EXPLAINED Worldwide Network and Applications Technology John M. Griffiths

British

ΕL

ECUM



ISDN Explained

Worldwide Network and Applications Technology

by

John M. Griffiths British Telecom

with Peter Adams Derek Rumsey Malcolm Jones Paul McDonald

John Hovell John Atkins Graham Hudson Kevin Woolland Paul Challener Geoff Morrison

JOHN WILEY & SONS Chichester · New York · Brisbane · Toronto · Singapore

Copyright © 1990 by John Wiley & Sons Ltd. Baffins Lane, Chichester West Sussex PO19 1UD, England

Reprinted November 1991

All rights reserved.

BT will donate the royalties from the sale of this book to charity.

No part of this book may be reproduced by any means, or transmitted, or translated into a machine language without the written permission of the publisher.

Other Wiley Editorial Offices

John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0012, USA

Jacaranda Wiley Ltd, G.P.O. Box 859, Brisbane, Queensland 4001, Australia

John Wiley & Sons (Canada) Ltd, 22 Worcester Road, Rexdale, Ontario M9W 1L1, Canada

John Wiley & Sons (SEA) Pte Ltd. 37 Jalan Pernimpin 05-04, Block B, Union Industrial Building, Singapore 2057

Library of Congress Cataloging-in-Publication Data:

ISDN explained : worldwide network and applications technology / edited by John M. Griffiths. p. cm.

 ISBN 0 471 92675 2

 I. Integrated services digital networks.

 I. Sciffiths, John M.

 TK5103.7.1835
 1990

 621.382—dc20
 90-35078

 CIP

British Library Cataloguing in Publication Data:

ISDN explained. I. Telecommunication systems. Networks. Integrated services. I. Griffiths, john M. 621.382

ISBN 0 471 92675 2

Contents

Preface			ix
Biographies	I		xi
Chapter 1	Ne	twork Evolution	1
	1.1	The Telegraph	1
	1.2	The Telephone Arrives	1
	1.3	Telex	. 5
	1.4	The Computer and the Transistor	6
	1.5	Data Networks	7
	1.6	Digital Telephony	8
	1.7	The ISDN	9
	1.8	Early ISDNs	10
	1.9	International Standards	12
	1.10	Open Systems Interconnect (OSI); the 7-layer Model	13
		References	18
Chapter 2	Puł	se Code Modulation	19
-	2.1	Sampling	19
	2.2	Coding	20
	2.3	Compression	20
	2.4	64 kbit/s Channels	22
		Reference	23
Chapter 3	The	Integrated Digital Network	25
-	3.1	Multiplexing	25
	3.2	Transmission	26
	3.3	Switching	29
	3.4	Remote Units	31
	3.5	Signalling	33
		References	35
Chapter 4	Loca	al Network Digital Transmission	
	Peter	Adams	37
	4.1	The Local Network	37
	4.2	Cable Characteristics	39
	4.3	Modulation	42

	m train	
4.4	Equalization	44
4.5	Timing Recovery	46
4.6	Duplex Transmission	46
4.7	A Comparison of Transmission Techniques	49
4.8	Crosstalk Analysis	50
4.9	Impulsive Noise	51
4.10	Echo Canceller Structures	52
4.11	Echo Canceller Adaptation	54
4.12	Non-linearity	55
4.13	Jitter Performance	55
4.14	Dynamic Range	55
4.15	Transmitter Realization	57
4.16	Receiver Realization	57
4.17	Line Interfacing	57
4.18	USA Standard System	58
4.19	Modulation Format	58
4.20	Transmit Signal Specification	60
4.21	Local Network Transmission Outside USA	63
	Reference	63

Chapter 5	The Basic Rate ISDN Customer's Interface		65
	5.1	The Customer's Installation John Hovell	65
		5.1.1 Power feeding	66
		5.1.2 Layer 1 of the user-network interface	67
		5.1.3 The functions of Layer 1	67
		5.1.4 Binary organization of Layer 1 frame	68
		5.1.5 Layer I D channel contention procedure	70
		5.1.6 Layer 1 activation/deactivation procedure	71
		5.1.7 Layer I physical implementation	72
	5.2	Layer 2 Kevin Woollard	73
		5.2.1 Layer 2 addressing	75
		5.2.2 Layer 2 operation	78
		5.2.3 Layer 2 error control	79
		5.2.4 Disconnecting the LAP	81
		5.2.5 TEI allocation	81
	5.3	Layer 3 Derek Rumsey	82
		5.3.1 Layer 3 supplementary services	88
	5.4	Relationship with Signalling System CCITT No 7	89
	5.5	Maintenance	89
	5.6	Testing and Approvals	91
		References	93
Chapter 6	Prin	ary Rate ISDN Access	95

•		I TIMALY RATE IS DIA ACCESS		
	6.I	Background	95	
	6.2	Signalling	96	
	6.3	The UK Solution	97	
	6.4	International Standards for Inter-PABX Signalling	100	
		References	100	

Chapter 7	Frame Mode Services John	Atkins 103
	7.1 Store-and-forward Switching	g 103
	7.2 Datagrams and Virtual Circu	lits 104
	7.3 Flow and Congestion Contro	bl 104
	7.4 Standards	106
	7.5 Support of X.25 by an ISDN	107
	7.6 New ISDN Frame Mode Ser	vices 108
	References	112
Chapter 8	ISDN Customer Premises E	quipment and
	Applications	115
	8.1 High Quality Speech Paul Ch	allener 117
	8.1.1 Speech coding using A	DPCM 118
	8.1.2 Sub-band ADPCM for I	high quality speech [2]
	8.1.3 Embedded data chanr	nei 123
	8.2 Facsimile Malcolm Jones	124
	8.2.1 Group 4 facsimile for IS	SDN 126
	8.2.2 Group 4 compression a	algorithm 126
	8.2.3 Facsimile terminal app	paratus 131
	3.3 Photographic Videotex Grah	am Hudson 132
	8.3.1 Coding	132
	8.3.2 Discrete Cosine Transf	orm (DCT) 135
	8.3.3 Functionality of the da	ta compression technique 139
	8.3.4 ISO DCT conformance	levels [4]
	3.4 Video Geoff Morrison	142
	8.4.1 Source format	144
	8.4.2 Source coding	145
	8.4.3 Video multiplex coding	g 149
	3.5 Customer Premises Equipments	ent 153
	References	155
Chapter 9	Broadband ISDN	157
-	.I N × 64 kbit/s	157
	0.2 H Channels	160
	.3 Higher Rate Interfaces	161
	9.4 Synchronous Digital Hierarch	ny Paul McDonald 162
	9.5 Asynchronous Transfer Mode	e (ATM) 169
	.6 The Future	170
	References	171

Preface

To the telecommunications network engineer it is straightforward evolution to allow the tide of digital working to ripple out to the periphery of the network. To the user the arrival of the tide means a change of the network interface, that interface which had been stable for decades.

King Canute is often depicted as a fool who claimed he could stop the tide. In fact scholars say he was a wise man who was simply demonstrating to his followers the limits of his powers. Even if not surprised he must have found it uncomfortable when the first cold waves lapped his feet. The ISDN is lapping at our feet and the tide is still coming in.

Past disbelievers who originally wrote off ISDN as 'Innovations Subscribers Don't Need' have moved to a more ambiguous 'I still Don't Know'. Some have become converts and are saying 'I Smell Dollars Now' but have been held back because the arrival of the ISDN was 'Irritatingly Slow, Definitely Needed'.

This book is designed to help those who have noticed the arrival of the ISDN and need the technical background. It is pitched at a level deeper than the general 'gee-whizz, isn't it wonderful' by going into the principles of all facets of the ISDN operation. It does not pretend to go down to the detail needed to build network or customers' premises equipment components. The full details would require a book several times the size of this. The object of this book is to give an overview either to set the operations in context, or to act as an introduction to detailed Standards, Recommendations and specifications.

The work recorded in this book is the result of many people's inspiration. I would like to thank the contributors to this book who have filled in their own areas of expertise. The author of Section 8.3 would also like to thank Gregory Wallace for his assistance.

I would also like to compliment Peter Adams, who wrote Chapter 4, for his leadership of the team at British Telecom Research Laboratories whose work on local network transmission and the 2BIQ line code has led to what is likely to become *the* local network transmission system.

Finally I would like to thank Sue Hornsby for her patience in typing and retyping this book.

John Griffiths Martlesham January 1990

Biographies

John Griffiths (who wrote all the unattributed sections and edited the book) was educated at Ealing Grammar School and the University of Manchester whence he graduated with a 1st in Electrical Engineering. Out of perversity he then spent a year on flying training in the Royal Air Force before joining British Telecom (then the Post Office) in 1967, working on early PCM systems. In 1971 he was promoted and bypassed offers of jobs in the proper subjects of switching and transmission, to take a job in the wilderness (as it was then) of Local Telecommunications. From early work on CATV and Local Network Transmission, this evolved to ISDN and its protocols and has progressed to broadband ISDN. He is now Division Manager of the Digital Network Division at British Telecom Research Laboratories, Martlesham, and a Fellow of the Institution of Electrical Engineers. On behalf of that organization he is the UK representative on the Council of the International Symposium on Subscriber Loops and Services (ISSLS). He is married with two children and for hobbies he is a flying instructor and train spotter.

Peter Adams (who wrote Chapter 4) joined British Telecom in 1966 as a student apprentice. After obtaining a BSc in electronic engineering from Southampton University in 1970 he joined British Telecom Research Laboratories where he worked on telecommunications applications of digital signal processing. From 1979 he has headed a group concerned with speech-band data moderns and subsequently local network transmission and maintenance. Currently he leads a Local Network Exploitation section.

John Hoveli (Section 5.1) joined British Telecom as an apprentice in 1967. He moved to the Research Department in 1974 after graduating with an honours degree in electrical and electronic engineering where he worked on line card design for early digital exchanges. In 1981 he obtained an MSc in telecommunications systems and went on to lead a group responsible for the ISDN user/network interface. Recently this work has expanded to include testing, in particular the application of Formal Methods to the proving of signalling protocols. He is a chartered engineer and member of the IEE.

Kevin Woollard (Section 5.2) graduated with a BSc degree in physics from Bath University in 1983. He returned to British Telecom to work on network related aspects of ISDN and in particular has implemented a Layer 2 handler for an experimental ISDN exchange. He was involved in ISDN conformance test specifications as a member of the European group responsible for the WAN-CTS conformance tests for ISDN equipment working on aspects of automatic test case generation from protocol specifications.

Derek Rumsey (Section 5.3) joined British Telecom in 1965. Following a period working on the development of microwave radio links, he joined a division responsible for the system design of data networks where he specialized in packet network protocols. In 1981 he became involved in the development of protocols for ISDN access and currently heads a group responsible for the specification of signalling systems for access to British Telecom's ISDN.

John Atkins (Chapter 7) is Head of the Data Communication Networks Section at British Telecom Research Laboratories. His background includes R&D in dgital switching systems for voice and data services, exploiting both circuit and packet modes of operation. His present interests lie mainly in the networking of non-voice services, and span LANs, MANs, and public wide area networks, with special reference to the ISDN.

Paul Challener (Section 8.1) gained a 1st class honours degree in electronic engineering from Nottingham University in 1974 and became a Chartered Engineer in 1981. He is well known internationally in the speech coding community and has contributed extensively in IEE and IEEE activities both at home and abroad. He is currently Head of the Voice Terminals Section at British Telecom Research Laboratories, concentrating on developments involving interactive speech and high quality audio processing. He has worked on high quality speech and music coding since 1982.

Malcolm Jones (Section 8.2) joined the Research Department of the Post Office in 1963 and worked on communications systems for government departments until 1970 when more specialist activities on maximizing the capacity of coaxial cables for submarine and land/line applications were undertaken. This was followed by data modem design using adaptive equalizers for higher speed operation over the PSTN. This led to general terminal development and international standardization work for the PSTN particularly in the areas of facsimile and teletex. Current activities are still in the area of terminal development including for the ISDN in the form of PC based 64-kbit/s terminal products and for satellite based networks with their unique broadcasting ability.

Graham Hudson (Section 8.3) started his career in British Telecom as a Technician Apprentice in 1966, and then, with a BT scholarship, he obtained an honours degree in electrical and electronic engineering at London City University. In the early 1980s, as Head of the Advanced Videotex Terminals Group, he was responsible for the development of ISDN photovideotex terminals. During this time he was chairman of the ESPRIT PICA project and

the ISO JPEG, both concerned with the development and selection of photographic coding techniques. He is now Head of the Home and Office System Section at British Telecom Research Laboratories. He is a chartered engineer and a member of the IEE.

Geoff Morrison (Section 8.4) joined the Visual Telecommunications Division of the British Post Office Research Laboratories after graduating from St John's College Cambridge in 1971. He has worked on video terminal and transmission aspects, both analogue and digital, mainly for videotelephone and videoteleconferencing services. He is head of a group responsible for low bit-rate video coding standards and was a UK representative in the CCITT Specialists Group which formulated Recommendation H.261. His current interests include coding for storage devices and for asynchronous transfer mode networks.

Paul McDonald (Section 9.4) became a British Telecom sponsored student while studying electronic and microprocessing engineering at Strathclyde University. After graduating with a Bachelor of Engineering degree in 1987 he then joined the British Telecom Research Laboratories to work in the Digital Networks Division. He has been involved in systems and network studies for wideband network interfaces and his work has encompassed research into narrowband ISDN interfaces and standards, SDH and SONET transmission standards, ATM techniques, and broadband ISDN issues. He is currently working on the problems of interworking mobile networks with the ISDN.

Chapter 1

Network Evolution

1.1 THE TELEGRAPH

The earliest telecommunications systems were optical. Although the use of beacon fires to warn of anticipated dangers is widely known, it is difficult to regard as telecommunication a system which only allowed the transmission of one bit of information over a period of several weeks. There were several attempts at more flexible systems using large shutters or movable arms which could be mounted on tops of hills (Figure 1.1). These were observed from adjacent hills and relayed on. The system which came into virtual universal use was the French semaphore system invented in 1792. By 1830 much of Europe was covered by such networks.

In 1837 the first practical electrical telegraph appeared. In Europe Cooke and Wheatstone produced their five-wire telegraph (Figure 1.2). In the same year Samuel Morse invented his well-known system, although it was another seven years before it was put into service. For the next 40 years telegraphy evolved technically and in its services. The five wires of the Wheatstone system were complex and expensive but the ability of unskilled operators to use it made it in some situations more attractive than the morse system. The ideal was a system which only used two wires (or one and an earth return) and was simple to operate. The eventual result of this was the teletypewriter developed in 1915 by Kleinschmidt and Morkrum in the USA.

In functional terms by 1875 the telegraph network was simple with European countries each offering a state monopoly telegram service. In the USA, while not an actual monopoly, the telegraph network was dominated by Western Union. Hence a resident of that time could have considered that he or she had an Integrated Services Digital Network (ISDN), there only being one network for all the services available at that time and this being digital in nature.

1.2 THE TELEPHONE ARRIVES

The coming of the telephone in 1875 removed both the integration and digitalization of the network. Bell's early attempts at telephony involved



Figure 1.1 The semaphore.

operating a make-break contact by a microphone diaphragm in a binary way, but his early success relied on the change of magnetic circuit impedance and hence magnetic field in a coil due to movements of a diaphragm. However this lacked sensitivity. Elisha Gray made real progress when he made the contact softer using a diaphragm connected to a wire immersed in a fluid giving a resistance change proportional to the audio pressure change (Figure 1.3). Edison's carbon microphone invented in 1878 was a more convenient way of achieving the same effect.



Figure 1.2 Wheatstone's telegraph.

The scene was therefore set for two clearly separate telecommunications networks. On the one hand, the telephony network carrying signals whose magnitude was fairly linearly related to the sound pressure incident upon the sending microphone and were hence called 'analogue'. Its big advantage is that it can be used by unskilled operatives; communication simply involved speaking and listening and what can be more natural than





Elisha Gray. Instruments for transmitting and receiving vocal sounds telegraphically. Caveat filed 14 February 1876.



Figure 1.4 Telegraph repeater 1901. that? On the other hand, the digital telegraph network had the advantage that a hard copy of the communication is produced. In technical terms the telegraph network had further advantages; at a very early stage a means of relaying (Figure 1.4), or repeating signals was developed such that operation over long distances was possible. This has always presented a problem for telephony and was initially solved by using wires of extraordinary thickness. The invention of the triode by Lee de Forest in 1906 meant that attenuation of telephony signal could be compensated for, but any noise or non-linear distortion once acquired by the signal was there for good. Even the use of amplification presented problems in that, if it was not carefully controlled, systems could become unstable and oscillate. In a telegraph network one merely had to identify whether a signal was present or not.

Thus the science of telephony became one of balancing the economies to be gained by concentrating traffic into a small number of large switching centres, against the need to maintain an acceptable speech quality by having switching centres close to the customer and hence numerous.

The convenience of the telephone meant that its numbers grew very rapidly and, because of the pressure of demand, technical problems were at least made manageable if not overcome. For transmission purposes the most fundamental change was to make the trunk network into a four-wire system so that the two directions of transmission were separated and could be amplified without the compromises required for full duplex operation on one pair of wires. The other innovation in the first decade of the twentieth century was the introduction of automatic switching. Although this increased the complexity of operations required of the user, it meant that the person power and accommodation required to provide switching could be kept within reasonable bounds in an ever-expanding network.

1.3 TELEX

The expansion of the telephone network was such that in the early 1900s the public telegraph ceased to be the dominant form of telecommunication. On the other hand it provided an essential service to those without access to telephones or where a written record was important. The need for the written record and the development of the page teleprinter in 1928 led to the evolution of the telex service in which the telegraph was extended to customers' premises. Initially this was offered over the existing telephone network using what would now be called moderns which, at the customer premises, converted the d.c. teleprinter signals into modulation of a 1500 Hz signal for transmission across the telephone network. In practice this presented problems as connections which were entirely satisfactory for telephony purposes often gave an excessive error rate for telex purposes. For this reason all countries eventually followed the lead of Germany, which had in 1936 set up a separate switched telex network directly carrying the d.c. teleprinter signals. An additional advantage is that between the switching nodes 24 telex channels could be carried on one 4 kHz telephony bearer. The narrow bandwidth requirement meant that it was easier to extend telex networks overseas. Even though the telex network was intended to create a simultaneous two-way connection, it was also relatively easy to interface it to a store and forward facility where traffic or (in the case of HF radio) circuit conditions dictated. This further extended the attraction of telex to the business community.

Thus by the 1950s in most countries there were three major types of public networks.

- (a) The public telegraph network, by then in decline but with 30 years still to go.
- (b) The telephony network, by far the most popular owing to its ease of use, based on a nominal 4 kHz bandwidth and using analogue techniques, which in the customers access link had not changed in transmission principle since 1880, and in signalling principle since about 1910.
- (c) The 50 baud telex network, proving indispensable to the commercial world because of its provision of a written record and the ability to work unattended.

In the trunk network telephone signals were frequency division multiplexed to carry 12 channels (a 'Group') on a single pair in the range 60– 108 kHz. These groups were then assembled into supergroups of five Groups for transmission on coaxial cables. These supergroups could then be further assembled into master groups. Switching was by electromechanical switches, either a developed form of Strowger's original step-by-step switch of 1898 or motor driven switches. The crossbar switch of 1938 is a more refined way of connecting lines and its principles (but not its mechanism) carry through to the present day. Intercontinental connections were largely by HF radio. In 1956 North America and the UK were first linked by telephone cable, and in 1965 by geostationary satellite. In the 1960s there was also a general extension of the use of microwave links as an alternative to coaxial cable and the sealed reed relay also became an attractive alternative to the open contact of previous generations of switch.

1.4 THE COMPUTER AND THE TRANSISTOR

In the 1950s the next two significant players came on the scene. First the concept of the electronic computer; at this stage it was really only the

concept which was significant, because the implementation using vacuum tubes, while an advance on previous electromechanical techniques, would not survive long. However, the concept of being able to store and manipulate large amounts of data would surely also lead to the need to transfer a lot of data from place to place.

Secondly the invention of the transistor in 1948 meant that computers of then unimagined storage capacity and processing power could be realized in a small size and consuming little power. These would be able to absorb enormous amounts of data communications capacity. The computer could also be used to control telephone exchanges.

Although telex networks might have been the obvious candidates for upgrading to higher data rates for transferring computer data, their roots lay in electromechanical technology and very little was done in this direction; instead the telephone network was pressed into service. As in the early telex days, modems were designed which could take the baseband data signals and modulate them upon audio carriers to be transmitted across the telephony network. The first modems were available in the early 1960s and operated at 100 baud. Owing to improved modulation schemes, and improvements to the telephone network itself, the error performance of the modem was acceptable and the problems of the early telex modems did not recur in general although particular circuits gave difficulty. Over the years modem performance has been steadily improved until by the end of the 1980s, 9.6 kbit/s can be carried over a typical switched telephony circuit. However, this progress cannot go on indefinitely. Shannon's famous theorem gives an upper bound to the information capacity of a channel, based on its bandwidth and signal-to-noise ratio.

$$I = F \log_2 \left(1 + \frac{S}{N}\right)$$

where F is the bandwidth available and S/N is the signal-to-noise ratio. For a telephony channel the useful bandwidth is about 3 kHz and the signal-tonoise ratio may be about 30 dB. Hence the capacity is of the order of 30 kbit/s. Experience has shown that to get within a factor of 2 or 3 of Shannon's limit is a notable achievement and little more can be expected in that direction.

1.5 DATA NETWORKS

Even at these data rates the telephony network leaves much to be desired for intercomputer communications. In particular call set up on electromechanical systems takes several seconds, and with computer processors operating at millions of operations a second, the time to set up a call is often embarrassing, even when a processor is able to delegate such tasks to

peripheral buffers. For this reason in the 1970s consideration was given to further specialized networks for intercomputer communication. These fell into two categories. The traditional solution was a circuit switched network similar in principle to the telex network, but able to operate at kbit/s rather than bit/s, and with fast call set-up. The other solution was more revolutionary and considered more deeply the needs of the computer community. The concept could be compared with the postal service in which a packet is dropped into a mailbox. This packet could be a postcard or a small parcel and is routed to its destination on the basis of the address on the outside. In telecommunications packet networks, data is preceded by a header containing an address by which the packet is routed to its destination. This is discussed in more detail in Chapter 7. Although this is a very flexible solution, it does require a lot of high speed intelligence applied to each packet at each switching node which has its cost and throughput penalties. The consensus is that where capacity is of greatest consequence then circuit-switched networks are best. Where flexibility is important packet-switched networks are best.

Both packet- and circuit-switched networks have been implemented. Early examples of packet networks are Spain (Iberpac), UK (PSS), France (Transpac), Germany (Datex P), Switzerland (Telepac), USA (GTE Telenet, Tymnet) and Canada (Teleglobe). Circuit-switched networks were established in the Nordic countries (Norway, Sweden, Denmark and Finland) and the Federal Republic of Germany.

Most developed countries now have such data networks. In the case of circuit-switched services they will offer data rates typically of 9.6 kbit/s, with call set-up times of a few hundred milliseconds. In the case of packet networks, connections may be made at up to 48 kbit/s but throughput is limited by the flow control mechanisms in the network, depending on network or destination capacity.

By the end of the 1980s the telegraph networks had virtually ceased to exist. In some countries the telegram service was no more; in others it made use of the telex network. Telex networks were booming but probably at their peak, coming under pressure from facsimile services operating over the telephone network and also (in some countries) teletex services operating on public data networks. The public switched data networks were expanding but with competition from the omnipresence of the telephone network and its modems.

1.6 DIGITAL TELEPHONY

For the next stage in the proceedings one must go back to 1937 when Alec Reeves conceived pulse code modulation (PCM) of speech signals, specifically to overcome the aforementioned problem of the accumulation of distortion and noise in analogue networks. The process is given in detail in the next chapter; in outline it consists of measuring the analogue speech signals 8000 times a second, encoding them into one of 256 levels represented by 8 binary bits, and hence converting the analogue speech into a binary stream at 64 kbit/s. The advantage of this is that if the signal is regenerated before noise is sufficient to make the two binary states difficult to distinguish, then there is no accumulation of distortion. When first invented no devices were available which could make PCM economic. Wide-scale use of PCM for telephony transmission did not occur until the late 1960s. The individual 64 kbit/s streams were multiplexed by interleaving them to carry 24 (North America) or 30 (Europe) channels on pair-type cables. Primary rate multiplexes could then be further interleaved to carry more channels on coaxial cables or microwave radio systems.

In the 1970s it was recognized that Very Large Scale Integrated Circuits (VLSI) and optical fibre transmission would make PCM the obvious choice for future telephony networks, for transmission and switching purposes. The developments of the 1970s came to fruition in the 1980s, with all major network operators replacing their networks and switching nodes by digital facilities. The switches were entirely electronic with computers (usually called processors) to control them, and interprocessor links (called common channel signalling) to control call set-up. Note, however, that the digital PCM operation does not extend to include the local access loop to the customer, which remained the analogue two-wire link with a telephone whose transmission antecedents go back to Edison, with a pulse signalling system introduced in about 1910 or possibly a multifrequency keypad signalling system first used in the USA in 1960. Nevertheless, conversion of the telephony main network to digital has enormous implications. It is now an infrastructure of universal extent which not only can carry telephony, but can carry data at 64 kbit/s. As data and telephony can now be integrated on the one network it is called an Integrated Digital Network or IDN.

1.7 THE ISDN

The next stage is to take the digital channels all the way to the customer so that all services can be integrated on one bearer. This is the Integrated Services Digital Network or ISDN. CCITT defined the ISDN as:

A network evolved from the telephony IDN that provides end-to-end digital connectivity to support a wide range of services, including voice and non-voice services, to which users have access by a limited set of standard multipurpose customer interfaces. Two challenges faced the ISDN proponents.

- 1. To develop a system which would carry the 64 kbit/s digital channels to the customer. The local network mainly consists of copper pairs to provide the final link to the customer which are designed primarily to carry signals up to 4 kHz. The 4 kHz limit is not absolute (except in some North American situations where loading coils were inserted), but attenuation of the pair and crosstalk from other pairs increases as frequency increases. One solution is to substitute the copper pair by an optical fibre. However even though fibres are comparable in price to copper pairs, the cost involved in the changeover is prohibitive unless the demand is very significant and this is difficult to substantiate in early days. Copper pairs also have the advantage of being able to carry power for a telephony terminal, which is desirable for reliability and emergency situations. Chapter 4 considers the use of copper pairs as the ISDN bearer.
- 2. The interface to the customer should be appropriate for a multiplicity of services. Data services are normally terminated in a multipin socket into which a single device can be plugged. Several telephones may be connected in parallel across the incoming analogue telephone line. What is needed is a system which could allow several terminals of different types (e.g. VDU, telephone, personal computer, facsimile machine) to be simply connected to the network, and a signalling system which allowed the terminals to be appropriately called and make calls. It was also decided that with this multiplicity of services it was unreasonable to restrict customers to the use of only one channel at a time. For this reason a basic access offering two 64 kbit/s channels is standardized upon, allowing the use of two terminals at a time. These channels are called 'B' channels. A 16 kbit/s signalling channel is also offered, which may also be used to provide access to a packet-switched service. This channel is called a 'D' channel. The interface identified is a four-wire bus structure to which up to eight terminals can be connected in parallel, whose operation is described in detail in Chapter 5.

The ISDN is a complex mix of network capability and customers premises equipment supporting applications which can exploit the capability. It is the standard exchange line of the future.

1.8 EARLY ISDNs

Although discussions on the ISDN had been in progress throughout the 1970s, it was not until 1984 that CCITT published recommendations for the

standards to be adopted for ISDN services and interfaces. Prior to that several administrations were in a position where they were installing an extensive network of digital exchanges which could be used for ISDN service. Rather than delay the exploitation of the network they introduced services using proprietary standards.

The first service was introduced in June 1985 by British Telecom. Marketed as IDA (Integrated Digital Access) pilot service, it was initially based on one exchange in London. By the end of 1985 it was extended to four exchanges (two in London and one in Birmingham and Manchester) with remote multiplexers as shown in Figure 1.5. By 1988 it was available in 60 towns throughout Britain. It offers two simultaneous traffic channels: one at 64 kbit/s and the other at 8 kbit/s. The original terminal apparatus included digital telephones with data ports and a termination which could offer a whole series of standard data ports. However, because of the requirements of the regulatory regime in the UK, only a standard X.21 interface was offered from 1986.

Another early pre-standard ISDN service was project Victoria, offered by Pacific Bell in 1986 in Danville, California. This offers to a customer five RS232C ports which can be configured to operate at various rates from 9.6 kbit/s down to 50 bit/s, together with an analogue telephone. The whole is multiplexed on to an 80 kbit/s line rate.



Figure 1.5 Pilot ISDN coverage in the UK 1986.

In France since 1985 the Transcom service offers 64 kbit/s interconnect over the digital telephony network using X.21 and V.35 interfaces to the customer. Also in 1986 in the USA the local 'Public Switched Digital Service' combined with the nationwide Accunet 56 service has offered a similar service, but restricted to 56 kbit/s owing to the continued presence of circuits from which bits are stolen for signalling purposes. This will be mentioned again in Section 3.1.

The first CCITT ISDN recommendations formally appeared in 1984 in their 'red book' and systems more or less to these standards have been put into service in Oak Brook, Illinois, USA at the end of 1986, in Mannheim and Stuttgart, West Germany in 1987, and also in Brittany, France in 1987. By 1988 about forty different trials or services to Red Book standards were in progress.

However the next generation of CCITT recommendations issued in their 1988 'blue book' do contain significant changes and clarifications which will lead to truly international standards.

1.9 INTERNATIONAL STANDARDS

The cost of developing modern equipment is very high. The design and verification of a single VLSI chip costs millions of pounds or dollars. however once designed, the unit production costs are low. For software the situation is even more extreme; development costs are very high but replication costs are almost zero. For this reason it is important that any product gets the widest possible use to spread development costs. Worldwide standards against which all telecommunications and users may purchase are therefore very important. This also eases interworking between different networks. In telecoms the main standards are developed and issued by the CCITT (the French abbreviation for the International Telephone and Telegraph Consultative Committee). The recommendations are issued in several series. These are the ones which are of greatest relevance to ISDN:

- V series for data/communications over the switched telephone network.
- X series for dedicated data communications networks.
- T series for terminal equipment and protocols for telematic services.
- Q series for signalling systems.
- G series for telecommunication transmission systems.
- E series for operations, numbering and routing.
- I series for general ISDN matters.



Figure 1.6 ISDN reference configuration.

Most of this book will refer to the I series of recommendations and these are listed in Table 1.1. Where there is an alternative number in brackets this means that the I Series recommendation refers out to a recommendation in another series. A particular feature of the ISDN is the establishment of a reference configuration for the ISDN connection between the customer and the exchange. This is shown in diagrammatic form in Figure 1.6. TE1 is a terminal which complies with ISDN User Network Interface recommendations such as a digital telephone, data terminal equipment or an integrated workstation. TE2 is also a terminal which complies with some other interface standards (e.g. V or X series) at reference point R, and hence needs a terminal adapter (TA) to adapt to the ISDN User/Network Interface. NT2 is some sort of customer premises switching function such as a PABX or LAN, or it may be a null function if it is not required. Reference points S and T apply to physical interfaces which are defined in the I series of Recommendations. NT1 is on the customers premises and provides the proper physical and electromagnetic termination of the network. The transmission line (or U reference point) is terminated by the NT1. Chapter 4 considers the interface at the U reference point and Chapters 5 and 6 consider the interface at the S and T reference point.

1.10 OPEN SYSTEMS INTERCONNECT (OSI); THE 7-LAYER MODEL

OSI is a term for the agreed International Standards by which systems should communicate. In telecommunications, as in most other activities, it takes teamwork to provide and utilize the facilities offered. Most basically transmission media such as optical fibres or copper pairs must be provided, in addition signalling and speech coding must also be provided, and all this will be wasted if there are no customers who wish to talk to one another. Thus there is a natural layering of the telecommunications process. The International Standards Organization (ISO) has formalized this into seven layers for the interworking of computers, terminals and applications. In the 7-layer model it is assumed that one has a physical connection such as an optical fibre, copper pair or coaxial line. Upon this is built:

Layer 1 — the physical layer defines the characteristics of the signal to be transferred over the bearer. It covers such things as pulse amplitudes, line coding, transmission rates, connectors, and anything else needed to transfer digits satisfactorily.

Layer 2 — the link layer provides discipline for the assembling of the digits. It provides error detection and corrrection by assembling the digits into frames. At present all Layer 2 formats are derived from a standard known as High Level Data Link Control (HDLC).

Layer 3 — the network layer ensures that messages are routed to the appropriate destinations, and also provides mechanisms to ensure the appropriate control and acknowledgement of messages.

Layer 4 — the transport layer. This is the terminal-to-terminal layer. Data may be carried across the networks using various forms of Layers 1, 2 and 3 (e.g. via a LAN and ISDN) but the terminals must have information at appropriate rates.

Layer 5 — the session layer. This defines the way in which applications running at the two ends of the link intercommunicate, including initiation and termination of sessions and co-ordinating their activity during the session.

Layer 6 — the presentation layer. This establishes the common format which is to be used between terminals, using common rules for representing data.

Layer 7 — *the application layer*. This is the task to be performed, for example file transfer, airline booking, message handling.

The lowest three layers are called the network service layers. In the ISDN there are two types of channel. First the 64 kbit/s 'B' circuit-switched channel for customer use. For straightforward data purposes in simple terms the ISDN only defines the Layer 1 attributes of this channel, the customer being free to use the bits provided for his own higher layer protocols. The 16 kbit/s 'D' signalling channel includes all three layers and these are discussed in Chapter 5.

The usual diagram for describing these sets of protocols is shown in Figure 1.7. The layers are stacked vertically. Two stacks are shown, the control 'plane' for signalling and the user 'plane'. The lines passing vertically through the centre of each stack show the progress of the relevant channels. The user information line carries applications (Layer 7) from terminal to terminal, only changing its physical form (Layer 1) at





intermediate points such as the NT1 or local exchange. However the control information in the form of signalling has only three defined layers. As will be seen in Chapter 5, Layers 2 and 3 of the protocol are terminated in the exchange. This is much simplified but illustrates the type of diagram used to illustrate message flows. A further complication arises from the fact that the 7-layered model is in no way absolute and sub-layers can often be identified. Hence diagrams with dotted intermediate levels are often encountered. Similarly other 'planes' are possible.

Table 1.1 CCITT Recommendations on ISDN

Rec. No	
1.110	Preamble and general structure of the I-Series Recommendation for the integrated services digital network (ISDN).
I.111	Relationship with other Recommendations relevant to ISDNs.
1.112	Vocabulary of terms for ISDNs.
L113	Vocabulary of terms for broadband aspects of ISDN.
1.120	Integrated services digital networks (ISDNs).
1.121	Broadband aspects of ISDN.
1.122	Framework for providing frame mode bearer services.
1.130	Method for the characterization of telecommunication services supported by an ISDN and network capabilities of an ISDN.
I.140	Attribute technique for the characterization of telecommunication services supported by an ISDN and network capabilities of an ISDN.
1.141	ISDN network charging capabilities attributes.
1.200	Guidance to the I.200-Series of Recommendations.

Table	1.1	(continued)	
CCITT	Rec	ommendations on	ISDN

Rec. No	
1.210	Principles of telecommunication services supported by an ISDN and the means to describe them.
1.220	Common dynamic description of basic telecommunication services.
1.221	Common specific characteristics of services.
1.230	Definition of bearer service categories.
1.231	Circuit-mode bearer service categories.
1.232	Frame-mode bearer services categories.
1.240	Definition of teleservices.
1.241	Teleservices supported by an ISDN.
1.250	Definition of supplementary services.
1.251	Number identification supplementary services.
1.252	Call offering supplementary services.
l.253	Call completion supplementary services.
1.254	Multiparty supplementary services.
1.255	Community of interest supplementary services.
1.256	Charging supplementary services
1.257	Additional information transfer.
1.310	ISDN—Network functional principles
1.320	ISDN protocol reference model
1.324	ISDN network architecture
1.325	Reference configurations for ISDN connection types
1.326	Reference configuration for relative network resource requirements
1.330	ISDN numbering and addressing principles
1.331(E.164)	Numbering plan for the ISDN area
1.332	Numbering principles for interworking between ISDNs and dedica-
	ted networks with different numbering plans
1.333	Terminal selection in ISDN
1.334	Principles relating ISDN numbers/subaddresses to the OSI refe-
	rence model network layer addresses
1.335	ISDN routing principles
1.340	ISDN connection types
1.350	General aspects of quality of service and network performance in digital networks, including ISDN.
I.351(G.821/2)	Recommendations in other series concerning network performance objectives that apply at reference point T of an ISDN
1.352	Network performance objectives for connection processing delays in an ISDN.
1.410	General aspects and principles relating to Recommendations on ISDN user-network interfaces.
I.411	ISDN user-network interfaces-reference configurations
1.412	ISDN user-network interface-interface structures and access capa- bilities.
1.420	Basic user-network interface.
1.421	Primary rate user-network interface

Table	1.1 (continued)
CCITT	Recommendations on ISDN

Rec. No	
1.430	Basic user-network interface-Layer 1 specification.
1.431	Primary rate user-network interface—Layer 1 specification.
1.440(Q.920)	ISDN user-network Interface data link layer—general aspects.
l.441(Q.921)	ISDN user-network interface, data link layer specification.
I.450(Q.930)	ISDN user-network interface Layer 3—general aspects.
1.451(Q.931)	ISDN user-network interface Layer 3 specification for basic call control.
l.452(Q.932)	Generic procedures for the control of ISDN supplementary services.
I. 460	Multiplexing, rate adaption and support of existing interfaces.
I.461(X.30)	Support of X.21, X.21 bis and X.20 bis based data terminal equip-
	ments (DTES) by an integrated services digital network (ISDN).
I.462(X.31)	Support of packet mode terminal equipment by an ISDN.
I.463(V.110)	Support of data terminal equipments (DTES) with V-series type
	Interface by an integrated services digital network (ISDN).
I.464	Multiplexing, rate adaption and support of existing interfaces for restricted 64 kbit/s transfer capability.
L465(V.120)	Support by an ISDN of data terminal equipment with V-series type
	interfaces with provision for statistical multiplexing.
1.470	Relationship of terminal functions to ISDN.
1.500	General structure of the ISDN interworking Recommendations.
1.510	Definitions and general principles for ISDN interworking.
1.511	ISDN-to-ISDN Layer 1 internetwork interface.
1.515	Parameter exchange for ISDN interworking.
1.520	General arrangements for network interworking between ISDNs.
1.530	Network interworking between an ISDN and a public switched
	telephone network (PSTN).
I.540(X.321)	General arrangements for Interworking between circuit switched public data networks (CSPDNs) and integrated services digital networks (ISDNs) for the provision of data transmission.
I.550(X.325)	General arrangements for interworking between packet switched public data networks (PSPDNs) and integrated services digital networks (ISDNs) for the provision of data transmission.
l.560(U.202)	Requirements to be met in providing the telex service within the ISDN.
I.601	General maintenance principles of ISDN subscriber access and subscriber installation.
1.602	Application of maintenance principles to ISDN subscriber installa- tions.
1.603	Application of maintenance principles to ISDN basic accesses.
1.604	Application of maintenance principles to ISDN primary rate accesses.
1.605	Application of maintenance principles to static multiplexed ISDN basic accesses.

NETWORK EVOLUTION

REFERENCES

18

The CCITT recommendations are published by the International Telecommunication Union, Geneva.

The 7-layer OSI model is described in the International Standards Organization (ISO) Standard 7498.

Chapter 2 Pulse Code Modulation

The International Standard ISDN is based upon 64 kbit/s circuit switched channels and transmitted through a network which, for the foreseeable future, will predominantly be used for telephony purposes. Thus, although the great opportunities offered by ISDN may be in the fields of data and information technology, it is appropriate to consider the *raison d'être* of the network on which it is founded. The 64 kbit/s channel is regarded as virtually fundamental, but in some ways is a historical accident. As mentioned in the previous chapter, PCM was conceived by A. H. Reeves but, apart from some military applications in the Second World War, the first application was developed in the USA in the early 1960s for increasing the capacity of existing copper pairs between switching nodes. A diagram of the modulation process involved is shown in Figure 2.1.



Figure 2.1 PCM coding.

2.1 SAMPLING

The analogue signal is band limited to 3.4 kHz and sampled 8000 times per second. This sampling rate arises from the process which gives rise to sidebands around frequencies which are integer multiples of 8 kHz as

shown in the diagram. If the sampling rate is less than twice the bandwidth of the analogue signal, then the sidebands will overlap (known as 'aliasing') and subsequent decoding is not possible. This is Nyquist's sampling theorem. Note that there is another strong reason for sampling at a multiple of 4 kHz, owing to the fact that at the time of introduction of the PCM systems there existed frequency division multiplex (FDM) systems in the network which carry telephony on sinusoidal carriers at multiples of 4 kHz. Inevitably there is some leakage of the PCM sampling waveforms into the FDM systems which could give rise to whistles unless they were multiples of 4 kHz, in which case the beats would appear as very low frequencies and be inaudible.

2.2 CODING

The amplitude of the samples is then measured and encoded into a binary number. Obviously the precision with which this encoding process can take place depends on the number of bits in the binary number. Any lack of precision is known as quantizing distortion. Empirically it was found that to give adequate speech quality 12 bits were needed in this encoding process. That is to say ± 2048 levels could be identified. There is an economy that can be made in these bits. With the straightforward 12-bit encoding the distortion would be the same whether the speaker was whispering or shouting. However, at higher speech levels much greater quantizing distortion can be tolerated as it is masked by the speech itself.

2.3 COMPRESSION

To exploit this masking a non-linear compression function is introduced to translate the 12-bit code into an 8-bit code. The ideal form of compression is logarithmic as this characteristic leads to a constant signal to quantization distortion ratio. This is shown in Figure 2.2 which shows diagrammatically how the uncompressed code could be mapped on to a compressed code for both polarities of signal. The problem arises for very low level signals as the logarithm of very small numbers becomes negative. Somehow the compression law has to be forced to pass through the origin. Two solutions were found. In North America the two logarithmic curves were displaced towards the central vertical axis giving a transfer function of the form $y \propto \log (1 + \mu x)$. This is known as the 'mu' law. In Europe a line was drawn tangentially to the two curves and hence, by symmetry, it passes through the origin. Thus the curve is of the form $y \propto Ax$ over the central part of the





range and $y \propto 1 + \log (Ax)$ at the extremes of the range. This is known as the 'A' law.

To simplify further the translation process, the actual curves are approximated by linear segments each of half the slope of the previous. Although the theoretical derivation of the compression laws has been described, in



Figure 2.3 μ and A law companding.

practical terms the standard exists in the form of a tabulation for the conversion in CCITT Recommendation G.711. These are plotted in Figure 2.3. This reveals the exaggerations of Figure 2.2. In practice the two curves are very similar. Note that a symmetrical code is used so that the first bit indicates the polarity of the sample and subsequent bits indicate the magnitude of the sample. Additionally, in the European A-law case alternate bits in the code word are inverted; this is because in practice channels are frequently idle and, without the inversions, the idle signal would contain a lot of zeros which may make derivation of timing information difficult in line transmission systems. Similarly the North American system inverts all binary digits which increase the density of 1s for transmission purposes. The decoding process is the inverse of the encoding process and is again summarized in Figure 2.4.



Figure 2.4 PCM decoding.

2.4 64 KBIT/S CHANNELS

The process described above based on 8-bit coding of samples at 8 kHz gives rise to the 64 kbit/s fundamental channel rate. Earlier in the chapter I referred to it as an historical accident. This is for two reasons:

- (a) The A-law and the mu-law PCM standards were chosen so that the transmission systems could be embedded anywhere in an analogue network. Owing to the need to ensure that analogue networks are stable, it is not possible to compensate for all the network attenuation and hence levels will vary by some 15 dB according to the location in the network. In an ISDN, speech encoding and decoding are always performed at the customers' terminal and the only level variation is due to speaker differences. Without the variations in level due to the network, only about 6 bits are required for PCM coding to give an adequate signal to quantization distortion. In addition the system is designed to permit several conversions from analogue to digital and back again with acceptable quantizing distortion.
- (b) There are more efficient ways of encoding speech than PCM. With human speech, samples are not totally independent of one another. In
fact there is considerable correlation between successive samples and encoders have been developed which exploit this fact. Previous samples can be used to predict the next sample, and then only the errors in the prediction have to be transmitted. Section 8.1 describes how this may be exploited.

For these reasons a digital telephony network could be constructed which offered telephony quality at 32 kbit/s or even lower. Nevertheless, the 64 kbit/s channel is likely to remain the fundamental for the terrestrial ISDN. The reasons are:

- (a) Optical fibres offer large bandwidths.
- (b) Switching and multiplexing costs tend to be related to the number of channels rather than the bit rate.
- (c) VLSI components have been developed for use at 64 kbit/s.
- (d) 64 kbit/s gives a degree of headroom above modem data rates which should suffice for a reasonable period, and more sophisticated encoding techniques can be used to give higher quality speech.
- (e) Avoiding interworking problems between parts of the network operating at different rates. The one standard channel size means that clear channel working between terminals can easily be guaranteed. Mobile systems operating within the bandwidth constraints of free space will, however, use the low bit-rate coding.

REFERENCE

CCITT Recommendation G.711. Pulse Code Modulation (PCM) of voice frequencies.

Chapter 3

The Integrated Digital Network

The Integrated Digital Network consists of local switching nodes (usually called exchanges in Europe and central offices in North America) joined by digital transmission links carrying 64 kbit/s digital channels multiplexed on to a copper pair, coaxial pair, microwave or optical fibre bearer. The switching nodes are controlled by processors which intercommunicate through signalling links. In this chapter I will cover in outline these matters, but only in so far as they affect the ISDN.

3.1 MULTIPLEXING

The advantage of encoding speech into PCM is that it may be processed as would any other data stream, using digital integrated circuits. The multiplexing process involves taking the PCM blocks of 8 bits which are generated every 125 μ s (i.e. 8000/sec) and interleaving them with blocks from other PCM encoders to give a Time Division Multiplex (TDM) of channels. Of course, in an IDN the 64 kbit/s sources need not be speech encoders but could be other data sources. In the 1970s North America and Europe chose different ways of assembling the multiplex. North America interleaved 24 channels giving an aggregate rate of 1.536 Mbit/s (24 \times 64 kbit/s). At the start of each frame of 24 channels a marker bit F, is included which adds a further 8 kbit/s to give a total rate of 1.544 Mbit/s. This is shown diagrammatically in Figure 3.1. The F bit follows a defined sequence so that it is unlikely to be imitated by data in the traffic channels. The 24-channel PCM systems are widely used for connections between central offices in North America and for this purpose an 'associated channel' signalling system is used where in every 6th frame the least significant bit of the 8 bits in the channel is 'stolen' for signalling relating to that particular channel. Although a convenient system for the original point-to-point purpose this does mean that the channel available is not the full 64 kbit/s for ISDN purposes, and also is not readily adaptable to common channel interprocessor signalling. For these reasons in an ISDN



Figure 3.1

Frame structure of 1.544 Mbit/s interface.



Figure 3.2

Frame structure of 2.048 Mbit/s interface.

environment the 'bit-stealing' process is being abandoned and the 24 channels are available for full 64 kbit/s transmission. Signalling is carried as one of the 64 kbit/s stream leaving 23 traffic channels.

In Europe a different format was adopted. Thirty 64 kbit/s channels were interleaved with a further 64 kbit/s signalling channel and a 64 kbit/s synchronizing channel. This gives an aggregate rate of 2.048 Mbit/s. This is shown diagrammatically in Figure 3.2. Channel 0 is used for the synchronizing (and also some maintenance) channel. Channel 16 is used for signalling. Once again in the original form the signalling channel had fixed bits allocated to particular traffic channels (i.e. channel associated), but in an ISDN environment channel 16 is used for a common channel signalling system.

The 1.544 Mbit/s and 2.048 Mbit/s are known as primary rate multiplexes. Further interleaving of several primary rate multiplexes can take place giving rates of 6, 45 and 274 Mbit/s in North America and 8, 34, 139, and 560 Mbit/s in Europe. As far as the user of the ISDN is concerned, such rates are not visible and will not be discussed further here.

3.2 TRANSMISSION

It was mentioned in Section 1.6 that the earliest use of PCM was to increase the capacity of links on copper pairs between switching nodes. Previously two pairs of copper wires could carry one audiofrequency telephone conversation if the two directions of transmission were separated, or two audiofrequency telephone conversations if the two directions of transmission were on the same pair (but this presents problems with amplification as mentioned in Section 1.2). The two pairs could carry 24 (North America) or 30 (Europe) conversations if primary rate PCM systems were used.

However transmission at 1.544 or 2.048 Mbit/s gives rise to two main problems.

- (a) Attenuation. The loss of typical pairs is of the order of 15 dB/km at the relevant frequencies.
- (b) Crosstalk. There is no screening between the pairs in the cable. They rely on their symmetry and various twists to prevent a signal in one pair from being coupled to another pair.

For these two reasons regeneration of the signal is required approximately every 2 km. In fact the argument was actually taken in reverse. The PCM systems replaced audio systems whose transmission performance was improved by periodically inductively loading the copper pair at intervals of 6000 feet. Hence additional underground capacity and cable access was available at these sites which could also be used for PCM regenerators. Working backwards through the cable characteristics, this defined the transmission rates chosen. Europe chose its higher rate several years after North America and indicates a degree of optimism based on experience of earlier systems.

The simplest way of transmitting binary digits is to represent the 1 and 0 by two voltage levels. For line transmission this simple process is rarely used for two reasons:

- 1. Maintenance of balance. It is convenient to include a high pass filter at the input and output of regenerators so that power may be fed along the same pair and separated from the signal. This high-pass filter is usually a transformer. Transformers also give protection against surges. If long strings of signals of the same polarity are transmitted then after passage through the high-pass filter they will be severely distorted. The simplest way of balancing the signal is to transmit '0' as 0 volts and '1' as alternately a positive and negative signal as shown in Figure 3.3. This is known as 'bipolar', 'pseudo-ternary' or 'alternate mark inversion (AMI)'.
- 2. Timing extraction. The regeneration process always involves retiming the signals and this is done by averaging the transitions over a considerable period and then using this extracted clock to retime subsequent pulses. In early regenerators the AMI signal was simply rectified and applied to a low loss tuned circuit. More recently phase locked loops are used. For this process to work a certain minimum density of transitions is required. The AMI line code does not ensure this in the case of a string of



(note: the version of AMI used on the basic rate ISDN interface uses full width pulses and the binary is negated; see section 5.1.4)

Figure 3.3

AMI with half-width pulses.

0s. For encoded speech this is rarely a problem taking the precautions mentioned in Section 2.3. However the use of channels for data can give rise to such unfortunate patterns and the European systems and modern North American systems modify the basic AMI coding. In Europe the modified coding is called HDB3 (high density bipolar3). The basic AMI code requires that successive 1s are transmitted with alternating polarity. That is to say 10111 would be transmitted at +0-+-. If two 1s of the same polarity were transmitted then this would be a 'violation' of the rules. HDB3 operates by replacing any group of four 0s by a sequence consisting of three 0s followed by a violation of the coding rules (i.e. a symbol of the same polarity as the previous). Thus 11000001 would be sent as +-000-0+ or the inverse. The violation indicates the four 0s. It is possible with this simple rule that successive violations would be of the same polarity and a low frequency content added to the signal. To avoid this the four 0s can be substituted by B00V (B represents a normal bipolar pulse) to ensure that violations alternate in polarity; thus 10001100001 would be coded as +000+-+00-+. Note that the 3 in HDB3 indicates the longest run of zero voltage symbols which can occur in the coded stream. In the North American system the modification to AMI is called B8ZS (Bipolar with 8 zeros substitution). Whenever eight 0s occur they are coded 00B0VB0V. As this code contains two violations separated by an ordinary bipolar pulse it is inherently balanced. Up to seven consecutive '0's can be transmitted.

These coding techniques were developed in the 1960s when VLSI was not available and hence regenerators were simple by modern standards. With more complex technology the above constraints are much less important and in Chapter 4 it will be seen that coding effort is mainly directed to spectrum control, to minimize the effect of crosstalk and attenuation for local network applications. The 1.544 Mbit/s and 2.048 Mbit/s are firmly established and in the future may be more relevant as interface standards between equipment rather than as standards for signals on copper pairs.

3.3 SWITCHING

If one had several terminals connected to individual pairs, which one wished to interconnect, one would arrange to connect the pairs to a switch matrix and close those cross points which allow the appropriate terminals to intercommunicate. Figure 3.4 shows a 6×6 matrix. The crosses indicate the closed switches and indicate that input terminal 1 is connected to output terminal 3, 2 to 4, 3 to 2, and so on. This is known as a space switch. The data from terminals could have been time division multiplexed on to a single pair, shown diagrammatically in Figure 3.5 by the rotating switch on the left. The data channels will now appear as sequential timeslots (each containing 8 bits) on the single pair. To connect to the appropriate terminals on the right-hand side needs a reordering of the channels and this is performed by the switch in the centre. The process involves putting the bits in each timeslot into memory and reading them out in the appropriate order. This function is called a time switch. The demultiplexing switch shown symbolically on the right will now be connected to the appropriate terminals on the left. In a real digital switch both space and time switching functions are required. Figure 3.6 shows a block diagram of a typical digital switch. So that the channels are in a convenient order 24 or 30 channel



Figure 3.4 A space switch.







Figure 3.6 TST digital switch.

digital links are first time-switched individually. A space switch then allows timeslots to be transferred from digital link to digital link. Notice that during each timeslot there is a different configuration of the space switch so that each timeslot may be routed to a different digital link. After the space switching there is a further stage of time switching to reorder the channels as required. The most obvious economy in a time divided digital switch compared with an analogue switch lies in the space switch, which is used in as many different configurations as there are timeslots. In the example of Figure 3.6 the space switch is being used to switch $3 \times 24 = 72$ or $3 \times 30 = 90$ channels instead of the three channels it could switch standing alone as an analogue switch. For this reason in real digital switches there is an additional stage of multiplexing of digital links to increase the number of channels time division multiplexed and offered to the space switch. The switch described has a Time-Space-Time structure (TST). Other structures are possible but the TST now predominates.

3.4 REMOTE UNITS

Figure 3.7 shows the basic components of an exchange. The concentrator is a switching stage which connects the subscribers to the number of internal trunks which are needed to carry the traffic that the customers generate. Typically the number of internal trunks may only be one tenth of the number of customers. The route switch connects the call to an external trunk or back to the concentrator for own exchange calls. The whole is controlled by the processor.

The use of digital transmission means that impairments which were introduced by analogue links no longer exist. Processor control and signalling links mean that the concept of an exchange as a single indivisible piece of equipment is not essential. Instead, if economics support it, pieces of the



Figure 3.7 Exchange outline.

exchange may be separated. For example the processor may control several concentrator and route switches at remote sites. Similarly parts of the concentrator may be separated. One form which is widely used in North America consists of a unit which may be remotely sited from the exchange, possibly in 'street furniture' such as roadside cabinets, and typically has 96 line cards to which customers may be connected. The customers are then multiplexed on to four 1.544 Mbit/s bearers. These may then be regarded as a replacement for the 'front end' of the concentrator stages. Such systems when used for telephony purposes are often known as 'digital loop carrier' systems. They have the advantage of concentrating % customers on to eight copper pairs (four in each direction) or two optical fibres, which is particularly advantageous where cable pairs are scarce and difficult to provide. It is also a means of providing service to a small community, where reach considerations mean that they cannot be served directly. The digital links from remote unit to exchange may be of any length provided appropriate regeneration is provided.

A further use of remote units which is particularly relevant in an ISDN context is to introduce a service in advance of that service being generally available. Two aspects are significant.

- (a) Out of area service. In the early days of the ISDN only some local exchanges may be suitable (e.g. the digital ones). Customers on other exchanges may be served by small remote units until their local exchange is replaced by a suitable exchange. This is how the British Pilot ISDN service (see Figure 1.5) was introduced so widely in 1986.
- (b) Limited demand. In the early days of ISDN the demand will inevitably be small in comparison with the demand for conventional analogue telephony and it may not be economical to go to the complexity of modifications to the whole exchange for the sake of a small number of ISDN connections. In this case a remote unit may provide the ISDN interface to the small number of initial customers, putting off the more extensive exchange changes until demand is more significant. This has a further advantage that when tracking changes in International Standards only a small piece of equipment need be modified, rather than the whole exchange. In this case the 'remote' unit may actually be colocated with its parent exchange.

A further development of the remote is the 'adjunct' switch. This could be used at the front end of an analogue exchange. Any ISDN voice calls are routed to the analogue exchange, but calls requiring 64 kbit/s are routed to a digital network.

3.5 SIGNALLING

All modern switches are controlled by a central processor. These are usually called Stored Program Controlled (SPC) exchanges. Information flows between processors by means of Common Channel Signalling Systems (CCS) and their use is fundamental to the provision of an IDN and hence an ISDN. This title distinguishes them from earlier systems where the signalling was carried along with the speech channel to which it related, which were called Associated Channel Signalling Systems (ACS). The disadvantage of ACS in an SPC environment is that each channel into the exchange (of which there may be several thousand) has to have associated with it some form of processor access. CCS only needs, in principle, as many links as there are processors with which it needs to communicate. In practice these links will be replicated for traffic and reliability reasons.

Early Common Channel Signalling systems were CCITT No. 6 in Europe and the similar North American CCIS (Common Channel Interoffice Signalling). For ISDN purposes these are now being superseded by CCITT No. 7. CCITT No. 7 Signalling system is primarily intended to be carried between processors on a 64 kbit/s bearer. It has a layered structure, similar to, but not the same as, the OSI layered structure described in Section 1.10. The lower levels are called the Message Transfer Part (MTP) whose purpose is to transfer signalling messages over the network in a reliable manner. It has a 3-level hierarchical structure.

Level 1—encompasses the physical signalling data link, which in the digital network consists of a 64 kbit/s timeslot in a digital multiplex. In principle any timeslot can be used but in the European system it is traditional to use timeslot 16; in the North American system timeslot 24 is used.

Level 2—comprises functions which facilitate the reliable transfer of messages over a single signalling link. These functions ensure that messages are delivered in the correct order, without loss or duplication. To each message this level adds 16 check digits whereby the equipment receiving the message can check whether any errors have been introduced. Sequence numbers are also added to identify the current message and to acknowledge those messages which have already been received.

Level 3—ensures the reliable transfer of signalling messages even in the event of network failures. Functions and procedures are included which inform the remote parts of the signalling network of the consequences of failures and reconfigure the routing of messages through the signalling network to overcome them. Carried by the MTP, and hence hierarchically above the MTP, are the user parts. These are the messages which are used to set up, clear down and otherwise control calls. Included in this area is the facility to access nonswitching nodes such as intelligent network databases, and also to support network operations and maintenance activities.

The user part for controlling setting up telephony calls is called the Telephony User Part (TUP). In Britain there is a national variant of this known as National User Part (NUP). For ISDN purposes a new user part has been defined called Integrated Services User Part (ISUP) but as an interim measure in Europe an enhanced version of the TUP, called TUP+, will be used for early ISDN purposes. Ultimately it is planned that ISUP will supersede the earlier systems. Examples of some ISUP messages are:

Initial Address Message (IAM). A message sent in the forward direction to initiate seizure of an outgoing circuit and to transmit number and other information relating to the routing and handling of a call.

Subsequent Address Message (SAM). A message that may be sent in the forward direction after an initial address message, to convey additional called party number information.

Information Request Message (INR). A message sent by an exchange to request information in association with a call.

Information Message (INF). A message sent to convey information in association with a call requested in an INR message.

Address Complete Message (ACM). A message sent in the backward direction indicating that all the address signals required for routing the call to the called party have been received.

Call Progress Message (CPG). A message sent in the backward direction indicating that an event has occurred during call set-up which should be relayed to the calling party.

Answer Message (ANM). A message sent in the backward direction indicating that the call has been answered. This message is used in conjunction with charging information in order to: (1) start metering the charge to the calling customer, and (b) start measurement of call duration for international accounting purposes.

Facility Request Message (FAR). A message sent from an exchange to another exchange to request activation of a facility.

Facility Accepted Message (FAA). A message sent in response to a facility request message indicating that the requested facility has been invoked.

Facility Reject Message (FRJ). A message sent in response to a facility request message to indicate that the facility request has been rejected.

User-to-User Information Message (USR). A message to be used for the transport of user-to-user signalling independent of call-control messages.

Call Modification Request Message (CMR). A message sent in either direction indicating a calling or called party request to modify the characteristics of an established call (for example, change from data to voice).

Call Modification Completed Message (CMC). A message sent in response to a call modification request message indicating that the requested call modification (for example, from voice to data) has been completed.

Call Modification Reject Message (CMRJ). A message sent in response to a call modification request message indicating that the request has been rejected.

Release Message (REL). A message sent in either direction, to indicate that the circuit is being released because of the reason (cause) supplied and is ready to be put into the IDLE state on receipt of the release complete message. In case the call was forwarded or is to be re-routed, the appropriate indicator is carried in the message together with the redirection address and the redirecting address.

Release Complete Message (RLC). A message sent in either direction in response to the receipt of a release message, or if appropriate, to a reset circuit message, when the circuit concerned has been bought into the idle condition.

Signalling will be discussed further in Chapter 5.

REFERENCES

CCITT Recommendations:

G.702. Digital hierarchy bit rates.

G.703. Physical and electrical characteristics of hierarchical digital interfaces.

Q.700-Q.795. Specification of Signalling System No. 7.

American National Standards Institute: T1.102–1987. Digital hierarchy--electrical interfaces.

Bell Communications Research: TR-TSY-000057. Functional criteria for digital loop carrier systems.

Chapter 4

Local Network Digital Transmission

Peter Adams

In Section 1.7 it was highlighted that a fundamental problem of establishing an ISDN was in making use of the existing local network copper pair infrastructure to carry channels at 64 kbit/s. In this chapter we shall consider the problems and how they can be overcome.

4.1 THE LOCAL NETWORK

The local network has a tree structure with cables radiating from the local exchange. The cables branch at permanently connected joints and at crossconnection flexibility points. A typical section of a European local line network is shown in Figure 4.1. It is often referred to in two parts-main and distribution-the dividing line being the primary cross-connection point. The main network consists of pressurized cables and the distribution network of petroleum jelly filled cables each thereby being resistant to the ingress of moisture. A variety of cable sizes and conductor gauges exist; generally the larger size, thinner gauge cables being used in the main network. Figure 4.2 shows an alternative way of providing flexibility by arranging that pairs may appear at several distribution points. This is more common in the USA and has dramatic effects on attenuation and impedance. The planning limits in force when the network was laid out varied. The attenuation limit is variously specified by different administrations as being 6 dB at 800 Hz or 10 dB at 1600 Hz or corresponding figures for intermediate frequencies. More modern exchanges may have limits of 10 dB at 800 Hz and 15 dB at 1600 Hz. The resistance limit depends on the signalling limits for the exchange. This ranges from about 800 ohms for early exchanges to 2000 ohms for modern exchanges. The cumulative distribution of the UK and USA networks is shown in Figure 4.3. Loop resistance is not of primary interest in determining the transmission characteristics of a cable although it does correlate well with cable loss





Local line network with flexibility points.

especially for 0.4 mm (26 AWG) and 0.5 mm (24 AWG) pairs. However, subscriber loops are often required to be line powered and under maximum voltage limitations the loop resistance will affect the amount of power available at the subscriber's end. The changes in planning limits will cause a very gradual upward drift in average loop length which has to be borne in mind in the design of future transmission systems. Thus, in terms of structure and types of cable, the local network is extremely variable and so the electrical characteristics which affect digital transmission also vary dramatically from connection to connection.



Figure 4.2 Local line network with multiple teeing.



Figure 4.3 Cumulative distribution of loop resistances.

4.2 CABLE CHARACTERISTICS

The characteristics of particular interest for the design of digital transmission systems are:

- (a) Attenuation. The variation of attenuation with frequency is plotted in Figure 4.4 for the gauges of conductor found in the local network. At frequencies below about 10 kHz and above about 100 kHz the attenuation varies with the square root of the frequency; in between there is a reduced attenuation slope. The attenuation decreases with increasing conductor gauge.
- (b) Phase. The phase frequency characteristics of the various gauges of cable are shown in Figure 4.5. All the cables exhibit a linear phase characteristic with a constant phase offset above 10 kHz but are nonlinear at lower frequencies.
- (c) Characteristic impedance. The modulus of the characteristic impedance of the various cable types is shown in Figure 4.6. At higher frequencies it approaches a constant resistive value but at lower frequencies varies markedly with frequency and from cable to cable. The phase of the



Figure 4.4 Attenuation of typical cables of various gauges.



Figure 4.5 Phase coefficient of copper pairs.



Figure 4.6

Characteristic impedance of typical cables of various gauges.

characteristic impedance also varies with frequency at low frequencies but at high frequencies approaches zero.

- (d) Crosstalk. Two cases of crosstalk are considered. First where the transmitter and receiver are at the same end of the cable, called Near End crosstalk (NEXT). Secondly where transmitter and receiver are at opposite ends of the cable, called Far End crosstalk (FEXT). The crosstalk characteristics vary markedly between pairs in the same cable and with frequency. For the situation where many systems operate over one multi-pair cable we can average the crosstalk characteristics and then it is found that the Near End crosstalk power sum attenuation follows the familiar $f^{1.5}$ law. For the cables of the UK network it has been found that 98 per cent of all connections have NEXT couplings with a crosstalk power sum attenuation of better than 55 dB at 100 kHz. For FEXT the situation is complicated by the tree structure of the local network but fortunately FEXT is less of a problem than NEXT and can be ignored where NEXT is the limiting factor.
- (e) Impulsive and random noise. Impulsive noise, arising from electromechanical switching transients in adjacent cable pairs and electromagnetic interference from nearby electric traction systems and power lines, is the most serious kind of noise found in the local network. It is very illdefined and is usually characterized by peak amplitude and rate



Figure 4.7

Margin against decision errors (threshold) in mV required for 99 per cent error-free seconds expressed as a cumulative distribution.

occurrence. A few measurements have been made in the UK; the plot in Figure 4.7 shows the results obtained at one local exchange. The plot shows the cumulative percentage of measurements for which a given threshold was not exceeded for 99 per cent for seconds during an approximately 10-minute measurement period. Random interference in the form of thermal noise is not significant in the local network. In certain geographical locations, however, radio interference could be a problem.

4.3 MODULATION

Digital transmission requires the modulation of the digital information on to a suitable electrical signal; the most common way of doing this is to use shaped pulses and to modulate them by the data. Thus a pulse would be sent normally to represent a '0' and would be sent inverted to represent '1'. More complex arrangements are possible and a particular solution is to take two binary digits so that the one binary digit would cause the shaped pulse to be sent erect or inverted: the other binary digit would cause the pulse to be sent normal amplitude or a larger amplitude. Thus two binary digits are encoded as one symbol which can have four states. The process is called pulse amplitude modulation (PAM) and as four states are allowed



Figure 4.8 1200 ohm loop pulse responses.

it is called quaternary. Modulated pulses are sent over the cable and, ideally, are detected without error at the far end. The cable characteristics result in pulse distortion which spreads the pulses ('symbols') and causes Intersymbol Interference (ISI), and crosstalk and noise which cause errors if the signal level is insufficient. The degree to which these factors affect transmission is a function of the line code and/or the pulse shaping used. Pulse shapes which have a spectrum concentrated at low frequencies suffer severe distortion because of the rapidly changing frequency response characteristics; however, they are relatively insensitive to crosstalk. Conversely, pulse shapes with little low frequency energy suffer little distortion but are more susceptible to crosstalk interference. This trade-off is clearly illustrated by Figure 4.8 where the responses of a typical cable to three pulse shapes is shown. The unbalanced pulse of width 2T seconds is severely dispersed but, as the spectral plot in Figure 4.9 shows, the spectrum peaks at zero frequency. The (1, -1) pulse suffers far less dispersion but is attenuated and has more spectral energy at higher frequencies. The WAL 2 pulse is virtually dispersion-free but suffers the highest attenuation and has a much wider spectrum.

The well-behaved frequency response at higher frequencies suggests that on shorter connections where crosstalk is not a problem (because of





the lower line loss) modulation of the digital information on to a sinusoidal carrier can confer some advantages. At frequencies between about 10 kHz and 100 kHz the flatter attenuation response will introduce less distortion especially to a double-sideband amplitude modulated signal. Also, the process of demodulation will automatically remove the phase offset and the remaining linear phase introduces no distortion.

4.4 EQUALIZATION

As we have seen, the judicious use of pulse shaping or linear line coding can give varying degrees of reduction in dispersion and the resulting ISI. However, the more crosstalk tolerant pulse shapes do suffer ISI and, because of the variability of the cable characteristics, the distortion is not easily predicted with precision. It is, therefore, necessary to use some form of adjustable (and preferably automatic) equalization of the cable frequency response. There are two fundamentally different ways of achieving equalization, illustrated in Figure 4.10.

(a) Linear equalizer. Linear equalization can be provided in two ways: by equalizing over the whole bandwidth of the line signal or, after sam-





(b) decision feedback equalisation

Figure 4.10 Types of equalizer.

pling, over the sampled-signal spectrum. The former can be achieved by an adjustable continuous time analogue filter or by an adaptive fractional-tap (FT) discrete time equalizer in which the waveform is sampled at least at twice the highest frequency of the line signal; the latter by a T-spaced adaptive equalizer where the waveform is sampled at the symbol rate. By using an appropriate adaptation algorithm the FT equalizer is also capable of optimally filtering any noise present with the received signal. However, the T-spaced equalizer is easier to implement and, provided a good sampling phase is used, gives almost as good performance. Both forms of linear equalization introduce gain to combat the frequency dependent loss of the cable and so cause noise enhancement. A further disadvantage is that the processing required in the equalizer involves the multiplication of accurately represented quantities.

(b) Decision feedback equalizer (DFE). The DFE has the advantage of not enhancing noise because noise does not enter the filter. Also, because the filter is driven by detected data, if a transversal filter is used the multiplications involved are each between a coarsely quantized value and only one accurately represented value. However, because detected data is fed back, if an error does occur the resulting magnification rather than cancellation of ISI can cause error propagation. The received pulse shapes of Figure 4.8 suggest that, as all the ISI can be made to arise from previously transmitted pulses, a pure DFE can be used to cancel it. However, the large, slowly decaying, response of the unbalanced pulse would mean that a long transversal filter would be required and the error propagation mechanism of the DFE would generate long bursts of errors. The unbalanced pulses, however, require a relatively short filter and the error propagation is far less severe. Effectively, the DFE can be used to advantage provided some compromise linear equalization of the cable is included. Line coding, or pulse shaping, is one way of doing this; alternatively a fixed linear equalizer at the receiver can be used.

4.5 TIMING RECOVERY

The form of timing recovery depends on the type of equalization adopted. If an ideal analogue equalizer is used then the receiver samples optimally at the peak of the resulting impulse response. For the FT linear equalizer any stable sampling phase will suffice as the performance is nearly independent of sampling phase. For the T-spaced linear equalizer a near-optimum sampling phase is given by a timing recovery scheme which achieved Band-Edge Component Maximization (BECM). The pure DFE requires that the pulse response is sampled so that any pre-cursor ISI is negligible.

4.6 DUPLEX TRANSMISSION

Digital subscriber loop transmission systems are required to operate in duplex mode over a single pair of wires. Theoretically, this is possible, as shown in Figure 4.11, by the use of a perfect hybrid balancing network. If Z_{b} is identical to the input impedance of a cable then the transmitted signal is completely removed from the input to the receiver. However, as we have already seen, the characteristic impedance of a cable is a complicated function of frequency and is difficult to match exactly. Also the coupling to line may be by a transformer creating a further impedance variation at low frequencies. Typically, a balance return loss of about 10 dB for low frequency pulse-shaped signals is all that can be achieved reliably with a simple compromise balance impedance. For modulated signals operating at a frequency where the characteristic impedance is nearly resistive typically a 20 dB balance return loss may be reliably obtained. Over all but the shortest cables the locally transmitted signal will seriously interfere with the received signal and so methods are required to remove the local echo (and any others from impedance mismatches on a mixed cable gauge





connection) from the received signal. There are three techniques available (Figure 4.12): frequency division, time division and echo cancellation.

- (a) Frequency Division Duplex (FDD). In FDD the signals for the two directions of transmission occupy separate frequency bands and are separated by appropriate filters. The technique is very simple and inherently provides good tolerance to NEXT. However, the tolerance to FEXT and noise in general is degraded because the upper frequency channel suffers more attenuation. In addition, very high performance band-separation filters are required. These disadvantages have led to the exclusion of FDD as a candidate for digital subscriber loops.
- (b) Time Division Duplex (TDD). Also known as 'burst mode' and 'pingpong' transmission, this technique separates the two directions by transmitting and receiving in different timeslots as illustrated in Figure 4.12. The transmitted signal can be removed from the receive signal path by gating it out with an appropriately chosen timing waveform. As the received signal is not continuous the receiver timing recovery circuit has to either reacquire timing rapidly on each new burst or, exploiting the periodicity of the bursts, be heavily damped so that a 'flywheel' operation is possible. TDD is very simple to realize; no complex filtering is required and by using a well-balanced pulse shape

no adaptive equalization is necessary. However, in order to accommodate duplex transmission, the bit rate during a burst is at least twice the required continuous rate. Therefore, the tolerance to crosstalk and noise is degraded. It is argued that by synchronizing the bursts at the local exchange end, the effects of NEXT can be avoided but it is doubtful whether such a constraint is acceptable from an operational point of view and the degradation in tolerance to noise and FEXT still remains. Usually the burst bit rate is significantly greater than twice the continuous bit rate because of the effect of the propagation delay of the cable. If the delay is t_g , the continuous bit rate B, the burst bit rate B_b and there are M bits per burst then B_b must be greater than 2 $MB/(M - 2Bt_g)$. If there are additional bits included in the burst for synchronization purposes then the value of M in the numerator is increased making the required bit rate higher.

On shorter connections the burst mode technique offers a simple and effective means of duplex transmission. For this reason it is favoured for internal loops which terminate on PABXs. For public network use, however, the poor reliable transmission range makes it less attractive.



Figure 4.12 Duplex transmission techniques.

(c) Echo cancellation. Also known as the adaptive hybrid technique, echo cancellation involves adaptively forming a replica of the echo signal arriving at a receiver from its local transmitter and subtracting it from the signal at the input to the receiver. As shown in Figure 4.12 the signals sent in each direction can occupy the same frequency band and are continuous; therefore the disadvantages of the FDD and TDD techniques are avoided. The penalty is that an adaptive echo canceller requires a complex hardware implementation. However, modern VLSI technology enables this complexity to be implemented at acceptable cost so that it is now recognized that echo cancelling duplex systems offer the best solution to subscriber loops with longer reach.

4.7 A COMPARISON OF TRANSMISSION TECHNIQUES

Before comparing transmission techniques it is instructive to evaluate the limits to digital transmission in the local network. The fundamental limitation to the communications capacity is imposed by crosstalk. Applying Shannon's theory (see Section 1.4) the capacity as a function of cable length can be calculated. Using a value of NEXT power sum attenuation of 55 dB



Figure 4.13 Transmission capacity and 0 dB signal to crosstalk frequency.

at 100 kHz and typical measurements of cable frequency response the graph in Figure 4.13 was plotted. The communications capacity of local network cables is clearly cable length dependent, but even over long lengths is sufficient to provide the data rates required for digital subscriber loops. Also plotted is the frequency at which the NEXT attenuation and cable attenuation are equal; this gives a bound on the maximum useful frequency above which more crosstalk is introduced than signal, suggesting that to achieve the highest performance the signal spectrum should be confined to low frequencies.

4.8 CROSSTALK ANALYSIS

Much effort has gone into the optimization of transmission systems in the local network. Making plausible assumptions about the local network it is possible to calculate the performance of equipment operating with different modulation schemes. Results are shown in Figure 4.14. The plots clearly show that pulse amplitude modulation is better than sinewave



Figure 4.14 Crosstalk limited performance.

amplitude modulation, decision feedback equalization is better than linear equalization and that multilevel coding allows greater data rates and/or greater reach.

4.9 IMPULSIVE NOISE

It is important that a system is designed to have the best tolerance to impulse noise. Such measurements of impulsive noise as have been done are usually in terms of the peak magnitude and frequency of occurrence, the implication being that impulses above a certain magnitude will cause errors. Unfortunately this simplistic approach neglects the effect of the receive filter and linear equalizer; to calculate the peak value of an impulse at their output a knowledge of impulse shape is required. In the absence of such information the best we can do is to compare systems by their tolerance to various carefully chosen impulse shapes.

Figure 4.15 shows the calculated tolerance of the various systems to a matched impulse when the transmit power is 10 mW into a 140 ohm load. PAM and decision feedback equalized systems have a clear advantage, but multilevel coding appears to decrease the tolerance to impulses. It should



Figure 4.15 Maximum error-free impulse amplitude for 144 kbit/s transmission.

be noted, however, that the impulses required to cause errors in the fourlevel case have twice the energy of the impulses required in the two-level case. If we normalize the plots on an equal energy basis then the four-level systems are better than the two-level systems. This discussion highlights the difficulty of comparing systems if no precise knowledge is available on the shape of practical impulses.

4.10 ECHO CANCELLER STRUCTURES

Figure 4.12(c) shows a general echo canceller which operates on the output of the transmitter to produce an echo replica. The echo to be cancelled is a function of the current transmitted data, past data and the electrical characteristics of the transmitter, cable and the line interface. The latter normally result in an echo impulse response which decays with time and so the adaptive network can be constrained to model a finite duration channel only apart from some insignificant residual echo. Any filters in the transmitter will cause the amplitude of the input to the network to be of a continuous nature; simpler echo cancellers result if a data driven structure is used as shown in Figure 4.16. The echo replica at time instant *iT* is given by

$$e(iT) = f(d_i, d_{i-1}, d_{i-2}, \ldots, d_{i-N+1})$$

where d_i is the current data element and d_{i-j} , j = 1 to N = 1 are the past N = 11 data elements. In a binary system the d_i are binary valued; in a multilevel system of L levels the d_i are represented by B bits, where B is equal to or greater than log₂ L. A straightforward implementation is to use the data elements as the address lines to look-up table which contains the 2^{NB} distinct values of e(iT). An echo canceller realized in this way is known as a table look-up echo canceller and has the great advantage of simplicity. Also, there are no constraints on the function f(.). However, the implementation has two disadvantages: the memory size increases exponentially with N and, as we shall see later, it is slow to converge when made adaptive. Both these disadvantages are gradually diminished if the look-up table is subdivided to produce partial outputs $e_i(iT)$ which are summed to give e(iT). Extending this process to a look-up table per data element and then assuming a linear relationship between each value of d_i and its corresponding partial output $e_i(iT)$ then a linear transversal filter echo canceller results; i.e.

$$e(iT) = \sum_{j=0}^{N-1} e_j(iT) = \sum_{j=0}^{N-1} d_{i-j} h_j$$



Figure 4.16 General data-driven echo canceller.



Figure 4.17 Transversal filter echo canceller.

where the values h_j are the samples of the echo impulse response. This structure is shown in Figure 4.17. The amount of memory required has been reduced to the minimum of N locations, but the processing increased to one multiplication and addition per delay line tap.

In the section on timing recovery it was noted that conventional timing recovery schemes require operations on the whole of the received signal. For such schemes to work correctly the echo must be removed from the whole of the received signal spectrum. Therefore, the echo canceller must operate at a sample rate that is at least twice the highest signal frequency in the received signal. Such fractional-tap spaced echo cancellers are easily obtained by multiplexing m (m greater than or equal to 2) of the structures of Figures 4.16 and 4.17 such that outputs are obtained every T/m seconds.

4.11 ECHO CANCELLER ADAPTATION

To adjust the echo canceller coefficients to their optimum values the error signal at the output of the subtractor is minimized by an appropriate adaptive algorithm. The most commonly used are the least-mean-squared (LMS) algorithms which seek to minimize the mean squared error signal. If $e^*(iT)$ is the actual echo at time iT then the error signal is

$$r(iT) = e^*(iT) - e(iT)$$

In the look-up table case the algorithm is trivially simple; each location in the table is updated, when it is addressed by the algorithm.

$$e((i+1)T) = e(iT) + \mu r(iT)$$

e((i + 1)T) then replaces e(iT) in the appropriate look-up table location. The gain constant (μ) determines the rate of convergence. For random data each location is addressed on average only 2^{-NB} of the time so convergence can be very slow.

For the linear transversal filter the algorithm is a little more complicated because the stored variables are not the values of e(iT) but the coefficients h_j . The commonest algorithm is given by

$$\mathbf{H}((i+1)T) = \mathbf{H}(iT) + \boldsymbol{\mu}.\boldsymbol{r}(iT).\mathbf{D}(iT)$$

where H(iT) and D(iT) are the vectors of tap coefficients and data elements respectively at time iT. This is known as the stochastic gradient algorithm. If the data elements are mutually uncorrelated then the optimum values of the coefficients h_j are the first N samples of the echo impulse response. Note that, unlike the look-up table echo canceller, each coefficient is updated every iteration of the algorithm and so it converges faster. When the echo canceller is working in the presence of the signal from the far end, that signal is superimposed on the error signal and will cause the echo canceller coefficients to fluctuate about their true values. This fluctuation will not affect the mean value of the coefficients provided the signal is not correlated with the transmitted signal, a property usually made highly likely by the use of different scramblers for the two directions of transmission.

4.12 NON-LINEARITY

As already stated the look-up table echo canceller can model a non-linear echo response. For the linear echo canceller any non-linear distortion products must be well below the level of cancellation required. Achieving the desired degree of linearity can be a problem for higher performance, longer reach systems implemented in low cost circuitry. At the same time such systems may require large echo canceller spans making the pure lookup table approach impractical. Fortunately, techniques are known that, by exploiting a knowledge of the actual non-linearity, allow the non-linear distortion to be coped with by the addition of extra echo canceller taps or by a combination of look-up tables and linear taps.

4.13 JITTER PERFORMANCE

Timing jitter is a well-known source of degradation in digital transmission systems, but in echo cancelling duplex systems it can be much more serious. If the sampling of the received signal and/or the generation of the transmitted signal is subject to jitter then the echo replica e(iT) formed by the echo impulse response is sampled at a slightly different phases. By having a timing recovery circuit that introduces known phase increments an echo canceller can be designed which corrects for the error.

4.14 DYNAMIC RANGE

Echo cancellers are required to suppress echoes to well below the level of the received signal from the far end. As the received signal is often highly attenuated and the hybrid circuit usually does not give a great deal of trans-hybrid loss the dynamic range requirement of the echo canceller is severe. We can write the minimum dynamic range of the echo replica signal as the product of:

peak transmit level, decision distance attenuation, trans-hybrid loss, quantizer peak signal to rms noise ratio

The decision distance in the back-to-back case is defined as unity and so the decision distance attenuation is the pulse loss of the cable, transmit and receive filters and the linear equalizer (if used). The noise used in calculating the S/N ratio is the sum of any external noise arriving at the input to the receiver, any uncancelled echo and ISI and the quantization noise arising from either the DAC or ADC used in the implementation.

The dynamic range is plotted in Figure 4.18 for a number of transmission system types and assuming a S/N of 20 dB (corresponding to a 6 dB margin against a 1 in 10^6 error rate for Gaussian noise). It can be seen that for long loops baseband systems are essential and even then the dynamic range requirement is severe. In a digital realization the digital accuracy (1 bit per 6 dB) is acceptable, but the provision of accurate DACs and ADCs constitutes a significant practical problem.



Figure 4.18 Minimum system signal-to-noise ratio for 144 kbit/s transmission.

4.15 TRANSMITTER REALIZATION

The transmitter is perhaps the easiest part of the duplex modem to realize. It does, however, in a data-driven echo cancelling system have some severe constraints imposed on its performance. The S/N ratio of the transmitted signal must be such that noise is acceptably below the received signal level at the input to the receiver as this cannot be removed by the echo canceller. As mentioned in Section 4.12, if a linear echo canceller is used any non-linearity in the transmitter must be insignificant. It is also advantageous if the transmit filter is constrained to be minimum phase. A non-minimum phase filter will tend to create an echo pre-cursor so that additional echo canceller taps will be necessary.

4.16 RECEIVER REALIZATION

The realization of the receiver function depends on its complexity. A receiver not requiring an adaptive equalizer may be readily realized in analogue form. The realization of an adaptive equalizer, whilst not quite as challenging as an echo canceller because of the reduced dynamic range requirements can, nevertheless, be facilitated by digital signal processing.

4.17 LINE INTERFACING

The line interface connects the unbalanced transmitter signal to the balanced 2-wire line and converts the balanced receive signal to an unbalanced form for the input to the receiver. In addition, if the subscriber loop is required to be line powered, it must allow separation of the power from signals. A line transformer with split windings can be used for power feeding or extraction. Inductors can be used to prevent power circuitry noise from interfering with transmission. The transformer split windings must be matched to ensure that any imbalance to ground does not cause the crosstalk characteristic to worsen.

Again the line interface must not introduce any significant level of noise and be sufficiently linear. In addition the values of inductor and transformer inductance have a marked effect on the cable pulse and echo responses and need to be carefully designed.

4.18 USA STANDARD SYSTEM

As discussed in Section 1.7, one of the perceived needs for the ISDN was to have an interface to which many terminals could be connected. In Section 1.9 the CCITT ISDN reference configuration is described and it was tacitly assumed that the customer would take responsibility for the connection at the T reference point, and that the network would be terminated by the NT1. However in the USA it was determined legally that the NT1 would be the responsibility of the customer and that the defined interface would be at the U reference point, to provide a 'wires only' interface. For this reason a U interworking interface has to be defined. Such an interface has to define the meaning of all the bits on the interface, their temporal order and their electrical characteristics. In particular, the line code and the modulation scheme are as we have already seen of great importance in determining the performance obtained.

In August 1986 the American National Standards Institute chose to define an interface based on 2B1Q line code, PAM at 80 kBauds initially proposed by British Telecom Research Laboratories at Martlesham. The fundamental modulation process is to convert two binary digits into one quaternary element (quat). The choice was based on both performance and complexity issues. It is convenient to represent the four possible quat states as +3, +1, -1, -3 as this indicates symmetry about zero, equal spacing between states and convenient integer magnitudes.

4.19 MODULATION FORMAT

The significance of all the bits transmitted, the transmission rate and the line coding together are known as the modulation format. 2B1Q is a PAM system; because each symbol represents two binary digits to carry the 144 kbit/s of the basic rate access a modulation rate of at least 72 kBaud is required. In practice additional 'housekeeping' channels are required for framing, maintenance and control functions and hence a modulation rate of 80 kBaud was chosen. Six kBauds are allocated for framing, leaving 4 kbit/s for maintenance and control.

A frame structure was chosen based on the use of a block synchronization word (SW) of nine quaternary elements repeated every 1.5 ms:

It has an autocorrelation function with a peak of approximately 4 relative to the normalized rms value of a random 2B1Q signal and the off-peak values are all equal to or less than 0.2 relative to the peak. These values are
a measure of the tolerance of the SW to ISI and of its efficacy at channel estimation for timing recovery respectively.

Alternative decision directed timing recovery schemes do not require a block SW and have the advantage that all the received signal is used for timing phase estimation. Therefore, they are less sensitive to noise and hence give better jitter performance than schemes which use only the block SW.

One clear advantage of the block SW is that it can be used to curtail error propagation events in a DFE-based transceiver by feeding a locally generated replica of the SW through the DFE at the appropriate time each frame. In conjunction with this it can also be exploited as an alternative to jitter compensation or an analogue phase locked loop (PLL), to enable timing phase adjustment to be made with a digital PLL such that the resulting echo cancellation error does not corrupt user data.

The frame structure of the line signal consists of 120 quats (i.e. 240 bits). Eight frames together make a multiframe. The first frame of the multiframe is indicated by sending the Inverted Synchronization Word (ISW). At the end of each frame are some maintenance and control bits, including provision of an embedded maintenance and control channel, activation bit, cold start indication bit, power status bits, and far-end error check bits.

Summarizing the frame structure is:

bits	quats	
1–18	1–9	Inverted Synchronization Word (frame 1) Synchronization Word (frames 2–8)
19–26	10-13	B-channel 1
27-34	14–17	B-channel 2
34-36	18	D-channel
37-44	19-22	В
45-52	23–26	В
53-54	27	D
55-62	28-31	В
63-70	32-35	В
71–72	36	D
73-80	37-40	В
81-88	41-44	В
89-9 0	4 5	D
91 -9 8	46-49	В
99–106	50-503	В
105–108	54	D
109-116	5558	В
117-124	59-62	В
25-126	63	D
27-134	64_67	B

bits	quats	
135-142	68–71	В
143-144	72	D
145152	73–76	В
153-160	77-80	В
161-162	81	D
163–170	82-85	В
171–178	86-89	В
179–180	90	D
181–188	91–94	В
189–196	95 98	В
197–198	99	D
199–206	100-103	В
207–214	104–107	В
215-216	108	D
217-224	109–112	В
225–232	113–116	В
233-234	117	D
235240	118-120	Maintenance and control

The frame repeats every 1.5 ms.

As mentioned in Section 4.11, for proper operation of the echo cancellers there must be no correlation between the signals in opposite directions and this is achieved by inserting scramblers. For the USA system these consist of shift registers with modulotwo additions as shown in Figure 4.19. The synchronization words are not scrambled, but are offset by half a frame in each direction of transmission to avoid correlation.

4.20 TRANSMIT SIGNAL SPECIFICATION

It is important that transceivers of different manufacture can interwork without any significant loss of performance. To achieve this, whilst still allowing a reasonable spread of component tolerances, the following features of the line signal are specified.

Transmit signal power

To minimize the effect of impulsive noise it is desirable to have as high a signal level as possible. The signal level has no effect on crosstalk between





ISDN systems as the interfering level and the interfered signal level will be similarly affected. The signal level is limited by two factors. First the need to avoid crosstalk interference to other non-ISDN services in the cable. Secondly to keep within the range of practical VLSI circuit output powers. The compromise of 13.5 \pm 0.5 dBm into 135 ohms is specified.

Pulse shape

The shape of the pulse implies a power spectrum and peak voltage which are chosen to enable the ISDN to coexist with other services. A pulse mask is defined which controls the variation between transceivers of different manufacture. The peak signal level is 2.5 V representing a +3 quat.

Non-linearity

It is possible for output pulses to lie within the pulse mask but still to be significantly different from a linear function of the generating quat. As discussed earlier, it is possible for equalizers and echo cancellers to cope with non-linear distortion but only at the expense of greater complexity and other degradations. To put a bound on the level of this non-linearity it is specified that the non-linear portion of the line signal power must be at least 36 dB below the signal power.

Frequency offset and jitter

Although the nominal modulation rate is 80 kBaud there are two effects which change this. First the clock to which the network is tied may be in error. In practice, although figures of 5 p.p.m. tolerance may be specified, the actual frequency offset is normally considerably less than this. The other source is jitter which is short-term variation of the timing waveform which arises from many sources, including the imperfect extraction of timing information from digital line systems.

The jitter is described in terms of its magnitude measured in the width of a symbol, known as a 'unit interval' (UI). The magnitude of the tolerable jitter varies with its frequency. Below 0.5 Hz it can be 0.3 UI pk-pk. Above 0.5 Hz the permitted jitter decreases inversely with frequency to 0.008 UI pk-pk at 19 Hz. Typically the free-running clocks in the NT1 are within 50 p.p.m. of the standard frequency. The low frequency jitter is approximately equivalent to a 6 p.p.m. frequency offset. Together with the possible network error this means that the NT1 must cope with a possible 61 p.p.m. offset (50 + 6 + 5).

At the customer's end the pulses generated by the NT1 will derive their timing from those received from the exchange. As clock recovery is not perfect, additional jitter may be added, which must be tracked by the exchange line termination. The additional jitter must be less than 0.04 UI pk-pk and 0.01 UI rms when measured with an 80 Hz high pass filter and must be less than 0.05 UI pk-pk and 0.015 UI rms when measured in a band between 1 and 40 Hz; in all cases the filters roll-off at 6 dB/octave. The variation with time between the output pulse phase and input pulse phase must not exceed 0.1 UI.

Start-up procedure

As discussed in Section 4.11, the echo canceller is self-adapting to the signal presented to it, but the rate of convergence of the process can very much depend on its detailed design. Similar considerations apply to convergence of DFEs and clock extraction circuits. Steps can be taken to minimize this start-up time including the use of special training sequences initially to which the echo canceller and DFE can easily converge, and the

storing from one call to another of information so that only a small update is required to take into account changes which may have taken place (e.g. cable temperature). The start-up procedure is defined flexibly and simply so that different transceiver designs are possible and so that the start-up time may be used in different ways according to the particular design. For a cold start (i.e. when the equipment is first used) 5 s are allowed for the NTI and 10 s for the exchange. As these are additive a total of 15 s may elapse. A warm start is allowed a total of 300 mS. In this case the equipment will have been used previously and echo canceller coefficients, DFE coefficients and exchange timing phase may have been stored from the previous use and may be used to facilitate a faster transceiver start-up.

4.21 LOCAL NETWORK TRANSMISSION OUTSIDE USA

As mentioned in Section 4.18, in the USA legal constraints have been imposed which require the customer interface to be at the U reference point. This was not originally foreseen by CCITT, who assumed that the customer interface would be at the T reference point and that the NT1 would be regarded as part of the network. This is mainly because the interface at the T reference point is more easily defined and not dependent on the implementation and characteristics of the particular administration's local network. In the future the local network copper pairs may well be replaced by optical fibres. The inclusion of the NT1 in the network shields the customer from the technological change elsewhere. For this reason there is no international standard for the U reference point. On copper pairs, Germany is using an echo cancelling system which uses a line code which converts four binary digits into three ternary digits (4B3T). The UK is using a 3B2T system but is planning to move to 2B1Q. Other countries are using AMI line code; however this is very restricted on range.

Nevertheless, with the widespread availability of 2B1Q devices to meet the USA market, it is likely that these will also be used by other administrations, where service is still provided on copper pairs.

REFERENCE

American National Standards Institute: ANSI T1.601–1988, ISDN Basic Access Interface for use on Metallic Loops for application on the Network side of the NT (layer 1 specification).

Chapter 5

The Basic Rate ISDN Customer's Interface

It has already been mentioned in Section 1.7 that CCITT decided to offer two 64 kbit/s B channels as the basic access to the customer plus a D signalling channel of 16 kbit/s. It is called a 'basic' access to distinguish it from other formats such as 30B + D channels in Europe and 23B + Dchannels in North America which are known as 'primary rate' accesses, and from reading Section 3.1 will be recognized as being of the same form as the inter-exchange transmission systems. Primary rate access is discussed further in Chapter 6.

The basic rate access is intended to be equivalent in ISDN terms to the simple telephone in analogue terms. The connection of a multiplicity of analogue terminals (e.g. telephones, modems) is achieved by simply connecting them in parallel. The requirement for the ISDN interface is also to be able to simply connect terminals in parallel with the additional feature that terminals may be individually addressed by type or other identifier.

5.1 THE CUSTOMER'S INSTALLATION John Hovell

The CCITT reference configuration for basic rate access described in Sections 1.7 and 1.9 is represented pictorially in Figure 5.1.

The two-wire transmission line from the local network is terminated by the NT1. Although this NT1 is located on the customer's premises it is on the network side of the CCITT defined user/network interface and hence the responsibility of the network operator. The exception to this is in the USA where the user/network interface is at the network side of the NT1 as explained in Section 4.18.

Into this NT1 is connected a four-wire bus, often known as the 'S' bus because it is at the CCITT S reference point. This bus is designed to operate over normal twisted pair internal cable as may be provided for traditional analogue sources. The bus can operate in two modes, point-to-point or point-to-multipoint. In the point-to-point mode one Terminal Equipment (TE) is connected at the end of up to a maximum of about 1 km of cable; the





actual limit is an attenuation of 6 dB at 96 kHz. In the point-to-multipoint mode up to eight terminals can be connected in parallel anywhere along the bus, but the bus length is now limited to about 200 metres by timing constraints. The ends of the bus are terminated by resistors. Connections are made to the bus via 8-pin plugs and sockets developed from the standard 6-pin American telephone connectors.

Over this bus passes the two transparent 64 kbit/s B channels, the 16 kbit/s D signalling channel and other bits used for miscellaneous purposes such as frame synchronization and collision avoidance. This contributes another 48 kbit/s making a total of 192 kbit/s.

The B channels contain the user data which is switched by the network to provide an end-to-end transmission service. A B channel path is established by signalling messages in the D channel.

In a multi-terminal situation all terminals have access to the D channel by the use of an access procedure but each B channel is allocated to a particular terminal at the time of call set-up and is not capable of being shared between terminals.

5.1.1 Power feeding

Power is provided across the user/network interface to provide a basic telephone service in the event of a local mains failure. The physical arrangement for power feeding is shown in Figure 5.2. The centre of the diagram shows a basic 4-wire bus consisting of two twisted pairs. A power source 1 is available within the NT which can supply, via the centre tapped transformers (known as a phantom feed), at a nominal 40 volts, up to 1 watt of power. The power under these normal arrangements will probably be derived at the NT1 from a local mains source. Under mains failure conditions the power will be limited to 420 mW, and to indicate this state to



note 1: this symbol refers to the polarity of framing pulses

note 2: this symbol refers to the polarity of power during normal power conditions (reversed for restricted conditions)

Figure 5.2

Reference configuration for signal transmission and power feeding in normal operating mode.

the terminals the polarity will be reversed. In this case the power will be derived from the network, and is only intended to power a single digital telephone for emergency use. Note that there are an additional two optional pairs allocated for alternative power feeding arrangements, but it is not clear the extent to which these will be used.

5.1.2 Layer 1 of the user-network interface

Layer 1 is responsible for transferring information between terminals and the NT1. The transmission medium on which it must achieve this is not perfect and errors may well be introduced. For the D channel signalling ('C' plane) between terminals and the network it is Layer 2's responsibility to handle any error detection and correction. For user information on the B channels (the 'U' plane) the network offers no error control although it must be stressed that the error rate is expected to be very low.

5.1.3 The functions of Layer 1

In order to provide this transport service Layer 1 must support the functions outlined below.

B channels

Layer 1 must support for each direction of transmission, two independent 64 kbit/s B channels. The B channels contain user data which is switched by the network to provide the end-to-end transmission source. Note there is no error correction provided by the network on these channels.

D channels

Layer 1 must support for each direction of transmission, a 16 kbit/s channel for the signalling information. In some networks user packet data may also be supported on the D channel.

D channel access procedure

This procedure ensures that in the case of two or more terminals on a point-to-multipoint configuration attempting to access the D channel simultaneously, one terminal will always successfully complete the transmission of information. This procedure is described in more detail later.

Deactivation

Deactivation permits the TE and NT equipments to be placed in a low power mode when no calls are in progress; activation restores the equipment to its normal power mode. A low power mode is required for two purposes. First to save power. This is significant from the network point of view as the NT1 is line powered. Secondly, deactivation is also provided to reduce electromagnetic radiation and hence crosstalk interference.

The interface may be activated from either the terminal and network side, but deactivated only from the network side because of the NT1's multi-terminal capability. This procedure is described in more detail later.

5.1.4 Binary organization of Layer 1 frame

The structures of the Layer 1 frames across the interface are different in each direction of transmission. Both structures are shown in Figure 5.3.

A frame is 48 bits long and lasts $250 \ \mu$ s. The bit rate is therefore 192 kbit/s and each bit approximately $5.2 \ \mu$ s long. Figure 5.3 shows that there is a 2-bit offset between transmit and receive frames. This is the delay between a



Figure 5.3 Frame structure. frame start at the receiver of a terminal and the frame start of the transmitted signal.

Figure 5.3 also illustrates that the line code used is Alternate Mark Inversion (AMI); a logical 1 is transmitted as zero volts and a logical 0 as a positive or negative pulse. Note that this convention is the inverse of that used on line transmission systems (see Section 3.2). The nominal pulse amplitude is 750 mV.

A frame contains several L bits; these are balance bits to prevent a buildup of d.c. on the line. The frames are split into balanced blocks as represented by dots on the diagram. For the direction TE to NT, where each B-channel may come from a different terminal, each terminal's output contains an L bit to form a balanced block.

Examining the frame in the NT to TE direction, the first bits of the frame are the F/L pair, which is used in the frame alignment procedure. The start of a new frame is signalled by the F/L pair violating the AMI rules. Obviously once a violation has occurred, to restore correct polarity before the next frame, there must be a second violation. This takes place with the first mark after the F/L pair. The FA bit ensures this second violation occurs should there not be a mark in the B1, B2, D, E, or A channels. The E channel is an echo channel in which D-channel bits arriving at the NT1 are echoed back to the TEs. There is a 10-bit offset between the D channel leaving a terminal, travelling to the NT and being echoed back in the E channel.

The A bit is used in the activation procedure to indicate to the terminals that the system is in synchronization. Next is a byte of the B2 channel, a bit of the E channel and a bit of the D channel, followed by a M bit. This is used for multiframing, a facility not provided in Europe, but which may be used elsewhere. The M bit identifies some FA bits which can be stolen to provide a management channel.

The B1, B2, D and E channels are then repeated along with the S bit which is a spare bit.

5.1.5 Layer 1 D channel contention procedure

This procedure ensures that, even in the case of two or more terminals attempting to access the D channel simultaneously, one terminal will always successfully complete the transmission of information. The procedure relies on the fact that the information to be transmitted consists of Layer 2 frames delimited by flags consisting of the binary pattern 01111110. Layer 2 applies a zero bit insertion algorithm to prevent flag imitation by a Layer 2 frame. The interframe time fill consists of binary 1s which are represented by zero volts. The zero volt line signal is generated by the TE transmitter going high impedance. This means a binary 0 from a parallel terminal will overwrite a binary 1. Detection of a collision is done by the terminal monitoring the E channel (the D channel echoed from the NT).

To access the D channel a terminal looks for the interframe time fill by counting the number of consecutive binary 1s in the D channel. Should a Binary 0 be received the count is reset. When the number of consecutive 1s reaches a predetermined value (which is greater than the number of consecutive 1s possible in a frame because of the zero bit insertion algorithm) the counter is reset and the terminal may access the D channel. When a terminal has just completed transmitting a frame the value of the count needed to be reached before another frame may be transmitted is incremented by 1. This gives other terminals a chance to access the channel. Hence an access and priority mechanism is established.

There is still the possibility of a collision between two terminals of the same priority. This is detected and resolved by each terminal comparing its last transmitted bit with the next E bit. If they are the same the terminals continue to transmit. If, however, they are different the terminal detecting the difference ceases transmission immediately and returns to the D channel monitoring state leaving the other terminal to continue transmission.

5.1.6 Layer 1 activation/deactivation procedure

This procedure permits activation of the interface from both the terminal and network side, but deactivation only from the network side. This is because of the multi-terminal capability of the interface. Activation and deactivation information is conveyed across the interface by the use of line signals called 'Info signals'.

Info 0 is the absence of any line signal; this is the idle state with neither terminals nor the NT working.

Info 1 is flags transmitted from a terminal to the NT to request activation. Note this signal is not synchronized to the network.

Info 2 is transmitted from the NT to the TEs to request their activation or to indicate that the NT has activated as a response to receiving an Info 1. An Info 2 consists of Layer 1 frames with a high density of binary zeros in the data channels which permits fast synchronization of the terminals.

Info 3 and Info 4 are frames containing operational data transmitted from the TE and NT respectively.

The principal activation sequence is commenced when a terminal transmits an Info 1. The NT activates the local transmission system which

indicates to the exchange that the customer is activating. The NT1 responds to the terminals with an Info 2 to which the TEs synchronize. The TEs respond with an Info 3 containing operational data and the NT is then in a position to send Info 4 frames. Note that all terminals activate in parallel; it is not possible to have just one terminal activated in a multi-terminal configuration. The network activates the bus by the exchange activating the local network transmission system. Deactivation occurs when the exchange deactivates the local network transmission system.

5.1.7 Layer 1 physical implementation

Semiconductor manufacturers are introducing devices to perform the functions described above. This is often in the form of two chips as illustrated in Figure 5.4, one providing the transmission functions and the other digital logic. The first chip contains the transmitters and receivers with associated timing extraction. The other chip contains a frame formatter, channel multiplexing and demultiplexing and a microprocessor interface for control of both chips. This second chip often also contains the lower layer functions of the Layer 2 HDLC processing for the D channel. To complete the circuitry a line transformer is used to provide a balanced line signal and d.c. isolation. To protect the silicon from induced transients



Figure 5.4 Typical Layer 1 Implementation—terminal side.

on the bus, protection diodes are required and, ideally, some series impedance to dissipate the power.

Also becoming available are d.c.-d.c. power converters specifically designed for this ISDN application. The NT power controllers provide the 40 V phantom line feed from a 50 V source. The TE power controllers convert the received line voltage into 5 V for driving the TE circuitry.

5.2 LAYER 2

Kevin Woollard

The Layer 2 recommendation describes the high level data link (HDLC) procedures commonly referred to as the Link Access Procedure for a D channel or LAP D. The objective of Layer 2 is to provide a secure, error-free connection between two end-points connected by a physical medium. Layer 3 call control information is carried in the information elements of Layer 2 frames and it must be delivered in sequence and without error. Layer 2 also has responsibility for detecting and retransmitting lost frames.

LAP D was based originally on Lap B of the X.25 Layer 2 recommendation. However, certain features of LAP D give it significant advantages. The most striking difference is the possibility of frame multiplexing by having separate addresses at Layer 2 allowing many LAPs to exist on the same physical connection. It is this feature that allows up to eight terminals to share the signalling channel in the passive bus arrangement.

Each Layer 2 connection is a separate LAP and the termination points for the LAPs are within the terminals at one end and at the periphery of the exchange at the other. Layer 2 operates as a series of frame exchanges between the two communicating, or peer, entities. The frames consist of a sequence of eight bit elements and the elements in the sequence define their meaning as shown in Figure 5.5. A fixed pattern, called a flag, is used to indicate both the beginning and end of a frame. Two octets are needed for the Layer 2 address and carry a service identifier (SAPI), a terminal identifier (TEI) and a command/response bit. The control field is one or two octets depending on the frame type and carries information that identifies the frame and the Layer 2 sequence numbers used for link control. The information element is only present in frames that carry Layer 3 information and the Frame Check Sequence (FCS) is used for error detection. A detailed breakdown of the individual elements is given in Figures 5.6 to 5.8.

What cannot be shown in the diagrams is the procedure to avoid imitation of the flag by the data octets. This is achieved by examining the serial stream between flags and inserting an extra '0' after any run of five '1' bits. The receiving Layer 2 entity discards a '0' bit if it is preceded by five '1's.

octet 1	0 1 1 1 1 1 1 0	Opening flag		
octet 2	Address octet 1			
octet 3	Address octet 2			
octet 4	Control octet 1	The structure of the control field depends on the frame type and		
octet 5	Control octet 2*	is described in Figure 5.7		
octet 6	Layer 3 information	Layer 3 information is only present in Layer 2 ' Information frames '.		
octet n-3				
octet n-2	FCS octet 1	Frame check sequence		
octet n-1	FCS octet 2	Frame check sequence		
octet n	01111110	Closing flag		

*The second octet of the control field is not always present

Figure 5.5 Layer 2 frame structure.

The two octets of the address field are coded as shown:

A	7	_	E				•	
•				•	3	۷.	•	
		S A	ΡI			C/R	EAO	octet 2
			TEI				EA1	octet 3

- SAPI Service access point identifier
- TEI Terminal endpoint identifier
- C/R Command/response bit
- EA0 Extended address bit 0
- EA1 Extended address bit 1



Control field format for information frames



Control field format for supervisory frames

	1	2	3	4	5	6	7	8
octet	1	1	M	M	P/F	M	M	M

Control field format for unnumbered frames

- N(S) Transmitter send sequence number
- N(R) Transmitter receive sequence number
- S Supervisory function bit
- M Modifier function bit
- P/F Poll bit when issued as a command,
 - final bit when issued as a response
- X Reserved and set to 0

Figure 5.7

Control field formats. The control field may be 1 or 2 octets in length depending on the frame type. The detail of the control field formats are given in the figure.

5.2.1 Layer 2 addressing

Layer 2 multiplexing is achieved by employing a separate Layer 2 address for each LAP in the system. To carry the LAP identity the address is two octets long and identifies the intended receiver of a command frame and the transmitter of a response frame. The address has only local significance and is known only to the two end-points using the LAP. No use can be made of the address by the network for routing purposes and no information about its value will be held outside the Layer 2 entity.

	05000005	ENCODING									
COMMANL	HESPUNSE	8	7	6	5	4	3	2	1		
INFORMAT	TION FRAMES										
					N(S)				0	4	
IFRAME -			N(R)								
SUPERVIS	ORY FRAMES										
	PD	0	0	0	0	0	0	0	1	4	
nn	nn		N(R)								
DND		0	0	0	0	0	1	0	1	4	
nnn	BNB		N(R)								
BEI	DEI	0	0	0	0	1	0	0	1	4	
HEJ HEJ		N(R)								5	
UNNUMB	ERED FRAMES	5									
SABME	-	0	1	1	P	1	1	1	1	4	
-	DM	0	0	0	F	1	1	1	1	4	
U	-	0	0	0	P	0	0	1	1	4	
DISC	-	0	1	0	P	0	0	1	1	4	
-	UA	0	1	1	F	0	0	1	1	4	
-	FRMR	1	0	C	F	0	1	1	1	4	
IFRAME I RR I RNR I REJ I SABME	nformation frai Receiver ready Receiver not re Reject Set asynchronx belanced mod	me ady ous		D U Fi D	ISC A RMR M	Disco Unnu ackr Fram Disco	nnect mbere iowiec e rejec innect	id Igemi t mode	ent e	~	

Figure 5.8

Layer 2 frame control field codings. Three Layer 2 frames types have been identified in Figure 5.7. By assigning different combinations to the S and M bits many frames may be defined. The figure shows the control fields of all the Layer 2 frames and indicates whether a frame exists as a command, response or both.

The Layer 2 address is constructed as shown in Figure 5.6. The details of each element will be described in the next section but it is useful to introduce the concepts associated with the SAPI and TEI fields.

The Service Access Point Identifier (SAPI) is used to identify the service that the signalling frame is intended for. An extension of the use of the D channel is to use it for access to a packet service as well as for signalling. Consider the case of digital telephones sharing a passive bus with packet terminals. The two terminal types will be accessing different services and possibly different networks. It is possible to identify the service being invoked by using a different SAPI for each service. This gives the network the option of handling the signalling associated with different services in separate modules. In a multi-network ISDN it allows Layer 2 routing to the appropriate network. The value of the SAPI is fixed for a given service.

The Terminal Endpoint Identifier (TEI) takes a range of values that are associated with terminals on the customer's line. In the simplest case each terminal will have a single unique TEI value. The combination of TEI and SAPI identify the LAP and provide a unique Layer 2 address. A terminal will use its Layer 2 address in all transmitted frames and only frames received carrying the correct address will be processed.

In practice a frame originating from telephony call control has a SAPI that identifies the frame as 'telephony' and all telephone equipment will examine this frame. Only the terminal whose TEI agrees with that carried by the frame will pass it to the Layer 2 and Layer 3 entities for processing.

As it is important that no two TEIs are the same, the network has a special TEI management entity which allocates TEI on request and ensures their correct use. The values that TEIs can take fall into the ranges:

0–63	non-automatic assignment TEIs
64–126	automatic assignment TEIs
127	global TEI

Non-automatic TEIs are selected by the user; their allocation is the responsibility of the user. Automatic TEIs are selected by the network; their allocation is the responsibility of the network. The mechanism is discussed in Section 5.2.5. The global TEI is permanently allocated and is often referred to as the broadcast TEI.

Terminals which use TEIs in the range 0–63 need not negotiate with the network before establishing a Layer 2 connection. Nevertheless the rule that all TEIs on the customer's line must be different still applies and it is the responsibility of the user to ensure that he does not connect two terminals using the same non-automatic TEI.

Terminals which use TEIs in the range 64–126 cannot establish a Layer 2 connection until they have requested a TEI from the network. In this case it is the responsibility of the network not to allocate the same TEI more than once to any given time.

The global TEI is used to broadcast information to all terminals within a given SAPI; for example a broadcast message to all telephones, offering an incoming telephony call.

5.2.2 Layer 2 operation

The function of Layer 2 is to deliver Layer 3 frames, across a Layer 1 interface, error free and in sequence. It is necessary for a Layer 2 entity to interface to both Layer 1 and Layer 3. To highlight the operation of Layer 2 we will consider the operation of a terminal as it attempts to signal with the network.

It is the action of attempting to establish a call that causes the protocol exchange between terminal and network. If there has been no previous communication it is necessary to activate the interface in a controlled way. A request for service from the customer results in Layer 3 requesting a service from Layer 2. Layer 2 cannot offer a service unless Layer 1 is available and so the appropriate request is made to Layer 1. Layer 1 then initiates its start-up procedure and the physical link becomes available for Layer 2 frames. Before Layer 2 is ready to offer its services to Layer 3 it must initiate the Layer 2 start-up procedure known as 'establishing a LAP'.

LAP establishment is achieved by the exchange of Layer 2 frames between the Layer 2 handler in the terminal and the corresponding Layer 2 handler in the network (the peer entity). The purpose of this exchange is to align the state variables that will be used to ensure the correct sequencing of information frames. Before the LAP has been established the only frames that may be transmitted are unnumbered frames. The establishment procedure requires one end-point to transmit a Set Asynchronous Balanced Mode Extended (SABME) and the far end to acknowledge it with an Unnumbered Acknowledgement (UA).

Once the LAP is established Layer 2 is able to carry the Layer 3 information and is said to be the 'multiple frame established state'. In this state Layer 2 operates its frame protection mechanisms.

Figure 5.9 shows a normal Layer 2 frame exchange. Once established the LAP operates an acknowledged service in which every information frame

terminal r	etwork
SABME	->
<	UA
IFRAME ·····	•>
<	RA
IFRAME	->
<	RR
< IFRA	ME
RR	->

Figure 5.9 A normal Layer 2 frame exchange.

(I frame) must be responded to by the peer entity. The most basic response is the Receiver Ready (RR) response frame. Figure 5.9 shows the LAP establishment and the subsequent I frame RR exchanges. The number of I frames allowed to be outstanding without an acknowledgement is defined as the window size and can vary between 1 and 127. For telephony signalling applications the window size is 1 and after transmitting an I frame the Layer 2 entity will await a response from the corresponding peer entity before attempting to transmit the next I frame. Providing there are no errors all that would be observed on the bus would be the exchange of I frames and RR responses. However Layer 2 is able to maintain the correct flow of information, in the face of many different error types.

5.2.3 Layer 2 error control

It is unlikely that a frame will disappear completely but it is possible for frames to be corrupted by noise at Layer 1. Corrupted frames will be received with invalid Frame Check Sequence (FCS) values and consequently discarded.

The frame check sequence is generated by dividing the bit sequence starting at the address up to (but not including) the start of the frame check sequence by the generator polynomial $X^{16} + X^{12} + X^5 + 1$. In practical terms this is done by a shift register as shown in Figure 5.10. All registers are preset to 1 initially. At the end of the protected bits the shift register contains the remainder from the division. The 1's complement of the remainder is the FCS. At the receiver the same process is gone through, but this time the FCS is included in the division process. In the absence of transmission errors the remainder should always be 1101 0000 1111.

The method for recovering from a lost frame is based on the expiry of a timer. A timer is started every time a command frame is transmitted and is stopped when the appropriate response is received. This single timer is thus able to protect both the command and response frame as the loss of either will cause it to expire.

When the timer expires it is not possible to tell which of the two frames has been lost and the action taken is the same in both cases. Upon the timer



Figure 5.10 Generation of frame check sequence.

expiring, the Layer 2 transmits a command frame with the poll bit set. This frame forces the peer to transmit a response that indicates the value held by the state variables. It is possible to tell from the value carried by the response frame whether or not the original frame was received. If the first frame was received, the solicited response frame will be the same as the lost response frame and is an acceptable acknowledgement. If however the original frame was lost, the solicited response will not be an appropriate acknowledgement and the Layer 2 entity will know that a retransmission is required.

It is possible for the same frame to be lost more than once and Layer 2 will retransmit the frame three times. If after three retransmissions of the frame the correct response has not been received, Layer 2 will assume that the connection has failed and will attempt to re-establish the LAP. Figure 5.11 illustrates a Layer 2 sequence recovering from two frame losses.

Another possible protocol error is the arrival of an I frame with an invalid send sequence number N(S). This error is more likely to occur when the LAP is operating with a window size greater than one. If, for example, the third frame in a sequence of four is lost the receiving Layer 2 entity will know that a frame has been lost from the discontinuity in the sequence numbers. The Layer 2 must not acknowledge the fourth frame as this will imply acknowledgement of the lost third frame. The correction operation for a Reject (REJ) frame to be returned with the receive sequence number N(R) equal to N(S) + 1, where N(S) is the send variable of the last correctly received I frame, in this case I frame two. This does two things; first it acknowledges all the outstanding I frames up to and including the second I frame, and secondly it causes the sending end to retransmit all outstanding I frames starting with the lost third frame.



Figure 5.11 Layer 2 recovery.

The receipt of a frame with an out of sequence, or invalid, N(R) does not indicate a frame loss and cannot be corrected by retransmissions. It is necessary in this case to re-establish the LAP to realign the state variables at each end of the link.

The Receiver Not Ready (RNR) frame is used to inhibit the peer Layer 2 from transmitting I frames. The reasons for wanting to do this are not detailed in the specification but it is possible to imagine a situation where Layer 3 is only one of many functions to be serviced by a microprocessor and a job of higher priority requires that no Layer 3 processing is performed.

Another frame specified in Layer 2 is the FRaMe Reject frame (FRMR). This frame may be received by a Layer 2 entity but may not be transmitted. It is included in the recommendation to preserve alignment between LAP D and LAP B. After the detection of a frame reject condition the data link is reset.

5.2.4 Disconnecting the LAP

After Layer 3 has released the call it informs Layer 2 that it no longer requires a service. Layer 2 then performs its own disconnection procedures so that ultimately Layer 1 can disconnect and the transmission systems associated with the local line and the customer's bus can be deactivated.

Layer 2 disconnection is achieved when the frames disconnect (DISC) and UA are exchanged between peers. At this point the LAP can no longer support the exchange of I frames and supervisory frames.

The last frame type to be considered is the Disconnect Mode (DM) frame. This frame is an unnumbered acknowledgement and may be used in the same way as a UA frame. It is used as a response to a SABME if the Layer 2 entity is unable to establish the LAP, and a response to a DISC if the Layer 2 entity has already disconnected the LAP.

5.2.5 TEI allocation

Because each terminal must operate using a unique TEI, procedures have been defined in a Layer 2 management entity to control their use. The TEI manager has the ability to allocate, remove, check and verify TEIs that are in use on the customer's bus. As the management entity is a separate service point all messages associated with TEI management are transmitted with a management SAPI.

TEI management procedures must operate regardless of the Layer 2 state and so the unnumbered information frame (UI) is used for all management Terminal

Network

UI TEI REQUEST ·····> (REF No. X)
<ui allocate<="" td="" tei=""></ui>
(REF No. X)
SABME>
< UA
IFRAME>
< RR

etc.



messages. The UI frames have no Layer 2 response and protection of the frame content is achieved by multiple transmissions of the frame.

In order to communicate with terminals which have not yet been allocated TEIs a global TEI is used. All management frames are transmitted on a broadcast TEI which is associated with a LAP that is always available. All terminals can transmit and receive on the broadcast TEI as well as their own unique TEI. All terminals on the customer's line will process all management frames. To ensure that only one terminal acts upon a frame a unique reference number is passed between the terminal and network. This reference number is contained within an element in the UI frame and is either a number randomly generated by the terminal, or 0 is the TEI of the terminal, depending on the exact situation. Figure 5.12 shows the frame exchange required for a terminal to be allocated a TEI and establish its data link connection.

5.3 LAYER 3

Derek Rumsey

This layer effects the establishment and control of connections. It is carried in Layer 2 frames as can be seen in Figure 5.5. The messages are passed between the terminals and the exchange and vice versa. The first procedures defined relate to the control of circuit-switched calls. However later work defines the use of ISDN for packet-switched calls in the D channel, for exchange of information between users, and the invoking of supplementary services. The general structure of Layer 3 signalling messages is shown in Figure 5.13.

The first octet contains a protocol discriminator which gives the D channel the capability of simultaneously supporting additional



Figure 5.13 Signalling message structure.

communications protocols in the future. The bits shown are the standard for user-network call control messages.

The call reference value in the third octet is used to identify the call with which a particular message is to be associated. Thus a call can be identified independently of the communications channel on which it is supported. This feature is particularly important in connection with incoming call offering procedures on a passive bus arrangement since the channel is only allocated to the called terminal after answer.

The message type code in the fourth octet describes the intention of the message (e.g. a SETUP message to request call establishment). These are listed in Table 5.1. A number of other information elements may be included following the message type code in the fourth octet. The exact contents of a message are dependent on the message type, however the coding rules are open ended and in principle it is a simple matter to include additional information elements to satisfy any requirement which may be identified in the future. By way of example, Table 5.2 lists the information elements which may be contained in a SETUP message. This is probably the most complex message yet defined.

The message sequence for call establishment is shown in Figure 5.14. In order to make an outgoing call request, a user must send all of the necessary call information (i.e. called party number and supplementary service requests) to the network. Furthermore the user must specify the particular bearer service required for the call (i.e. Speech, 64 kbit/s unrestricted or 3.1 kHz Audio) and any terminal compatibility information

Me	essaę	e ty	pes			_		
8	7	6	5	4	3	2	1	
0	0	0	0	0	0	0	0	 escape to nationally specific message typ
0	0	0	-	-	-	-	-	Call establishment message:
			0	0	0	0	1	- ALERTING
			0	0	0	1	0	CALL PROCEEDING
			0	0)	}	1	- CONNECT
			0	1	1	J	1	CONNECT ACKNOWLEDGE
			0	0	0	1	I.	- PROGRESS
			0	0	ł	0	1	— SETUP
			0	١	1	0	I	SETUP ACKNOWLEDGE
0	0	ł	-	-	-	-	-	Call information phase message:
			0	0	1	1	0	— RESUME
			0	I	1	1	0	RESUME ACKNOWLEDGE
			0	0	0	I.	0	RESUME REJECT
			0	0	ł	0	1	SUSPEND
			0	1	1	0	I	
			0	0	0	0	I	
			0	0	0	0	0	USER INFORMATION
0	1	0	-	-	-	-	-	Call clearing messages:
			0	0	1	0	1	- DISCONNECT
			0	ł	1	0	1	— RELEASE
			1	1	0	1	0	- RELEASE COMPLETE
			0	0	1	I	0	RESTART
			0	I	I	ł	0	RESTART ACKNOWLEDGE
0	1	l	-	-	-	-	-	Miscellaneous messages:
			0	0	0	0	0	SEGMENT
			ł	1	0	0	1	CONGESTION CONTROL
			1	I.	0	ł	I	- INFORMATION
			0	0	0	L	0	FACILITY
			0	I.	ł	ł	0	NOTIFY
			1	ì	1	0	1	— STATUS
			1	0	1	0	1	STATUS ENQUIRY

which must be checked at the destination. A low layer compatibility information element may be used to specify low layer terminal characteristics such as data rate. Where applicable the non-voice application to be used on the call may be specified via the high layer compatibility information element (i.e. Group 4 FAX, Teletex, Videotex or Slow Scan Video).

The initial outgoing call request may be made in an en bloc or overlap manner. Figure 5.14 illustrates the call establishment procedures. If

Table 5.1

	Direction	Length (octets)	Comments
Protocol discriminator	Both		
Call reference	Both	2-3	
Message type	Both	-	
Sending complete	Both	_	Optional; included if user or network indicates that all
			information is included in this SETUP message
Bearer capability	Both	6-8	Indicates CCITT telecommunications service
Channel identification	Both	2-2	To identify a channel within an ISDN interface
			controlled by these procedures
Network specific facilities	Both	2-?	Optional
Display	Network-User	2-82	Optional: IA5 (ASCII) characters for display at terminal
Keypad	User-Network	2-34	Alternative for transferring called party number
			Keypad may be used for other information
Calling party number	Both	2-2	Optional
Calling party sub-address	Both	2-23	Optional
Called party number	Both	2-5	In user-network direction alternative to use of keynad
Called party sub-address	Both	2-23	Optional
Transit network selection	User-Network	2- <i>2</i>	Optional
Low layer compatibility	Both	2-16	Optional
High layer compatibility	Both	2-4	Optional
User-user	Both	2-131	Optional, when the calling user wants to pass user
			information to the called user

overlap sending is used then the SETUP message must contain the bearer service request but the facility requests and called party number information may be segmented and conveyed in a sequence of INFORMATION messages as shown. Furthermore if a speech bearer service is requested and no call information is contained in the SETUP message, then the network will return in-band dial tone to the user until the first INFORMATION message has been received.

Following the receipt of sufficient information for call establishment, the network returns a call PROCEEDING message to indicate that it is attempting to establish a connection to the called party.

At the called side a SETUP message is delivered to the called party via the broadcast data link. All terminals connected to the NT1 can examine the SETUP message to determine whether or not they are compatible with the calling party. This compatability check must be performed by examining the bearer capability and low layer compatibility information





Message sequences for basic call establishment and clearing.

element. Furthermore an indication of the application required (e.g Group 4 Fax and Teletex) will be found in the high layer compatibility information element if this has been provided by the calling party. In this particular example two terminals have identified that they are compatible (TE15 and 8) and return an ALERTING message to the network. At the same time, if appropriate, the terminal should generate a local indication of the incoming call (e.g. ring bell). When a terminal detects that the call has been answered it forwards a CONNECT message to the network. The network allocates the call to the first terminal to return a CONNECT and instructs that terminal to connect to the appropriate B channel by returning a CONNECT ACKnowledge containing the B channel identification. All other terminals which had responded to the incoming call will receive a RELEASE message from the network indicating that they should clear and return to the idle state. Following the receipt of a CONNECT message from a terminal at the called user's installation, the network will advise the calling party that the call has been answered by sending a CONNECT message. Call charging can then commence if appropriate.

A user may clear a call at any time by initiating a three message sequence (DISConnect, RELease and RELease COMPlete). In certain circumstances clearing may be initiated through the sending of RELease or RELease COMplete.



Figure 5.15 Suspend/resume procedures.

In addition to call establishment and clearing it is possible to suspend a call in order to move a terminal between sockets on the passive bus. Following such an operation, the call may be resumed from the new position by sending a RESUME message to the network with the call identity in the form of two IA5 characters as defined by the user at the time of call suspension. The message sequences for suspend/resume are shown in Figure 5.15.

5.3.1 Layer 3 supplementary services

Two ways of invoking supplementary services have been identified.

Stimulus

This makes use of information elements KEYPAD and DISPLAY. KEYPAD information elements may be used in the direction user to network to convey strings of IA5 characters representing the results of key depressions at the man/machine interface presented by the terminal.

Through the use of separators such as '*' and '#', these character strings can be arranged as facility requests comprising a facility code representing the supplementary service required followed by parameters to identify the various options selected. Thus a facility request may be represented in the following form:

Facility Request:

<+> <Facility Code> <+> <parameter 1> <+> <parameter n> <#>

Facility requests may be conveyed during outgoing call establishment within a SETUP/INFOrmation message sequence or during an established call via the INFOrmation message. Although these character sequences may be generated directly from a numeric keypad with additional characters '*' and '#', they may also be generated through the depression of a pre-programmed key or by an automatic terminal without direct human intervention.

The network response to facility requests will be generally in the form of a sequence of IA5 characters in a DISPLAY information element or an inband tone or announcement. The arrangement of IA5 characters within a DISPLAY information element is intended to be suitable for display to a human operator. No change of call state is ever implied by the transfer of DISPLAY information elements. They are usually conveyed across the user/network interface within INFOrmation messages which in themselves do not give rise to a call state change. The details of the syntax of this technique are not the subject of CCITT standardization. It is known as 'stimulus' as its operation does not in principle require the user equipment to have any knowledge of the supplementary service involved.

A development of this technique uses 'feature activation' and 'feature indication' information elements included in the SETUP and INFORMATION messages and is known as a 'feature key management protocol'.

Functional

The functional protocol is based on the facility information element and FACILITY message, and requires the knowledge of the supplementary service by the user equipment. It can therefore operate without human intervention, the user equipment processing it on its own. Messages may be exchanged between user equipment and network to synchronize resources between them. The detailed coding of the supplementary services will continue as new services are identified.

5.4 RELATIONSHIP WITH SIGNALLING SYSTEM CCITT No 7

In Section 3.5 the use of CCITT No 7 in the IDN was discussed. The Layer 3 ISDN signalling between the terminal and the local exchange is carried across the network by CCITT No 7 and at the distant end is reconverted to the ISDN signalling. Figure 5.16 shows the message flows of a possible ISDN call set-up. Note that the diagram shows the case where the call set-up information is sent from the customer in an overlap fashion, rather than *en bloc* in a single SETUP message. Transit exchange 2 has sufficient information that it knows how much routing information is needed and hence reassembles it into a single IAM. Some information is only transferred to the terminating exchange when it is ready to receive it, requested by the INR message and returned in the INF message.

5.5 MAINTENANCE

It is essential that the network administration has some means whereby it can centrally test customers' connections. On analogue links to the customer d.c. or low frequency tests are customary. Historically this has been provided over metallic copper test pairs from an exchange test desk.



THE BASIC RATE ISDN CUSTOMER'S INTERFACE





Even for analogue links this is becoming less desirable as it inhibits the use of remote subscriber interface units operating over multiplexed links particularly where the physical connection is provided by optical fibres. When the final link to the customer is digital, the use of metallic pairs to provide a test link is even less satisfactory as digital streams observed over a test link of any length are so modified by the attenuation of the link that useful information is not obtained. For this reason testing of the ISDN is based on the principle of looping back signals from the exchange in the terminal apparatus so that the signals are returned to the exchange. Other loopbacks could return the channel to the user. There they can be observed and any errors monitored.

Figure 5.17 shows possible locations of loopbacks. The dotted loopbacks are regarded as optional but loopbacks labelled 2 and 3 are 'recommended' and 'desirable' respectively. Loopback 2 is in the NT1 as near as possible to the T reference point. It will cause B channels to be looped back and the D channels also, in so far as that is practical. It will be under the control of Layer 1 signal from the exchange. Loopback 3 is as near as possible to the S reference point and may either be used for local maintenance or be controlled by Layer 3 messages in the D channel.

5.6 TESTING AND APPROVALS

As may be imagined, proving that equipment performs correctly to the protocols described in this chapter is important. There are several needs.

(a) The initial specification must be complete, self-consistent and unambiguous. This has always been a serious problem and the tendency is to go to more formal methods to describe the actions of terminal and network, with a mathematically rigorous background. Hence much of the protocol description in the CCITT documentation is in the graphical Specification and Description Language (SDL). More formal is the ISO language LOTOS.

- (b) The equipment must be tested for conformance to the specification. The protocols are so complex that it is theoretically impossible to investigate every possible sequence of messages, but as a minimum every possible path through a call must be explored. However this is not sufficient; it is important that non-permitted sequences be applied to equipment to ensure that no mechanism is available which would cause systems to lock-up or in some other way degrade the network. The non-permitted sequences must be ignored or rejected to protect against data corruption on the line or faulty apparatus. Because of the large number of sequences involved, a powerful processor with software to match and the necessary hardware interfaces is needed. For Layer 1 testing an automated rig of oscilloscopes, power supplies, spectrum analysers and network analysers is needed.
- (c) The equipment performance must be checked. It is not sufficient for equipment simply to conform to specification at low calling rates, but this conformance must be maintained at the design call rate and under overload. It is only too easy for buffers to overflow into one another in ram causing unpredictable results which will not show up in conformance tests. Test gear must be able to operate at such speeds as will show up these effects.
- (d) The approvals regime must be established to ensure that only satisfactory equipment is connected to the network. It may be felt that full conformance is not necessary but as a bearest minimum electrical safety is necessary. Approval to connect to a public ISDN network interface must be obtained by a manufacturer.

In Europe the approval specifications are known as 'Normes Européennes de Télécommunication' (NET). The following NETs are relevant.

NET3—Basic Rate ISDN connection. NET5—Primary Rate ISDN connection. NET7—Terminal Adapters. NET33—Digital Telephony.

The guiding principle under which the NETs are written is the need to ensure that essential requirements are met. These include user safety, the safety of employees of public telephone network operators, the protection of networks from harm and (in justified cases) the interworking of terminal equipment. In the USA approval must be sought under the Federal Communications Commission (FCC) Rules Part 68, the objectives of which are very similar to the NETs. Part 15 covers radio-frequency radiation.

REFERENCES

CCITT Recommendations:

I.430	Basic user-network interface—layer 1 specification
Q.920	ISDN Usernetwork interface data link layer general aspects
Q.921	ISDN User-network interface—data link layer specification
Q.930	ISDN User-network interface layer 3—general aspects
Q.931	ISDN User-network interface layer 3 specification for basic call control
Q.932	Generic procedures for the control of ISDN supplementary procedures
Z.100–104	S.D.L.

Other documents are supplied by network operators which will include details of national variants and options. Examples are:

In the UK-British Telecom Network Requirement-(BTNR) 191-I
series interfaces for ISDN Access.
In Germany—FTZ – Richtlinie 1R6.
In the USA—Bellcore TR-TSY-000268—ISDN Basic Access Call Control
Switching and Signalling requirements.
In France—France Telecom Spec VN2.

NETs are published by CEPT Liasion Office, Seilerstrasse 22, CH-3008 Bern.
Chapter 6

Primary Rate ISDN Access

6.1 BACKGROUND

Domestic and small business customers having a need for a small number of connections at 64 kbit/s will be served with ISDN by the basic rate access described in Chapter 5. However, larger customers may require access to tens or hundreds of connections at 64 kbit/s for PABX or computer bureaux type service; some way of offering service in larger blocks is required. For this reason the interfaces used within the network have been modified and pressed into service. The interfaces operate at 1.544 Mbit/s in North America and 2.048 Mbit/s in Europe and the Layer 1 format is described in Section 3.2. In addition to the use of providing multichannel ISDN access to customer, the primary rate interface may be used to access remote units providing basic rate access to customers (see Section 3.4).

The frame structures for ISDN access are as shown in Figure 6.1 and 6.2. Note that the D-signalling channel is at a rate of 64 kbit/s and is in timeslot 24 in the 1.544 Mbit/s system and timeslot 16 in the 2.048 Mbit/s system. It is possible that one D channel may carry signalling for more than one primary rate interface in which case all 24 and 31 timeslots can be used for B channels. The electrical characteristics of the interfaces are pulses of 3 volts nominal amplitude on a symmetric pair into a resistance of 120 ohms (2.048 Mbit/s) or 100 ohms (1.544 Mbit/s). An alternative interface in use at 2.048 Mbit/s in the United Kingdom is 2.37 volts peak nominal on a coaxial line of 75 ohms impedance. The pulses are half-width; that is to say that each pulse only occupies half the pulse period (see Figure 3.3, unlike the pulses on the basic rate interface which are full width. The coding is HDB3 (2.048 Mbit/s) and B8ZS (1.544 Mbit/s) as described in Section 3.2. It must be emphasized that these are the signals presented to the customer. The signals on the transmission line to the customer may be the same, or may be quite different. For example in the UK a more complex 4B3T line code may be used which, by reducing the modulation rate of a 2.048 Mbit/s signal to 1.536 M baud, can extend the reach available to the customer. Alternatively optical fibre transmission may be used which employs a totally different coding technique.



Figure 6.1

Frame structure of 1.544 Mbit/s primary rate interface.



Figure 6.2

Frame structure of 2.048 Mbit/s primary rate interface.

6.2 SIGNALLING

Differences between the primary rate interface and the basic rate interface are:

- (a) No provision is made in the primary rate interface for multipoint working in the customer's premises; the connection will simply be between the network and a single PABX or other piece of equipment. Of course the individual channels across the interface can be routed independently to whatever destination is required.
- (b) No provision is made in the primary rate interface for deactivating the link to economize in power.
- (c) Any customer who uses as many circuits as provided by the primary rate access may be expected to have both switched and private circuits and it is desirable that the interface be able to support a mix of both.

The first two points simplify design of the signalling protocols. It is relatively straightforward to adopt the basic rate protocols for use in primary rate service. The 64 kbit/s D channels provides additional throughput to support the larger number of channels. Such services were introduced in West Germany in 1988. In the USA AT&T introduced such a service for access to its long haul network services. Australia, Belgium, Finland, France and Japan were close behind, as were the Local Exchange Carriers in the USA. The final point on the mixing of private and public switched connections is more difficult. A particular solution has been adopted in the United Kingdom.

6.3 THE UK SOLUTION

In Section 1.8 mention was made of the pilot ISDN service offered by British Telecom in 1985. Its specification was based on discussions which were taking place at CCITT forums around 1982. Hence they incorporated ideas which were later developed to form the 1984 Red Book and 1988 Blue Book recommendations. The signalling used on the pilot service basic rate access was called DASS (Digital Access Signalling System). This was subsequently developed in two ways.

- (1) As DASS2, to provide for primary rate access to the ISDN.
- (2) As a Digital Private Network Signalling System (DPNSS) for the interworking of PABXs over private circuits.

Owing to their common ancestry the two forms of signalling can coexist in channel 16 of the primary rate access and so private inter-PABX circuits can be mixed with channels connected to the public switched network on the same 30 channel multiplex.

	1 octet	2 octets	1 octet	up to 45 octets	2 octets	1 octet	
	flag	address	control	layer 3 message	FCS	flag	

Figure 6.3

DASS2/DPNSS Layer 2 frame.

The Layer 2 protocols are HDLC based. The frame is shown in Figure 6.3. The signalling operates in a 'compelled' mode. That is to say frames are transmitted continuously until an acknowledgement of correct reception is received from the distant end. Layer 2 frames containing messages on behalf of each of the 30 traffic channels are interleaved on channel 16. Access to channel 16 is given on a cyclic basis, thus giving equal priority to each traffic channel.

Layer 3 provides for call control. In principle the main difference between DASS2 and DPNSS lies in the symmetry of operation. DPNSS intercommunicates directly between PABXs of equal rank in the hierarchy, whereas DASS2 is used to communicate with a large national and international network. This is illustrated by comparing the message flows for a simple call as shown in Figures 6.4 and 6.5. The following abbreviations are used for message contents for simple calls in both DASS2 and DPNSS:

ISRM—Initial Service Request Message. CA—Called Address.



Figure 6.4

Establishment of a simple call-DPNSS.

DA-Destination Address.

SIC-Service Indicator Code (indicates voice, data fax).

NAM-Number Acknowledge Message.

CCM-Call Connect Message.

CRM-Clear Request Message.

Those messages particular to DPNSS are:

CLC—Calling Line Category (e.g. ordinary extensions).

OLI-Originating Line Identity (e.g. number of extension A).

Those messages particular to DASS2 are:

CS—Channel Seize. ICI—Incoming Call Identity. CAM—Call Accepted Message. CIM—Clear Indication Message. EOC—End Of Call.



Figure 6.5 Establishment of a simple call—DASS2.

DASS2 provides for several supplementary services including 'user-touser signalling' where PABXs are able to interact co-operatively, and 'closed-user-groups' so that members of groups are able only to communicate within that group.

DPNSS provides for interconnection to be made between PABXs of different manufacture, so that to the user the PABXs will appear as a single large PABX with all supplementary services available wherever the extension. The services defined are listed in Table 6.1. In practice not all PABX manufacturers are in a position to offer all the supplementary services which appear in the specification, but in the event of a user on one PABX invoking a service which is not available on another manufacturer's PABX, then a sensible fallback option is automatically offered where possible. The ability of DPNSS to handle this situation allows new services to be added progressively without the need to simultaneously upgrade all the PABXs in the network.
 Table 6.1

 Features specified for DPNSS

Simple telephone call	Series call		
Circuit switched data call	Three party takeover		
Swap	Night service		
Call back when free	Centralized attendant		
Executive intrusion	Traffic channel maintenance		
Diversion	Remote alarm reporting		
Hold	Add-on conference		
Three party	Time synchronization		
Call offer	Call back when next used		
Non-specified information	Do not disturb		
Service independent strings	Remote registration of diversion		
Call waiting	Remote registration of do not disturb		
Bearer service selection	Priority breakdown		
Route optimization	Call back messaging		
Extension status call	Loop avoidance		
Controlled diversion	Forced release		
Redirection	Textual messaging		

6.4 INTERNATIONAL STANDARDS FOR INTER-PABX SIGNALLING

DPNSS is the first 'open standard' digital inter-PABX signalling system, and has been implemented by many PABX manufacturers. It is therefore to be found far beyond the UK where it originated. There are other activities to produce inter-PABX signalling. In Europe the European Computer Manufacturers Association (ECMA) have been working on such a standard. Telecom Australia has defined their TeLinc standard. All these are based on the I.451 (Q.931) standard. Several manufacturers have proprietary systems either based on early versions of Q.931 or on CCITT No 7. Eventually they will be superseded by an international standard which is compatible with the basic rate interface signalling standards, but obviously it is desirable to develop an evolutionary path such that private networks can contain both systems simultaneously during a transition period. This could be achieved by ensuring that the Layer 3 information flows are compatible.

REFERENCES

CCITT Recommendation:

I.431—Primary rate user-network interface—Layer 1 specification. I.451 (Q.931)—ISDN User-network interface—Layer 3 specification for basic call control.

British Telecom Network Requirement (BTNR) 190-DASS2.

British Telecom Network Requirement (BTNR) 188-DPNSS.

Chapter 7

Frame Mode Services

John Atkins

Early demand for switched data services, during the 1960s and early 1970s, arose largely from the high cost of computers. To make most effective use of the expensive data processing equipment, time-sharing by remote terminals became an attractive choice for many users. The ubiquitous public switched telephone network (PSTN) was pressed into service to provide the necessary switched access. As computer technology developed, however, it became clear that the PSTN would not be adequate for many of the new applications and that dedicated switched data networks would be required.

In 1967 Donald Davies at the National Physical Laboratories in the UK and Larry Roberts of the Department of Defense in the USA independently proposed packet switching as the most economical technique for switched data networks. The basic ideas for packet switching had been developed several years earlier by Baran and co-workers at the Rand Corporation to meet military requirements, but their work was shelved and apparently forgotten.

7.1 STORE-AND-FORWARD SWITCHING

Communication between computers or between computers and terminals usually involves the transfer of 'blocks' of data. Packet switching exploits the idea that data blocks may be transferred between terminals without setting up a continuous end-to-end connection. Instead they are transmitted on a link-by-link basis, being stored temporarily at each switch *en route* where they are queued for transmission on an appropriate outgoing link. Routing decisions are based on control information contained in a 'header' prefixing each data block. The term 'packet' refers to the header plus data block.

In fact the idea of switching by storing and forwarding is older than packet switching. It had been used in a variety of forms to provide message switched services in which users could exchange complete messages--- often very long messages—with the advantages of delayed delivery and broadcast options and retransmission if the message was garbled or lost in transmission. The distinctive feature of packet switching is that the packets, and consequently the queueing delays, are sufficiently short to permit interactive transactions.

7.2 DATAGRAMS AND VIRTUAL CIRCUITS

The simplest form of packet-switched service is the 'datagram' mode in which each packet is regarded as a complete transaction in itself. The network provides an independent switched path for each datagram. This implies that the header contains a complete set of information to route the datagram packet and that the network will have no pre-awareness of the connection.

Many transactions, however, involve the transfer of sequences of packets for which the 'virtual call' mode is more appropriate. In this case a virtual circuit is established by an initial exchange of signalling packets between the communicating terminals and the network. Full addresses are only used for this initial call set-up phase. The setting up of a virtual call has the advantage that headers may be simpler, only indicating the virtual call identity. The disadvantage is that each node has to keep records of the routing relating to the call. The existence of two types of packet (signalling and data) means that each packet must be inspected to see whether it is for network control or for the customer. Data transfer then takes place and at the end of the transaction the call is cleared. During the data transfer phase the network tries to create the illusion of a real 'connection' whilst achieving the economies of dynamically shared resources.

Although the virtual call mode of operation is in general use, the debate continues as to the merits of the two manners of operation.

7.3 FLOW AND CONGESTION CONTROL

In a packet-switched network packets compete dynamically for the network resources (i.e. buffer storage, processing power, transmission capacity). A switch accepts a packet from a terminal largely in ignorance of what network resources will be available to handle it. There is therefore the possibility that a network will admit more traffic than it can carry, with a corresponding degradation in service. Controls are therefore needed to ensure that such congestion does not happen too often and that the network recovers gracefully when it does.

The importance of good flow control and congestion control is illustrated in Figure 7.1. For a given delay performance a packet-switched network may be regarded as having a notional maximum throughput, T packets/ second. An ideal network would accept all traffic up to its throughput limit; beyond this some traffic would be rejected but the network's throughput would remain high (curve A). In a network with poor flow and congestion control, however, the throughput, whilst initially increasing with offered traffic, would eventually fall rapidly as as result of congestion (curve B). In practice good flow and congestion control maintain high throughput whilst leaving an adequate margin of safety against network



Figure 7.1 The importance of flow control.

failures and the imperfect operation of practical algorithms (curve C). Any flow control imposed by the network is in addition to end-to-end flow control applied by the terminals to match differences in speed. In fact endto-end flow control is itself effective in avoiding network congestion since any increase in cross-network delay caused by the onset of congestion will delay the return to the sending terminal of permission-to-send, effectively reducing the rate at which it may send further packets.

The design of flow and congestion controls that are both effective and fair is difficult and remains the subject of active research. In general, flow and congestion control specify rules for the allocation and reservation of buffer storage and the selective rejection of incoming packets by switches when certain allocation thresholds (e.g. queue lengths) are exceeded.

7.4 STANDARDS

Growing interest in packet switching stimulated vigorous activity in CCITT, and in 1976 an international standard was agreed for packetswitched services. This is Recommendation X.25, which has provided the foundation for the development of Packet Switched Public Data Networks (PSPDNs) around the world and shaped the thinking of a generation of switching engineers. It is a virtual circuit protocol, structured after the ISO OSI 7-layer model in three layers, as shown in Figure 7.2.



Figure 7.2 The three levels of X25.

7.5 SUPPORT OF X.25 BY AN ISDN

The ISDN arrived in an environment in which PSPDNs already existed. During the early years therefore ISDNs will depend largely on these separate PSPDNs to support packet mode bearer services. A comprehensive standard has been developed by CCITT which sets up arrangements for interworking between an ISDN and a PSPDN. This standard, CCITT Recommendation X.31 (also designated I.462), takes X.25 as the standard packet layer protocol, either in an X.25 terminal with a terminal adapter or in conjunction with the I.420 user-network interface to form a packet-mode terminal equipment. In this case, the ISDN simply provides a transparent 64 kbit/s circuit between the packet terminal and a packet port on the associated PSPDN. This access circuit may be a semipermanent or switched on demand B channel (Figure 7.3).

In addition to access to the services of a PSPDN, X.31 also defines what are called ISDN Virtual Circuit Bearer Services based on X.25 packet handling capability contained within the ISDN itself. Access to the packet handler may be obtained either via a user's B channel or via the D channel as shown in Figure 7.4, and again may be either semi-permanent or switched on demand. The packet handler provides interworking with a PSPDN. This interworking may use the standard protocol X.75 which is designed for internetwork use, or may use a proprietary equivalent.

Depending on the application, packet access via the D channel may offer a number of distinct advantages, including:

- -unlike the B channels the D channel is essentially non-blocking;
- --- it leaves the B channels available for other services.



Figure 7.3 Access to PSPDN services.





But D channel access also imposes some service limitations. The 16 kbit/s basic access D channel cannot support higher data rates and LAPD (see Section 5.2) restricts X.25 packets to a maximum of 260 octets compared with 1024 for X.25. Furthermore, the additional delay caused by the slower D channel may be significant in some applications.

7.6 NEW ISDN FRAME MODE SERVICES

The term 'data' is traditionally associated with terminal-to-computer or computer-to-computer communications. But as new applications emerge and penetrate the workplace (and in due course the home) the scope for 'data' services is growing rapidly and it is becoming more appropriate to talk of non-voice services rather than data. Indeed, with the introduction of integrated multi-media applications even the distinction between voice and non-voice may become spurious.

Many of the new applications are 'bursty' in their requirement for transmission capacity (e.g. variable-bit-rate video) or involve shortduration transactions. Applications differ also in their sensitivity to delay or variations in delay, and in their ability to tolerate packets lost or corrupted because of transmission errors or congestion. There is also a general trend towards higher data rates (e.g. for video applications or LAN-LAN interconnection).

X.25 was developed specifically as a robust packet transfer service for 'traditional' data applications, but its complexity and lack of transparency (due to timeouts and flow control interactions) mean that it cannot support



Figure 7.5 ISDN switched mode services.

many of the new applications. It is equally clear that the ISDN should support new forms of frame mode bearer service that are able to offer the necessary flexibility. To this end a new family of ISDN frame mode bearer services is defined by CCITT.

One of the key features of the ISDN is the clear logical separation of signalling information and data. In CCITT parlance signalling belongs to the control-plane or C-plane, whilst users' information belongs to the user or U-plane. In effect the ISDN may be viewed as two logically separate subnets—a signalling subnet and a switched information subnet, as shown in Figure 7.5. The new ISDN frame mode bearer services follow this established ISDN principle. Virtual calls are set up and controlled using the same out-band ISDN signalling procedures as circuit mode connections, unlike X.25 which carries signalling and data inseparably in the same logical channel.

Two distinct ISDN frame mode bearer services have been defined, frame relaying and frame switching. The same signalling procedures are used for both, but they differ in the protocol supported by the network in the user-plane during the data transfer phase. A user may set up a number of virtual circuits and/or permanent virtual circuits simultaneously to different destinations.

Frame relaying

Figure 7.6 shows the protocol architecture for frame relaying, the simplest of the new frame mode services. Signalling in the C-plane uses the standard ISDN signalling at Layers 2 and 3 enhanced to support the frame mode parameters, the same as for circuit mode services. But in the user-plane the network supports only a very small part of the link layer protocol, generally referred to as the 'core' functions. These are:

- -separating the frames using HDLC flags;
- -forcing data transparency by zero bit insertion and extraction;
- checking frames for transmission errors and length—any frames found to be in error are discarded;
- -multiplexing and demultiplexing the frames associated with different virtual calls on the basis of their Layer 2 address. Many virtual calls may be supported simultaneously to different destinations.

The basic service provided by frame relaying is the unacknowledged transfer of frames between ISDN terminals. At its simplest the network creates an entry in a routing look-up table at call set-up time using the



Figure 7.6 Frame relay frame mode.

signalling procedures, a Layer 2 address being allocated to each virtual call for its duration. In the data transfer phase the network simply indexes the routing table using the Layer 2 address of the incoming frame, and queues it for transmission on the appropriate outgoing route.

In the data transfer phase terminals operate higher layer protocols on an end-to-end basis without involving the network.

Frame switching

The protocol architecture for frame switching is shown in Figure 7.7. Again signalling in the control-plane follows standard ISDN procedures but in this case the network operates the complete Link layer protocol in the user-plane during data transfer. Frame switching therefore provides acknowledged transfer of frames between terminals. The network detects and recovers from lost or duplicated frames and frames with errors and it operates network-enforced flow control. As with frame relaying, multiplexing, demultiplexing, and routing at each switching node is done on the basis of Layer 2 address. Again, higher layer protocols are operated by the terminals on an end-to-end basis.



Figure 7.7 Frame switching frame mode.

The new ISDN frame mode bearer services represent a significant extension to the capabilities of the emerging ISDN. The architectural integration of frame mode services with the complementary circuit-mode services will provide the flexibility needed to meet the growing range of users' requirements. The lightweight data transfer protocol of frame relaying will support applications which cannot tolerate the additional delays that would be caused by error-correction-by-retransmission. And it means that frame relaying will provide a high degree of protocol transparency, there being a minimum of interaction with higher layer protocols. Frame switching on the other hand will offer the robustness against errors usually associated with X.25 but without the per-packet overhead of Layer 3.

REFERENCES

The standard adopted for most public packet-switched networks-CCITT recommendation:

X.25 Interface between data terminal equipment and data circuit terminating equipment for terminals operating in a packet mode and connected to public data networks by dedicated circuit.

There are three other standards relating to the connection of terminals, often referred to as the 'triple-X' protocol:

- X.3 Packet assembly/disassembly facility (PAD) in the public data network.
- X.28 DTE/DCE interface for a start-stop mode data terminal equipment accessing the packet assembly/disassembly facility (PAD) in a public data network situated in the same country.
- X.29 Procedures for the exchange of control information and user data between a packet assembly/disassembly (PAD) facility and a packet mode DTE or another PAD.

X.25 on the ISDN is described in:

X.31 Support of packet mode terminal equipment by the ISDN.

The internetwork protocol is:

X.75 Terminal and transit call control procedures and data transfer system on international circuits between packet-switched data networks.

New ISDN frame mode services:

- I.122 Framework for providing frame mode bearer services.
- 1.232 Frame mode bearer service categories.

Chapter 8

ISDN Customer Premises Equipment and Applications

The ISDN as described so far only considered the provision of control and traffic channels for customers use. In CCITT terms this is equivalent to an ISDN bearer service, defined as follows:

A provision of such a service to the customer allows him the capability for information transfer between ISDN access points according to CCITT standards. The higher layer (i.e. above Layer 3) terminal functions are defined by the customer.

That is to say the customer is free to make whatever use of the 64 kbit/s bearer he or she may wish, without constraint. However in practical terms such liberalization leads to anarchy rather than communications. It would be farcical, for example, for every manufacturer to choose their own technique of 64 kbit/s speech encoding so that only customers using equipments of the same manufacture could speak to one another. Nevertheless such situations have arisen in the past between facsimile equipments and other video and data exchange equipment. To make maximum use of the bearers it is necessary that the higher layers of the customer-to-customer protocol are defined so that interworking is possible. Such a defined service is known by CCITT as a 'teleservice', defined as:

This type of service provides the user with the complete capability, including defined terminal equipment functions, for communication with another user of the service according to protocols established by CCITT.

The intention of the CCITT in this categorization is to distinguish services according to what technical standards are needed for their specification and implementation such that international interconnection is possible. It was not the intention to categorize services according to different possible arrangments of equipment ownership or the provision to the customer of the means to support the services.

By examining the functional picture of the access arrangement to the ISDN (Figure 8.1), it is possible to have a clearer understanding of the differences and the relationship between these two categories of services.





Customer access to services supported by an ISDN.

Access points 1 and 2 are the access points for the bearer services supported by an ISDN. At access point 4, depending on the types of terminal adapters provided, other CCITT standardized services may be accessed. These could be called dedicated network bearer services such as the circuit-switched or packet-switched data services associated with usernetwork interfaces according to Recommendations X.21 or X.25 respectively.

Access points 3 and 5 are the access points for teleservices. In this chapter we shall look at the applications of ISDN. In particular we shall highlight those applications which particularly benefit from the availability of 64 kbit/s bearers.

Services can benefit at three levels from the availability of a 64 kbit/s bearer.

- 1. Those which benefit in a quantitative way but whose principles are basically the same as when operating at lower speeds. The obvious examples are file transfer and electronic mail. Changing from 9.6 kbit/s modem based bearers to 64 kbit/s bearers means that the transfer rate is increased almost seven-fold and times correspondingly reduce. Nevertheless the principles are the same. Note that it may be that the seven-fold improvement is not readily achievable as error control protocols which are optimized for the lower data rates may be less than optimal at higher speeds owing to software limitations, and the fact that the number of bits delay in the link is much higher on the ISDN connection both due to the higher bit rate and the switching mechanisms employed.
- 2. Those which benefit to such a degree that the totally new uses of the application appear. Examples are:

(a) Facsimile transmission is obviously possible over the analogue network, but the 20 or 30 seconds per page means that it is not an

attractive proposition for more than a few pages. On ISDN, at 64 kbit/s, pages can be transmitted in 3 or 4 seconds and this moves the processes into the realm of the 'slow photocopier' where tens of pages could be transmitted with little effort. It is arguable that in practice the ISDN facsimile service will be applied entirely differently to the analogue service.

(b) Videotex services such as Prestel (UK), Minitel (France), Bildschirmtext (Germany) will be very different services when they are moved from 1200 bit/s to 64 kbit/s. Pages can be written instantly and browsing is possible. Colour pictures can be included in the pages which adds a whole new dimension for marketing and information distribution.

3. Those whose provision is totally dependent on the availability of 64 kbit/s links, which are not realizable at lesser data rates or on analogue bearers. Particular examples:

(a) Videophone whose performance may be acceptable at 64 kbit/s, which at lower rates does not at present appear usable.

(b) High quality speech which uses a coding algorithm which is more complex than PCM to convey better quality speech, or even music. No equivalent service is possible over analogue networks.

As discussed earlier in the chapter, such applications must be standardized so that terminals from a wide range of sources can interwork. In the next sections the operation of systems in categories 2 and 3 above will be discussed.

8.1 HIGH QUALITY SPEECH

Paul Challener

The process of encoding speech into PCM is described in Chapter 2. PCM is a simple technique consisting of filtering to avoid the overlapping of sidebands in the sampling process (known as aliasing), sampling at an 8 kHz sampling rate and 256-level quantization. At each sampling instant, the quantizer outputs an 8-bit codeword corresponding to the quantization level chosen. Hence the overall bit-rate is 64 kbit/s. The coded signal bandwidth is 300 Hz to 3.4 kHz which results from the anti-alias filter. This may not be seen as particularly speech specific, but the design of the quantizer does reflect understanding of the speech signal. The 8-bit codewords result from a 12-bit to 8-bit compression following a logarithmic law (A-law in Europe and μ -law in USA), such that the quantization noise encountered on coding large signal segments is generally larger than that encountered when coding small signals. Thus the signal-to-noise ratio

remains approximately independent of the signal level. The advantage of this technique for coding speech is that the signal hides most of the quantization noise so that it is not perceived by the listener; this is known as 'auditory masking'.

Algorithms have been developed which exploit redundanices in the speech signal more than PCM, resulting in further reductions in the bitrate necessary to code telephone quality speech. In the last few years such techniques have also been applied to wide bandwidth, high quality speech signals in an attempt to code them into the standard 64 kbit/s channel of ISDN. This has reached acceptance in a CCITT standard G.722 which covers conversion of a 7 kHz bandwidth speech signal into a 64 kbit/s bearer. The rest of this section is intended to provide an understanding of this standard with an insight into the techniques employed and the properties of speech which are exploited in achieving them. However, as the workings of a technique known as Adaptive Differential Pulse Code Modulation (ADPCM) are central to an understanding of the standard, this will be described first.

8.1.1 Speech coding using ADPCM

A major step forward in coding speech more efficiently than PCM came in 1984 when ADPCM was accepted by the CCITT for coding telephone quality speech at 32 kbit/s (G.721). In this standard the target quality was identical to that of 64 kbit/s PCM on a subjective level. In other words, it would not matter if there was a measurable decrease in signal-to-noise ratio if this could not be perceived by the listener.

The principal differences between PCM and ADPCM are contained in the 'A' and the 'D' which stand for 'adaptive quantization' and 'differential encoding' respectively.

An 'adaptive quantizer' differs from the standard PCM quantizer in two significant ways. First, at any given sampling instant, all the quantization levels are uniformly spaced with an interval S; they are not logarithmically spaced. Secondly, when signals are quantized towards the limits of the quantization range the step size S is increased; when they are quantized towards the centre of the range, S is decreased; elsewhere the step size hardly alters. As speech signals from a single talker do not change in level significantly from one word to the next, this kind of quantizer can quickly adapt to a range suitable for the actual signal conditions and then remain almost perfectly scaled until the speaker stops talking. Figure 8.2 illustrates the operation of the adaptive quantizer. Speech which is initially linearly encoded into 12 bits (i.e. 4096 levels) is then compared with 16 quantizing levels spaced S apart and the nearest quantization level is chosen to





represent the speech. As there are 16 levels, the chosen level can be encoded as 4 bits.

The expansion or compression of the quantization range is achieved by simply multiplying S by a number slightly greater than unity or slightly less than unity dependent only upon the quantization level selected to code the previous signal sample. Such a dependence only upon past events to determine current values is termed a 'backward adaptive process'. The reason for this name can easily be seen from Figure 8.2 where the direction of the stepsize adaptation process opposes the normal signal flow through the coder.

Backwards adaptation is an elegant technique which can be simultaneously performed at the encoder and decoder without the need for external side information to keep the processes in step; they are each driven from the output of the coder. Furthermore, it is possible to include the concept of 'leakage' in the quantizer to improve its robustness to channel errors; this means that in addition to the control on S exerted by the output of the decoder, S is also multiplied by an additional factor to cause it to tend to a fixed value. This ensures that departures from synchronization are leaked away to zero. In this way, the encode and decode processes will automatically align themselves at start-up or after bursts of errors have corrupted the transmitted signal values. Hence all the signal bits fed to line are truly used to encode speech efficiently, not to control terminal processes, and the result is a good mixture of simplicity, robustness and elegant design. 'Differential encoding' implies that the quantizer input is not obtained from the signal itself, but from the difference between current signal values and previous ones. In ADPCM, this concept is progressed further by the inclusion of a predictor circuit in the differential loop. Figure 8.3 shows this; the outer loop implements the differential process. Predictors are adaptive filter circuits which use previous sample values to construct an estimate of the signal at the current sampling instant. When the real current value arrives at the coder input, the estimate is then subtracted from it leaving only a tiny difference signal which is fed to the quantizer. Hence, the differential process reduces the size of the signal fed to the quantizer and thus enables fewer bits to be used in the quantizer giving rise to a second significant bit-rate reduction over standard PCM.



Figure 8.3 32 kbit/s ADPCM encoder.

The differential prediction process is also a backwards adaptive process which again requires no side information to keep the decoder in synchronism with the encoder. Clearly, the closer the predictor estimate is to the actual current signal sample, the smaller the size of the signal fed to the quantizer and hence the fewer bits needed to encode it. The theory of predictors is a difficult and demanding task which is outside the scope of this book. Suffice to say that some knowledge of the signal statistics is necessary in order to design the predictor filter appropriately. In the case of 32 kbit/s this knowledge is obtained by modelling the signal spectrum by a transfer function whose six poles and two zeros are determined by the long-term and short-term characteristics of the signal respectively. The property of speech being exploited is that it is highly predictable for much of the time; departures only occur at boundaries between the elements of speech which constitute real information. On ISDN, the B channel rates are 64 kbit/s and so there is little benefit in using 32 kbit/s ADPCM for single channel telephony applications. However, when higher quality audio is required, there is clearly some scope within the 64 kbit/s bit-rate to widen the bandwidth and improve the dynamic range outside the limits of 64 kbit/s PCM. However, before considering what techniques to use the first question is what signals do we wish to transmit?

There are obviously applications for systems which transmit compact disc quality audio (50 Hz to 20 kHz), but such a target would currently be impossible to achieve within a single 64 kbit/s B channel. However, in many possible ISDN applications such as conferencing or commentary broadcasting, the only audio signal is speech, and tremendous benefits can be achieved if this is coded to a subjectively high quality. So the next question to consider is what is the real bandwidth of speech?

The answer to this surprises many as it does not match the auditory bandwidth. Speech coding experts generally consider it to be about 50 Hz to 7 kHz. The low frequency limit is due mostly to the presence of low pitch frequencies in some voice speech, being particularly relevant in 'fruity' voiced male speech. The high frequency limit is largely due to limits on the mechanics of the human speech production equipment. It requires muscle control to move the tongue, lips and teeth to create plosives or stops (e.g. 'B' or 'T') and hence these do not contain very high audio frequencies. Only the fricatives (hissing sounds like SSS, FFF) contain energy towards the 7 kHz region and even then the power at these frequencies is small. Hence, if the spectrum from 50 Hz to 7 kHz can be faithfully reproduced we can consider we have achieved a high quality speech system. This was the target which the CCITT set themselves in the late 1980s and the result was the standard G.722. It was so successful that it proved possible to generate additional data channels alongside the audio with little or no apparent degradation to the speech signal. The operation of this standard is described below in conjunction with Figure 8.4.

Input speech is first sampled at 16 kHz and digitized to at least 14 bits before being fed into the first digital signal processing block marked 'Transmit QMF' in Figure 8.4. 'QMF' stands for Quadrature Mirror Filters and consists of a pair of digital bandpass filters which separate the signal into two bands, a low band and a high band, with frequency ranges 50 Hz to 4 kHz and 4 kHz to 7 kHz. They are not 'brickwall' filters but deliberately overlap each other in the 4 kHz region in a predetermined manner. The two bands are then downsampled (by taking alternate samples) to 8 kHz as each contains less than 4 kHz signal bandwidth and the reduced sample rate signals are then coded separately. Some slight aliasing occurs because of the QMF filter overlap at 4 kHz but this will not ultimately cause





concern. In each band, the coding process is ADPCM, but in the case of the low band, where most of the speech energy lies, 6 bits are allocated to the quantizer whereas for the high band only 2 bits are allocated. (Hence the 48 and 16 kbit/s data streams shown in Figure 8.4 for the two bands.) These two datastreams are combined in a multiplexer and sent to line at a composite 64 kbit/s rate.

In the decoder, the reverse processing occurs-demultiplexing of the two bands, decoding the ADPCM in each and then upsampling to a 16 kbit/s sampling rate and finally recombining through a receive OMF filter bank. Cleverly, in this final process, the alias terms which were introduced on downsampling are cancelled and the result is very high quality speech. An interesting fact is that this process could never have been performed with purely analogue processing; the degree of alias cancellation is a result of precise filter characteristics implemented to a tolerance beyond the dreams of analogue designers; they are also not subject to drift through ageing or temperature variation and are therefore truly a product of the digital era! Hence, in addition to the properties of speech exploited by pure ADPCM, sub-band ADPCM allows us to take account of the spectral shape of speech, allocating most bits to parts of the spectrum with most energy and less bits to those with least, without affecting the overall bit rate. Furthermore, any quantization noise which occurs due to quantizing a signal in one band will be filtered by the QMF filters so that it also lies in that band. In this way, the noise will tend to be well masked by the signal which caused it as they must be reasonably close in frequency.

8.1.3 Embedded data channel

So much for the high quality speech signal, but what about the low bit rate data signals mentioned previously? This is the final part of the 7 kHz algorithm and is often seen as the icing on the cake! However, it should also be seen as the vital key to applications on ISDN. One example is in an audio or video conferencing, where in the real world participants will often arrive with last-minute paperwork relevant to the discussion. This could easily be distributed to all participants if only fax were available alongside the conference; the algorithm will permit fax signals to be sent on the single B channel whilst the conference progresses with no interruption to the audio channel. A second example in outside broadcasting occurs due to the need for a 'control circuit' independent of the broadcast channel. Control circuits are used to allow direction of the outside broadcast team independently from the primary circuit which is 'on air'; the data channel permits low bit-rate speech to be relayed without affecting the primary circuit.

In Figure 8.4, it can be seen that the ADPCM encoder for the low band is more complex than that for the high band. The technique adopted to provide the data channel is to steal one or two of the 6 bits from the low band at each sample instant, giving rise to an 8 kbit/s or 16 kbit/s data channel. The multiplexer allows external data to be inserted when these channels are available and ensures appropriate interleaving in specific modes of operation. The trick employed here is that the low band ADPCM coder is what is called an 'embedded' coder. This implies that whether external data is present or not, the most significant bits of the speech low band signal are always present in the same positions in the multiplex. Only the least significant bits are dislodged to allow external data to be sent and then only for the time when external data is actually present; once the external data input ceases, all 6 bits per sample are given back to speech. In this way a flexible speech and data system is ensured and even if the multiplex mode is misaligned due to transmission errors, the received speech is only affected in least significant bits which are scarcely audible.

Another particularly important reason for providing external data channels is to support sophisticated signalling protocols which will allow complex terminal interaction with no need for human intervention. Protocol G.725 allows terminals to rapidly synchronize to incoming signal streams, adjust their modes appropriately and confirm successful communication to the originating terminal. If appropriate responses are not received by the originating terminal it can be made to switch to simple PCM communication as a fall back.

What of the future? Research activity around the world is focused on efficient coding of high-quality music signals; a number of algorithms are now available for coding 15-20 kHz music into 128 kbit/s data streams. However it seems unlikely that this will result in 64 kbit/s CD-quality audio.

8.2 FACSIMILE

Malcolm Jones

Facsimile image transmission, a method of transmitting unstructured images using a raster-scanning technique, pre-dates telephony by several decades. It was first patented in 1843 by Alexander Bain as a development of the master-slave-clock time distribution system which provided a method of synchronizing two pendulums over a communications channel.

The image transmitting station scanned an electrically conducting image, such as a page of metallic type, with a stylus on the end of its pendulum and the receiving station applied ink to paper from the end of a synchronized pendulum. In 1850 the pendulums were replaced by the cylinder and screw scanning mechanism and in 1902 the photo-electric scanning technique was patented by Dr Arthur Korn of Germany. By 1910 there was an international facsimile service between London, Paris and Berlin using Korn equipment.

Until the 1960s the market for facsimile equipment was small and specialized (e.g. phototelegraphy for the newspaper industry). In 1968 the CCITT produced a recommendation for group 1 facsimile equipment, as distinct from phototelegraphy, which defined an analogue FM transmission system for sending a grey-scale image of a page via the PSTN in 6 minutes at a vertical resolution of 3.85 scan lines per millimetre. In 1976 an improved system was defined which reduced the transmission time to 3 minutes using a special analogue VSB transmission technique. This was followed by the definition of the first digital form of facsimile, group 3, in 1980.

Group 3 facsimile broke with the tradition of previous CCITT standards by not defining a method of supporting grey-scale image transmission. The group 3 image is first quantized on a pel-by-pel basis (pel = picture element) into one of only two 'colours', either black or white. This dichotomization allowed a run-length coded compression scheme to be applied to the picture data which resulted in a reduced page transmission time of 1 minute when using a 4800 bits per second communication system. Group 3 offers a digital horizontal pel resolution of 1728 pels per 215 millimetres or 8.03 pels per millimetre. The basic vertical scanline resolution is 3.85 lines per millimetre as in the earlier groups but there is an optional higher vertical resolution of double the basic value, 7.7 lines per millimetre.

The introduction of a compression technique into any system leads to some new problems. These are principally the magnification of the effect of communication errors and the possibility that the transmitter may over-run the receiver. To deal with the error extension problem in group 3 the compression algorithm is designed to operate on only a very limited number of scanlines at a time and each compressed scan line is terminated with a long robust end-of-line code. The number of scanlines processed at any given instant is referred to as the k factor and is basically 1 but may be negotiated between the terminals up to a maximum of 4. This small number is necessary because the original version of group 3 did not include any error correction in its communication system. Despite this limitation the compression system is reasonably efficient giving compression factors between 7 and 12:1 on most documents.

The communication system for group 3 facsimile operated at 4800 bit/s but optional speeds of 9600, 7200 and 2400 bit/s were also defined. When operating at 9600 bit/s a single A4 page image takes approximately 30 seconds to transmit depending upon the complexity of the image. One factor which has contributed to the success of group 3 was the introduction of the flat-bed scanner which replaced the old cylinder and screw mechanism and allowed multi-sheet scanners to be built into facsimile machines at reasonable cost.

The group 3 facsimile terminal is one of the most successful telecommunication terminals ever devised with the number installed being several times greater than the telex terminal population, for example. Facsimile is now the second service on the PSTN after telephony and it is generating very large volumes of traffic particularly on international routes where the office hours at each end do not overlap. The figures for 1989 suggest that some 10 million terminals had been installed with a growth rate over 30 per cent per annum and no sign of market saturation.

8.2.1 Group 4 facsimile for ISDN

When a new facsimile group was being discussed for use on the ISDN in the early 1980s this phenomenal success was not yet visible and a single solution was being sought for all kinds of non-voice traffic under the generic title 'Telematics'. This unified approach to telecommunications services matched the universal concept for communications protocol design as defined by the International Standards Organization (ISO) in their 'Open Systems Interconnection' (OSI) concept. Thus group 4 facsimile was defined in 1984 as one of a family of services which also included basic teletex with its inherent access to the telex service, and mixed mode, a form of document coding which supported both character and fascimile image blocks within a single page. All these services used the same communications protocol stack and only differed at the application layer. Teletex and group 4 facsimile had interworking arrangements included by creating three classes in the group 4 definition.

Class 1 only supports basic facsimile without interworking to teletex.

- Class 2 is compatible with class 1 but it is also capable of receiving basic teletex and mixed mode calls.
- Class 3 is compatible with class 1 but it is also capable of both transmitting and receiving basic teletex and mixed mode calls.

All types of Telematic terminal employ a 7-layer communications protocol stack similar to the ISO's OSI.

8.2.2 Group 4 compression algorithm

The group 4 facsimile system is fully digital in all respects including having an error-correcting communications protocol. Therefore it can employ a redundancy removing compression algorithm. The coding concept is based upon the identification of run lengths of the same colour between adjacent scan lines and is therefore referred to as two-dimensional coding or 2D coding. It is also called Modified Modified Reed (MMR) coding.

The coding of the run lengths is simply performed by selecting variable length Huffman codes. The principle of Huffman coding is to allocate short codes to those codewords which occur most often and longer codes to less frequently occurring codewords. If there is any redundancy in the original sequence then the resulting data is reduced compared with the original. For a sequence of known statistics an optimum code translation can be identified. In this case a pre-defined dictionary is used. The dictionary is made up of two parts. One part contains the short run-lengths from 0 to 63 pels in steps of one pel which is referred to as the 'terminating codes table' and the second part contains the make-up codes for run-lengths from 64 to a full scanline of 1728 pels, in the A4 paper case, in increments of 64 pels. The dictionary contents are pre-defined and fixed. There is no dynamic updating process.

Identification of the run lengths is performed in three different ways depending upon the precise pel pattern and its position on the page. Note that facsimile scanners overscan the paper, to ensure that all the image is captured, and therefore all scan lines start with a white interval plus the first scan line is an imaginary all white line.

The three coding modes are referred to as:

- (a) Pass mode (see Figure 8.5). This mode is identified when the position of b_2 lies to the left of a_1 . However, the state where b_2 occurs just above a_1 , is not considered as a pass mode.
- (b) Vertical mode (see Figure 8.6). When this mode is identified, the position of a_1 is coded relative to the position of b_1 . The relative distance a_1b_1 can take on one of seven values V(0), $V_R(1)$, $V_R(2)$, $V_R(3)$, $V_L(1)$, $V_L(2)$ and $V_L(3)$, each of which is represented by a separate codeword. The subscripts R and L indicate that a_1 is to the right or left respectively of b_1 and the number in brackets indicates the value of the distance a_1b_1 .



Figure 8.5 Pass mode.



Figure 8.6

Table 8.1

Vertical mode and horizontal mode.

(c) Horizontal mode. When this mode is identified, both the run lengths a_0a_1 and a_1a_2 are coded using the codewords $H + M(a_0a_1) + (Ma_1a_2)$. H is the flag code word 001 taken from the two-dimensional code table. $M(a_0a_1)$ and $M(a_1a_2)$ are codewords which represent the length and 'colour' of the runs a_0a_1 and a_1a_2 respectively and are taken from the appropriate white or black run-length code tables.

The coding procedure is shown in the flow diagram, Figure 8.7. The coding procedure identifies the coding mode that is to be used to code each changing element along the coding line. When one of the three coding modes has been identified according to Step 1 or Step 2, an appropriate codeword is selected from the code table given in Table 8.1.

A further feature of group 4 is the inclusion of an extension mode which allows a portion of the page to be coded in an uncompressed format. There

Code table				
Mode	Elements to be coded b_1, b_2 ntal a_0a_1, a_1a_2		Notation	Codeword
Pass			 Р	0001
Horizontal			н	$001 + M(a_0a_1) + M(a_1a_2)$
Vertical	a_1 just under b_1	$a_1b_1 = 0$	V(0)	1
	a_1 to the right of b_1	$a_1b_1 = 1$	V _R (1)	011
		$a_1b_1 = 2$	V _R (2)	000011
		$a_1b_1 = 3$	V _R (3)	0000011
	a_1 to the left of b_1	$a_1b_1 = 1$	V _L (1)	010
		$a_1b_1 = 2$	V _L (2)	000010
		$a_1b_1 = 3$	V _L (3)	0000010
Extension			-	0000001xxx



Figure 8.7 Coding flow diagram.

is no mandatory requirement to switch into uncompressed format if the compression algorithm is actually producing an expansion of the data rather than a compression. The extension mode could be used to introduce other types of coding but so far no others have been standardized.

Only pel transmission density is defined, as the resolution capabilities of the actual mechanisms used to create or to print the image are not restricted to allow the maximum freedom of implementation. It is mandatory to provide 200 lines per inch in both dimensions, and optional to provide 240, 300 or 400 lines per inch. Terminals supporting optional pel transmission densities are required to provide a pel density conversion algorithm which results in low impairment to the image. The basic image format is ISO A4 in vertical orientation and a reproducible area which allows for both European and American paper sizes. (European = 210 mmwidth by 297 mm length, American = 215.9 mm width by 279.4 mm length.) The guaranteed reproducible area is the area of the original image that is reproduced at the receiver and is not lost due to mechanical tolerances, mainly at the printing station, including the tolerance in the scanning and printing resolution of ± 1 per cent. This area is the same for both group 3 and group 4 facsimile. In the case of European paper it is 196.6 width \times 281.46 length in mm.

This close approximation to a full European page is achieved by overscanning the original by 37 pels on both edges, at 200 lpi, and by treating the input paper as being of unknown length until the end of the page is detected, which is usually at the end of the scanning process. The full overscanned image of 1728 pels width is transmitted to the receiver. The unlimited length of page assumption is not a problem for a receiver equipped with a roll-fed printer but for a cut-sheet printer there may be a need to print the image on two sheets occasionally. When an American size image is being received, 215.9 mm width by 279.4 mm long, there is no problem with the length and the full overscanned image width is available at the receiver but the worst case printable area is still the same, therefore there may be some loss of edge detail. However, as the image is centre-line referenced this small loss is not usually a problem and machines are available which make a feature of being able to print the full image width using wider paper, for example.

A terminal identification is required which enables each terminal to be uniquely identified. This TID is exchanged between the terminals at the start of every connection. It may be used to form the call-identification-line which may be printed at the top or bottom of the first or every page of the transmitted document. Group 4 facsimile terminals are required to support automatic answering.

To handle terminal-to-terminal interactions, Group 4 facsimile provides powerful data communication. The link layer is based upon HDLC LAPB with a few modifications for the circuit-switched type of connection. The
network layer is X.25 plus enhancements to bring it in line with the OSI's connection mode network layer definition as far as the ISDN is able to support it. The transport layer is X.224 class 0 and the Session layer is X.225. Above these layers there is the DTAM, Document Transfer and Manipulation, protocol and above that the Open Document Architecture (ODA) which defines the details of the document. All these protocols have a section in their definitions describing how to maintain backward compatibility with the 1984 version of group 4 class 1 facsimile. As long as the link layer parameters allow over half a second of data to be outstanding and awaiting acknowledgement, then group 4 works quite efficiently via satellite channels because it generates such large amounts of data.

8.2.3 Facsimile terminal apparatus

Group 4 terminals are not limited to 64 kbit/s or to the ISDN network by the recommendation but most commercial offerings to date have operated at 64 kbit/s. They have been equipped with either a circuit switched public data networks interface and a PSTN interface or just an ISDN interface.

Given the high resolutions available in group 4 facsimile and a highspeed network such as the ISDN to transfer the large amount of data created by such resolutions, it would seem logical to employ high-quality printing mechanisms in the terminal, e.g. a pre-cut sheet, plain paper, laser-xerographic type. Many terminals have already appeared with this type of printer and the quality of the output image at 400 lpi is for all practical purposes as good as the 300 lpi laser printed original. Unfortunately, there are some disadvantages with this type of printer:

- (a) It can only print a full page and not a string of partial pages. Therefore the terminal has to wait until the whole page has been received and decoded before it can transfer that image information to the printing engine and commence printing. This results in a relatively slow printing process of approximately 10 to 15 seconds per page which is between half and a third of the speed needed to keep up with the 64 kbit/s transmission speed.
- (b) If a 400 lpi laser printer is adopted then it is common practice (for compatibility with group 3) to support only 200 and 100 lpi resolutions as these images are printed by the simple expedient of doubling or quadrupling the ratio of the number of print lines to scan lines. However, this also means that it takes just as long to print a high resolution page as a medium or standard resolution page.
- (c) Compared with thermal print engines laser printers are rather expensive. Some manufacturers have retained the thermal printing

techniques of their group 3 equipment but with the higher resolutions of group 4 and thus produced smaller, cheaper group 4 terminals. However, the quality and stability of the thermally printed image is generally poorer. The limitation of a slow printer can be avoided by installing a large memory to receive the image at maximum speed and then carry out the printing off-line. Some group terminals have been launched with Winchester disc storage for this purpose and to support mailbox facilities within the terminal.

8.3 PHOTOGRAPHIC VIDEOTEX

Graham Hudson

Basic videotex services, providing, via a telecom network, text and graphics information from a central database on a low cost terminal, have become well established, particularly in Europe. Videotex was invented by British Telecom Research Laboratories in the early 1970s and the first public service, Prestel, was launched in 1979.

The deployment of videotex has been helped by the co-operation between PTTs in the European standards body, the Conference of European Postal and Telecommunication Administrations (CEPT). A very comprehensive alpha-mosaic display standard was produced in 1983. This standard also contained an option for photographic quality pictures. British Telecom Research Laboratories demonstrated photographic videotex on the PSTN in 1980 and with full screen pictures via the ISDN in 1983 (see Figure 8.8).

8.3.1 Coding

Alpha-mosaic text and graphics displays are character coded; that is each character rectangle on the display is represented by a single code (7 or 8 bits). The information displayed in the character rectangle is produced with reference to a look-up table or a character generator. A full display screen of 24 rows of 40 characters requires 960 bytes which can be transmitted in 6.4 seconds on the PSTN, but on the ISDN this can be achieved in less than a second on the D channel or less than one tenth of a second on the B channel.

With photographic displays each dot (or pixel) has to be independently defined. Facsimile coding is a simple example of photographic coding where each dot can only have one or two values, black or white. With colour still picture photographic coding each pixel can be any colour from a very large range, typically 16 million colours. Each pixel is represented by values of three primary colour components, typically having an accuracy of

ESTEL 64781469k Born 1925. Educated Grantham High School and Somerville College,Oxford. M.P. for Finchley 1959-74. M.P. for Finchley and Barnet 1974-. Prime Minister and First Lord of Treasury 1979-. Key # for picture 388414a VISIT BANGKOK WITH TRANSWORLD'S SOUTH EAST ASIA EXPRESS TICKET PHONE 01 733 6210 FOR DETAILS O TRANSWORLD INDEX FOR PICTURE

Figure 8.8 PSTN (1980) - Hadamaad transform and ISDN (1983) DPCM photovideotex screens.



Figure 8.9 Character and photographic coding.

8 bits per component. A studio quality picture on a videotex screen has of the order of 262 k pixels and so would require 786 kbytes to be fully represented (see Figure 8.9).

The availability of technologies such as optical storage, graphic display controllers and high-speed networks, such as the ISDN, have made it possible to contemplate the storage and transmission of photographic pictures. Even on the ISDN, however, such a picture would take several minutes to transmit which is unacceptable for interactive services. There has been intense activity in recent years in signal processing research to find the best data compression technique for standardization to make such services possible.

At the end of 1986 experts from ISO and CCITT met to form the Joint Photographic Experts Group (JPEG). They set themselves the task of selecting a high performance universal compression technique. Applications that were kept in mind during the studies, as well as photographic videotex and photographic teletext (broadcast systems), were still pictures for teleconferencing, slow scan pictures for security, medical images, newspaper pictures and satellite weather and surveillance pictures.

A technique was chosen by specifying the main functionality required, setting a number of tests and then evaluating the competing techniques. The selection process took place between 1987 and 1988. Twelve proposals

were registered covering most of the established compression techniques from predictive coding, block coding, discrete cosine transform, vector quantization and combinations of these techniques. Techniques were evaluated on functionality, complexity, but most importantly on subjective quality at fixed compression values.

8.3.2 Discrete Cosine Transform (DCT)

The DCT technique produced the best subjective quality results (Figure 8.10) and it was demonstrated that this technique could be economically decoded at 64 kbit/s by both a hardware (DSP or multiplier) or a microprocessor.

By performing certain mathematical transforms on the spatial pixel values in a block of the image it is possible to produce a set of less correlated coefficients in the transform domain. The general twodimensional transform is defined by the equation:

$$F = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) g(x, y, u, v)$$

where g(x, y, u, v) is the forward transformation kernel.

This may be presented in matrix form as:

F = [T] [P] [T]'

where [F] is the array of coefficient value in the transform domain

$$= \begin{bmatrix} f_{00} & f_{01} & f_{02} & f_{03} & \cdots \\ f_{10} & f_{11} & f_{12} \\ f_{20} & f_{21} & f_{22} \\ f_{30} \\ \vdots \\ \vdots \\ \vdots \end{bmatrix}$$

and [P] is the array of pixel data in the spacial domain

$$= \begin{bmatrix} p_{00} & p_{01} & p_{02} & p_{03} & \cdots \\ p_{10} & p_{11} & p_{12} \\ p_{20} & p_{21} & p_{22} \\ p_{3^{0}} \\ \vdots \\ \vdots \\ \vdots \end{bmatrix}$$



Figure 8.10

Pictures produced by DCT at 0.08, 0.25, 0.75 and 2.25 bits per pel. (Note: originals were in full colour.)

[T] is the transform basis matrix and [T]' is the transpose of the matrix.

The mathematical definition of the forward discrete cosine transform used by ISO for 8×8 pixel blocks is

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

where C(u), $C(v) = \sqrt{\frac{1}{2}}$ for u, v = 0 and C(u), C(v) = 1 otherwise.



The transform basis matrix for an 8 \times 8 DCT

	0.3536 0.4904 0.4616 0.4517	0.3536 0.4157 0.1913 -0.0975	0.3536 0.2778 -0.1913 -0.4904	0.3536 0.0975 -0.4619 -0.2778	0.3536 -0.0975 -0.4619 0.2778	0.3536 -0.2778 -0.1913 0.4904	0.3536 -0.4157 0.1913 0.0975	0.3536 - 0.4904 0.4619 - 0.4157
=	0.3536	-0.3536	-0.3536	0.3536	0.3536	-0.3536	-0.3536	0.3536
	0.2778	-0.4904	-0.0975	0.4157	-0.4157	-0.0975	0.4904	-0.2778
	0.1913	-0.4619	-0.4619	-0.1913	-0.1913	0.4619	-0.4619	0.1913
	0.0975	-0.2778	-0.4157	-0.4904	0.4904	-0.4157	0.2778	-0.0975

In the spatial domain the image energy (square of the pixel values) is normally evenly distributed over the pixel block. Following the transform the majority of the energy resides in a minority of the coefficients. There is much less correlation between coefficients in the transform domain than there was between pixels in the spatial domain. The coefficient f(0,0) the top left-hand corner term, is dominant and contains most of the energy from the pixel block. This zero frequency or d.c. term represents a scaled mean value of the pixel data. The high order coefficients, moving towards the bottom right-hand corner generally contain less energy. Data compression is achieved by reducing the coding accuracy of the less significant coefficient and possibly ignoring some altogether.

The coding process first involves splitting the image into 8 by 8 blocks and then performing the two-dimensional DCT on each block (Figure 8.11).



Figure 8.11 Block diagram of cosine transform encoder.

In the ISO standard all coefficients are linearly quantized. Quantization is achieved with reference to an 8 by 8 quantization matrix Q.

Increasing horizontal frequency →

16, 11, 10, 16, 24, 40, 51, 61, 12, 12, 14, 19, 26, 58, 60, 55, Increasing 14, 13, 16, 24, 40, 57, 69, 56, vertical 14, 17, 22, 29, 51, 87, 80. 62. frequency 18, 22, 37, 56, 68, 109, 103, 77. 24, 35, 55, 64, 81, 104, 113, ſ 92, 49, 64, 78, 87, 103, 121, 120, 101, 72, 92, 95, 98, 112, 100, 103, - 99,

Coefficient quantization matrix

Each value Q(u,v) represents the quantizer step size for that position of coefficient. The step size has been derived according to the perceptual threshold of the visual contribution of the cosine basis function. A different quantization matrix may be used for each colour component.



Figure 8.12

Code structure and Huffman code table.

The d.c. coefficient f(0,0) of the block because of its high correlation with adjacent blocks, is differentially encoded with respect to the d.c. term in the previous block. The 63 a.c. coefficients in the matrix are ordered into a one-dimensional sequence according to a zig-zag scan.

Increasing horizontal frequency \rightarrow

	(DC)	1	5	6	14	15	27	28	1
	2	4	7	13	16	26	29	42	
Increasing	3	8	12	17	25	30	41	43	ł
vertical	9	11	18	24	31	4 0	44	53	I
frequency	10	19	23	32	39	45	52	54	ł
Ļ	20	22	33	38	46	51	55	60	Í
	21	34	37	47	50	56	59	61	
l	35	36	48	49	57	58	62	63 .	
	Z	'ig-2	zag	sca	n p	ath	ma	atrix	

Both a.c. and d.c. terms are Huffman variable length encoded using a two-stage process—modelling and coding. Each non-zero a.c. coefficient is modelled in combination with the run length of the zero value a.c. coefficients proceeding in the zig-zag sequence. Each run length/non-zero coefficient combination is represented by two codes (see Figure 8.12).

8.3.3 Functionality of the data compression technique

The main requirements of an image data compression technique are that it should be efficient (high compression), economic (low cost of implementation), fast (capable of real transmission time decoding) and be applicable to a broad range of natural monochrome and colour picture services.







DCT picture quality for different compression values.

It was felt that the maximum acceptable waiting time for a picture to be transmitted was 5 s. For videotex purposes with a display area of the order of 512 \times 512 pixels a compression of the order of one bit/pixel is required. The ISO DCT provides good quality pictures with a compression of 0.5–1 bit/pixel (Figure 8.13).

Most traditional coding schemes, particularly those delivering to output devices such as printers, produce a sequential picture build-up. That is the picture is constructed in full resolution and colour pixel by pixel from left to right and top to bottom of the image. Such a scheme is ideal for file transfer between computers and for applications such as ISDN photovideotex where the picture build-up on the screen is very fast (Figure 8.14).

For very high resolution pictures on the ISDN and for transmission over slower data rate networks a progressive coding scheme is very desirable. The method provides a crude picture quickly which can, if required, be subsequently improved in several stages until the final quality is achieved. A picture of increasing resolution and accuracy is provided which is useful in database applications where it can match the amount of information transmitted to the capabilities of the output device. This also allows for easy scanning/browsing of databases by only viewing either a low quality full size picture or a reduced size (sub-sampled) picture.

Easily recognizable pictures can be achieved with a compression of less than 0.1 bit/pixel. This compresses an image to 3.3 kbytes and can be transmitted on the B channel in under half a second and in 3 s on the D channel.



Progressive build-up

Figure 8.14 Progressive and sequential picture build-up.

Pictures that are subjectively indistinguishable can be obtained with a compression of the order of 2 bit/pixel (64 kbytes, 8 s on a B channel). Some applications such as the transmission of medical images and surveillance pictures require the final image to be identical to the original. Typically this can be achieved with a compression of the order of 8 bit/pixel.

Not so long ago it was thought that the complexity/cost of implementation of a high-performance compression technique such as a cosine transform would be prohibitive for high-speed real-time transmission schemes. High-speed microprocessors (16 bit, >20 MHz), digital signal processors or VLSI matrix multipliers can each be used to decode DCT at 64 kbit/s.

8.3.4 ISO DCT conformance levels

The DCT compression technique is encapsulated in an ISO standard and a CCITT recommendation. The ISO standard has three conformance levels a baseline system, an optional extended system and an independent function level.

The baseline system must be implemented by every ISO encoder and decoder. It also serves as the default mode and guarantees image communication between any standard encoder/decoder. The system is limited to sequential picture build-up for 8 bit/pixel and does not allow for identical (lossless) coding. Default Huffman tables and quantization matrices based on the CCIR 601 colour space (Y, C_R , C_B) are specified. For a description of this colour space see Section 8.4. There are two default quantization matrices, one for the Y component and the other for the C_R

and C_B components. There are four Huffman tables— $Y_{d.c.}$, $Y_{a.c.}$, $C_R/C_{Bd.c.}$ and $C_R/C_{Ba.c.}$.

The extended system includes all the baseline system features but allows for a number of significant enhancements. Improved coding efficiency can be provided by using arithmetic coding or non-default Huffman coding. A variety of progressive build-up features are allowed for including hierarchical and non-hierarchical coding, bit slicing (increasing accuracy) and spectrum selection (increasing resolution). Reversibility is provided by sending a difference signal between the non-exact DCT image and the original image.

The independent function allows for a stand alone method with reversible (lossless) compression for applications requiring only this capability and uses a simple two-pixel predictor.

8.4 VIDEO

Geoff Morrison

This section describes how moving video pictures are digitally coded and compressed for ISDN bit rates. The algorithm was standardized by CCITT as Recommendation H.261 and is primarily intended for use at bit rates between about 40 kbit/s and 2 Mbit/s. For videotelephony and videoconferencing the coded video bit stream would be combined with audio in a multiplex structure conforming to Recommendation H.221 which provides framing and various housekeeping channels. Some 46.4 kbit/s video and 16 kbit/s audio can thus be carried in a single ISDN B channel. Where higher bit rates are available 7 kHz audio coded to Recommendation G.722 (see Section 8.1) is conveyed within a 64 kbit/s H.221 package leaving the remainder for video. An overall block diagram of the video codec is given in Figure 8.15.

A normal broadcast television picture is built up from 625 or 525 lines repeating at 25 or 30 times per second respectively. These two scanning standards originated in Europe and the USA many years ago and have since been adopted throughout the rest of the world, each country choosing one of the two. A colour television picture can be considered a combination of three signals representing red (R), green (G) and blue (B) constituents. For black and white displays a luminance signal Y can be derived.

Y = 0.30R + 0.59G + 0.11B

Colour information is then carried in two colour difference signals C_B and C_R : $C_B = B - Y$ and $C_R = R - Y$.

By suitable combinations of Y, C_B and C_R a return can be made to R, G and B for colour displays. The purpose of this somewhat devious process is





to take into account the fact that the human eye is rather more sensitive to luminance detail than to colour detail and hence the colour differences can be allocated less bandwidth. A straightforward way, such as specified in CCIR Recommendation 601, to digitally encode such a picture is to sample Y at 13.5 MHz and each of C_B and C_R at 6.75 MHz, each sample being represented with 8 bits resulting in an overall bit rate of 216 Mbit/s. This exceeds the capacity of a 64 kbit/s ISDN B channel by a factor of 3375. Obviously such massive compression needs concerted effort.

The coding scheme achieves its compression by a combination of several techniques. Some, for example entropy coding with variable length codes, achieve data reduction in a lossless way; that is, the coding is completely reversible and achieves an exact reproduction of the original. The others are lossy compression methods, such as low-pass filtering, which are not reversible and produce only an approximation to the original input. These always give an objective loss, but subjectively the distortion can range from being invisible to highly annoying and part of the science of video codec design is in the exploitation of the properties of the human visual system to minimize the subjective loss of quality.

8.4.1 Source format

The first technique used is the simple one of sacrificing a little of the picture sharpness. In the horizontal direction, the luminance is sampled at 6.75 MHz, half of CCIR 601, allowing a bandwidth of just over 3 MHz to be obtained.

For vertical resolution, the CCITT sought a recommendation for worldwide use and therefore needed to take account of the 625 and 525 formats. Rather than have two different versions of the recommendation the decision was taken to specify just one based on a compromise in which the number of lines bears a simple relationship to the 625 system and the repetition rate comes from the 525 system. The adopted number of luminance lines is 288, being half the number of the active lines in 625. In that system some 575 lines are available for the picture itself, the others giving a retrace interval to allow the scanning spot to move from the bottom back to the top of the picture. For 525 sources and displays the conversion is a little more complex, but the ratio 5 : 3 allows 480 lines to be covered. Theoretically this is just a few lines less than the active height of a 525 line picture but in practice displays are always overscanned and the missing lines would not be seen anyway. The picture refresh rate adopted comes directly from the 525 system, being approximately 29.97 Hz.

For the colour differences a further reduction of spatial resolution in each direction to half of the luminance is employed, taking advantage of the eye's lower acuity to colour detail mentioned above. The objective spatial resolution of the coding format is hence much less than CCIR Recommendation 601. However, the subjective loss is not as drastic as the numbers might suggest. The horizontal bandwidth of more than 3 MHz is on a par with the resolution of the colour display tubes typically found in domestic television sets and is better than many home video cassette recorders. Also, in the vertical direction resolution is better than the 1/2 of 625 or 3/5 of 525 could imply because those two formats use interlace whereas H.261 does not. The explanation of this is outside the scope of this book.

CCIR Recommendation 601 specifies a digital active line of 720 pels of luminance and 360 of each of the two colour difference signals. Simple 2 : 1 reductions would give 360 and 180. However, for reasons which will become obvious below it is desirable to have numbers which are an integer multiple of 16 for luminance and 8 for colour differences. To achieve this some pels are discarded at the edges of the picture to leave 352 and 176. The loss of picture information is minimal as the ignored pels frequently fall outside the real picture area in the part allocated to horizontal flyback of the scanning spot and even if not would not be visible on normally overscanned displays.

The format came to be known as 'Common Intermediate Format' and referred to as CIF. Subsequently, to permit simpler codecs for use with smaller screens and low bit-rates, for example ISDN videotelephones, an additional format was introduced whose spatial resolution is one half of CIF in each axis. This factor of four prompted the term 'Quarter CIF' (QCIF).

PCM encoding of CIF with 8 bits per pel would need $352 \times 288 \times (1 + 1/4 + 1/4) \times 29.97 \times 8 = 36.5$ Mbit/s.

Before leaving the source coding format, it should be noted that H.261 does not specify how CIF or QCIF is derived from the coder input video standard nor how a decoder should convert the decoded images in CIF or QCIF form to a suitable format for displays.

8.4.2 Source coding

A block diagram of the source coder and decoder is given in Figure 8.16. The function of the coder is to achieve data compression by redundancy reduction. The switches are shown in the position they occupy most of the time to take advantage of the fact that the television signal is a series of snapshots and that each one is very much like the one before it. Indeed for stationary parts of the scene the difference between successive pictures is zero and no information needs to be transmitted at all. The block marked P is a memory which holds the previous picture. By making the delay through it plus the other functions in the loop from the output of the



subtractor back to its lower input exactly equal to the period of one picture, the corresponding points in successive pictures are compared. Only in places where there have been changes such as motion will there be a nonzero output which needs to be transmitted.

A further refinement is motion compensation. If a moving object is undergoing purely translational motion then that part of the picture can be constructed by teproducing an offset part of the previous picture. The offset will depend on the direction and magnitude of the movement and is called a motion vector because it has two components, left-right and updown. The vector is used to adjust the delay of P so that corresponding points on a moving object are fed to the subtractor. Transmitting a motion vector for individual pels would require too many bits so the picture is divided into blocks of 16 by 16 luminance pels and one vector is used for it and also for the two spatially corresponding 8 by 8 colour difference blocks.

This process is known as motion compensated inter-picture prediction. The prediction is based on the previous picture and the whole process is recursive. The subtractor output, the prediction error, will not always be zero. Moving objects can change shape or rotate and background can be uncovered at the object edges so that one single vector cannot apply perfectly for all the block. However the energy of the prediction error is much less than the original signal and hence more easily transmitted. The process is completely reversible. By accumulating all the error signals and using motion vectors a decoder could reconstruct the original signal exactly. However, to get even more compression it is necessary to apply lossy compression to the prediction error in the quantizer Q. Because the predictions in the encoder and decoder must be the same and the decoder only has decoded pictures to predict from, it is necessary for the encoder to use the same modified versions for its predictions. This is the reason for the decoding boxes Q^{-1} and T^{-1} and the adder being required in the coder.

The method of determining motion vectors at the coder is not defined. The problem is essentially one of pattern matching and can be very expensive computationally since the vectors are allowed to have components in the range -15 to +15. This means that for each block up to 225 comparisons of 16 by 16 blocks could be needed to find the one with the smallest prediction error. Some schemes use simplified methods such as tree searches.

Because the coding scheme is recursive there are the problems of startup and transmission errors. Under those conditions the loops at encoder and decoder do not contain the same information and as the decoder is essentially a perfect integrator the differences would remain forever. To deal with this the switches are occasionally flipped over so that the video signal itself instead of a difference is sent. However this intra-coding mode requires many more bits and so can only be invoked for a small fraction of the time.

The block marked T performs a Discrete Cosine Transform (DCT) on 8 by 8 blocks according to the formula as set out in Section 8.3.2.

This reversible mathematical process converts the 64 numbers representing the levels of 64 pels to another set of 64 numbers, transform coefficients, but the information is contained in them in a different form such that coding gain can usually be achieved. These transform coefficients represent the amplitudes of spatial frequencies in the block. For example, F(0,0) is the average level of the block and F(1,0) gives a measure of the difference between the left and right halves. Many blocks contain only a few frequency components or several of them have low enough amplitudes to be ignored such that it is not necessary to transmit all 64 of them.

The quantizer, Q in Figure 8.16, introduces lossy compression by forcing the transform coefficient amplitudes to adopt a restricted set of values which require fewer bits to differentiate between them. Of course this introduces an error in the coefficient but the inverse transform spreads it over all the pels in a block so the effect is much less than if the pels themselves had been subjected to the quantization process. Q actually contains 31 different quantization laws. All are linear but have different step sizes so that more or less effect can be obtained. This facility is most useful to allow the intrinsic non-constant bit rate of the coder to be matched to the fixed bit rate transmission channel as will be mentioned later.

Since the forward transform is only in the encoder it does not have to be precisely specified. However, an inverse transform (IDCT) is present in both the encoder and decoder and if the two do not give the same outputs then the differences will accumulate and the loops diverge with visible effects such as a stationary granular noise pattern on the pictures at the decoder. Unfortunately the IDCT, like the DCT, contains cosine terms and many of them cannot be represented exactly with finite arithmetic. Errors will therefore be introduced by rounding or truncation in real systems. To allow various implementations of the IDCT function to be used, the CCITT decided not to specify a unique fixed precision IDCT but to leave a small error tolerance. Long-term accumulation of the allowed small mismatch errors between different IDCTs at encoder and decoder is controlled by stipulating that a block must be intra-coded at least once in every 132 times it is transmitted.

The filter, F in the diagram, is a two-dimensional slow roll-off low-pass filter which can be used to suppress noise introduced by the quantizer to prevent it circulating in the loop. However, it cannot be left active in all blocks because stationary detail would be affected giving a poor prediction and an increase in the number of transmitted blocks with a consequent increase in bit consumption and therefore poorer rendition of the genuinely changed areas. In many coders, therefore, the filter is linked to the motion vector so that it is switched to all-pass when the vector is zero.

The final item in Figure 8.16 is coding control. This determines the coding parameters such as the selection of the quantizer and the use of intra-mode. Additionally, especially at low channel rates the scheme is not able to transmit the input picture rate of 30 Hz and picture dropping is invoked. Since the decoder only needs to know the decisions made but not the reasons for them the coding control strategy is not specified.

8.4.3 Video multiplex coding

The function of the video multiplex is to serialize the various data streams such as those containing addresses of changed areas of pictures, motion vectors, quantized coefficients and so on into one bitstream. It also achieves further lossless data compression by statistical coding of the data from the source coder. The multiplex is arranged in a hierarchical structure with the four layers—Picture, Group of Blocks (GOB), Macroblock (MB), Block—as shown in the syntax diagram of Figure 8.17.

The Picture Start Code (PSC) is a specific 20-bit word which allows the decoder to identify the start point of the picture. Next comes the 5-bit Temporal Reference (TR) which is a form of time stamp to allow the decoder to put pictures to the output screen at the correct times if picture dropping has been invoked by the coder. PTYPE contains 6 bits of information which relate to the complete picture such as whether it is CIF or QCIF format. PEI is a single bit which indicates whether PSPARE of 8 bits and another PEI bit follow. This will permit a simple linked list structure to be used at a future date for enhancements but initially only the first PEI bit is included and indicates that PSPARE is absent. Information for the GOB layer then follows.

A group of blocks comprises one twelfth of the CIF or one third of the QCIF picture areas as shown in Figure 8.18. A GOB thus covers 176 pels by 48 lines of Y and the spatially corresponding 88 by 24 of each of C_B and C_R . The Group of Blocks Start Code (GBSC) is a 16-bit word and is immediately followed by 4 bits containing the GOB number (GN) to identify which part of the picture the GOB is from. GQUANT is 5 bits to signal which of the 31 quantizers is in use in the GOB. GEI and GSPARE allow a linked list extension structure in the same manner as in the picture layer. Then comes the macroblock data.

Each GOB is divided into 33 macroblocks as shown in Figure 8.19. A macroblock relates to 16 lines by 16 pels of Y (four 8 by 8 blocks) and the spatially corresponding 8 by 8 of each of C_B and C_R as in Figure 8.20. Macroblock Address (MBA) is a variable length codeword indicating the



Figure 8.17

Syntax diagram for the video multiplex coder.



Figure 8.18 Arrangement of GOBs in a picture.

1	2	3	4	5	6	7	8	9	10	11
12	13	14	15	16	17	18	19	20	21	22
23	24	25	26	27	28	29	30	31	32	33

Figure 8.19

Arrangement of macroblocks in a GOB.

1 2 5 6 3 4 5 6	Y	Ce	C _R
	13141 L	L	L
	1 2	5	6

Figure 8.20

Arrangement of blocks in a macroblock.

position of a macroblock within the GOB. MBA is the differential address between transmitted macroblocks. Macroblocks are not transmitted when that part of the picture is not to be updated. The variable length code table for MBA contains an extra codeword called MBA stuffing. This should be discarded by decoders as its function is merely to pad the bitstream as may be necessary when pictures with little motion are coded at the higher bit rates.

MTYPE is a variable length codeword to give information about the macroblock and which data elements follow. For example, it indicates whether the macroblock is inter- or intra-coded, whether there is a motion vector or not and whether the filter is applied in the prediction or not. It can also signal that a different quantizer will be brought into effect from this macroblock onwards. For a macroblock which can be perfectly reconstructed just by moving part of the previous picture only Motion Vector Data (MVD) is necessary. When the motion vector is zero this is explicitly signalled by MTYPE and MVD is not included. MVD is not needed for intra-coded macroblocks since these do not use predictive coding. For data compression reasons MVD is a variable length codeword representing the vector difference from the neighbouring macroblock since there is a high probability that its motion vector will be similar.

Coded Block Pattern (CBP) is a variable length codeword which tells which of the six 8 by 8 transform blocks have coefficient data sent for them in the block layer. CBP is not needed by all macroblock types. For example it is always the case that intra-coded ones have coefficients for the six blocks.

Data for a block consists of codewords for quantized transform coefficients (TCOEFF) in the zig-zag order given in Section 8.3.2, followed by an End of Block (EOB) marker. The marker is needed as the number of

coefficients may be between 1 and 64 because the trailing zero amplitude ones are not sent. The zig-zag follows the path of increasing probability that a coefficient is zero. A two-dimensional variable length coding method is adopted in which the 64 most commonly occurring combinations of successive zeros (RUN) and the following level (LEVEL) are encoded with words of between 2 and 14 bits. The remaining combinations of (RUN, LEVEL) are encoded with 20 bits consisting of a 6-bit ESCAPE, 6-bit RUN and 8-bit LEVEL.

The nature of the coding algorithm means that the rate of generation of bits is not constant but varies depending on the input pictures. Pictures with more motion have more changed areas and the statistics of these changes will be reflected in the lengths of the variable length codewords transmitted. Furthermore, the motion is rarely distributed evenly over the picture area so that the generation is spasmodic. A buffering process is necessary at the encoder to match this varying rate to the fixed channel rate and a corresponding inverse process at the decoder. Such buffers introduce delays and for conversational services such as videotelephone and videoconference these must be bounded at a few hundred milliseconds. The buffer, therefore, can only perform short-term smoothing and the long-term average data rate must be controlled by altering parameters in the coding loop. As mentioned earlier coarse control can be achieved by dropping complete pictures and fine control by selection of the quantizer, so most coders incorporate a feedback mechanism from the buffer fullness to regulate those functions.

The coded bitstream is rather vulnerable to transmission errors because of its heavy use of variable length codewords. In addition to corrupting one word, a single bit error can make one transmitted variable length codeword appear to be two, or vice versa. This results in a decoder losing track of the correct data sequencing and interpreting all subsequent data wrongly, usually until the next GOB Start Code. Because of the recursive nature of the algorithm any disturbances remain on the screen for many picture periods. Further the use of motion vectors can make corrupted regions move around the screen and grow larger. To offer protection against this, the transmitted bit-stream contains a BCH (511,493) Forward Error Correction Code which permits two random errors or a burst of up to six errors in a 511-bit block to be corrected. To allow the video data and error correction parity information to be identified by a decoder an error corrector framing pattern is included. This consists of a multiframe of 8 frames, each frame comprising 1-bit framing, 1-bit fill indicator, 492 bits of coded data and 18 bits parity as shown in Figure 8.21. The fill indicator can be set to zero to signify that the following 492 bits are not video but merely stuffing bits which the decoder should discard. This can be used in a similar manner to MBA stuffing described above.

The algorithm is intended for use over a 40 to 1 range of bit rates and the



Figure 8.21 Error correcting frame.

resultant overall quality will depend markedly on the operating rate and on the characteristics of the input video signal. In addition, the H.261 Recommendation only specifies those items which are necessary to guarantee compatibility and leaves considerable freedom to coder designers to achieve what they feel to be the optimum balance of the various disortions that must be introduced to achieve a given bit rate. This means that even when coding the same pictures at the same bit rate two coders can yield significantly different results. Unfortunately, even with photographs, it is not possible in a book to show examples of the quality that can be obtained because the motion is such an important part of the subjective evaluation. However, at 1.5 or 2 Mbit/s the pictures compare favourably with home video recorders even when coding typical broadcast television programmes. At 64 kbit/s a fixed camera position is really required and pictures can become noticeably jerky due to the necessarily low coded picture rate, but they are suitable for the intended purpose of interactive audiovisual communication.

8.5 CUSTOMER PREMISES EQUIPMENT

There is a chicken and egg relationship between network and terminal availability. The lead time for network developments is many years, whereas terminals come and go in a year or two. Thus any review of terminals can only be transitory.

ISDN terminals can be classified into several categories.

Telephony and other voice

There is no immediate pressure to adopt ISDN for voice transmission purposes. The reasons for the development of ISDN telephony terminals are fourfold:

- 1. Without the analogue local loop the quality is marginally improved, and for loudspeaking telephones the absence of a two-wire link means that one source of feedback and instability is removed. Telephones are available with 7 kHz speech as described in Section 8.1.
- 2. The improved signalling capability gives easy access to facilities on the public network similar to those found on modern PABXs. This is particularly relevant for CENTREX operation where the public exchange is used to emulate a PABX.
- 3. The telephone may incorporate a terminal adapter function offering a range of data ports.
- 4. To complete the range of equipment offered by a supplier which can be associated with other 1.420 products.

Terminal adapters

This is a terminal which offers data services only. It is used to interface non-ISDN terminals to the ISDN I.420 (or primary rate) interface. Equipment is currently available and interfaces include IBM 3270, IBM Twinex, V.24 (RS232), V.25 bis, V.35, X.21, X.25.

Computer products and dedicated terminals

The most popular initial product in this area is a card to plug into the IBM personal computer or its clones. These are available from many manufacturers and may incorporate other functions, such as providing an analogue port for a telephone handset. However, many are tied to the early ISDN realizations described in Section 1.8. Group 4 facsimile machines are available, some of which connect directly to an I.420 bus, and others which require a different interface (e.g. X.21) and need a terminal adapter. Workstations are able to exploit the 64 kbit/s channel and several are now available with this facility. Video terminals have been developed but the early ones do not use the international standard form of coding described in Section 8.4 and hence are limited in their application.

Switches and multiplexers

Many PABXs are available which can interface to the ISDN. The larger PABXs are more appropriately connected to primary rate accesses, and the smaller key systems will use a few basic rate accesses. Early PABXs tend to use analogue or proprietary digital access to the extensions, but with the advent of low cost I.420 devices the trend will be towards the use of standard ISDN interfaces for extensions. Multiplexers are rather simpler devices which break down a basic or primary rate interface to provide access to a multiplicity of interfaces and hence in complexity are midway between a terminal adapter and a PABX. A particular application for this equipment is to provide a means of backing up private circuits, in which private circuits may automatically be transferred to a dial-up ISDN circuit if the private circuit fails.

REFERENCES

CCITT Recommendations:

I. 42 0	Teleservices supported by an ISDN.
G.721	32 kbit/s adaptive differential pulse code modulation
C 700	(ADFCM). 7 kHz audio andina within 64 khit/a
G./22	/ KHZ audio coding within 64 kolt/s.
G.725	System aspects for the use of the 7 kHz audio codec within 64 kbit/s.
T.4	Standardization of Group 3 facsimile apparatus for document transfer.
T.563	Terminal characteristics for Group 4 facsimile apparatus.
T.503	A document application Profile BTO for the interchange of
	Group 4 facsimile documents.
T.6	Facsimile coding schemes and coding control functions for
	Group 4 facsimile apparatus.
X.224	Transport protocol specifications for open systems interconnection for CCITT applications.
X.225	Session protocol specification for open systems interconnection for CCITT applications.
T. 521	Communication application profile BTO for document bulk transfer based on session service.
H.221	Frame structures for a 64 to 1920 kbit/s channel in audiovisual services.
H.261	Video codec for audiovisual services at $p \times 64$ kbit/s.
T.81	Encoding of still photographic images.

T.431–433Document Manipulation and Transfer protocol (DTAM).T.411–418Open Document Architecture (ODA).

International Standards Organization:

ISO 9282-2 Still picture compression technique.

International Radio Consultative Committee:

CCIR 601-1 Encoding parameters of digital television for studios.

CEPT Recommendation

T/TE 06 series videotex services presentation layer data services.

Chapter 9

Broadband ISDN

Up to this point we have only considered ISDN based on 64 kbit/s B channels and 16 or 64 kbit/s D channels. However for many applications even higher bit rates would be useful. Although video telephones may be acceptable at 64 kbit/s using a very small screen, for video conferencing, where a normal television size screen is needed, 384 kbit/s (6×64 kbit/s) is more attractive. Entertainment TV must be able to cope with situations in which successive frames are very different to meet the artistic aspirations of producers. For this reason there is little in the way of redundancy reduction possible and bit rates in the tens of megabit/s are needed. High definition television requires bit rates in the hundreds of megabit/s.

On the business side the interconnection of high speed LANs used for computer-aided design may well generate traffic at high rates. Of course there is also the general drift of people's expectation. The ability to handle facsimile pages in 4 seconds compared with the 30 seconds of the pre-ISDN era may be widely appreciated, but how long will it be before people expect the fax machine to operate at the same speed as the office photocopier, and in full colour? Thus not only can the need be foreseen for higher rates, but also the technology to offer service is available in the form of optical fibre and devices.

Figure 9.1 shows how the data capacity of a public switched telephone network connection has increased as modems have developed and with the availability of the ISDN. The dotted lines give some plausible extrapolations for the future. The curved line indicates how modems have evolved under the constraint of Shannon's limit described in Section 1.4. The straight line offers a demand for the future without the Shannon limit. Figure 9.2 summarizes the data rate needs of various services. The provision of channels above 64 kbit/s is generally referred to as Broadband ISDN or BISDN but perceptions as to what constitutes a BISDN and the applications that must be carried vary enormously.

9.1 N × 64 KBIT/S

At the lowest end of BISDN comes the concatenation of several 64 kbit/s channels. On the face of it this could be achieved on the ordinary ISDN by





The rise of switched data rates.

setting up several 64 kbit/s calls to the same destination on a primary rate interface and concatenating the channels at the terminal. Unfortunately the network does not provide for channels set up in such a way to be subject to uniform delay. Time switching stages (see Section 3.3) may introduce up to a frame's delay in one channel relative to another. More seriously there



Figure 9.2 Bandwidth requirements.

is no guarantee that channels will follow the same, or even similar, paths through the network. Different channels may be routed via different transmission line plant on different routes with different delays; the most extreme situation would be where one channel is carried by a satellite link and another is carried on a terrestrial link. The provision of alternative routing strategies in the network may also mean that one channel may pass through an extra switching node with its associated delays. There are two ways of overcoming these problems.

The terminal solution

Obviously by providing appropriate buffers at the terminals the delays in each channel can be padded to be equal. This will necessitate an initial 'investigation' period by the terminals to establish the delays. One assumption that may reasonably be made is that relative delays will not change during a call as, in general, networks do not normally reconfigure established calls.

The network solution

In this case the exchange processors would ensure that all channels are kept within a single time division multiplex and hence follow a common route. Apart from the need for additional software, the major problem is in design for traffic carrying capacity. Routes are dimensioned to give an acceptable quality of service when channels are required singly. For example using Erlang's formula (implying the usual assumptions) a group of 30 or 23 trunks could be expected to carry 19.03 or 13.42 respectively erlangs of traffic and only 1 in 200 calls would be unable to find a free channel (i.e. a grade of service of 0.005). For a channel operating at this traffic level Table 9.1 shows the probability of not being able to accept a multichannel call of N channels. This is the grade of service for $N \times 64$ kbit/s circuits in an environment where multichannel calls are rare (and hence the chance of two arising is negligible) and single 64 kbit/s channels predominate in the design consideration. These tables show that the probability of being able to find six channels free on a 30-channel system is about one seventh of the basic grade of service. With 23-channel systems the same probability drops to one tenth of the basic grade of service. The grade of service offered for single 64 kbit/s channels, after setting up the 6 \times 64 kbit/s channel, is to be found under the heading of N = 7 (i.e. 0.050 or 0.072 for 30 and 23 channels respectively). For these reasons it might be felt that N = 6 is about the largest multichannel service that could be offered in this way. In many cases it may be felt that even this value of N might need

Table 9.1

The probability of not being able to provide an N \times 64 kbits channel on demand.

30-channel system N =2 3 5 12 1 4 8 Q 10 11 6 7 0.005 0.008 0.012 0.018 0.026 0.037 0.050 0.066 0.086 0.108 0.134 0.163 23-channel system N =١ 2 3 4 5 10 11 12 6 7 8 9 0.005 0.008 0.014 0.023 0.034 0.051 0.072 0.098 0.129 0.166 0.207 0.253

a special provision. Of course in an exchange and trunk network there will be more than one 30- or 23-channel group from which to choose to give the N channels. However on a single primary rate access there is no advantage from this source.

9.2 H CHANNELS

Referring to Figures 9.1 and 9.2, one might identify some specific steps in the rates offered above 64 kbit/s. These have been identified by CCITT as:

- (a) H0 at 384 kbit/s (i.e. 6 × 64 kbit/s). This would be specifically attractive for video conference codecs and hifi sound. The potential effect of this on traffic design has already been discussed in Section 9.1. It is assumed that the interface to the customer for H0 channels will be the primary rate interface. The 1.544 Mbit/s primary rate can accommodate three H0 channels in timeslots 1–6, 7–12, 13–18. If a signalling channel is not required then a fourth H0 channel can be accommodated in timeslots 19–24. At 2.048 Mbit/s five H0 channels can be accommodated and the preferred assignment is to use channels 1–6, 7–12, 13–19 (excluding 16), 20–25, 26–31. Any timeslots not used for H0 channels can be used for B channels.
- (b) H1 channels. Two forms of H1 channel have been identified:

H11 at 1.536 Mbit/s. This may be carried on a 1.544 Mbit/s primary rate interface but a signalling channel would have to be provided separately. Alternatively it can be carried in timeslots 1–25 (excluding 16) of a 2.048 Mbit/s primary rate interface. Signalling can then be carried in channel 16.

H12 at 1.920 Mbit/s. This may be carried in timeslots 1-31 (excluding 16) of a 2.048 Mbit/s primary rate interface. Channel 16 is used for signalling as usual.

Because of the large chunks of capacity that H1 channels would absorb in the switching fabric of a switch based on the switching of 64 kbit/s channels, it is most likely that either a new switching network will be provided or special provision will be made within switches for them.

(c) CCITT has also identified some other H channels: H21 around 34 Mbit/s, H22 around 55 Mbit/s, H4 around 135 Mbit/s.

9.3 HIGHER RATE INTERFACES

Optical fibres offer virtually unlimited bandwidths and their use in the local network is widely predicted. Given the demand the technical and economic problems of providing service to customers can undoubtedly be overcome. Even if it is unacceptable to use 1 fibre per connection, then wavelength division multiplexing offers enormous capacity, particularly if associated with coherent detection techniques.

Various channel rates have been identified above, but it is not necessary that there are interfaces to the customer operating at that rate. The whole concept of the ISDN is that there should be a minimum number of interfaces identified so that a wide range of compatible equipment should become available. So far only two interfaces have been identified-the basic rate and the primary rate (although this comes in the North American and Rest of World versions). Continuing to higher rates, once again there is pressure to have only two more interfaces; the proposal is that these should operate at about 150 and 600 Mbit/s. This does not mean that only channels at these rates would be available. In fact, as will be seen, the plan is that many channels of a wide range of rates will be multiplexible on to these interfaces, so that a wide range of terminal equipment can be connected to them. At the 150 Mbit/s rate it can be seen that all services except high definition TV can be accommodated. 600 Mbit/s could accommodate several standard TV channels or HDTV, simultaneously with lower rate services. When studying Figure 9.2 it is important to appreciate that the horizontal scale is logarithmic; on a linear scale the small capacity required for the lower rate services would be more easily appreciated because the ranges would be invisible on the left-hand side. The real problem is how services with such a wide range of rates could be efficiently multiplexed on to a common bearer and two solutions are being followed. One is based on synchronous multiplexing, but with a format which can be configured to match the needs of the user under his control; this is called the Synchronous Digital Hierarchy (SDH). The other extends the use of packet mode services to a very lightweight protocol called Asynchronous Transfer Mode (ATM) which because of its simplicity can be implemented at these high speeds.

9.4 SYNCHRONOUS DIGITAL HIERARCHY Paul McDonald

This standard has evolved from the American optical interface standard, SONET. This has been developed over the last few years by the American National Standards Institute (ANSI) to provide a standard inter-exchange interface for American network operators in response to the desire of these operators to have managed optical interconnections in their networks. SONET was later reworked in order to make it more suitable for carrying channels in the European hierarchy. This was put forward to the CCITT who have enhanced it and incorporated it into its current set of recommendations.

Using a new system such as this has the following advantages over existing multiplexing structures:

- (a) Direct access to lower rate channels without having to demultiplex the entire signal.
- (b) Enhanced maintenance facilities.
- (c) Simplified evolution to higher interface rates.
- (d) The ability to carry new broadband channels that may appear.

In SDH the signal is constructed from STM-1 (Synchronous Transport Module level 1) frames. This structure has a bit rate of 155.520 Mbit/s and is the basic unit from which higher interface rates are constructed. This is done by byte interleaving *n* of these STM-1s to form the higher rate STM-*n* signal. At present the only values of *n* defined are n = 1 and n = 4, giving interface rates of 155.520 Mbit/s and 622.080 Mbit/s respectively.

The STM-1 frame is organized as 9 rows by 270 columns of 8-bit bytes (Figure 9.3). This gives 2430 bytes to be transmitted in $125 \,\mu$ s. This structure is transmitted from top left to bottom right, scanned in a similar method as reading a book. For those just seeking a broad overview of wideband possibilities this is probably a good point to move on to the next section. If, however, more details are sought then read on!

The first 9 columns of each row, except row 4, are allocated to section overhead functions such as framing and maintenance channels. The rest of the available capacity carries the payload in Administrative Units (AU). To prevent large sequences of consecutive '1's or '0's a frame synchronous scrambler operates on all of the bytes except the first row of section overhead. This scrambler uses the polynomial $1 + x^6 + x^7$ and is reset to all ones after row 1 of section overhead.

The AU is a mechanism that allows efficient transportation of payloads that are not in exact synchronization with the STM-1 signal. The payload and STM-1 frame may have different phases and any slight frequency





differences do not cause frame slips; that is to say it can carry plesiochronous signals.

The data to be transported by the administrative unit is held inside a virtual container (VC) which can 'float' inside the STM-1 payload area. The start of the virtual container is indicated by an administrative unit pointer. This indicates the number of bytes, ignoring section overhead and administrative pointers, until the start of the next virtual container. It also contains frequency justification information that allows the start position of the virtual container to move inside the STM-1 frames if required. Figure 9.4 shows the largest virtual container, VC-4, floating inside the STM-1 frames.

Generally the AU pointer consists of three bytes; H1, H2 and H3. H1 and H2 are taken together to form the actual pointer, holding the offset to the start of the next virtual container. It can also act as a flag to indicate that the pointer value is to be updated. H3 is a negative justification opportunity and can hold valid data in frequency justification operations. Also there is a byte defined immediately after the pointer for a positive justification opportunity used in frequency justification.

If the payload rate is too slow with respect to the STM interface rate, then periodically the payload pointers will be incremented and the positive justification opportunity that is defined for that AU is filled with null data. If the payload rate is too fast, the AU pointer will be decremented. At all times the overall size of the VC remains constant, only the time taken to transmit it has altered.

There are three types of administrative unit defined. These allow three different sizes of virtual container to be carried by this means.



* AU pointer



VC-4 floating inside STM-1 frames.

AU-4. This consists of a single VC-4 which requires the total payload area available in the STM-1 frame. Its single AU pointer is held in row 4, columns 1 to 9 of the STM-1 frame (Figure 9.5).

Figure 9.6 shows the allocation of the AU pointer bytes.

This AU pointer calculates the offset to the following VC-4 in units of 3 bytes, i.e. offset 0 is the byte after the H3 bytes, offset 1 is 3 bytes after this (column 13). Frequency justification operations move the VC-4 in units of three bytes to maintain valid offset values. This also means that there are three H3 bytes required and similarly three positive justification opportunities defined.



Figure 9.5 VC-4 in an AU-4.





AU-32. This consists of three VC-32s inside the STM-1 payload area. Thus three AU pointers are required and these are held in row 4, columns 1 to 9 of the STM-1 frame (Figure 9.7).

When the three VC-32s are multiplexed into the STM-1 payload area they are byte interleaved so that a byte of the first VC-32 is followed by a byte of the second VC-32 then the third. The starts of each individual VC-32 may differ, though, and this is pointed to by its AU pointer. When each container is multiplexed into the STM-1 frame two columns of padding bytes are added to the VC-32 to fill out the available capacity. These padding columns are added between columns 29 and 30, and columns 57 and 58 of the VC-32 (Figure 9.8).

Each pointer consists of three bytes, H1, H2 and H3, and these are byte interleaved with the other AU-32 pointers. There is one positive and negative justification opportunity for each AU-32 respectively. The offset value counts the offset to the start of the next VC-32, counting only the bytes that could be occupied by that VC-32, that is location 0 for the first VC-32 is column 10, location 1 for it is column 12, etc.



Figure 9.7 Three VC-32s in STM-1 frame.





AU-31. This consists of four VC-31s inside the STM-1 payload area. There are four AU pointers, one for each VC-31, and these are stored in columns 11 to 14, rows 1 to 3 of the STM-1 frame. The other AU pointer location is row 4, columns 1 to 9 are filled with a null pointer indication (NPI) to indicate that this pointer is not in use. Column 10 is not required by the AU-31s and is filled with padding bytes (Figure 9.9).

The four VC-31s are multiplexed into the STM-1 payload area using byte interleaving so that a byte of the first VC-31 is followed by a byte of the second VC-31 then the third and fourth. Each individual VC-31 may start at a different point, though, and this is pointed to by its AU pointer (Figure 9.10).

Each AU-31 pointer consists of three bytes, H1, H2 and H3, arranged vertically. There is one positive and negative justification opportunity for each AU-31 respectively. The pointer offset value holds the offset to the start of the next VC-32, counting only the bytes that could be occupied by



Figure 9.9 Four VC-31s in the STM-1 frame.




that VC-31, that is location 0 for the first VC-31 is column 15, location 1 for that VC-32 is column 19, etc.

Virtual containers consists of two parts:

- 1. The path overhead (POH) which includes maintenance channels and control information.
- 2. The container which holds the data or lower rate virtual containers carried in tributaries.

Tributaries are lower rate services, e.g. 2.048 Mbit/s CCITT signals, mapped into higher rate containers. The service is first placed inside a virtual container of sufficient bandwidth which is then used to turn a tributary unit (TU) (Figure 9.11).

Tributary units may be locked or floating. In locked mode the position of the information is fixed with respect to the start of the TU. This makes the location of parts of the tributary simpler. In floating mode the position of the information is allowed to 'float' inside the TU, just like a VC inside an AU structure.

VC-3 tributaries are held in TU-3s. These VCs can 'float' inside a VC-4 in a similar way as in an AU-3 structure. The TU-3 has similar pointers to the AU-3 and they are held at different locations inside the VC-4.

The lower rate tributary units, TU-11, TU-12, TU-21, TU-22, allow VC-1s and VC-2s to be carried inside the higher rate VC-3s and VC-4s. The tributary units are first placed into a tributary unit group (TUG) which has a fixed size and holds a single type of TU, e.g. all TU-12s. This group is then put into a VC-3 or VC-4, multiplexing it with other TUGs that may contain different types of TU. Byte interleaving is used throughout this process.



Figure 9.11 SDH multiplexing hierarchy.

9.5 ASYNCHRONOUS TRANSFER MODE (ATM)

The previous section outlined a method of synchronously but flexibly multiplexing together channels of a range of rates. However, it may be deduced that even in outline the overall process turns out to be complex. A full description of the processes involved is several times the length allocated to Section 9.4. However the complexity primarily lies in the control mechanisms which are required to operate at low speeds of a few kbit/s rather than the interface rate.

An alternative strategy is to adopt message based techniques of the type described in Chapter 7. Using such techniques it is straightforward to accommodate a wide range of channel rates on a single bearer. The penalty is that much of the control mechanism has to operate at the bearer rate. Present day technology only allows the simplest processing activities at rates of 150 Mbit/s, and at 600 Mbit/s even the simplest of operations are challenging.

ATM uses short, fixed-length cells with minimal headers, to allow calls to be routed at high speed by means of hardware implemented routing tables at each switch. International agreement is that the header of each cell will consist of 5 bytes and that the cell information field will consist of 48 bytes, making a total cell length of 53 bytes. Note the basic unit is called a 'cell' to avoid confusion with 'packets' of lower bit rates. The header will consists of bits to provide two main routing functions.

- (a) Virtual Path Identifier (VPI). A path is the equivalent of a route in a circuit based environment, permanently connecting two points together. In an ATM environment the path would not have a fixed capacity. The 'virtual' tag indicates that cells would be routed from node to node on the basis of the VPI in whatever way seemed most appropriate.
- (b) Virtual Call Identifier (VCI). These calls would be set up as required over the virtual path indicated by the VPI.

The structure of the header at the user-network interface is shown in Figure 9.12. It contains the following sub-fields:

- -Generic flow control field (4 bits).
- Routing field (VPI, VCI); 24 bits are available, however the maximum number of virtual connections that can be active is determined on a subscription basis, never exceeding 2²⁰ virtual connections. The VPI/VCI 4-bit field can be assigned to either the VPI or the VCI, leading to a VPI field of 12 or 8 bits and a VCI field of 12 or 16 bits respectively.
- -Payload Type (PT) field (2 bits).
- -Reserved field (2 bits).
- -Header Error Control (HEC) field (8 bits).





The ATM network would be able to flexibly handle all types of traffic. Particular examples are:

- (a) Voice. 64 kbit/s voice can be assembled into ATM cells. With 48 bytes of information in each cell then each cell can contain 6 mS of speech. For some purposes the delay this introduces may be excessive and only partially filled cells may be used.
- (b) Video. As discussed in Section 8.4, the main technique for reducing the capacity required for a video signal is to transmit only information relating to the changes in picture content. This varies with the action in the picture and so is ideally suited to the variable capacity of ATM.
- (c) Signalling. This would be based on the standards for narrow band ISDN, but assembled into cells. Some of these cells may have to have high priority (e.g. flow control) particularly under congestion conditions.

9.6 THE FUTURE

The aim of telecommunication has long been to simplify its use to the same level as that required for electricity supply. For the domestic and office environment power comes from a socket which is suitable for a wide range of equipment ranging from a clock consuming less than a watt, up to a kettle consuming a kilowatt or two. It has long been held as a dream that a similar arrangement could exist for telecommunications—a universal socket which could be used for anything from low speed data for telemetry and control, right up to high definition television. The promise of broadband ISDN is that this vision will become reality. The only question is when?

REFERENCES

CCITT Recommendations:

- G.707 Synchronous digital hierarchy bit rates.
- G.708 Network node interface for the synchronous digital hierarchy.
- G.709 Synchronous multiplexing structure.
- I.113 B-ISDN Vocabulary.
- I.121 Principles of B-ISDN.
- I.150 B-ISDN ATM functional characteristics.
- I.211 B-ISDN Service aspects.
- I.311 B-ISDN General network aspects.
- I.321 B-ISDN Protocol Reference Model.
- I.327 B-ISDN network functional architecture.
- I.361 ATM layer specification for B-ISDN.
- I.362 Adaptation layer functional description for B-ISDN.
- I.363 Adaptation layer functional specification for B-ISDN.
- I.413 B-ISDN User-network interfaces.
- I.432 B-ISDN UNI-PMD layer specification.
- I.610 B-ISDN OAM principles.

Index

1.544 Mbit/s 25, 95, 160 2.048 Mbit/s 26, 95, 160 2B1Q 58, 63 4B3T 63, 95 64kbit/s 9, 10, 11, 14, 19, 22, 83, 95, 107, 116 7 layer model 13, 106 A-law 21, 117 Accunet 12 Activation 68, 71 Adaptive Differential Pulse Code Modulation (ADPCM) 118 Adjunct switch 32 Administrative Unit (AU) 162 Aliasing 20, 117 Alternate Mark Inversion (AMI) 27, 63, 70 American National Standards Institute (ANSI) 58, 63, 162 Applications 14, 103, 115, 157 Approvals 91 Associated channel signalling 25, 33 Asynchronous Transfer Mode (ATM) 161, 169 Attenuation 5, 10, 27, 37, 39, 91 B channel 10, 14, 59, 65, 67, 91, 107 B8ZS 28, 95 Basic rate interface 10, 65, 92, 155 Bearer service 115 Bipolar see AMI Bit stealing 12, 26 Blue Book 12, 97 Broadband 157 Burst mode 47 CCITT 10, 12, 15, 65, 97, 115, 134, 142, 160 CCITT no 7 33, 89, 100 Cells 169 Circuit switch 8, 19, 29, 116 Coaxial cable 6, 9, 95 Colour space 141 Common channel signalling 9, 26, 33 Common Intermediate Format (CIF) 145

Compression 20, 125, 134, 144 Computer 6, 9, 95, 103, 154 Conferencing 123, 134, 142, 157 Control plane (C-plane) 14, 67, 109 Copper pairs 10, 26, 37, 89 Crossbar 6 Crosstalk 27, 41, 50 Customer interface 10, 13, 65, 95, 161 Customers premises equipment 115, 153 D channel 10, 14, 59, 65, 67, 82, 91, 95, 107 Data 7, 103, 108, 116, 123 Datagram 104 Deactivation 68, 71 Digital Access Signalling System (DASS) 97 Digital Private Network Signalling System (DPNSS) 97 Discrete Cosine Transform (DCT) 135, 148 Duplex transmission 46 Echo canceller 49, 52, 60, 62 Echo channel 71 Equalization 44, 59, 62 Erlang 159 Facsimile 8, 10, 84, 116, 124, 157 Federal Communications Commission (FCC) 93 Flag 73, 110 Flow control 104, 170 Frame Check Sequence (FCS) 73, 79 Frame mode 108 Frame relaying 110 Frame switching 111 Framing see Synchronisation Frequency division 6, 20 Frequency division duplex 47 H channels 160 HDB3 28, 95 High definition TV 157, 161, 171 High Level Data Link Control (HDLC) 73, 97, 110, 130

High quality speech 117 Huffman coding 127, 139, 141 Hybrid 46 IDA (Integrated Digital Access) 11 Impedance 39 Impulsive noise 41, 51 Integrated Digital Network (IDN) 9, 25, 89 International Standards Organisation (ISO) 13, 92, 134 Intersymbol Interference (ISI) 43 ISDN User Part (ISUP) 34 Jitter 55, 62 Justification 163 Key system 155 Link Access Protocol (LAP) 73, 81, 108, 130 Link layer 14, 73 Loading 10, 27 Local Area Network (LAN) 108, 157 Local network 10, 37 Maintenance 58, 89 Message Transfer Part (MTP) 33 Microwaves 6, 9 Mobile 23 Modem 5, 7, 23, 65, 116, 157 Modulation 42 Motion compensation 147 Mu-law 20, 117 Multifrequency signalling 9 Multiplexing 6, 25, 32, 73, 155 Music 124 $N \times 64$ kbit/s 157 National User Part (NUP) 34 Network layer 14, 82 Network service layers 14 Normes Européennes de Télécommunications (NETs) 92 NT1 13, 58, 63, 65 NT2 13 Nyquist 20 Open Systems Interconnect (OSI) 13 Optical fibre 9, 23, 32, 91, 95, 161 PABX 48, 95, 97, 100, 154 Packet handler 107 Packet switch 8, 103, 116, 169 Phase characteristics 39 Physical layer 14, 67 Ping-pong 47

Power 66, 96 Presentation layer 14 Prestel 117, 132 Primary rate 26, 92, 95, 155 Processor 9, 26, 30, 33 Pseudo ternary see AMI Pulse Amplitude Modulation (PAM) 42, 58 Pulse Code Modulation (PCM) 8, 19, 25, 117 Quadrature mirror filters 121 Quantizing 20, 118, 138 Quaternary 43, 58 Red Book 12, 97 Reference configuration 13, 58 Remote unit 30, 95 RS232 11, 154 S reference point 13, 91 S-bus 65 Sampling 19, 44, 118, 144 Satellite 6 Scrambler 55, 60, 162 Service Access Point Identifier (SAPI) 73, 76 Session layer 14 Setup 83 Shannon 7, 49, 157 Signalling 10, 33, 65, 96, 154, 170 Slow scan 84, 134 Space switch 29 Specification and Description Language (SDL) 92 Stored program control 33 Strowger 6 Subband ADPCM 121 Supplementary services 88 Synchronization 26, 58, 70, 162 Synchronous Digital Hierarchy (SDH) 162 Synchronous transport module 162 T reference point 13, 63, 91 Telegraph 1, 8 Telephone 1, 10, 65, 77, 92, 154 Telephony User Part (TUP) 34 Teleservice 115 Teletex 8, 84, 126 Telex 5, 7, 8, 126 Terminal adaptor 13, 92, 107, 154 Terminal Endpoint Identifier (TEI) 73, 77, 81, 87 Testing 91 Time Division Duplex (TDD) 47 Time Division Multiplex (TDM) 25, 29

Time switch 29 Timing 22, 27, 46, 47, 55, 62 Transistor 7 Transmission 26, 37 Transport layer 14 Tributary Unit (TU) 167

U reference point 13, 58, 63 User plane (U-plane) 14, 67, 109

V.35 12, 154

Very Large Scale Integrated Circuits (VLSI) 9, 12, 23, 28 Victoria 11 Video 84, 108, 117, 123, 142, 145, 157, 170 Videotex 84, 117, 132 Virtual Call Identifier (VCI) 169 Virtual circuit 104 Virtual container 163 Virtual Path Identifier (VPI) 169

X.21 11, 116, 154 X.25 73, 106, 108, 112, 116, 131, 154

SDN EXPLAINED

Worldwide Network and Applications Technology

John M. Griffiths

with the staff at British Telecom, UK

ISDN (Integrated Services Digital Network) is an exciting new communications development. John Griffiths presents a highly readable overview of the principles of all facets of the subject, giving an international perspective.

The telephone network is now operating digitally: the resulting ability to carry data at higher speeds means that a new range of interfaces has been developed (in line with internationally agreed standards) which will supersede all existing telephony and data interfaces. This book discusses the evolution of both networks and terminals, the nature of the interfaces, and the way in which terminals interact with the network.

Contributions from various members of the highly experienced team at British Telecom cover the users' needs and technology, pulse code modulation, digital networking, local area transmission and ISDN basic rate interface. The final section is concerned with the applications: high quality speech, fax, video and packet services.

Readers with a general knowledge of electronics will be able to understand the operation of ISDN, and will be provided with an introduction to detailed texts such as international standards.



JOHN WILEY & SONS Chichester · New York · Brisbane · Toronto · Singapore