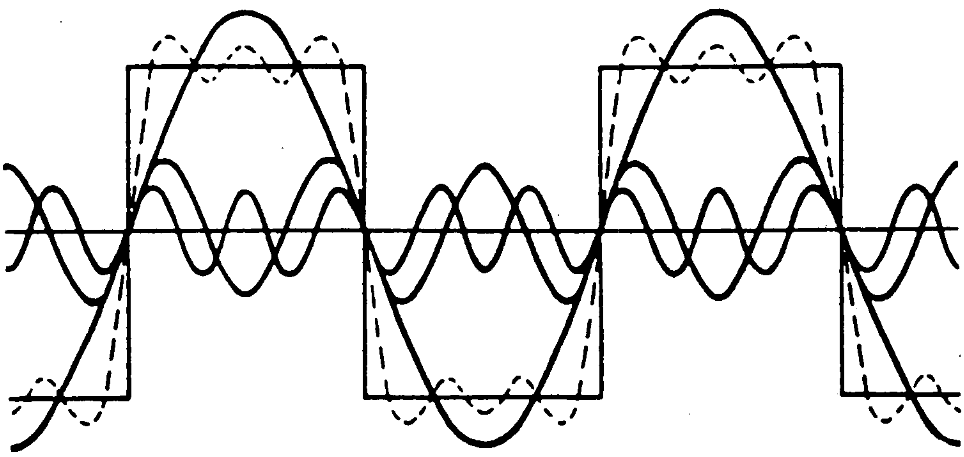
Transformation Series

Collection Transformation



BUILDING A DIGITAL NETWORK: DATA COMMUNICATIONS AND DIGITAL TELEPHONY, 1950–1990

Bryan Dewalt

2

Transformation Series

"Transformation," an occasional paper series published by the Collection and Research Branch of the National Museum of Science and Technology is intended to make current research available as quickly and inexpensively as possible. The series presents original research on science and technology history and issues in Canada through refereed monographs or collections of shorter studies, consistent with the Corporate framework, "The Transformation of Canada," and curatorial subject priorities in Agriculture and Forestry, Communications and Space, Transportation, Industry, Physical Sciences and Energy.

The Transformation series provides access to research undertaken by staff curators and researchers for development of collections, exhibits and programmes. Submissions will be considered from academicians or professionals working in areas complementary to curatorial subject priorities of the Collection and Research Branch.

Papers are printed in the language of submission (French or English) with a bilingual abstract.

Enquiries regarding the submission of manuscripts may be sent to the Managing Editor at the address below.

To Order

Additional copies of this publication are available for $10.00 (includes GST), while quantities permit. A cheque or money order payable to the Receiver General for Canada must accompany all orders and may be addressed to:

Finance Division – Transformation Series
National Museum of Science and Technology
P.O. Box 9724, Ottawa Terminal
Ottawa, Ontario
K1G 5A3

Transformation Series

Collection Transformation

2

Building a Digital Network: Data Communications and Digital Telephony, 1950-1990

Bryan Dewalt

National Museum of Science and Technology
Ottawa, Canada
1992

ISSN 1188-2964

Canadian Cataloguing in Publication Data

Dewalt, Bryan, 1959-

Building a digital network : data communications
and digital telephony, 1950-1990

(Transformation series, ISSN 1188-2964 ; 2)
Includes an abstract in French.
Includes bibliographical references.
ISBN 0-660-12018-6

1. Digital communications -- History. 2. Digital
telephone systems -- History. 3. Data transmission
systems -- History. I. National Museum of Science
and Technology (Canada). II. Title. III. Title: Data
communications and digital telephony, 1950-1990.
IV. Series.

TK5103.7D58 1992 621.382 C92-099614-0

Cover photo: Analysis of the first three harmonics of a
square wave. (Martin, Telecommunications and the
Computer, [1976], 151.)

Contents

Abstract

The period between about 1960 and 1990 was marked by major changes in the technology and organization of telecommunications in Canada and the rest of the industrialized world. The key to this change was the evolution of a versatile, multi-purpose network. By converting all types of information to digital pulses, this network could accommodate everything from telephone conversations, to written messages, to computer data, to photographs and moving pictures.

While this convergence of technologies is not yet complete, it is clear that the line between data and voice communications systems has disappeared. Designers of data communications networks have used the telephone network to gain access to offices and homes. And operators of the public telephone network have introduced new digital technologies to serve the growing market for information transfer. Many observers believe that newspapers, television and other information and entertainment services will also be integrated into this system.

But consumers have not yet found the need for the type of omnibus service being proposed. Much information is still conveyed by print media, and entertainment is still delivered through the physically separate cable television network. Despite technical integration, the telecommunications network in 1990 consists of a universal basic telephone service overlaid with and connected at key points with specialized networks serving data users.

At least in the short term, this multitier structure will persist. It is nevertheless clear that changes in telecommunications have been felt throughout society and that these changes are intimately tied to the simultaneous growth in the use of computers. This report, therefore, concludes with an analysis of the literature of social change that has developed in the wake of the computer and telecommunications boom. It emphasizes the economic and political forces that have driven and shaped the development of the new information technologies.

Résumé

La période de 1960 à 1990 a été marquée par des changements majeurs dans la technologie et la structure des télécommunications, tant au Canada que dans le reste du monde industrialisé. La clé de ces changements a été le développement d'un réseau polyvalent. En convertissant tous les types d'information en impulsions numériques, il est devenu possible d'intégrer à ce réseau aussi bien les conversations téléphoniques que les messages écrits, les données informatiques, les photographies et les images en mouvement. Bien que cette convergence de technologies ne soit pas encore achevée, il apparaît évident que la frontière entre les systèmes de transmission des données et ceux de communication vocale a désormais disparu. Les concepteurs de réseaux informatiques ont utilisé le réseau téléphonique pour assurer les liaisons aussi bien au bureau qu'à domicile. De plus, les exploitants du réseau téléphonique public ont mis en œuvre de nouvelles techniques numériques pour servir le marché croissant du transfert d'information. De nombreux observateurs sont d'avis que les journaux, la télévision et d'autres services d'information ou de divertissement seront aussi intégrés à ce système. Mais les consommateurs n'ont pas encore ressenti le besoin du type de service « tout compris » qui leur est proposé. Une grande partie de l'information est toujours transmise au moyen de l'imprimé, alors que le divertissement passe encore par le réseau de télévision par câble, matériellement distinct du réseau des télécommunications. Malgré son intégration technique, le réseau des télécommunications se compose toujours en 1990 d'un service téléphonique universel de base, auquel se greffent en des points névralgiques des réseaux spécialisés au service des utilisateurs de systèmes de transmission de données.

À brève échéance, tout au moins, cette structure à plusieurs niveaux persistera. Il est cependant clair que les changements survenus dans le secteur des télécommunications se sont fait sentir dans l'ensemble de la société, et que ces changements sont intimement liés à l'accroissement simultané de l'utilisation des ordinateurs. En conséquence, le présent rapport se termine par une analyse de la documentation disponible sur l'évolution sociale observée dans le sillage de cette formidable expansion de l'informatique et des télécommunications. Nous y mettons en lumière les forces économiques et politiques qui ont entraîné et conditionné l'émergence des nouvelles technologies de l'information.

Foreword

The development of the electric telegraph and the invention and spread of the telephone in its first century are relatively well known stories.

The invention of the transistor in 1948, and the subsequent development of the integrated circuit have brought about dramatic revolutions in the scope and capability of communications systems and devices. For example, they have reduced the massive computer and the expensive facsimile machine in size and price to the point where they are available to the average person for the at-home office. Information is now routinely exchanged between such devices over the regular communications network.

These same ICs have also made the wholesale conversion of our telecommunications system from analogue to digital circuitry possible. This revolution has not yet been documented in any one place and is only covered in the specialized press.

When the National Museum of Science and Technology decided to produce an exhibit on the subject of communications, one of the objectives was to try to bring together the story of this revolution with particular reference to Canada. This paper was researched and written in 1989/90 and in addition to becoming the basis for the modern end of one section of the exhibit, it will be used to identify key developments and thereby guide the future development of the National Collection in this area.

The communications field is changing so rapidly that both the exhibit and the document will need regular updating to remain current.

January 1992

E. A. DeCoste,
Senior Curator,
Communications and Space

Introduction

The twentieth century has seen the spread of the public telephone system to all corners of the earth. It is the most accessible, flexible, and comprehensive communications system we have devised. Its technology dates from the last quarter of the 19th century and in some respects it operated until the 1960s much as it did almost a hundred years earlier. A person picked up the telephone, sending a signal along a pair of copper wires to a central office, or exchange, requesting connection to another subscriber. At the exchange, where all wires for a local area converged, an operator or a machine arranged a physical connection between the subscriber's line and that of the intended recipient of the call. If that person responded to the call, a conversation would take place. This conversation depended on the conversion of a sound wave into an analogue electrical signal, which passed along the copper wires, through the exchange and to the listener's telephone. There another conversion turned the electrical signal back into a sound wave. The conversation, which passed over the "local loop" that connected each telephone subscriber to the central office, could be extended across the city, across the country, or around the world by sending it through a complex web of trunk cables and toll exchanges that connected virtually every telephone on earth.

By 1960 there had already been many technical innovations in the telephone system. Manual switching by human operators was gradually replaced by electromechanical and then electronic switches controlled by machine logic. Long distance transmission was extended around the earth by the use of complex and delicate vacuum tube amplifiers placed along wire lines and by piggybacking telephone signals on a radio wave for transmission across the ocean or from one microwave tower to another. As telephone traffic increased, telephone companies developed ways to combine, transmit, and separate numerous conversations in order to share transmission lines or radio channels. This technique was called frequency division multiplexing because it involved transmitting over the same medium separate conversations in adjacent high frequency bands of the radio spectrum.

Beside this telephone network, another system based on an even older technology was also evolving. Telegraph systems could not compete with the telephone network in sweep or ubiquity. But certain types of information, most notably the written word or numbers of any kind, could be cheaply and easily converted to coded pulses of electromagnetic energy and transmitted to a distant destination. Telegraphy therefore met a demand by offering a variety of services. These included the telegram, now only seen in old movies or at wedding receptions, and stock and commodity price reports. The most sophisticated telegraph services were TELEX and TWX, switched networks of automatic teletype stations once found in many offices, and the facsimile transmission of pictures. The exploitation of both, however, was limited to specialized users in business, government, and journalism.

In the 1950s and 1960s, two things brought about a change in this state of affairs. On the one hand governments, the military, universities, and large corporations discovered the utility of the computer and the need for a means of passing digital data back and forth between them over great distances. Most of this traffic passed over circuits leased from the established telephone and telegraph companies. At the same time, telephone companies turned to digital techniques to increase the capacity of their existing cables for voice transmission and to streamline the immensely complex task of switching. Instead of modifying a continuous wave or electrical signal, digital telephony would convert messages into pulses.

Beginning slowly in the 1960s and picking up momentum in the 1970s and 1980s, the telephone system became the nexus for a new trade in coded pulses. Information of any kind was converted into bits and megabits and was shunted over a variety of voice and data networks that used all or part of the public telephone system. Increasingly, observers talked of the development of a single, integrated network that would transmit all kinds of information: voice, numerical data, printed information, television signals, and anything else that could be converted to a digital code.

This phenomenon was not unprecedented. Telegraph companies had often leased circuits from their telephone counterparts. The latter had from time to time offered their own telegraph services over the same lines as their voice services. And analogue television signals were already sent over the same kinds of transmission media as telephone calls. But many observers identified something socially revolutionary in the technological trends of the 1960s and 1970s. This was only partly related to developments in telecommunications.

For some, the key was the relative decline of heavy industry and manufacturing as an employer of the work force in industrial countries. More and more people seemed to be finding work in service trades and in occupations that emphasized the processing of information and application of abstract knowledge. At the same time, the computer was becoming a common fixture. Entering the popular imagination as a capricious monolith of sheet metal and flashing lights, it was immediately identified as a new kind of machine. It operated according to strict rules of logic, had a huge memory for detailed information, and could perform mind-numbing routines of calculation in a split second. At the same time its inner workings were mysterious, involving invisible impulses that shot through arrays of electronic devices. But the computer could be made to communicate in our own language and to process valuable information for us. It was therefore perceived to have human-like qualities and, in the case of its memory and computing speed, qualities superior to humans.

The development of cheap, mass produced integrated circuits in the 1960s met the technical demand for a further expansion of electronic information processing. In the late 1970s the first personal computers were introduced into homes and offices. Most users of these new machines remained ignorant of their internal workings. But by accepting them as tools that could help perform important tasks, they affirmed just how central computers had become to the way they worked.

Aided by the same integrated circuits and microprocessors that ran computers, telecommunications was evolving to bring together these two other trends—computers and the changing workforce. Information workers in modern corporations and bureaucracies had to be able to talk to each other, sometimes over great distances. They also had to have access to computers across the hall or halfway around the world and to exchange increasing quantities of data with them. In providing the transmission and switching services for this interaction, the telecommunications network itself became a huge information processing system. It became a part of what many perceived to be the infrastructure of a new "post-industrial" or "information" society. Computers and telecommunications together came to be called "information technology." Today it is everywhere: in the automatic money machines that have replaced human contact in most bank transactions; in the computerized pricing and inventory systems at the supermarket checkout; in the computers that now sit on so many office desks; in the "smart" telephones that store phone numbers and the telephone branch exchanges that record and relay voice messages; in the inexpensive fax machines that read a sheet of paper and send an image of it as streams of coded pulses.

The impact of information technology and the nature or even existence of the information society has become a subject of debate. Some observers marvel at the capabilities of the technology and the solutions it offers to cumbersome problems of production or information access. Others study the possibilities for social control inherent in a technology that is able to monitor individuals and store and exchange information on them. Entire careers have been made as consultants to perplexed government and corporate leaders. While many industry studies focus on the justifiable importance of technical developments, most notably low-cost integrated circuits and microprocessors, we should not neglect the economic and political forces that have driven and shaped the development of the new information technologies. This report will therefore conclude with an analysis of the debate over the role of information technology in our society.

But before this is possible it is important to trace the technical history. This report is concerned with the communication systems that form one half of the information technology concept. It will begin with a discussion in Part 1 of the general technical considerations and specialized language that underpin an understanding of recent telecommunications development. The history of specialized data networks will be covered in Part 2. Part 3 will address the digitization of the telephone network and Part 4 will look at the integration of voice and data networks.

1 Communication Theory and Channel Capacity

Communication is the process of generating a message in one place and reproducing it in another. This can be as simple as a speaker being heard and understood by a listener sitting across the room. Or it can involve a person reading a letter received in the mail from a friend in another city. Since the mid-19th century, information passing from the source to its destination has relied on increasingly sophisticated technologies employing coded signals composed of electromagnetic waves or pulses. The speaker speaks into a telephone and is "heard" miles away. The letter writer types at a data communications terminal or feeds a handwritten sheet into a fax machine. Or, an actor appears before a television camera while the viewer watches and hears an electronic reproduction of the actor on his or her television. Communication may even take place between a human being and a computer or between two computers. We remain, however, the ultimate destination of almost all information passing through computers. In this sense, they are only a stage in a communication process between people, a location where information is stored and processed for human ends.

Any communication system must reproduce a message in a reasonable amount of time, with reasonable accuracy, and at a reasonable cost. All three of these criteria can vary, depending on our needs or expectations. In considering these criteria, designers of a communication system must also understand a fundamental constraint imposed on all communication systems: the limited capacity of any channel to carry a specific amount of information in a specified time. Taking into account the purpose of the system and its intended users, a system designer must deal with:

- the amount of information in a given time that a source generates for transmission;
- the most efficient coding system for converting this information into a signal;
- the physical properties of the transmission medium that limit capacity and cause signal losses;
- "noise" on the channel that distorts the signal and prevents the accurate reproduction of the message;
- and the signal power, derived from the transmitter and amplifiers or repeaters, needed to overcome noise and compensate for signal losses.

Our standards for acceptable cost, fidelity, and speed will help determine the best solution to these technical problems.

The common phenomenon that is actually transmitted in a communication system is electromagnetic energy, be it electric current through a solid conductor, radiation through space, or light through a light guide. As either a digital or an analogue signal, this energy is modified to embody a message. Digital signals are chains of discrete pulses. A common type is the binary signal, one which can exist in one of two states: eg. on/off or high voltage/low voltage. Digital signals are valued for their simplicity and flexibility in representing diverse types of information. Analogue signals consist of continuously varying values of energy that rise and fall in a wave-like pattern. Their main advantage is that their wave pattern can be made similar, or analogous, to many natural phenomena that are transmitted as information, especially sound and light waves. Telegraphy and modern data communications are digital systems, while telephony and television are, at least in their early histories, analogue.

Given appropriate encoding, we can transmit virtually any message as electromagnetic energy in either analogue or digital form. Theoretically then, a communication system can handle any message, whether it be a telephone conversation, a data transmission, or a television signal. Different messages, however, contain different amounts of information, and the capacity of a channel will determine how fast this information can be sent.

Information Can be Quantified

From an engineer's point of view, the content or meaning of a message is irrelevant to the problem of designing a system. The engineer examines the source and its information mathematically, to determine an abstract amount of information generated. While we can easily distinguish between the relative information content of a one-word telegraph message and that of a telegraphed version of *War and Peace*, it is less easy to distinguish the quantities of information in a telephone conversation or a photograph. Information theory, first comprehensively stated by Claude Shannon of Bell Laboratories in 1948, addresses this very problem by translating all messages into binary numbers.

When we generate a message, we are actually selecting one from a finite set of possibilities. When we speak, for example, we chose from a limited number of sounds that it is possible for a human to make. But because most utterances involve a complex series of sounds in various combinations, we cannot easily predict the exact message before it is spoken. According to Shannon, the more complex, or unpredictable, the message from the source, the more information it carries. A source that produces only one of two possible messages (eg. "yes" or "no") with equal probability, has less information value than a source that might, again with equal likelihood, select one of twenty-six messages (the number of letters in the alphabet). A source that invariably produces the same message has no information value because it is totally predictable.

Because of the frequent use of binary digital signals in communications systems, Shannon used binary numbers to characterize the information produced by a source.[1] This can be represented by the on/off switch of a simple telegraph key. In the case of a "yes/no" source, a single relay and a binary coding system of 1 (on) for "yes" and 0 (off) for "no" could easily embody all the information. To choose one of the twenty six letters of the alphabet, each letter would require a unique pattern of several 1's and 0's to carry the information needed to reproduce the letter without ambiguity. Automatic telegraphy has long used a five digit binary code that provides 32 different combinations of 1's and 0's, enough for each letter of the alphabet, several printer control characters, and, by using "shift" keys, numerals and most punctuation. In North America, much modern data communications employs a seven digit code that allows for 128 configurations. Each binary digit represents one "bit" of information. The larger the number of possible choices at the information source, the more bits are required to represent each possible choice.[2]

1. Unless otherwise stated, binary codes and signals are assumed throughout this report.

Most messages with which we are familiar require more than a single choice of yes/no or A,B, C to Z. A sentence or paragraph of text requires a choice for each letter, with each letter embodying several bits. A page-long message sent by TELEX or electronic mail page might contain 10,000 bits of information. A facsimile page that also reproduces the original size, shape, style, and intensity of each letter, would require 200,000 bits Hence a facsimile of a page from a Gutenberg Bible will require many more bits than a simple transcription of its text. A newspaper photograph, which embodies many choices about the blackness or whiteness of each fraction of the image, is estimated to contain 100,000 bits of information. A colour television frame requires approximately 1 million bits. And a high-quality colour photograph takes twice as many.

The capacity of an information channel is really a measure of how many information bits it can carry in a given period of time. Channel capacity is generally measured in bits per second (b/s). Returning to the elementary example above, if one second is assumed to be an acceptable delivery time, then the one bit required to send the message "yes" will require a channel capacity of at least one bit per second. But if the letter E is represented by seven binary digits, the channel must have a capacity of seven bits per second. Data rates for teletype, a system of automatic telegraphy that has been in use for decades, range from 50 to 200 bits per second.[3] Modern data communication systems have capacities ranging in the thousands and millions of bits per second; the reproduction of a letter E requires just a fraction of a second. Units for high capacity circuits are stated as: kilobits (Kb/s) for thousands, megabits (Mb/s) for millions and gigabits (Gb/s) for billions. Where possible this report uses the simpler expressions thousand b/s, million b/s, and billion b/s.

2. A five-bit binary code has 2^5 possible combinations ($2 \times 2 \times 2 \times 2 \times 2 = 32$). A seven-bit binary code has 2^7 possible combinations ($2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 = 128$).
3. This rate is determined both by the limited capacity of telegraph circuits used for transmission and by the relative slowness of standard electromechanical transmitters and receivers.

Character		Number of symbol	5-unit international code No. 2 (used by telex machines)
Letters shift	Figures shift		
A	–	1	Z Z A A A
B	?	2	Z A A Z Z
C	:	3	A Z Z Z A
D	Who are you?	4	Z A A Z A
E	3	5	Z A A A A
F	•	6	Z A Z Z A
G	•	7	A Z A Z Z
H	•	8	A A Z A Z
I	8	9	A Z Z A A
J	Bell	10	Z Z A Z A
K	(	11	Z Z Z Z A
L	)	12	A Z A A Z
M	.	13	A A Z Z Z
N	,	14	A A Z Z A
O	9	15	A A A Z Z
P	0	16	A Z Z A Z
Q	1	17	Z Z Z A Z
R	4	18	A Z A Z A
S	'	19	Z A Z A A
T	5	20	A A A A Z
U	7	21	Z Z Z A A
V	=	22	A Z Z Z Z
W	2	23	Z Z A A Z
X	/	24	Z A Z Z Z
Y	6	25	Z A Z A Z
Z	+	26	Z A A A Z
< Carriage return		27	A A A Z A
≡ Line feed		28	A Z A A A
↓ Letters		29	Z Z Z Z Z
↑ Figures		30	Z Z A Z Z
■ Space		31	A A Z A A
(Not used)		32	A A A A A
Signal repetition			–
Signal α			(Permanent A polarity)
Signal β			(Permanent Z polarity)

•: Not used internationally; reserved for national allocation.

EQUIVALENT MEANINGS OF THE ABOVE A AND Z CONDITIONS

	A Condition	Z Condition
Bits:	0	1
Start-stop code:	Space (start condition)	Mark (stop condition)
Holes (perforations) in paper tape:	No hole	Hole
Single-current signaling:	No voltage	+ ve voltage
Double-current signaling:	– ve voltage	+ ve voltage
Amplitude modulation	Tone-off	Tone-on
Frequency modulation	High frequency	Low frequency
Phase modulation with reference phase	Opposite phase to reference phase	Reference phase
Differential phase modulation	Inversion of the phase	No inversion

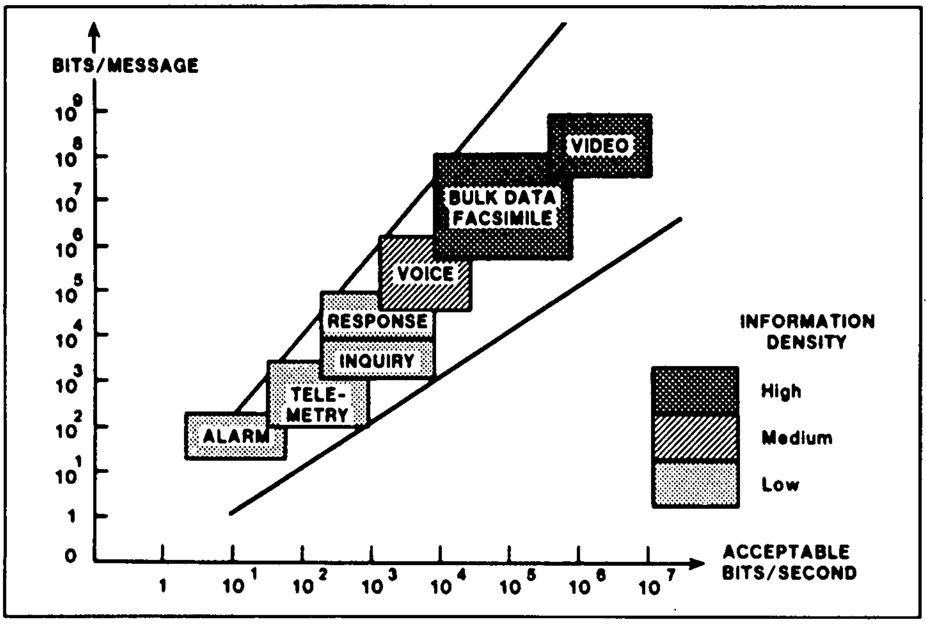
Five-bit CCITT international code for TELEX. The code for letter E is ZAAAA, more commonly expressed as 10000. (Martin, Telecommunications and the Computer *[1976], p.59)*

		Bit positions 5, 6, 7:							
		000	100	010	110	001	101	011	111
		0	1	2	3	4	5	6	7
Bit positions 1, 2, 3, 4: 0000	0	NUL	DLE	SP	0	@	P	`	p
1000	1	SOH	DC1	!	1	A	Q	a	q
0100	2	STX	DC2	"	2	B	R	b	r
1100	3	ETX	DC3	#	3	C	S	c	s
0010	4	EOT	DC4	$	4	D	T	d	t
1010	5	ENQ	NAK	%	5	E	U	e	u
0110	6	ACK	SYN	&	6	F	V	f	v
1110	7	BEL	ETB	'	7	G	W	g	w
0001	8	BS	CAN	(	8	H	X	h	x
1001	9	HT	EM	)	9	I	Y	i	y
0101	10	LF	SUB	*	:	J	Z	j	z
1101	11	VT	ESC	+	;	K	[	k	{
0011	12	FF	FS	,	<	L	\	l	¦
1011	13	CR	GS	–	=	M	]	m	}
0111	14	SO	RS	.	>	N	^	n	~
1111	15	SI	US	/	?	O	_	o	DEL

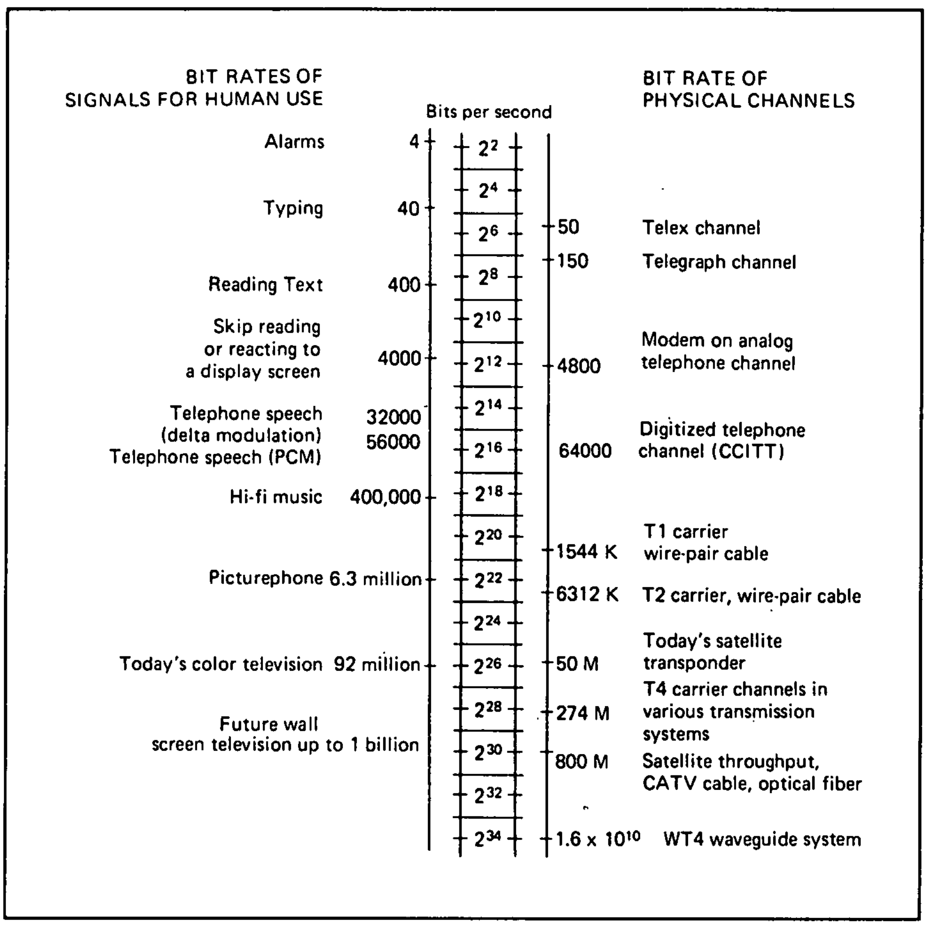
The U.S. ASCII code for data exchange is a seven-bit code. The code for letter E is 1010001. (Martin, Telecommunications and the Computer *[1976], p. 66)*

Message Type	Bits
1. A high-quality color photograph	2 million
2. A newspaper-quality photograph	100,000
3. A color television frame	1 million
4. A Picturephone frame	100,000
5. A brief telephone voice message	1 million
6. A vocoder telephone voice message	100,000
7. A voice message of code-book words	400
8. A document page in facsimile form	200,000
9. A document page in computer code	10,000
10. A typical interoffice memo	3,000
11. A typical telegram	2,000
12. A typical flip chart	1,000
13. A typical computer input transaction	500
14. A typical electronic fund transfer	500
15. A typical airline reservation	200
16. A coded request for a library document	200
17. A fire or burglar alarm signal	40

Typical numbers of bits needed for different types of message. (Martin, Communication Satellite Systems *[1978], p. 216)*



Acceptable transmission speeds for different types of message. (Reprinted from Telephony, *25 October 1982. ©1982 by Telephony Publishing, Chicago, Ill. All rights reserved)*



A comparison of the speeds of physical channels with speeds of signals employed by humans. (Martin, Future Developments in Telecommunications *[1977], p. 65)*

Sine Waves, Harmonics and Bandwidth

Until quite recently, almost all information was transmitted by analogue signals. A sound or light wave generated at the source was converted in the transmitter (of say, a telephone or a TV camera) into a complex electrical wave. In its irregularly varying frequency and amplitude, this electrical wave embodied the same pattern of variations in the original wave. The ability of a channel to accurately transmit this wave was determined by "Fourier analysis." According to Jean Baptiste Joseph Fourier, a mathematician of the early 19th century, any complex wave can be analyzed as the sum of countless simple sine waves of appropriate frequency and amplitude. A musical note played on an instrument, for example, comprises a fundamental tone plus a series of higher tones, each of a different frequency, or pitch, and a different amplitude, or degree of loudness. Each of these pure tones may be represented mathematically and graphically as a sine wave. Combined, these "harmonics" produce a unique, complex sound wave.[4] A complex electrical wave, a current whose amplitude varies with time at varying frequencies, can also be analysed as the sum of a number of sine waves. By making the analogy between sound waves and electrical waves, in fact, Alexander Graham Bell gained the essential insight that evolved into the telephone.

If we are to exactly receive a signal generated at the transmitter, our communication channel must accommodate a signal containing all frequencies in the range between the lowest and highest frequency waves in the signal. This range is called the bandwidth. While an absolute rule, communication system designers have, in fact, always bent this rule to suit the varying degrees of fidelity required in the received signal. While a sensitive ear can hear vocal sounds ranging in frequency from 30 Hz to 20,000 Hz, voice recognition requires reproduction only of frequencies between 300 and 3100 Hz. In order to economize, therefore, telephone engineers design transmission channels with a bandwidth of about 3000 Hz. Transmission with higher fidelity, perhaps for an FM stereo broadcast, would use a bandwidth more closely approaching that of the original.

4. Because each instrument produces some harmonics stronger than others, an identical note will sound different on, say, a trumpet or a piano.

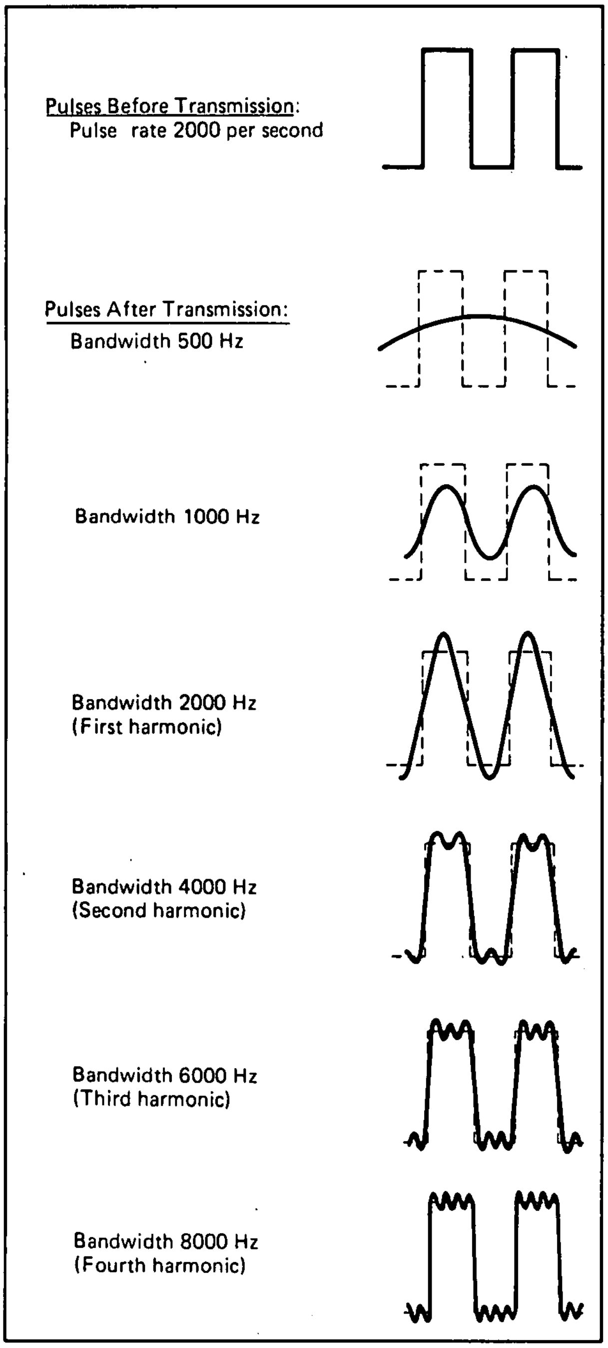
We can also apply Fourier analysis to a digital signal, making a connection between the two measures of channel capacity, bandwidth and bit rate. A digital signal of electrical pulses may be considered a square wave (amplitude rising and falling abruptly) and analyzed as the sum of an infinite range of harmonics. The fidelity of the received signal depends on the number of harmonics transmitted, which in turn is limited by the bandwidth of the circuit. Rather than send all harmonics in order to exactly reproduce the square wave, we overcome the bandwidth limitation by sending only enough to allow recognition and regeneration of the original bit pattern. A slow bit rate requires relatively few frequencies in order to be recognizable. But by increasing the bit rate, the spectrum of the resulting square wave is increased and a wider bandwidth is required for acceptable transmission quality. In digital transmission, therefore, the quantity of information a channel can carry is directly proportional to its bandwidth. With binary coding, each time we double the capacity in bits per second we must also double the bandwidth.

Noise is an Inherent Limit on Channel Capacity

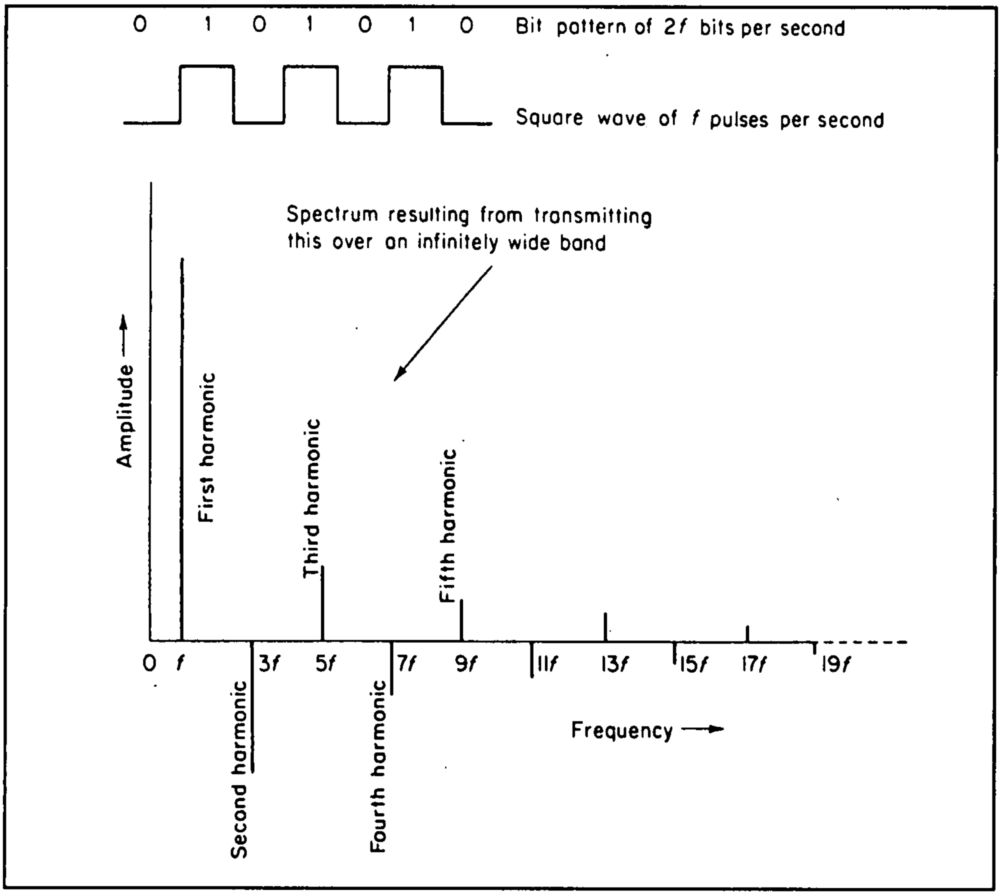
Various transmission media have different physical properties that determine the range of frequencies they can carry. Some physical media like coaxial cable or radio may have larger bandwidths than a pair of copper wires, but none can avoid one limitation on any channel's capacity. This limitation is noise. Just as conversation in a crowded party can be drowned out by the surrounding hubbub, so too electrical signals must compete to be heard.

In communication usage, noise is as any random electrical signal that disturbs a communication channel and obscures the clarity of the received message signal. Variously called white noise, Gaussian noise, or thermal noise, we find these random signals in any communication channel. They are caused by the vibration of electrons in all substances, which causes the emission of electromagnetic waves at all frequencies. Some noise is caused by the actual electronic components in a system. Improved design can minimize but not eliminate this noise, and there remains a certain amount of noise that is naturally occurring and constant.

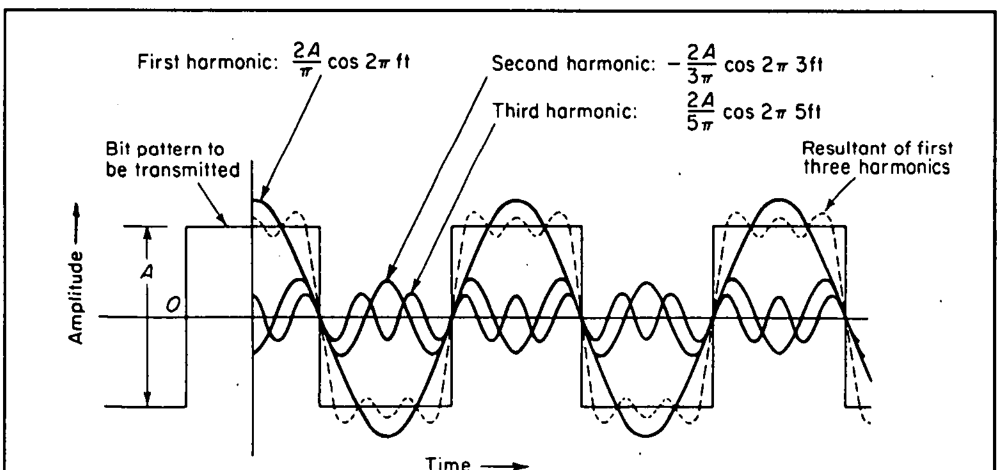
The system designer must ensure that the power of the received signal is sufficient to ensure its recognition above the noise. In digital transmission, for example, the power of signal pulses must



The effect of increasing bandwidth on the quality of a pulse transmission. (Martin, Telecommunications and the Computer *[1976], p. 149)*



The spectral components of a square wave representing a bit pattern of 010101, etc. Over a limited bandwidth only the lower harmonics can be transmitted. (Martin, Telecommunications and the Computer *[1976], p. 150)*



The first three harmonics composing a bit pattern 010101, etc. The sum of the first three harmonics only approximately represents the bit waveform transmitted. (Martin, Telecommunications and the Computer *[1976], p. 151)* be strong enough to distinguish them from random noise pulses. The difference between signal power and noise power is called the signal-to-noise ratio.

Shannon proved mathematically that any channel of a given bandwidth and a given signal-to-noise ratio can transmit only a limited number of information bits per second. To increase the bit rate we must reduce the noise power or increase the signal power. At a noisy party we convey a message to a friend across the room by asking for silence, raising our voices, or passing a message through intermediaries who repeat it for us. Just as in a party, noise power in any communications system cannot be reduced below a certain minimum. We can increase signal power by spacing repeaters closer together or by increasing the power of the transmitter. Both solutions, however, are limited by cost and, because transmitters must often be attached to the telephone system, by the electrical capacities of the system's circuitry. Another option is to tolerate a lower signal-to-noise ratio. This is really only possible with digital transmission, where inexpensive, frequently spaced repeaters can recognize the information in a noisy signal and regenerate a clean pulse. Finally, we can also use various methods and media to increase channel bandwidth. But because some noise increases proportionately to bandwidth, the bit rate does not increase as fast we might hope. Nevertheless the last two solutions, exploitation of easily regenerated digital signals and the development of large bandwidth transmission media, have been used extensively in modern telecommunications networks.

Coding

Despite the absolute limitation of noise, channel capacities in actual communication systems still leave considerable room for improvement. Binary signalling is a common technique in telecommunications, but the information rates it allows are only a fraction of Shannon's maximum. Multilevel coding techniques, which allow more information to be embodied in fewer signal elements, come closer to the ideal. A binary signal exists at two possible voltage levels and can represent one bit of information at any instant. A signal with four possible states, on the other hand, can simultaneously represent two bits of information. Furthermore, a signal with eight possible voltage levels can convey three bits of information at a time. The advantages of more efficient coding for increasing channel capacity are obvious. But regardless of the system used, capacity can still never exceed Shannon's theoretical maximum for a given bandwidth and signal-to-noise ratio. Furthermore, a more elaborate coding requires more complex and expensive transmitters and demands a longer encoding time. At a certain point, excessive coding time cancels out gains in transmission rate.[5]

5. Information for this report was derived from: James Martin, *Telecommunications and the Computer*, Second Edition (Englewood Cliffs, N.J.: Prentice-Hall, 1976), pp. 58–64, 132–86, 302–14; and Claude E. Shannon and Warren Weaver, *The Mathematical Theory of Communication* (Urbana, Ill.: University of Illinois Press, 1949), pp. 3–32.

2 Evolution of Data Networks in North America

Early Data Communications

In September 1940, George R. Stibitz of Bell Telephone Laboratories demonstrated an electromechanical calculating machine at the annual meeting of the American Mathematical Society in Hanover, New Hampshire. By the standards of later electronic computers, this was a primitive machine and a technological dead end. But Stibitz was a computer pioneer in a more enduring sense. During the Hanover demonstration, Stibitz's computer had remained at Bell Laboratories in New York City, 250 miles away. Using a communication system developed by his colleague, Samuel B. Williams, Stibitz transmitted numbers for calculation from a teletypewriter terminal in Hanover. The data passed over telephone lines of AT&T's Teletypewriter Exchange Service (TWX) to New York, and the results returned along the lines and were printed out on Stibitz's machine. The entire process took almost one minute, a glacial speed by modern standards. The Hanover demonstration was the first example of the control of a central computer from a remote terminal, an arrangement often used in the first two decades of electronic computer use. It also anticipated a common approach to computer communications—the use of the existing telecommunications network for this new, specialized task.[6]

UNIVAC I, the first commercially available electronic computer, was introduced by Remington-Rand (later Sperry Rand) in 1952. True data communications networks did not immediately follow. For one thing, computers were so few, and their application was so limited, that there was really little need for communication, let alone a network. "Batch processing," in which data was coded on punched cards or punched tape and handed to a computer operator for entry and processing, was the established method of interaction between people and computers. The first application of electrical communication to data transport did not alter this basic approach to the use of computer time. In the 1950s, the computer companies devised terminals that would accept punched cards, punched tape, or magnetic tape for electrical transmission of data to a distant computer centre. There, a receiver repunched the card or tape or converted the received signal into magnetic tape for feeding into the computer. This was still essentially batch processing, but it allowed a large corporation to install a single centralized computer that would receive bulk data, usually financial records, from a number of branch locations for processing. Transmission was over voice bandwidth lines leased from telephone, or telegraph, companies for exclusive, point-to-point use.

The public telephone network has two great advantages for data communications that were recognized in the 1950s and continue to be exploited today. First, it is already in place, so that the costly construction of a special-purpose data line is unnecessary. Second, the ubiquity and interconnectability of the telephone system offer the possibility of creating a network of terminals communicating with a central computer, or of computers communicating with each other. But in their original, or baseband form, data signals consist of digital pulses of direct current. Before transmission over telephone lines, they must be converted to a form suited to the characteristics of the analogue telephone plant. Analogue telephone circuits in subscriber loops are designed to transmit a bandwidth of about 3000 Hz, in a range between 300 and 3300 Hz in the spectrum. They do not accommodate the energy of signals falling below 300 Hz or above 3300 Hz. Much of the energy of a baseband data signal, however, falls

6. George R. Stibitz, "Early Computers," *A History of Computing in the Twentieth Century*, N. Metropolis et al. eds. (New York: Academic Press, 1980), p. 481; George R. Stibitz & Evelyn Lovejoy, "Relay computers at Bell Labs," *Datamation* (April 1967), p 44; E.G. Andrews, "Telephone Switching and the Early Bell Laboratories Computers," *Bell System Technical Journal*, 42,2 (March 1963), pp. 346–7.

below 300 Hz and would be lost in a telephone circuit.[7]

Modulation is therefore used to convert the baseband direct current data signal into a form for transmission through the analogue telephone network. For transmission in the local loop, modulation involves selecting an alternating current carrier in the voice band and modulating it with the pattern of the baseband signal. The carrier is a tone of constant frequency (that is, a sine wave) that falls safely within the voice band, thus behaving in a manner similar to a voice transmission. The device for generating the carrier and modulating it with the digital signal and then reversing the process after reception, is called a modem (a contraction of modulator/demodulator). Three forms of modulation using a sine wave carrier are generally used, each modifying a different characteristic of the wave in order to convey information.

In amplitude modulation (AM), the amplitude of a carrier wave varies in accordance with the pattern of the digital signal. In its simplest form, the carrier is switched on and off; binary 1 causing the carrier to be transmitted, binary 0 causing it to stop. A slightly more complex version is for the carrier to have two amplitude levels, one for binary 1 and the other for binary 0. The carrier switches from one amplitude level to another depending on the information in the baseband signal. Amplitude modulation has been used for data transmission since the first telegraph carrier systems were developed in the 1920s. But is rarely used today because it is sensitive to noise interference causing amplitude changes that lead to transmission errors.

Far more common is frequency modulation (FM), in which the carrier is switched between two frequencies as the baseband signal shifts between 1's and 0's. This system is also known as frequency shift keying. Because the amplitude of the carrier does not change, and thus carries no information that might be distorted by accidental amplitude changes, FM is not noise-sensitive. Developed in the 1930s and introduced into telegraphy in the 1940s, it is commonly used in low-speed (i.e. less than 2400 b/s), asynchronous (stop-start) transmission. An additional advantage is that the modem is relatively inexpensive. For years it has been the most common data transmission system. It is, for example, well-suited for communication between personal computers.

The third major form of carrier system is phase modulation, in which the phase of the carrier wave, that is, the point at which the wave begins its cycle, is altered with each change in the baseband signal. Phase modulation is the most complex and expensive form of modulation, but because of its reliability and its ability to carry a "clocking" signal, it is used in all high-speed, synchronous transmission. It use will likely increase as demand for such services increases.[8]

The first large and sophisticated data communications system of the 1950s was designed for the U.S. military by the Lincoln Laboratory at the Massachusetts Institute of Technology (MIT). Begun in 1953 and placed in operation in June 1958, the Semiautomatic Ground Environment (SAGE) was an air defense system interconnecting hundreds of radar stations to regional data-processing centres. Radar information at each station in a district was converted to a digital bit stream, modulated, and transmitted to the data centre, where it was automatically combined, calculated, and plotted on maps. If a surprise bomber attack were detected, SAGE also provided for ground-to-air voice communications for directing jet interceptors. Of particular interest, SAGE data was transmitted over simple telephone circuits leased from AT&T. Lincoln Laboratory and AT&T cooperated in designing a modem and "conditioning" the needed lines to reduce the error rate caused by impulse noise and the electrical characteristics of the line. Knowledge gained in this project laid the basis for AT&T's civilian data transmission services as this demand arose.[9]

Though perhaps not appreciative of the scale this industry would eventually assume, existing

7. James Martin, *Telecommunications and the Computer* (Englewood Cliffs, N.J.: Prentice-Hall, 1976), pp. 135, 142–6.

8. Andrew S. Tanenbaum, *Computer Networks* (Englewood Cliffs, N.J.: Prentice-Hall, 1981), p.100; Ken Sherman, *Data Communications: A User's Guide*, 2nd Edition (Reston, VA: Reston, 1985), pp. 178–181; Martin, *Telecommunications and the Computer*, pp. 205–28.

9. J.Z. Millar, "Data-Transmission Equipment," *Communication System Engineering Handbook*, Donald H. Hamsher, ed. (New York: McGraw-Hill, 1967), pp. 12-1–12-3; Alex Curran & Dave Hamilton, "The Evolution of Data Communications – A Canadian Experience," *Telesis*, 4,1 (Spring 1975), p. 9; M.D. Fagen (ed.), *A History of Engineering and Science in the Bell System: National Service in War and Peace (1925–1975)* (Murray Hill, N.J.: Bell Labs, 1978), pp. 546, 573–580; George E. Valley, Jr., "How the SAGE Development Began," *Annals of the History of Computing*, 7,3 (July 1985), pp. 200, 205, 219.

common carriers were aware of the business potential in data communications. In 1962 the president of Bell Canada told company shareholders of the opportunity that data communications offered for diversification, and stated that "our aim is to carry over the telephone network virtually every kind of information that can be translated into suitable electrical signals." The existing teletype networks, TWX, operated by telephone companies, and TELEX, run by telegraph companies, already constituted simple public data communication links. But these were very low speed systems, ranging from 45 to 75 b/s, or about 60 to 100 words per minute. Higher speeds were possible in the voice band of telephone circuits using electronic modems, which several companies began producing for commercial use in the late 1950s and early 1960s. The earliest modems for a **public** data transmission service were introduced for AT&T's Dataphone, which began offering service on a limited basis in 1958. Dataphone was introduced by Bell Canada in 1961. By the 1970s, modems offered data rates between 600 and 4800 b/s, depending on the modulation system and the design of the modem. By the 1980s, a maximum rate of 9600 b/s was standard, though this required using multilevel coding and, hence, more sophisticated modems.

Much higher bit rates requiring more than voice bandwidth were possible over leased, or in some cases public and switched, channels provided by the telephone and telegraph companies. These broadband channels were generally groups of analogue circuits designed to carry numerous frequency division multiplexed telephone calls or teletype messages. Instead, they could be used to transmit one or a few high-speed data communications requiring wider bandwidth than that offered by a single voice circuit. A variety of transmission media, most notably coaxial cable, microwave, and satellites were used. Bit rates possible with wideband facilities were as high as 230,400 b/s.[10]

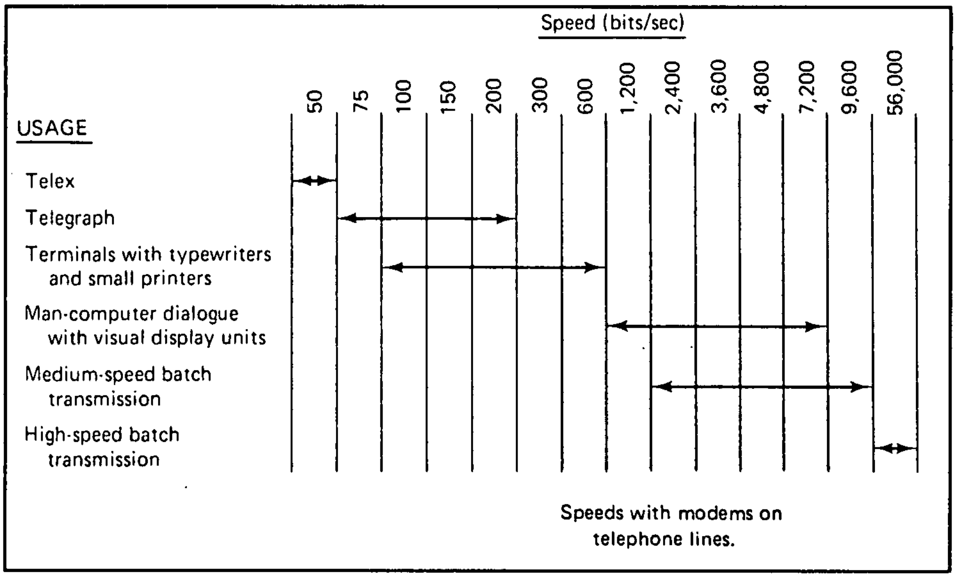
10. Millar, p. 12-6–12-10; Martin, *Telecommunications and the Computer*, pp. 244–61, 344–5; William R. Bennett & James R. Davey, *Data Transmission* (New York: McGraw-Hill, 1965), pp. 247–50; *Communications —Understanding Computers* (Alexandria, VA: Time/Life, 1986), pp. 18–20; Bell Canada, *Annual Report* (1961), p. 12; Bell Canada, *Annual Report* (1962), pp. 2, 9.

Time Sharing Systems in the 1960s

The 1960s were marked by developments both in the way that computer power was allocated and in the manner that communication was managed. The major innovation was time sharing, which solved the problems that had arisen with batch processing as demands on computers increased. Before time sharing, the only means of gaining regular, reliable access to data processing services for anything but bulk data transfer was to locate near one of the still large and expensive computers. This limited computer use to individuals working in the same building or neighbourhood as the data centres of a few large corporations and universities. Even in these cases, users had to deliver punched cards to data centres and wait hours or days for the processing of a problem or program that might only run a few seconds. Furthermore, minor programming or keypunch errors could not be detected until after the program was run and could only be rectified by repunching, resubmitting, and waiting.

Time sharing changed this situation by removing the bottleneck at the access point to the computer's processing functions. The problem had never been in the actual processing speed; the computer could perform most tasks in a few seconds or less. Far more time consuming was the business of interacting with a "slow" human user who would input data or commands sporadically, every few seconds or minutes. The use of pre-punched cards had been a means of efficiently using the computer's capabilities by separating the creative, and hence slow and hesitant, human thought process from the computer's rapid, but essentially mindless, execution of routines. But most potential users naturally preferred to communicate with the computer in "real time," getting immediate information from the computer on the progress or outcome of their task as they worked on it.

In a time sharing system, users were connected to the computer via their own keyboard-equipped terminals. After "logging in" to the computer and giving their passwords, numerous users could virtually simultaneously work with it in real time, though in fact the computer at any one moment was handling only one user's task. A "supervisory" program managed the allocation of the computer from one user to another, leaving the bulk of the computer's capacity to the actual processing of data. By switching very rapidly from one user to the next, processing all or a portion of the task



Commonly used data transmission speeds over the public telecommunications network. (Martin, Telecommunications and the Computer *[1976], p. 58)* before moving on, the computer gave the user the impression of conversing with him or her exclusively. Most time sharing systems included a variety of other software that could be made available to users for word processing, accounting, inventory control, or more specialized technical or scientific procedures. Other software allowed two users to communicate through the computer or to leave messages.

First proposed in 1959, the first model of a time sharing system was demonstrated at MIT on an IBM 709 computer in 1961. In the next two years several more systems went into operation at MIT and a number of other research centres. Some early installations in Canada were at Northern Electric's labs in Ottawa (now Bell-Northern Research), at Sir George Williams University in Montreal (now Concordia University), and at the University of Alberta in Edmonton. Initially designed for users affiliated to the institution owning the "host" computer, time sharing was soon exploited for its commercial potential. In 1964 Dartmouth College, which had hosted Henry Stibitz's long distance computer demonstration in 1940, began selling computer time on its system to other universities. Soon, computer service companies began offering time sharing to businesses interested in computerized accounting, payroll, ordering, and inventory control. Too small to afford their own computers, these companies could rent time on a computer and gain access to it via a teletypewriter, a modem, and the switched telephone network. The large computer manufacturers followed with their time sharing computers and programs. By 1968, General Electric had built a national network in the United States serving 50,000 customers from 31 regional time-sharing centres.

In addition to such general purpose networks, special purpose networks were also instituted, the most notable of which was SABRE (Semi-Automated Business Research Environment), developed by IBM for airline reservation systems. When introduced in about 1964, it was the largest commercial real-time computer network in the world, serving 2000 terminals across the United States connected by leased high-capacity telephone circuits to a host computer in Briarcliff, New York. With a response time of less than three seconds, it allowed an airline agent at any of the terminals to inquire about space on a flight and make or cancel reservations.[11]

The efficiency of these time sharing networks was enhanced in the late 1960s and early 1970s by the introduction of digital transmission in the public telecommunications networks. Digital transmission offered lower costs and lower error rates because of less noise distortion. It also enabled the use of higher and more flexible transmission rates suited to many different terminal speeds. Until then, data travelling over the public telephone network passed either at TELEX or TWX rates or at rates possible with a modem over a single voice circuit. Private leased lines offered a limited choice of higher rates derived from the multiplex groups used in long distance analogue voice transmission.

In Canada, the main data communications market was in serving time sharing systems, which used low-speed TELEX-type transmission. Most of the clients for these services were small-businesses and, partly because of the high cost of trans-Canada transmission, many of them subscribed to American time-sharing services. In response to demands to lower its data transmission rates, the Trans-Canada Telephone System (now Telecom Canada) instituted the world's first nationwide commercially available digital data communications system in 1973. "Dataroute" accommodated a range of speeds from less than 110 b/s to 50,000 b/s (50kb/s). It was essentially a private line service, though access was possible through the public telephone network. Long-haul transmission was over high-speed digital radio links on the existing trans-Canada microwave network. Customer signals travelled at low or medium speeds over separate telephone lines to a "node" in a central office where they were interleaved in a single bit stream (time division multiplexing) and transmitted at the higher rate over the radio channel. The national system was kept in synchronization by a master atomic clock in Toronto, which controlled two back-up clocks and individual tertiary and terminal clocks in every node of the system. Dataroute brought cost savings to users of

11. R.M. Fano & F.J. Corbato, "Time-Sharing on Computers," *Scientific American* (September 1966), pp. 129–132; *Communications (Understanding Computers)*, (Alexandria, VA: Time-Life, 1986), pp. 39–43; G.N. Boyd, "An Inter-City, Multi-Access, Time Sharing Computer System," *Telesis*, vol. 1 (September 1968), pp. 93–7; Stinson, *Wired City*, p. 158; Curran & Hamilton, p. 9.

over 50 per cent for high-speed circuits and a dramatic 90 per cent for low-speed lines.[12]

Packet Switching and Value Added Networks

Time sharing provided for more efficient use of computing resources. But the proliferation of time sharing systems created its own problem when demand began to arise for interconnecting the computers in the various independent systems. In addition, the gap between the cost of computing power and the cost of transmission circuits narrowed in the 1960s with the development of integrated circuits. The number of data centres in the General Electric time sharing system, for example, reflected an emerging fact of life in computers and communications: the cost of the long distance communications circuits was so burdensome that it was actually more attractive to purchase and install numerous regional computers rather than lease long distance lines to connect all users to one centre. While the timely introduction of digital data transmission lowered the cost and improved the performance of long distance circuits, data communications users still had an incentive to consider alternative methods of building a data communications network.

One of the main problems with transmission over the public telecommunications network was the inefficient way that circuits were allocated to users. Telephone companies followed standard practice in allocating a single circuit, be it a wire pair, a frequency band, or a time slot, before transfer of information began and reserving it until the transaction was complete. Transmission of data did not begin until a path had been established through the labyrinthine switched telephone network. While suited to the voice traffic for which it was designed, this system, known variously as circuit switching or bandwidth pre-allocation, entailed two inefficiencies for data users. First, almost all data transactions over the public network were shorter than the average voice call of five minutes. In fact, the several seconds it took to set up a connection might be longer than the duration of the data transmission for, say, a credit card check. This meant that a disproportionate amount of time and circuit capacity was consumed by signalling rather than carrying data. The second major problem with circuit switching was that data transmissions were not like a continuous telephone conversation; they consisted of short bursts of information followed by pauses as the operator awaited a response from the computer or formulated his or her next thoughts. But because the circuit was established for the duration of the transmission, the circuit capacity during these pauses was wasted, being unavailable to another user. The cost of the unused capacity was borne by the user and the cost per user for a transmission was therefore greater. In addition, the occupation of circuits not fully used increased the possibility of blockages in the system because all circuits to a given destination could easily be engaged.

One alternative to circuit switching was message switching, which originated in telegraph practice. It involved the sending of a completed message from the source to a switching office, where it was stored and then forwarded when a circuit became available. A message might pass through several intermediate switching offices, each one storing it, before reaching its destination. Message switching allowed transmission circuits to be used to their maximum capacity because circuits could be loaded for continual use. Through the 1950s and 1960s, message switching evolved from using the simple "torn tape" technology of telegraphy to the use of magnetic disks and tape for storage. But the rise of real time computer systems highlighted the delay factor in message switching, which was rarely less than a few minutes and at busy times could run to several hours. While suited to transactions where time was not essential, message switching was unsuitable for users wishing to "converse" with a computer.

The answer to the limitations of circuit and message switching was packet switching, a computerized, speeded-up version of message switching with an important modification. Packet switched networks used computerized switching nodes and high-speed transmission to facilitate transmission in "real time." The modification that distinguished it from message switching was that long messages were broken up into smaller "packets." Each packet was addressed and sent as if it were a separate message, passing along one or a number of routes and being reassembled at the destination. Although a user was only assigned a circuit for the moment a packet was being sent, the high speed of the system gave the impression of a dedicated circuit. The central switching elements in a packet network were the nodes, minicomputers that received, acknowledged, checked for errors, stored,

12. Dave Horton, "New Digital Network Speeds Canadian Data Streams," *Telesis*, 3,3 (Fall 1973), pp. 66-9; Curran & Hamilton, pp. 10-3.

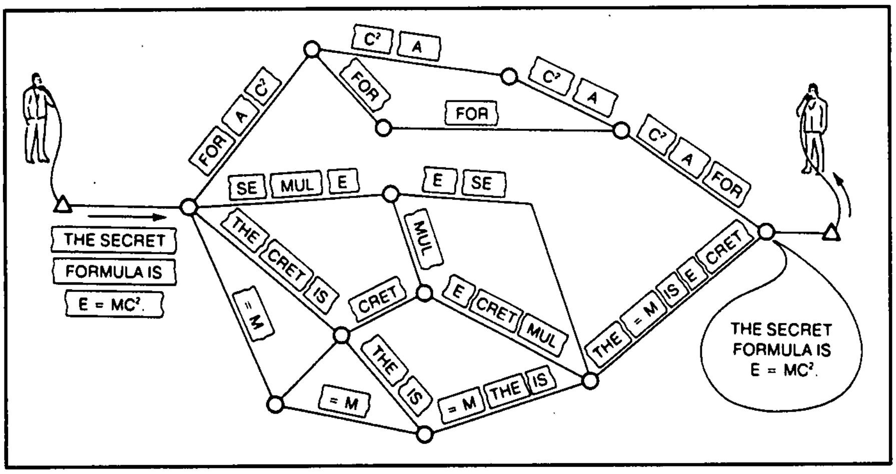
and forwarded packets along their way. A packet would pass through any number of nodes on its route. In the "datagram" variation on packet switching, each packet in a message could even be routed by a different path, depending on the availability of circuits when the packet was ready to be sent. On the other hand, the "virtual circuit" version sent packets along the same route. While making it unnecessary for nodes to make routing decisions for each packet, a virtual circuit required a delay for setting up the route before data transmission could begin.

Public packet switched data networks usually offered a number of "value-added" features that increased their attractiveness to users. Telecommunication carriers traditionally simply established circuits for users and, by reducing noise and circuit distortions, attempted to deliver messages with a minimum of error. The client, not the carrier, was responsible for error detection and for ensuring that the terminal devices were compatible for communicating. The most important value-added service offered by many packet networks was the enabling of communication between otherwise incompatible devices and systems. This involved the translation of the data codes used by various computer and terminal manufacturers, the conversion of protocols governing the interaction for devices within a system, and the provision of buffers (memories) to facilitate the matching of data rates between, say a high-speed computer and a low-speed terminal. Value-added networks also provided error detection and correction. Another important feature was a packet network's ability to automatically find alternate routes if its first options were blocked or had failed. Finally, value added networks provided a user with network control and management, including monitoring, accounting, and statistical analysis of the user's traffic.

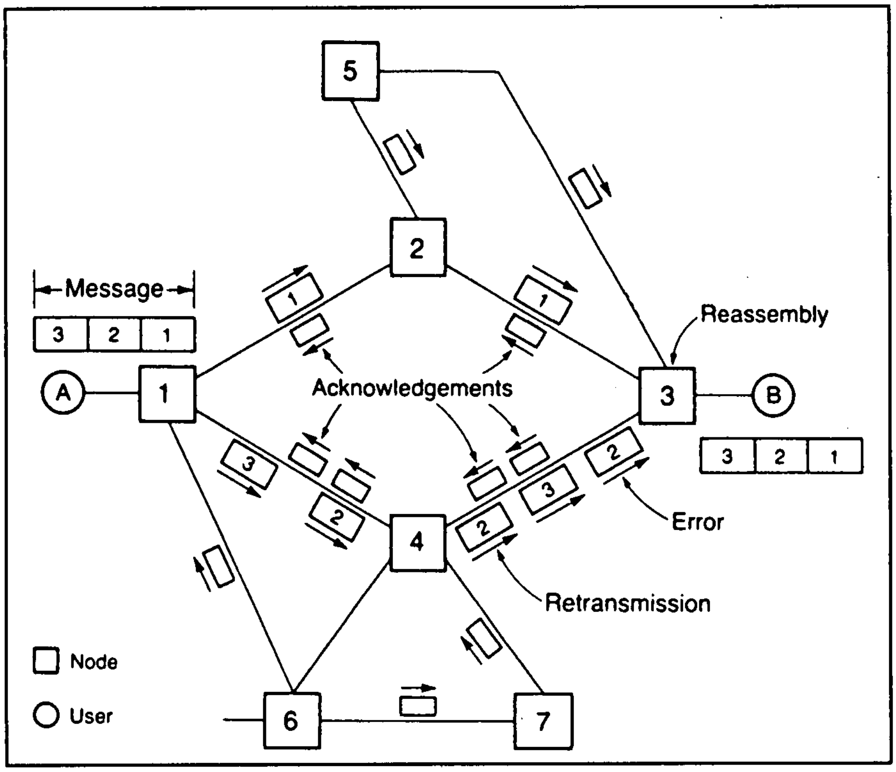
Packet switching was independently conceived in the United States and the United Kingdom in the mid-1960s. Indicative of the problems it was intended to solve, it was developed not by the telecommunications industry but by the computer industry. In 1964, Paul Baran of the Rand Corporation proposed packet switching for the U.S. Air Force, which had requested a decentralized communication system able to survive damage to a portion of the network. A year later, Donald Davies of the National Physical Laboratory (NPL) in the United Kingdom proposed a packet switching system. Davies's system, conceived before he became aware of Baran's proposal, was the first to use the term "packet" to describe the blocks of data switched through the network. Though essentially similar to the actual networks later built, Davies was only able to implement the system as a local area network, built for the NPL in 1970 and modified over the following few years. The first large-scale packet network was constructed in the United States for the Advanced Research Projects Agency (ARPA), an agency of the U.S. Defence Department that wished to link the various research labs and university labs working on defence projects into a single network. ARPANET, as the system was called, began operations with four nodes in December 1969, and by April 1971 was supporting 23 host computers across the country. ARPANET continued to grow rapidly over the following years, though its use was restricted to institutions performing military research. Packet switching was not implemented on a larger scale until increasing demand for its type of service sparked the interest of computer users and telecommunications companies.

According to one expert, the year 1969 marked a turning point in the success of packet switching. That year, the cost of computer power fell below the cost of transmission lines for a data communication system. From this point on, dynamic allocation systems like packet switching became cheaper than pre-allocation systems like circuit switching.

The first packet switched networks were private systems operated by groups of research labs and universities or by corporations for their own uses. A flurry of announcements by public telecommunications carriers in 1973 and 1974 reflected a sudden awareness of its market potential and marked the beginning of a period of major packet switch network construction. The first public packet network was introduced in the United States in August 1975 by TELENET, a communication company expressly created to exploit packet switching. In early 1977, the TransCanada Telephone System brought a national packet switched network called DATAPAC into service. DATAPAC initially operated with nodes in Calgary, Toronto, Ottawa, and Montreal, with nodes in other cities being installed as traffic increased. The state



A model packet switched network for secure voice communications. An analagous system is used for data messages of all kinds. (Rosner, Packet Switching *[1982], p. 31)*



Basic operation of a packet switched data network. User A is transmitting a three packet message (3-2-1) to User B. After receiving a packet, each node along the route sends back an acknowledgement. If a node detects a transmission error it instructs the previous node to retransmit that packet. (Rosner, Packet Switching *[1982], p. 77)* telephone monopolies in Britain, France, and Japan also completed national packet networks in the late 1970s.[13]

Protocols and Networks

Just as in human communication, interconnection of computers requires agreements on common languages and rules of interaction. In addition to ensuring basic physical and electrical compatibility of terminals, provision must be made for code translation and for rules, or protocols, for addressing one another, initiating exchange of information, acknowledging receipt, correcting errors, and ending the exchange. The task of interconnecting computers has been greatly complicated by the different standards used by the various computer manufacturers. In the 1960s, two basic steps were taken in North America to facilitate interaction between devices produced by different companies. In 1964, the American Standards Association (ASA) agreed to a universal code for the representation of data for transmission. The American Standard Code for Information Interchange (ASCII) was a 128-character, seven-bit code assigning the correct number and pattern of bits for representing letters, numbers, punctuation marks, such control commands as DELETE and BACKSPACE, and signals marking the beginning and end of a message. The second important standard was RS-232-C, established by the Electronics Industries Association in 1969. This created a common method for interconnecting computers and modems, dictating voltage levels and the number of wires in the connecting cable, and establishing protocols for deciding the mode and speed of each transmission. In the 1970s, standards for digital transmission rates over the public telephone network were also set.

Packet network development was hastened by the adoption in March 1976 by the International Telegraph and Telephone Consultative Committee (CCITT) of an international protocol known as X.25. This dictated standard physical and electrical connections between host computers and the network, standard error detection codes, and uniform procedures for packet formatting and routing. X.25 allowed easy interconnection of the packet networks in various countries. It was adopted largely because of pressure by Canadian carriers, who feared that standards would be set by computer giants like IBM, who were encroaching on communications services. Most operators and builders of public data networks, including the TransCanada Telephone System (TCTS) with DATAPAC, subsequently adopted the X.25 standard. In early 1978 DATAPAC and TELENET in the United States were interconnected, an event that fulfilled the intent of those who pioneered packet switching as a network builder.

As networks expanded in both their extent and their complexity, so too did the scale and complexity of the software and hardware needed to allow communication between incompatible computers and devices. Many such components grew haphazardly to meet new demands as they emerged, and systems as a whole became unwieldy, with changes to one portion of the system often having unexpected results elsewhere. The task of devising software and hardware for each new type of interconnection was, already by the 1970s, proving costly and complicated. The result of this situation was the adoption of standard "architectures" for computer networks that established the various components of a network, their forms, and their modes of operation. The two most important standard architectures were the Systems Network Architecture (SNA), promulgated by IBM in 1973, and the Open Systems Interconnection model (OSI), agreed to by the International Standards Organization (ISO) in 1977.

Both these conventions divided communication functions into several hierarchical layers. Each layer performed a specific function needed to communicate with the other system, for example code translation or packet formation, then passed the data to the next layer (to give some indication of the scope of such architectures, CCITT's X.25 conformed to just the bottom three layers of the seven-layer OSI model). Each layer in one machine appeared to communicate directly with the corresponding layer (its "peer") in another, according to certain protocols. But in fact, data from one user's terminal travelled down through one layer after another to the physical transmission layer, where

13. Joseph V. St. Amand, *A Guide to Packet-Switched, Value-Added Networks* (New York: Macmillan, 1986), pp. 1–24; Roy D. Rosner, *Packet Switching: Tomorrow's Communications Today* (Belmont, Cal.: Wadsworth, 1982), pp. 3–39; *Communications: Understanding Computers*, pp. 43–57, 65–77; Lawrence G. Roberts, "The Evolution of Packet Switching," in *Satellites, Packets, and Distributed Telecommunications*, Roy D. Rosner, ed. (Belmont, Cal.: Wadsworth, 1984), pp. 110–114; Martin Campbell-Kelly, "Data Communications at the National Physical Laboratory (1965–1975)," *Annals of the History of Computing* 9,3/4 (1988), 224–39; Bill Clipsham & Max Narraway, "Datapac: a Public, Shared Data Network for Canada," *Telesis*, 4,5 (April 1976), pp. 130–5; Curran & Hamilton, p. 14.

it was carried to the receiving end and passed up through each corresponding layer to the receiving terminal. The work performed by each layer at the transmitting end allowed the data to be received, understood, and processed by the corresponding layer at the other end of the transmission. Standard architectures did not stipulate how each layer performed its functions, only that the data be passed to the next layer in a certain condition and that all functions be performed in a particular order. In fact, the internal operation of each layer was invisible to the layers above and below. Through coded "headers" following prescribed protocols and attached to outgoing data, each layer communicated with its peer, as, for example, the user appeared to communicate directly with his or her counterpart.

Layered architectures reduced the complexity of communications software and hardware by breaking up systems into more comprehensible components. Because the objectives of each layer were standard while the means for achieving them were flexible, changes could be made in one layer without affecting others. Design and development of network components could, in fact, be undertaken by different teams of programmers and engineers. In addition, individual standards could be established for each layer as the need or the consensus arose.[14]

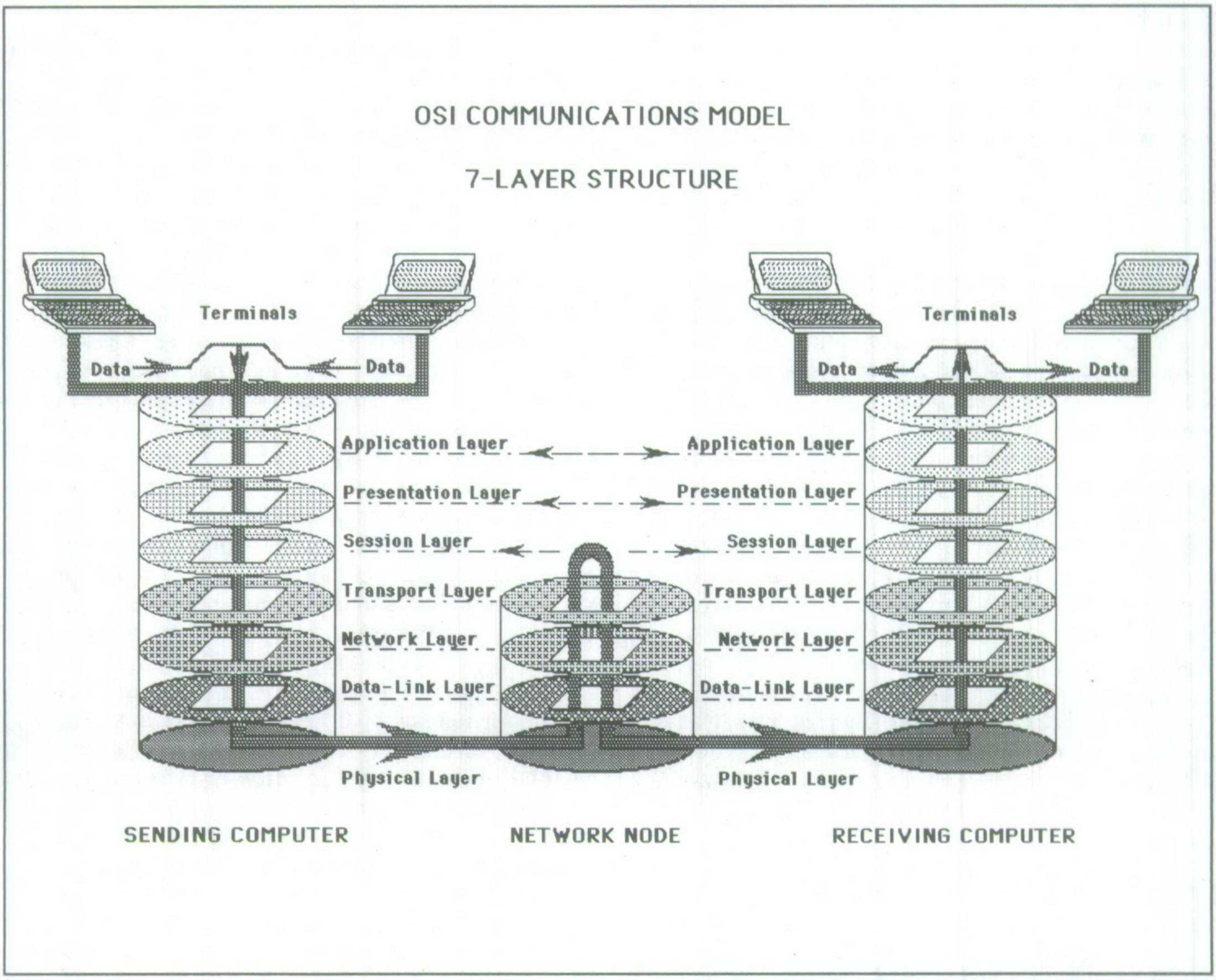
Local Area Networks (LANs)

It has been estimated that more than 60 per cent of information generated in an office is destined for use on the same site. Yet the internal dissemination of this information had, up to the 1980s, relied on decades- and century-old technology: most specifically the telephone, the typewriter and the photocopier. As part of the broader economic restructuring taking place in the 1980s, employers attempted to cut costs in the traditionally labour-intensive field of office work by increasing output per employee. Just as manufacturers had turned to automation as one strategy to increase output per worker, so office employers introduced new information technology (made cheap by the evolution of microprocessors) to reduce the labour cost of managing the suffocating volumes of information and paperwork generated by modern corporations and government bureaucracies. On the one hand, they sought to reduce the time that support staff spent on such tasks as typing and retyping documents. At the same time, they attempted to reduce managers' and professionals' "wait time and waste time" between various stages of a job by speeding up internal communication and access to information with such services as electronic mail and shared databases. Just as the 1970s were distinguished by the construction of national-scale data networks, therefore, the 1980s saw the linking of personal computers and other computerized office equipment in local area networks.

A local area network (LAN) is privately owned and generally links computers and related devices via a common transmission medium over a limited area like an office building, a factory, or a university campus. It is ideally a relatively inexpensive, high speed link, giving a transmission rate in the neighbourhood of 1-100 million b/s (Mb/s) with a low error rate. Local area networks range from relatively primitive systems linking a few personal computers to larger office automation systems and more high speed networks serving computer data centres and research institutes. A more specialized kind of LAN is a factory network linking various small computers and automated production machines in order to coordinate the manufacturing process.

The advantages of local area networks are obvious. They not only perform the traditional role of linking remote terminals to central processors and databases, they allow the interchange of data and messages between desk-top computers. In addition, they allow the sharing of expensive resources like processing capacity, software, memory discs, and printers. If properly designed, a decentralized local network can also be more reliable and survivable than one relying on a central processor because functions are dispersed and backup is possible in the event of a local failure. Finally, a properly designed local network allows changes to be made incrementally to single components, thus limiting disruption to other parts of the network. With adequate software, hopefully provided by industry protocols, equipment from various suppliers can be connected as desired. Similarly, a network allows flexibility in locating equipment by

14. *Communications: Understanding Computers*, pp. 49-56, 105-21; Lawrence G. Roberts, "The Evolution of Packet Switching," pp. 113-4; Clipsham & Narraway, p. 135; "A New Standard for Data Networks," *Telesis*, 4,6 (June 1976), p. 191. A clear discussion of multilayered architectures for non-experts may be found in: Andrew S. Tanenbaum, *Computer Networks* (Englewood Cliffs, N.J.: Prentice-Hall, 1981), pp. 10-21. See also: Uyless D. Black, *Data Communications and Distributed Networks*, 2nd ed. (Englewood Cliffs, N.J.: Prentice-Hall, 1987), pp. 252–6.



The OSI seven-layer model architecture for a computer network. Like a traditional bureaucracy where messages must pass up and down through layers of directors, secretaries, translators, mail clerks and couriers, data communication is arranged in a hierarchy of functions. Information travels down through each layer from the originating "director" and up each layer to the receiving "director." Despite only communicating through intermediaries, each layer in the hierarchy <u>*appears*</u> *to communicate directly with its counterpart in the other office. We may think of the intermediate layers of the "network node" as a postal station that sorts and routes messages between offices. (Based on Time/Life Books,* Communications: Understanding Computers *[1986] pp. 106–7)* providing a standard interface with a shared transmission medium. Finally, some LANs, most notably high capacity, or broadband, networks, can provide voice and video channels in an integrated system.

The disadvantages of LANs are similar to any other data network. Their usefulness is limited by the lack of protocols and other standards among the diverse suppliers of LANS and equipment. Often, for example, the ability to exchange data does not ensure that the devices of different vendors can be used cooperatively. Each computer might use different file formats and control characters, making full interoperability (to allow, say, the editing of one file on another machine) impossible without first converting formats. There is also the problem of cost control: decentralized control of data equipment leads individual units within organizations to acquire equipment for their own needs that cumulatively exceed the needs of the entire organization. But most importantly, the main advantage of distributed systems, easy access to information, is also their main disadvantage, the loss of control over information. If any one of a number of sources can make changes to a single data base, multiple data sources might weaken integrity of data bases as well as endanger their security and privacy.

In the 1970s and 1980s, a wide variety of LAN designs were developed and sold, reflecting the lack of industry-wide standards among the various computer manufacturers who each hoped to sell proprietary systems compatible only with their devices. The number of possible system types was augmented by competition from traditional suppliers of office telephone networks: the manufacturers of private branch exchanges (PBXs). These companies joined the computer vendors in offering data communication packages to business users. This section will only describe those systems devised by the former. Because they at least imply the integration of voice and data, digital PBXs will be presented in Part 4.

The LANS designed by computer manufacturers tend to transmit data as packets over a shared transmission medium. This medium can be laid out to link stations in a number of different forms, or topologies. As there is usually no central processor or node supervising the system, a variety of methods are used to control access to the medium. In response to the bewildering proliferation of LAN designs, the Institute of Electrical and Electronic Engineers (IEEE) established three standards in the 1980s that embraced the major methods on the market. This number was likely to increase in view of the continuing fragmentation of the market.

The first IEEE standard, based on the "Ethernet" LAN devised by the Xerox Corporation in 1975 (but based on an earlier data network, ALOHA, developed at the University of Hawaii), consists of a single linear coaxial cable, a "bus," to which all stations are attached. Any packet transmitted on the bus, which allows a data rate of 10 million b/s, is received by all stations in "broadcast" manner but is only copied by the station to which it is addressed. Each station "listens" to the network and only transmits when it is idle. If two stations accidentally transmit simultaneously, they detect a "collision" of packets, "back off" for a random period, then again attempt transmission. A variation on the bus format is the "tree," which is a bus split into a number of branches.

The "token ring" describes both the form and the access method of the second IEEE standard, which has been promoted by IBM since it introduced its version in 1985. In a ring network, each station is attached to a looped transmission medium. At each station's access point is a repeater that receives the packet circulating on the loop and relays it to the next station. Stations are equipped to recognize and copy packets addressed to them. Access to the common medium is controlled by a "token" packet that continually circulates around the ring when it is not in use. To transmit, a station receives the token as it passes through and replaces it with its data packet addressed to another station. This packet is transmitted around the ring to the receiving station, which copies the packet and allows it to return in full circle to the sender. On receipt of its packet, the sending station replaces the token on the ring and thus makes it available to other stations wishing to transmit. The data rate for a coaxial cable token ring is 4 million b/s, though by using fibre optics a token ring can be designed to transmit at 100 million b/s.

The third LAN type, the "token bus," is laid out in a bus format but uses a token for access control. Because stations do not receive the token in a sequence determined by their position in a ring, each must be assigned a place in an arbitrary sequence that dictates when a station may receive the token. Stations that require more frequent access to the bus may be assigned more than one place in the sequence, ensuring access within a guaranteed time interval. The token bus has found a specialized use in factory automation systems, both because of its simple layout and because of its ability to guarantee data delivery times. The latter is important for manufacturing process control signals between automated elements in a single production sequence.[15]

15. John Williamson, "Office Automation: PABXs and LANs," *Telephony* (24 September 1984), pp. 51–2; Maris Graube, "Local Area Nets: A Pair of Standards," *IEEE Spectrum* (June 1982), pp. 60–64; W. Scott Currie, *LANs Explained: A Guide to Local Area Networks* (Chichester, U.K.: Ellis Horwood, 1988), pp. 13–17, 36–46, 54–62, 72–3, 84–5, 97–8; William Stallings, *Local Networks: An Introduction* (New York: Macmillan, 1984), pp. 1–12, 53–76, 88–90, 111–130.

3 Converting the Telephone Network to Digital

Digital Transmission—Pulse Code Modulation (PCM)

At the same time that they were introducing the first data transmission services in the early 1960s, telephone companies also began a major change in the way they treated voice services. The era of digital telephony was beginning. The first stage in the transition was the introduction of digital transmission circuits on heavily-used urban trunk cables. Digital transmission involved converting the continuous electrical wave of an analogue voice signal into a series of discrete, digital pulses that in turn could be interleaved with those of other signals in a process called time division multiplexing. The telephone companies were attracted to digital transmission by a number of advantages it possessed over traditional analogue methods.

Digital transmission offers a more rugged and retrievable signal than analogue, which effectively increases the capacity of a digital channel over that of an analogue channel using the same medium. Noise in analogue transmission is amplified with the signal as it passes through each amplifier on a circuit, thus becoming cumulative. But with digital transmission, each repeater station recognizes the remnant pulse amid the noise and, rather than amplify the entire signal, simply regenerates a clean pulse. This allows a digital signal to be transmitted through a much more noisy medium without distortion, loss, or error. While digital voice requires a much higher bandwidth than analogue, frequent repeater spacings allow for a lower signal-to-noise ratio and hence a higher bit rate. Each wire pair in an underground cable, for example, can carry a maximum of 12 analogue channels. But with digital transmission using repeaters every 6000 feet, 32 channels can be carried; even more channels, that is a higher bit rate, are possible with a closer repeater spacing.[16]

16. Martin, *Telecommunications and the Computer*, pp. 264–5.

One of the other attractive features of digital transmission is the ease of multiplexing the signals using time division methods. It is becoming economical, through the use of microwave radio, satellites and fibre optics, to build channels of higher bandwidth. But frequency division methods of exploiting their capacity create high multiplexing and switching costs. Frequency division multiplexing (FDM) first involves modulation, impressing the electrical pattern of a voice signal onto a high frequency "carrier" wave. Several other voice signals are modulated at slightly higher or lower frequencies, and then all are transmitted together. At each switching point, this complex signal carrying calls to different destinations must be demultiplexed, demodulated, switched, then modulated and multiplexed again. Each stage uses expensive analogue circuit components: oscillators, modulators, amplifiers and filters. Multiplexing and switching costs, therefore, become a greater proportion of network cost as channel capacity is increased. The cost of digital multiplexing circuitry, on the other hand, is falling rapidly with the use of large-scale integrated circuits. Switching can be integrated into the transmission process using time division methods.[17]

Full digital transmission also allows data to be transmitted over voice lines without modems, lowering terminal equipment costs. And pulse code modulation (PCM) also allows voice, data, fax, and video to be transmitted as a single bit stream. They will not interfere with one another and will not require different engineering of different specialized channels. This eventually might result in an integrated network.[18]

17. Martin, *Telecommunications and the Computer*, pp. 266–7.
18. Martin, *Telecommunications and the Computer*, pp. 267; D.A. George & S.T. Nichols, "Telecommunications Technology," *Telecommunications for Canada: An Interface of Business and Government*, H. Edward English, ed. (Toronto: Methuen, 1973), pp. 274, 289–90; Prescott C. Mabon, *Mission Communications: The Story of Bell Laboratories* (Murray Hill, N.J.: Bell Labs, 1975), p. 42.

The first step in all digital, pulse modulation systems is the sampling by an electronic switch, or gate, of the analogue signal at periodic intervals. This produces a series of pulses whose amplitudes correspond to the amplitude of the wave at the instant of sampling. A sampling rate of at least twice the highest significant frequency in the signal will ensure the samples contain all the information of that signal and will therefore allow it to be completely reconstructed. With a voice frequency signal, a sampling rate of 8000Hz (8000 samples per second) is used—about one sample every 125 microseconds.[19]

The simplest pulse modulation system is pulse amplitude modulation, in which each sample is transmitted as a regularly spaced pulse whose amplitude corresponds to that of the sample. In pulse duration modulation, the varying amplitudes of the samples determine the duration of pulses. And in pulse position modulation, the amplitude of each sample varies the time of occurrence of each pulse.

Pulse code modulation goes further than other pulse systems by encoding the samples as a series of digital pulses of uniform amplitude, duration, and position. This requires further processing of the signal before transmission, but makes the signal less sensitive to interference, easier to regenerate, and more compatible for transmission with other types of signals like digital data.[20]

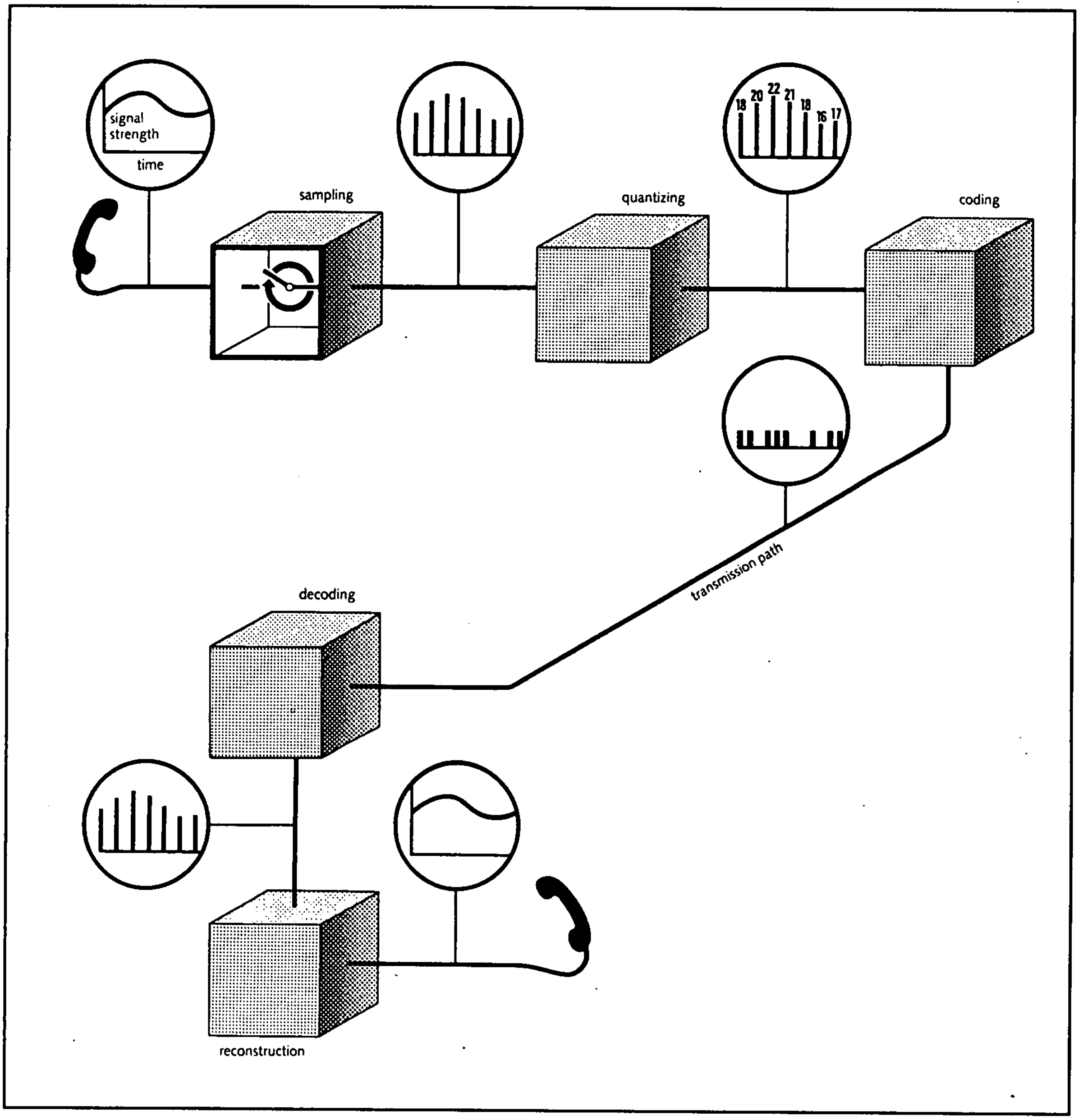
The pulses resulting from sampling can have an infinite number of amplitude values within a range between the lowest and highest amplitude in the signal. But in a process called quantizing, the PCM encoder circuit rounds off each sample to one of 128 discrete values. The human ear cannot distinguish an infinite range of amplitudes, and therefore an adequate reconstruction of the signal is possible from a discrete range of values approximating the real range. Each quantized pulse is then encoded. With a binary code, seven bits are required to represent each of the 128 possible values; an eighth bit is added for signalling and routing. Thus the data rate for one PCM channel is 64,000 bits per second (8 bits x 8000 samples per second.) PCM entails sending eight times as many pulses as other pulse systems, therefore requiring more bandwidth, but this is compensated by the simplicity of detecting and regenerating a simple series of on-off pulses.[21]

Time division multiplexing using PCM involves the rapid sampling in precisely timed sequence of several analogue voice signals. Each sample is then quantized and encoded with a seven bit binary code and the eighth signalling and routing bit is added. Each coded sample is then transmitted separately in an assigned, recurring time slot within a "frame" containing one sample from each of the other signals. At the receiving end, each incoming frame is broken up by gates operating on the same timing as the transmitter and the samples for each signal are reassembled. AT&T's T1 system, which sets the North American standard, interleaves 24 channels. It samples each of 24 channels 8000 times a second. As each sample is coded into seven bits with an eighth information bit added, a frame contains eight bits for each of the 24 samples. To this frame of 192 bits is added one synchronization bit for a total of 193. Thus the total capacity for 24 channels is 1.544 million b/s (193 bits per frame x 8000 samples per second). Other systems use higher bit rates.[22]

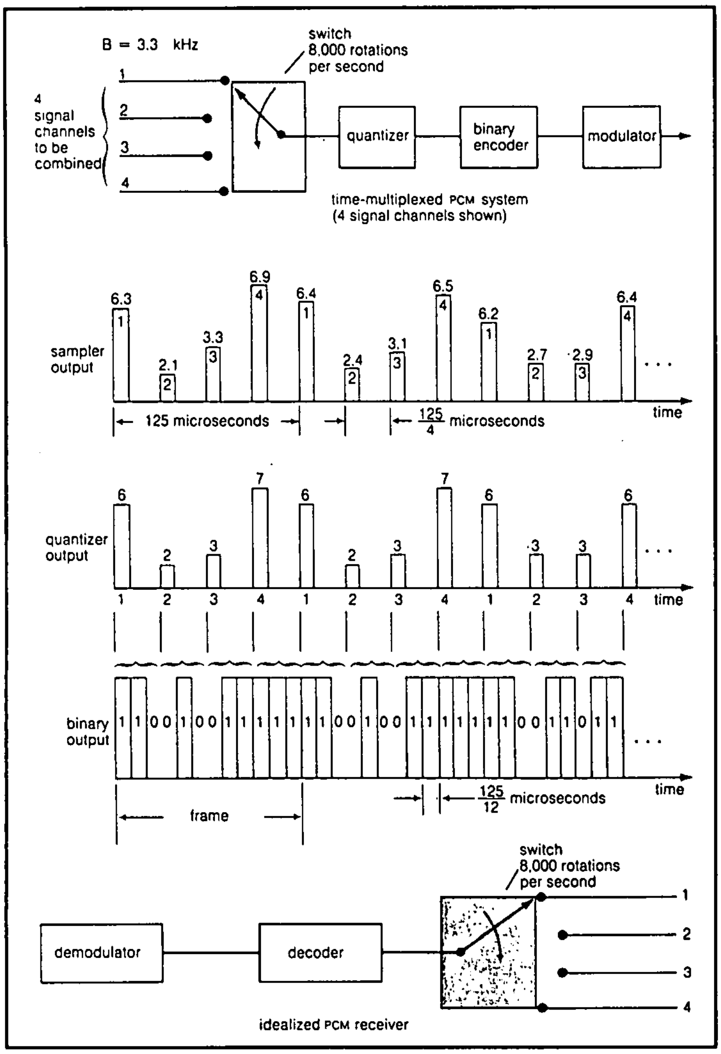
Pulse code modulation was first proposed by Alec H. Reeves of ITT's Paris Lab in 1937 as a means of making telephone signals more "rugged", i.e., less susceptible to degradation or destruction by noise. He proposed sampling the amplitude of a speech electrical waveform and representing these values by a binary code of on-off pulses. His system was not developed because at the time the necessary electronic circuits did not exist.[23]

The development of pulse code modulation for voice communications was assisted by the demand for secret global military communication during World War 2. In the United States, Bell Labs performed work on secret speech transmission systems for the National Defence Research Committee. Begun in October 1940 and known as Project X (or Sigsaly), the first systems began operation in mid-1943 in Washington, London, and North Africa. Others were later installed in Paris, Hawaii, Australia and the Philippines. Sigsaly involved a form

19. Simon Haykin, "Communication Systems, Civilian," *Encyclopedia of Telecommunications*, Robert A. Meyers, ed. (San Diego: Academic Press, 1989), p. 70; James Martin, *Telecommunications and the Computer* (Englewood Cliffs, N.J.: Prentice-Hall, 1976), pp. 267–70.
20. Haykin 70–71; David Talley, *Basic Carrier Telephony* (Rochelle Park, N.J.: Hayden, 1977), p. 161.
21. Haykin, p. 72; Talley pp. 161–3.
22. Talley, pp. 159–60; Martin, *Telecommunications and the Computer*, pp. 238–41, 271.
23. Eryl Davies, *Telecommunications: A Technology for Change* (London: Science Museum/HMSO, 1983), p. 48.



Pulse code modulation of a single telephone circuit. (Davies, Telecommunications: A Technology for Change *[1983], p. 49)*



A time-multiplexed pulse-code-modulation (PCM) system. At the top, the analogue signal for each channel is successively sampled, quantized, and then converted to a binary output. After decoding in the receiver, the reconstructed samples are separated and transferred to output channels by a switch synchronized with the transmitter sampling switch. (Encyclopaedia Britannica *[1984], vol. 18)* of sampling and quantizing, the use of a six-level code, and digital transmission by frequency shift keying (FSK) over radio circuits. According to a Bell Labs historian, Sigsaly was "the pioneering digital speech transmission system employing a form of pulse code modulation. It was one of the starting points of the digital transmission age that followed."[24]

Another military communication system, designed by the British and the Americans, was a microwave relay system using pulse modulation and time division multiplexing. The American system, AN/TRC-6, was designed by Bell Labs and used pulse position modulation and time division multiplexing to transmit eight voice signals. A similar system developed by a subsidiary of ITT was installed in a telephone microwave link between Nova Scotia and Prince Edward Island in 1948, the first of its kind in commercial operation.[25]

Bell Labs scientists became aware of Reeves's work on pulse code modulation after the issuing of his U.S. patent in 1942. As a result, Bell Labs turned its attention to binary PCM; one byproduct was Shannon's publication of information theory in 1948. In 1947 and 1948, multichannel PCM systems were demonstrated at Bell Labs. But PCM was not immediately economically feasible, both because it required more bandwidth than analogue transmission and because the precision and speed of operation that it required could not be provided at reasonable cost, given the electronic devices then available. PCM became feasible with invention of the transistor and other semi-conductor electronic devices and their development into reliable high-speed switching components in the mid-1950s. At about the same time, abundant bandwidth became available through new transmission media like microwave radio and coaxial cable.[26]

AT&T introduced T1, the first commercial application of PCM in the telephone system, in 1962. It was an inexpensive alternative to installing additional cable in large cities in order to increase traffic capacities. Still in use in 1990, T1 transmitted 24 simultaneous conversations at approximately 1.5 million b/s (1.5 Mb/s) for distances up to 50 miles using a single copper pair for each direction of transmission. Northern Electric manufactured T1 in Canada for the first time in 1964, and by the following year had installed a large number on urban cables in order to increase their capacities.[27]

In 1973 AT&T introduced T2 for intercity service up to 500 miles. T2 carried four multiplexed T1 bit streams as a single stream of 6.3 million b/s (6.3 Mb/s). A typical system had 48 T2 systems in two 50-pair cables, providing 4608 two-way channels. T2 was designed to eventually link into a national network, the gaps temporarily being bridged by analogue microwave and coaxial circuits. T2 was also designed to carry Picturephone signals, one signal per bit stream. In 1975 AT&T introduced T1C, which had twice the capacity of T1: 48 conversations at 3.152 million b/s. In the same year, it introduced T4M, a digital carrier for coaxial cable. It could carry 32,000 two-way conversations on an 18-tube coaxial. It operated at 274 million b/s, allowing 4032 voice channels to be carried. T4M was to relieve congested telephone and data routes in metropolitan areas. Its first application between Newark and New York City rerouted circuits destroyed by a fire in the Second Ave. central office. The most notable Canadian development in the same period was LD-4, a coaxial cable carrier for long haul transmission with a capacity of 274 million b/s. Its first major application was on the busy communications corridor linking Montreal, Ottawa, and Toronto.

As a result of AT&T's technical dominance of the industry, a standard North American hierarchy of basic digital carrier rates was established in the 1970s. The basic rate on a single subscriber loop was 64 Kb/s, sufficient for one voice signal. T-1 was set at 1.544 Mb/s, T-2 at 6.312 Mb/s, T-3 at 45 Mb/s, and T-4 at 274 Mb/s. Similar, though not identical, standards were set in Europe and in Japan. A variety of different transmission media and multiplexers could be used to achieve these levels, which became the standard for both voice

24. M.D. Fagen, ed., *A History of Engineering and Science in the Bell System: National Service in War and Peace (1925–1975)* (Murray Hill, N.J.: Bell Labs, 1978), p. 297–303, 310–15; Roland Mueser, ed., *Bell Laboratories Innovation in Telecommunications, 1925–1975* (Murray Hill, N.J.: Bell Labs, 1979), p. 72.
25. Fagen pp. 336–8; Davies, p. 34; MT&T *Monthly Bulletin*, 41,11 (Nov. 1948).
26. Fagen, pp. 316–7; Mueser, p. 15; Mabon, p. 42.
27. Mueser, p. 157; Northern Electric, *Annual Report* (1965), p.8. The full system name was T1/D1. D1 is the digital channel bank that converts analog voice signals into pulses and reconstructs them at the receiving end.

and data transmission over the public telephone network.[28]

Time Division (Digital) Switching

The increasing use of PCM-TDM makes time division switching possible and desirable. The term "time division" switching will be used here, rather than the common but ambiguous term "digital switching," because digital elements have long been present in telephone exchanges. Switching systems are generally divided into control elements and network elements. Digital techniques have been used for the control function of switching since the first automatic switching systems were introduced at the turn of the century. In addition, digital signals may pass through networks designed for analogue voice signals. The innovation in so-called digital switching is to change the operating principles of the switching network from "space division" to "time division," thus taking advantage of the discontinuous nature of digital signals.

All previous telephone switching systems, from step-by-step up to and including the first electronic, stored program controlled systems, have depended on networks that create a separate path for each signal through the exchange for the duration of a conversation. Time division switching, on the other hand, allocates each call a separate time slot on a shared path. By sampling, encoding, and then multiplexing a number of separate analogue signals, each call can share a single physical path, or "bus," through the exchange while remaining separated into individual "channels." Switching involves shifting each coded sample of a signal from an incoming time slot to an outgoing time slot, first between various intermediate stages in the network, then onto a line leading to the call's destination.

The switching from input time slots to output time slots can be accomplished either by storing the samples, essentially delaying them until the desired time slot, or, if the time slot is not to be changed, by establishing a direct connection without delay to an output line. Essential to Time Slot Interchange (TSI) is a memory system or buffer that allows selective delay of individual samples before assignment to new time slots. For example, the digits of one sample in a frame are stored in the first slot of memory, the next sample in the second slot, etc. The central processing unit then reads out the information in a new order, depending on the route it has devised for the signal. The contents of slot 1 are read out, for example, into slot 4 of the output.

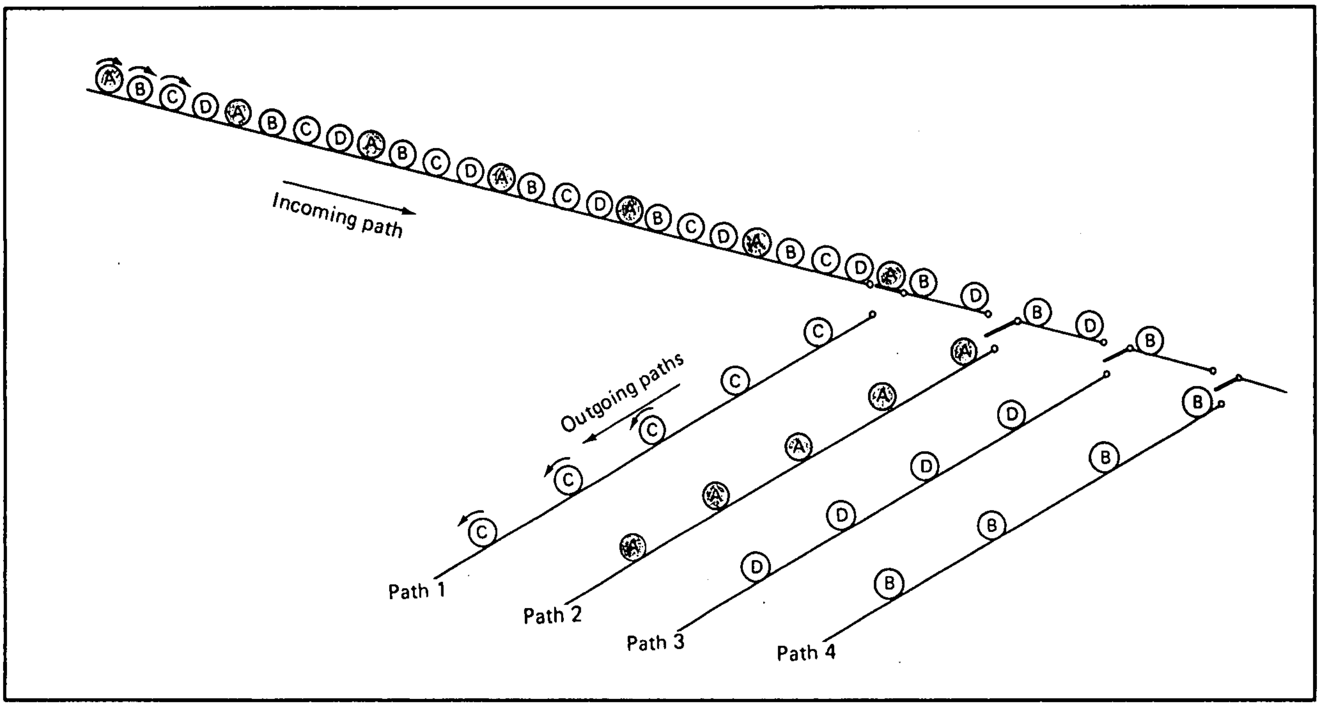
Large switching systems use many parallel time slot interchanges linked by buses. Space division techniques (i.e. switches, gates, or relays) enable connections between these buses. The space division stages are extremely rapid, only connecting circuits for the instant required to pass the bits in one time slot. They therefore differ from traditional space division switches that remain in place for the entire conversation. Because of their rapid action, large numbers of lines may share a relatively small number of space switches. Exclusively space division exchanges must have a large number of crosspoints to accommodate the numerous possible simultaneous connections through the network.

The process of receiving signal information and setting up the complex connection through time and space is carried out by an electronic computer using stored program control. The switching network is, like the control, entirely composed of miniature, semiconductor, integrated circuits. The process of time and space switching is carried out by standard machine logic procedures.[29]

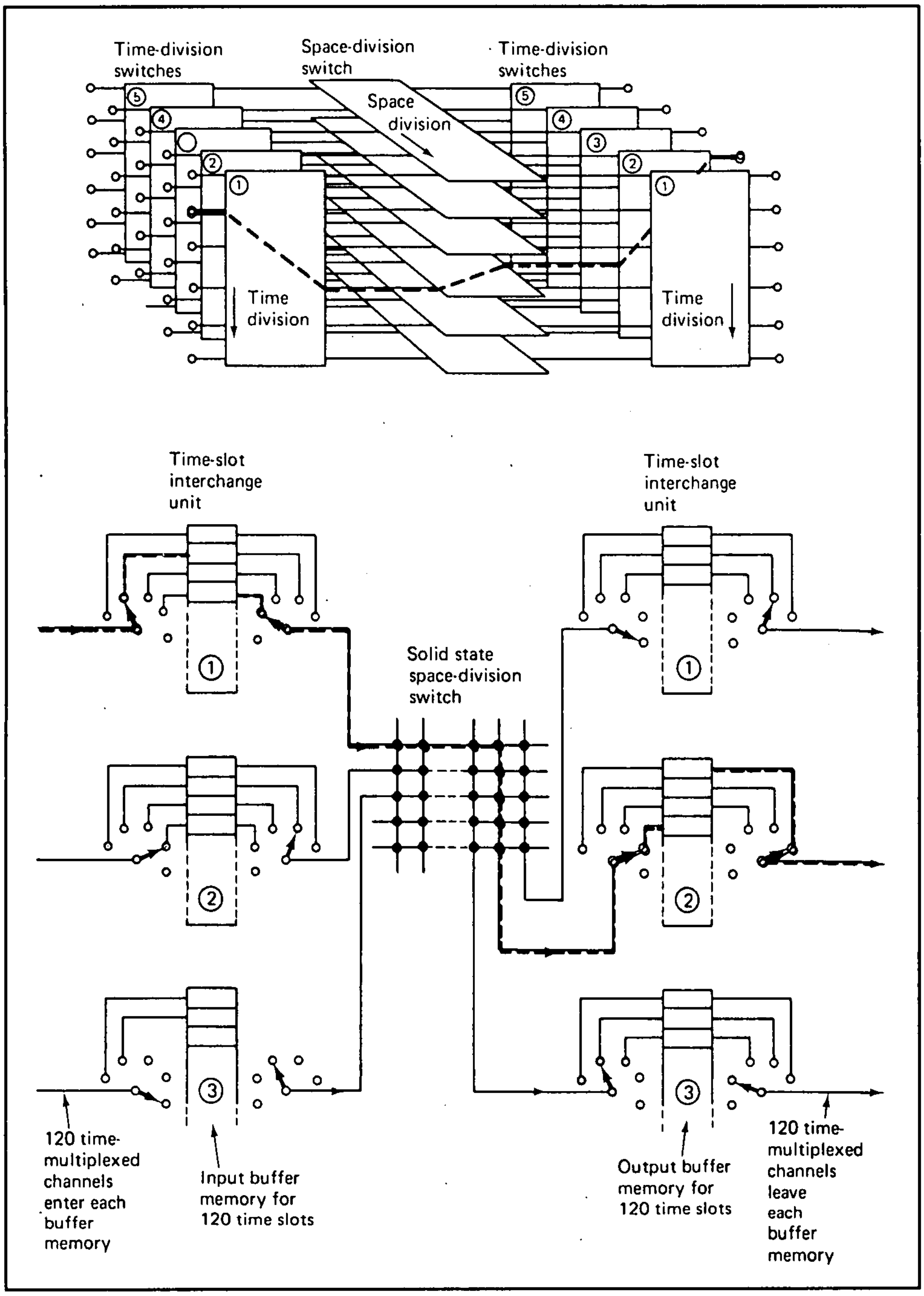
One of the most important features of time division switching is its flexibility for all kinds of communications use. It allows switching of different types of users requiring different data rates. Analogue space switches provide users with the same channel, regardless of whether voice, low speed data, high speed data or video are transmitted. This does not make efficient use of the capacity and flexibility of digital transmission, which

28. Mueser, pp. 168, 170; Martin, *Future Developments in Telecommunications* (Englewood Cliffs, N.J.: Prentice-Hall, 1977), pp. 511, 517–8; George & Nichols, p. 282; Bill Deacon, "New Telephone Plant of the 1970s," *Telesis*, 2,7 (Winter 1972), p. 5; Gordon E. Inns, "Cost Savings on the Digital Road," *Telesis*, 1976/3 (August 1976), p. 196; James Martin, *Future Developments in Telecommunications* (Englewoood Cliffs, N.J.: Prentice-Hall, 1977), p. 501.

29. Amos E. Joel, Jr., "Digital Switching – How It Has Developed," in *Electronic Switching: Digital Central Office Systems of the World*, Amos E. Joel, Jr., ed. (New York: IEEE Press, 1982), pp. 2–5; Joel, "Towards a Definition of Digital Switching," in Joel, pp. 14–15; Martin, *Telecommunications and the Computer*, pp. 489–93; Sam G. Pitroda, "Telephones go digital," *IEEE Spectrum*, (October 1979), pp. 58–9; Frank Banks and John Hopkins, "Preparing for the Digital World," *Telesis*, 5,2 (April 1977), p.36; *McGraw-Hill Encyclopedia of Science and Technology* (1982), s.v "Switching systems (communications)."



A simple time division switch showing a time-multiplexed bit stream being switched from a bus to four outgoing channels. The separate switch for each path could also be represented by a single rotary switch synchronized to the bit stream, as in the preceding figure. No delay, or time slot interchange, is used here. (Martin, Telecommunications and the Computer, *p. 484)*



The ESS No. 4 switch can switch calls between more than 100,000 trunk lines by a set of time-division switches interconnected by a space-division switch. Time slot interchange is essential to this system. The dotted line shows the path of a typical call through the exchange. (Martin, Telecommunications and the Computer *[1976], p. 492)* allocates only enough channel capacity as is needed for a transmission. Time division switching therefore, is well adapted to a future integrated telecommunications network carrying a variety of signals in digital form. It has also been noted, however, that time division switching systems designed for the relatively slow (64,000 b/s) bit rates of telephone and data signals may not function at the high rates needed to process complex signals like video. In this regard, time division switching has the same bandwidth limitations as time division multiplexing.

But this does not disparage the real advantages that time division switching holds for the telephone network as it becomes increasingly reliant on digital transmission. In order to switch digital multiplexed signals to separate destinations through an analogue exchange, the signals must first be demultiplexed and converted to analogue, then often recoded and remultiplexed afterward. This digital/analogue interface represents extra equipment cost and causes the introduction of noise and crosstalk. Hence where the first digital transmission facilities have been in urban interoffice trunk lines, the first time division offices have been used to connect them.

But the interface problem also works the other way. Time division switching requires a sampled signal. Unsampled analogue signals (for example, those incoming from the local loop) must be sampled, coded, and then multiplexed before they can be processed. The necessity of interfaces between analogue signals and digital switches is the biggest technical and economic barrier to the introduction of digital switching into a largely analogue telephone plant. But as more digital transmission is introduced, digital switching becomes economically and technically justifiable. As more time division offices are introduced, there are further incentives to convert to digital transmission, which in turn promotes more time division switching at other levels in the network. This is especially true as telecommunications demand increases and costs must be contained.

When designed in modular form as part of switching "families," time division switching systems can cost less to build, house, maintain, and expand than comparable analogue space division facilities. A single network design can be adapted to serve anything from a small local office to a large city exchange serving 100,000 lines or more. The use of common, modular hardware means that common parts may be manufactured and kept on hand for all systems. Initial capital costs are kept low by providing enough common control equipment for future growth but installing only enough network modules for current demand. Additional network modules are only installed as growth requires, meaning no idle capacity is kept on hand against future requirements. Modular software design, meanwhile, allows teams of engineers and programmers to quickly develop software systems and introduce software changes as technology and demand evolves. Modular design also means less diversity of equipment, which leads to ease of maintenance. The use of maintenance, fault detection, and diagnostic programs, which can take up almost half of the programming in a system's software, also lowers maintenance costs.

The use to time division to share space switches among many circuits results in less expensive and smaller equipment and greatly reduces floor space needed to house it. At the same time, time division switching centres can have a larger capacity than space division systems. One machine can serve a larger number of lines and a larger geographical area, meaning fewer local switching centres are needed. This lowers operating and capital costs.

A variety of distributed terminals and remote switching locations controlled by one central processor allow the extension of direct-dial and enhanced services to new areas without the necessity of constructing a new exchange or extending long lines of copper wire pairs. Digital subscriber carriers in the subscriber loop can digitize and multiplex signals for transmission to a distant central office, bringing city-type service to rural communities without the need to open a new central office. On the same principle, rapid growth areas, like a high-rise apartment complex, are easily accommodated without need for new copper wire pairs. In addition, old, small analogue switching centres can be closed and replaced by digital remote units that sample and multiplex subscriber calls and transmit them to a large, central digital exchange. Similarly, remote concentrators using PCM allow, say 100 subscribers, to share 20 PCM channels leading to a switching office. This again avoids the need for lengthy wire pair connections for each terminal.

The heavy reliance on software in time division switching systems reflects the versatility of stored program control that was recognized by the developers of the first electronic switching systems in the 1950s. Recently, however, the reliability of a highly integrated telecommunications network increasingly dependent on complex software control has been brought into question. On 15 January 1990 a mysterious flaw in newly installed software prevented 50 million long distance telephone calls from going through. About half of all attempted calls failed to be completed during the nine hour service breakdown, the worst in the history of AT&T. This was far above the service standard in the industry, which stipulates no more than two hours downtime in the entire forty year service life of a system. The AT&T incident demonstrates how the operation of the entire network becomes endangered by errors in programming.

On the whole, telephone companies have judged the merits of time division switching and digital transmission to outweigh the drawbacks.[30]

The first application of time division switching was in small private branch exchanges (PBXs) that used pulse amplitude modulation (PAM). In 1958 Bell Labs began development of a small electronic private branch exchange using time division. This PBX, later renamed No. 101 ESS, was first put into commercial operation in November 1963 in Cocoa Beach, Florida. The switching network and control unit of No. 1 ESS were entirely composed of semiconductor electronic devices. A compact switch unit, commonly serving 200 lines, was the only apparatus located on customer premises. It was connected by a data link to the much larger control unit located at the central office. This unit received signalling information and issued switching instructions to the network. Like later electronic-mechanical hybrid systems, No. 1 ESS used stored program control. Information was stored on magnetically encoded cards that were easily accessed and recoded for new functions. The control unit could accommodate 32 separate switch units and a total of 3200 lines. This was to allow control costs to be distributed over as many lines as possible.

Because it served so few lines, the network design of No. 1 ESS and similar PAM time division switches was much simpler than the central office systems later introduced. Electronically gated speech samples from incoming lines were multiplexed on a common path within the switch called a "bus." The central control unit would connect a subscriber by instructing the switch of the called party's line to connect to the bus in the appropriate time slot.[31]

The Federal Communication Commission's 1968 "Carterfone" decision, which allowed telephone subscribers to attach their own apparatus to telephone lines, created a new U.S. market for PBXs. Northern Electric began development in 1970 of a low-cost, electronic PBX that resulted in the introduction in 1971 of SG-1 PULSE. This was similar in principle to No. 101 ESS, being a time division switch using pulse amplitude modulation. But control functions were self-contained, rather than located in the central office. This made it, according to its developers, "the first self-contained commercial PBX on the TDM principle." In the mid-1970s, time division PBXs using pulse code modulation began to appear, with Northern Telecom introducing SL-1 in 1975.[32]

While digital transmission was first introduced on a large scale in the 1960s, PCM time division switching was a technology of the 1970s. In the late 1960s and through the 1970s, telecommunications companies in most industrialized countries developed digital time division switching systems, some being used for military purposes. Some of these systems never progressed past lab experiments or field trials, but many proceeded to commercial installation. The advent of low cost electronic components in the form of integrated circuits reduced the price of such systems to levels competitive with metallic space division switches.

Telephone companies chose two different paths of development. Some installed digital facilities in local areas near telephone subscribers. In France in 1970 the Centre national d'études des télécommunications (CNET) introduced its E 10 system. Later manufactured by CIT-Alcatel, this was a 30,000 line switch integrated with digital transmission in the local area. It initially did not use stored program control, but this and other improvements were added through the 1970s.

30. Inns, p.198; Banks and Hopkins, pp. 36–7; Ernst Munter and Raju Patel, "DMS-200: a toll switch for the Digital World," *Telesis*, (Feb. 1979), p. 3; Martin, *Telecommunications and the Computer*, p. 482; Joel, "Digital Switching – How it has Developed," pp. 5–6; Joel, "Towards a Definition of Digital Switching," pp. 14–5; F.F. Taylor, "Telecommunication Switching," *Encyclopedia of Telecommunications*, pp. 438–443; *Globe and Mail*, 17 January 1990; B.J. Eckhart and G.L. Rainey, "DMS-200 System Evolution," in Joel, p. 93.

31. Amos E. Joel, Jr., *A History of Engineering and Science in the Bell System: Switching Technology (1925–1975)* (Murray Hill, N.J.: Bell Labs, 1982), pp. 201–2, 228–9, 252–7, 461–5.

32. Tony Stansby, "Economical Electronic Switching for the Small Business User," *Telesis*, (Winter 1971), p. 10; Joel, "Digital Switching – How It Developed," p. 9.

Other companies began with trunk switching and provided an "overlay network" at levels above the subscriber loop and local office. Telephone companies in the United States and Canada, which had invested considerably in short-haul trunk digital transmission (T1), chose this latter path for North America. The first such switch to proceed beyond the laboratory was the Empress system, field tested by the British Post Office in 1968. It was a tandem, or intermediate, office designed to accommodate six 24-channel TDM lines. The first system in commercial operation was AT&T's No. 4 ESS, a time-division toll switch introduced in 1976.

Progress in integrated circuit technology led to a flurry of announcements in 1976 of local switching systems ranging in size from several hundred to several thousand lines. Able to interface with analogue lines at the local office or at remote locations, they began entering service in 1977–78. The first among these was Northern Telecom's DMS series, which followed its "Digital World" announcement of March 1976. This "family" was promoted as "a single range of compatible machines covering the total range of telephone operating company requirements."[33] It used modular hardware and software to allow for low initial cost and easy expansion, simple feature enhancement, and technological evolution of individual modules without complete redesign.

The DMS family featured several levels of switching centre. DMS-10, a community dial office developed by Northern Telecom's American subsidiary and introduced in 1977, served local switching centres in the 300-6000 line range. DMS-100 was a standard design for larger offices that could be configured for a variety of switching functions. Within the DMS-100 group were the DMS-200 toll office (1979), the DMS-100 large local office (1980), and the DMS-300 international gateway switch for overseas service (1981).[34]

A fundamental element in the DMS family was a system for extending the digital network almost onto the subscriber's premises. DMS-1, introduced in 1977, was designed to serve two functions. As a digital subscriber carrier system between rural areas and a distant local office, it allowed low cost one-party service by multiplexing locally and thus saving on copper wire and the need for a local office. In a digital network, DMS-1 served as a remote terminal for a digital central office. Instead of analogue/digital interface at the central office, subscriber analogue signals would be digitized at the remote terminal.[35]

The transition to digital time division switching was swift. In 1977 the number of analogue space division switches on the market exceeded digital time division systems. By 1983, no fully analogue systems remained for sale in competitive world markets. Moreover, there was an increased uniformity among the digital systems as manufacturers standardized their products to meet international specifications for interconnection through Common Channel Signalling (CCS). This was an essential prelude to the Integrated Services Digital Network (ISDN).[36]

Increasing Bandwidth with Digital Transmission

Digital transmission techniques, both pulse code modulation and time division multiplexing, can be applied to any transmission medium: a simple pair of copper wires, a coaxial cable, a radio wave, or, more recently, optical fibre. In the 1970s, increasing demand by corporations for high-speed data transmission, coupled with a continuing increase in the use of long distance telephone, forced telecommunications companies to find new ways of increasing capacity for data services without losing voice circuits. Short-haul transmission and switching was already becoming digital, so it also made sense for new long-haul techniques to be digital. The first tendency, aimed at meeting demand into the 1980s, was to apply digital technology to existing transmission media.

In 1970, for example, Northern Electric carried out extensive development work on a digital coaxial cable carrier system for long-haul voice, data, and television transmission. Bell Canada began construction of a route for LD-4 in the busy corridors between Montreal, Ottawa, and Toronto in 1972. This system provided a total capacity of about 274 million b/s (Mb/s) on a standard 12 tube cable,

33. Munter and Patel, p. 3.
34. Munter and Patel, p. 3; Real Gagnier, "The DMS-100 in Perspective – An Overview," *Telesis*, 4 (1980), pp. 2, 4–5; Neil Kaden and Jim McGee, "DMS-300: Gateway of the Digital World," *Telesis*, 4 (1981), p. 22; "Editorial," *Telesis* (August 1978), p. 289.
35. Mike Frame, "DMS-1 is part of a digital family," *Telesis* (August 1977), pp. 99–101.
36. Graham Langley, "Modern Switching Means Digital Switching," *Telephony* (9 July 1984).

compared to 1.544 Mb/s on a standard T-1 circuit.[37]

LD-4 was developed in response to the increasing demands on the long-haul Canadian microwave system. During the 1970s, this radio technology was also adapted for digital transmission. Data Under Voice (DUV), developed by a number of communication companies in the early 1970s, was a stop-gap measure designed to squeeze a 1.544 million b/s data channel from an unused frequency band under the microwave system's voice channels.

A more fundamental change was the introduction of "digital microwave." This involved the use of pulse code modulated, time division multiplexed signals as the basis for modulating a radio carrier wave, rather than assigning each data or voice signal a separate frequency band, as in frequency division multiplexing. The RD series of digital microwave systems (91 million b/s) were designed by Bell Northern Research for Northern Telecom and Bell Canada in the 1970s and early 1980s. As deregulation and the break-up in the American Bell system stimulated the construction of competing long distance networks in the United States, this type of long-haul digital microwave became very attractive as an inexpensive, and hence profitable, means of providing high-bandwidth digital communications. Digital microwave was also used for short distance applications, both by telecommunications companies as a way of providing temporary and emergency links, and by large private users wanting more bandwidth at a price below that offered by telephone companies. By using short-haul microwave, these companies could essentially by-pass the local loop, connecting dispersed offices and possibly connecting to a long-haul microwave network or a satellite.[38]

Satellites links, like microwave and other transmission media, can be adapted for digital transmission. Until the mid-1970s, most satellite communication systems used frequency modulated signals that were frequency division multiplexed. Several of these signals from a number of earth stations, each on a different carrier frequency, would usually be processed simultaneously by a single satellite transponder in a process called Frequency Division Multiple Access (FDMA).

As demand for more digital transmission increased, and the cost of digital circuitry declined, new satellite systems were designed to operate digitally. Individual signals were time division multiplexed and the radio carrier, the same frequency for all earth stations, was modulated using phase shift keying. A single satellite transponder processed a number of signals from different earth stations in sequence, each signal reaching the satellite as a high-speed burst in an assigned time slot.

Time division multiple access (TDMA) with digital modulation makes more efficient use of the satellite transponder's bandwidth and power, allowing for a higher capacity than possible if the same satellite used an FDMA system. FDMA requires gaps or, "guardbands," to be placed between each carrier band to prevent interference, taking up valuable transmission bandwidth. While each TDMA burst requires control and synchronization bits and gap between it and the next burst, this "overhead" can be minimized with proper system design. Another limitation of FDMA is that the electronic characteristics of the transponders cause the carriers to modulate each other. This

37. "Canada's New Coaxial Cable Communications System," *Telesis*, 2,3 (Winter 1971), p. 28; Doug Carruthers, "Planning for a Long-Haul Digital Transmission System," *Telesis*, 2,5 (Summer 1972), p. 27; Bob Johnston and Wally Johnston, "LD-4: A Digital Pioneer in Action," *Telesis*, 5,3 (June 1977), pp. 66–7; Bell Canada, *Annual Report* (1974), pp. 29, 35; Northern Electric, *Annual Report* (1970), pp. 16, 19; (1974), p. 8; (1976), p. 3.

38. Joe Baart and Dave Baker, "DG 1: Equipment for Parallel Data Under Voice Systems," *Telesis*, 4,2, (Summer 1975), pp. 46–7; Don Silverthorn, "Digital Radio – Tommorrow's Microwave Technique," *Telesis*, 2,6 (Fall 1972), p. 25; Pierre Hervieux, "RD-3: An 8 GHz Digital Radio System for Canada," *Telesis*, 4,2 (Summer 1975), pp. 53–9; Ivan Godier, "Digital Microwave Radio – A Proven Case," *Telesis*, 5,6 (December 1977), pp. 162–5; "First Step to a Cross-Canada Digital Radio System," *Telesis*, 5,11 (October 1978), p. 352; Steve Barber and John McNicol, "The RD series of Digital Microwave Radio Systems," *Telesis*, 12,1 (1985), p. 42; RD-6B: Bryan Godfrey et al. "On the World's Wavelength," *Telesis*, 14,2 (1987), pp. 37–9; Larry Lannon, "Is Short-haul Microwave's Future, Well, Short?" *Telephony Transmission Special* (October 1988), pp. 67, 69; James E. Innes, "The Microwave Factor in Disaster Recovery," *Telephony* (9 October 1989), pp. 38, 42.

can only be alleviated by operating the transponder at less than maximum power levels. On the other hand, because TDMA uses a single carrier frequency and allows only one user access to the transponder at a time, all the transponder's power may be used. TDMA is also more flexible than FDMA. Transponder capacity is more easily and quickly reassigned to meet the constantly shifting demand by users for channels of differing capacity. Rather than reassigning earth station carrier frequencies as demand changes, circuits may be allocated by simply reassigning time slots and lengthening or shortening the duration of each earth station's pulse.

The first commercial TDMA service in Canada was offered by Telesat Canada in 1975.In the United States, the expansion of satellite capabilities and an increase in number of satellites, both meeting and anticipating demand, made private use of dedicated channels by individual users possible. This was especially the case for large users exploiting a variety of services like digital facsimile, electronic mail, and closed-circuit television. Some large users by-passed the local loop, whose low bandwidth was a bottleneck in a satellite transmission system. They installed microwave dishes at their places of business to transmit and receive signals via satellite circuits or a terrestrial microwave network provided by a supplier other than the local telephone company.[39]

Fibre Optics

In one way, fibre optics may be seen as a logical extension of the drive for increased bandwidth. In the twentieth century, telecommunications researchers have exploited higher and higher frequencies for their greater bandwidth and attendant improved capacity and efficiency. In this sense, the use of light waves is a progression from microwaves, which themselves were improvements over the radio waves used in multiplexing voice-band signals. The rapid adoption of fibre optic communications reflects this trend of developing higher frequency bands in response to the increasing demand for bandwidth from telecommunications users. From the time that fibre optics first became technically feasible in 1970, only seven years were to elapse before its first commercial trial. And by the end of the 1970s, telecommunications companies all over the world were making large-scale installations. This rapid implementation of fibre optics coincided, as is seen above, with attempts in the 1970s to increase the capacities of existing transmission technologies. Both these movements were responses to the demand by large telecommunications users, most notably businesses, for greater capacities to accommodate the increasing flows of information passing between telephones, between computers, and between other business machines (like facsimile), both locally and across the increasingly far flung reaches of their operations.

But at the same time that fibre optics can be seen as a logical extension of previous developments, it is also a radical departure from existing technology. For in its use of light as a carrier of information, it challenges electronics as the basis for practice in the telecommunications industry. A simple fibre optic communication system consists of an electrical input signal, usually but not necessarily digital, a light source that converts the signal to light pulses, an optical fibre lightguide to act as the transmission medium, and an optical detector that receives the pulses and converts them back into an electrical signal. Two way communication, as in telephony, requires two of these circuits.

The light source is a device made of semiconductor material (gallium aluminum arsenide or GaAlAs) that, when an electrical current is passed through it, can be caused to emit a narrow beam of photons, or "particles" of light energy, in waves of uniform frequency. A digitized electrical telephone signal will, for example, cause the source to emit light during the presence of a binary '1' in the device, and to not emit in the presence of a binary '0.' The two possible devices for use as light sources are lasers and light emitting diodes (LEDs). LEDs are only suited for shorter transmission distances and slower information rates because only a portion of its wide beam can be coupled to the fibre. And this usable light is spread over a wide range of frequencies that propagate in the

39. S.J. Campanella, "Satellite Communications," in *Encyclopedia of Communications*, 317–24, 330; James Martin, *Communications Satellite Systems* (Englewood Cliffs, N.J.: Prentice-Hall, 1978), PP. 234, 243–4, 251, 254–6; "Modulation and Multiple Access Techniques," in *Communication Satellite Systems: An Overview of the Technology*, R.G. Gould and Y.F. Lum, eds. (New York: IEEE Press, 1976), pp. 81–3; Robert K. Kwan, "The Canadian Domestic System," in *Communication Satellite Systems: An Overview of the Technology*, p. 35; Ken Sherman, *Data Communications: A Users Guide* (Reston, Virg.: Reston, 1985), pp. 306–8; Tom Forester, *High-Tech Society: The Story of the Information Technology Revolution* (Cambridge, Mass.: MIT Press, 1987), pp. 91, 108–9; *Communications (Understanding Computers)*, pp. 95–100.

fibre at different rates (chromatic dispersion), leading to a distortion of the light pulse as it passes down the fibre. The laser, on the other hand, emits light in a much narrower beam with much lower spectral spread, making it ideal for high-speed fibre optic communications. However, it requires more sophisticated electronics, is more expensive and less tolerant of temperature variations, and has a shorter service life. The detector is also a semiconductor device. When light emitted by the source and carried by the fibre strikes the detector, electrons are caused to flow through it. This electrical current then passes into the conventional communication system.

The medium between the source and detector is a hair-thin strand of extremely pure glass. The fibre is actually composed of two layers of glass. The core has a slightly higher refractive index than the cladding, which causes light passing through the core to be reflected inward when it strikes the cladding. This gives the fibre total internal reflection. Three different types of fibre are available: step index multimode, graded index multimode, and step index single mode. Their difference lies in the degree of difference in refractive indexes between core and cladding and in the rate of transition from one to the other. This determines the number of possible paths, or modes, the light may take in reaching the end of the fibre, which in turn will determine the rate at which pulses can be sent through the fibre without error. The more paths that the light takes, the more the pulse becomes spread during transmission, thus lowering the possible signal rate. Step index single mode fibre allows the highest transmission rate but is more difficult to operate at low loss levels. More amenable to telecommunication uses is graded index multimode fibre.

Fibre optic communications have several advantages over conventional copper and radiowave transmission systems. Fibre's lower level of transmission loss allows repeaters to be spaced farther apart, and even eliminated in some interoffice trunks. Fibre also has an extremely high information capacity far exceeding that attainable on a wire pair using advanced digital techniques. A cable of bundled fibres carries billions of bits per second. In addition, glass fibre is electrically nonconducting, making it immune to noise from electromagnetic interference, including crosstalk from adjacent lines, and to power surges in the network. For the same reason, it cannot be easily tapped, making communications more secure from eavesdropping. Finally the light weight and small diameter of fibre cables makes them easy to install in existing ducts. As a large number of fibre cables may fit in the space taken by a smaller number of bulky copper cables, underground capacity in congested urban areas can be increased without major new construction. The main disadvantage of fibre is the care required in handling, installation, and splicing in order to ensure the glass structure is not minutely altered and that joints are perfectly aligned. This, however, can be overcome with the proper training and equipment for installers.[40]

Fibre optic communication is the product of a number of previously unrelated developments in the fields of communications, physics, electrical engineering, and glass technology. Optical communications date back to the visual signalling and telegraph systems that have existed in various forms for centuries. These reached a level of sophistication in the mid-19th century with military use of the heliograph for focusing sunlight into a narrow beam using it to transmit coded messages. Scientists at the same time were learning about the behaviour of light in liquid and in glass, and attempts were made to transmit light through internally reflective pipes. In addition, a method was devised for drawing fine fibres of uniform thickness from molten quartz. While these developments were important to later research, they were not done primarily with a view to inventing an optical communication system. Closer to conventional communication uses, Alexander Graham Bell in 1890 demonstrated the Photophone, which optically transmitted the wave pattern of speech sounds through air and received them on a telephone receiver equipped with a crystalline selenium detector. Bell's device was hampered by the high interference with light beams passing through air.

In the 1930s, Norman R. French of AT&T patented a system for transmitting voice signals on beams of light travelling through a "light cable," though his main interest was in hollow reflective

40. Doug Kulm, "Fibre Optics for Telephony: A Tutorial," *Telephony*, 19 November 1984, pp. 84–98; Dave King & Otto Szentesi, "Fibre Optics: A New Technology for Communications," *Telesis*, 5,1 (February 1977), pp. 3–8; Jeff Hecht, "Making Light Work," *New Scientist – Inside Science*, 17 June 1989, pp. 2–4; Jan Conradi et al., "Fibre Optic Technology: Present and Future," *Telesis*, 8,2 (1981), pp. 15–7; Ronald G. Ajemian & Albert B. Grundy, "Fibre Optics: The Medium for Audio," *Journal of the Audio Engineering Society*, 38,3 (March 1990), pp. 162–70.

tubes and not glass fibres. Work on glass fibres was directed to other uses, as for example, research in the 1950s by Dutch scientists on the use of glass fibres for the transmission of images in submarine periscopes.[41]

Despite this work, it did not derive from nor lead to the conception of an optical communication system using glass fibres. This did not occur until the demonstration in 1960 of an effective laser that could generate a coherent, narrow beam of light. The effectiveness of Theodore A. Maiman's laser, a product of U.S. military-sponsored research and the melding of heretofore disparate knowledge from physics and electrical engineering, sparked an immediate interest in using it as a transmitter in a communications system. But it would be another decade before scientists at Bell Labs could develop a laser that would operate reliably at room temperature.

In the meantime during the 1960s, glass technology was improving. In response to proposals in 1966 for fibre optic transmission system from the British Post Office and Britain's Standard Telecommunications Laboratories, the Corning Glass Company in the United States set out to develop a glass fibre of sufficient clarity with a sheathing of a balanced refraction index to reduce transmission loss to a minimum. The method developed by Robert D. Maurer and his team at Corning was to coat the inside of a tube of fused silica with a mist of silica "doped" with an agent to reduce its reflectivity below that of the surrounding tube. The internally coated tube was then heated until it collapsed around the coating and together they were drawn into a single fibre. This method, which a historian has argued "laid the basis for the optical communications industry," was announced in 1970. At Bell Labs in the same year, scientists achieved both a laser operating at room temperature and a light-emitting diode suitable for short transmission. By 1970, the essential devices and knowledge were in place for fibre optics.

The next few years involved efforts by a number of telecommunications and computer companies to develop materials rugged and reliable enough to withstand commercial operation. In Canada, Northern Electric scientists did pioneering work on LEDs in the late 1960s, and in the early 1970s began theoretical analysis of fibre optic communications systems. Canadian work in fibre optics beyond this level probably dates from 1973, when Bell-Northern Research (BNR) began construction of a fibre drawing facility that, by the following year, was producing fibres by vapour deposition. Also in 1974, BNR, in conduction with Litton, installed a fibre link on a Canadian naval destroyer. This was followed in 1976 by a secure communications system installed by BNR in the National Defence Headquarters in Ottawa.[42]

The first outdoor field test of a fibre optic system was successfully undertaken by Western Electric at Atlanta, Georgia in 1976. In April of the following year, GTE established the first commercial system with a link between its Long Beach, Cal. toll centre and a local exchange 8 kilometres away. Using a light emitting diode, this system operated at 1.544 Mb/s and offered 24 voice channels. Later in the year AT&T began its first commercial trial in Chicago. In Canada, Northern Telecom ran a successful trial between two Montreal switching offices in 1977. In 1978, it ran the Yorkville Fibre Optics Trial in Toronto, offering residents a variety of services possible with the new capacity offered by cable. A similar trial was run with the Manitoba Telephone System in the rural communities of Elie and St. Eustache, near Winnipeg, in 1981. Northern Telecom installed its first commercial system in 1979, an FA-1 analogue fibre link relaying INTELSAT video signals from Teleglobe Canada's satellite earth station to a Bell Canada microwave tower for relay to Montreal. By 1980, it had developed a complete "family" of fibre optic systems providing a variety of standard channel capacities suitable for short-haul, long-haul, and local loop applications.[43]

41. Sami Faltas, "The Invention of Fibre-Optic Communications," *History and Technology* 5,1 (1980), 32–6; *The Bell Telephone: The Deposition of Alexander Graham Bell in the Suit Brought by the United States to Annul the Bell Patents* [1908] (New York: Arno Press, 1974), pp. 190–1; "Bell's Photophone," *New York Daily Graphic*, 20 November 1880 in L.B. McFarlane Scrapbook, Bell Canada Archives, cat. no. 12016.

42. Faltas, 37–43; Trudy E. Bell, "Fibre Optics," *IEEE Spectrum*, 25,11 (1988), 98–100; Alauddin Javed et al. "Fibre Optic Transmission Systems: The Rationale and Application," *Telesis*, 8,2 (1981), p. 5; Koichi Abe et al. "Optical Fibre and Cable: Past, Present, Future," *Telesis*, 8,2 (1981), p.9.

43. Bell, pp. 100–1; Abe et al., p. 9; Manfred Ficker, "FA-1: Video over Optical Fibres," *Telesis*, 6,4 (August 1979), pp. 21–2; Mario Larose, "Practical Experience with Optical Fibre Systems," *Telesis*, 8,2 (1981), pp. 30–5; Kenneth B. Harris, "Rural Canadians First to Learn Full Fibre Optic Capability," *Telephony*, 23 November 1981.

The installation of fibre optic lines in North America, Europe, and Japan after 1979 was rapid. The first large-scale commercial installation of fibre in Canada was undertaken by the Saskatchewan government's Sask Tel in 1980. Completed several years later, the 3200 km network was intended to deliver cable television to cities and towns, where coaxial cables would carry the signal into homes. But Sask Tel predicted fibre would eventually be hooked into telephone switching centres to provide an integrated high-capacity communications network. Other large-scale fibre projects were completed by the end of the decade. In March 1990, after three years of construction, the member companies of Telecom Canada opened the last link in a 7000 km buried fibre line between Halifax and Vancouver. The most difficult portion of the construction involved blasting solid granite in order to carry the line through the rugged Canadian Shield of Northern Ontario. Built in order to accommodate increasing business demand for data communications, Telecom Canada called it the world's longest fibre optic network.[44]

In the United States, fibre optic system construction was spurred by the break-up of the American Bell system in 1984, as competition in long distance telephone service encouraged investment in an economical, high capacity transmission system. As production and installation of fibre increased, the cost of fibre fell significantly. According to one estimate, the cost of a meter of fibre dropped between 1977 and 1985 from US $3.50 to just twenty-five cents. Both AT&T and MCI, the major long distance carriers, installed nation-wide networks of fibre optics that, by the end of the decade, had superseded microwave as the basic transmission medium in the public long distance system. The same massive installation of fibre occurred in Britain after the government's sale of British Telecom and the licensing of a competitor, Mercury Communications. By 1989, fibre carried 65 per cent of British Telecom's trunk traffic, and Mercury's new national "backbone" network was almost entirely fibre.[45]

Fibre optics also began making a contribution to trans-oceanic communication with the laying of TAT-8, a joint venture of AT&T, Teleglobe Canada, and several American and European telecommunications carriers, across the Atlantic Ocean in 1988. This cable experienced some difficulty in its first year of operation, as fishing trawlers cut the cable on several occasions, necessitating a complete shutdown of service in order to make repairs. While perhaps an indication of optical fibres' fragility, these incidents did not, apparently, discourage the laying of more cables. In April 1989 a trans-Pacific cable was laid between Point Arena, Calif., Chikura, Japan, and the American possession of Guam. Also in 1989, TAT-9, a second trans-Atlantic cable was announced, with the link between Nova Scotia and several points in Europe being complete by 1991. By 1989, a message could travel from Switzerland to the United States to Japan entirely on fibre.[46]

While the introduction of fibre was rapid in long-haul operations and in short-haul applications like urban interoffice trunks, fibre optics have not yet been introduced into the local loop leading to most businesses and homes. In the United States, the local loop accounts for fully 90 per cent of all telephone line mileage. One of the reasons for this is technical. Copper wire retains the huge advantage over fibre of providing a metallic connection between the central office and each subscriber's apparatus. This allows the central office to supply the power necessary to operate telephones via the subscriber loops. Assuming telephones in a fibre optic network were powered by the local electric utility, each telephone would still need a back-up power source in the event of a power failure. As yet, telephone companies have not reached a consensus on the least costly and most efficient method of supplying back up power, or of who would pay the cost of powering subscriber apparatus.

44. "Fibre Optics in Saskatchewan," *Telesis*, 7,2 (1980), p. 24; Kenneth J. Head & Trevor J. Truman, "Installation Progresses as FO Network Stretches across Canadian Prairie," *Telephony*, 23 May 1983, pp. 28, 38; Bernard Houle, "Against the Odds," *Telephony*, 6 June 1988, pp. 26–7; Ottawa *Citizen*, 14 March 1990.

45. Forester, pp. 100–1; "AT&T Announces Plans for National Network," *Telephony*, 19 November 1984, pp. 13, 21; James E. Innes, "The Microwave Factor in Disaster Recovery," *Telephony*, 19 October 1988, p. 38; John Williamson, "Big Business Drives U.K. Local Fibre," *Telephony*, 28 August 1989, p. 28.

46. Forester, p. 101; "FCC Gives Nod to TAT-8 Fibre Optic Cable," *Telephony*, 4 June 1984, p. 17; "First Pacific Cable Links U.S., Japan," *Telephony*, 24 April 1989, p. 9; *Globe and Mail*, 6 June 1989; *Data Communications*, January 1990, pp. 45–8.

The other main hindrance to local fibre is its cost. The manufacturing cost of fibre has fallen dramatically, and installation costs have fallen to a point comparable with copper for, say, a new subdivision or office building. But the complete conversion of local loops to fibre will require the removal of old copper, the installation of fibre circuits, and the overhaul of all local equipment. This will require billions of dollars, and cannot be justified without significant demand for the increased bandwidth capabilities of fibre optics. So far, telecommunications companies have failed to interest most residential phone subscribers in services that require more than the copper plant already in place. One early attempt, videotext, has been commercially unsuccessful in most areas where it has been introduced. Another possible means of creating bandwidth demand is by promoting the use of second phone lines and electronic "workstations" by people working out of their homes.

But the main hope of for increasing residential subscriber demand appears to be in the field of entertainment. The possibilities here include pay-per-view television programs, video-game libraries and high definition television (HDTV), a much-promoted but yet commercially-untried technology. In addition, telephone companies are also casting envious glances at the cable television industry. A combination of telephone and basic cable television service might tip the cost balance in favour of fibre. But it would likely meet objections from the local cable monopolies and government regulators, who would see it as little more than technology substitution and a grab for extra business. And certainly any dependence on the repackaging and sale of commercial entertainment will tarnish the futurists' optimistic predictions, made just a decade ago, about knowledge-rich lifesyles in the Information Society.

The only immediate prospects for fibre subscriber loops would appear to be with business. In North America, some new office buildings are being "wired" with fibre, though demand for its flexible, high capacity capabilities seems to be lagging behind supply, except among high-tech companies. In the U.K., British Telecom has been promoting "flexible access," fibre optic systems in London's City and Docklands financial districts. Exploiting the huge capacity of a fibre cable, flexible access allows users to add and subtract voice and data lines as needs change throughout the day without requiring a rewiring of circuits.

The prospects for fibre in the local loop might also be improved by a reconception of the network architecture. Presently, telephone systems are based on a star configuration, in which each subscriber is served by a separate copper pair radiating out from a central office. A more economical design for a fibre optic network might be a "bus" configuration, in which a large number of subscribers are linked to single circuit, or bus, that carries their multiplexed messages to the central exchange. It appears that a completely fibre optical telecommunication system is still decades away, though telecommunication manufacturers like Northern Telecom have begun to offer new equipment that adapts existing digital switching offices to the high-speed standards of fibre optics, easing the process of transition substantially. In addition, recent developments in high-speed optical processors indicate that optics will continue to assume a greater role in the entire field of information technology.[47]

47. Dustin J. Becker, "Power Problems in the Fibre Loop," *Telephony*, 15 January 1990, pp. 46–50; Les Hewitt & Mark Pitchford, "Making the Transition: Fibre Winds its Way Home," *Telephony*, 15 February 1988, pp. 34–5; Jeff Hecht, "Optical Fibres Find Their Way Home," *New Scientist*, 3 March 1988, pp. 63–6; Susan Ubis, "Coming to a Residence Near You," *Telephony Transmission Special*, October 1988, pp. 33, 35, 39; Ottawa *Citizen*, 8 November 1988; *Globe & Mail*, 3 May 1990; John Williamson, "Big Business Drives U.K. Local Fibre," *Telephony*, 28 August 1989, pp. 28, 31; Pallab Ghosh & Jeff Hecht, *New Scientist*, 14 July 1988, pp. 36–7; *Globe & Mail*, 13 October 1989, 30 January 1990; 15 May 1990; 17 May 1990.

4 Merging of Voice and Data Networks

The "Intelligent Network"—Not Yet All Things to All People

During the 1970s, the improvement of data communication networks completed the technological developments necessary to, at least theoretically, make computer power and masses of computer-stored data available to anyone with a proper terminal connected to the public telephone network. Already, the popularity of time sharing systems among specialized users encouraged some in the computer and telecommunications industries to envisage the development of huge computer service utilities that offered the general public access to remote computers and databases through terminals in their homes and workplaces. Planners with telephone companies, meanwhile, saw their increasingly digitized networks as the core for a "wired city," served by an all-purpose communication system. James Martin, a prolific writer on computers, stated in 1972 that the development of computers and digital telecommunications marked a "fundamental step forward in civilization" and predicted:

> *Data transmission will become as indispensable to city-dwelling man as his electricity supply. He will employ it in his home, in his office, in shops, in his car. He will use it to pay for goods, to teach his children, to obtain information, transportation, stock prices, items from the shops, and sports scores; he will use it to seek protection in the crime-infested streets. The best potentials of data transmission will give man more knowledge, more power, more leisure time; he will have less mandatory travel; his job will be more interesting. In many ways his life will be richer.*[48]

In addition to the creation of a digital infrastructure, the 1970s saw several ambitious experiments in public information access via videotex technology.

By the end of the decade more pieces were about to be put in place. At this time voice and data communications networks were still functionally separate. Though they often shared the same transmission circuits over the public telephone network, data users and voice users were channelled through different switching systems that had been specially designed for voice or data traffic. But the conversion of the telephone network to digital switching and transmission, the growing use of computers and microchip-based devices, and the exploitation of high bandwidth transmission media like digital radio and fibre optics, leant new realism to what was now called, more ambitiously, the "digital world." Telephone companies were actively planning and constructing the components of a future integrated digital network.

Despite the fact that the telephone network beyond the local loop was by 1980 primarily digital, the technical complexity of creating an integrated system was still great. As has already been seen in the discussion of packet switching, different types of information—voice, interactive data, bulk data, telemetry (alarms and meter readers), fixed images (fax, computer graphics, freeze frame TV), audio, and video—have specific characteristics that make them difficult to reconcile in a single network. The most important of these characteristics are the required bit rate, the degree of "burstiness," the acceptable set-up and response times, the tolerable error rates, and the call, or session, length. Moreover, an integrated network must ensure compatibility among devices provided by different manufacturers. Without standards for universal interconnection an integrated public voice and data network will not be possible. Much of the 1980s was consumed in overcoming these technical and political problems; the technical, though perhaps not commercial, feasibility of an integrated network was established.[49]

48. James Martin, *Introduction to Teleprocessing* (Englewood Cliffs, N.J.: Prentice-Hall, 1972).

49. Frank Banks & John Hopkins, "Preparing for the Digital World," *Telesis*, 5,2 (April 1977), pp. 35, 37; Wayne J. Felts, et al., "Bell's Concept of the ISDN," *Telephony* (25 October 1982), p. 43.

But it is still unclear if an integrated and universal network will be realized beyond the technical level. There have been some noted failures. But these might only be temporary aberrations in a long term trend toward integration and mass use of data communications services. Analysis is complicated because we are still in the midst of events, and it is difficult to separate important developments with long-term implications from the marketing patter of the various companies hoping to profit from a new field of business. This process has especially been intensified by the increase in telecommunications competition resulting from deregulation in the United States and the United Kingdom. In Canada, the telephone companies have attempted to forestall an encroachment on their business by launching their own ventures into integrated voice and data services. In either case, there have been rather extravagant claims made for the new products and services. The visionary language of the business, for example, has suffered from an extended bout of rhetorical inflation. The Wired City begat the Digital World, while the Integrated Digital Network became the Integrated Services Digital Network, which spawned the Intelligent Network, the Advanced Intelligent Network, and, more recently, the Intelligent Universe. But it is not yet clear whether the new services offered under these banners will be embraced by more than a few specialized users. We might question whether the integrated network will indeed be universal, let alone ask the more metaphysical question raised by the marketers: can we make the universe intelligent?

Videotex—Data Communications for the Mass Market

Those who predicted that data communications might one day become as indispensable as our power supply were encouraged by the rise of publicly accessible databases in the 1970s. The first databases were developed in the early 1960s, being as old as the first time sharing systems. They provided mostly bibliographic services for people conducting research in the voluminous technical literatures of the sciences, engineering, and medicine. Also available were various types of stored numerical information specific to the various scientific and technical fields. The number of these databases increased as time sharing computers became more common, low-cost and fast terminals were manufactured, telecommunications costs fell, and random access storage devices became cheaper and more capacious. In the 1970s, the number of institutional databases serving specialized clientele was augmented by more publicly accessible databases, often based on existing time sharing services, which stored huge archives of legal case texts, news and magazine "clippings", and encyclopedia articles, as well as offering rapidly updated news and financial information. In North America, these were accessed over the expanding value-added packet networks, whose rapid communication and protocol conversion services made them, in one author's words, "electronic flea markets" on which the database vendors sold their products.[50]

The overall growth in networks, time-sharing computers, and public online databases was enough to encourage some journalists, sociologists and experts in the computer and telecommunications industries to envision a future society in which everyone was linked into a public computer network to conduct their work and communicate for work and leisure. Indeed, many in the computer industry hoped to ride this trend by supplying the necessary terminals and mainframe computers anticipated, and by those in telecommunications looking for a more profitable sideline to the steady but unspectacular growth of the "plain old telephone service" (POTS).

But the fact was that in the 1970s, the existing computer and data communications technology was only exploited by a relatively small number of people in business, science, and technology. Many believed that the major obstacle to broader use of computer networks and databases was the specialized content of the services offered and the difficulty of using them without knowledge of and experience with computers. Information was not tailored to general users but to specialists. And access codes, protocols, and special programs were needed to gain access to and effectively use the databases. Though databases were arranged in logical files under topics and subtopics, an information specialist who knew the proper commands and the structural idiosyncrasies of each database was often needed to find desired information.

One solution to this perceived barrier between computers and the general population was videotex (also called viewdata), the product of a strange mix of technological utopianism, democratic sentiment, business pragmatism and marketing folly. Videotex

50. Jerome Aumente, *New Electronic Pathways: Videotex, Teletext, and Online Databases* (Newbury Park, CA: Sage, 1987), pp. 76–83.

was not so much a new technology as an attempt to popularize existing technology and offer information of general interest. Home users were to have access to a huge package of information and services held in a central computer and transmitted via the public telephone system to specially adapted television sets or single-purpose videotex terminals. A telephone call and a few simple commands made on a primitive keypad gave access to the system, simple "menus" led users through the information and service choices, and the data itself was attractively presented in a "page" format with colourful computer graphics. The content, which depended on what was offered by private information providers, varied from nightly entertainment listings, news items, and stock market listings to directories, home shopping catalogues, correspondence courses, and computer games. Some videotex systems, like Bell Canada's Vista project, offered the use of the remote computer for calculations and the storage of personal files, recipes, and mailing lists.

Much early videotex research was carried out by government telecommunications corporations in Europe and, in Canada, by the federal Department of Communications (DoC). In the early 1970s, the British Post Office began developing Prestel, with its first field trials in 1976 and commercial introduction in 1979. In France, a state communications research centre developed the ANTIOPE system for displaying text and graphics on televisions or computer terminals. Field tests were launched in the late 1970s. In the early 1980s, the French Direction général des télécommunications (DGT) began giving away free "Minitel" terminals in order to promote its videotex venture, known as Télétel. In Canada, researchers at the DoC devised a sophisticated graphics system, christened Telidon, whose features were later incorporated into the North America standard for videotex graphic displays. DoC also sponsored a number of videotex-Telidon trials across Canada in 1980 and 1981 in cooperation with various telephone companies. In the United States, where telephone companies were prevented by law from providing computer services, the first commercial videotex ventures were launched by newspaper companies, who saw videotex as a potentially lucrative extension of their publishing businesses.[51]

Despite the large number of field trials and commercial ventures and the millions of dollars invested, videotex was in the 1980s a commercial disappointment, if not an outright failure. One of the few exceptions to this case was France, where the DGT had promoted videotex by withdrawing printed telephone directories and giving away free terminals; telephone subscribers searching for a number consulted an electronic directory offered as part of the videotex service. In other countries, consumers preferred to consult existing print media for published information already available at a lower price. For those few willing to use computers for more ambitious applications than mere information retrieval, microcomputer technology evolved rapidly to outstrip videotex. Relatively complex calculations could be made with handheld calculators costing a few dollars. And personal computers combined increasing processing power and memory, a limited communication ability, plentiful useful and recreational software, and a growing sophistication in graphics. For hobbyists and frequent home computer users, PCs were the overwhelming preference over expensive videotex terminals and service charges.

The few successes for videotex came in business, not home consumer, applications. Britain's Prestel, for example, provided various services to the travel industry that had not, as they had in North America, existed beforehand. In North America, success went not to the colourful, graphics-oriented, mass-market ventures, but to the more mundane online database providers. In fact, the boundary between videotex and these services blurred considerably. Aside from colour and graphics, online databases increasingly offered most of the advantages of videotex—news, information, transactional services, simple menus and indexes, page formats, and frequent updates—but offered their services to specialized markets. In the United States, the *Wall Street Journal*'s Dow Jones News/

51. Aumente, pp. 14–19, 27–8, 32–7, 39–40; Efrem Sigel, ed. *The Coming Revolution in Home/Office Information Retrieval* (White Plains, N.Y.: Knowledge Industry, 1980), pp. 10–1, 18–9, 57–69, 113–122; Vincent Mosco, *Pushbutton Fantasies: Critical Perspectives on Videotex and Information Technology* (Norwood, N.J.: Ablex, 1982), pp. 1–9, 71–87; Keri Sweetman, "The Changing Face of Videotex: Vista Marries Telidon," *Telesis*, 7,1 (1980), pp. 7–11; William W. Seelinger, "Videotex in the United States," *Telephony* (23 April 1984), pp. 38, 42–3.

Retrieval service offered business news, subject-labelled "clipping" files, and market listings. CompuServe and The Source also provided news and business services, but were especially successful at attracting personal computer users with a variety of electronic mail, special interest bulletin board, and other communication services.[52]

Reflecting consumer resistance to videotex but the modest popularity of online databases, some telecommunication companies shifted emphasis from being videotex service operators to acting as intermediaries between users and the various independent, special-interest databases. Their gateway services essentially added additional "intelligence" to the existing value-added public data network. By dialling gateway switches in telephone central offices, subscribers to such services as Telecom Canada's iNet 2000 received an electronic directory of the various databases. To each it offered automatic connection and access without need of registration or a password, instructions for use, and some limited translation of search commands into a single standard vocabulary. Rather than deal with separate providers, a subscriber dealt with the telephone company as their agent, from which it received a single monthly bill for all services. Though more limited in their ambition than the videotex ventures, the online databases and gateways seemed to serve a need that would justify their existence.[53]

In the final analysis, the videotex ventures of the early 1980s failed because they did not recognize this. A few technophiles aspired to be on the "cutting edge" of technology for its own sake. But as one expert wrote in 1986:

> *Aside from computer hobbyists..., the vast majority of home consumers are not enamored with the technology[,] but what it can do to enhance their lives with minimal costs or disruption to established media habits and life-styles.*[54]

Or, as one journalist stated:

> *Most people will avoid a new gadget unless it meets at least three conditions. First, it must cost no more than a few hundred dollars; second, it must make life simpler; third, in some manner rather hard to define, it must be fun.*[55]

The appeal of cheapness, convenience, and fun has marked such recent videotex ventures as Bell Canada's Alex, which began commercial operation in Montreal, Ottawa, and Toronto in 1990. Operators of the new systems have attempted to repeat the Minitel phenomenon by distributing cheap terminals or videotex adapters for personal computers. They have also tried to sell the systems for their convenience, hoping that consumers' increasingly busy lives and their new acquaintance with personal computers and banking machines will have broken their resistance to home banking, home shopping, and other transactional services. The general trend has been to abandon pure information retrieval, or "electronic publishing," in favour of transactional, communication, and entertainment services. Electronic bulletin boards and classified ads have been popular. And if all else has failed, videotex operators have opened their systems to providers of video games and various personal and "erotic" message services.[56]

Because the possible profits of unregulated communication and information services are so attractive, telecommunication companies are again attempting to use videotex to turn data communications into a mass market. But it is unclear whether they will be more successful in the 1990s than they were in the previous decade.

52. Aumente, pp. 15, 44–6, 51–74; Roy D. Bright, "Videotex: An Interactive Information System," *Telephony* (2 January 1984), pp. 90–1; Julian Hewett, "Viewdata is Maturing Rapidly in Europe," *Telephony* (23 April 1984), pp. 32–5; Larry Lannon, "Videotex: The Time Has Come to Produce Profits but Can It?" *Telephony* (7 May 1984), pp. 102, 104, 138; Bick Truet & Mark Hermann, "A Skeptic's View of Videotex," *Telephony* (10 July 1989), pp. 26–7; *Economist* (10 March 1990), pp. 28–9.
53. Aumente, pp. 16, 45; A. Jones, "Videotex: The Failure of the Publishing Concept in Britain?" *Videotex and the Press* (Oxford: Learned Information, 1982), p. 36; Ian Cunningham & John Raiswell, "Gateway to Online Information Services," *Telesis*, 10,1 (1983), pp. 2–7; Frederic Saunier, "The Public Network Goes On-Line," *Telephony* (3 April 1989), pp. 26–8, 37; George Doyle, "Strange Bedfellows?" *Telephony* (1 January 1990), pp. 24–6.
54. Aumente, p. 58.
55. *Economist* (10 March 1990), p. 26.
56. Aumente, pp. 70–3; Charles Mason, "Videotex Service Operators Push for Cheap Terminals," *Telephony* 11 September 1989, pp. 8–10; *Computing Canada* (5 January 1989); *Financial Post* (16 April 1990); *Globe & Mail* (5 January 1989), (19 September 1989), (1 May 1990); Ottawa *Citizen* (27 May 1988), (19 July 1989).

Digital PBX and Integrated Office Systems

While a mass market for personal data services was slow in emerging, the same was not true in the market for internal business communications. The same drive for office automation that produced local area networks also encouraged an alternative, the integrated voice/data office communication system. This was based on a new generation of private branch exchanges (PBXs) that emerged in the early 1980s and used time division switching and the latest microprocessors and memory devices. The new PBXs not only switched signals as digital pulse streams, making them inherently adaptable to both voice and data, but also could be programmed to offer a number of services to business users. In addition, memory and microprocessors were increasingly installed in telephones themselves, allowing them to store telephone numbers and to be programmed to request certain services from the PBX through use of the telephone keypad. An ordinary analogue telephone could be accommodated in these PBXs by installing an analogue/digital converter, or codec, at the point where its line input met the PBX. But the trend, as evidenced by the development of programmable telephones, was to transform the actual loop from the telephone to the PBX into a digital circuit by installing a codec in each telephone. By making each telephone a digital terminal, and each loop a digital line, it became possible to attach data terminals or personal computers directly to the telephone system without the need for modems. The logical step in this evolution was to design integrated voice/data terminals.

These systems offered a number of telephone features to the productivity-conscious office manager that were dependent either on software and memory-equipped telephones, processing and storage of special information at the PBX, or both. In general, these services were designed to reduce time spent trying to reach parties who were busy, unavailable, on another line, or at another phone. Such features included: hold, forward, or transfer of calls; central answering services with store and forward of voice messages; the ability to "camp on" a busy line with an automatic call back when the line was free; storage of frequently-used numbers and speed dialling by using a one or two digit code; redial of last number dialled; etc. In addition to reducing "telephone tag," these systems were intended to offer an alternative to group meetings through the use of conference calling. Finally, "intelligent" PBXs allowed the monitoring and control of outgoing telephone calls in order to, for example, limit long distance calling to authorized lines.

Digital PBX also offered several advantages to computer users who might have hesitated to invest in a separate local area network for internal data communications. The wiring cost of installing a LAN in an existing building could amount to 50 per cent of the total installation cost. With a PBX, wire pairs for telephones were already in place. Often, PBXs had back-up pairs for each pair in actual operation, which could be appropriated to attach microcomputers and peripheral devices. In addition, various techniques (discussed in the next section) could be used to carry voice and data on the same pair. Thus, the wiring cost of providing data switching over an existing PBX network could be much less than a LAN. A study in 1983 confirmed that for most applications where high data rates were not required, the cost of attaching to a PBX was significantly cheaper than to a LAN.

PBXs were inadequate for local networks requiring high data rates for such broadband features as high-resolution graphics, video, or the rapid transfer of large files. In the mid-1980s, PBXs had a maximum rate per terminal of 64,000 b/s (64 Kb/s), while LANs could accommodate terminals at several million b/s. But for most office applications with conventional data devices, 64,000 b/s would appear to the user accustomed to "high-speed" modems capable of 9,600 b/s (9.6 Kb/s) to be, as one expert noted, "equivalent to infinity." Another drawback to early digital PBXs was that if all paths through the switch were engaged, attempts at data transmission could be "blocked." Blocking usually occurred because a few data circuits could be occupied for hours, rather than the minutes assumed for telephone conversations. Because they were packet switched, LANs were non-blocking, though they could suffer from data delay during periods of high circuit demand. Some newer PBXs were designed to be non-blocking by providing enough excess capacity for both voice and data during peak periods. Others adopted packet switching for both voice and data, while still others were hybrid systems using packet switching for data and circuit switching for voice with both coordinated by shared processors.

Largely for reasons of cost and least disruption, and because the voice network was still more critical than the data network, many organizations in the 1980s found it preferable to upgrade their existing telephone system by installing a digital PBX with a data switching capability rather than to build a special purpose LAN. Other organizations requiring a greater data capability installed both a LAN for data and a PBX for voice. In some of these dual networks, interfaces between the two were arranged to use the voice PBX for access to the public telephone network and to a shared pool of modems and protocol software necessary for transmission over the public network. It seemed likely that neither the PBX nor the LAN would be found ideal for all voice and data uses, and that while a digital PBX might provide for all the needs of some users, a LAN was still a useful option for specialized and/or high volume data users. Finally, as the decade closed, digital central office-based systems began to offer features comparable to both PBXs and LANs, and it appeared possible that the telephone network would prove flexible enough to handle both private and public integrated data and voice services. This concept increasingly became embodied in a new network architecture promoted by the telephone companies—the integrated services digital network (ISDN).[57]

Integrated Services Digital Network (ISDN)

By the 1980s, the telephone networks were well on their way to fully digitizing their switching systems, transmission lines between switching offices, and long distance networks. This had been undertaken largely to reduce operating costs and cheaply increase system capacity. But the capabilities of this digital technology, coupled with increasing demand for digital data transmission services, encouraged planning of an integrated digital network based on the public telephone network. In the 1980s these plans coalesced in ISDN, defined by the International Telegraph and Telephone Consultative Committee (CCITT) as a network evolving from the digital telephone system that would provide end-to-end digital connectivity for voice, data, and other services. Users plugged terminal apparatus into the system through a single, CCITT standard interface.

ISDN was intended to evolve gradually from the existing network. In addition to existing digital switching and transmission technology and standardized voice/data terminal apparatus, there were two important technical features fundamental to the first phase of ISDN development. The first was a means of digitizing and increasing the capacity of existing local loops to carry voice and data simultaneously. The second was a means of improving the signalling ability of the network in order to make it easier to provide new services from central computers.

Bringing digital connectivity into the customer premises via the local loop eliminates the need for modems and increases the circuit capacity from a maximum of 9,600 b/s with an advanced modem to the pulse code modulation standard voice rate of 64,000 b/s. This new increased capacity lowers the transmission time of a business letter to .3 seconds from 2 seconds, while the time required to send a facsimile of one page with a photo is cut to 9.8 seconds from more than one minute.

But because pulse code modulation requires two wires for one-way transmission, a method is required to allow two-way conversation or interactive data transmission without having to add a second wire pair to each customer terminal. Two techniques have been adopted. Time compression multiplexing (TCM), which has been exploited in integrated voice/data PBXs, doubles the capacity of the wire pair by storing and compressing each pulse into half its usual length, thus allowing twice as many pulses to be transmitted in the same period of time. Each end in the circuit then takes turns transmitting, according to a procedure nicknamed "ping-pong protocol." Instead of the standard 64 Kb/s PCM capacity of a digital loop, the effective capacity is now 128 Kb/s. The second, more recently developed, technique is echo cancellation with hybrid (ECH), which allows simultane-

57. Michael P. Ludlow, "An Office Communication System for the '80s," *Telephony* (7 June 1982), pp. 40–2; Rod Adkins, et al., "Displayphone: Telephone and Terminal Combine in a Compact Desk-top Unit," *Telesis* 9,4 (1982), pp. 2–7; Daniel A. Pitt, "Interaction Between Voice and Data Elements in a Local Area Network," *Telephony* (7 March 1983), pp. 40–2; Thomas B. Cross, "The Future of the PBX: Promises and Problems," *Telephony* (2 May 1983), pp. 84–5; Donald W. Barrett, "Integrated Office Systems - Fact or Fiction?" *Telephony* (6 February 1984), pp. 41–6; John Williamson, "Office Automation: PABXs and LANs," *Telephony* (24 September 1984), p. 52; Christopher Ellis, "The Integration of Voice and Data: The PABX as a LAN," *Telephony* (24 September 1984), pp. 60–62; M.A. Paradesi, "PBXs and Data Communications: What's Available in the Market," *Telephony* (1 October 1984), pp. 119, 121; Tracy S. Storer, "Voice/Data Integration in Small Offices," *Telephony* (8 October 1984), pp. 32, 36, 40; Stallings, pp. 70–2, 191–201; James Quarforth, "Centrex to the Rescue;" Carol Wilson, "Centrex II: The Telcos' Revenge," *Telephony* (17 July 1989), pp. 23-31; Bob Vinton, "WANs, MANs and CO LANs," *Telephony - Transmission Special* (November 1989), pp. 14–5.

ous rather than alternate two-way transmission on the same pair. To do this, a means must be found to distinguish a station's own signal from that of the other station when both are transmitting simultaneously. This is accomplished by having a digital signal processor at each transmitter "learn" to identify the characteristics of its own signal. The learned signal is subtracted from the received signal, leaving as the difference the signal transmitted by the other station. The CCITT has adopted TCM as the interim solution for ISDN. ECH, for whose functions reliable microchips were only developed in the 1980s, has been designated as the preferred technique in the long term.

CCITT has designated two standard transmission services to be offered over the ISDN. Basic rate access (BRA) on a single wire pair consists of two 64 Kb/s channels for voice and/or data and one 16 Kb/s "D" channel for packet data and signalling information. Primary rate access (PRA), designed for trunks connecting PBXs to the public network, provides 23 64 Kb/s voice/data channels and one 64 Kb/s "D" channel. In both cases, the D channel allows interconnection between the telephone network and packet switched data networks. But in addition, D channel supports out-of-band signalling, the second new technological pillar of ISDN.

Conventional telephony relies on in-band signalling; call set-up information is transmitted to switching centres as dial pulses or touch-tone signals over the same channel as the voice or data call. The circuit is occupied during the call-set up process and once the connection is made no signalling is possible. ISDN, on the other hand, uses out-of-band signalling, designated by CCITT as Signalling System No.7 (SS7) but also known as common channel signalling (CCS). CCS separates the signalling channel from the call channel and, by extension, the various ISDN service functions from the basic circuit switching function. Signals from subscribers wishing to place a call or to request a special service pass as messages over a functionally, but not always physically, separate high-speed network running to the telephone central office and between central offices.

CCS greatly decreases the time required for call set-up, increases routing flexibility, and allows signalling during a conversation or data transaction. In addition, it allows services, and the stored data required to implement them, to be handled by a conventional computer separate from the telephone switch and its controls. Because there is a separate signalling network, computers need not be located in every central office, and different tasks may be assigned to different computers. And to change a service or feature requires a simple reprogramming of the computer's software. In addition, large amounts of information about a customer—name, address, phone number, the services subscribed to, long distance credit account, etc.—may be kept in central databases and quickly accessed by the network operator or, in some cases described below, by a service subscriber.

Common channel signalling and end-to-end digital connectivity allow telephone companies to more easily and cheaply provide existing and new services to subscribers. The most attractive services appear to be for business users, most obviously: faster and cheaper transmission of information, transparent access to specialized data networks, a single loop for all communication services, and cheaper equipment due to universal standards. For those more willing to innovate, voice and data circuits for teleconferences will be available on demand without prior arrangement of a dedicated channel. ISDN could also enhance teleconferencing by offering the ability to easily pass data among the computer screens of all participants. As well, the capacity of ISDN circuits would be sufficient for a form of videoconferencing using slow-scan, or freeze-frame, television.

The enhanced data storage and retrieval abilities of ISDN allows the telephone companies to provide numerous services, many of which, apparently, have not yet been devised. They clearly perceive, however, that the applications of the information they hold can be profitable and nearly limitless. For example, customer records can be used not simply to validate a company's long distance calling card calls but also to answer any other credit card inquiries. This moreover, can be extended to the offering of credit itself—AT&T recently introduced its Universal card, a combination calling card and general credit card competing directly with the banks' offerings of VISA and Mastercard.

One particular service to business and residential users, called variously Custom Local Area Signalling System (CLASS) or Call Management Service (CMS), has proven controversial. In addition to offering such services available in PBXs as call return and automatic recall on busy lines, call management enables the subscriber to identify a caller before answering the telephone by displaying his or her number, and in some cases name, on a small screen. By punching a few buttons on an ISDN phone, subscribers may also return the last call placed or received, screen out all calls from particular numbers, and automatically record the numbers of annoying or abusive callers in a central computer. These services are currently being marketed as conveniences and as protections to privacy from anonymous or undesirable callers. Call management is an attempt to exploit several aspects of modern life that are felt intensely by many people—the lack of time, the absence of privacy, the contradictory desire for human contact and the strict control of one's social and business contacts, the fear of the malevolent stranger. But some, among them the American Civil Liberties Union, have perceived these same services as a threat to privacy, as the caller loses his or her anonymity and the security of an unknown location. Callers to stores might find sales people returning their calls for unwanted sales pitches or find their phone numbers placed on lists or in databases of interest to marketers or credit agencies. A 1990 court decision in Pennsylvania highlighted another disturbing use to which caller identification might be put: the tracking down of abused women by the abusers they have fled.

But rather than specific fears, a good deal of opposition to call management is based on a general social unease about the use of computers and the telecommunications network for the increasingly efficient collection, storage, and retrieval of information about individuals. Impersonal, computerized credit and security checks have long since replaced trust in good character and public reputations. Automated banking and debit cards, which depend on knowledge of a person's personal finances, are promoted by banks to replace cash, whose value is independent of the identity and wealth of the bearer. Mailing lists, customer records, and consumer profiles are bought and sold by corporations for sales pitches that are increasingly focused on small target groups and individuals. Social insurance numbers are used by governments and banks for identification and access to diverse personal files. Though the motives behind this trend are generally pecuniary or bureaucratic rather than authoritarian, it is nevertheless disturbing that private life seems increasingly tied into the use of personal information for profit or for the smoother operation of the State apparatus. Opposition to the electronic invasion of another area, confidential telephone communication, should surprise no one.

Ironically, one of the reasons that call management has emerged as a prominent ISDN service is the seeming absence of any broadly-based demand for sophisticated telecommunications services. The telephone companies have yet to find a service with wide enough appeal to recoup, or "prove-in," the expense of introducing ISDN throughout the telephone system. The early promise of universal ISDN as a "nostrum for the communication ills of users" seems to be disappearing, and telephone companies are casting about for applications that will appeal to enough users to justify it. In Europe, where ISDN has been promoted heavily by national telephone monopolies, hopes for a residential market have been abandoned and energies directed at recruiting business and professional users.[58]

This highlights a growing trend, the emergence of specialized networks serving limited user groups. The only universal network is the one that serves residential users, while the other networks, overlaid on the physical plant of the first, serve

58. Robert Kenedi, "Plotting a Strategy for the Emerging ISDN," *Telephony* (22 June 1981), pp. 22–3; Felts, et al., pp. 43, 45–6; Irwin Dorros, "Shaping the ISDN: The Public Network's Role," and Del Myers, "The Cloudy Future for the ISDN," *Telephony* (24 October 1983), pp. 38–40; John H. Farrell, "The Reality of ISDN," *Telephony* (9 July 1984), pp. 60, 64, 66; Matthew J.J. Vea, "Moving Toward CCITT No. 7 Common Channel Signalling," *Telephony* (9 July 1984), pp. 68–72; Michael Shepperd & W.S. Szeden, "What's it Going to be – TCM or ECH?" *Telephony* (13 June 1988), pp. 31–2; Larry Lannon, "Carriers Drive Network Toward SS7 Services in '88," and Michael Warr, "The Next Step in SS7," *Telephony Transmission Special* (October 1988), pp. 11, 24–5; Jean-Claude Pennanec'h, "French Impressions of ISDN," *Telephony* (26 June 1989), pp. 42–4, 48–9; John Williamson, "European ISDN: Problems and Promises," *Telephony* (23 January 1989), pp. 30–1; Carol Wilson, "SS7 ... on the Fast Track," *Telephony* (20 November 1989), pp. 49, 52–4; Carol Wilson, "What's in a Calling Name? More Privacy," *Telephony* (26 March 1990), p. 3; Paul Travis, "Banks Attack AT&T's Discount on Card Calls," *Telephony* (21 May 1990), p. 3; Paul Connolly et al., "Network Services in the Age of Common Channel Signalling," *Telesis*, 15,3 (1988), pp. 5–7, 10; Cho Lun Wong & Rob Wood, "Implementation of ISDN," *Telesis*, 15,3 (1988), pp. 5–8, 12; John S. Ryan, "ISDN Goes Commercial," *IEEE Spectrum* (February 1990), pp. 32–3; R. Liebscher, "Data Transport Services on ISDN," *Electrical Communication* 62,2 (1988), pp. 128–33; *Economist* (10 March 1990), pp. 25, 27; Roy Leahy, "Demystifying ISDN," *Computerdata* (November 1988), p. 36 and "ISDN: Potential Appears Endless," *Computerdata* (December 1988), p. 25; Ottawa *Citizen* (23 May 1990, 1 June 1990, 14 June 1990); *Globe & Mail* (7 June 1990).

special business users. The demand for capacity and services is coming not from residential users, who are understandably quite content with their technological lot, but from businesses increasingly obsessed with the rapid movement of information and with the use of network intelligence for privacy, convenience, and information gathering. The attempt to attract residential users to ISDN reflects a desire to spread the costs of these networks over a broader market, thus lowering the cost to suppliers and to their primary consumers—business.

Broadband ISDN

A similar situation is likely to emerge if telephone companies attempt, as they project, to convert local loops to optical fibre and offer a form of broadband ISDN. The prospects for such a system have increased since the establishment in the United States of standards for a single "synchronous optical network," or SONET, adapted for multiple users, equipment suppliers, and service vendors. Northern Telecom, AT&T and other suppliers have announced plans to supply switching equipment capable of handling the bit rates possible with fibre, rates set by SONET at between 51.48 Mb/s and 13.22 Gb/s. This amounts to a huge increase in transmission capacity, from North American ISDN's basic rate of 144,000 b/s and primary rate of 1.54 million b/s to SONET's range of 51.48 million b/s to 13.22 billion b/s. Network planners hope to find a ready business market for broadband ISDN (BISDN) by connecting high-speed LANs for data transmission and by enabling videoconferencing as a substitute for business travel. But unless a popular residential service can be found to justify the fibre loop, broadband ISDN will be no more universal than the emerging ISDN.

Attempts to promote pay-per-view home movies and high definition television (HDTV) reflect the chicken-and-egg problem of quickly converting the telecommunication network to a new technology. The telephone companies are competing with private networks for data traffic and with other entrepreneurs for potentially profitable new communication fields. In order to reduce operating costs, they prefer to deal with universal networks and to maintain technological compatibility, if not uniformity, in the network. A universal, public broadband network, where services are purchased from a supplier as needed, seems like a cost-effective alternative to maintaining a private network, where a company pays for facilities and whether or not they are fully used, or to a special purpose network, where only a few clients bear the costs. But this assumes the costs of the public network are spread over many users. This would be the case in a universal system, if universally used, but the cost of constructing such a network from scratch is astronomical. One estimate of the cost of extending BISDN to just half of American homes is $100 billion. In the long term, it might be economically feasible to convert all local loops to fibre. But for the next few decades, in the absence of significant demand for new services, telephone companies might have difficulty justifying the replacement of existing copper loops in residential areas.

If broadband ISDN is to become a universal network sooner than this, the residential market must be exploited, either by creating new demand or by invading existing communication markets. As there is little residential demand for high speed transmission of facsimile, computer data, or videophone signals, telephone companies are pinning their hopes on home entertainment—the demand for diversion appearing to be limitless. The services will likely take the form of video games and, more importantly, an enhancement and/or appropriation of television service, the largest single user of transmission capacity into the home. This will have to involve either introducing commercially unproven HDTV or digital television, perhaps in competition with video movie rentals, or taking over the role of cable television companies. It is not yet apparent that consumers are willing to pay for expensive, advanced television receivers. And it is clear that both the cable television companies and government regulators, at least in Canada and the United States, oppose any attempts by the telephone companies to offer residential television transmission. Barring a change in consumer habits or in regulatory policy, therefore, broadband ISDN will not soon become a universal network.[59]

59. Ronald L. Weindruch, "Thinking About the Post-ONA Environment," *Telephony* (8 February 1988), p. 34; Larry Lannon, "Bandwidth Challenge," *Telephony - Transmission Special* (October 1988), pp. 49, 51; Jim Rice, "A Fiber Optic Broadband ISDN Network? It's up to Congress," *Telephony* 20 February 1989), pp. 32-4, 38, 40; David L. Wenner, "Are You Ready for Residential Broadband?" *Telephony* (22 May 1989), pp. 84-6; Lawrence K. Vanston, et al., "How Fast is New Technology Coming?" *Telephony* (18 September 1989), pp. 50-2; "Northern Telecom Jumps on Sonet Bandwagon," *Telephony* (16 October 1989), p. 8; Anthony Lavia, "The Metamorphosis of the Network," *Telephony* (25 December 1989), pp. 28, 30; *Economist* (10 March 1990), p. 29; *Globe & Mail* (18 June 1990).

5 Information Technology and Society: Do Liberties of Action Make Technologies of Freedom?

One social scientist has called the opportunities presented by a new technology "liberties of action." By this he refers to the new power a technology gives to overcome constraints on human action. Communication technology, for example, grants certain powers of mobility by allowing physical separation while maintaining social contact. But the term liberties of action can also have a political connotation, for the notions of both liberty and action imply human choices in the making and remaking of our world. This suggests that a technology can be examined not just to assess its technical merits, its ability to fulfil the technical goals set for it, but also to determine its proper place in society.

Any discussion of the past, present, or future impact of a technology leads to a number of questions about the purported power of technology over people and of people over technology. Some wonder how possible it is to control a new tool, once unleashed, and ask if a new tool can change society in unintended, possibly undesired, ways. Others choose to look not so much at relations between people and their tools, but at relations among people themselves. Accepting as given the new liberties of action, they tend to examine who controls the development and diffusion of new technologies, and who they are designed to benefit.

The "information revolution," if it does in fact exist, is still unfolding, and it is difficult to write a history of the social change that might be accompanying it. Instead, this section offers a summary of various points of view on the likely direction of change. Technologists, sociologists, and historians do not agree on the social meaning and importance of the growing use of computers and the increasing sophistication of data communications systems. Some envision a technology-driven electronic utopia of small communities or an efficient market of high-tech entrepreneurs linked into a broad communication network. Others, equally impressed by the power of information technology, write of an emerging post-industrial or information revolution that will bring both benefits and problems for society. Others are more sceptical about both the ability of a technology to single-handedly transform a society and of the motives behind the application of information technology within our own.

The powers offered by information technology, to organize vast amounts of information and convey it instantly over great distances, have been embraced by some who believe that the technology might transform a society. This tendency can be found in libertarian advocates of modern information technology; the liberty of action offered by a device or technique makes it a "technology of freedom." American political scientist Ithiel de Sola Pool believes that modern information technology will, if left to realize its inherent nature, provide an information pluralism that will contribute to a free society. Individual freedom is best served, Pool believes, by a "free marketplace of ideas" within which individuals are free to choose and express their beliefs. Fundamental to this marketplace is the widest possible access to information and communication resources.[60]

Paradoxically, for a theory grounded on notions of human freedom, Pool subscribes to a "soft technological determinism" about the historical interaction between communication technology and free speech. "Freedom is fostered," according to Pool, "when the means of communication are dispersed, decentralized, and easily available," while "central control is more likely when the means of communication are concentrated, monopolized, and scarce."[61] Due to their relative ubiquity, printing presses and microcomputers are thus technologies of freedom, while data networks oriented on central computers are not. Rather than a hierarchical communication system focused on a central, electronic "brain," Pool suggests modern information technology "promotes a trend toward

60. Ithiel de Sola Pool, *Technologies of Freedom* (Cambridge, Mass.: Harvard University Press, 1983), p. 219.
61. Pool, p. 5.

distributed processing throughout the system and against a central brain." A corporation, for example, will disperse processing among several nodes in order to save transmission costs to a central processor. This would give each node a degree of autonomy from the head office. Furthermore, the introduction of relatively cheap microcomputers in the 1970s and a decline in the cost of data transmission to levels approaching television, radio and newspapers will broaden the opportunities for free public expression. Pool envisions terminals in homes and offices allowing people to electronically exchange educational materials, news, and ideas. The system could, in fact, become a medium for large-scale discussion and debate.[62]

Pool is primarily concerned with legislative or regulatory restrictions on the ownership and use of information technology and the information it provides. Policy, according to Pool, has not kept pace with advances in technology. Government regulation of telecommunications was once justified by the belief in a scarcity of frequencies in the electromagnetic spectrum, or by the need to ensure unrestricted access to a natural monopoly like the telephone network. Pool believes that digital technology now offers, through cable television networks and the prospect of fibre optic lines into the home and office, ample capacity for a relatively free market to operate. While the need for universal connectivity and economies of scale might dictate a single physical network, government should do no more than ensure nondiscriminatory access to all who might wish to buy time or spectrum space in the system. Like publishing, which was early granted constitutional protection in the United States, anyone may gain access by "reasonable sacrifice and effort." Competition, fuelling continued improvements in technology to provide more capacity, would prevent the monopolization of resources by the highest bidder and ensure some access by the poor.[63]

Pool assumes that a free market economic system is the most efficient and least authoritarian means of distributing resources within a society. But economic power, concentrated in a few hands, is quite clearly a form of political power. This is likely to be even more pronounced in the control of information. Though all might theoretically be allowed access to information resources, the current degree of corporate concentration even in the supposedly accessible publishing industry suggests that control falls to the few who can afford to buy it. This gives them disproportionate influence over the political decisions made by any society.

While Pool argues for essentially a deregulated version of the status quo, others see not simply libertarian tendencies but revolutionary potential in information technology. Appealing to an audience in the British "alternative technology" movement, John Garrett and Geoff Wright argue that information technology makes possible a "Golden Age" of small communities, popular democracy, and small-scale, self-managed production. Microprocessors will allow machines to be continually reprogrammed to produce small batches of varied products in worker-controlled workshops. In place of central computers and hierarchical systems leading to them, new network architectures will use parallel or ring structures or switched systems to eliminate central control. All homes and offices will be linked by an interactive teletext network that, if it contained "all available economic, social and political information," would "make possible decentralized, democratic, economic management on the feedback principle." The communications network would, in fact, provide a "very fast and efficient decision-making system" that would include everyone, not just a few elected representatives. These technical means are available, Garrett and Wright argue, but their potential cannot be "fully exploited in the industrial societies we live in." The creation of a new system will require a political decision.[64]

Pool and Garrett and Wright perceive technology as a tool arriving from science fully-formed, its function inherent in its design and physical properties. All that remains for society is to take the tool and exploit its potential. While they acknowledge that political decisions will determine the way that technology is implemented, they are far more enamoured with the technology itself than concerned with the complexities of political power and the tensions of social change. In addition, they choose not to critically examine the type of society that has produced the technology and the clues this might give about its likely future uses.

Less sanguine than the utopians are liberal sociologists in the tradition of the pioneers Max Weber and Emile Durkheim. For them, technology

62. Pool, pp. 98–100, 229–230.
63. Pool, pp. 5–8, 223–4, 234–40, 246.
64. John Garrett and Geoff Wright, "Micro is Beautiful," in *The Microelectronics Revolution*, Tom Forester, ed. (Oxford, U.K.: Blackwell, 1980), p. 490–6.

is inextricably bound up with the society that both created and was formed by it. The historical watershed in this analysis is the Industrial Revolution, which saw the transformation of society from one based on agriculture, hand craft, and an undeveloped division of labour to one based on factory production, rationality, mechanical technique, and a detailed division of labour. The institutional model of this rationalistic, specializing society was the impersonal bureaucracy. The culture that emerged with this society evolved from one built on religion, tradition, and the close-knit rural community, to one built on individualism, social interaction free of traditional constraints, and urbanization. To the political system, responsible for maintaining order according to social consensus on the proper and most effective means of governing society, fell the task of reconciling the demands of various interests groups formed according to occupational, income, or cultural allegiances.

Technology occupies a central but slightly ambiguous position in this analysis. According to British sociologist Colin Cherry, a communications specialist steeped in the tradition of Durkheim and Weber, technology is more than a neutral tool, a material "artefact" or "thing." Its nature is essentially "active" and "social," with no meaning and no power outside the social context in which it is used. Though a particular artefact may offer certain "liberties of action" it does not determine its own use. Its power "arises from its possession by people who live within some particular social, legal, political, economic, framework which decides the mode of that possession."[65] Cherry argues that today, for example, technology is "virtually inseparable from the great institutions of industry." And though science is increasingly used in business, the decision to produce a new product is not a purely scientific one, but one based on company policy. It is an essentially political decision.[66]

The artifacts of technology, therefore, did not alone cause the Industrial Revolution. Nor will new forms of information or communication technology, Cherry writes, directly cause an "information" or "post-industrial" revolution. Political, economic, religious, and other factors will also play a role. But any technology clearly has a transformative role because it changes the way people feel about themselves and their world, thus altering the values that they use to manage and mould their society.

> *The sort of person you feel yourself to be depends very much upon the technology that you possess; if you have nothing but a stone axe you feel as one kind of person, but if you belong to [a] society which has computers, aircraft, cars, telephones and modern technology you feel quite different. You will feel about yourself and your rights, and about other people and their place in the world quite differently.*[67]

Communication technology is embedded in the highly organized structure of industrial society. It is, in fact, an essential tool of organization in a specialized, fragmented, but essentially interdependent social system. A new technology of communication and organization can thus be expected to have a major social impact. But Cherry believes it is difficult to anticipate the future use, significance, or value of a technology when it is first introduced. When the telephone was first introduced, few could conceive of it as anything but a means of broadcasting public entertainments or as a substitute for existing techniques—for speaking tubes and buzzers in large offices or bell-pulls for summoning domestic servants. More important than the invention of the telephone, Cherry argues, was the development of the exchange linking all telephone users into a single network. This was revolutionary because, for the first time, people were offered a choice of who they might communicate with and where they might wish to be when doing so. This has supposedly favoured "democratic social relations," unlike broadcasting, which involves an "authoritative centre" transmitting selected information to a relatively captive audience who cannot respond individually.[68]

Because the telephone transmits human speech, with all the values and nuance that it conveys, Cherry believes the telephone will never be supplanted by new inter-active data networks like videotex. But these systems, in their bi-lateral nature, will have a democratizing tendency similar to that of the telephone network.[69] Unlike early computer networks, which were built around large expensive computers, "remote and shrouded in mystery," new information systems are cheap, simple to operate, and give the individual a weapon against experts, business, or government who "threaten him by virtue of their organized 'bureau-

65. Colin Cherry, *The Age of Access: Information Technology and Social Revolution* (London: Croom Helm, 1985), p. 79.
66. Cherry, p. 23.
67. Cherry, p. 37.
68. Cherry, p. 64.
69. Cherry, p. 86.

cratic' powers for dealing with very complex 'rational' matters."[70] They offer individuals access, at least in principle, to information and to each other. Also like the telephone, they allow for interaction and coordination at the same time that they allow dispersal. The result might well be a more decentralized society, in terms both of geography and power. Cherry is clear, however, that this outcome is subject to political decisions. Telecommunication and computer networks:

> *...can assist centralisation by enabling a central authority to obtain information from, and to direct the activities of dispersed units, or, on the contrary, they may encourage decentralisation by offering the various organs of industry, business and government, etc. the liberty of dispersal, with the choice of operating in concert.*[71]

Cherry believes that the trend today is toward decentralization, but argues that this is conditional on the means that society determines for distributing information resources.

Describing information as a common resource, Cherry takes the opposite approach to de Sola Pool's free market treatment of information as a commodity. By communicating information to another person, an individual does not lose information but shares it. Thus information and knowledge are not, by nature, commodities to be bought and sold. They can only be made so by "legal and man-made devices." "The economic principles of operating all our various communication and information services," Cherry writes, "must differ radically from the economic principles of trade and exchange of goods."[72]

Cherry demonstrates that technology is a social process that is intimately tied to the ways that humans have arranged their economic and political institutions and pinpoints the liberties of action that a technology might offer. Such a view allows for human initiative in directing the evolution of technology, leading us to believe we can control information technology and use it to create a better society. According to Cherry, information technology can be used to either centralize or decentralize power in society. Control over the direction of technological change clearly lies in political decisions and not in the technology itself. But if this is so, Cherry appears little concerned with the way that political decisions on technology are made in our society. He alludes to the close relationship between technology and "the great institutions of industry," and calls business decisions essentially political. But he does not pursue this issue. Rather than analyze who controls the objectives and deployment of new technology and predict what impact this might have on its future development, he seemingly trusts in the political system and its dubious ability to override the decisions of business.

This weakness is also apparent in the theorists of the information society, or post-industrial society, who are less willing than Cherry to admit that technology is subject to social control before, as well as after the fact. In its extreme form, this analysis borders on fatalism and technological determinism. American sociologist Daniel Bell identifies several phenomena in the post-industrial society that he believes emerged in the United States in the 1950s and 1960s. The first, which is based on statistical research in 1960s by Fritz Machlup and Marc Porat, is a shift from a goods-producing to a service economy. In a post-industrial society, the majority of the population are engaged in services: health-care, education, social services, or "professional services" like design, systems analysis, computer programming and information processing. In conjunction with this employment trend has been the rise in new science-based industries and the ascendancy of a class of professionals and technical experts. The rise of this class reflects the fact that "theoretical knowledge" has become the "strategic resource and transforming agent" in society. In the place of old rules of thumb and intuitive judgements, decision makers increasingly use systems analysis, computer models, and well-defined problem-solving rules (algorithms) to manage the "organized complexity" of this new society.[73]

Central to Bell's analysis is an analytical separation of social reality into largely independent realms that operate according to their own principles. Bell perceives the major changes in our society taking place within the "social structure," consisting essentially of the economy and technology. This realm operates by its own principles of "functional rationality and efficiency." Conflict erupts when the principles of the social structure meet those of interest groups steeped in a "culture" that emphasizes equality, self-actualization, and

70. Cherry, p. 68.
71. Cherry, p. 69.
72. Cherry, p. 47.

73. Daniel Bell, *The Coming of Post-Industrial Society* (New York: Basic Books, 1973), pp. 14–33.

self-indulgence. It is left to the "political system," the third realm in Bell's model of society, to mediate the conflicts and arrive at new policies that balance the logic of the social structure with the aspirations inspired by the culture. Government will be forced to take on more of a planning role in society, but will be increasingly hard-pressed to reconcile demands for social equality and popular democracy with a social structure that requires specialized theoretical knowledge and creates privileged technical elites. The decisions will be essentially moral, a matter of weighing questions of social welfare and public good against questions of efficiency and private good. Once the moral decisions are made, government must use technical means to achieve its ends.[74]

New communications technology is a complicating factor in the process of "social management." Writing in 1979, Bell called the rapidly merging technologies of telephone, television, and data transmission "the really major social change of the next two decades."[75] Earlier in the 1970s, Bell had already identified two attributes of electronic communications that hindered the ability of government to operate. The first was the "eclipse of distance," which, while allowing rapid diffusion of ideas and the immediate, direct response to social issues, also reduced the "insulating space" between an isolated regional conflict and the nation and the central government. The immediacy of television, and to some extent other media, brings "emotional issues" to the national stage, permitting "the setting off of chain reactions which may be disruptive of civil politics and reasoned debate."[76]

In addition to the type of information that it may transmit, communications technology also has increased the sheer volume of information that must be digested. This "communications overload" leads to "enormous problems" for society and the political process. Individuals will have great difficulty in finding privacy or "open spaces," as a "relief from the stresses created by these incessant 'messages' out of the blue." Politically, Bell observes the "terrifying" number of problems that flow automatically to the central government for resolution in "real time" and asks, "can such a system continue without breakdown?"[77] Bell offers no solutions to the psychological problems of communications overload. Politically, he advises government to stand above conflicts and reconcile the demands of interest groups by the increasing use of experts and theoretical know-how.

The problems of the emerging society, Bell believes, will be not technical, but political. The main policy question is the type of "infrastructure" that will evolve from the merging of computers and communications. At the economic level, governments must, in order to encourage further increases in productivity, ensure the "more efficient distribution of necessary knowledge" for research and development. By this he envisages the creation of a national network linking universities, research labs, and other technical and scientific centres and the automation of data banks to allow for direct retrieval of information at a distance.[78]

The second infrastructure policy question is broader and more vexatious. First of all, Bell acknowledges that control over information is a source of power and that "access to communication is a condition of freedom." Secondly, he observes an explosion in the volume of news, statistical data, and other information that will require new means of access to make them useful to people. In deciding how to operate any integrated communication system for this purpose, government will have to determine if it should be unregulated, like computers and the print media, or regulated, like telephones and the electronic media. This will involve deciding which type of regime is best designed to be efficient, meet consumer needs, and allow for continuing technological improvement.[79]

Unlike Pool, Bell is not specific about the extent of government intervention in these matters. As the justification for government action, he has relied on Adam Smith's conception of public goods. These are goods that are beneficial to all society but which no one enterprise is willing or able to profitably produce. A communication infrastructure may fall into this category. But at the same time, Bell defines the role of government planning to be not to "direct" society but to "facilitate desired social changes."[80] And he believes a market system is superior in "economic efficiency and technological

74. Bell, *Coming of Post-Industrial Society*, pp. 12–13, 337, 345, 358–67, 477, 480, 487.
75. Daniel Bell, "The Social Framework of the Information Society," in *The Microelectronics Revolution*, Tom Forester, ed. (Oxford: Blackwell, 1980), p. 533.
76. Bell, *Coming of Post-Industrial Society*, p. 316.
77. Bell, *Coming of Post-Industrial Society*, p. 317.
78. Bell, "Social Framework," pp. 528–31, 535–7.
79. Bell, "Social Framework," pp. 515, 533–4.
80. Bell, *Coming of Post-Industrial Society*, pp. 302–4, 313.

responsiveness" to one based on state control. But he also perceives a limited role for government, as social equity may require subsidies to those who, because of low incomes and social discrimination, cannot afford to pay market prices. Likewise, Bell is aware that "institutional restraints" will be needed to prevent abuse of information and invasion of privacy by government agencies. For with the merging of computer data banks with rapid communications, "police and political surveillance of individuals is much more possible and pervasive."[81]

While Bell does place importance on political options, his definition of what is political is so narrow that the options themselves are quite limited. He restricts all debate and conflict over technology to the political realm, separate from the economic realm that arguably is the source of technological innovation, and conflict, in our society. For Bell, however, economy and technology are inherently rational and efficient; hence there is no conflict here, only the search for greater rationality and efficiency. The only remaining political question is that of establishing the infrastructure for the fairest and most efficient use of what is already there. But other, equally real, political decisions, to produce a certain product or expand a business and hence use more communication resources, have already been made by business.

Though Bell would argue the contrary, this position comes close to a form of technological determinism. In response to this criticism, Bell has emphasized that "a change in the techno-economic order (...the realm of information) does not determine changes in the political and cultural realms of society but poses questions to which society must respond." The "societal effects" of the expansion of the communications system have "consequences [that] must be understood in order for a society to make intelligent policy choices."[82] But if the economy and technology operate according to principles of rationality and efficiency, then the introduction of a new technology is inevitable, provided it meets these criteria. Short of changing the economic rules, the only remaining political option is to respond to the social consequences of its introduction.

The logic of this argument is seen in a 1980 report prepared for the Canadian government by the Department of Communications. Entitled *The Information Revolution and its Implications for Canada*, its authors, Shirley Serafini and Michel Andrieu, subscribe to Bell's notion of the post-industrial society and to the idea that we may be living through a social transformation of similar magnitude to the Industrial Revolution. Serafini and Andrieu identify two major manifestations of this revolution. The first is a shift in occupational patterns in Canada similar to those in the United States. The second manifestation is in the rapid innovations in information technology over the past forty years. In this sense they rely much more than Bell on the technological nature of the revolution. The varied applications of information technology will, according to Serafini and Andrieu, increase productivity, provide new goods and services, make information more widely available, enrich jobs, and eliminate dangerous work. But the new technology will also "affect all aspects of our lives and raise serious economic, social, legal and political issues."[83]

To their credit, Serafini and Andrieu present a detailed list of issues that might arise from the information revolution. At the level of the technology itself, increasing reliance on computers and data networks entails greater risk of "economic and social disruptions" if the system fails due to natural disaster, espionage, crime, or terrorism. At the economic level, problems might include a centralization of economic decision-making, which might lead both to increased regional economic disparities and an erosion of national sovereignty in the face of increasingly autonomous multi-national corporations. Additional economic effects might include an increase in unemployment resulting from the introduction of labour-saving technologies and from the decline of traditional industries. Additional effects on individuals might include the use of information technology for surveillance and the collection and exchange of personal data. Control of information might become concentrated in fewer hands, and national culture could be submerged in the inundations of electronic information flowing across borders. Individuals might become "electronic hermits" who interact more with machines than with people.[84]

Despite these possibly severe social consequences, however, Serafini and Andrieu believe

81. Bell, "Social Framework," pp. 534, 541.
82. Bell, "A Reply to Weizenbaum," in *Microelectronics Revolution*, pp. 573–4.
83. Shirley Serafini & Michel Andrieu, *The Information Revolution and Its Implications for Canada* (Ottawa: Canada, Department of Communications, 1980), pp. 8–11, 21–2.
84. Serafini & Andrieu, pp. 27–41.

there is no turning back. Invoking both economic competition and a certain "technological imperative," they argue that "like the industrial revolution, the information revolution is unavoidable." The most that can be done is adapt to the new regime by retraining unemployed workers, ensuring some form of income equity between the unemployed and the skilled, employed elite, and passing laws protecting personal privacy. Beyond these modest techniques of adaptation to the inevitable, Serafini and Andrieu advise that "the objectives of public policy should be not to prevent the revolution from occurring, but rather to turn it to our advantage." The irresistible force of the revolution must be welcomed, competitive pressures must be met. The authors advise the Canadian government to encourage capital investment in new technology, arguing that "if we are to gain the enormous benefits offered by the new technologies, barriers to their diffusion must be overcome."[85]

This ethic of adaptation, where human beings must respond to, rather than control the introduction of technologies, was attacked in the 1950s by the French sociologist Jacques Ellul. Ellul called "technique" the supreme principle of modern society. Far more than a given method of doing things or as any particular machine or piece of hardware, technique is "the totality of methods rationally arrived at and having absolute efficiency...in every field of human activity." Technique is not "an isolated fact in society," as the liberal sociologists might argue, but is pervasive, "related to every factor in the life of modern man." At the core of technique is the rejection of spontaneity, custom, and traditional morals in favour of the conscious use of reason to determine the "one best means" of accomplishing a certain end. But over time, all sense of the ends of technique have been lost in the maze of sophisticated and powerful means that have been created. Integral parts of the system, humans now serve an autonomous, pervasive technique, rather than make technique serve them.[86]

Ellul believed humanity was vaguely aware of this, but was as yet incapable of breaking out of technique. Most solutions have themselves been forms of technique, attempts to alleviate human anxiety about their world, adapting people to be more comfortable with technique. "Thinking machines" (computers) have been devised that allow humans to better communicate in the manner of the world of technique. And technique has been shaped to meet human physical and psychological needs, which have been identified in a numerical mechanistic way that makes them amenable to technique. But this process "despoils" humans of their "essence." Rather than fully autonomous and whole actors, they are reduced to the objects of technique.[87]

One step toward preventing Ellul's bleak vision from being realized is to reintroduce real people and the notion of free choice into the analysis of post-industrial society. The distinguishing feature of most writings on the post-industrial society, of which Bell's has the most scholarly renown, is its high level of abstraction. Individual human beings are invisible, and society is conceived as an arid model of separate categories with only limited and formalized connections between them. But in observing any particular society, it is not possible to separate the economic-technological realm from the political and the cultural. Individuals do not live their lives in isolated realms but in a totality. Their experiences at work and at home, their cultural memories and practices, and their political attitudes intermingle and together form impressions about their individual lives and their societies. The innumerable decisions of individuals, modified by the personal and historical circumstances in which they live, are the essence of any human society. New technologies are introduced by the decisions of individuals or groups who have an interest in introducing them. They do not invent and diffuse themselves.

In questioning the analysis of the information society theorists, we might also question the notion that information activities are somehow different from other activities in society, a premise that underlies the notion of a fundamentally different post-industrial society emerging on the grave of industrialism. In making his economic arguments, Bell identifies as shift in emphasis in post-industrial society from the application of manpower and machinery for the production of goods to the creation of information and theoretical knowledge. But even in industrial society, knowledge and information were central to production, being embodied in the experience of a skilled worker or the mechanisms of a machine, rather than in the theories or systems of the scientist and engineer. Furthermore, information in post-industrial society is not

85. Serafini & Andrieu, pp. 13, 33.
86. Jacques Ellul, *The Technological Society* (New York: Vintage Books, 1964), pp. xxv–xxvi, 4–6, 20–21, 430.
87. Ellul, p. 431.

an end in itself. The only reason it might appear so is that it is increasingly treated as a product that may be bought and sold. But the value of information remains its function as an aid to the production and sale of goods or to the organization of the system that performs these functions. And the production of goods has not ceased, though its location may have been separated from that of conception, control, and exchange. Information and knowledge cannot be abstracted from the human motives and human labour that are central to their creation. We might ask who sets the objectives for the production of information, who determines the conditions under which intellectual workers perform their labour, and who owns the product of their work. We might well find similar individuals and social groups to those who made decisions before the transition to a post-industrial society.

This approach, pursued by Canadian sociologists Vincent Mosco and Heather Menzies, examines not simply the process of production, but also the purpose of production. Information, once abstracted from the world in which is really functions, is placed back in its context. The creation and exchange of knowledge becomes not an isolated activity, a rational, technical issue, but also a political and cultural one. By questioning the ends of information activities, the means may also be examined more critically.

Rather than witnessing the transition from an industrial to a post-industrial society, Mosco argues that we are experiencing "a deepening and extending capitalism," as information is transformed from common property to private property through the use of information technology. Mosco criticizes Bell's conception of information as a "strategic resource" because it does not embrace the system of control or ownership of the resource that is characteristic of any society. The system that has defined our society has been the increasing primacy of private over common property. The post-industrial society analysis

> *...ignores what is arguably the most significant movement in society over the past four hundred years: the development, first, of widespread private property ownership and, more recently, the rise of national and transnational businesses that shape the production, distribution and use of resources, including information.*[88]

88. Vincent Mosco, *The Pay-per Society: Computers and Communication in the Information Age* (Toronto: Garamond Press, 1989), p. 25.

While Bell and Pool perceive a market approach to information as simply the most efficient policy for allocating increasingly necessary resources, Mosco sees the growing significance of information as inherent in the process of capital accumulation. Capitalism grows by expanding the marketplace, either "extensively" over more and more of the earth, or "intensively," by incorporating more and more activities into the marketplace that were once outside. Geographic expansion of the marketplace has been abetted by telecommunications, which overcomes space and time constraints that heretofore limited markets to a local, regional, or national domain.[89]

But it is the intensive enlargement of markets that Mosco focuses on, the transformation of information into a commodity. In the past, information has been difficult to define and quantify and consequently its exchange difficult to measure. These attributes are necessary for the creation of a commodity.[90] Print technology and copyright laws have always enabled some information to be "commodified," though this has always been undermined by public lending of books and by unauthorized copying. Digital telecommunications and data systems have made the process much more comprehensive.

> *The reduction of information to a common digital code and the ability to process and transmit this coded information instantaneously make it possible to measure information and monitor information transactions with considerable quantitative precision. These developments have augmented the possibilities for packaging and repackaging information in a marketable form.*[91]

Rather than being contained in a book on a library shelf, free for any member of the community, information might become only accessible, for a fee, through electronic connection to a private database.[92]

There are other components to Mosco's analysis of commodification. For example, information technology not only measures and packages information for sale. It collects information which allows the more rigorous measurement, packaging,

89. Mosco, p. 26.
90. Mosco, p. 26.
91. Mosco, pp. 26–7.
92. In a recent example of this process, a library at the University of Toronto removed most books from its shelves and replaced them with computer terminals linked to an online database. For access, students and business people will be charged for computer time and for each page copied. *Globe & Mail* (27 June 1990).

and delivery of audiences to product advertisers who pay communication carriers for audience attention. While previously, audience and market research relied on faulty written or spoken information provided by audiences, "people-meters" wired from residential televisions into a central computer can now record not only which programs are being watched, but also the age and sex of those watching. Information technology also elaborates the commodification of time and space. Under capitalism, time has been transformed from the religious, seasonal, or generational cycles of feudalism into a linear process that is increasingly more precisely measured, budgeted, and assigned monetary value. Finally, information lends new dimensions to the commodification of space. Pressure mounts for the sale of satellite orbits or electromagnetic spectrum to the highest bidder, and privately owned "teleports" are established to provide large-volume users with preferred line-of-sight locations for satellite earth stations and microwave towers.[93]

In recent years, the commodification of information has been played out politically through the debate over regulation of telecommunications. "Deregulation" is already well-advanced in the United States, where government policy has opened long distance telephone and data communications, as well as a large variety of information services, to competition. This was the result of pressure from large telecommunication users, mostly large national and international corporations with widespread operations, and from companies hoping to enter the lucrative information field. Deregulation (or rather "reregulation" according to different principles) is, according to Mosco, "recognition that telecommunications and its related informatics and communications sectors, have come to occupy a central place in the capital accumulation process."[94]

While deregulation has caused a decline in long distance rates, these have been at the expense of increases in local rates. This new pricing regime has harmed the interests of the poor, who use the telephone for local calls. In a more general sense, the movement to deregulate and privatize public information services, like public libraries, public schools, government postal services, and government-owned telephone systems, harms the interests of the poor and less powerful people in our society. Mosco favours information policy remaining in the domain of democratic, representative bodies guided by the same principles that established public education. They should ensure universal access at affordable rates to local and long distance telephone service, and to "basic information about health care, education and other community services" and offer opportunities to "respond electronically to verbal communication" and "signal for emergency services, information and other vital communications."[95]

The main attributes of telecommunications, its ability to overcome space and time, and that of computers, to receive, store, and execute coded instructions, are the basis for Mosco's second theme in the study of information technology and capital accumulation—control. Information technology allows the dispersal of corporate functions, either within a workplace, across town, across the country, or around the world, while retaining or enhancing their centralized control by management. While Bell and other communication theorists have also stressed the use of information as an instrument of control, Mosco's class analysis leads him to ask: "who or what controls whom or what"—and to what end?[96]

Mosco is sceptical of the claims of "technological democrats" like Pool who envision a technology-driven trend toward democracy. He argues that "forces at work in the economies of those nations that lead in computer development make it difficult to identify or foresee democratic tendencies." He cites, for example military sponsorship of advanced information technology research in the United States, research not intended for peaceful, democratic applications. More importantly for a country like Canada, are broader tendencies toward centralized planning and control, electronic monitoring of work and financial transactions, centralized information control, and the treatment of information as a commodity. Class divisions are likely to deepen as a gulf develops between the information rich, who can afford new information commodities, and the information poor.[97] Mosco is also alarmed by the possibilities for surveillance created by information technology. While not doubting that policies and technologies can be designed to ensure individual privacy, he is more concerned that information about group behaviour will be used by

93. Mosco, pp. 32–34.
94. Mosco, p. 103.
95. Mosco, p. 37.
96. Mosco, pp. 27–8.
97. Mosco, pp. 78–80.

governments and corporations to manage or manipulate, either by ensuring political popularity for certain policies, or by ensuring "stable growth in consumption patterns and a cooperative workforce."

> *Electronic communication and information systems—including those that measure and monitor phone transactions, bank deposits and withdrawls, credit or debit card purchases, keystroke counts in the workplace, etc.—gather massive amounts of information about the choices of large or small, amorphous or precisely defined, collectivities, for the purpose of more effectively managing and controlling their behaviour.*[98]

This pervasive group or social surveillance by those with power is a "threat to freedom, to a self-managed life, or to a life in which people choose their own form of collective management."[99]

While a communication system is central to a democratic political process, which depends on "widespread distribution of information and the ease of social contact," Mosco argues that technology itself will not lead to democracy. Such a society will not be built on a technology shaped by a powerful elite for undemocratic ends. Rather, "democracy grows from movements committed to its realization."[100]

Mosco examines one facet of our lives that unfolds in a profoundly undemocratic manner, our work, and examines two processes that information technology exacerbates. On the one hand, information technology facilitates the "direct exercise of coercive power" over workers. Managerial control of the production process and workers is increased by the elimination of jobs by technology and the shift of jobs from union to non-union positions. International telecommunications allows the shift of some jobs to low-wage, anti-union countries or regions while maintaining control at the head office. And "deskilling" of trade occupations through the use of programmed machines "takes the power of conception away from the individual worker and vests it in machines and management."[101]

Secondly, power is more subtly exercised by "controlling the values that people use to define themselves and their place in everyday life." Our expectations, our attitudes, our ways of seeing, talking, and thinking are influenced by the way we work. As more workers are made to work at electronic terminals, monitoring of their work becomes more possible and more likely. Some office workers already subject to electronic monitoring include word processors, data-entry clerks, telephone operators, customer service workers, telemarketing and other sales people, insurance claims clerks, mail clerks, and bank clerks. As rewards are linked to meeting numerical performance counts—for keystrokes per hour, or whatever—Mosco believes we might internalize the discipline of surveillance and reinterpret the purpose of our jobs according to the abstract, quantifiable standards of efficiency set by management.[102]

Heather Menzies takes up this theme, envisioning the "enclosure of people as individuals and as a whole society within a relentlessly programmed environment." This process derives from the installation of computer systems in workplaces and the ascendancy of the "universal techno-logic pervading them."[103] This logic, or "technical, quantifiable way of knowing" values systematism and economic efficiency and reduces individuals and groups of people to abstract categories within systems. The transformation of the workplace into a technological system is accompanied by the wide application of systems logic throughout society. Already,

> *The perception of technological change and economic restructuring has been transformed from a social and political question into a technical one; from explosive issues of massive unemployment, the demeaning of work, and the loss of democratic control and personal autonomy, into a de-personalized, disembodied technocratic puzzle: how to "manage" the "impacts" of restructuring; how to "adjust" people—as if they were numbers on a flow chart.*[104]

This will, Menzies fears, lead to an "inevitable point of closure when society at large will be transformed in the image of that logic and culture." The "most fundamental philosophical and ethical questions" will be reduced "to questions of efficient cost-effective management."[105] The depths of this process may be found in biotechnology, where technique is applied to the control and modification of life itself in order to achieve abstract goals of efficiency or productivity.

98. Mosco, p. 38.
99. Ibid.
100. Mosco, pp. 177, 81.
101. Mosco, pp. 115–9.
102. Mosco, pp. 115, 119–21.
103. Heather Menzies, *Fast Forward and out of Control: How Technology is Changing Your Life* (Toronto: Macmillan, 1989), p. ix.
104. Menzies, p. xi.
105. Menzies, p. xiii.

While Menzies's rhetoric reveals an affinity to Ellul, her research is grounded in an understanding of the recent applications of technology in our society. In this sense, she provides empirical support for Mosco's theories. In particular, she clearly connects information technology with the ongoing restructuring of the national and global economies and examines its impact on our social structure. In fact, she argues that it is "fruitless" to try to "disentangle the influence of technological change *per se* from economic restructuring."[106] This restructuring has been based on the rise of global corporations that use computers and the international telecommunications network to effect an international division of labour. In established industrial countries like Canada, technology is being used to reduce labour costs, boost the value added to new products, and create new goods and services with new markets. Labour-intensive, "dirty" work is relocated to poor Third World countries, while corporations retain knowledge-intensive functions like management, financial planning, and research and development at the home office. A global corporation may, in fact, be able to better manage and control time because it can rotate information-based activities through the telecommunications network around the clock and around the globe, taking advantage of changing time zones and perhaps lower wage rates for professional services in certain countries. Finally, telecommunications allows a global economy in information-based services and a financial system operating on instantaneous transactions that bear increasingly less connection to the actual production of goods.

In Canada, economic restructuring has been felt as a growing polarization of society into rich and poor, fully and marginally employed, information rich and information poor, the socially involved and the disenfranchised. There has been a breakdown in the post-World War II "social contract" that saw workers accept new technology in return for higher wages, higher economic growth and an increase in employment. In the 1980s, increased productivity has not brought these returns. There have been massive job losses in traditionally high-paying, unionized industries like mining, forestry, and manufacturing as a result of plant closures, corporate restructuring, and technological change. In addition, corporations have reduced middle-management "layers" and clerical positions by contracting out services or introducing automated systems that allow fewer people to do more work. Using government employment statistics, Menzies demonstrates that in the 1980s, the only net new job growth in Canada was in the service sector, with many of these jobs the most poorly paid in the country. In manufacturing, the only jobs created were at the very lowest and the very highest pay levels. In all sectors, the middle ground shrank. Increasingly, Menzies says we may speak of two Canadian economies: one global, highly technological, and oriented to growth and efficiency; and the other local, consisting of small businesses or non-profit community projects in support services, and subordinate to the global economy.[107]

Within industry, Menzies detects an evolving division of labour between those who have conceptual skills and the power to control their and others' work, and those who perform procedural work according to rigid rules set by someone else. This process, which has been going on since the 19th century, "has meant the abstraction of work from the larger context of its purpose or mission, and the separation of the work function from the creation or control function."[108] This process has been reinforced by microprocessors, as work procedures increasingly are embodied in a computer-controlled automated systems. Even in high technology industries, this division of labour exists, with the subordinate role taken by technicians who have a variety of skills in operating the latest automated equipment but who do not have the credentials or training to control the conception of their work.

While management theorists talk of increased worker involvement and job enrichment resulting from the introduction of new technology, the reality is that managerial prerogative to control conception of work is not challenged. And often, job enrichment consists merely of an enlargement of job duties, which might offer variety but no overall increase in skill or control. In fact, as procedural instructions are increasingly programmed into automated systems, workers have seen a decline in their discretionary powers on the job.

106. Menzies, p. 17.

107. Menzies, pp. 6–23, 27–9, 38, 122–3, 204–6.

108. Menzies, p. 138.

Where previously the head of the boss was separate from, and therefore not in full control of, the hands that ran the machines to make what the boss had ordered, now the head and the hands are joined electronically in a system that envelops the workers. Only a few people work with the system as if it were a tool—that is, free to conceive of what the system should be directed to do as well as how to do it. More people simply work for the system, as programmed as the robots they look after.[109]

The issue of power over information and conception might also in future be played out at a regional and national level. The increasing economic importance of information will mean that whoever can afford to control the means for gathering and processing information will hold great power. Conversely, those without access to information resources will be disadvantaged. "The time advantage" offered by information technology, writes Menzies, "lets those who use it occupy the future, while relegating others to a permanent catch-up position." As information becomes an increasingly critical resource, it becomes the subject of more and more intense competition, and there is an increasing danger of it falling under control of a few monopolistic companies. This might entail the loss of local, regional, and even national economic autonomy, as important information is held in a few remote hands. For Canada, the threat comes from the transfer of company data from Canadian subsidiaries to American parent companies.[110]

Like Mosco, Menzies rejects the notion that information technology, or any technology, is self-determining or "implicitly progressive." Rather, "technological change and economic restructuring" are processes that are "shaped by and serve the values and priorities of whoever controls them."[111] But she does distinguish between technology as a tool and technology as a system. A tool implies that the user may use the technology freely as one sees fit. A system, on the other hand, prescribes the legitimate uses of the technology, either through proprietary restrictions, secrecy, government regulation, or more subtly and often, through job descriptions and the division of labour. Initially, the use and application of a new idea are open, but as "certain specific uses become institutionalized" a system develops. At this stage, the technology becomes "almost deterministic as it embodies the choices and priorities of those who have designed, manufactured, and sold it." Once established, the system seems the only normal, rational use of the technology, and other options are forgotten or overlooked. At this point the current uses of the technology seem to have developed inevitably.[112]

Mosco provides an alternative to this despairing vision, arguing that the process of domination and hegemony through new technology creates contradictions that undermine the functioning of the system. First, increased managerial control, emphasis on quantitative productivity over product quality, and the elimination of skilled production workers is often achieved at the price of increased malfunctions and lost flexibility. While meeting short-term competitive pressures for low-cost production, this strategy causes long-term problems for a corporation. But by pursuing the alternative, the employment of skilled, knowledgeable workers, a corporation creates a powerful group able to oppose or undermine management's control of production. Secondly, the lower labour costs achieved by unemployment and deskilling will in the long run undermine mass consumption, which has been a foundation of modern capitalism. And a decline in social equality and middle-class living standards in countries like Canada, the United States and Britain might lead to a "crisis in legitimacy" and demands for a redistribution of social power and resources.[113] Opposition to the current uses of information technology will likely grow, writes Mosco, with the "accelerating gap...between the potential of the technology to democratize power and distribute economic gains more equitably, and the current drive merely to centralize control and profit."[114]

Anticipating this conflict, Menzies calls for a "redrafting of the technological-change and economic-restructuring agenda" to allow popular participation in defining new goals that go beyond the values of "cheaper" and "faster." Among her

109. Menzies, p. 148.
110. Menzies, pp. 61, 92–101, 224.
111. Menzies, p. xix.

112. Menzies, pp. 51–7.
113. Mosco, pp. 122–4. Perhaps one example of the contradictory and opposable nature of information technology is last year's decision by Bell Canada, under union pressure, to end "average work time" monitoring of most telephone operators. The company has also shifted management emphasis from speed to quality of service. Operators are reported to be more satisfied with their jobs as a result, though the company has not detected any increase in productivity. Reflecting its reluctance to give up control in the workplace, Bell Canada continues to monitor first-year operators and keeps average work time records on entire offices: *Globe and Mail*, 22 Feb. 1990.
114. Mosco, p. 124.

recommendations are worker participation and initiative in and for technological change that will enrich their work and further social, not just profit ends. At the local level, community groups and unions should work for local economic initiatives that can link into the national and global economies.[115]

> *The issue is not are you for or against technology ... [T]he issue is really what vision of society, what conception of progress, de we share against which the current program of restructuring and application of technology can be evaluated, judged, and where warranted, redesigned? To what premises, what priorities should the design and application of technology be accountable? In what values should economic restructuring be grounded?*[116]

Conclusion

This section has largely dealt with predictions and policy recommendations. In one sense, it is not yet possible to write an account of the social impact, or rather social history, of information technology. If nothing else, it is clear that this technology is still in the process of becoming. So too, and continually, is our society. But a historical perspective is exactly what is needed in the present situation. The main flaw with liberal sociologists studying information technology and post-industrial society has been their scant reference to the ways that information and its facilitating technologies have evolved in our own society in the recent past, including questions of who has used the new technology and for what ends. At worst, they have placed their faith in some decentralizing logic of the technology or in the virtues of free markets. At best, they have examined partial phenomena—for example, statistical changes in employment patterns, the relative importance of certain economic sectors within a particular national economy, the role of government in regulating a communication infrastructure.

But Bell's influential work, for one, was written before the massive economic changes of the 1980s. These changes cannot be separated from the new technologies that, disseminated to serve the needs of business, reached deeper into society in the decade.

We should look at the recent past for clues about the current tendencies of these technologies. In Canada, the early part of the 1980s was dominated by the worst recession since the 1930s, with unemployment exceeding 10 per cent for four years. This was followed by the longest period of economic growth since World War II. But by the end of the decade, unemployment remained at around 7 per cent and real incomes had shown almost no increase over 1980. In addition, the length of the average work week had increased, one estimate placing it at 20 per cent more in 1990 than fifteen years earlier. The Economic Council of Canada, meanwhile, has confirmed Menzies's conclusions about the polarization of Canadian society in the 1980s into low-paid service workers in temporary and part-time jobs and high-paid, skilled workers in stable jobs. The middle class has dwindled.[117] At the same time, the geographical distribution of wealth and economic power was increasingly centralized in Toronto and the surrounding region of southern Ontario.

Internationally, the economic integration of national economies into a single global economy had been accepted as conventional wisdom. Increasing international competition has led to the consolidation of firms into fewer, larger corporations with global capabilities. A United Nations study has found that just 600 companies, each with sales exceeding $1 billion, produce 25 per cent of all products in the world's market economies. In the United States and Britain, these large companies account for 80 to 90 per cent of all exports. Symbolic of this country's integration into an economic world largely outside its control, Canada signed a free trade agreement with the United States in 1989.[118]

Information technology has not caused these economic changes, but it has facilitated them by allowing businesses rapid transmission and interconnection around the globe. We have seen little evidence of any inherent decentralizing tendencies in the technology. Information technology may well offer certain liberties of action that may be used for human liberation. But at the moment, its current uses, and users, do not suggest this is happening.

The two opposing points of view examined in this section reflect a difference of opinion not only about information technology but about society. Both hold that information and communications simply facilitate other processes in society. Believ-

115. Menzies, pp. 245–64.
116. Menzies, p. 262.

117. *Globe and Mail* 6 Feb. 1990, 19 Feb. 1990, 15 Feb. 1990; Ottawa *Citizen*, 15 Feb. 1990; *Financial Post*, 10–12 Feb. 1990.
118. Martin Mittelstaedt, "Business Goes Global," *Report on Business Magazine* (February 1989).

ing that society is essentially a free association of individuals bound by mutual self-interest, a liberal will emphasize the role of information technology in facilitating a free exchange of ideas, goods, etc. A Marxist like Mosco, on the other hand, bases his analysis on a class conception of society. Under capitalism, information and communications are commodities and tools of control, exploited by business for the further accumulation of capital.

Some observers see information and communications as aids to free association, others see them as tools of power. The reality is that communication is both. As Cherry writes, we cannot foresee the uses to which a technology will eventually be put. And as Mosco argues, any society will have internal contradictions, forces that, while operating in one direction, also set up forces that work in the opposite. While one tendency may seem dominant at any one time, one cannot assume that it is permanent and immutable. It is therefore important to continually reconsider the uses and interests to which information technology might be turned. We must adopt a critical attitude toward any technology and insist that new technology respect common human ends, and not simply those of profit or efficiency. And we should understand that perhaps nothing can do this better than a society that values face to face contact, a sense of tradition, and a sense of limitation on human power over nature and humanity.

By emphasizing the historical nature of any inquiry into technology and society, I am rejecting attempts to predict the shape of a future society that might arise from the logic of any particular new technology. No one can predict the future because no one can see what does not yet exist. In any event, a democratic society does not assume we can predict the future. It assumes we create it.

National Museum of Science and Technology

Musée national des sciences et de la technologie

Canada