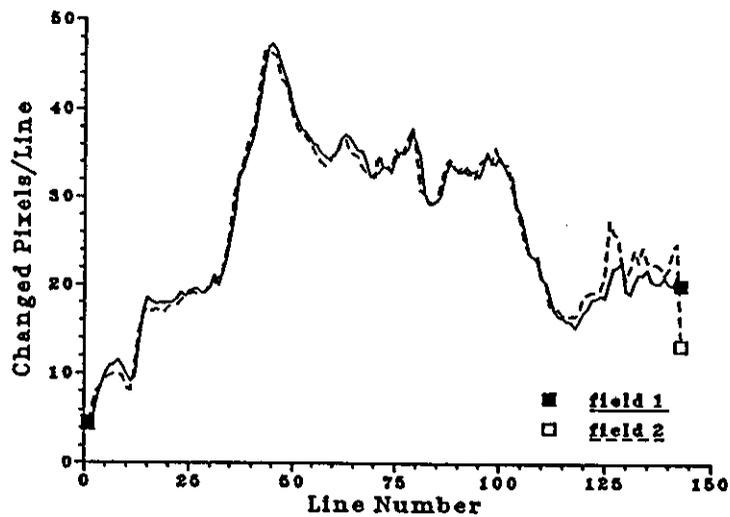


Fig. 4.2 The distribution of changed pixels within a frame averaged over a sequence.

shown were obtained by averaging over the entire sequence. As expected, the changed pixels were distributed almost identically in the two fields. The interesting feature of the distribution is that three distinct regions of varying degrees of activity can be identified: a low, a medium and a high activity regions.

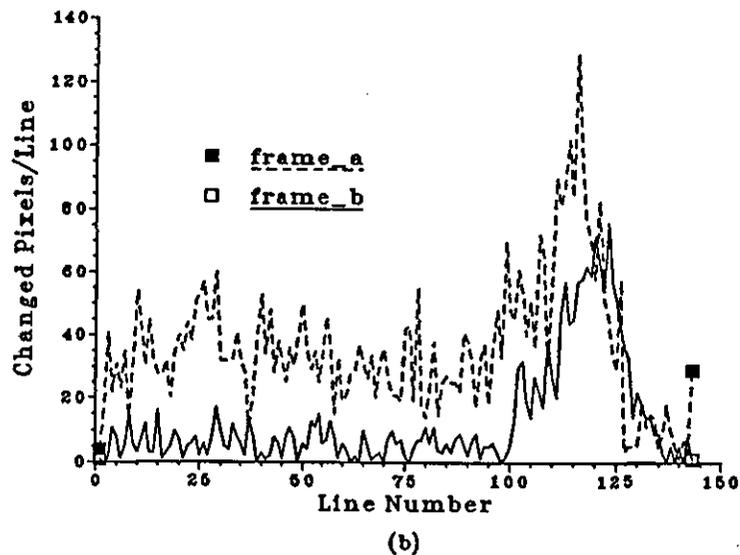
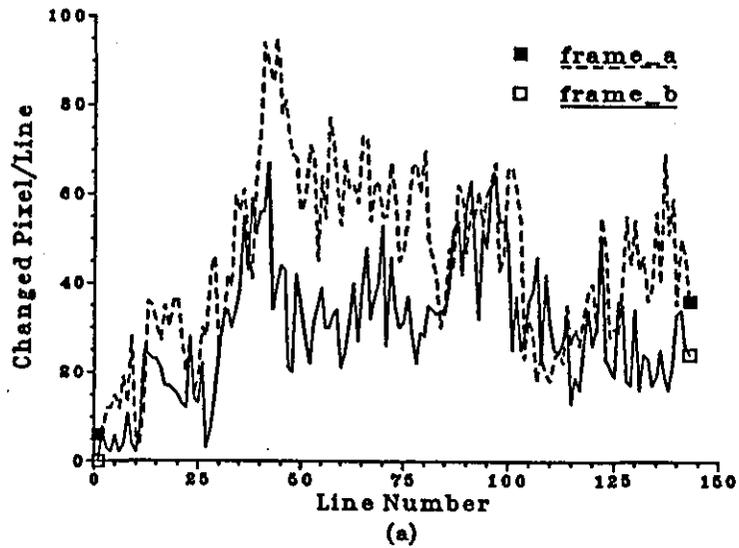
The distributions of changed pixels for two frames that are widely separated in time, for the same sequence, are presented in figure 4.3, along with those from a different sequence (sequence 1). Only a single field is shown for each frame. The three regions of different levels of activity can still be recognised in figure 4.3(a), and they can also be identified in figure 4.3(b). The level of activity of each region tends to vary in proportion to the FD; when the FD is very small or very large, the three regions become almost contiguous and indistinguishable.

The three regions can be related directly to the content of the scenes. For instance, figure 4.3(a) is a head and shoulders scene where the low activity region corresponds to the top of speaker's head; the face area constitutes the high activity region across the middle of the screen, while the moderate motion of the shoulders makes up the medium activity region. Figure 4.3(b) is a waist-up type picture, and in this case, the medium activity region covers the head and chest of the speaker, while his hand motion generate the largest FD and consequently form the high activity region; the low activity region corresponds to the relatively still waist area. In general, the shape of the distributions can be derived intuitively from the content of the scenes, and most distributions can be broken down into three activity regions, although the level of activity and the number of lines in each region vary for different scenes.

It is clear from these results that the instantaneous data rate varies quite significantly within a frame, especially if the field synchronisation period is to be taken into account. The data rate could rise from zero to twice that of the mean frame data rate, and the peak could last for several milliseconds. The data rate variation within a frame is therefore an important factor to be considered when investigating the transportation of VBR video over any ATM network.

4.4 Modelling of the Source Behaviour

The various aspects of the behaviour of the videophone type source, which were considered to be important characteristics of the source and which may have serious impacts on the efficient transfer and control of VBR video over an ATM network, have been analysed. These features were incorporated in the model, although some simplifications were necessary due to the lack of information on some aspects of the source characteristics and the constraints in computing resources. The model must however retain those features of the



Fig(s). 4.3 The distribution of changed pixels within a frame of 2 widely separated frames for 2 different sequences (only a single field is shown for each frame).

source behaviour which are most testing from the point of view of the networks, so as to enable more meaningful results to be derived from the subsequent ring simulation study. These features and their modelling will be considered.

4.4.1 Probability Density Function (PDF) of Frame Differences

Results from the preceding chapter and from Seyler^[51] have indicated that the PDFs of the FD of most picture material (notably the head and shoulders type), exhibit mixed characteristics of the Normal and Gamma distributions. Since the actual PDF of the FD is very complex and difficult to generate, and since there is a lack of in-depth understanding of the relationship between the two distributions and the picture material, it was decided that only the Gamma nature of the PDF would be modelled.

Gamma distribution was chosen as the basis for generating the FD because the long tail of the distribution is considered an important feature of the signal characteristics, a feature which could have a substantial impact on the network loading and cell loss performance. Furthermore, as can be seen from figure 3.10, the Gamma distribution is probably the best single mode distribution for approximating the experimental results, and the approximation improves as the variety of motion in the picture sequence increases.

However, since Gamma distributed samples are rather difficult to generate, and since the underlying PDF of the generated samples would eventually be distorted by the filter used for correlating successive samples, samples with an Erlang distribution were generated instead. The Erlang distribution^[52] is a subset of the Gamma distribution (from equation 3.3, α is a real number for Gamma and an integer for Erlang), and it thus retains most of the required qualities of the latter. Samples with an Erlang distribution are easier to generate and furthermore, a facility already exists in the DEMOS package for this purpose. The development of the model was thus greatly simplified. It should be pointed out that the curve fitting in figure 3.10 was, in fact, carried out using the Erlang distribution, thus validating the use of this distribution for the model, although it does not have the range of the Gamma distribution. The scope of the model is thus restricted in this respect.

4.4.2 Auto-Correlation of the Frame Differences

The results obtained in section 4.2 indicate that the auto-correlation function of the FD follows a first order Markov process. A first order recursive filter, as shown in figure 4.4, can thus be used to introduce the required auto-correlation into the FD samples drawn from an Erlang distribution. The equation of the filter is as follow:

$$x_k = \alpha x_{k-1} + \beta u_k$$

where x_k and u_k are the k th generated and drawn samples respectively, α is the correlation coefficient and $\beta = \sqrt{1 - \alpha^2}$.

From the results previously obtained, α was found to be in the range 0.7-0.9, assuming there were no scene changes. It is expected to vary within this range with time but the exact nature of this variation is not known. However, for simplicity, α is assumed to be constant throughout with the mean correlation coefficient of 0.8.

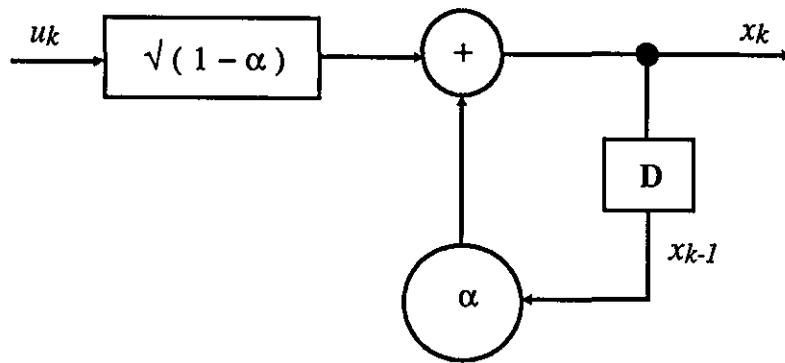


Fig. 4.4 A first order recursive filter.

The filter has a transfer characteristic given by:

$$H(j\omega) = \frac{\beta}{1 - \alpha \exp(-j\omega)}$$

which gives rise to a D.C. gain of

$$\frac{\beta}{1 - \alpha}$$

With $\alpha = 0.8$, the D.C. gain is 3, which means that the sample mean at the output of the filter is three times that of the input samples, while the variance remains unaffected. As a result, when the output samples are scaled down by the gain factor, their variance is reduced by a factor of $(3)^2$. The implication of this is that, for an output sample sequence with a

required mean and variance, the input samples must either have a mean of one third of the required output mean, or a variance $(3)^2$ times larger. The second approach was adopted in this model.

A further characteristic of this filter is that the output samples will have a PDF which tends to a Normal distribution, i.e. less skew, whatever the PDF of the input samples. This undesirable effect is, to some extent, compensated for by the fact that subsequent coding and addressing of the changed pixels on the COST211 codec have a similar effect on the output bit rate (see section 3.3.3), and in that respect, this effect of the filter is a positive aspect.

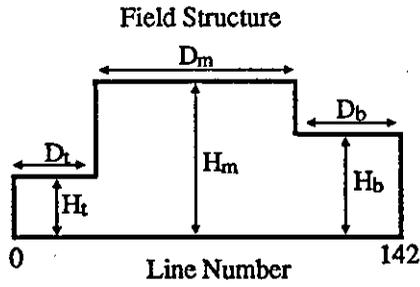
4.4.3 Distribution of Changed Pixels within a Frame

The distribution of changed pixels within a frame was modelled according to the findings of section 4.3, that is, the frame was partitioned in the vertical direction into three regions of high, medium and low activity; with the added proviso that the top region should not be the high activity region and the middle region not the least active, since scenes with these types of pixel distribution were considered highly unlikely, although not impossible. The two fields within a frame were assumed to have identical distributions, with the number of changed pixels in the frame equally divided between the two fields. Hence, only a single field with half the FD needs to be modelled, and a frame can be obtained by repeating the same field. The flow chart of a scheme for generating the intra-field distribution of changed pixels is given in figure 4.5.

Since there is no simple relationship linking the number of lines or the number of changed pixels per line in the three regions, these were chosen randomly. Only the number of lines for two regions needs to be determined and the third is automatically fixed since there is a fixed number of video lines in each field. It could, however, be assumed that the top and bottom regions are more likely to consist of only a few lines, especially in the head and shoulders scene. As such, the number of lines in these two regions could be assumed to have a Negative Exponential or Erlang distribution.

Having partitioned the field into three regions, the activity levels (number of changed pixels per line) for the three regions were then chosen randomly based on a frame with a FD of 10%, and subject to the constraint on the relative activity levels of the regions stipulated earlier.

The scheme enables an intra-field distribution of changed pixels of the required general shape to be constructed in block form. For detailed modelling, the inter-line variation should then be superimposed on the basic model. Some analyses have been carried out on the inter-line variation behaviour, but it was decided that such details would not be necessary - the reasons being that the variation in the number of changed pixels from line to line in a



mpl : mean number of changed pixels per line
 N : total number of changed pixels in a frame with 10% FD
 mxp : maximum number of pixels per line (306)

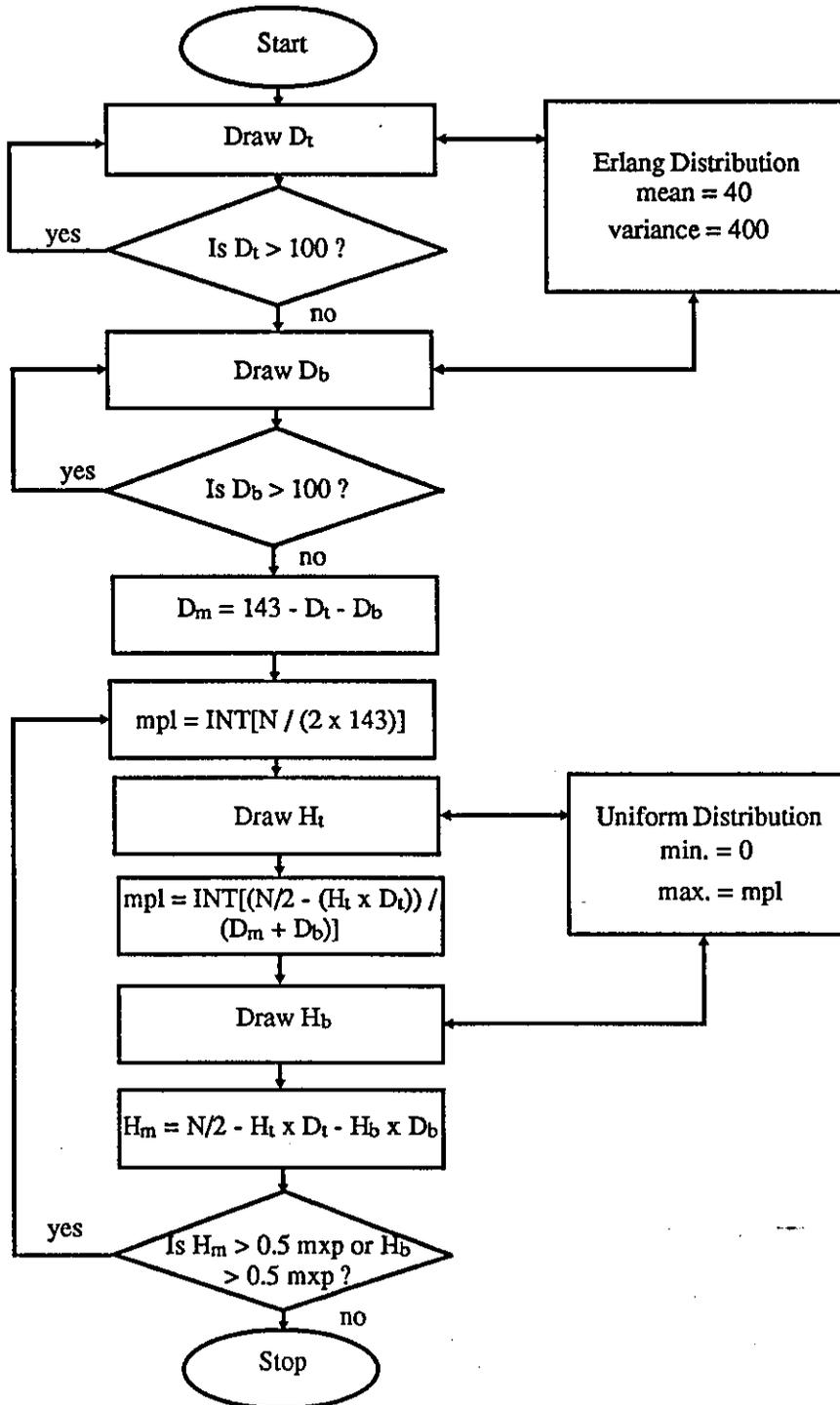


Fig. 4.5 Flow chart for generating intra-field changed pixels distribution.

region is relatively small, and the bursts are generally of very short duration. As there will almost certainly be a small amount of buffering present at the source output, such variation can be easily mitigated. For instance, a maximum variation of about 40 pixels from the regional mean in the high activity region was observed in figure 4.3(a), and assuming a 2:1 compression, this would represent a 160 bits fluctuation. When compared to the proposed 32 byte (256 bits) and 64 byte (512 bits) ATM cell information fields, considerable smoothing can be achieved by the packetisation process alone.

A further reason for not going into such detail is the computing overhead required: a sample would have to be drawn for each line and fed through a filter. Considerable processing time can be saved since line activity occurs at very short intervals, in this case, 128 μ s.

A different intra-field distribution of changed pixels was generated for different sequences, and for simplicity, the same distribution was used for the entire duration of a sequence. The activity level of the regions were scaled in proportion to the FD of each frame, and care was taken to ensure that the level of the high activity region did not exceed the maximum number of pixels per line, i.e. 306. The simplification made is justified provided no scene change is allowed over the duration of the sequence.

4.4.4 Coding Performance and Addressing Overhead

The modelling of the three aspects of the VBR source behaviour discussed above have provided sufficient details for a basic source model to be developed. However, there are still some minor factors which need consideration: the coding and the addressing of the changed pixels.

Without considering any subsequent coding on the changed pixels, the output bit rate of the model would simply be the FD converted into the number of bits per frame, since each pixel is PCM coded. It becomes more complicated when subsequent coding is considered, because coding performance is not constant for different values of FD as has been deduced in section 3.3.4. The coding performance is a function of the mean cluster length, which is in turn related to the FD. Unfortunately, the exact relationship between the MCL and the FD is picture dependent, although it is generally true that the MCL increases with the FD. However, the variation of the coding performance is only significant for very large and very small FD; for moderate FD, the coding performance does not vary greatly. Therefore, for simplicity, the coding performance was considered constant.

The addressing efficiency has a very similar characteristic to the coding performance, in that the proportion of the output data which is addressing information varies with the FD. Again, for similar reasons as those for coding performance, addressing overhead was

considered to be a fixed proportion of the output data. With these two parameters fixed, the FD could be scaled according to the amount of compression and addressing required to obtain the output bit rate.

The effects of coding and addressing on the underlying PDF of the output bit rate have been, to some extent, accounted for by the recursive filter; thus no further refinement was attempted.

4.4.5 Field Synchronisation

Field synchronisation refers to that part of the video signal which is devoted to synchronisation purposes and contains no visual data. In the COST211 codec, the field synchronisation period (during which no data is generated) extends over 13 video lines or about 2 ms. Field synchronisation is considered an important feature of VBR video as it increases the variation of the instantaneous data rate. Furthermore, its duration is significant. Consequently, it was recommended that this feature be incorporated into the model. It is acknowledged that some codecs may use this period for coding purposes to alleviate the processing speed constraint. As a result, a 'blank' period will not appear in the output data stream of such codecs.

4.5 A Simple Source Model

A simple model was developed based on the source behaviour modelling concepts outlined above. A block diagram of the model is shown in figure 4.6. The FD Generator consists of a random number generator with an Erlang distribution which has a mean of 10% and a variance of 100. The samples were generated at a rate of 25 samples/s on the simulated time scale, corresponding to the full frame rate of 25 frames/s.

These samples were fed into the first order recursive filter described earlier in order to induce the auto-correlation required into the samples (first order Markov process with $\alpha=0.8$). The samples were then scaled down to remove the undesirable D.C. gain, and the variance was reduced, as a consequence, to 11.1. This value is slightly lower than those shown in Table 3.1 for a sequence with a mean FD of 10%, and accounts for the effects of subsequent coding and addressing. These samples were then scaled according to the sequence mean, which was drawn from another random number generator with a pre-determined probability distribution, and the variance of the samples was also varied accordingly, which is desirable. Unfortunately, the nature of the distribution of the sequence mean was not known as there was insufficient data available for this to be deduced. In this model, it was proposed that either a Normal or a Uniform distribution be adopted. The former is quite probable since most random processes are Normal in nature, while the latter was chosen for its simplicity.

The sequence mean was taken to lie mainly in the range 8%-22% as an approximation to the results in Table 3.1. The scaled output samples then become the generated FD of a VBR video source with a given sequence mean.

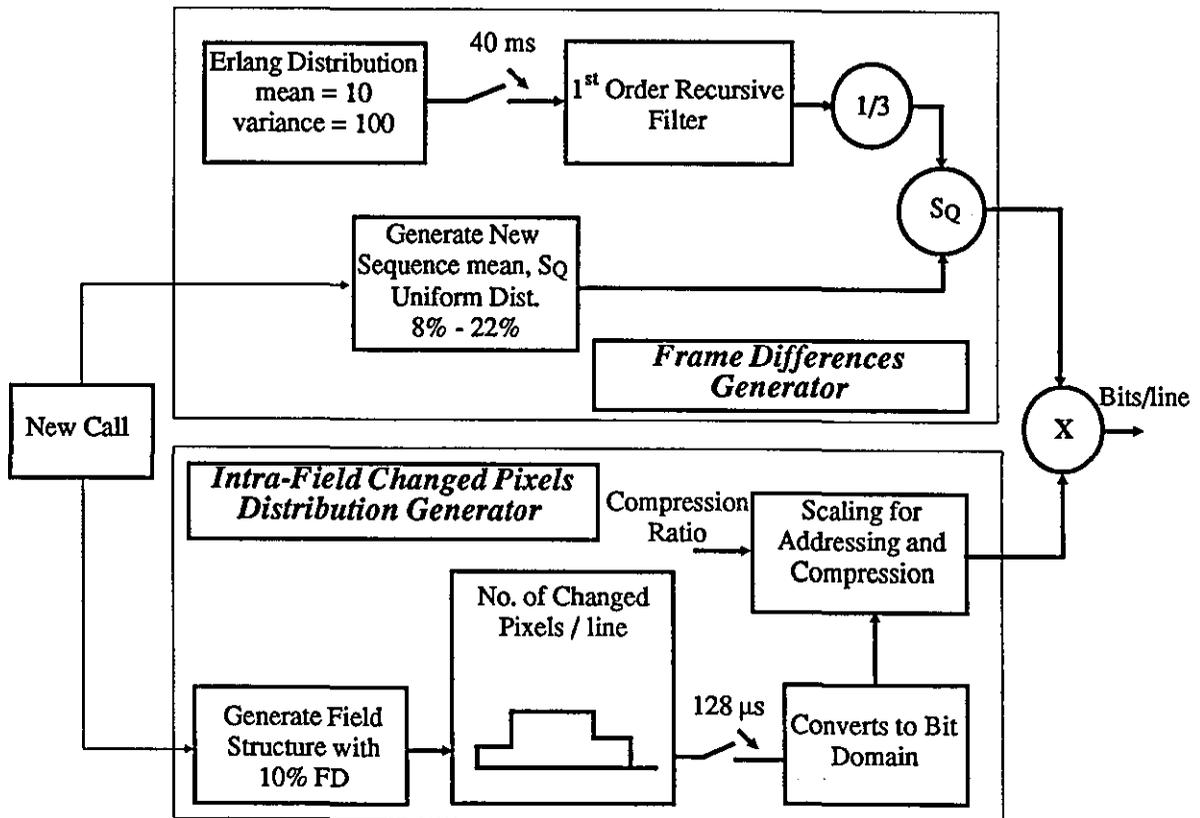


Fig. 4.6 Block Diagram for the VBR Video Source Model.

The Intra-field Changed Pixels Distribution Generator has already been described in the preceding section. The distribution generated, based on a frame with a FD of 10%, was converted and expressed in number of bits per line. This was then scaled down according to the compression specified, and subsequently scaled up by a factor which corresponds to the pre-determined proportion of addressing overhead (20% in this case). The compression ratio was left to user specification in order to allow for user control of the required output bit rate.

The final output bit rate of the model was obtained on a line-by-line basis. Each line was taken in order at line intervals ($128 \mu\text{s}$) from the Distribution Generator, and scaled in proportion to the FD generated by the FD Generator. The distribution, which was field-based, was repeated to emulate the two fields in a frame before the FD was updated, and 13 blank lines were inserted between fields for field synchronisation.

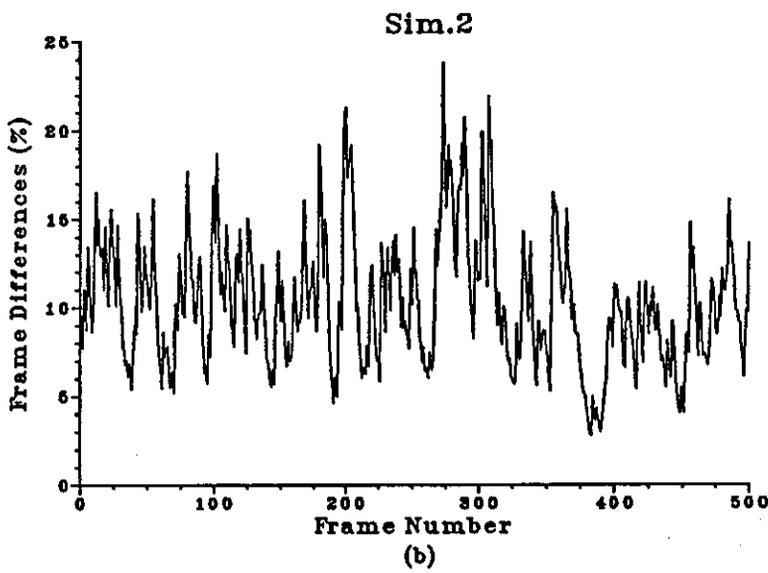
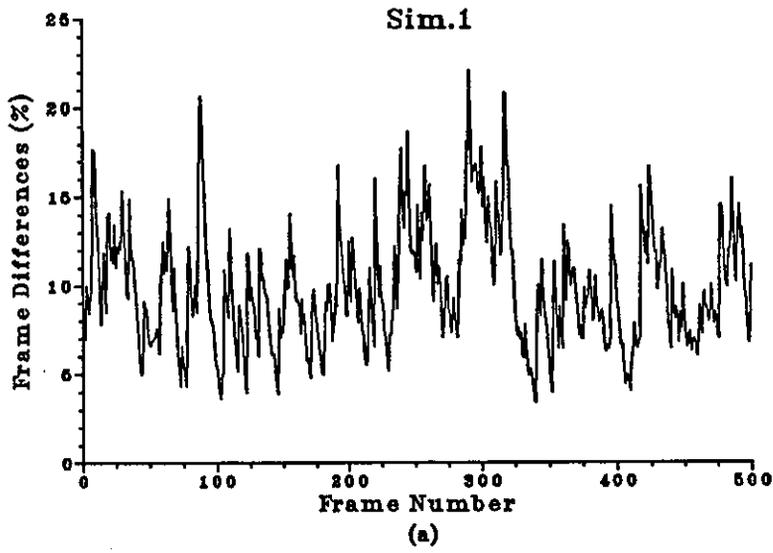
Examples of the variation of the generated FD with a sequence mean of 10% are shown in figure 4.7. The sequences bear a good resemblance to those shown in figure 3.5, and peak-to-mean ratios of 2-3 were observed, which is consistent with the practical results. Video-spurts of long duration were clearly visible. The auto-correlation functions of the generated FD shown in figure 4.8 is of the appropriate characteristic. The PDF of the generated FD is shown in figure 4.9, and as can be seen, the PDF still retains some of the skewness of the Erlang distribution except that the tail has been reduced. Thus the effect of the recursive filter was not as drastic as originally expected. The performance of the FD Generator of the model is considered very satisfactory given the constraints associated with the simplicity of the model.

The most probable intra-field distributions of changed pixels generated are illustrated in figure 4.10. These, in general, are the type of distributions expected from the generator, and they are only as good as the assumptions on which they are modelled. On that premise, the performance of this generator is also satisfactory.

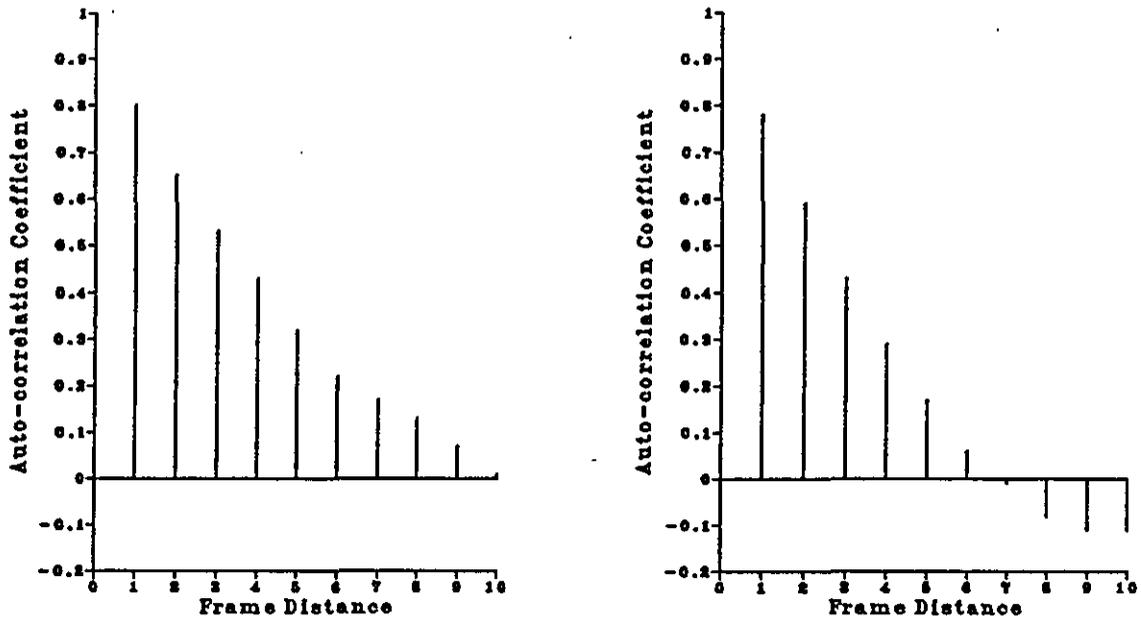
The model can be used to generate VBR video sources exhibiting a range of characteristics. Each time a VBR video source is generated from the model, its characteristics will be different from others generated from the same model. Each source will be different in terms of their sequence means and FD variances, FD variations and intra-field distributions of changed pixels. The relationship between the sequence mean and the FD variance is, however, fixed, and other aspects such as the auto-correlation of the FD, the proportion of addressing overhead and the compression ratio are the same for all sources generated.

Scene changes can be implemented quite easily. Since they occur randomly in a sequence, they can be implemented as different generated sources concatenated to each other, intervened only by sharp peak changes. They are, however, not implemented because of doubts over whether scene changes should be transmitted in full.

One other factor which hitherto has not been considered is the way in which the first frame of a sequence should be treated. As this corresponds to a scene change situation, it would generate the maximum data rate. However, it may not be necessary for this frame to be transmitted in real time. Rather, it can be transmitted over a duration of a few frames, thus averaging the peak data rate which might otherwise be difficult to handle if every video connection is to start with a peak. The slow picture build-up of the first frame need not necessarily be noticeable. In this model, the first frame is assumed to be transmitted over a pre-specified number of frames at a fixed bit rate.



Fig(s). 4.7 Variation of Frame Differences with time for 2 simulated video sources with a mean of 10% (25 frames/s).



Fig(s). 4.8 The auto-correlation coefficient functions of 2 simulated video sources.

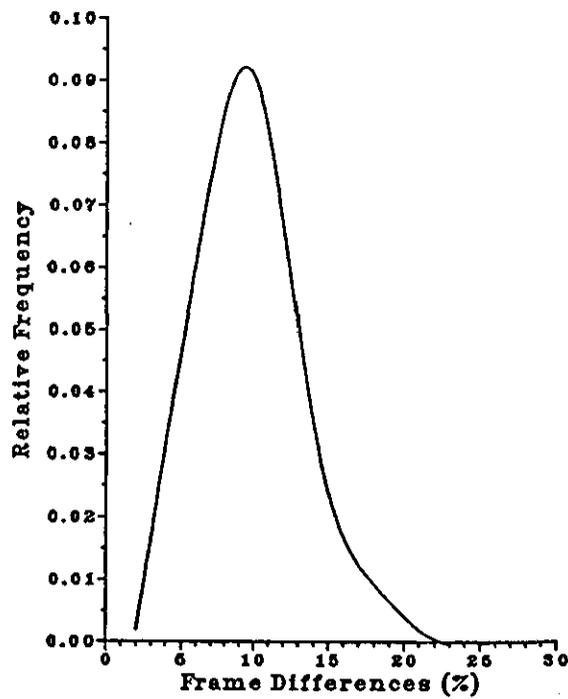
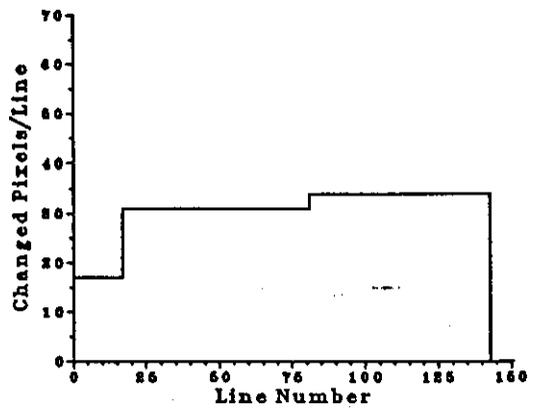
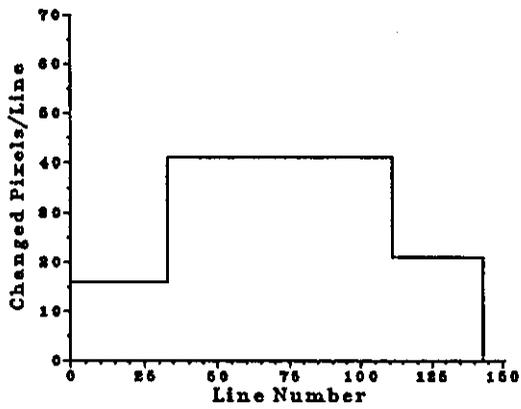
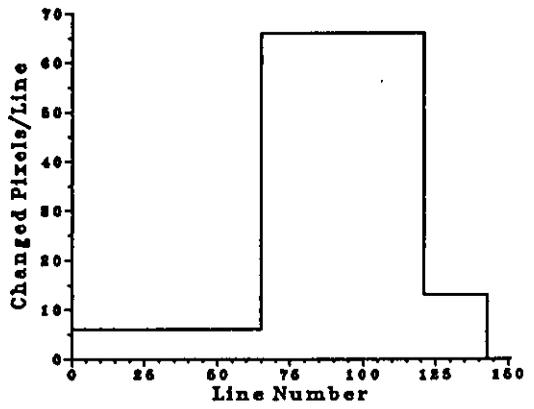
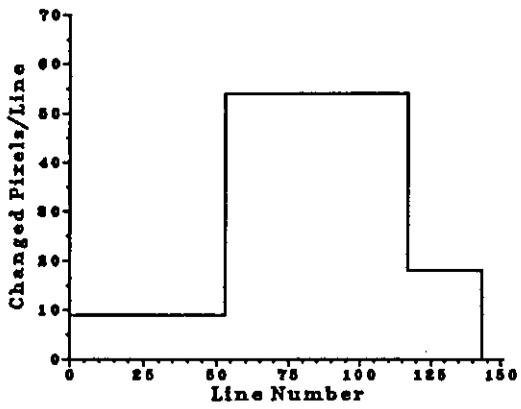
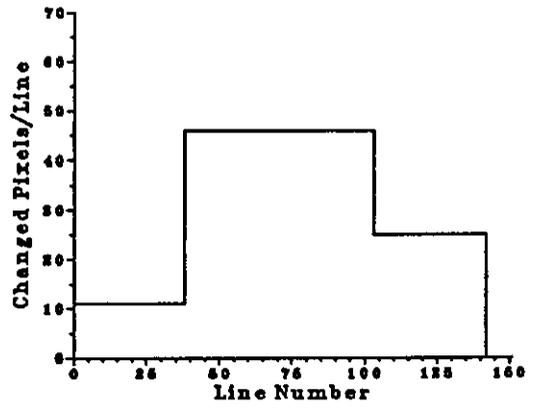
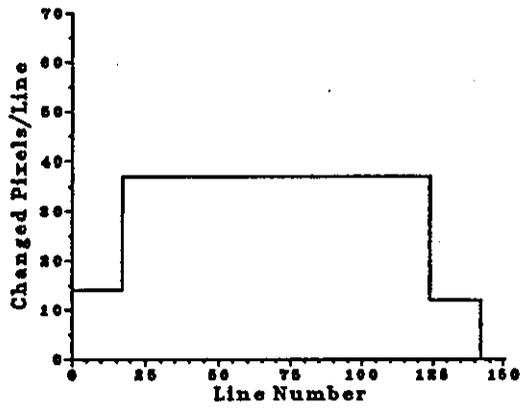


Fig. 4.9 The PDF of Frame Differences for a simulated video source.



Fig(s). 4.10 Examples of the possible frame structure generated.

4.6 The Model in the Ring Simulation Environment

The model developed next had to be incorporated into the ring simulation software, to enable the interaction of VBR video and the Orwell Ring to be studied. This section describes the integration of the model into the ring simulation.

Both the ring simulation and the model were developed in Simula with DEMOS facilities, which supports the object-oriented programming paradigm. In this environment, the source model simply forms an 'object' (or 'entity' in DEMOS) in the ring simulation program, and a copy of this object can be generated whenever a VBR video source is needed. Each of these generated objects represents a different video source with different characteristics (as described in the preceding section), and each interacts independently with the 'node object' to which it is attached. When a connection is terminated, the object can be modified to become a different source with different characteristics and relocated on a different node.

In the model, data are generated on a line-by-line basis, which means that the output comes in bursts at the end of each line (128 μ s interval). This time interval is large when compared to the critical time interval of 125 μ s on the ring (see Chapter 5), and will lead to an unrealistic and highly variable loading situation of regular large data bursts separated by quiet periods. This will not only create a false temporary overload, but will also distort the results of the ensuing simulation studies. Data generation must therefore be spread over the line rather than being concentrated at the end of each line; to this end, a video line is divided into 4 phases of equal interval, each 32 μ s, and data are assumed to be distributed equally over the 4 phases and are output at the end of each phase.

This feature can also be used to ease the tracking of the progress of each VBR video connection. For this purpose, the simulation time is regarded as consisting of cycles of 4 phases, where each cycle corresponds to a line interval. Each new connection is assumed to start at the beginning of the phase in which it is generated, and its subsequent line and field synchronisation are maintained with respect to that phase. This approach avoids the need to track each connection separately, and all connections with the same starting phase can be processed in groups, thus saving valuable computing time.

Since the ring is an ATM network, data must be packetised into cells before being transferred to the node for subsequent launching onto the ring. Whenever a cell is filled, it is output at the end of that phase; a cell which is not filled at the end of a phase is carried over to the next phase for continued filling. An unfilled cell at the end of a line is carried forward to the next line, and a partially filled cell at the end of a field is transmitted incomplete. No buffer is provided on the source such that cells are output at the end of the phase in which they are assembled. In practice, at least a buffer of one cell would be needed on the source to resolve contention for the node buffer.

A further modification to the source model was necessary in view of the short simulation time for which the ring simulation was performed - typically in the order of a few seconds in real time for economic and practical reasons. This constraint limits the simulated mean call holding time of VBR video connections to a small fraction of the simulated time, and this means a source will generate, on average, only a few frames of data before being terminated. As a result, the effects of data rate variation will not be significant. In order to mitigate this short-coming, video frames were considered to consist of only a single field so that more frames could be generated over the duration of a call. This modification was carried out by not repeating a field and by increasing the FD sample generation rate to 50 samples/s. Consequently, more data rate variation can be generated and this provides a better basis for the simulation study.

4.7 Summary

In this chapter, two more aspects of the videophone type signal behaviour, namely, the auto-correlation of the frame differences and the distribution of changed pixels within a frame, were examined. These, together with the statistics presented in the preceding chapter, formed the basis for a simple VBR video source model to be developed for videophone and some video-conferencing applications. The source was modelled down to the video line level.

The scope of the model was, however, limited by the amount of data available on the signal characteristics and the type of codec from which they were gathered, although results from other workers were also considered. It was further limited by the numerous simplifications that were necessary in view of the constraints in computing resources. Notwithstanding these limitations, the model was considered to have adequately emulated the video source upon which it was modelled, and would enable some valuable insights into the interactions of VBR video with the Orwell Ring to be gained in subsequent simulation studies. The integration of the model into the ring simulation software necessitated some further modifications to the model.

Although primarily based on videophone applications, results obtained from the ring simulation with this model should also provide information about the performance of other VBR video services over the ring, since many aspects of the VBR video signal behaviour are quite similar for the different video-communication applications.

Chapter 5 :

Simulation of VBR Video on an Orwell Ring

5.1 Introduction

It has been suggested that video services can be better and more efficiently provided, in terms of enhanced picture quality and improved network performance, if video signals were VBR coded and transmitted over an ATM based BISDN. It has also been suggested that such benefits can only be realised if the outputs of a large number of statistically uncorrelated VBR video sources are multiplexed in the network, and effective policing or control of these bursty traffic can be exercised to achieve a small cell delay and minimal cell loss, while maintaining the stringent performance requirements of the high priority services and attaining a high network utilisation factor.

With the developments in ATM networks still in an early stage and the concept of VBR video evolving alongside, there is no practical means of investigating these suggestions as there is no large scale working system on which experiments and measurements can be conducted; simulation was therefore used for this study. In this chapter, the results obtained from the simulation of VBR videophone service on an Orwell Ring, using the VBR video source model developed earlier and an Orwell Ring simulation developed at BTRL, are presented.

Before the results are presented, a brief description of the Orwell Ring will be given with emphasis on some key parameters which are to be investigated. An account of the simulation and the aspects of the system performance which are of interests are also given.

5.2 The Orwell Ring - A Brief Description

An Orwell Ring is basically a high speed slotted ring operating the Orwell Medium Access Control (MAC) protocol. Nodes on the ring concentrate traffic from a wide range of services, for example video, voice and data, onto the ring. The ring has an integer number of slots circulating in it; all information to be exchanged over the ring is organised into fixed size data transfer units called cells, which are carried in the slots. The cells contain a header field and an information field. The header part contains routing and control information which enables the cells to be steered to the correct destinations.

A cell is launched into an empty slot when the latter arrives at an active node. It is retrieved by the node, or group of nodes in the case of multicasting, which recognises the address in the cell header. The slot is then converted back to an empty slot which can be used by any node downstream. This destination release feature is unique to the Orwell protocol when compared to other slotted ring protocols^[30]. Cells are delayed at a node while waiting for an empty slot, but the protocol limits this cell delay to an upper bound for delay-sensitive services.

With the destination release policy, a slot can deliver more than one cell during a single rotation on the ring. As a result, the ring has been reported to be capable of supporting an information transfer rate 1.5 times its own speed^[6]. This policy may, however, result in 'hogging', where nodes downstream from a high usage node may be completely deprived of their share of the empty slots or bandwidth. This could lead to an unacceptable cell delay and eventual cell loss as a result of buffer overflow. In order to prevent such occurrence, the protocol limits the number of cells that a node may launch onto the ring before being re-initialised. This number is termed the 'd' value of the node, and it is dynamically adjusted to meet the demand of the node subject to bandwidth availability. Whenever all the active nodes have exhausted their 'd' allocations such that at least one empty slot successfully circulates the ring, a reset slot is generated to re-initialise the nodes, and the whole process repeats itself.

The interval between two consecutive resets is known as the 'reset interval (RI)'. The RI is clearly influenced by the load on the ring: it is short when the load is light and vice versa. However, when the ring is heavily loaded, the RI is only allowed to increase up to a pre-defined maximum value. In order to maintain the RI below this maximum, each node uses a load monitor to measure the average RI over a period of time, or alternatively, to keep a count of the number of reset occurrences over the same period (reset rate, RR). Based on these measurements, call requests are rejected if they may cause the maximum RI to be exceeded (or the RR to fall below a minimum). There are therefore, call acceptance thresholds associated with the different services, and the threshold levels depend on the bandwidth requirements of the services. When the RR falls below any of these thresholds,

call requests of the corresponding service will be rejected. With the RI maintained below the maximum value, every node gets its 'd' allocation within a specified period, thus its guaranteed bandwidth allocation.

The above summarises the basic operation of the Orwell Ring. For further details on the ring see references [1,2,3,22]. However, it is considered necessary that some of the Orwell terminologies used in the simulation study should be clarified to facilitate discussion.

5.2.1 Service Queues With Priority

In order to meet the different requirements of the various classes of service, cells originating from the different classes of service are treated with different priorities at the nodes and are assigned separate service queues. Cells from higher priority services are always given preference in accessing the empty slots. In order to prevent excessive cell delay to the low priority services, each queue has its own 'd' allocation which works in a similar manner to the node 'd' allocation. High priority services are not subjected to overload control such that their 'd' allocations will always remain intact; this will be explained in more detail later.

Currently four service queues have been identified which cater for the 2 Mb/s CBO video, 64 Kb/s CBO voice, data and VBR video services (in order of priority). The allocation of 'd' to the queues for the CBO services is deterministic while that for the VBR services is a subject of research.

5.2.2 Masked Resets

A method of 'd' allocation for the data service was proposed^[1], by which a small fixed 'd' value would be provided on each node to the data queue, thereby guaranteeing the data service a minimum amount of bandwidth at all times. However, this background allocation may not always be fully utilised, and this could lead to a false estimate of the unused ring bandwidth. As a result, excess calls may be accepted onto the ring with an increased risk of ring overload when a data burst occurs.

Similarly, whatever the 'd' allocation method for the VBR video, the fact that bandwidth is allocated in blocks (for instance, with a maximum RI of 125 μ s and a cell information field of 16 bytes, a 'd' of 1 is equivalent to 1 Mb/s of bandwidth) suggests that there will always be excess bandwidth being allocated which could cause overload. An additional control mechanism is thus required to cope with this overload condition. Such a mechanism is also necessary to cope with a ring failure in a Torus system.

This additional control is achieved with the concept of Masked Resets (MR), where some of the resets are masked from the low priority VBR queues in the event of an imminent overload, thus effectively removing some of the bandwidth allocated to these services. The underlying principle of the method is that, until the RR drops to a certain value of r resets

per $125 \mu\text{s}$ (the current maximum RI), no resets are masked from the VBR queues. Below this value, resets are masked at a rate which is adjusted to the observed RR, reaching a maximum when there is only one reset every $125 \mu\text{s}$.

Masked resets operates over a $500 \mu\text{s}$ clock cycle. The nodes are organised into four groups, each starts at a time which is offset by $125 \mu\text{s}$ from the previous one. If r_o is the observed number of resets in a clock cycle and c , a counter which contains the difference $(4r - r_o)$ at the end of a clock cycle, then while c remains positive, the resets are masked from the VBR queues in the next clock cycle except for the first reset. For each reset masked, a value $(r - 1)$ is deducted from c , and masking ceases when c becomes zero or negative.

With this mechanism, the ring can reduce the throughput from the VBR queues down to 25% of that allocated as the system approaches overload, and thus provide the extra protection to the high priority services since they are unaffected.

5.2.3 Auto-Resets

Auto-reset is designed to protect the high priority services against two critical events: corrupted reset slots and sudden large surges in VBR load. Although the masking mechanism could guard against the latter occurrence, it could take up to $500 \mu\text{s}$ to react, hence auto-resets are necessary for the interim period.

Auto-reset operates as follows: If a node experiences no reset within a certain interval T , it automatically resets its own 'd' allocation. Currently, T is set equal to the maximum RI of $125 \mu\text{s}$. The node makes no attempt to reset others, instead each node resets itself when its own time-out expires. Auto-resets are not included in the reset count by the load monitor and thus will not interfere with the normal operation of the system. Only the VBR queues on those nodes with a zero c value will receive the auto-resets.

It must be noted that the Orwell protocol is still undergoing the standardisation process; therefore a number of features and parameters described here may be subjected to future modification as more is learnt about the protocol. It is hoped that this simulation study will contribute towards this process by providing information on the ability of the protocol in coping with VBR video traffic.

5.3 The Orwell Ring Simulation

The simulation incorporates all the features of the Orwell protocol. Briefly, the simulated system may be described as follows:

A slotted ring with randomly distributed nodes is subjected to a traffic load comprising CBO video, CBO voice, data and VBR video cells. The voice and video services are two-way calls, each with an inter-call arrival interval and a call holding time which are Negative Exponential distributed. The CBO video and voice are 2 Mb/s and 64 kb/s connections respectively and generate cells at regular intervals, while the VBR video connections generate cells at an irregular rate with a mean data rate which is user specified (see section 4.5). All data connections are assumed to have the same channel rate such that cells are generated at a regular rate over the duration of a message; the inter-message arrival interval on a node and the message length are assumed to have Negative Exponential distributions.

The nodes act as traffic concentrators for the range of traffic described. Access priority is in the order of CBO video, voice, data and VBR video. The data service was given a higher priority than the VBR video (despite being more delay and loss tolerant) in order to prevent its allocation, which is small, from being totally pre-empted by the VBR video service. Priority handling is such that if a cell from a low priority service is waiting, its access to the ring can be interrupted by the arrival of a cell from a higher priority service. Only when no more cells are waiting in the higher priority queues or the queue allocations have been exhausted, will a cell from a lower priority queue be launched.

Only a 34 Mb/s ring with 10 nodes attached was investigated. Two slots, each with a 4 byte header and a 16 byte information field, circulate in the ring. The maximum RI was 125 μ s and the RR was averaged over an interval of 2 ms. The threshold below which masking commences was initially recommended to be 32 resets/2ms (or 2 resets/125 μ s).

The CBO video and voice services were assumed to have mean call holding times of 0.25 s and 0.1 s respectively.* In view of the low speed of the ring, the CBO video was modified to have a smaller throughput of 1 Mb/s so that more calls could be supported and load arising from this service more evenly distributed. The delay requirements for both services were to be less than one cell assembly time, i.e. 125 μ s for video and 2 ms for voice, hence only a buffer of a single cell was allocated to each connection. The maximum number of each type of calls allowed on a node simultaneously was unrestricted. The allocation of 'd' is straight forward for the video service where a 'd' of 1 was allocated for every connection. For the voice service, a load smoothing mechanism was built in such that exactly 1/16 of a 'd' was allocated to each connection.

*see note on page 112

The data service was assumed to have a mean message length of 1000 bits, and data devices were assumed to have a channel rate of 19.2 kb/s. A buffer of 20 cells was allocated on each node for this service. The 'd' allocation method used was simply a fixed allocation of one 'd' per node. The VBR video service was assumed to have a mean call holding time of 0.27 s^{*}, with the first frame being transmitted at a quarter of the peak bit rate over the first 20 ms. The VBR video connections were assumed to have a population mean bit rate of 350 kb/s, with the individual connections having a mean bit rate between 180 kb/s and 530 kb/s and a possible peak in excess of 1.5 Mb/s (intra-frame variation could cause this to be even higher). The population mean approximates the CCITT 384 kb/s video-conferencing standard. The maximum number of simultaneous VBR video connections allowed on a node was also unrestricted.

With most of the parameters of the ring and the offered load fixed, the simulation study was carried out under different loading conditions by varying the call arrival rates and the data message arrival rate. A symmetrical loading of the ring was always assumed. The auto-reset mechanism was disabled so that the performance of the protocol could be studied in the presence of the VBR video traffic without the influence of auto-resets, which in any case should only be a contingent measure and should not be activated frequently. Such a study could help to determine an optimum expiry time for the activation of the auto-resets.

5.4 Aspects of Interest

As the objective of this work was to study the effective transfer and control of VBR video over the Orwell Ring, the main emphasis of this simulation study was on the various aspects related to this subject. The following aspects of interest were identified for investigation.

Ring Dimensioning In order to provide a pre-specified quality of service (QOS) in terms of call blocking and cell loss to the users, the ring must have a certain amount of capacity available for a given offered load. Since the ring speed is fixed in this case, the offered load must be dimensioned to fit the ring capacity. For CBO services, this can be done with the established Erlang Capacity Table^(4,18), but with VBR video, the Table cannot be used directly because each call has a different bandwidth requirement which is time varying. A modified dimensioning method is therefore necessary for the VBR video service in order to achieve the required QOS.

Ring Bandwidth The bandwidth available on the ring is significantly higher than its source release counterpart, but the exact gain is not known since it varies depending on the distribution of the load. An estimate of the mean bandwidth available will therefore need to be made, and it will form the basis for subsequent work. Some possible ways of improving the bandwidth gain are considered.

Call Acceptance Thresholds An optimum call acceptance threshold for the VBR video service must be determined, such that the acceptance of an additional call will not increase the risk of ring overload beyond a required level in line with a given cell loss performance, and yet achieve high network utilisation. Considerations must be given to the bursty nature of the ring load and the non-deterministic nature of the new calls. The thresholds for other services should be assigned with respect to that of the VBR video.

A 'd' Allocation Scheme A satisfactory 'd' allocation scheme is required which will allow VBR video to access the unused network bandwidth as the needs arise, and to restrain them when excess bandwidth is unavailable, hence reducing the risk of ring overload. This would form part of the traffic control strategy to complement reset masking and call blocking.

Masking Threshold This threshold must be correctly set to ensure that resets are not masked unnecessarily from the VBR queues, and so avoid unwarranted cell loss due to unfair overload control.

Observation Interval The observation interval over which the RR is measured is currently set at 2 ms. This has been based mainly on considerations for CBO services; with VBR video, the ring load is highly bursty and a longer observation interval may be necessary in order to provide a better estimate of the RR, and hence, the unused bandwidth on the ring.

Buffer Size The buffer on each node must be sufficiently large to absorb some of the irregularity of the VBR video load arriving at the nodes, and to sustain some short-term ring overload due to the bursty nature of the load on the ring, thus minimising cell losses. It must not, however, lead to an excessive cell delay.

Cell Loss and Cell Delay These measurements for the VBR video service with different levels of offered load and different settings of the traffic control parameters, are particularly useful in providing information on the best way of transmitting VBR video over the Orwell Ring and on the performance of the system as a whole. Cell delay for the other services in the presence of VBR video is also of great interest. There should not be any cell loss on the CBO services even under overload conditions if proper control has been exercised.

Auto-resets It has been tentatively decided that auto-resets should be activated if no reset is encountered within an interval of 125 μ s (which corresponds to the proposed maximum RI). Considerations must be given to the fact that only an average RI of 125 μ s is maintained by the protocol, for instance over a period of 2 ms; the individual RIs could still exceed the maximum. The choice of a 125 μ s expiry interval may therefore result in many unnecessary auto-resets.

5.5 Dimensioning of the Ring for VBR Video

In order to provide a certain QOS on a network both from a cell loss and call blocking perspective, sufficient network capacity must be made available to meet the anticipated user traffic level. For CBO services such as voice, the network capacity required for a given load can be obtained from the well-established Erlang 'B' Capacity Table (Appendix III) for the desired call blocking performance. Cell loss other than those due to bit error is inhibited by the protocol and need not be considered in this case.

However, with VBR video, the situation is more complex and there is no established method of dimensioning the network for a given load and a required QOS. The difficulties stem from the non-deterministic nature of the calls and their variable bandwidth requirements, and as such, no fixed channel capacity can be allocated to calls individually. The bursty network load complicates the issue further, particularly in the case of the Orwell Ring, where call blocking is based on a dynamic bandwidth estimation mechanism. Cell loss performance is an extra factor which requires consideration in addition to call blocking, when compared to networks carrying only CBO services. With VBR video, the effect of statistical multiplexing of the traffic also has to be considered.

In view of the probabilistic nature of the traffic and the bandwidth available on the ring, a simple relationship linking the offered load, the network capacity and the call blocking and cell loss performance is difficult to derive. Hence, only an approximate method for dimensioning the ring is suggested, and it is only intended to provide a rough estimate of the network capacity required for a given load and for an approximate QOS. The method is as follows:

If C is the anticipated call arrival rate, and T the mean call holding time, then the traffic intensity is given by E , where

$$E = C \times T \quad \text{erlang} \quad \dots(5.1)$$

Using the Capacity Table, the number of simultaneous two-way connections, N , which may be accepted onto the ring before call blocking commences can be obtained. The maximum number of calls the ring is required to support is thus $2N$ for E erlang of traffic. However, there will be an overload probability of P in terms of call attempts where the excess calls will have be blocked.

Under static traffic conditions, the maximum VBR video load on the network would have a mean, assumes source independence, of

$$\eta_c = \sum_{i=1}^{2N} \eta_i \quad \dots(5.2)$$

and a variance of

$$\sigma_c^2 = \sum_{i=1}^{2N} \sigma_i^2 \quad \dots(5.3)$$

where η_i and σ_i^2 are the mean and variance of the bandwidth requirement of the individual calls, and are normally expressed on a frame basis, i.e. bits/frame. An even load is assumed within the frames.

However, η_i and σ_i^2 are different for different calls, and they are unknown to the network. Consequently η_c and σ_c^2 cannot be calculated at the time of dimensioning. But if the underlying probability distributions of η_i and σ_i^2 are assumed constant for a particular class of VBR video service, then by the sampling theory, $\eta_c/2N$ and $\sigma_c^2/2N$ will approach the population means of the two parameters respectively as N increases.

It is therefore useful to define a standard video source with characteristics η_c and σ_c^2 , which are the population means of the two parameters for a given class of VBR video source, and which can be obtained by monitoring real video traffic over a network.

Then if N is large, which is the case for the VBR video gain to be realised, η_c and σ_c^2 can be approximated from η_i and σ_i^2 respectively. For simplicity, the uncertainties in the values of η_i and σ_i^2 , which diminish as N increases, are ignored in the approximation.

From the central limit theorem, we know that as the number of sources increases, the probability distribution of the total load will become Normal. Thus from the table for areas under the Normal curve [Appendix III], we can determine the amount of bandwidth, $k\sigma_c$, which must be provided on top of η_c in order to achieve a given probability f , that the load exceeds the allocated bandwidth. This, however, does not explicitly indicate the actual probability of cell loss or cell loss rate. An approximation of the cell loss rate can be made by first obtaining the mean amount of excess load (bits/frame-time) by

$$L_x = \int_B^{\infty} p(n) \times (n - B) dn \quad \dots(5.4)$$

where $p(n)$ is the probability of the load being n bits/frame-time and $B = \eta_c + k\sigma_c$.

If the cell has an information field of c bits, then the mean load and mean excess load expressed in number of cells will be η_c/c and L_x/c , and the cell loss rate r will be

$$\begin{aligned} r &= \frac{L_x/c}{\eta_c/c} \\ &= \frac{L_x}{\eta_c} \end{aligned} \quad \dots(5.5)$$

This is only true for the static case of $2N$ calls. However, since there is only a probability of P that there are $2N$ calls established on the network, and if cell loss is assumed negligible when the number of calls in progress is less than $2N$, the cell loss rate r_o must be

$$r_o = P \times r \quad \dots(5.6)$$

The bandwidth required for E erlang of VBR video with an overload probability of $(P \times f)$ or a cell loss rate of r_o would be approximately

$$B = \eta_c + k\sigma_c \quad \dots(5.7)$$

and ideally call blocking should commence when the mean network capacity utilisation reaches η_c . This mean utilisation is, however, not always possible to determine in practice because of the dynamic nature of the system. This problem will be discussed when considering call acceptance threshold (section 5.6.3).

On a *per source* basis, the capacity required for a single call would be

$$\begin{aligned} (\eta_c + k\sigma_c)/2N &= [2N\eta_c + k(2N\sigma_c^2)^{1/2}]/2N \\ &= \eta_c + k\sigma_c/(2N)^{1/2} \quad \dots(5.8) \end{aligned}$$

It can be seen that as N increases, the extra capacity required per channel decreases. This is consistent with the earlier suggestion that the VBR video gain is improved when a large number of sources are multiplexed in a shared resource environment.

An approximate method of dimensioning the network capacity for VBR video traffic has been described. It must be noted that many of the assumptions are based on the statistical multiplexing effect which only applies exactly for a large number of uncorrelated sources. Therefore this method will only work well with a relatively high level of VBR video traffic. These assumptions need to be treated with care when the traffic level is moderate or low. First of all, the uncertainties of the estimated η_c and σ_c^2 , which were not accounted for in the simple method above, may be significant and may need to be considered. Secondly, the assumption that the multiplexed load has a Normal distribution is not strictly true, particularly in the tail region if the number of sources is small. Yet this region is important in the cell loss estimation. Furthermore, the video load has been assumed to be smoothed within frames which is not entirely correct; thus the instantaneous bandwidth may be larger and may result in more cell loss. The actual cell loss rate will also depend on other factors such as the bandwidth allocation scheme and the traffic control strategy employed, as well as the size of the buffer allocated.

From the above discussion, it would appear that the method is deficient in many respects. This is because of the highly non-deterministic nature of the VBR video traffic and the difficulties in accounting for external factors such as buffering. It must be stressed that the method was only meant to provide an estimate of the amount of network capacity required for a given traffic level and QOS.

The advantage of this method is that call control is based on some measure of the bandwidth utilisation, rather than on the number of established calls and their individual characteristics; the latter approach is unsuitable for the Orwell Ring where call control is completely distributed. There are, nevertheless, some problems which need to be resolved: the most important being the correct determination of the mean bandwidth utilisation in a dynamic load environment; with the Orwell Ring, there is a further complication in that the ring capacity is non-constant. The accurate dimensioning of the ring is therefore unnecessary.

5.6 The Simulation Study

The ring simulation was run with different traffic mix and traffic levels, and with different settings of the load control and overload control parameters. The simulation results were collected over 2.5 s of simulation time after a warm-up period of 1.25 s. This represents the minimum simulation time considered adequate to provide sensible results for the call holding times used.

The results would of course be more reliable with longer runs or with a large number of runs, but this is not possible in this work given the computing resources available. Thus when investigating the effects of the parameter settings, only a few runs were made; they usually included the observed worst cases in terms of cell loss, the use of which was to enable ways of reducing the cell loss to be studied. More runs were only made with parameter settings which were deemed optimum or where further confirmation was required on the results obtained. Some of these results have been published in two recent papers^[13,14].

Table 5.1 gives a description of the headers used in the tables of results presented in this section.

| <i>Column Heading</i> | <i>Description</i> | <i>Unit</i> |
|-------------------------|---|--------------|
| T_a | call acceptance threshold | resets/2ms |
| T_c | ceiling threshold | resets/2ms |
| T_m | masking threshold | resets/2ms |
| Offered Load | offered traffic load : (video & voice) | calls/s |
| | (data) | Mb/s |
| Buffer Size | size of the common node buffer provided for VBR video | cells |
| Min./Mean RR | minimum and mean observed reset rate | resets/2ms |
| No.of MR | number of masked resets total over all nodes | - |
| d Alloc. Denied | requests for 'd' allocation increment from VBR video which are rejected, total over all nodes | - |
| CLR | cell loss rate | (ratio) |
| Call Blocking | call blocking rate | (ratio in %) |
| Monitoring Interval | dynamic 'd' allocation monitoring interval, over which the cell arrival rate is measured | ms |
| Max./Mean Cell Delay | maximum and mean cell delay | μ s |
| Observation Interval | interval over which reset rate is measured | ms |

Table 5.1 Description of the column headings used in the tables of results in this chapter.

5.6.1 Ring Bandwidth Estimation

The problem with deriving this bandwidth using analytical methods arises from the difficulty in determining the magnitude of the statistical gain obtainable from the destination release policy. This gain is not constant but varies with the distribution of the load and the traffic pattern; it also depends on the ring configurations⁽⁵⁵⁾, notably the number of nodes attached. Nevertheless, the lower and upper bounds of the ring bandwidth can be calculated. The lower bound corresponds to a unity gain situation where each slot delivers exactly one cell per ring revolution. For a two-slot 34 Mb/s ring with a slot structure of 20 bytes of which 16 form the information field, the usable bandwidth available would be:

$$34 \text{ Mb/s} \times 16/20 = 27.2 \text{ Mb/s}$$

But some of this bandwidth is required for the trial and reset operations. Assuming both slots are not carrying cells during the trial period, they go to waste. A further slot is required for reset. Thus the total number of slots wasted every 125 μs , i.e. during full load, would be three. The maximum proportion of the bandwidth required for these two operations is therefore

$$\begin{aligned} & D/125 \times 100\% \\ & = 11\% \qquad \text{where } D \text{ is 3 slot-time in } \mu\text{s} \end{aligned}$$

leaving a net bandwidth available for information transfer of only about 24 Mb/s. This figure represents the lower bound of the usable bandwidth. Even if the above assumption is relaxed, 2 slots would still be wasted, and the bandwidth available would be 25 Mb/s.

On the other hand, if each slot delivers on average 2 cells per revolution, the upper bound of the bandwidth available would be around 50 Mb/s. However, this bandwidth can only be achieved if both slots released are reused almost immediately⁽⁴⁰⁾. In practice, this is not the case for several reasons. First of all, the ring cannot be loaded to such an extent to ensure immediate slot reuse, especially with the 125 μs maximum RI. Secondly, the nodes are not allowed to transmit all the time because of their 'd' allocations, as a result, empty slots have to hunt for an active node and the time spent is translated into wasted bandwidth. Furthermore, there must be a large number of nodes and a well distributed load to ensure a high degree of slot-reuse such that slots do deliver on average 2 cells per ring revolution.

In view of the many uncertainties in ascertaining the average bandwidth gain, the ring simulation was run with the intended configuration to estimate the average capacity available on the 34 Mb/s ring. Only voice traffic was used since a higher loading-factor could be achieved, and its constant bit rate characteristic enabled the bandwidth available on the ring to be deduced from the Erlang Capacity Table. The results are presented in Table 5.2, and from the Capacity Table, an average ring bandwidth of about 30 Mb/s was estimated. This figure represents a gain of 25% above the minimum capacity of 24 Mb/s; it must be pointed out that this gain fluctuates with time as the load varies.

| Number of Nodes | Offered Load (calls/s) | Original Ring | | Split-Node Ring | |
|-----------------|------------------------|-------------------|---------|-------------------|---------|
| | | Call Blocking (%) | Mean RR | Call Blocking (%) | Mean RR |
| 10 | 2230 | 1.31 | 22.7 | 0.97 | 23.7 |
| | 2250 | 1.62 | 22.3 | 1.37 | 23.0 |
| 20 | 2230 | - | - | 0.05 | 30.9 |
| | 2250 | - | - | 0.00 | 30.0 |
| | 2500 | - | - | 0.27 | 24.0 |

Table 5.2 Call blocking performance for voice traffic with different configurations of the ring and a call acceptance threshold of 17.

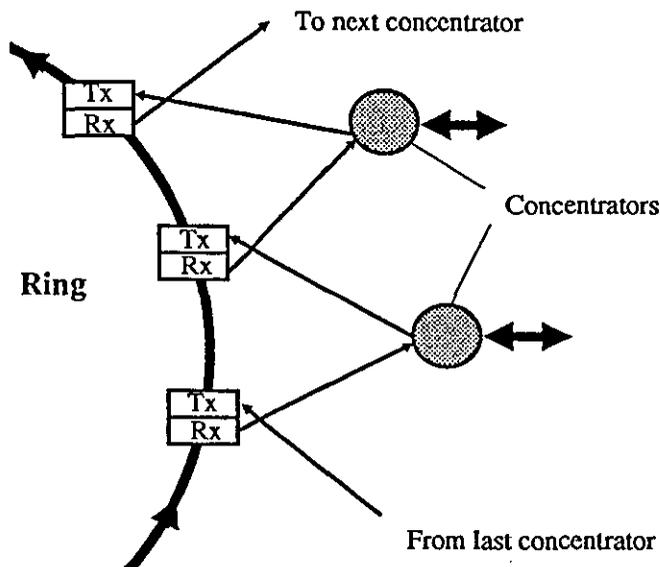


Fig. 5.1 The Split - Node Concept.

This bandwidth is, however, only strictly applicable to a highly loaded ring with a well distributed voice traffic. With wider bandwidth services, this condition does not hold because fewer calls can be supported on the ring, with the consequence that the load is less well-distributed and an asymmetric loading condition arises. Under this circumstance, the gain would be attenuated as the probability of slot-reuse is reduced.

A split-node concept was proposed to alleviate the problem of asymmetric loading. With this concept, a node will have its transmit function located on a downstream node while the receive function remains on the original node. The transmit function in the original node will be used by an upstream node and so on. All connections are connected to a pair of adjacent nodes as shown in figure 5.1. This concept is practical in the backplane version of the Orwell Ring where the nodes are very close together, and to some extent, in the distributed ring as well since nodes are generally clustered together.

The principle behind this concept is to overcome the objection of a node using a slot it releases. This limitation comes about from an implementation consideration. By removing this restriction, improvement in the bandwidth gain can be anticipated since some of the slot 'hunting time' can be reduced. However, the advantage of this concept will only be apparent with a highly asymmetric two-way traffic, where the gain could approach a factor of 2 instead of a unity gain with the original configuration. With an even or slightly asymmetric load, the gain is not expected to increase significantly, but it may help to maintain the gain when large bandwidth services are considered.

This concept was implemented in the ring simulation but no comprehensive comparative study was conducted between the new and the original ring configurations. Results obtained with voice load are shown in Table 5.2 and indicate only marginal improvement in the gain - possibly in the order of 500 kb/s. The average capacity available on the ring remains at about 30-31 Mb/s (25%-30% gain).

It was suggested earlier that with a larger number of nodes, a greater degree of slot-reuse could be achieved with a subsequent improvement in the bandwidth gain. As a matter of interest, a simulation was carried out with 20 nodes and the results are shown in Table 5.2. A substantial gain in the ring bandwidth was observed, and the magnitude was estimated to be in the region of 4 Mb/s, taking the ring capacity up to around 35 Mb/s (46% gain). However, the simulation does require more computing power and the gain may not be very important for this work. It was therefore decided to work only with a 10 node ring for economy in computing resources.

5.6.2 The Reset Rate - Ring Load Characteristic

Since the reset rate assumes a key role in the operation of the protocol, its relationship with the amount of load on the ring must be established. This relationship between the RR and the ring load is shown in figure 5.2. Four curves are shown which correspond to the original and split-node ring with 10 nodes, the split-node ring with 20 nodes, and the projected worst case when one slot carries exactly one cell per ring revolution. The variation in the bandwidth gain for the different configurations can clearly be seen, and as expected, the gain increased with the ring load. But more strikingly, the gain appeared to accelerate after a load of about 28 Mb/s (which corresponds to a RR of 22). The curves tend to level off after this point, suggesting that more load could be carried on the ring without affecting the RR significantly. However, the current setting of a 125 μ s maximum RI (a minimum RR of 16) means that a large proportion of the bandwidth goes to waste compared to the earlier recommendation of a maximum RI of 2 ms^[22].

In figure 5.2 the region between the RR of 32 and 16 is of particular interest. This is the critical region where the ring is progressively approaching an overload condition. In this region, the curve for the split-node ring with 10 nodes is approximately linear, such that an increase in the load of about 1 Mb/s leads to a decrease in the RR by 2. This relationship was used extensively when considering threshold settings. The ring characteristic will obviously depend on the speed and the configuration of the ring, and will need to be determined *per se* or be approximated by some means.

5.6.3 Traffic Control and Thresholds Considerations

Traffic control on the Orwell Ring is achieved with three mechanisms: call blocking, 'd' allocation and masking. These form the subjects of the current investigations and this section deals primarily with VBR video traffic. Traffic control on CBO services is relatively straight forward and will be considered in the next section.

First of all, the VBR video traffic must be dimensioned for the ring. The standard source can be estimated in this case since the characteristics of the model is known (assuming a Uniform distribution for the mean bandwidth requirements of the calls and a known relationship between the means and the variances of the frame differences). The standard source has a η , of 350 kb/s which is the population mean. The relationship between the mean and the variance of the FD of a source is given by^[51]:

$$\sigma^2 = \eta^2/3^2 \quad \dots(5.9)$$

where η is the mean FD expressed in percentage.

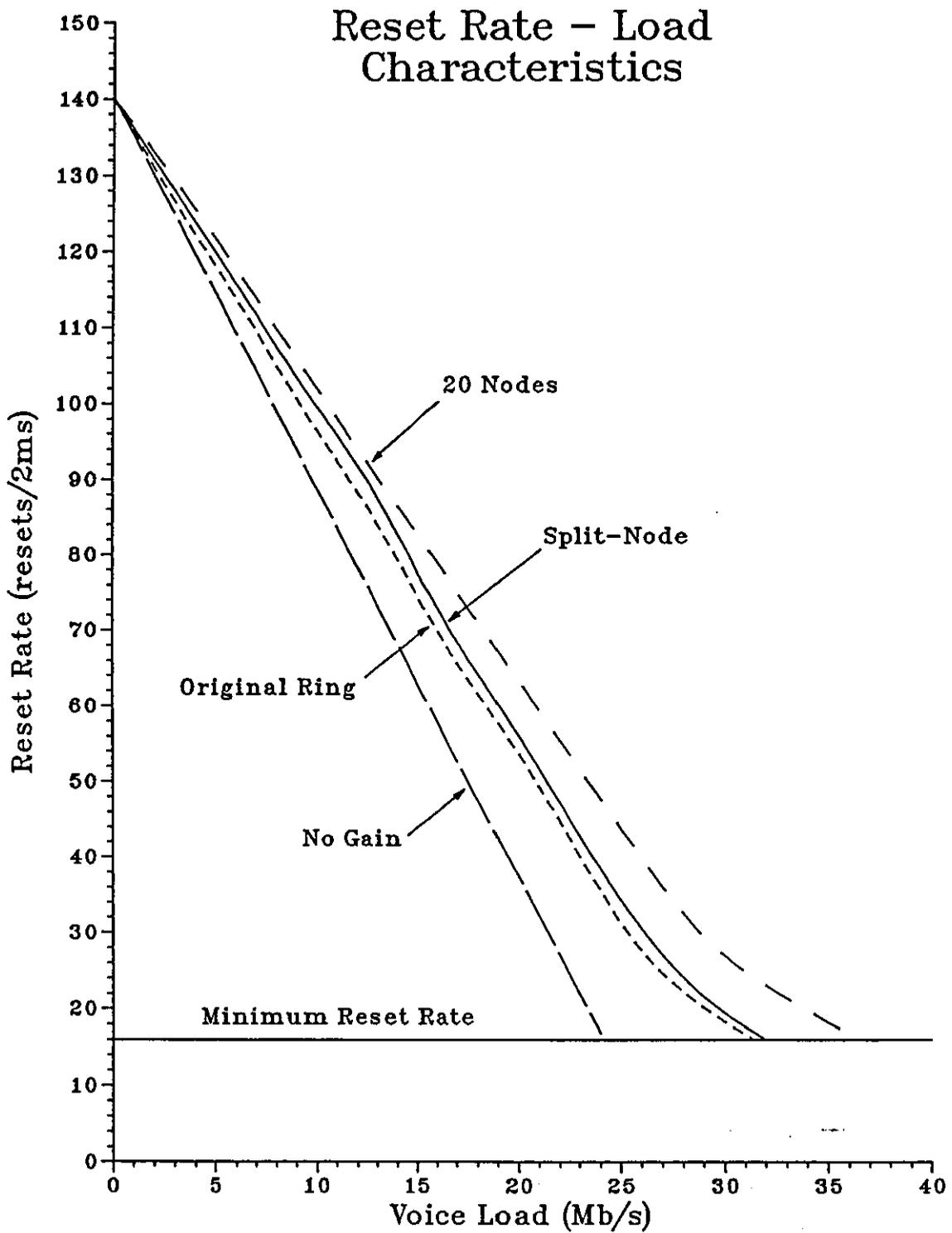


Fig. 5.2

The mean variance (σ_r^2) for a Uniformly distributed mean FD between 8% and 22% is therefore

$$\sigma_r^2 = \int_8^{22} p(\eta) \times \eta^2 / 3^2 d\eta \quad \dots(5.10)$$

Since the mean FD of the sources are Uniformly distributed, $p(\eta)$ is a constant; by integrating the expression and converting the FD to bits per second, σ_r^2 was calculated to be 43556 (kb/s)².

From equation 5.7, the bandwidth B for a call blocking rate of 1% and an overload probability f of 10^{-4} (i.e. with $k = 3.7$) is thus

$$B = 2N \times \eta_r + k \times (2N \times \sigma_r^2)^{1/2} \quad \dots(5.11)$$

Solving for N with B = 30 and 31 Mb/s gives 33 and 35 two-way connections respectively. These are equivalent to 23 E and 24 E of VBR video traffic. With a call holding time of 0.27 s, these represent call rates of 85 and 90 calls per second. The actual cell loss rate (CLR) would be in the order of 10^{-8} (from equations 5.4-5.7). In view of the many uncertainties which were not taken into account in these calculations, such as the uncertainties in η_r and σ_r^2 and the non-constant ring capacity, a more cautious choice would be to work with an offered load of 23 E.

From equation 5.7, the mean bandwidth occupied during full load on the ring is about 23 Mb/s, while the excess bandwidth required is 6.3 Mb/s - representing a 27% overhead. This will only affect the gain of VBR video slightly, and the proportion of this overhead will diminish with more VBR video load on a larger system. But with a 34 Mb/s ring, the VBR video bandwidth gain is reduced by about 20%.

5.6.3.1 Call Acceptance Threshold and Call Blocking

With an offered VBR video load of 23 E, calls should be blocked when the mean load on the ring reaches about 23 Mb/s for the cell loss and call blocking performance specified. Using the reset rate-load characteristic in figure 5.2, this load corresponds to a RR in the region of 40, and hence the call acceptance threshold (T_c). This would be the case if the load on the ring can be correctly estimated from the RR, but this is not strictly true. Furthermore, the uncertainty in the reset rate-load characteristic can be significant at this loading level.

The difficulty in estimating the load from the RR arises from finding an optimum observation interval over which the RR can be determined. A short interval, such as the 2 ms interval used, will provide an estimate of the most current short-term load condition but not the mean load; on the other hand, although a long observation interval provides a better mean load estimate, it may fail to account for the most current load variation due, for instance, to

call establishments or terminations. This gives rise to a contradicting situation and the search for an optimum observation interval will require extensive simulation work. But because of the dynamic nature of the ring load, a simple load estimation mechanism using only a single RR measurement may never be sufficient to provide an accurate estimate. Figure 5.3 shows the variation of the RR with time measured at 2 ms interval, and helps to illustrate this point.

Thus when working with the 2 ms observation interval, a T_r of 40 may result in a significantly higher call blocking rate than is expected ($>1\%$) because of the bursts in the ring load; allowance must also be given to the uncertainty in the reset rate-load characteristic. Consequently, it was decided to lower the T_r assuming that the observed RR of around 40 is, in most cases, an over-estimate of the mean load on the ring - since there is only a small probability that the mean load would exceed 23 Mb/s. This, nevertheless, also increases the risk of ring overload when excess calls are being accepted by the use of a lower threshold, especially when overload in terms of call attempts arises.

An investigation into the QOS for the VBR video with different T_r s for a given load was conducted. The 'd' allocation scheme and the other parameter settings used are discussed in subsequent sections. The results are tabulated in Table 5.3. These results were obtained with 5 sample runs for each threshold. In hindsight, these are probably not the most representative of the samples, as they represent some of the worst case situations both in terms of cell loss and call blocking. However, the emphasis here is on the relative QOS performance with different thresholds, and not on the absolute performance.

| T_r | Call Blocking(%) | CLR | No. of MR | d Alloc. Denied |
|-------|------------------|----------------------|-----------|-----------------|
| 22 | 0.63 | 1.3×10^{-4} | 64 | 230 |
| 24 | 0.72 | 1.3×10^{-4} | 69 | 225 |
| 26 | 0.90 | 4.8×10^{-5} | 32 | 152 |
| 28 | 1.00 | 4.8×10^{-5} | 32 | 152 |
| 32 | 1.25 | 3.9×10^{-5} | 19 | 85 |
| 40 | 3.24 | 0.0 | 1 | 0 |

Table 5.3 Results for the variation of call acceptance threshold (T_r) with a VBR video load of 85 calls/s; with masking threshold (T_m)=16, ceiling threshold (T_c)=18 and buffer=10 cells.

The thresholds were chosen such that there would be sufficient capacity to support the mean bandwidth requirement of the new connection as well as provide room for bursts in the ring load. If the thresholds were to be used in a mixed services environment, provision must also be given to the background 'd' allocation of the data service, which is also bursty in nature. The minimum threshold investigated was 22; at this RR, there is approximately 3 Mb/s available for the new two-way connection (0.7 Mb/s mean) and any bursts in the ring load that may arise.

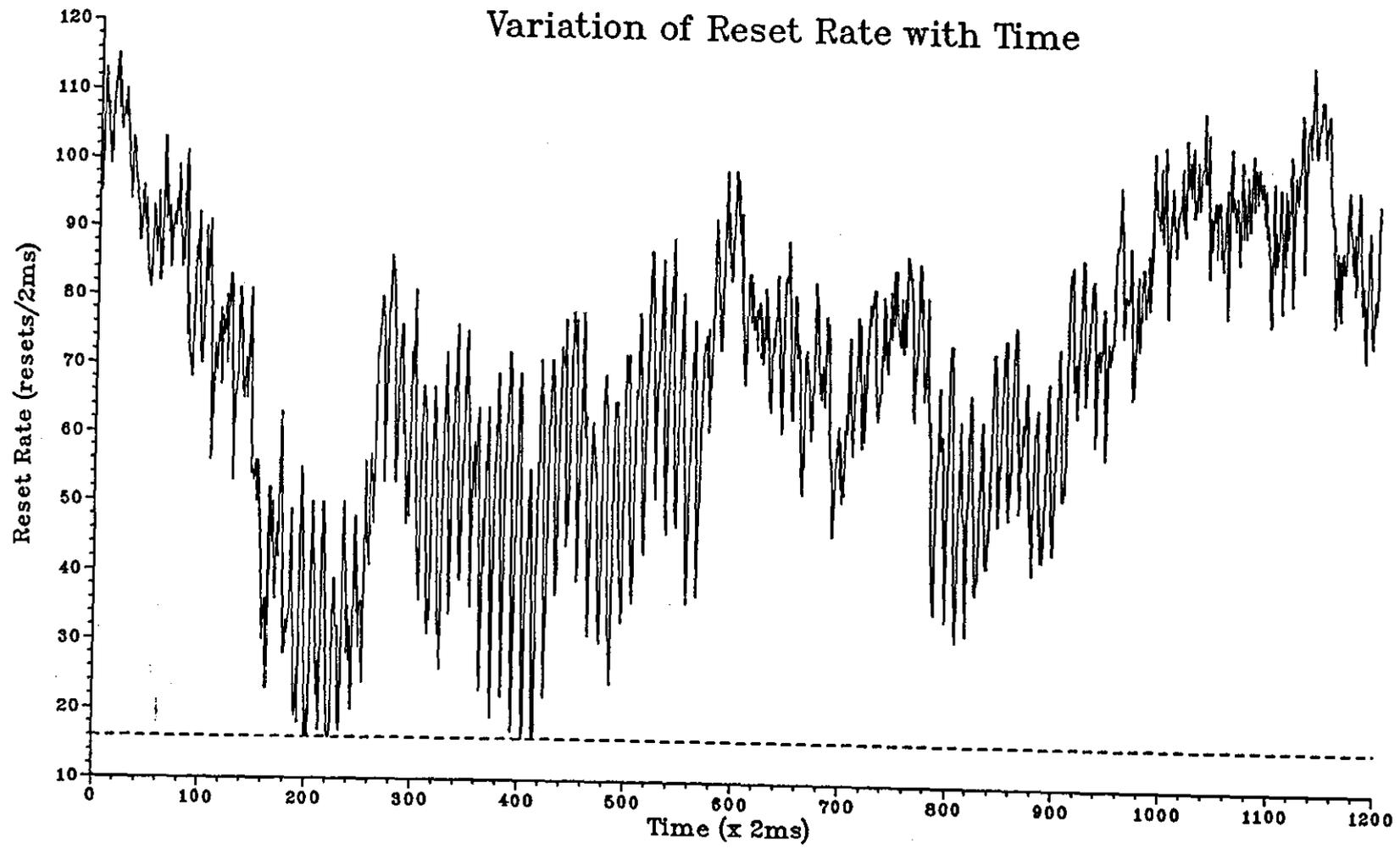


Fig. 5.3

From Table 5.3, it is obvious that a T_r of 40 is inappropriate. The QOS is quite similar for the threshold range of 22 to 32, with the expected increase in the call blocking rate and decrease in the cell loss rate (CLR) as T_r was increased, but the differences are small. The threshold of 22 is, however, not considered suitable because it leaves very little room for bursts in the ring load, especially if data service is also present. A threshold of 24 is regarded as being the minimum with at least 4 Mb/s remaining on the ring. Higher thresholds would provide a greater margin of safety but attain a proportionally smaller bandwidth utilisation, which may not be objectionable. The call blocking performance with a T_r of 32 or below was within specification, and more runs were considered necessary to enable a final choice of the threshold to be made (see section 5.6.5).

The CLR's appeared to be unexpectedly high in all cases; this could be a consequence of the use of low T_r 's, which would inevitably lead to the occasional acceptance of extra calls onto the ring, thus increasing the risk of cell loss. More importantly, because of the short simulation time used (2.5 s per run), these CLR's could represent the short-term cell loss performance. Since losses tend to occur in bursts, a high CLR can be expected when losses occur in a run. Other factors which could contribute to these high CLR's are the instantaneous data bursts within the frames, which were ignored in the cell loss estimation, as well as the limitations of the 'd' allocation scheme and the small buffer used (10 cells). The high CLR was investigated further to determine the actual cause of these losses.

5.6.3.2 The 'd' Allocation for VBR Video

Since the characteristics of the VBR video sources differ from one source to another and their instantaneous bandwidth requirements vary with time, a deterministic way of allocating bandwidth is not considered suitable for the Orwell Ring. However, a deterministic bandwidth allocation scheme was briefly examined whereby each connection was allocated a 'd' value which corresponded to half its peak bandwidth requirement. Thus when the ring is lightly loaded such that the RR is greater than 32, a connection can obtain bandwidth up to its peak requirement. As the ring is gradually loaded, the connection will eventually be limited to the bandwidth it was allocated, i.e. half the peak, at the minimum RR of 16. There are several problems with this method. First, it is difficult to determine the peak bandwidth requirement of a connection at call set-up time if this is not to be the absolute peak. Second, the total amount of the ring bandwidth already allocated is not known to the nodes as call control is completely distributed and based only on an estimate of the ring load; the total bandwidth allocated across the ring could therefore be far greater than the ring capacity. The allocated, yet unused, bandwidth could cause an overload if a surge in the video load occur in one or more nodes. Furthermore, connections with a peak-to-mean ratio of less than 2 will

suffer heavy losses at full load. The study on this scheme in a mixed services environment has also revealed its weakness in failing to protect the CBO services against cell loss under overload conditions.

A better approach, which was adopted in this work, is a dynamic 'd' allocation scheme, whereby the bandwidth requirement of the combined VBR video sources on a node is continuously being monitored and the VBR video 'd' allocation updated accordingly, subject to there being sufficient bandwidth remaining on the ring. This way, the nodes are only given the exact amount of bandwidth they require at any instance, thus avoiding the build-up of unused bandwidth as in the deterministic bandwidth allocation scheme above. This scheme also has an inherent load control capability which denies further 'd' allocation to the nodes when the ring is approaching an overload condition.

The realisation of this scheme was by the use of a counter on each node to count the number of cells arriving in a given interval. The 'd' required, which is simply the average number of cells arriving every 125 μ s, is then compared with the last 'd' allocation. If the current 'd' requirement is greater than the last 'd' allocated and the RR is above a pre-specified threshold (called the ceiling threshold, T_c), the 'd' allocation for the VBR video queue will be incremented by 1. The 'd' allocation is only allowed to increment in steps of 1 to avoid overloading the ring and to allow for a better sharing of the available bandwidth with other nodes. Conversely, if the current 'd' demand is less than the last 'd' allocation, the current 'd' will become the new 'd' allocation of the queue. A minimum 'd' allocation is always maintained on each node to avoid a node from being deprived of its VBR video bandwidth allocation when its output data rate is temporarily low. This minimum 'd' should be related to the expected VBR video load on a node, but in this work it was conveniently set to one (1 Mb/s), in order to reduce the complexity of the scheme. The flow chart for this scheme is shown in figure 5.4.

The two parameters which could affect the performance of this scheme are the ceiling threshold (T_c) and the interval during which the node counts the arriving cells. Their effects were investigated and the results are presented in Tables 5.4 and 5.5 respectively.

The setting of the T_c should be as low as possible so that the ring capacity can be utilised fully. But since one of the main functions of the scheme is load control, the provision of more bandwidth to a node must not cause the ring to overload to such an extent as to incur cell loss on the CBO services. There must also be sufficient bandwidth between the call acceptance threshold and T_c to enable a new call to acquire the bandwidth it needs. The range of T_c s of interest for a minimum call acceptance threshold of 24 and a counter interval of 2 ms was 16 to 22. Table 5.4 shows improvement in the cell loss performance with decreasing T_c as expected. There was however only a small improvement in the cell loss

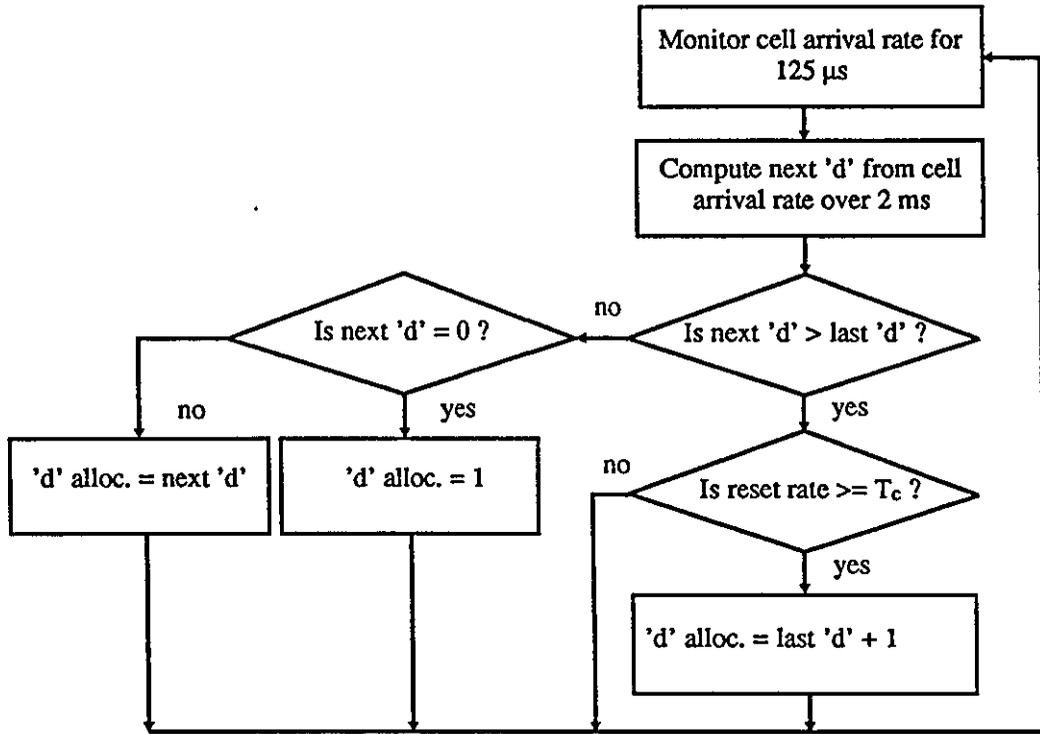


Fig. 5.4 Flow Chart for the Dynamic 'd' Allocation Scheme.

| T_c | Call Blocking(%) | CLR | No. of MR | d alloc. denied | Min. RR |
|-------|------------------|-----------------------|-----------|-----------------|---------|
| 16 | 0.72 | 1.26×10^{-4} | 85 | 21 | 15 |
| 18 | 0.72 | 1.30×10^{-4} | 69 | 225 | 16 |
| 20 | 0.72 | 3.52×10^{-4} | 25 | 760 | 15 |
| 22 | 0.72 | 3.92×10^{-4} | 31 | 1614 | 16 |

Table 5.4 Results for the variation of the ceiling threshold (T_c) with a VBR video load of 85 calls/s; with $T_s=24$, $T_m=16$ and buffer=10 cells.

| Monitoring Interval | Call Blocking(%) | CLR | No. of MR | d Alloc. Denied |
|---------------------|------------------|----------------------|-----------|-----------------|
| 2 ms | 1.10 | 1.9×10^{-4} | 295 | 74 |
| 1 ms | 1.00 | 8.6×10^{-5} | 476 | 112 |

Table 5.5 Results for the variation of the 'd' allocation monitoring interval with $T_c=18$; other settings were as Table 5.4.

performance by decreasing T_c from 18 to 16, which corresponds to an increase in the allowance between T_c and T_s from 3 Mb/s to 4 Mb/s. An interesting point to note is that the main traffic control mechanism used in the two cases was different.

With the T_c at 18, the dominant traffic control mechanism used was the 'd' allocation scheme which denied the provision of more 'd' to the nodes when the ring was approaching overload. Whereas with the T_c at 16, 'd' was allocated up to the verge of overload; traffic control was mainly by the masking mechanism which was activated when the RR fell below 16 (note that this is different from the recommended masking threshold of 32). The inter-change of the two traffic control mechanisms is evident from Table 5.4. It is also obvious from the results that there was still a large demand for more 'd' allocations after a RR of 18 but very little after a RR of 16; this suggests that the ring was correctly loaded such that the bandwidth was utilised efficiently and the RR rarely fell below 16 as required. RRs of 15 were observed but only very rarely, and they occurred in isolation and are thus of no particular concern.

A T_c of 18 thus appears to be premature; but there is a higher risk of overload with a T_c of 16 and rigorous tests are required to verify its suitability in a mixed services environment. The advantage of using a T_c of 16 is that the nodes get their bandwidth demand most of the time and the ring capacity is fully utilised. This could lead to a better cell loss performance. T_s s greater than 18 are not recommended as they result in high CLRs by stopping 'd' allocation prematurely as can be seen in Table 5.4.

It must also be stressed that because of the minimum RI, bandwidth can only be allocated in steps of 1 Mb/s. This implies that the nodes would almost certainly be allocated excess bandwidth, for example, a node requiring slightly more than 1 Mb/s would be given 2 Mb/s. The excess, but unused, bandwidth would have a similar effect as the background 'd' allocation of the data service in bringing about ring overload. It also implies that cell delay for VBR video will be very small.

The interval over which the node counts and averages the cell arrivals is another factor which could affect the effectiveness of the 'd' allocation scheme. In the previous section, this interval was conveniently set to 2 ms to coincide with the observation interval of the RR. It should be noted that these intervals are not discrete blocks of 2 ms, but are implemented as a window of 2 ms which slides along in steps of 125 μ s. With the 2 ms interval, there will be a delay, up to 2 ms, in the counter response to load variation and thus a delay in the 'd' allocation update. As a result, a small buffer (10 cells in this case) may not be sufficient to absorb the excess cells when the load is rising quickly. This becomes worse if the incoming load shows a sudden large increase while the 'd' allocation is limited to step increments of 1. There are two ways to alleviate this problem: one is to use a smaller counter interval, another is to use a larger buffer.

The first option is considered here. Table 5.5 shows the comparison of using a 2 ms and a 1 ms counter intervals for the 3 worst case runs in terms of cell loss. The results show a significant improvement in cell loss performance without affecting the call blocking performance. This indicates that node overload (in contrast to ring overload) caused most of the cell losses that were observed in all the previous runs, and resulted in the high CLRs. Even with a 1 ms interval, node overload could still occur. An interval of this duration, or shorter, although very responsive could result in a very fast changing 'd' allocation - being extremely sensitive to even small spikes in the load variation. This can lead to a highly variable load on the ring.

A better approach is perhaps to increase the buffer size in view of the cheap memory available nowadays. This has the added advantage in that the load on the ring can be partially smoothed by using a longer counter interval, i.e. a slower 'd' adjustment, and the use of the buffer to absorb small load variations on the node. This smoother ring load will in turn enable a better estimate of the load on the ring and reduce the risk of overload. This approach is examined in section 5.6.4.

5.6.3.3 Masking Threshold

Masked resets (described in section 5.2.2) is a concept used for overload control on the ring. It works by masking some of the resets from the queues of low priority bursty services in groups of nodes in turn, thus reducing the frequency of their 'd' allocations during overload. The number of resets to be masked is adjusted to the ring loading condition such that a sufficient amount of bursty load can be blocked from the ring to avoid serious overloading which could affect the CBO services. The ring loading level above which this mechanism commences operation is referred to as the masking threshold (T_m), and it is expressed in terms of the RR as this is the unit by which the ring load is measured.

Since the masking mechanism requires T_m to be given as an integer number of resets per 125 μ s, means that T_m can only be specified in steps of 16 when expressed on a 2 ms basis. Furthermore, the mechanism as detailed in section 5.2.2 does not provide for a T_m of 16. The reason being: If r is the masking threshold on a 125 μ s basis, then for every reset masked, the mechanism will subtract from c - the number of resets to be masked - a quantity equal to $(r - 1)$. But with a T_m of 16, r is 1, which suggests that once masking is activated, all the resets within the masking interval except the first, will be masked irrespective of the state of the ring load, i.e. maximum masking whenever overload occurs. This is rather undesirable as masking should be adjusted in accordance to the degree of overload, otherwise it could result in drastic changes in the ring load and cause other complications.

Since a T_m of 16 is of great interest, a slight modification to the masking mechanism was necessary. With the modification, the masking mechanism subtracts 1 from c for every reset masked if r is unity, otherwise it operates as normal. This results in a more gradual reduction in the bursty load on the ring when masking begins at a RR of 16 - a mechanism which is now adaptive to the loading conditions.

The initial proposal was for a T_m of 32 to be used. It was considered that at such a loading level, the ring would quickly approach an overload condition, and some precautionary actions would be required; although no bandwidth allocation would actually be removed from the bursty services until the RR has fallen below 24. However, masking at such an early stage (at a RR of 24, there is still about 4 Mb/s of unused bandwidth) could result in unnecessary cell loss to the bursty services. Furthermore, with a T_s of 24 for VBR video, and a generally lower T_s for voice, the effect is that bandwidth would gradually be removed from the existing bursty connections in order to support new connections - especially voice calls. A new VBR video connection may not get the bandwidth it requires since a 'd' in this case does not necessarily represent 1 Mb/s; this could lead to a high CLR to the bursty connections. The ring is therefore an unfair system in this respect, unless the T_s s for all the services are raised well above the 24 resets/2ms mark. A T_m of 32 could also lead to inefficiency since the ring bandwidth may not be fully utilised. For instance, the observed RR never fell below 21 as a result of masking, leaving about 2 Mb/s of bandwidth unused.

It is therefore suggested that a lower T_m of 16 should be used. This implies that masking will only begin when overload has actually occurred; the bursty connections would be able to obtain all their allocated bandwidth up to this point. Beyond this, the masking mechanism would quickly damp down the surplus bandwidth allocated as explained earlier. Temporary overloads due to the delay in the response of the masking mechanism is not expected to disrupt the high priority CBO services. Admittedly, this threshold is a little low - a T_m of 20 may be a better choice, but the present implementation of the masking mechanism excludes this option. It can however be achieved with some modifications to the present implementation. This option would only be explored if the T_m of 16 proved unsuitable.

Simulation runs were carried out with VBR video traffic as well as with mixed traffic with a T_m of 16 and 32 respectively. Table 5.6 shows the results for the runs under normal load and overload conditions (in terms of call attempts). The results are in accord with the suggestions on the effects of the use of the two threshold values. With a T_m of 32, the call blocking performance of all the services was improved at the expense of the cell loss performance. There was no cell loss on the CBO services, while all the cell losses on the VBR video were due to the masking mechanism as no 'd' allocation request was denied since the RR never fell below 21. Cell delay was increased but the difference was negligible. With a T_m of 16, the cell loss performance was greatly improved but the call blocking performance

was degraded; no cell loss was observed on the CBO connections even under extreme loading conditions. The effectiveness of this overload control mechanism and the suitability of the T_m setting of 16 were thus demonstrated.

| Offered Load | | VBR video=85 calls/s | | Voice=200 calls/s; Data=5 Mb/s; VBR video=50 calls/s | | CBO video=10 calls/s; Voice=100 calls/s; Data=2 Mb/s; VBR video=85 calls/s | |
|-------------------------|-----------|----------------------|----------------------|--|----------------------|---|----------------------|
| T_m | | 16 | 32 | 16 | 32 | 16 | 32 |
| Call Blocking (%) | CBO video | - | - | - | - | 0.79 | 0.00 |
| | Voice | - | - | 0.41 | 0.00 | 4.35 | 0.00 |
| | VBR video | 1.19 | 0.46 | 0.62 | 0.42 | 3.36 | 1.68 |
| VBR video CLR | | 2.2×10^{-4} | 6.8×10^{-4} | 0.00 | 2.8×10^{-4} | 0.00 | 5.5×10^{-4} |
| No. of MR | | 64 | 867* | 25 | 644* | 2 | 725* |
| Min. RR | | 16 | 21 | 16 | 22 | 16 | 21 |

*only resets masked when the RR fell below 24 were included.

Table 5.6 Results for the variation of the masking threshold (T_m) with different loading conditions ($T_c=18$ and for T_s see section 5.6.5.2).

The necessity of the masking mechanism was also tested by removing it. Under normal loading conditions, no CBO services were affected and the cell loss performance of the VBR video was improved. However, under overload conditions, cell losses on the CBO services were observed, thus underlining the importance of this overload control mechanism in safeguarding CBO services against unforeseen ring overload.

This section on traffic control and threshold considerations may be summarised thus: A dynamic 'd' allocation scheme was outlined and the various threshold settings investigated. It was found that a call acceptance threshold in the range of 24-32 would be satisfactory from a call blocking performance perspective for the two-way VBR video connections with a mean bandwidth requirement of 0.7 Mb/s. But the cell loss performance was worst than expected. This could, however, be just the short-term CLR, which can be high since cell loss tends to occur in bursts. Some of the losses were attributable to the limitation of the 'd' allocation scheme and the size of the buffer used, as well as to the difficulty in estimating the ring load and thus in the correct setting of T_c . Cell loss can be minimised with a 'd' allocation ceiling threshold of 16-18 and a masking threshold of 16. Due to the delay in response in the 'd' allocation scheme, improvement to the cell loss performance can be achieved with a shorter counter interval, for instance, 1 ms; this may however lead to a very bursty ring load and is

thus not preferred. An alternative is to use a larger buffer, which helps to absorb the intra-frame data bursts as well as the asynchronism between the incoming load and the 'd' allocation on a node.

5.6.4 Buffer Size

As mentioned in the previous section, most of the cell losses were actually due to insufficient buffering in the nodes to accommodate the delayed response of the 'd' allocation scheme and the subsequent node overload. The initial choice of a buffer of 10 cells per node was based on the number of simultaneous connections expected on a node and the assumption of an instantaneous response from the 'd' allocation scheme. With an offered load of 23 E, the number of connections on a node could vary up to about 12; ideally, if the 'd' allocation matches the instantaneous cell arrival rate, a buffer of no more than 12 cells in size would be needed, but this is not the case. Furthermore, the ring dimensioning method assumed constant output within a field (i.e field buffering was assumed). A buffer of 10 cells is thus insufficient to meet the cell loss performance predicted.

Table 5.7 shows the results for buffer size variation with a given load, and figure 5.5 shows the graphs of CLR versus buffer size. The results were obtained from an aggregate of 4 runs, most of which were known to suffer cell loss. The call blocking performance is identical in almost all cases, indicating the same loading on the ring in terms of the number of calls and load build-up. Call blocking was slightly higher for the worst case run with a buffer larger than 55 cells; this is because the large number of cells which were originally lost were gaining access onto the ring, resulting in a higher load and therefore more calls being rejected.

With a T_c of 18, no cell loss was observed with a buffer of 75 cells or larger; with a T_c of 16, all cells gained access onto the ring with a buffer of 55 cells. Thus a buffer of about 100 cells would be more than adequate to achieve a CLR lower than 10^{-6} . Even a buffer of this size is considered small in practice. A much larger buffer can be implemented without heavy cost since memory chips are relatively cheap nowadays. A buffer of 100 cells only corresponds roughly to 30% of a field worth of video data (a complete field generates 340 cells) and it is shared by up to 12 connections, whereas a much larger buffer will be required to smooth a VBR video into a constant bitstream output if at all possible. It has been suggested that a buffer of 1000 cells in size is practical. The cell loss performance with a buffer of this size will be negligible such that the call blocking performance becomes the only service criterion which is of concern.

There was an increase in the number of resets masked and 'd' allocation requests denied since more cells could now gain access onto the ring. The mean cell delay was virtually unaffected while the maximum cell delay increased from about 400 μ s to 1.50 ms.

| Buffer Size(cells) | Call Blocking(%) | CLR | No. of MR | d Alloc. Denied | Mean Cell Delay/us | Max. Cell Delay/us |
|--------------------|------------------|----------------------|-----------|-----------------|--------------------|--------------------|
| 10 | 1.10 | 1.8×10^{-4} | 74 | 295 | 11 | 480 |
| 20 | 1.10 | 7.8×10^{-5} | 103 | 330 | 12 | 630 |
| 30 | 1.10 | 3.7×10^{-5} | 97 | 342 | 12 | 908 |
| 55 | 1.22 | 9.5×10^{-6} | 74 | 331 | 12 | 1420 |
| 75 | 1.22 | 0.0 | 74 | 330 | 12 | 1570 |

Table 5.7(a) Results for buffer size variation with VBR video call rate=85 calls/s, $T_s=24$, $T_c=18$ and $T_m=16$.

| Buffer Size(cells) | Call Blocking(%) | CLR | No. of MR | d Alloc. Denied | Mean Cell Delay/us | Max. Cell Delay/us |
|--------------------|------------------|----------------------|-----------|-----------------|--------------------|--------------------|
| 10 | 1.10 | 1.6×10^{-4} | 101 | 21 | 11 | 418 |
| 20 | 1.10 | 6.1×10^{-5} | 136 | 16 | 12 | 580 |
| 30 | 1.10 | 2.2×10^{-5} | 134 | 18 | 12 | 740 |
| 55 | 1.22 | 0.0 | 106 | 7 | 12 | 1420 |

Table 5.7(b) Results for buffer size variation for VBR video with call rate=85 calls/s, $T_s=24$, and $T_c=T_m=16$.

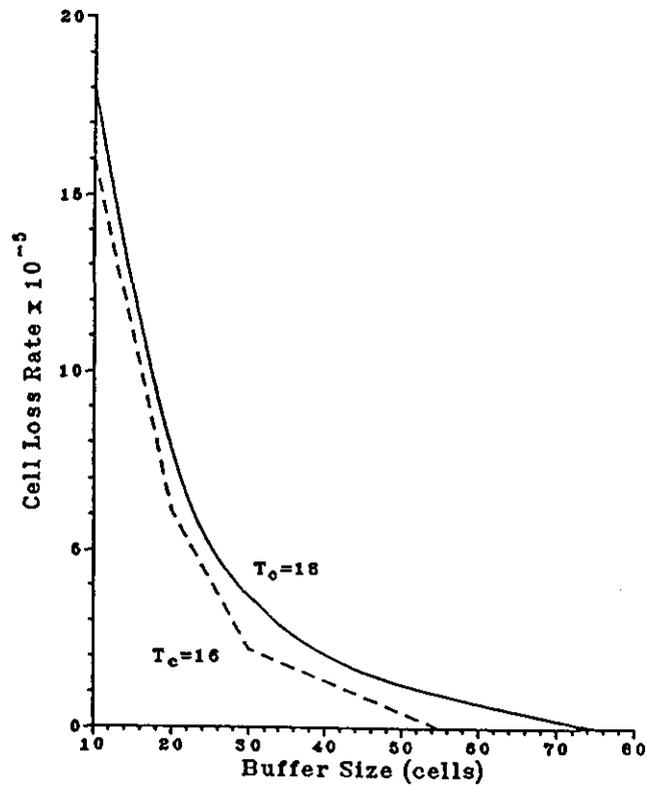


Fig. 5.5 Cell Loss performance with varying buffer size.

This delay is still small when compared to the 20 ms video field time, and is far from the 120 ms video lag mentioned in Chapter 2. This small delay is a consequence of the way in which the 'd' allocation scheme operates: It provides surplus bandwidth to the VBR video connections most of the time. Another contributing factor is the spare capacity provided on the ring to cater for the occasional bursts in call attempts or in the ring load.

A further point which requires consideration is that when a very large buffer, for example 1000 cells, is used, one must ensure that cells are not held up in the buffer for a prolonged period. It may therefore be necessary to make a slight modification to the 'd' allocation scheme such that the 'd' allocation on a node will only be decremented if the buffer is filled below a given level. This would ensure that the buffer is cleared sufficiently quickly to avoid any prolonged cell delay. With a larger buffer, a longer counter interval can be used in the 'd' allocation scheme, which could then lead to a smoother load on the ring. Preliminary investigations with a buffer size of 100 cells and a counter interval of 4 ms revealed that a small amount of smoothing was indeed possible, but this requires further study.

5.6.5 Service Performance on the Ring

5.6.5.1 VBR Video Traffic

I. With Varying Offered Load

The first aspect of interest in this section is the performance of the system under varying VBR video load conditions. The varying load can also be interpreted as the varying amount of capacity provided on top of, or short of, the mean bandwidth requirement of each VBR video connection. The T_s was set at 24 - the minimum recommended value, and a T_m of 16 was used. The study was conducted with the two different T_s s of 16 and 18, and with buffers of 10 cells and 100 cells. Only 3 runs each were made because of the large number of runs required for the different combinations.

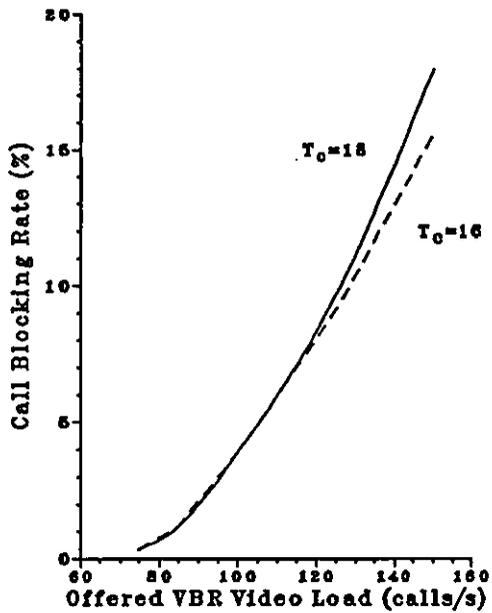
The results are presented in Tables 5.8 and 5.9, and in figure 5.6. From figure 5.6, it can be seen that with a buffer of 10 cells, which has been established to be inadequate, the CLR increases exponentially; with a T_s of 16, cell loss performance was significantly better especially in heavy loading conditions. The results indicate that load control by call blocking is only partially effective from a cell loss perspective - primarily because of the low T_s used, which inevitably led to excess calls being admitted by error when the call attempts were high. It must be stressed that with such a small buffer, some of the losses were due to factors other than actual ring overload.

| Call Rate (calls/s) | Extra Capacity(%) | Call Blocking(%) | CLR | No. of MR | d Alloc. Denied | Min. RR |
|---------------------|-------------------|------------------|----------------------|-----------|-----------------|---------|
| 75 | 43 | 0.36 | 0.0 | 2 | 0 | 18 |
| 85 | 30 | 1.19 | 2.2×10^{-4} | 69 | 225 | 16 |
| 90 | 22 | 1.98 | 6.7×10^{-5} | 67 | 267 | 16 |
| 100 | 13 | 3.90 | 1.8×10^{-4} | 71 | 650 | 15 |
| 125 | 4 (OL) | 9.68 | 5.5×10^{-4} | 383 | 2730 | 15 |
| 150 | 20 (OL) | 17.94 | 1.6×10^{-3} | 849 | 6139 | 15 |

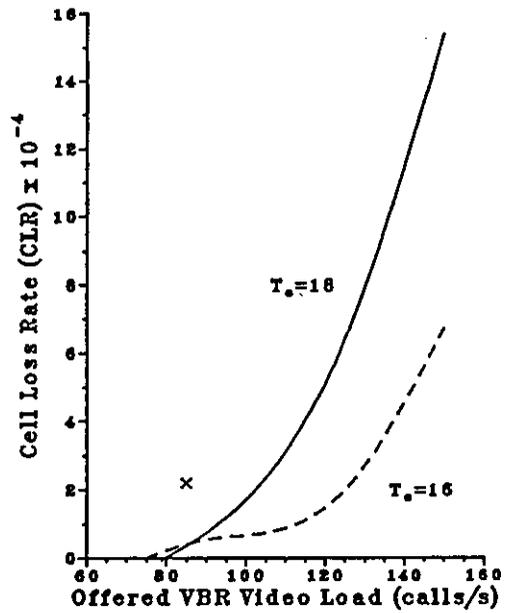
Table 5.8 Service performance and ring behaviour with varying VBR video load and $T_c=18$.

| Call Rate (calls/s) | Extra Capacity(%) | Call Blocking(%) | CLR | No. of MR | d Alloc. Denied | Min. RR |
|---------------------|-------------------|------------------|----------------------|-----------|-----------------|---------|
| 75 | 43 | 0.36 | 0.0 | 2 | 0 | 18 |
| 85 | 30 | 1.19 | 2.1×10^{-4} | 85 | 21 | 15 |
| 90 | 22 | 2.13 | 5.4×10^{-5} | 89 | 14 | 15 |
| 100 | 13 | 3.90 | 6.6×10^{-5} | 114 | 14 | 15 |
| 125 | 4 (OL) | 9.13 | 2.0×10^{-4} | 455 | 48 | 15 |
| 150 | 20 (OL) | 15.54 | 6.7×10^{-4} | 1210 | 333 | 15 |

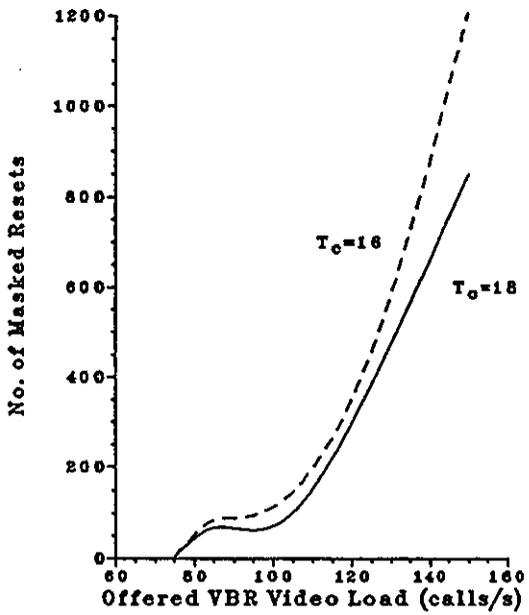
Table 5.9 Service performance and ring behaviour with varying VBR video load and $T_c=16$.



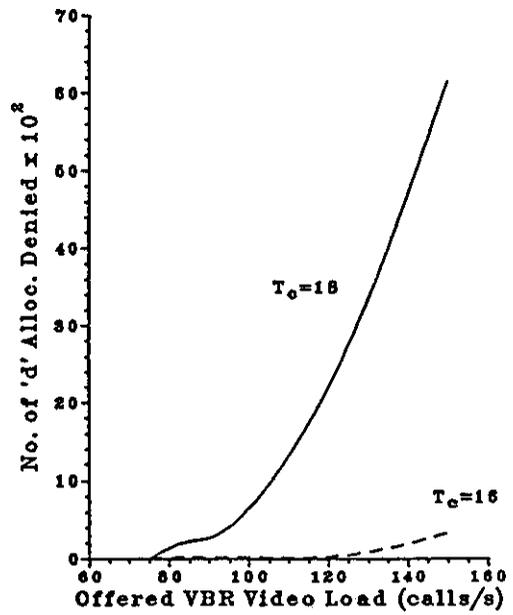
(a)



(b)



(c)



(d)

Fig(s). 5.6 Service performance and ring behaviour with varying VBR video load and different ceiling threshold (T_c) values.

When a buffer of 100 cells was used, no cell loss was observed even under extreme overload conditions, hence the results are not shown. This suggests that an overload is generally of a short duration and small magnitude, and can be easily contained with a buffer of a reasonable size. Furthermore, with a T_c of 16, the predominant load control was the masking mechanism, which tends to spread the burden of a ring overload over all the nodes so that each node only needs to bear a fraction of the excess load; this effect is evident when comparing Tables 5.8 with 5.9. The ring is thus a fully distributed system with the excess load being shared over all the nodes. This effectively provides, with 10 nodes, a total buffer capacity of 1000 cells! This should be more than sufficient to cope with any overload, which should have been, to some extent, limited by the call blocking mechanism.

Extreme load such as those investigated should not happen in practice, unless there has been a partial system failure or poor traffic forecast. The results again suggest that the cell loss performance is less of a concern than the call blocking performance. In order to meet the required call blocking performance of 1%, a reasonable offered load would be in the range of 80-90 call/s (23-24 E), which corresponds with the earlier estimate. And provided a reasonably large buffer is used, the CLR should be well below 10^{-6} . Even if extreme loads do occur, the CLR would still be very small as was the case with a buffer of 100 cells.

II. With Normal Offered Load

In all the previous simulations, the number of runs had been limited to 3-5 in each case in order to minimise the computing resources required; in each case, the observed worst cases were included in order to study the cell loss performance with a minimum number of runs. Although this may be desirable in some studies, it also resulted in a distorted view of the average service performance, especially with such short simulation time used per run.

It was therefore necessary to study the service performance with a larger number of sample runs. Twenty runs were made with an offered load of 23 E; T_a and T_m were set to 24 and 16 respectively. The results are presented in Table 5.10 for the T_c settings of 16 and 18, and with buffers of 10 and 100 cells. Figure 5.7 shows the negative cumulative distribution of the number of runs with different cell loss performance; only one result is shown because the results for the two T_c settings with a buffer of 10 cells are almost identical while no cell loss was observed with the larger buffer.

Most of the runs had zero or very small call blocking rates and CLR's, and load control of any form was rarely exercised. Figure 5.7 also shows that the probability of encountering a run with a high CLR is very small, indicating that the high CLR's observed in the earlier runs were indeed the short-term performance, where cell loss occurred in bursts during the rare video surges. The long-term results indicate a call blocking rate of about 0.4%, which is just under half the specified rate of 1%. The CLR is 4.39×10^{-5} and 4.67×10^{-5} for the T_c 's of 16 and 18 respectively with a buffer of 10 cells. With a larger buffer of 100 cells, these cell losses

| Tc | Buffer Size(cells) | Call Blocking(%) | CLR | No. of MR | d Alloc. Denied |
|----|--------------------|------------------|----------------------|-----------|-----------------|
| 18 | 10 | 0.38 | 3.5×10^{-5} | 98 | 333 |
| 16 | 10 | 0.38 | 3.3×10^{-5} | 122 | 21 |
| 16 | 100 | 0.40 | 0.0 | 98 | 369 |

Table 5.10 Service performance of VBR video with a call rate of 85 calls/s.

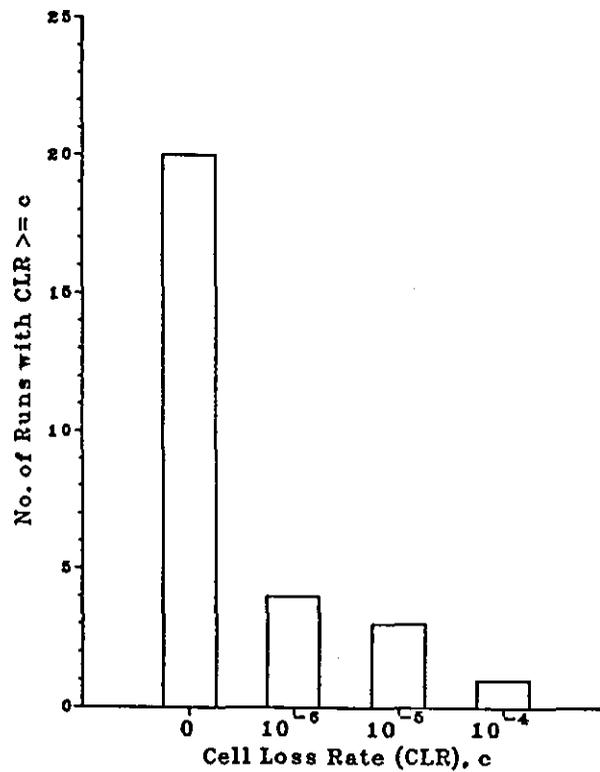


Fig. 5.7 Distribution of the simulation runs with a cell loss rate (CLR) $\geq c$.

can be completely avoided and the CLR could be well below 10^{-7} , which is near the cell loss performance expected (10^{-6}) or better. Unfortunately to verify the results with the prediction would require in excess of 100 runs!

It is noted that the call blocking performance is very much better than that specified. This suggests that the T_s of 24 is indeed too low; a higher T_s of 28-32 may be more appropriate and could help to prevent ring overload by reducing the risk of admitting excess calls onto the ring. A threshold of 32 would bring it into the range of the threshold values suggested in section 5.6.3, i.e. around 40. Alternatively, with a large buffer, the offered load may be increased to 90 calls/s while maintaining the T_s at 24. The results for these two alternatives are tabulated in Table 5.11. As expected, the CLR was reduced with a T_s of 32, while the call blocking performance was still within specification. By increasing the offered load to 90 calls/s and by providing a buffer of 100 cells, no cell loss was observed and the call blocking rate was also acceptable. Any further increase in the offered load would quickly degrade the call blocking performance, although the buffer may be able to sustain the cell loss performance for some further increase in the offered load (see Tables 5.8 and 5.9).

| Offered Load (calls/s) | T_s | Buffer Size (cells) | Call Blocking (%) | CLR | No. of MR | d Alloc. Denied | Min. RR |
|------------------------|-------|---------------------|-------------------|----------------------|-----------|-----------------|---------|
| 85 | 32 | 10 | 1.08 | 1.3×10^{-5} | 33 | 0 | 16 |
| 90 | 24 | 100 | 0.86 | 0.0 | 127 | 24 | 15 |

Table 5.11 Service performance of VBR video with i) a higher T_s and ii) a higher offered load, $T_c = T_m = 16$ in both cases.

5.6.5.2 Mixed Traffic

So far, the performance study has been confined to mainly one type of traffic, namely, VBR video, without considering the co-existence of other traffic. This simplifies the problem as the interaction between the different traffic can be ignored. However, the Orwell protocol was specifically designed to carry traffic from a mix of services, with emphasis on its capability to guarantee cell security and bounded cell delay for the CBO services even in overload conditions. Hence it is most important that the performance study be conducted in a multi-service environment. Thus, the 1 Mb/s CBO video, the 64 kb/s CBO voice and the data services were included in this part of the study.

The 'd' allocation for the CBO video and voice services were deterministic such that a 'd' of 1 was allocated to every video connection, while each voice call was allocated 1/16 of a 'd'. Data service was given a background 'd' allocation of 1 per node, and because the data service was not considered to be connection-oriented, there was no call control associated

with this service. Since a two-way VBR video connection requires 2 Mb/s of bandwidth, it would cause the RR to decrease by about 4 resets/2ms when accepted onto the ring near the full load condition. If at least 2 Mb/s were to be allowed for VBR video and data bursts, this would require that the T_s for the CBO video to be set at 24 or higher values. Using the same argument, the voice service would then be given a threshold of 21. The T_s s for all three services should not be set too far apart otherwise those with a lower T_s would effectively be given priority in call acceptance, thus achieve a better call blocking performance at the expense of other services. In this work, the thresholds were fixed at 24 for both the video services and 21 for the voice service; the thresholds could be raised en-bloc if more reserved bandwidth is required to cater for bursts in the ring load. A slightly higher T_s for the CBO video may also be appropriate since it occupies a relatively larger amount of bandwidth per connection.

A simple approach was adopted in estimating the various traffic for the ring: the ring capacity was partitioned into blocks of bandwidth which were dimensioned for the different classes of service. Dimensioning of the bandwidth for the VBR video was as before; for the CBO services, the Erlang Capacity Table was used, while the mean bandwidth requirement was used for the data service - although this is not strictly correct as some extra capacity should be provided to cater for data bursts. All the services were dimensioned for a call blocking performance of 1%. No account was provided for the fact that there could be some statistical gain in multiplexing the various services in a single network, especially when there is a large number of different services at low traffic levels³³¹. The simple approach used ignores this gain and the total extra capacity provided for each class of service to meet its service requirement could be substantial; this could be shared by all the services and may lead to a better overall performance than that specified.

I. Normal Loading

The results for a range of traffic mix with correct bandwidth dimensioning are presented in Table 5.12. The buffer used was 10 cells since a larger buffer was not considered necessary in this case. The performance of all the services were good with very little load control or overload control being necessary, and virtually no cell loss was observed on any services. The results appear to indicate that some bandwidth sharing among the various services had taken place, especially when CBO video was included; bearing in mind that no provision was given to the bursts in the data service when the ring capacity was dimensioned.

Cell delay for all the services, except voice, was very small: The mean cell delay was well under 20 μ s and the observed maximum was about 200 μ s. The large variation in the cell delay for voice was caused by the load smoothing mechanism implemented on the voice queue in the original ring model. It suffices to say here that this mechanism introduces large delay when the voice traffic level is low; this can be deduced from Table 5.12. The results on

| Offered Load | | | | Call Blocking | | | VBR Video CLR | Min. RR | Mean Cell Delay/us | | | |
|---------------------------|--------------------|----------------|---------------------------|---------------|-------|--------------|----------------------|------------|--------------------|-------|------|-----------|
| CBO Video (calls/s) | Voice (calls/s) | Data (Mb/s) | VBR Video (calls/s) | CBO Video | Voice | VBR Video | | | CBO Video | Voice | Data | VBR Video |
| 10 | 100 | 3 | 20 | 0.00 | 0.00 | 0.00 | 0.0 | 34 | 7 | 119 | 7 | 7 |
| - | 200 | 5 | 50 | - | 0.32 | 0.49 | 0.0 | 16 | - | 124 | 12 | 10 |
| - | 460 | 1 | 50 | - | 0.00 | 0.49 | 0.0 | 16 | - | 52 | 8 | 9 |
| - | 460 | - | 55 | - | 0.21 | 0.61 | 0.0 | 16 | - | 55 | - | 10 |
| - | 186 | - | 70 | - | 0.27 | 0.71 | 1.2×10^{-6} | 17 | - | 114 | - | 10 |
| - | 1080 | 2 | 20 | - | 0.01 | 0.00 | 0.0 | 30 | - | 29 | 11 | 11 |

Table 5.12 Service performances in a multi-service environment with normal loading.

| Offered Load | | | | Call Blocking | | | VBR Video CLR | Min. RR | Mean Cell Delay/us | | | |
|---------------------------|--------------------|----------------|---------------------------|---------------|-------|--------------|----------------------|------------|--------------------|-------|------|-----------|
| CBO Video (calls/s) | Voice (calls/s) | Data (Mb/s) | VBR Video (calls/s) | CBO Video | Voice | VBR Video | | | CBO Video | Voice | Data | VBR Video |
| 10 | 100 | 2 | 100 | 8.70 | 4.72 | 14.23 | 8.2×10^{-3} | 15 | 13 | 266 | 15 | 17 |
| 20 | 150 | 5 | 25 | 2.18 | 3.70 | 10.00 | 0.0 | 16 | 13 | 178 | 16 | 12 |
| - | 150 | 2 | 150 | - | 16.93 | 25.35 | 1.0×10^{-3} | 14 | - | 281 | 17 | 22 |
| - | 500 | 2 | 125 | - | 14.33 | 25.23 | 2.4×10^{-4} | 15 | - | 136 | 17 | 19 |
| - | 1000 | 2 | 100 | - | 21.06 | 31.37 | 1.3×10^{-3} | 14 | - | 85 | 19 | 26 |
| - | 2000 | - | 100 | - | 29.67 | 45.93 | 2.0×10^{-3} | 14 | - | 80 | - | 33 |

Table 5.13 Service performances in a multi-service environment under overload conditions.

the cell delay for voice at a low voice traffic level are therefore misleading. It would be most likely that there will be a large voice traffic on a practical system, as a result, the cell delay with a high voice traffic level is probably more representative. Furthermore, with the presence of VBR video on the ring, the load smoothing mechanism is unnecessary. It can be expected that with this mechanism removed, the cell delay for voice should be in the same order as the other services.

II. Overload Conditions

One of the most important features stipulated for the Orwell Ring is its capability to guarantee cell security for the CBO services under all loading conditions. The results above demonstrated this capability under normal loading conditions. However, the effectiveness of the mechanisms which are used to provide this capability cannot be fully tested under normal loading conditions. They must also be able to protect the CBO services even when the ring is heavily overloaded.

The ring was put to test by subjecting it to a wide range of traffic mix under extreme overload conditions. All the threshold and parameter settings were as before and the node buffer remained at 10 cells. The results are tabulated in Table 5.13. No cell loss was observed on any other service besides the VBR video. The CLR's for the VBR video were in the same order as those shown in Tables 5.8 and 5.9 for overload conditions with VBR video traffic only; it has also been shown that these losses can easily be prevented by the use of a larger buffer. Thus even under overload conditions, the cell loss performance of the system can still be maintained with minimal cell loss to the VBR video and no cell loss to the CBO services. The results on the number of masked resets and 'd' allocation requests denied are not presented since the figures vary greatly and do not convey much information. It suffices to say that the two control mechanisms were active for most of the time and were effective in maintaining the RR such that it rarely fell below 16. Although the RR did fall to 14 in some instances, these occurrences were however rare, and were thus inconsequential. Furthermore, the CBO services were also to some extent protected by the priority scheme. Hence no cell loss was observed on these services.

The cell delay for all traffic types was affected, but only by a small amount. With the cell delay for the voice service, a similar problem is encountered as described previously with regard to the load smoothing mechanism. With the CBO video, the mean cell delay was about 10 μ s, with more than 95% of the cells experiencing a delay less than 30 μ s. The data service experienced mean cell delays of no more than 20 μ s with more than 95% of the cells getting through under 80 μ s; these results would, nevertheless, depend on the amount of data load offered since the 'd' allocation for this service is fixed. With VBR video, the mean delay remained very small at less than 35 μ s with 95% of the cells delayed by less than 80 μ s. These results, when compared with those obtained with the ring under normal loading conditions,

show little difference. This suggests that the call blocking mechanism was sufficiently effective in not allowing too many excess connections onto the ring, and this is illustrated by the large number of call attempts of all types being rejected. Attention must be drawn to the fact that the T_s s for all the services were set at relatively low values, and it has been shown earlier that higher thresholds may be more suitable. With higher thresholds, the performance of the services will be better with more excess calls being rejected and more bandwidth being reserved to absorb the bursts arising from both the VBR video and the data traffic.

It has thus been demonstrated that the load control and overload control mechanisms provided by the Orwell protocol and the 'd' allocation scheme are adequate, not only in protecting the CBO services against cell loss under extreme overload conditions, but they can also protect the VBR video service against excessive cell loss provided a reasonably large buffer is available on each node.

5.6.6 Reset Rate Observation Interval

One of the ring parameters which has hitherto not been investigated is the observation interval over which the RR is measured. In all the previous studies, this was fixed at 2 ms which is the current recommended setting. This may be an acceptable interval when VBR video is excluded and the ring load is not so bursty, but in the presence of VBR video, a longer interval may be more appropriate. The reasoning behind this is that a longer interval could provide a better measure of the ring load by averaging the number of resets over a longer period, especially when the load is very bursty. However, a long interval may fail to recognise the latest ring loading condition. There appears to be no satisfactory solution to this dilemma, unless a more sophisticated load monitor is used. A compromised observation interval can be sought which could give a reasonably accurate load measurement as a sub-optimum solution.

The task of searching for an optimum interval will inevitably involve a very large number of test runs, and is beyond the capability of the computing resources available. Nevertheless, a number of runs were made with an observation interval of 4 ms to investigate the stipulation that a longer interval could provide a better measure of the ring load. The results are shown in Table 5.14. There appears to be a slight improvement in the cell loss performance with more excess calls being rejected, which would otherwise cause more overloading. The results, although not conclusive, seem to be in accord with the stipulation, that is, a better load estimate hence a better load control.

| Observation Interval | Call Blocking (%) | CLR | No.of MR | d Alloc. Denied | Min. RR |
|----------------------|-------------------|-----------------------|----------|-----------------|---------|
| 2 ms | 0.89 | 1.25×10^{-4} | 77 | 298 | 16 |
| 4 ms | 1.18 | 1.42×10^{-5} | 97 | 200 | 16.5 |

Table 5.14 Results for the variation of the reset rate observation interval with VBR video=85 calls/s, $T_r=24$, $T_c=16$, $T_m=16$ and buffer=10 cells.

5.6.7 Cell Delay and Cell Loss

The cell delay and cell loss performance for the various services on the ring were investigated under different loading conditions and the results have been presented under the corresponding sections. This section summarises the results.

The cell delay for the various services, except voice, was shown to be of a very small magnitude, and this was true even under heavy overload conditions (in terms of call attempts). The overall cell delay had a mean less than 30 μ s, and in many cases the mean delays were well below this; 95% of the cells gained access onto the ring within 80 μ s. This cell delay performance should meet any delay requirements that may arise for these services. On the other hand, the study on the cell delay for the voice service was complicated by the load smoothing mechanism. Although primarily designed to smooth the voice load under light voice traffic, it has the effect of introducing large cell delay, especially when the ring is heavily loaded with other types of traffic, although such traffic mix is not considered probable. As a result, the cell delay observed on the voice service showed a large variation, and could be unacceptably large under the loading situation just described: with a mean of 300 μ s and a maximum delay bordering the maximum allowable delay of 2 ms. This mechanism could lead to the possibility of cell loss on this service. However, such a smoothing mechanism is redundant and will not be implemented on a practical system; the cell delay for voice is thus expected to be in the same order as those for the other services under the same loading conditions.

Cell loss on the ring was limited to the VBR video service when a small buffer was used. No cell loss on the CBO services was observed although there was a small risk of losing voice cells, but this was entirely due to the load smoothing mechanism rather than the deficiency of the ring protocol. For the data service, data load was offered at less than the maximum bandwidth allocated, hence no cell loss was expected and none was observed. But should the load offered be equal or more than the allocated bandwidth, cell loss can be expected.

The cell loss performance for the VBR video was expected to be in the order of 10^{-8} if sufficient buffering is provided; but with the limited number of results available, it can only be deduced that the cell loss rate is at least below 10^{-7} with a buffer of 100 cells. The cell loss performance was further enhanced in a multi-service environment where the spare capacity allocated for each of the services provide further scope for resource sharing.

With a buffer of 10 cells, the CLR for the VBR video was quite high (in the order of 10^{-5}), and cell loss occurred in bursts as can be deduced from the results in Table 5.10 where only a few runs suffered large losses. Further analysis showed that, for the worst case, all the cell losses occurred within a period of 400 ms, and this corresponds to a CLR of 10^{-3} as compared to 10^{-4} over the entire run. Under normal loading conditions, the losses tend to concentrate mostly on a single node rather than being spread evenly over all the nodes. With this taken into account, a node could actually suffer a CLR of 10^{-2} over the short period. Cell losses, however, did not occur contiguously but intermittently. Furthermore, because of cell multiplexing in the node buffer, the maximum burst of cell loss contiguously for a connection was usually very small, and the longest burst observed under overload conditions was 13 cells (3.8% of a field).

It has also been demonstrated that with a small buffer, cell loss was bounded to about 10^{-3} under overload conditions. This can be attributed to the the call blocking mechanism and to some extent, buffer sharing induced by the masking mechanism. Cell loss of this order can easily be prevented with the use of a reasonably large buffer, or it can be reduced with the use of higher call acceptance thresholds, since low thresholds were used in this study.

5.6.8 Auto-Resets

Auto-reset is designed as the ultimate overload control when all other means fail. It is only meant for very short-term overload control while the other control mechanisms are being brought into effect. It is also designed as a safeguard against corrupted reset slots. Auto-reset is simply a timer which re-initialises a node's CBO services bandwidth allocation when the timer expires before a reset is received. This expiry time is currently 125 μ s although it requires further investigation. Auto-reset was not used in this work in order to allow for the determination of an optimum expiry time for this mechanism.

It is clear from the results obtained that the protocol is capable of maintaining the service requirements it is designed to meet under all loading conditions, even on occasions when the RR fell below the 16 resets/2ms minimum. But this by no means suggests that auto-reset is redundant; it is still required to guard against unforeseen overload, and most of all, against corrupted reset slots. The question is what should the expiry time be in order not to cause undesirable auto-resets.

Although the average RR was maintained at or above the 16 resets/2ms target most of the time, detailed analysis of the individual RIs revealed that a small but significant number of RIs exceeded the 125 μ s maximum specified, even under normal loading conditions. The maximum RI observed was 200 μ s in overload conditions. This is unavoidable due to the statistical nature of the load and the fact that the protocol works with the average RI and not the absolute value. In general, when the RIs were averaged over 2 ms, the average RI did not usually exceed 125 μ s even though some of the individual RIs were well over the limit.

Figure 5.8 shows the distribution of the RIs for 3 different loading conditions with no cell loss. It is apparent from these results that if the expiry time for the activation of the auto-reset is 125 μ s, a large number of unnecessary auto-resets would be triggered when the ring is still capable of sustaining the service requirements. This may upset the normal operation of the ring. A better expiry time, taking advantage of the insensitivity of the ring to temporary ring overload, is perhaps 150 μ s. Although this may lead to problems in the event of a prolonged ring overload when the masking mechanism fails to damp down the excess load, the situation is considered highly unlikely.

With an expiry time of 150 μ s, most of the unnecessary auto-resets can be avoided, although under overload conditions, there may still be a significant number of auto-resets. These are, however, justified and could help to improve the cell security of the CBO services under these conditions. The number of auto-resets can be further reduced with the use of higher call acceptance thresholds.

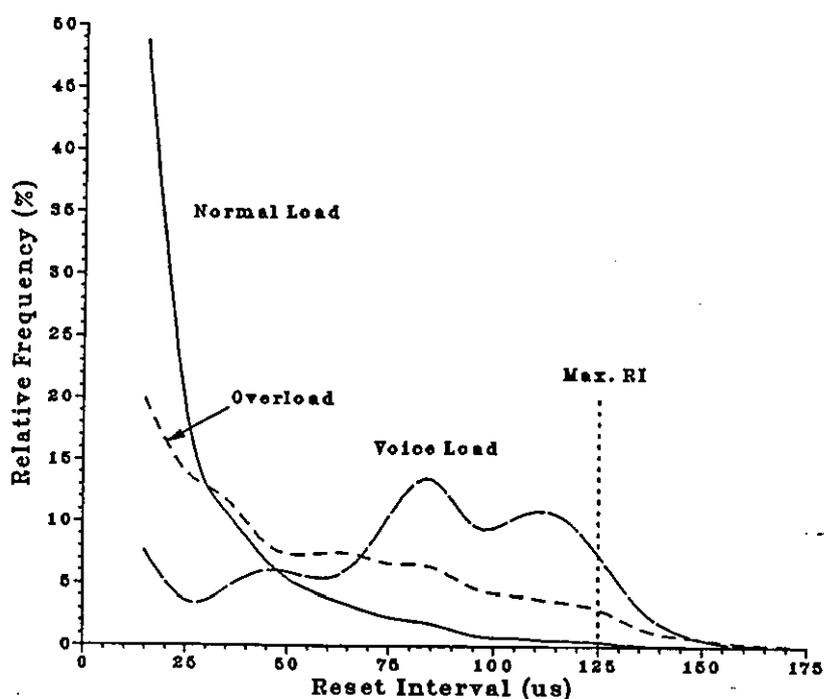


Fig. 5.8 Distribution of the Reset Interval (RI) for different loading conditions.

5.7 Discussion and Conclusions

In this chapter, the results from the simulation study on the transmission of VBR video over an Orwell Ring were presented and discussed. The results indicated the feasibility of using VBR video as a means of providing video services on such a network. With the correct dimensioning of the network, and the proper exercise of the load control and overload control mechanisms, a good QOS can be provided for all the services without having to provide excessive spare capacity for the VBR video traffic. This is, however, only true with a high VBR video traffic level. For instance, with 23 E, only about 20%-30% extra capacity was required for a call blocking rate of 1% and a cell loss rate in the order of 10^{-6} ; and less if a lower QOS is acceptable. This suggests that the bandwidth gain over CBO video quoted in Chapter 3 of about 2-3 can be more or less maintained. Furthermore, in a multi-service environment, particularly when there are wide bandwidth services present at low traffic levels, the accumulated spare bandwidth can be used to reduce the extra capacity required for the VBR video services.

The feasibility of VBR video services is further enhanced by the fact that most networks are planned for the future and may very often be replaced well before the end of their planned life span, thus leaving some spare capacity which bursty services can make use of. Overall, only a very small amount of extra capacity will actually be required for the VBR video services for a given QOS, thus greatly improving the viability of VBR video.

The Orwell protocol was shown to possess adequate traffic control capabilities to meet a variety of loading conditions without any loss of cells on the CBO services. Although the results obtained were those for a 34 Mb/s ring, they could equally be applied to a larger ring or a Torus. However, the call acceptance thresholds will need to be adjusted according to the expected traffic load and the characteristics of the system concerned. The bandwidth gain of the larger system may be higher and the reset rate may fall more slowly in near full load conditions, otherwise the behaviour of the larger systems are similar to the 34 Mb/s ring; the deductions from this work are therefore equally applicable. The results can also be extended to the inter-networking of a number of rings or Tori. If no long distance link is involved, the performance of the system should be similar since the bridges would act just as any ring node, with perhaps higher throughputs. Since the Orwell protocol operates on the bridges, each bridge can simply be regarded as a concentrator for traffic from one ring to an adjacent ring. The concept is thus not too dissimilar from the single ring system with nodes concentrating traffic onto the ring.

Where a link is necessary, an Orwell link multiplexor - which operates the Orwell protocol on a link - can be used so that uniformity can be maintained throughout the system. In this case, the characteristic of the link multiplexor must be determined, and since it would be different from the characteristic of the ring, some of the control parameters may need to be set differently to allow for a greater safety margin; hence, lower link efficiency. But this may

not be important with the immense capacity available on fibre optic links. A different control strategy can be formulated but all the control mechanisms of the protocol are available, and the proper exercise of these controls could result in a similar system performance as the 34 Mb/s ring. The results thus have a wide implication.

One of the most important issues concerning the use of VBR video has been the concealment of cell loss related errors due to the highly compressed nature of the data. Two methods are available: background update - where errors are corrected by a high priority background refresh data stream; and layered coding - where video information is coded into high priority basic information which is protected against cell loss, and low priority enhancement information. The first method is simpler but slower, and can only tolerate a small amount of cell loss, while the latter is more complex but very efficient even where the CLR is high.

Since a cell loss rate of less than 10^{-7} can be achieved efficiently on the ring, and the simultaneous overload over a network of rings is considered highly unlikely, the background update is probably sufficient for error concealment purposes. LC is not a necessity but it does offer better error concealment capability as well as other added advantages. With LC, no noticeable degradation would be observed even with a CLR much higher than 10^{-7} on the low priority data stream, and it is more effective against burst error. This suggests that even lesser extra capacity would need to be reserved for the video bursts, thus increasing the gain of VBR video.

Although both methods are feasible, LC would have a better performance in terms of better picture quality and higher bandwidth efficiency. It can also be used to overcome problems associated with scene changes, simply by not sending the low priority video data, thus reducing the peak bit rate. These are of course achieved at the expense of more hardware complexity in the codec and some control complexity since both data streams are VBR in nature. The small cell delay on the ring however relieves the problem of synchronising the two data streams. If necessary, the low priority data can be given a smaller buffer to prevent excessive delay.

Scene changes were not considered in this simulation study because it was suggested that full resolution at scene changes may be unnecessary. Even if full resolution is required, it is possible to absorb these scene changes with a sufficiently large buffer because they occur very infrequently with videophone type pictures and normally last for only a single frame. Nevertheless, video bursts with a maximum bit rate of up to 70% of the peak bit rate were generated. This corresponds to some very violent motion and is of more importance in this work.

To conclude, the simulation results indicate that with the correct dimensioning of the ring, the proper control of the VBR video traffic using the control mechanisms provided by the protocol and the 'd' allocation scheme, along with a suitable error concealment technique to overcome the rare but inevitable occurrence of cell loss, VBR video can be supported on the ring in a multi-service environment, while meeting the required QOS for all the services without sacrificing the gain of VBR video. VBR video thus appears to be a promising means to provide future video-communication services. However, these results will need to be validated and verified with a practical system which is presently unavailable.

note :- (from page 72)

Although these call holding times are unrealistically short, they are considered adequate for this simulation study and are extensively used at BTRL, in order to reduce the amount of computing resources required. The objective is to achieve a state of equilibrium in the shortest possible time so that the QOS, in terms of call blocking and cell loss, can be studied. On this premise, the short holding times are justified for the CBO connections. However, they are only barely adequate to bring out the VBR nature of the sources. As such, only single-field frames were used to increase the data rate variation, and under a state of equilibrium, the short holding time used should be justified. Longer call holding times will obviously be more appropriate if computing resources permit.

Chapter 6 :

Experimental Work on a Prototype Orwell Ring

6.1 Introduction

This chapter briefly describes the work carried out in this research project that deals with the practical aspects of the transmission of VBR video over an experimental Orwell Ring. This work was not intended as a means to verify the results obtained in Chapter 5 because neither a ring of that speed nor VBR video codecs similar to those simulated are available. Instead, the objective of this exercise is to demonstrate the inter-working of an experimental 15 Mb/s Orwell Ring and a simple VBR video source in the form of a modified COST211 codec. From this, insight into the performance of the protocol and of the system as a whole can be acquired, and experience in the interfacing of the two systems can be gained. One important issue which will be dealt with is the design of an addressing scheme suitable for transporting VBR video data in an ATM environment. For details of the hardware implementation see [26].

6.2 The Experimental Orwell Ring

The experimental Orwell Ring was developed at BTRL as a demonstrator for demonstrating the working of the Orwell protocol. The ring operates at 15 Mb/s with a single slot and a single node. The cell structure used conforms to the latest specifications for the protocol, i.e. a 21 byte cell with a 5 byte header.

On this experimental ring, not all the functions specified in the Orwell protocol have been implemented on the node. The functions not implemented are call control, masking and auto-reset. These are not necessary because the number of connections on the single node has been restricted to a one-way 2 Mb/s CBO video, a two-way 64 kb/s voice, a one-way VBR

video (to be implemented) and a data connections. The node 'd' allocation has been fixed such that the reset interval never exceeds 125 μ s. The VBR video and data connections can utilise all the bandwidth not required by the two CBO services.

A block diagram of the node layout is shown in figure 6.1. The use of a microprocessor as the controller provides the flexibility for those functions not implemented to be added as and when required. As illustrated, four service queues were provided for CBO video, CBO voice, VBR video and data, in the order of priority as enumerated. This is slightly different from the simulation exercise where the priority order of the two bursty services were reversed according to the latest proposal.

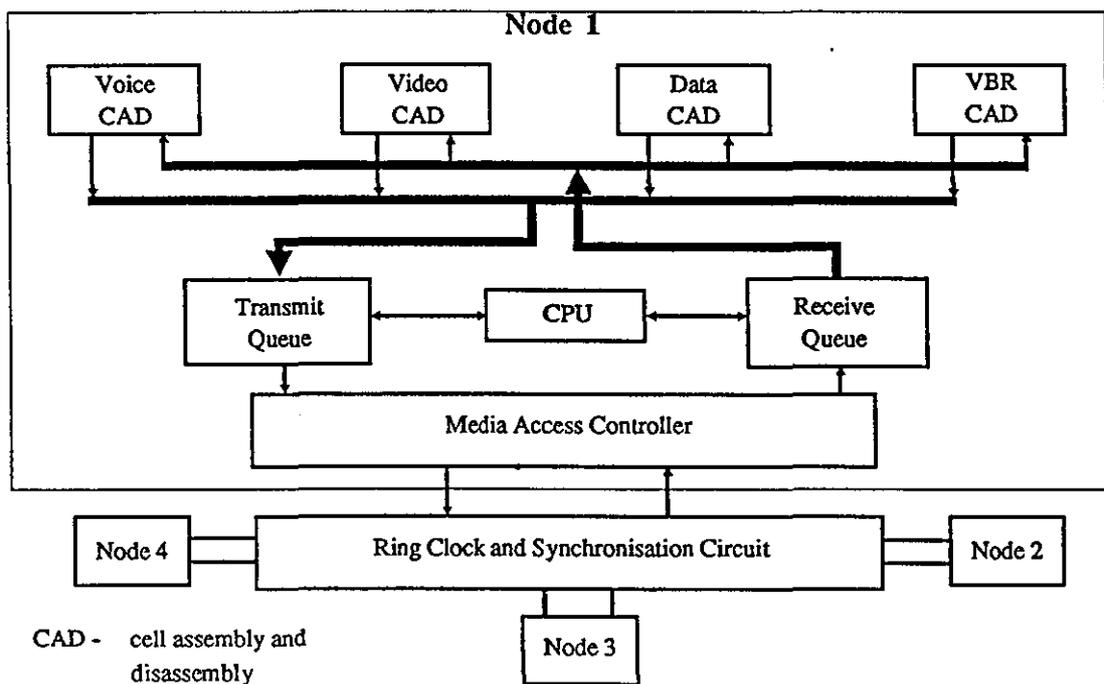


Fig. 6.1 Block Diagram of the Ring Node.

The single node ring has a very low effective bandwidth capacity. This is because the destination release policy of the protocol will not result in any gain in a one node ring, and the bandwidth required for the trial and reset processes will be excessive. On this ring, a node 'd' of 4 has been determined to be the ceiling value, which implies that 60% of the bandwidth is taken up by the two control processes since the node is unable to reuse a slot which it releases. This, together with 23.8% for the cell overhead, drastically reduces the information

transfer capability of the ring to just 4.57 Mb/s. This presents a major problem when an attempt is made to interface the modified COST211 codec, which has a much higher data rate, to the ring.

The ring has been tested by simultaneously connecting a CBO video, a two-way voice, a cell generator which allows the cell generating rate to be varied on line, and a teletext data connections. It has been shown that with the reset interval maintained at or below 125 μ s with the fixed node 'd' allocation, no cell loss arising from excessive delay was inflicted on the two CBO services as the ring load was increased with the cell generator to overload level. The teletext update rate was, however, reduced and eventually stopped, indicating that its allocation was being pre-empted. Although the Orwell protocol is not fully implemented here, the observation does demonstrate the effectiveness of the protocol provided the 125 μ s maximum reset interval can be maintained by those functions not implemented.

6.3 The VBR Video Codec

The COST211 codec and the modifications made to convert it into a VBR video codec have already been described in detail in Chapter 3. However, no consideration has been given to the transmission aspect of the codec in an ATM network environment. This section and the next will address this issue and describe any further modifications necessary to ensure the correct transfer of video data over the network, and to maintain synchronisation between the received data and the decoder.

As the codec was not designed to cope with the type of errors expected in ATM networks, such as burst cell loss and cell jitter, simply packetising the output of the coding loop for transmission will not provide the error recovery capability required. Synchronisation of the received data to the decoder will also present problems because of the irregularity of the data rate and cell delay; these problems will be dealt with here. The issue of error recovery will be discussed in the next section.

In the COST211 codec, frame stores are organised as closed loops of delay stages. Each stage represents one pixel data. Pixel data circulate within the loops which have only a single access point, thus making the timing of updating the frame stores critical. This does not represent a major problem in CBO video over a circuit switched network since transmission delay is constant throughout; but with VBR video in an ATM network, the irregular data rate and large cell jitter make the synchronisation of the received data to the decoder frame store very difficult. Inability to achieve synchronisation would result in the loss of video data. This problem would not be present if a random access frame store had been used instead.

It was decided that a secondary random access frame store would be used at the decoder in addition to its own frame store. The two frame stores were synchronised so that the frame store in the codec was continuously updated in its entirety with data from the

secondary frame store. With this arrangement, the received data were decoded as soon as they were received and the secondary frame store was updated accordingly without any timing constraint. The problem of synchronisation was thus avoided. The decoder frame store was effectively replaced from the main decoding loop by the secondary frame store.

This approach, however, meant that the secondary frame store had to be integrated into the decoding loop, a very difficult task in itself. Therefore, for simplicity and ease of implementation, it was decided that only the conditional replenishment coding with bridging would be considered in the initial work. As a result, only PCM video data were used, giving a peak throughput of 17.5 Mb/s. The mean throughput is picture dependent and 2.6 Mb/s is expected for a mean frame differences of around 15%. It is clear that while the experimental ring is capable of coping with the mean codec throughput, a moderate increase in the throughput would lead to cell loss. This problem was not envisaged at the early stage because a ring with a larger effective bandwidth was anticipated.

6.4 An Addressing Scheme for the Video Data

Simply adopting the addressing scheme used in the COST211 codec to address the video data would not provide adequate error recovery capability to counteract the burst cell loss problem encountered in an ATM network. Error recovery is taken to mean, in this context, the ability of a codec to recover synchronisation from an error condition quickly and with minimal degradation to the picture quality, and is different from error concealment. For this purpose, sufficient auxiliary information must be carried in the data stream to enable the codec to fulfil this requirement.

Cell loss, either singly or in bursts, introduces discontinuity in the data stream whereby synchronisation will be lost. An addressing scheme for the VBR video data in an ATM environment must therefore be designed to enable the detection of cell loss and the rapid re-synchronisation of the incoming data. Failure to do so will result in a significant degradation to the picture quality.

Most addressing schemes for transporting video data are designed for CBO video over a circuit switched network, and are not suited to the transport of VBR video over an ATM network.

The COST211 codec uses a scheme whereby field, line and cluster synchronisation are provided by different unique codes, with the cluster code being the shortest since it is the most frequently used (see Appendix II). A field code identifies the start of an odd or even field, and a line code is transmitted at the beginning of every line without specific reference to the line number and irrespective of whether the line contains any video data; only a

modulo 8 line sequence counter is used to enable line continuity to be checked. Cluster codes separate clusters from one another; luminance clusters are separated from the chrominance clusters using a colour escape code, and each cluster is addressed within a line.

The adoption of this scheme to transport VBR video over an ATM network would present several problems: the loss of several line codes in a row could make re-synchronisation very difficult since the exact line number is not known; the loss of the colour escape code could result in data being fed to the wrong picture area; and the loss of the field code could cause data to be placed in the wrong line, or even the wrong field. Some of these error conditions could result in a large degradation to the observed picture quality since recovery may take a long time.

A prototype addressing scheme based on the above scheme has been designed for the modified COST211 codec for transporting PCM video data over an Orwell Ring. The scheme remedies some of the shortfalls of the original scheme at the expense of a higher addressing overhead. It was decided that only 256 video lines (i.e. the first 128 lines of each field) would be used since they can be conveniently addressed with just one byte, and the maximum throughput of the codec can be reduced. This also has the advantage that the most significant bit of the line address will always indicate which field the line belongs to, and the field code can be made redundant, thus removing one source of error.

Each active line, that is a line that contains video data, is addressed with its line number following the line unique code; line code and line address for non-active lines are not sent. An additional piece of information indicating the number of cells associated with a line is also sent with the line code. The nibble allocated for this purpose would normally be adequate, but in the case of a whole line requiring updating, the number of cells generated would exceed 16. In this circumstance, the line code and line address - known alternatively as the continuation code - would be regenerated at the end of the sixteenth cell to cope with the extra data. Experiments have shown that continuation codes are rarely required. The line code and line address allows the decoder to re-synchronise quickly to the next active line in the case of cell loss and also provides for the detection of cell loss. In the event of mass cell loss, corrective action can be initiated, for instance, a complete frame update can be requested. However, in order to acquire and retrieve the information on the number of cells associated with a particular line, a line store would be required at both ends of the codec.

Each cluster starts with a unique code which indicates whether it carries luminance or chrominance data, followed by the cluster address within a line; so avoiding the misinterpretation of the data in the event of the loss of the colour escape code in the original scheme. In order to avoid excessive delay to the last cell generated in a field, a cell filler code is used to top up any incomplete cell for immediate dispatch rather than waiting for data from the following field.

This addressing scheme was designed for PCM data, and as a result, all codes are byte-oriented to ease implementation. The unique codes were selected from data values which are not used for coding pixel; pixel data use only levels 16 to 239, the rest are available for this purpose. The details of these codes are presented in Appendix II. The byte-oriented approach, however, requires redundant bits to be included to preserve the byte boundary, thus causing inefficiency, especially the use of a byte wide cluster code. Investigation into the efficiency of this addressing scheme has revealed that with very small frame differences, the addressing information could constitute up to 50% of the frame data, but with higher frame differences, for example 15%, efficiency of 70%-80% can be achieved. With this addressing scheme, the codec would have a peak throughput of about 16 Mb/s and a mean of about 3.5 Mb/s, which leaves very little room on the experimental ring for data bursts.

Admittedly, the prototype addressing scheme is not error proof. There are occasions when errors may slip through. One problem is that the unique codes, while unique with respect to the pixel data, overlap with the address space; this could lead to the occasional misinterpretation of the received data when part of the addressing information is lost. Nevertheless, the scheme is considered to have provided sufficient synchronisation information to enable the codec to recover quickly from this error condition.

The capability of the codec to make use of the synchronisation information to detect and recover from cell loss errors will depend upon the intelligence incorporated in the codec to analyse the received data. In the present implementation, only some very basic error detection logic was employed (extra circuitry was built for this purpose). These included:

1. checking for discrepancy between the actual number of cells received in a line and that stipulated in the line code;
2. checking for the relative positions of unique codes, for instance, a line code must be followed by a line address and cluster code; luminance cluster codes must precede chrominance cluster codes and so on;
3. checking the order of the line and cluster addresses. They must be received in ascending order unless intervened by a field change or a line change respectively;
4. checking for valid chrominance cluster address which must not exceed 51;
5. checking for valid pixel data value which must lie in the range 16-239;

When an error is detected, all subsequent cells will be discarded until the detection of a new line code. Error recovery is thus line-based with error propagation being limited to a single line unless cell losses are spread over several consecutive lines.

If the pixel data are to be DPCM and Huffman coded later, unique codes similar to those used by the COST211 can be adopted with the necessary modifications outlined above, together with a new cluster code for the chrominance data. The overall effect is an increase in the addressing overhead, the price to pay for better error recovery.

6.5 Codec to Ring Interfacing

The experimental 15 Mb/s Orwell Ring and the VBR codec have been described. The next step comprised of the linking up of the two systems, which required additional interfacing hardware.

It was recommended that the interface should access the transmit queue board on the ring node directly, with the priority three buffer as the service access point since higher level access points were not provided on the node for this service. This implies that the interface must provide the higher layer services, which include the assembly and disassembly of cells and the provision of cell headers.

Video data were packetised into blocks of 16 bytes as they arrived at the transmit interface, and a 5 byte cell header, the content of which was fixed in this case, was attached to the beginning of each block to form a complete cell. Cell filler codes were generated to complete the last data block in a field when indicated by the codec for immediate forwarding to the node buffer. Cell transfer between the interface and the node buffer was by way of hand-shaking depending on the state of the buffer, and cells were stored in the buffer awaiting transfer to the MAC board for launching onto the ring.

On the receiver side, the node provided an extra level of service to the VBR video connection in that cell headers were stripped off before the blocks of data were stored in the buffer. The interface would be alerted of the presence of received data in the node buffer, and data transfer between the buffer and the interface was again by hand-shaking. The data were then forwarded to the codec. Figure 6.2 illustrates the interfaced experimental system.

6.6 The Experimental Work

The ring and the codec were interfaced as in figure 6.2. Only a single codec was used with the encoder and decoder functioning as the transmitter and receiver respectively. They were configured to work independently over the ring with no other communication path between them; this is thus only a one-way connection. Two other services, namely the two-way voice and the teletext were also supported on the ring when required. The CBO video was excluded because of the limited node throughput.

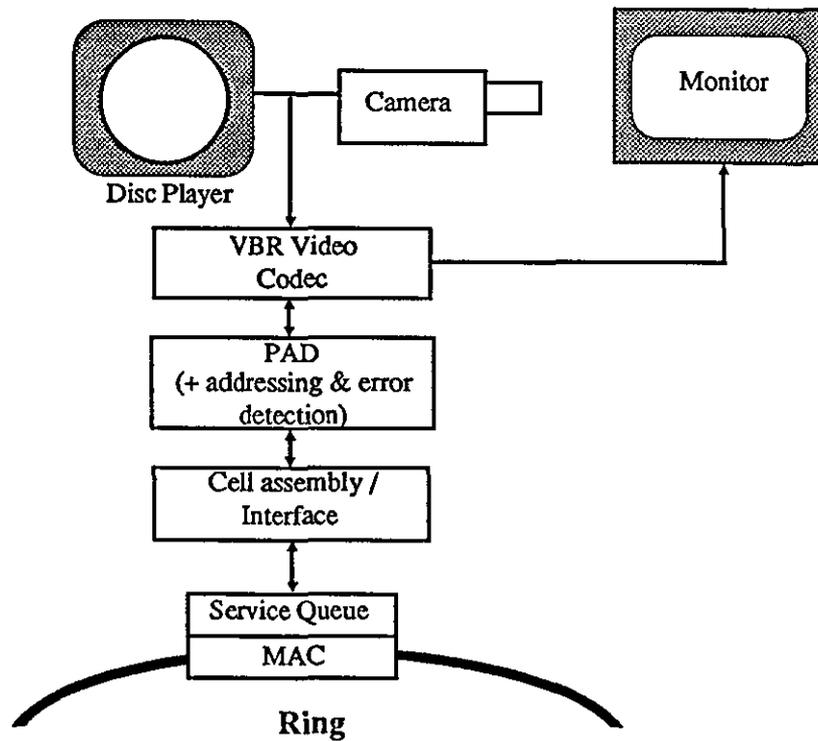


Fig. 6.2 Experimental System Configuration.

In view of the large throughput of the codec compared to the transmission capacity available on the node, initial tests were limited to using test signals generated by additional circuitry on the interface board and picture sequences which were known to generate low data rates. Many hardware and noise-related problems were ironed out to allow the test signals and the selected sequences to be transferred correctly over the ring, thus demonstrating the successful inter-working of the two systems.

Data corruptions were, however, clearly visible in the decoded picture at the receiver in the form of streaks scattered randomly over the frame. This is believed to be due to noise on the data bus linking the interface and the ring. Bit-errors in the addressing information could have caused the decoder to abandon decoding until the receipt of a line synchronisation code, which would have meant the loss of a sizeable amount of video data. Bit-errors in the pixel data that can be detected would also have a similar effect. Errors could also have arisen from corrupted addressing which was not detected, leading to incorrect placement of the pixel data. This problem of noise on the data bus is now being rectified.

The combined system has also been shown to work satisfactorily in a multi-service environment, with the ring supporting the voice and teletext connections as well. The voice

has been shown to be unaffected by data bursts in the VBR video while the teletext update rate was reduced, thus demonstrating the effectiveness of the protocol in meeting its design target.

A full demonstration of VBR video over the ring has not yet been achieved because of the limited throughput of the node in comparison to the peak data rate of the codec. At present, data bursts result in an unrecognisable picture at the receiver due to loss of cells arising from buffer overflow, and corruption in the picture background persists even after the data rate has fallen back to normal. This problem needs to be resolved urgently for demonstration purposes. Several proposals have been suggested: one would be to reduce the size of the picture but not the resolution, in order that the data rate can be contained within the capacity of the node most of the time; this could, however, involve quite a drastic reduction in the picture size and much effort in hardware modifications. A second approach would be to restrict the amount of movement in a picture, for instance, by using a camera or a higher movement detection threshold. This however defeats the purpose of VBR video. Both solutions are considered unsatisfactory, and other alternatives are being sought.

In the mean time, the codec has been demonstrated to work in a back-to-back mode, i.e. with the ring being replaced by a data bus, but with data transfer still being conducted at the cell level. This experiment has helped to demonstrate the correct working of the codec as well as the performance of VBR video, which has been found to be very impressive under error-free conditions.

6.7 Discussion

The modified COST211 codec used in Chapter 3 for signal statistics study was further modified with extra circuitries added to manage the transmission aspects of the codec in the ATM environment. These included the synchronisation of the received data to the decoder frame store and the implementation of an addressing scheme which allows for rapid error recovery.

For ease of implementation of the above requirements, the codec was configured to work in PCM mode using conditional replenishment coding with bridging as the only compression technique. This may be acceptable for the first implementation, but further compression with DPCM and Huffman coding should be included as soon as possible, because the transmission of video data in PCM format is not a likely supposition and it is inherently more error-resilient as a result of the well-defined byte boundary. The implementation of more compression would have a two-fold effect: Firstly, the peak data rate can be lowered, thus reducing the risk of overloading the node; secondly, the extent of the cell loss error can be assessed in a more realistic manner, since error detection will be more difficult and error propagation more serious in the absence of the byte boundary. With further compression, modification to the proposed addressing scheme is necessary insofar as the

unique codes are concerned. The effectiveness of the concept of the addressing scheme can then be assessed under a more rigorous and realistic environment. The addressing scheme can also be made more efficient without the restriction of byte boundary, by using shorter cluster codes and smaller chrominance addressing.

The addressing scheme is by no means error proof, and its effectiveness will largely depend upon the intelligence of the error detection and recovery logic in the codec. This logic can be complicated and will be more so in the absence of a byte boundary in the data stream. In the present implementation, only a minimum error detection and recovery intelligence was built in to allow recovery to the next valid video line; more logic could be provided to allow recovery to the next valid cluster, but the added complexity must be justified in terms of improved picture quality.

The 15 Mb/s ring used in this experimental study is a very basic model, with only the kernel of the Orwell protocol being implemented, i.e. the 'd' counter-based access mechanism with different priority queues. The objectives of this exercise at the present stage are to demonstrate the inter-working of the VBR video codec with the ring and the effectiveness of the protocol. It can be extended later to investigate the effect of cell loss. The experimental study was, however, hampered by noise problems on the combined system.

The goals of this part of the project have been largely achieved insofar as a VBR video codec has been successfully produced and shown to function correctly in a back-to-back mode. The inter-working of the codec and the ring has been successfully demonstrated even though some noise problems in the data stream have yet to be solved. The combined system was also tested in a multi-service environment and the Orwell protocol was shown to function as designed. The concept of VBR video over an Orwell Ring has thus been illustrated.

An urgent problem which requires immediate attention is the much higher output of the codec in comparison to the node throughput. This problem hampers the demonstration of VBR video over the ring and a satisfactory solution must be sought quickly. In the longer term, error concealment must also be considered. Preliminary studies have indicated that background update is the logical choice of error concealment for the COST211 based VBR video codec because of the coding strategy used. Furthermore, background update is required anyway to remove coding errors. It is envisaged that the background update data will be transferred on a more secure high priority channel as a low bandwidth CBO data stream.

Chapter 7 :

Discussion and Conclusions

7.1 Discussion

A number of issues associated with the transmission of VBR video over an Orwell Ring have been considered and investigated in some detail in this research. There are two aspects which are of special interest in this work. The first relates to the assessment of the relative merits of VBR video, and the second to the efficient transfer and control of VBR video over an Orwell Ring. Other related issues such as error conditions, error concealment and video coding in an ATM environment have not been specifically dealt with, but have been discussed generally in Chapter 2.

The research was conducted along the lines set out in Chapter 1. The work and results were reported in detail in Chapters 3 to 6. A general discussion of each of the main parts of the research is given here along with suggestions for any possible extension to the present work.

I : The study of signal statistics and assessment of the relative merits of VBR video. The merits of VBR video were assessed in relation to CBO video using the signal statistics that were collected. This, however, proved to be difficult as there was no single basis upon which a consistent assessment could be carried out. Assessments were therefore based on a number of aspects of the signal characteristics, and not all assessments could be quantified.

The bandwidth gain of 2-3 over CBO video has been based solely on peak-to-mean bandwidth ratio measurements without considering the absolute peaks caused by scene changes. Full resolution for scene changes may be regarded as unnecessary due to the low sensitivity of the eyes to spatial detail during abrupt changes, and some simple mechanism was assumed available to reduce the picture resolution. It was also assumed that the data rate is smoothed within frames. A higher gain figure can be obtained if the absolute peaks or the intra-frame data burst are taken into account.

However, it must be noted that the bandwidth allocated to a particular class of CBO video is not usually optimised to the requirements of the individual sources, i.e. the peak bandwidth requirements used in the PMR measurements are not usually the bandwidth allocated to the CBO video. Instead, the allocation is usually chosen to fit the transmission hierarchy and is therefore sub-optimum. This factor must be considered when interpreting the gain figure quoted above. Comparatively, VBR video is undoubtedly more efficient.

There is also little doubt that VBR video would provide a better picture quality than CBO video under error-free conditions. However, the assessment of picture quality is very subjective, and the improvement in the observed picture quality may not be apparent, especially to the untrained eyes of the majority of viewers, except perhaps in scenes with very large movement. Therefore, this is perhaps not a particularly strong point in favour of VBR video, but nevertheless an advantage in maintaining constant picture quality.

A third advantage of VBR video is that it does not require a complex feedback mechanism, which would otherwise be required to smooth the highly variable data rate and the video-spurts. This can lead to considerable reduction in the hardware complexity of the codec, and to some extent it reduces a source of error since no signalling information will be required for control purposes. Critics may argue that the extra hardware required in the codec to implement error resilience to cope with cell loss could offset this advantage, but in an ATM network, this is necessary even for CBO video (note although CBO services are guaranteed an almost loss-free path on the Orwell Ring, this is not necessarily true on ATM networks of different architectures), especially for those cells carrying synchronisation and control information. Furthermore, the highly error-resilient layered coding technique, which is particularly suitable for video coding in the ATM environment, can be more efficiently implemented with VBR video; the technique also offers other advantages such as user selectable picture quality. There is thus an overall reduction in hardware complexity.

One concerning feature brought out in this study is the low coding and addressing efficiency when the amount of frame differences in a picture is small. This is, however, a direct consequence of the cluster-based coding strategy used in the COST211 codec and not a general coding feature. It can be alleviated if block-based coding is employed, since the block size (unlike the cluster size) is fixed and does not vary with the amount of frame differences. Block coding already forms the basis of many modern codec standards such as CCITT Rec. H261 and ESPRIT Project 925, and the problems identified here need not cause any undue concern unless cluster coding is preferred.

As regards the PDF distribution of the frame differences, the nature of the PDF has not been analysed in any great depth, rather, it was only studied in relation to already established work⁽⁵¹⁾. The nature of the observed PDFs was largely in line with those shown in the reference, as well as in other recently published work. The reason for not conducting an in-depth study in this area is that the required information has already been

provided in the reference; thus this exercise only serves to collate the findings. Moreover, no definite conclusion on the nature of the PDF is possible since it varies with different pictures, each exhibiting some of their own characteristics. The result is thus a general conclusion on the generic shape of the PDF, which is very often a Normal distribution superimposed on a Gamma distribution. The PDFs were clearly affected to some extent when coding was applied.

The study on the statistical multiplexing of the outputs of uncorrelated VBR video sources has confirmed many of the expected effects: a smoother variation of frame differences, a smaller spread of the frame differences around the mean, and a combined PDF approaching the Normal distribution. Peaks occurring due to scene changes were effectively absorbed, making flow control on these peaks less critical. The result has been achieved with only seven sources. It is thus expected that with a larger number of sources, the resultant data flow would be even less varying and thus more manageable. Most of the VBR video bandwidth gain can be obtained since little extra capacity would really be necessary for data bursts.

This study has been confined mainly to one specific but important class of video services, namely the videophone/video-conferencing services. These services stand to gain significantly from VBR video because of the economy in bandwidth requirement (and hence lower charges), and this is important for the widespread acceptance of these services. The large base of uncorrelated sources from these services lend themselves well to benefit from statistical multiplexing. A video-conferencing type picture of the 'split screen' nature was briefly examined, but due to the lack of such picture material, no conclusion was drawn. The frame differences generated were however unmistakably bursty.

Overall, it is felt that the coding strategy used in the COST211 codec is becoming increasingly obsolete. The statistics generated with the codec undoubtedly provide a good foundation for the understanding of VBR video characteristics, but an in-depth understanding of the characteristics of the signals when advanced coding strategies are used, is of pressing importance. A common feature of the new codec standards is the use of hybrid block coding, involving motion estimation and transform coding. Further investigation into video signal statistics should be based on such coding techniques. Another area of equal importance is the study of the signal characteristics when layered coding techniques are used, since these techniques are growing in importance in anticipation of the arrival of future ATM networks.

II : Developing a simple VBR video source model. A simple source model was developed based on the statistics collected in (I). The term 'simple' must be emphasised since detailed modelling of a VBR video source is both extremely complex and unnecessary for this work. Given the constraint in computing resources, the aim was to develop a model

which is economical to generate and yet exhibits the characteristics of the video sources which are most testing to the ring, without being unduly excessive and unrealistic. These include video-spurts and intra-field data burst.

These requirements have led to some simplifications being necessary, thereby limiting the scope of the model. The Erlang distribution was chosen for the data rate PDF instead of the hybrid Gamma - Normal distribution since the former is much simpler to generate. The Erlang distribution is a subset of the Gamma distribution which means that the range of the PDF that can be generated is restricted. This is not considered to be an important set-back and should not have much bearing on the purpose of the model. The 'tail' characteristic of the frame data rate PDF, which is more important, is emulated to some extent. The auto-correlation of the data rate between adjacent frames is fixed for this model; this simplification is necessary because the exact nature of the variation of the correlation is not known. This correlation introduces the video-spurt characteristic required, and by fixing the correlation coefficient at a relatively high value of 0.8 (which corresponds to the observed mean value), the desired video-spurts of varying duration can be generated, thereby satisfying one of the design objectives. The actual variation of the correlation may not be important and the range of variation itself is small (0.7-0.9).

The segmentation of frames into three regions of different activities, where the data rate within each region is assumed constant, is acceptable for videophone type sources. Data rate variation within a region is relatively small around the regional mean, and can easily be absorbed by the packetisation process. A cell information field of 32 or 64 bytes provides the equivalent of a small buffer for smoothing these variations. In the present model, the method of segmenting the three regions may sometimes lead to unproportionally high data rates for the middle region for video sources with a large mean data rate, although this is probably unimportant. The model can be refined, if required, to take into account the fact that the three regions tend to become less distinguishable when the mean data rate of the source is high.

The other simplifications that have been made are the assumption of a fixed coding efficiency and a fixed proportion of addressing overhead. This again is not crucial in this work and their variations have already been partially accounted for in the generation of the frame differences.

The transfer of the first frame of a video sequence has been greatly simplified in the model, based on the assumption that it is not necessary to transfer this frame in real time or with full resolution, in order to relieve the demand placed on the network. The slow build-up of a picture at the beginning of a connection is considered acceptable. Furthermore, since it is effectively a scene change, the human visual system will have a delayed response to such a change. The assumption is thus valid so long as the full frame rate or frame resolution is restored quickly. The model assumed a picture build-up time over four frames, which should be more than acceptable. It is worth noting that with layered coding, this problem can be

easily solved by not transmitting the higher layer information, thus resulting in a reduced resolution picture transferred in real time. Full resolution can be resumed immediately after the first frame.

Notwithstanding the many simplifications made, the model is considered adequate for the purpose of this work. It possesses most of the characteristics of those sources upon which it is modelled, and most of all, those features of VBR video which are most demanding from the network point of view. Undoubtedly more can be done to refine the present model, but it is regarded as unnecessary and may not yield extra information in the context of this research, while only serving to increase the complexity of the model. The simplicity of the model will be appreciated more in a large network where a large number of sources may have to be generated. This simplicity could save considerable computing resources. However, where a very large number of sources are required, it might be better to develop a multi-source model which generates a multiplexed video output rather than a single source output. In this case, only a single model is required on each node and this can potentially yield a greater economy in computing resources, due to the fact that significantly fewer sources need to be generated and a simpler video data generating mechanism may be possible by virtue of the statistical multiplexing effect.

III : Simulation study of VBR video on an Orwell Ring. This part of the work has been greatly restricted by the limited computing resources available. As a result, only a 10 node, 34 Mb/s ring has been investigated. This is, admittedly, a sub-optimum configuration as it has been shown that a larger number of nodes could result in a significant improvement in the gain in the effective ring bandwidth. For instance, a gain of 45% can be achieved with 20 nodes as compared to 30% with 10 nodes, but this is still low when compared with, say, a 25 node, 140 Mb/s ring, which has been reported to have a gain of about 60%. A high capacity ring would be more appropriate where video connections are concerned due to their large throughput. In order to obtain more usable bandwidth on the ring, the cell header has been reduced to 4 bytes, a factor which would not affect the simulation objective. The mean VBR video bandwidth has also been reduced to 350 kb/s to allow for more connections on the ring in order to achieve a good multiplexing effect. This order of compression is highly unlikely with the COST211 codec for a reasonable picture quality.

With larger rings, or simply with more nodes on a ring, the reset rates tend to decrease much more slowly near the full load condition. This characteristic can be used to advantage in supporting VBR video as it helps to absorb some of the irregularity of the data flow and ease the traffic control problem. This is because the traffic control mechanisms on the ring rely on the reset rate measurement, and as it is less sensitive to load variation, it would appear relatively stable even though the actual load is bursty. This would enable a better measure of the ring loading condition and more effective traffic control. Consequently, the risk of ring

overload can be reduced. With the 10 node, 34 Mb/s ring, the drop in the reset rate with increasing load is steeper because the gain from the destination release policy is limited. In this case, traffic control is more critical.

An approximate ring capacity dimensioning method for VBR video traffic has been suggested, and it has been shown to work reasonably well in the simulation study. The pre-specified call blocking performance can be approximately attained with the call acceptance threshold derived from this method with appropriate allowance for the uncertainty in load estimation and variation in ring capacity. The cell loss performance has been shown to conform, to some extent, to the target requirement. There were, however, an insufficient number of runs to verify its accuracy. The prediction of the cell loss rate is only an approximation due to the many assumptions made in the dimensioning method, which have been greatly simplified. For instance, the data rate is not smooth within a frame; the effect of the common node buffer and the 'd' allocation scheme on cell loss is difficult to establish; conformity of the multiplexed video load to the central limit theorem is not strictly correct in the crucial tail region; and the system is dynamic, thus making accurate load measurement difficult. However, the method was from its inception devised to provide only an estimate of the amount of VBR video load that the ring can support with a certain QOS. Accurate dimensioning is not possible in view of the non-deterministic nature of the characteristics of the ring and the VBR video.

Load estimation on the ring was based on a single measurement of the reset rate over a short interval of 2 ms. However, measurement over such an interval could only provide a measure of the instantaneous load condition on the ring rather than the mean loading state. Long intervals are equally unsatisfactory because of the dynamic nature of the system. This problem in accurate load estimation makes load control by call blocking difficult since the success of the mechanism relies on an accurate measure of the load on the ring. A compromised reset rate measurement interval which could provide the best load estimate is highly desirable. For more accurate load estimation, more intelligence would inevitably be required in the load monitoring scheme and hence more complexity. Both options are venues for further investigation.

The dynamic 'd' allocation scheme tested in the simulation has proved to be very effective in controlling the VBR video traffic. It also provides the capability to carry out some load smoothing on the ring by adapting the 'd' allocation to a longer-term cell arrival rate, i.e. a slower response, but sufficient buffering must be made available to absorb the irregular load on the node. Insufficient buffering has been the main cause of many of the cell losses observed when a small node buffer was used. It is, however, not known what the performance of the scheme would be, had higher bit rate sources been used, since the step increment of 'd' in the scheme may not be able to cope with a sudden large increase in the load on a node. This situation may require the use of a larger buffer to sustain the temporary

discrepancy between the bandwidth required and the bandwidth provided, or else the scheme would have to be based on a short-term cell arrival rate, which may not be very satisfactory since it introduces large load variations on the ring. This is also an area for further work.

The exact relationship between the node buffer size and the cell loss rate is relatively unknown. With the dynamic 'd' allocation scheme, only a moderate amount of buffer should be required to cope with the delay in the 'd' adjustment and the simultaneous arrival of cells from several connections. With a reasonably large buffer, some of the temporary overload can also be absorbed. Thus the actual cell loss rate could be lower than that estimated from the ring dimensioning. There is also an added dimension to this : the masking mechanism activated during overload tends to distribute the excess load over all the nodes; the overload burden and all the buffers on the ring are thus effectively shared among all the nodes, and this could further reduce the risk of cell loss. As a result, with a reasonably large node buffer, the cell loss performance remains excellent even under heavy load, while the call blocking performance degrades below the specification, making the latter a more critical criterion in service provision.

The simulation study has shown that with proper dimensioning of the ring and proper exercise of the load control and overload control mechanisms, VBR video can be efficiently transferred over the ring, even in a multi-service environment, while still retaining its gain and achieving the required QOS for all services. Layered coding may or may not be necessary depending on the cell loss performance on the ring and the amount of picture quality degradation acceptable to the viewer. A better cell loss performance is of course achieved at the expense of lower network utilisation. It has also been shown that the ring is more tolerant to overload than was expected. Temporary overloads were easily sustainable without incurring cell loss on the CBO services, and with sufficient buffering, on the VBR video as well.

The results obtained from the simulation work have wide implications. Similar conclusions can be drawn from these results about larger rings or Tori systems. However, in order to assess the performance of a wide area Orwell network, the performance of another component of the network, namely, the link multiplexor, must be studied. Although the link multiplexor operates a similar protocol to the ring, the different architecture of the multiplexor would mean different system characteristics and would require a different setting of the traffic control parameters, or a different control philosophy since the node will have complete control over the link resources. There is, however, more leeway in traffic control on the multiplexor since a modern day point-to-point optical link has a vast capacity, and therefore bandwidth utilisation efficiency may not be an important issue. In addition, the ring should be studied with a cell structure conforming to the emerging ATM standard of a 32 or 64 byte information field.

Other venues for further exploration in this work would inevitably require the use of more computing resources which are not presently available. However, when they are available, the simulation should extend over a longer period. This will enable more consistent results to be obtained, and will also enable a longer call holding time to be used for the VBR video connections, and therefore a more realistic simulation that brings out the true characteristics of VBR video. Higher speed rings and higher bit rate VBR video connections should also be a topic for future investigation. Another area of interest would be to model a layered codec in order to study the transfer of two-layer VBR video traffic over the ring. This represents an important area of research.

IV : Experimental studies on a prototype ring. The objectives of this part of the research have been largely achieved. A VBR video codec has been successfully derived from a CBO video codec and the inter-working of the modified codec with the experimental ring has been successfully demonstrated, although some noise problems on the data bus have yet to be solved. The concept of VBR video over an Orwell Ring has also been demonstrated in a multi-service environment, with the high priority service fully protected by the protocol as desired.

The VBR codec, which is developed from a COST211 codec with ATM networks as the target carriers, uses only conditional replenishment coding with bridging in its present stage. Video data are therefore byte-oriented; this enables the ease of testing new addressing schemes and provides the codec with some error-resilient properties. The codec should be upgraded to include the full COST211 coding strategy as soon as possible, in order to enable the study of the full extent of the cell loss problem with highly compressed video and in the absence of a byte boundary in the data stream.

There are other pressing problems to be solved as well, mainly that of the much higher codec output in relation to the node throughput. This problem may be solved later when the full coding strategy is implemented, but a quick solution is required for the present demonstration of the inter-working of the two systems. Another problem is the implementation of a second channel for background refresh to clear errors in the receiver frame store.

7.2 Conclusions

A number of aspects related to the transmission of VBR video over an Orwell Ring have been examined in this research. Most importantly, the study of the relative merits of VBR video and the simulation study of the efficient transfer of such bursty traffic over an Orwell Ring have shown a strong favour in using VBR video as the basis for supporting future video services on the Orwell networks, particularly the videophone/video-conferencing services where bandwidth, or cost economy is a major factor for their widespread acceptance.

This thesis presents the research work in a progressive manner, by first providing a general background to the effect of efficient video coding on the underlying nature of the output signal bit rate, and then looking at the problems of transmitting and controlling this bursty traffic on ATM networks.

The study began with an investigation into the signal statistics of VBR video signals and an assessment of the relative merits of VBR video based on these results. The investigation was conducted mainly on head and shoulders videophone type pictures using a modified COST211 codec. The results indicated a possible bandwidth gain of 2-3 over CBO video based on the peak-to-mean bandwidth ratio measurements, but not including peaks due to scene changes since full resolution is not considered necessary under this condition. This is, however, not the only basis upon which the merits of VBR video were assessed. The highly variable nature of the signals means that a very sophisticated feedback mechanism would be required to smooth the signals, and adjustments of the picture quality are inevitable in the process. VBR video obviates the need for such a complex mechanism and therefore results in a potentially simpler codec design and a constant picture quality. Other limiting factors of the CBO video were also considered, for instance, the fact that bandwidth allocation for CBO video is not adaptive to the optimum bandwidth requirements of the individual video sources inevitably leads to inefficiency. There is, therefore, much to be gained from the use of VBR video for supporting video services.

The benefit of VBR video cannot, however, be realised unless a large number of uncorrelated sources are available for statistical multiplexing to be effective, thereby smoothing out the large data rate variation. The effects of statistical multiplexing were investigated, and the results demonstrated that data rate smoothing was achieved and peaks were effectively subdued with only seven sources. It is thus concluded that the VBR video gain can be realised with a reasonably large number of uncorrelated sources on an ATM network.

The study of the VBR video signal behaviour has also enabled a simple VBR video source model to be developed for the simulation of the transfer of VBR video over an Orwell Ring. For reasons of economy in computing resources and insufficient data on some aspects of the source behaviour, many simplifications were necessary in the model, thereby limiting its scope. Nevertheless, the model exhibits many of the desired properties of a VBR video source, such as the inter-frame correlation, frame data rate with an Erlang distribution and the intra-frame data burst. The model is considered to be adequate for the purpose of this work, where exact modelling of the source was not the objective.

Simulation of the transmission of VBR video over an Orwell Ring has shown that the Orwell protocol together with the dynamic 'd' allocation scheme provide sufficient traffic control mechanisms to allow for the effective control of the bursty VBR video traffic and the safe transfer of CBO traffic over the ring.

With correct dimensioning of the ring capacity and proper exercise of the control mechanisms provided in the protocol and the 'd' allocation scheme, the required call blocking performance can be achieved for all the services on the ring. The cell loss rate on VBR video can be kept negligible with only a relatively small buffer, and cell security for the CBO services can be maintained. The above can be achieved with little loss in the bandwidth gain of VBR video, and the gain approaches the expected value with a high VBR video traffic level. With 23 E, only 20%-30% of extra capacity was required for an estimated cell loss rate in the order of 10^{-8} on a 10 node, 34 Mb/s ring. Cell delay was minimal for all the services, with a mean delay of no greater than 30 μ s under normal ring loading conditions. The overall performance is expected to be even better with many different services on a network, especially in the presence of some wide bandwidth services at low traffic levels, because of the large amount of accumulated spare capacity reserved for each service to enable it to meet its service requirement.

The use of the dynamic 'd' allocation scheme for the VBR video enables a small node buffer to be employed. As a result, the use of a moderately large buffer was sufficient to cope with temporary overloads and helped to reduce the risk of cell loss. The ring was thus able to support a very small cell loss rate for VBR video without any further complexity in control, or the need for a very large buffer.

Even under overload conditions in terms of call attempts, the control mechanisms were most effective in protecting the high priority services such that no cell loss was incurred upon them. With sufficiently large buffers, losses on VBR video can also be minimised. This was due, in part, to the distribution of the excess load over all the nodes by the masking mechanism during ring overload, thereby avoiding any single node, or group of nodes, having to bear the full effect of the overload, and effectively making all the buffers on the ring available to absorb the excess load.

The ring has also been shown to be tolerant to temporary overloads. Cell security on the high priority services was maintained even though the maximum reset interval was exceeded quite frequently. The minimum reset rate of 16 resets/2ms was, however, only rarely violated. This has led to the conclusion that the activation of the auto-reset within 125 μ s when no reset is received, is inappropriate. The suitability of the use of 2 ms as the basis for the reset rate (or load) measurement is also doubtful.

The COST211 codec has been successfully modified to operate as a VBR video codec. The transmission aspects of the codec in an ATM network environment have been accounted for. These include the use of a random access frame store for data-decoder synchronisation and an addressing scheme which provides for rapid error recovery. The codec has been shown to work satisfactorily in a back-to-back mode and the performance of VBR video is impressive. The inter-working of the codec with the ring has been successfully demonstrated

and the effectiveness of the protocol in a multi-service environment (which includes VBR video) has been verified. The concept of VBR video over an Orwell Ring has thus been demonstrated.

In closing, this research has shown the benefits and feasibility of using VBR video as the basis for supporting future personal video communications over an Orwell-based BISDN. In particular, the more efficient utilisation of network resources achieved with VBR video transmission could reduce the cost of providing these services and help to gain their widespread acceptance by the user base. VBR video should therefore be seriously considered for applications in future video-communication services.

REFERENCES

1. Adams J.L., *Orwell: A Protocol for Carrying Integrated Services on a Digital Communications Ring*, Teletraffic Memorandum R15/12/85, BTRL 1986.
2. Adams J.L., *The Orwell Torus Communication Switch*, GSLB-Seminar on Broadband Switching, Albufeira, Portugal, Jan. 1987, p.215-224
3. Adams J.L., *The Orwell Torus Communication Switch*, Performance Eng. Memorandum 86/14, BTRL 1986
4. Bear D., *Principles of Telecommunication Traffic Engineering*, Institution of Electrical Engineers, 1980
5. Birtwistle G., *Demos Implementation Guide and Reference Manual*, Research Reports No. 81/70/22, Nov. 1981, The University of Calgary
6. Brandt D.(ed.), *Final Report of the COST202bis Expert Group on ATM Switch Structure*, Oct. 1988
7. Candy J.C. et. al., *Transmitting Television as Clusters of Frame-to-Frame Differences*, Bell Syst. Tech.J., Vol.50, July-Aug. 1971, p.1889-1917
8. Carr M.D. et. al., *The Integration of Television Standards Conversion into a Conditional Replenishment Codec for Visual Teleconferencing*, CSELT Technical Reports, Vol.XII, No.5, Oct. 1984
9. Chen H.W. et. al., *Adaptive Coding of Monochrome and Color Images*, IEEE Trans. Comm., Vol.COM-25, No.11, Nov. 1977, p.1285-1292
10. Chin H.S. et. al., *Statistics of the Video Signals for Viewphone Type Pictures*, 2nd Int. Workshop on Packet Video, Torino, Italy, Sept. 1988 Paper D4
11. Chin H.S. et. al., *Statistics of Variable Bit Rate Video Signals for Viewphone Type Pictures*, IERE 5th Int. Conf. on Digital Processing of Signals in Comm., Loughborough, England, Sept. 1988, p.141-148
12. Chin H.S. et. al., *Statistics of Video Signals for Viewphone Type Picture*, IEEE J.Select. Areas in Comm.. To be published in June 1989
13. Chin H.S. et. al., *A Simulation Study on the Transmission of Variable Bit Rate Video Over an ATM Network*, Singapore Int. Conf. on Networks '89, 17th-20th July, 1989
14. Chin H.S. et. al., *The Transmission of Variable Bit Rate Video Over an Orwell Ring*, IEE Colloquium on Packet Video, London, 16th May, 1989
15. Chow M.C., *Variable-Length Redundancy Removal Coders for Differential Coded Video Telephone Signals*, IEEE Trans. Comm., Vol.COM-19, Dec. 1971, p.923-926
16. Clark R.J., *Transform Coding*, Academic Press, 1985

17. Coudreuse J.P., *ATM Networks: State of the Art*, 2nd Int. Workshop on Packet Video, Torino, Italy, Sept. 1988, Paper A2
18. Cravis H., *Communications Network Analysis*, D.C. Heath and Company, 1981
19. Decina M., *Broadband ISDN, Whither ATM ?*, 2nd Int. Workshop on Packet Video, Torino, Italy, Sept. 1988, Paper A1
20. Degan N.D. et. al., *Performance Analysis of a Variable Rate Video Coder for Packet Networks*, 2nd Int. Workshop on Packet Video, Torino, Italy, Sept. 1988, Paper D2
21. European Broadcasting Union, *Relative Timing of the Sound and Vision Components of a Television Signal*, EBU Recom. R37, 1986
22. Falconer R.M. et. al., *Orwell: A Protocol for an Integrated Services Local Network*, Br.Telecom Technol.J., Vol.3, No.4, Oct. 1985 p.27-35
23. Gallager I.D., *Multi-Service Networks*, BT Technol.J., Vol.4, No.1, Jan. 1986, p.43-49
24. Ghanbari M. et. al., *Components of Bit-Rate Variation in Videoconference Signals*, IEEE Select. Areas in Comm.. To be published June 1989
25. Ghanbari M. et. al., *The Interaction Between Video Encoding and Variable Bit Rate Networks*, IERE 5th Int. Conf. on Digital Processing of Signals in Comms., Loughborough, England, Sept. 1988, p.125-131
26. Goodge J.W. et. al., *A Variable Bit Rate Video Server for an Orwell Ring*, Internal Report, Dept. of Electronic and Electrical Eng., University of Technology, Loughborough, March 1989
27. Haskell B.G., *Buffer Sharing by Several Interframe Picturephone Coders*, Bell Syst. Tech.J., Vol.51, No.1, Jan. 1972, p.162-289
28. Haskell B.G. et. al., *Interframe Coding of Videotelephone Pictures*, Proc.IEEE, Vol.60, No.7, July 1972, p.792-800
29. Heinrich A. et. al., *Broadband Communication and its Realisation with Broadband ISDN*, IEEE Comm. Mag., Vol.25, No.11., Nov. 1987, p.8-19
30. Hopper A., *Local Area Network Design*, Addison Wesley, 1986
31. Independent Broadcasting Authority, *Digital Coding Standard*, IBA Technical Review 16, March 1982.
32. Ishiguro T. et. al., *Television Bandwidth Compression Transmission by Motion Compensated Interframe Interframe Coding*, IEEE Comm. Mag., Vol.20, No.11, Nov. 1982, p.24-30
33. Johnson S.A., *A Performance Analysis of Integrated Communications Systems*, Br.Telecom Technol.J., Vol.3, No.4, Oct. 1985 p.36-45
34. Karlsson K. et. al., *Sub-Band Coding of Video Signals for Packet Networks*, Proc. SPIE Conf.of Visual Comm. and Image Processing II, Oct. 1987

35. Kennedy J.B. et. al., *Basic Statistical Methods for Engineers and Scientists*, Harper and Row Publishers, 1986
36. Kishimoto R. et. al., *Packet Loss Compensation for TV Conferencing Video Signals in an ATM Network*, IEEE Select. Areas in Comm.. To be published June 1989.
37. Koga T. et. al., *Statistical Performance Analysis of an Interframe Encoder for Broadcast Television Signals*, IEEE Trans. Comm. Vol. COM-29, No.12, Dec. 1981, p.1868-1875
38. Kruz J.J. et. al., *Statistical Switching Architectures for Future Services*, ISS '84, Florence, May 1984, Paper 43A.1
39. Limb J.O. et. al. *Digital Coding of Color Video Signals - A Review*, IEEE Trans. Comm., Vol.COM-25, No.11, Nov. 1977, p.1349-1383
40. Mitrani I., *Modelling of Computer and Communication Systems*, Cambridge University Press, 1987
41. Musmann H.G. et. al., *Advances in Picture Coding*, Proc.IEEE, Vol.73, No.4, April 1985, p.523-548
42. Netravali A.N., *Motion Compensated Television Coding: Part 1*, Bell Syst. Tech.J., Vol.58, March 1979, p.631-669
43. Netravali A.N., *Picture Coding: A Review*, Proc.IEEE, Vol.68, No.3, March 1980, p.366-406
44. Nicol R.C. et. al., *The Development of the European Video-Conference Codec*, Globecom 82, Miami, Nov. 1982, Paper D4.3
45. Nomura M. et. al., *Layered Packet Loss Protection for VBR Coding Using DCT*, 2nd Int. Workshop on Packet Video, Torino, Italy, Sept. 1988, Paper D7
46. Pannerer Y.M., *The RACE European Projects on Coding*, 2nd Int. Workshop on Packet Video, Torino, Italy, Sept. 1988, Paper B1
47. Poole R.J., *An Introduction to Programming in Simula*, Blackwell Scientific Publications, 1987
48. Pratt W.K., *Digital Image Processing*, John Wiley & Son, 1978
49. Robbins J.O. et. al., *Spatial Sub-Sampling in Motion Compensation Television Coders*, Bell Syst. Tech.J., Vol.61, No.8, Oct. 1982, p.1895-1910
50. Roese J.A. et. al., *Interframe Cosine Transform Image Coding*, IEEE Trans. Comm., Vol.COM-25, Nov. 1977, p.1329-1339
51. Seyler A.J., *Probability Distribution of Television Frame Differences*, Proc.IREE (Australia), Nov. 1965, p.355-367
52. Terada Y., *Trends in Packetised Multi-Media Communication Network Technologies in Japan*, 2nd Int. Workshop on Packet Video, Torino, Italy, Sept. 1988, Paper C1

53. Verbiest W. et. al., *VBR Video Coding and ATM Switching: A Bell-RC Lab. Experiment*, 2nd Int. Workshop on Packet Video, Torino, Italy, Sept. 1988, Paper D1
54. Verbiest W. et. al., *The Impact of the ATM Concept on Video Coding*, IEEE J.Select. Areas in Comm., Vol.SAC-6, No.9, Dec. 1988, P.1623-1632
55. Zafirovic-Vukotic M. et. al., *Performance Modelling of the Orwell Basic Access Mechanism*, ACM SIGCOMM Workshop on Frontiers in Computer Comms. Technology, Stowe, Vermont, Aug. 1987.

Video Discs

- V1. BBC Video, *The World's Greatest Paintings 1 : Adoration*, BBCV 1013L, BBC Enterprises Ltd., 1982
- V2. BBC Video, *The BBC Videobook of British Garden Birds*, BBCV 1005L, BBC Enterprises Ltd., 1981
- V3. BBC Video, *Great Railways 1*, BBCV 1008L, BBC Enterprises Ltd., 1981

APPENDIX I

Sequence 2 :



Frame 0 (0 sec.)



Frame 25 (1 sec.)



Frame 50 (2 sec.)



Frame 75 (3 sec.)



Frame 100 (4 sec.)



Frame 125 (5 sec.)

Sequence 4 :



Frame 0 (0 sec.)



Frame 25 (1 sec.)



Frame 50 (2 sec.)



Frame 75 (3 sec.)



Frame 100 (4 sec.)



Frame 125 (5 sec.)

Sequence 5 :



Frame 0 (0 sec.)



Frame 25 (1 sec.)



Frame 50 (2 sec.)



Frame 75 (3 sec.)



Frame 100 (4 sec.)



Frame 125 (5 sec.)

Sequence 6 :



Frame 0 (0 sec.)



Frame 250 (10 sec.)



Frame 500 (20 sec.)



Frame 750 (30 sec.)



Frame 1000 (40 sec.)



Frame 1250 (50 sec.)

APPENDIX II

A. Synchronisation and Addressing Scheme for the GEC/BT 2 Mb/s Video-conferencing Codec

1 Video Synchronisation

The transmission of video synchronisation is by means of Line Start and Field Start Codes (LSC and FSC).

1.1 Line Start Code

The LSC includes a synchronisation word, a 3-bit line number code and a sub-sampling bit to signal the presence of element sub-sampling. The LSC is 20 bits in length and has the form

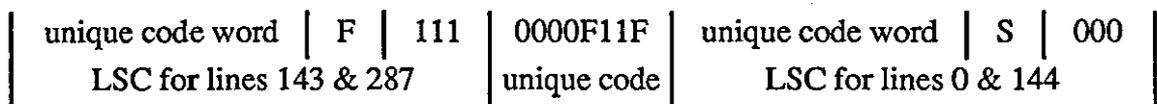


'S' is used to indicate the presence of horizontal sub-sampling on the video line that follows.

The line number code comprises the 3 least significant bits of the line number. Lines 143 and 287 are non-coded lines used for field synchronisation and line number continuity.

1.2 Field Start Code

The FSC indicates the start of a video field. Each FSC comprises the last LSC of a field, followed by an 8-bit code and then followed by the LSC of the first line of the next field. A FSC is as follows:



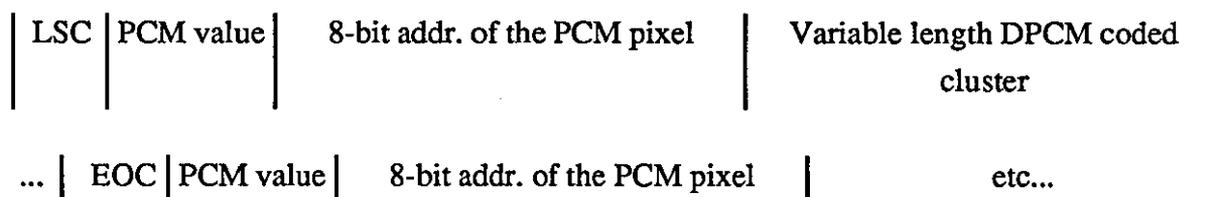
'F' indicates the field in which the following video data belongs, i.e. odd or even field.

2 Addressing of Pixel Clusters

2.1 Luminance Clusters

The positions of the clusters along each line are addressed by means of an address of the start of the cluster and 'End of Cluster' code (EOC).

The form of coding is :



The PCM value is the amplitude of the first pixel of the cluster. When there is no chrominance data, the EOC is omitted from the last luminance cluster of every line, i.e. both the LSC and FSC also signify end of cluster.

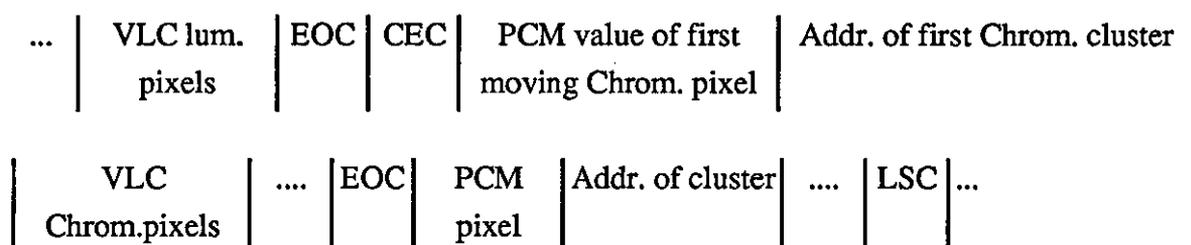
The EOC code is a 4-bit unique code.

The address indicates the sample number along the line of the first pixel of the cluster.

2.2 Chrominance Clusters

A colour escape code (CEC) is defined to allow extra space in the addressing range defined by 8 bits. This escape code is signalled by transmitting the EOC code followed by a unique code, and then the PCM value of the first colour moving pixel.

The form of addressing of chrominance clusters is shown below:



On video line where there is no chrominance cluster, the CEC is not transmitted. Where there is no luminance cluster, the CEC will appear directly after the LSC and precedes the PCM value of the first pixel of the first chrominance cluster.

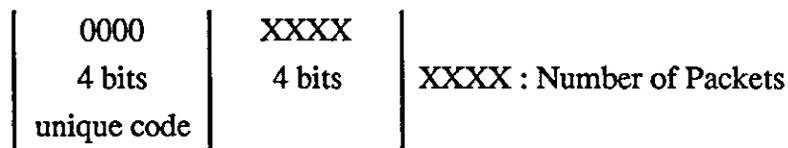
B. Addressing Scheme for the VBR Video Server

In order to preserve the byte boundary of the PCM video data stream, this addressing scheme has been designed to be byte oriented. The following details the individual feature of the addressing scheme. Each byte is identified with one of the three possible field types:

- i) control
- ii) addressing
- iii) data

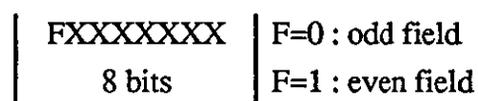
Unique codes are derived for various control functions in the scheme.

1.1 Line start / Continuation Code (LSCC)



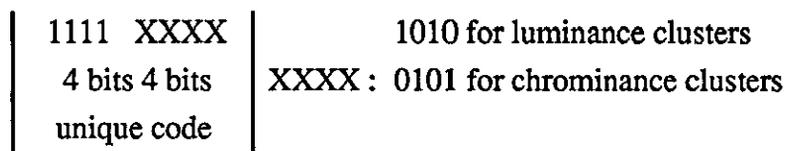
The upper nibble of this control byte is a 'unique' code and has a dual (mutually exclusive) role. When cluster data associated with a particular line is required to be transmitted, a LSCC is sent as the first in a sequence of control, addressing and data information. The lower nibble indicates the number of addition packets required to send the complete cluster information for a given line. If the total line cluster data can be accommodated within a single packet, this value is set to zero. Where the situation arises that more than 16 packets are required to send the line data, a Continuation Code (second role) is generated. Such is the case when a whole line is to be transmitted. Experiments have shown that Continuation Codes arise infrequently. Indeed, this code is only generated in the case where complete frame update or excessive picture movement occurs. In the absence of line cluster data, no LSCC codes are transmitted.

1.2 Line Address (LA)



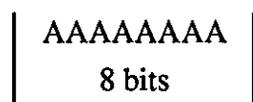
This byte directly follows the LSCC code, and simply indicates with which line the cluster data is associated. The most significant bit is used as a field indicator, and obviates the need for field start code.

1.3 Start of Cluster Code (SLCC, SCCC)



This control byte precedes the cluster data by one byte and in the case of the first cluster in a line, directly follows the Line Address. A 4-bit unique code is again used for the upper nibble; the lower one being used to differentiate between a luminance or chrominance cluster.

1.4 Cluster Address (CA)



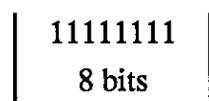
In order to locate precisely where in a line the cluster begins, a Cluster Address precedes the pixel data. In the case of the luminance data, address values lie within the range 1 - 255 and for chrominance addresses 1 - 51.

1.5 Pixel Data (PD)



Conditional update data is transported as 8-bit PCM for both luminance and chrominance pixels.

1.6 Packet Filler (PF)



This byte is used to fill up an otherwise incomplete packet arising at the end of a field, ensuring immediate dispatch. This avoids waiting for the next field's conditional update data to begin.

The following example illustrates how the octets would be assembled for a new line:

| | | | | | |
|-----------|-----------|-----------|-----------|-----------|-----|
| LSCC | LA | SLCC | CA | PD | ... |
| 0000 PPPP | FAAA AAAA | 1111 1010 | AAAA AAAA | DDDD DDDD | ... |
| UC NP | F | UC | | | |

where

- | - byte boundary
- UC - unique code
- NP - number of packets
- F - field indicator

For a new chrominance cluster, this would be assembled as:

| | | | | |
|-----------|-----------|-----------|-----------|-----|
| PD | SCCC | CA | PD | ... |
| DDDD DDDD | 1111 0101 | AAAA AAAA | DDDD DDDD | ... |

APPENDIX III

Table for areas under the normalised Normal curve:

$$Q(x) = 1 - F_x(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp\left(-\frac{t^2}{2}\right) dt$$

| | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|-----|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 0.0 | .50000 | .49601 | .49202 | .48803 | .48405 | .48006 | .47608 | .47210 | .46812 | .46414 |
| 0.1 | .46017 | .45640 | .45224 | .44828 | .44432 | .44038 | .43644 | .43251 | .42858 | .42465 |
| 0.2 | .42074 | .41683 | .41294 | .40905 | .40517 | .40129 | .39743 | .39358 | .38974 | .38591 |
| 0.3 | .38209 | .37828 | .37448 | .37070 | .36693 | .36317 | .35942 | .35569 | .35197 | .34827 |
| 0.4 | .34458 | .34090 | .33724 | .33360 | .32997 | .32636 | .32276 | .31918 | .31561 | .31207 |
| 0.5 | .30854 | .30503 | .30153 | .29806 | .29460 | .29116 | .28774 | .28434 | .28096 | .27760 |
| 0.6 | .27425 | .27093 | .26763 | .26435 | .26109 | .25785 | .25463 | .25143 | .24825 | .24510 |
| 0.7 | .24196 | .23885 | .23576 | .23269 | .22965 | .22663 | .22363 | .22065 | .21770 | .21476 |
| 0.8 | .21186 | .20897 | .20611 | .20327 | .20045 | .19766 | .19489 | .19215 | .18943 | .18673 |
| 0.9 | .18406 | .18141 | .17879 | .17619 | .17361 | .17106 | .16853 | .16602 | .16354 | .16109 |
| 1.0 | .15866 | .15625 | .15386 | .15150 | .14917 | .14686 | .14457 | .14231 | .14007 | .13786 |
| 1.1 | .13567 | .13350 | .13136 | .12924 | .12714 | .12507 | .12302 | .12100 | .11900 | .11702 |
| 1.2 | .11507 | .11314 | .11123 | .10935 | .10749 | .10565 | .10383 | .10204 | .10027 | .09853 |
| 1.3 | .09680 | .09510 | .09342 | .09176 | .09012 | .08851 | .08692 | .08534 | .08379 | .08226 |
| 1.4 | .08076 | .07927 | .07780 | .07636 | .07493 | .07353 | .07215 | .07078 | .06944 | .06811 |
| 1.5 | .06681 | .06552 | .06426 | .06301 | .06178 | .06057 | .05938 | .05821 | .05705 | .05592 |
| 1.6 | .05480 | .05370 | .05262 | .05155 | .05050 | .04947 | .04846 | .04746 | .04648 | .04551 |
| 1.7 | .04457 | .04363 | .04272 | .04182 | .04093 | .04006 | .03920 | .03836 | .03754 | .03673 |
| 1.8 | .03593 | .03515 | .03438 | .03362 | .03288 | .03216 | .03144 | .03074 | .03005 | .02938 |
| 1.9 | .02872 | .02807 | .02743 | .02680 | .02619 | .02559 | .02500 | .02442 | .02385 | .02330 |
| 2.0 | .02275 | .02222 | .02169 | .02118 | .02068 | .02018 | .01970 | .01923 | .01876 | .01831 |
| 2.1 | .01786 | .01743 | .01700 | .01659 | .01618 | .01578 | .01539 | .01500 | .01463 | .01426 |
| 2.2 | .01390 | .01355 | .01321 | .01287 | .01255 | .01222 | .01191 | .01160 | .01130 | .01101 |
| 2.3 | .01072 | .01044 | .01017 | .00990 | .00964 | .00939 | .00914 | .00889 | .00866 | .00842 |
| 2.4 | .00820 | .00798 | .00776 | .00755 | .00734 | .00714 | .00695 | .00676 | .00657 | .00639 |
| 2.5 | .00621 | .00604 | .00587 | .00570 | .00554 | .00539 | .00523 | .00508 | .00494 | .00480 |
| 2.6 | .00466 | .00453 | .00440 | .00427 | .00415 | .00402 | .00391 | .00379 | .00368 | .00357 |
| 2.7 | .00347 | .00336 | .00326 | .00317 | .00307 | .00298 | .00289 | .00280 | .00272 | .00264 |
| 2.8 | .00256 | .00248 | .00240 | .00233 | .00226 | .00219 | .00212 | .00205 | .00199 | .00193 |
| 2.9 | .00187 | .00181 | .00175 | .00169 | .00164 | .00159 | .00154 | .00149 | .00144 | .00139 |
| 3.0 | 1.35E-3 | 1.31E-3 | 1.26E-3 | 1.22E-3 | 1.18E-3 | 1.14E-3 | 1.11E-3 | 1.07E-3 | 1.04E-3 | 1.00E-3 |
| 3.1 | 9.68E-4 | 9.36E-4 | 9.04E-4 | 8.74E-4 | 8.45E-4 | 8.16E-4 | 7.89E-4 | 7.62E-4 | 7.36E-4 | 7.11E-4 |
| 3.2 | 6.87E-4 | 6.64E-4 | 6.41E-4 | 6.19E-4 | 5.98E-4 | 5.77E-4 | 5.57E-4 | 5.38E-4 | 5.19E-4 | 5.01E-4 |
| 3.3 | 4.83E-4 | 4.67E-4 | 4.50E-4 | 4.34E-4 | 4.19E-4 | 4.04E-4 | 3.90E-4 | 3.76E-4 | 3.62E-4 | 3.50E-4 |
| 3.4 | 3.37E-4 | 3.25E-4 | 3.13E-4 | 3.02E-4 | 2.91E-4 | 2.80E-4 | 2.70E-4 | 2.60E-4 | 2.51E-4 | 2.42E-4 |
| 3.5 | 2.33E-4 | 2.24E-4 | 2.16E-4 | 2.08E-4 | 2.00E-4 | 1.93E-4 | 1.85E-4 | 1.79E-4 | 1.72E-4 | 1.65E-4 |
| 3.6 | 1.59E-4 | 1.53E-4 | 1.47E-4 | 1.42E-4 | 1.36E-4 | 1.31E-4 | 1.26E-4 | 1.21E-4 | 1.17E-4 | 1.12E-4 |
| 3.7 | 1.08E-4 | 1.04E-4 | 9.96E-5 | 9.58E-5 | 9.20E-5 | 8.84E-5 | 8.50E-5 | 8.16E-5 | 7.84E-5 | 7.53E-5 |
| 3.8 | 7.24E-5 | 6.95E-5 | 6.67E-5 | 6.41E-5 | 6.15E-5 | 5.91E-5 | 5.67E-5 | 5.44E-5 | 5.22E-5 | 5.01E-5 |
| 3.9 | 4.81E-5 | 4.62E-5 | 4.43E-5 | 4.25E-5 | 4.08E-5 | 3.91E-5 | 3.75E-5 | 3.60E-5 | 3.45E-5 | 3.31E-5 |
| 4.0 | 3.17E-5 | 3.04E-5 | 2.91E-5 | 2.79E-5 | 2.67E-5 | 2.56E-5 | 2.45E-5 | 2.35E-5 | 2.25E-5 | 2.16E-5 |
| 4.1 | 2.07E-5 | 1.98E-5 | 1.90E-5 | 1.81E-5 | 1.74E-5 | 1.66E-5 | 1.59E-5 | 1.52E-5 | 1.46E-5 | 1.40E-5 |
| 4.2 | 1.34E-5 | 1.28E-5 | 1.22E-5 | 1.17E-5 | 1.12E-5 | 1.07E-5 | 1.02E-5 | 9.78E-6 | 9.35E-6 | 8.94E-6 |
| 4.3 | 8.55E-6 | 8.17E-6 | 7.81E-6 | 7.46E-6 | 7.13E-6 | 6.81E-6 | 6.51E-6 | 6.22E-6 | 5.94E-6 | 5.67E-6 |
| 4.4 | 5.42E-6 | 5.17E-6 | 4.94E-6 | 4.72E-6 | 4.50E-6 | 4.30E-6 | 4.10E-6 | 3.91E-6 | 3.74E-6 | 3.56E-6 |
| 4.5 | 3.40E-6 | 3.24E-6 | 3.09E-6 | 2.95E-6 | 2.82E-6 | 2.68E-6 | 2.56E-6 | 2.44E-6 | 2.33E-6 | 2.22E-6 |
| 4.6 | 2.11E-6 | 2.02E-6 | 1.92E-6 | 1.83E-6 | 1.74E-6 | 1.66E-6 | 1.58E-6 | 1.51E-6 | 1.44E-6 | 1.37E-6 |
| 4.7 | 1.30E-6 | 1.24E-6 | 1.18E-6 | 1.12E-6 | 1.07E-6 | 1.02E-6 | 9.69E-7 | 9.22E-7 | 8.78E-7 | 8.35E-7 |
| 4.8 | 7.94E-7 | 7.56E-7 | 7.19E-7 | 6.84E-7 | 6.50E-7 | 6.18E-7 | 5.88E-7 | 5.59E-7 | 5.31E-7 | 5.05E-7 |
| 4.9 | 4.80E-7 | 4.56E-7 | 4.33E-7 | 4.12E-7 | 3.91E-7 | 3.72E-7 | 3.53E-7 | 3.35E-7 | 3.18E-7 | 3.02E-7 |

The Erlang 'B' Capacity Table

| Number of trunks | 1 lost call in | | | | Number of trunks | 1 lost call in | | | |
|------------------------|----------------|---------------|----------------|-----------------|------------------------|----------------|---------------|----------------|-----------------|
| | 50 (0.02) | 100 (0.01) | 200 (0.005) | 1000 (0.001) | | 50 (0.02) | 100 (0.01) | 200 (0.005) | 1000 (0.001) |
| | <i>E</i> | <i>E</i> | <i>E</i> | <i>E</i> | | <i>E</i> | <i>E</i> | <i>E</i> | <i>E</i> |
| 1 | 0.020 | 0.010 | 0.005 | 0.001 | 24 | 16.6 | 15.3 | 14.2 | 12.2 |
| 2 | 0.22 | 0.15 | 0.105 | 0.046 | 25 | 17.5 | 16.1 | 15.0 | 13.0 |
| 3 | 0.60 | 0.45 | 0.35 | 0.19 | 26 | 18.4 | 16.9 | 15.8 | 13.7 |
| 4 | 1.1 | 0.9 | 0.7 | 0.44 | 27 | 19.3 | 17.7 | 16.6 | 14.4 |
| 5 | 1.7 | 1.4 | 1.1 | 0.8 | 28 | 20.2 | 18.6 | 17.4 | 15.2 |
| 6 | 2.3 | 1.9 | 1.6 | 1.1 | 29 | 21.1 | 19.5 | 18.2 | 15.9 |
| 7 | 2.9 | 2.5 | 2.2 | 1.6 | 30 | 22.0 | 20.4 | 19.0 | 16.7 |
| 8 | 3.6 | 3.2 | 2.7 | 2.1 | 31 | 22.9 | 21.2 | 19.8 | 17.4 |
| 9 | 4.3 | 3.8 | 3.3 | 2.6 | 32 | 23.8 | 22.1 | 20.6 | 18.2 |
| 10 | 5.1 | 4.5 | 4.0 | 3.1 | 33 | 24.7 | 23.0 | 21.4 | 18.9 |
| 11 | 5.8 | 5.2 | 4.6 | 3.6 | 34 | 25.6 | 23.8 | 22.3 | 19.7 |
| 12 | 6.6 | 5.9 | 5.3 | 4.2 | 35 | 26.5 | 24.6 | 23.1 | 20.5 |
| 13 | 7.4 | 6.6 | 6.0 | 4.8 | 36 | 27.4 | 25.5 | 23.9 | 21.3 |
| 14 | 8.2 | 7.4 | 6.6 | 5.4 | 37 | 28.3 | 26.4 | 24.8 | 22.1 |
| 15 | 9.0 | 8.1 | 7.4 | 6.1 | 38 | 29.3 | 27.3 | 25.6 | 22.9 |
| 16 | 9.8 | 8.9 | 8.1 | 6.7 | 39 | 30.1 | 28.2 | 26.5 | 23.7 |
| 17 | 10.7 | 9.6 | 8.8 | 7.4 | 40 | 31.0 | 29.0 | 27.3 | 24.5 |
| 18 | 11.5 | 10.4 | 9.6 | 8.0 | 41 | 32.0 | 29.9 | 28.2 | 25.3 |
| 19 | 12.3 | 11.2 | 10.3 | 8.7 | 42 | 32.9 | 30.8 | 29.0 | 26.1 |
| 20 | 13.2 | 12.0 | 11.1 | 9.4 | 43 | 33.8 | 31.7 | 29.9 | 26.9 |
| 21 | 14.0 | 12.8 | 11.9 | 10.1 | 44 | 34.7 | 32.6 | 30.8 | 27.7 |
| 22 | 14.9 | 13.7 | 12.6 | 10.8 | 45 | 35.6 | 33.4 | 31.6 | 28.5 |
| 23 | 15.7 | 14.5 | 13.4 | 11.5 | 46 | 36.6 | 34.3 | 32.5 | 29.3 |

| Number of trunks | 1 lost call in | | | | Number of trunks | 1 lost call in | | | |
|------------------------|----------------|---------------|----------------|-----------------|------------------------|----------------|---------------|----------------|-----------------|
| | 50 (0-02) | 100 (0-01) | 200 (0-005) | 1000 (0-001) | | 50 (0-02) | 100 (0-01) | 200 (0-005) | 1000 (0-001) |
| | <i>E</i> | <i>E</i> | <i>E</i> | <i>E</i> | | <i>E</i> | <i>E</i> | <i>E</i> | <i>E</i> |
| 47 | 37.5 | 35.2 | 33.3 | 30.1 | 74 | 62.9 | 59.8 | 57.3 | 52.6 |
| 48 | 38.4 | 36.1 | 34.2 | 30.9 | 75 | 63.9 | 60.7 | 58.2 | 53.5 |
| 49 | 39.4 | 37.0 | 35.1 | 31.7 | 76 | 64.8 | 61.7 | 59.1 | 54.3 |
| 50 | 40.3 | 37.9 | 35.9 | 32.5 | 77 | 65.8 | 62.6 | 60.0 | 55.2 |
| 51 | 41.2 | 38.8 | 36.8 | 33.4 | 78 | 66.7 | 63.6 | 60.9 | 56.1 |
| 52 | 42.1 | 39.7 | 37.6 | 34.2 | 79 | 67.7 | 64.5 | 61.8 | 56.9 |
| 53 | 43.1 | 40.6 | 38.5 | 35.0 | 80 | 68.6 | 65.4 | 62.7 | 57.8 |
| 54 | 44.0 | 41.5 | 39.4 | 35.8 | 81 | 69.6 | 66.3 | 63.6 | 60.3 |
| 55 | 45.0 | 42.4 | 40.3 | 36.7 | 82 | 70.5 | 67.2 | 64.5 | 59.5 |
| 56 | 45.9 | 43.3 | 41.2 | 37.5 | 83 | 71.5 | 68.1 | 65.4 | 60.4 |
| 57 | 46.9 | 44.2 | 42.1 | 38.3 | 84 | 72.4 | 69.1 | 66.3 | 61.3 |
| 58 | 47.8 | 45.1 | 43.0 | 39.1 | 85 | 73.4 | 70.1 | 67.2 | 62.1 |
| 59 | 48.7 | 46.0 | 43.9 | 40.0 | 86 | 74.4 | 71.0 | 68.1 | 63.0 |
| 60 | 49.7 | 46.9 | 44.7 | 40.8 | 87 | 75.4 | 71.9 | 69.0 | 63.9 |
| 61 | 50.6 | 47.9 | 45.6 | 41.6 | 88 | 76.3 | 72.8 | 69.9 | 64.8 |
| 62 | 51.6 | 48.8 | 46.5 | 42.5 | 89 | 77.2 | 73.7 | 70.8 | 65.6 |
| 63 | 52.5 | 49.7 | 47.4 | 43.4 | 90 | 78.2 | 74.7 | 71.8 | 66.6 |
| 64 | 53.4 | 50.6 | 48.3 | 44.1 | 91 | 79.2 | 75.6 | 72.7 | 67.4 |
| 65 | 54.4 | 51.5 | 49.2 | 45.0 | 92 | 80.1 | 76.6 | 73.6 | 68.3 |
| 66 | 55.3 | 52.4 | 50.1 | 45.8 | 93 | 81.0 | 77.5 | 74.3 | 69.1 |
| 67 | 56.3 | 53.3 | 51.0 | 46.6 | 94 | 81.9 | 78.4 | 75.4 | 70.0 |
| 68 | 57.2 | 54.2 | 51.9 | 47.5 | 95 | 82.9 | 79.3 | 76.3 | 70.9 |
| 69 | 58.2 | 55.1 | 52.8 | 48.3 | 96 | 83.8 | 80.3 | 77.2 | 71.8 |
| 70 | 59.1 | 56.0 | 53.7 | 49.2 | 97 | 84.8 | 81.2 | 78.2 | 72.6 |
| 71 | 60.1 | 57.0 | 54.6 | 50.1 | 98 | 85.7 | 82.2 | 79.1 | 73.5 |
| 72 | 61.0 | 58.0 | 55.5 | 50.9 | 99 | 86.7 | 83.2 | 80.0 | 74.4 |
| 73 | 62.0 | 58.9 | 56.4 | 51.8 | 100 | 87.6 | 84.0 | 80.9 | 75.3 |

