

Telecommunication Circuit Design, Second Edition. Patrick D. van der Puije
Copyright © 2002 John Wiley & Sons, Inc.
ISBNs: 0-471-41542-1 (Hardback); 0-471-22153-8 (Electronic)

Telecommunication Circuit Design

Telecommunication Circuit Design

Second Edition

Patrick D. van der Puije



A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.

Designations used by companies to distinguish their products are often claimed as trademarks. In all instances where John Wiley & Sons, Inc., is aware of a claim, the product names appear in initial capital or ALL CAPITAL LETTERS. Readers, however, should contact the appropriate companies for more complete information regarding trademarks and registration.

Copyright © 2002 by John Wiley & Sons, Inc., New York. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic or mechanical, including uploading, downloading, printing, decompiling, recording or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the Publisher. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0012, (212) 850-6011, fax (212) 850-6008, E-Mail: PERMREQ@WILEY.COM.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold with the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional person should be sought.

ISBN 0-471-22153-8

This title is also available in print as ISBN 0-471-41542-1.

For more information about Wiley products, visit our web site at www.Wiley.com.

CONTENTS

Preface	xiii
Chapter 1 The History of Telecommunications	1
1.1 Introduction	1
1.2 Telecommunication Before the Electric Telegraph	1
1.3 The Electric Telegraph	2
1.4 The Facsimile Machine	4
1.5 The Telephone	6
1.6 Radio	8
1.7 Television	9
1.8 The Growth of Bandwidth and the Digital Revolution	10
1.9 The Internet	11
1.10 The World Wide Web	13
References	15
Bibliography	16
Chapter 2 Amplitude Modulated Radio Transmitter	17
2.1 Introduction	17
2.2 Amplitude Modulation Theory	18
2.3 System Design	21
2.3.1 Crystal-Controlled Oscillator	22
2.3.2 Frequency Multiplier	22
2.3.3 Amplitude Modulator	22
2.3.4 Audio Amplifier	22
2.3.5 Radio-Frequency Power Amplifier	23
2.3.6 Antenna	23

2.4	Radio Transmitter Oscillator	23
2.4.1	Negative Conductance Oscillator	24
2.4.2	Classical Feedback Theory	26
2.4.3	Sinusoidal Oscillators	28
2.4.4	General Form of the Oscillator	28
2.4.5	Oscillator Design for Maximum Power Output	31
2.4.6	Crystal-Controlled Oscillator	35
2.5	Frequency Multiplier	37
2.5.1	Class-C Amplifier	40
2.5.2	Converting the Class-C Amplifier into a Frequency Multiplier	40
2.6	Modulator	47
2.6.1	Square-law Modulator	47
2.6.2	Direct Amplitude Modulation Amplifier	49
2.6.3	Four-Quadrant Analog Multiplier	51
2.7	Audio-Frequency Amplifier	54
2.7.1	Basic Device Characteristics	54
2.7.2	Class-A Amplifier	55
2.7.3	Class-B Amplifier	60
2.8	The Radio-Frequency Amplifier	67
2.9	The Antenna	71
2.9.1	Radiation Pattern of an Isolated Dipole	72
2.9.2	Monopole or Half-Dipole	72
2.9.3	Field Patterns for a Vertical Grounded Antenna	74
2.10	Classification of Amplitude Modulated Radio-Frequency Bands	75
	References	76
	Problems	76

Chapter 3 The Amplitude Modulated Radio Receiver **79**

3.1	Introduction	79
3.2	The Basic Receiver: System Design	79
3.3	The Superheterodyne Receiver: System Design	82
3.4	Components of the Superheterodyne Receiver	85
3.4.1	Receiver Antenna	85
3.4.2	Low-Power Radio-Frequency Amplifier	86
3.4.3	Frequency Changer or Mixer	90
3.4.4	Intermediate-Frequency Stage	99
3.4.5	Automatic Gain Control	101

3.4.6 Demodulator	103
3.4.7 Audio-Frequency Amplifier	105
3.4.8 Loudspeaker	105
3.5 Short-Wave Radio	107
References	108
Bibliography	108
Problems	108
Chapter 4 Frequency Modulated Radio Transmitter	111
4.1 Introduction	111
4.2 Frequency Modulation Theory	111
4.3 The Parameter Variation Method	115
4.3.1 Basic System Design	115
4.3.2 Automatic Frequency Control of the FM Generator	117
4.3.3 Component Design with Automatic Frequency Control	118
4.4 The Armstrong System	122
4.4.1 Practical Realization	125
4.4.2 Component Circuit Design	127
4.5 Stereophonic FM Transmission	137
4.5.1 System Design	137
Bibliography	138
Problems	139
Chapter 5 The Frequency Modulated Radio Receiver	143
5.1 Introduction	143
5.2 Component Design	146
5.2.1 Antenna	146
5.2.2 Radio-Frequency Amplifier	146
5.2.3 Local Oscillator	147
5.2.4 Frequency Changer	147
5.2.5 Intermediate-Frequency Stage	147
5.2.6 Amplitude Limiter	148
5.2.7 Frequency Discriminator	148
5.3 Stereophonic Frequency Modulated Reception	156
5.3.1 Synchronous Demodulation	158
5.3.2 Stereophonic Receiver Circuit	158
References	158
Problems	159

Chapter 6 The Television Transmitter	161
6.1 Introduction	161
6.2 System Design	162
6.3 Component Design	163
6.3.1 Camera Tube	163
6.3.2 Scanning System	167
6.3.3 Audio Frequency and FM Circuits	169
6.3.4 Video Amplifier	170
6.3.5 Radio-Frequency Circuits	179
6.3.6 Vestigial Sideband Filter	179
6.3.7 Antenna	179
6.3.8 Color Television	180
References	184
Bibliography	184
Problems	185
Chapter 7 The Television Receiver	187
7.1 Introduction	187
7.2 Component Design	189
7.2.1 Antenna	189
7.2.2 Superheterodyne Section	189
7.2.3 Intermediate-Frequency Amplifier	189
7.2.4 Video Detector	190
7.2.5 The Video Amplifier	191
7.2.6 The Audio Channel	191
7.2.7 Electron Beam Control Subsystem	191
7.2.8 Picture Tube	202
7.3 Color Television Receiver	203
7.3.1 Demodulation and Matrixing	203
7.3.2 Component Circuit Design	205
7.3.3 Color Picture Tube	207
7.4 High-Definition Television (HDTV)	209
Bibliography	210
Problems	210
Chapter 8 The Telephone Network	213
8.1 Introduction	213
8.2 Technical Organization	213
8.3 Basic Telephone Equipment	216

8.3.1	Carbon Microphone	216
8.3.2	Moving-Iron Telephone Receiver	218
8.3.3	Local Battery – Central Power Supply	220
8.3.4	Signalling System	220
8.3.5	The Telephone Line	221
8.3.6	Performance Improvements	222
8.3.7	Telephone Component Variation	224
8.4	Electronic Telephone	224
8.4.1	Microphones	225
8.4.2	Receiver	225
8.4.3	Hybrid	226
8.4.4	Tone Ringer	229
8.4.5	Tone Dial	230
8.5	Digital Telephone	243
8.5.1	The Codec	244
8.6	The Central Office	253
8.6.1	Manual Office	253
8.6.2	Basics of Step-by-Step Switching	253
8.6.3	The Strowger Switch	256
8.6.4	Basics of Crossbar Switching	257
8.6.5	Central Office Tone Receiver	259
8.6.6	Elements of Electronic Switching	261
	References	263
	Problems	263
Chapter 9	Signal Processing in the Telephone System	267
9.1	Introduction	267
9.2	Frequency Division Multiplex (FDM)	267
9.2.1	Generation of Single-Sideband Signals	268
9.2.2	Design of Circuit Components	269
9.2.3	Formation of a Basic Group	269
9.2.4	Formation of a Basic Supergroup	270
9.2.5	Formation of a Basic Mastergroup	270
9.3	Time-Division Multiplex (TDM)	272
9.3.1	Pseudodigital Modulation	273
9.3.2	Pulse-Amplitude Modulation Encoder	273
9.3.3	Pulse-Amplitude Modulation Decoder	279
9.3.4	Pulse Code Modulation Encoder/Multiplexer	286
9.3.5	Pulse-Code Modulation Decoder/Demultiplexer	286

9.3.6	Bell System T-1 PCM Carrier	286
9.3.7	Telecom Canada Digital Network	290
9.3.8	Synchronization Circuit	290
9.3.9	Regenerative Repeater	291
9.4	Data Transmission Circuits	295
9.4.1	Modem Circuits	298
	References	300
	Problems	301
Chapter 10	The Facsimile Machine	305
10.1	Introduction	305
10.2	Systems Design	307
10.2.1	The Transmit Mode	308
10.2.2	The Receive Mode	308
10.3	Operation	310
10.3.1	“Handshake” Protocol	310
10.4	The Transmit Mode	311
10.4.1	The CCD Image Sensor	311
10.4.2	The Binary Quantizer	315
10.4.3	The Two-Row Memory	317
10.5	Data Compression	318
10.5.1	The Modified Huffman (MH) Code	318
10.5.2	The Modified READ Code	318
10.6	The Modem	319
10.7	The Line Adjuster	319
10.8	The Receive Mode	319
10.8.1	The Power Amplifier	319
10.8.2	The Thermal Printer	321
10.9	Gray Scale Transmission: Dither Technique	322
	References	323
	Bibliography	323
	Glossary	323
	Problems	324
Chapter 11	Personal Wireless Communication Systems	325
11.1	Introduction	325
11.2	Modulation and Demodulation Revisited	326
11.3	Access Techniques	327
11.3.1	Multiplex and Demultiplex Revisited	327
11.3.2	Frequency-Division Multiple Access (FDMA)	328

11.3.3 Time-Division Multiple Access (TDMA)	328
11.3.4 Spread Spectrum Techniques	329
11.4 Digital Carrier Systems	331
11.4.1 Binary Phase Shift Keying (BPSK)	333
11.4.2 Quadrature Phase Shift Keying (QPSK)	334
11.5 The Paging System	338
11.5.1 The POCSAG Paging System	338
11.5.2 Other Paging Systems	341
11.6 The Analog Cordless Telephone	343
11.6.1 System Design	343
11.6.2 Component Design	343
11.6.3 Disadvantages of the Analog Cordless Telephone	347
11.7 The Cellular Telephone	347
11.7.1 System Overview	348
11.7.2 Advanced Mobile Phone System (AMPS)	348
11.8 Other Analog Cellular Telephone Systems	355
11.8.1 Disadvantages of Analog Cellular Telephone Systems	357
11.9 The CDMA Cellular Telephone Systems	358
11.9.1 System Design of the Transmit Path	358
11.9.2 Component Circuit Design for the Transmit Path	359
11.9.3 System Design of the Receive Path	361
11.9.4 Component Circuit Design for the Receive Path	362
11.10 Other Digital Cellular Systems	362
References	363
Bibliography	363
Problems	364
Abbreviations	364

Chapter 12 Telecommunication Transmission Media	367
--	------------

12.1 Introduction	367
12.2 Twisted-Pair Cable	367
12.2.1 Negative-Impedance Converter	370
12.2.2 Four-Wire Repeater	374
12.3 Coaxial Cable	375
12.4 Waveguides	376

12.5	Optical Fiber	376
12.6	Free Space Propagation	377
12.6.1	Direct Wave	379
12.6.2	Earth-Reflected Wave	379
12.6.3	Troposphere-Reflected Wave	379
12.6.4	Sky-Reflected Wave	379
12.6.5	Surface Wave	380
12.7	Terrestrial Microwave Radio	380
12.7.1	Analog Radio	381
12.7.2	Digital Radio	384
12.8	Satellite Transmission System	384
	References	389
	Bibliography	389
	Problems	390
Appendix A	The Transformer	391
A.1	Introduction	391
A.2	The Ideal Transformer	392
A.3	The Practical Transformer	394
Appendix B	Designation of Frequencies	397
	VHF Television Frequencies	397
	UHF Television Frequencies	398
Appendix C	The Electromagnetic Spectrum	399
Appendix D	The Modified Huffman Data Compression Code	401
Appendix E	Electronic Memory	405
E.1	Introduction	405
E.2	Basics of <i>S-R</i> Flip-Flop Circuits	405
E.3	The Clocked <i>S-R</i> Flip-Flop	407
E.4	Initialization of the <i>S-R</i> Flip-Flop	409
E.5	The Shift Register	409
E.5	Electronic Memory	411
E.6	Random Access Memory (RAM)	411
Appendix F	Binary Coded Decimal to Seven-Segment Decoder	415
Index		417

PREFACE

The first edition of this book was published in 1992. Nine years later it had become clear that a second edition was required because of the rapidly changing nature of telecommunication. In 1992, the Internet was in existence but it was not the household word that it is in the year 2001. Cellular telephones were also in use but they had not yet achieved the popularity that they enjoy today. In the current edition, Chapter 1 has been revised to include a section on the Internet. Chapter 10 is new and it covers the facsimile machine; I had overlooked this important telecommunication device in the first edition. Chapter 11 is also new and it describes the pager, the cordless telephone and the cellular telephone system. These are examples of a growing trend in telecommunications to go “wireless”.

This book is about telecommunications: the basic concepts, the design of subsystems and the practical realization of the electronic circuits that make up telecommunication systems. The aim of this book is to fill a gap that exists in the teaching of telecommunications and electronic circuit design to electrical engineering students. Frequently, courses on electronic circuits are taught to students without a clear indication of where these circuits may be used. Later in their career, students may take a course in communication theory where the usual approach is to treat subjects such as modulation, frequency changing and detection as mathematical concepts and to represent them in terms of “black boxes”. Thus the connection between the function “black boxes” and the design of an electronic circuit that will perform the function is glossed over or is completely missing.

The approach followed in this book is to take a specific communication system, for example the amplitude modulated (AM) radio system, and describe in mathematical terms how and why the system is designed the way it is. The system is then broken down into functional blocks. The design of each functional block is examined in terms of the electronic devices to be used, the circuit components and requirements for power. The effectiveness of each functional block is determined. In most cases, more than one circuit is presented, starting from the very elementary which usually illustrates the principles of operation best, to more sophisticated and practical varieties. The order in which the signal encounters the

functional blocks determines the order of the presentation, so new information is presented at an opportune moment when the interest of the student is optimal. Examples are provided to emphasize the link from concept, to design and realization of the circuits. Systems examined in this text include commercial radio broadcasting, television and telephone, with sections devoted to personal wireless communication, satellite communications and data transmission circuits.

This book was written with the final-year engineering undergraduate student in mind, so a clear and explicit style of writing has been used throughout. Illustrative examples have been given whenever possible to promote the active participation of the student in the learning process. However, the practical approach to electronic circuit design will no doubt be useful to people involved in the telecommunications industry for updating, review and as a reference.

Prerequisites to the course in telecommunication circuit design are university mathematics, basic electronics and some familiarity with communication theory (although this is not strictly necessary). In every case where communication theory has a direct impact on the design, enough background has been given to gain an understanding of the topic.

A number of specialized topics have been excluded in the interest of brevity. These include antennas, filters and loudspeakers. Many books are available on these subjects. Rather than presenting a cursory treatment of these very important subjects, I opted for a qualitative description of the operation and design of antennas, filters and loudspeakers in the hope that the reader can develop an appreciation of the outstanding features of these devices which can be built upon if they are of special interest. A list of available reading material has been given in the appropriate chapters.

Although most of the circuits discussed in this book can be found in integrated circuit form, I have, in general, avoided detailed discussion of integrated circuit design. This is because the “rules” of integration aim to reduce the area of the chip to a minimum and thus tend to increase the number of transistors or active components, which take up little area, at the expense of passive components such as resistors and capacitors, which take up relatively large areas. Furthermore, because the integrated circuit process can produce very closely matched transistors, the integrated circuit designer often uses symmetry to achieve circuit functions not possible with discrete devices. An explanation of how an integrated circuit works is therefore more complicated than its discrete counterpart. In every case where the simple and the modern have clashed, I have chosen the simple. However, integrated circuit design techniques have been discussed whenever they are relevant and do not distract the reader from a good understanding of the basic principles of circuit design.

In Chapter 1, a brief history of telecommunication is given. The last 150 years has been a time of tremendous growth and change in telecommunications, more than enough change to qualify as a “revolution”, perhaps the greatest revolution in the history of mankind — the Information Revolution.

Chapters 2 and 3 describe the amplitude modulated (AM) radio system and the electronic circuits that make it possible, from the design of the crystal-controlled

oscillator in the transmitter to the loudspeaker in the receiver. Chapters 4 and 5 repeat the process with the frequency modulated (FM) radio and include sections on stereophonic commercial broadcast and reception.

Television—the transmission and reception of images—is discussed in Chapters 6 and 7. The design of the circuits involved in the acquisition of the video signals, the processing, transmission, coding, broadcasting, reception and decoding are described.

In Chapters 8 and 9, the growth of the telephone system is traced from its humble beginnings to the world-wide network that it is today. The need to open up the system to an increasing number of subscribers has led to the development of sophisticated signal processing techniques and circuits with which to implement them.

Chapter 10 describes the facsimile machine as a system, as well as the design of its component parts.

In Chapter 11 the pager, the cordless telephone and cellular telephone systems are described.

Chapter 12 covers the development of channels in new transmission media, such as satellites and fiber optics, as well as improvements to hard wire connections made possible because circuit designers have produced the hardware at the right time and at the right cost. The growing traffic of “conversations” between machines of various descriptions has accelerated the trend towards “digitization” of signals in the telephone network. The design of circuits capable of accepting data corrupted by noise, restoring and retransmitting them is discussed.

This book started off as lecture notes for a senior college course in electrical engineering called “Telecommunication Circuits”. At the time I proposed the course, it was becoming increasingly clear that the knowledge of our graduating students of communication systems left much to be desired. Students seemed to think that anything analog (including radio, television and the telephone) was *passé*. Digital circuits (computers and software development), on the other hand, were considered “cutting edge”. It was necessary to bring some balance into this situation and I hope that this book helps to restore some semblance of symmetry. The seemingly simple task of changing a set of lecture notes into a textbook turned out not to be quite as simple as I had imagined. However, I have learned a lot from it and I hope the reader does, too.

The material contained in this book is more than can be presented in the normal 13-week term. However, the organization of chapters is based on the three major telecommunication networks: radio, television, and the telephone. It is therefore convenient to organize such a course around a group of chapters with minimal rearrangement of the material and still maintain coherence.

PATRICK D. VAN DER PUJE

*Ottawa, Ontario, Canada
September 2001*

WILEY SERIES IN TELECOMMUNICATIONS AND SIGNAL PROCESSING

John G. Proakis, Editor
Northeastern University

Introduction to Digital Mobil Communications

Yoshihiko Akaiwa

Ditigal Telephony, 3rd Edition

John Bellamy

ADSL, VDSL, and Multicarrier Modulation

John A. C. Bingham

Biomedical Signal Processing and Signal Modeling

Eugene N. Bruce

Elements of Information Theory

Thomas M. Cover and Joy A. Thomas

Practical Data Communications

Roger L. Freeman

Radio System Design for Telecommunications, 2nd Edition

Roger L. Freeman

Telecommunication System Engineering, 3rd Edition

Roger L. Freeman

Telecommunications Transmission Handbook, 4th Edition

Roger L. Freeman

Introduction to Communications Engineering, 2nd Edition

Robert M. Gagliardi

Optical Communications, 2nd Edition

Robert M. Gagliardi and Sherman Karp

Active Noise Control Systems: Algorithms and DSP Implementations

Sen M. Kuo and Dennis R. Morgan

Mobile Communications Design Fundamentals, 2nd Edition

William C. Y. Lee

Expert System Applications for Telecommunications

Jay Liebowitz

Polynomial Signal Processing

V. John Mathews and Giovanni L. Sicuranza

Digital Signal Estimation

Robert J. Mammone, Editor

Digital Communication Receivers: Synchronization, Channel Estimation, and Signal Processing

Heinrich Meyr, Marc Moeneclaey, and Stefan A. Fechtel

Synchronization in Digital Communications, Volume I

Heinrich Meyr and Gerd Ascheid

Business Earth Stations for Telecommunications

Walter L. Morgan and Denis Rouffet

Wireless Information Networks

Kaveh Pahlavan and Allen H. Levesque

Satellite Communications: The First Quarter Century of Service

David W. E. Rees

Fundamentals of Telecommunication Networks

Tarek N. Saadawi, Mostafa Ammar, with Ahmed El Hakeem

Meteor Burst Communications: Theory and Practice

Donald L. Schilling, Editor

Digital Communications over Fading Channels: A Unified Approach to Performance Analysis

Marvin K. Simon and Mohamed-Slim Alouini

Digital Signal Processing: A Computer Science Perspective

Jonathan (Y) Stein

Vector Space Projections: A Numerical Approach to Signal and Image Processing, Neural Nets, and Optics

Henry Stark and Yongyi Yang

Signaling in Telecommunication Networks

John G. Van Bosse

Telecommunication Circuit Design, 2nd Edition

Patrick D. van der Puije

Worldwide Telecommunications Guide for the Business Manager

Walter H. Vignault

1

THE HISTORY OF TELECOMMUNICATIONS

1.1 INTRODUCTION

According to UNESCO statistics, in 1997, there were 2.4 billion radio receivers in nearly 200 countries. The figure for television was 1.4 billion receivers. During the same year, it was reported that there were 822 million main telephone lines in use world-wide. The number of host computers on the Internet was estimated to be 16.3 million [1]. In addition to this, the military in every country has its own communication network which is usually much more technically sophisticated than the civilian network. These numbers look very impressive when one recalls that electrical telecommunication is barely 150 years old. One can well imagine the number of people employed in the design, manufacture, maintenance and operation of this vast telecommunication system.

1.2 TELECOMMUNICATION BEFORE THE ELECTRIC TELEGRAPH

The need to send information from one geographic location to another with the minimum of delay has been a quest as old as human history. Galloping horses, carrier pigeons and other animals have been recruited to speed up the rate of information delivery. The world's navies used semaphore for ship-to-ship as well as from ship-to-shore communication. This could be done only in clear daylight and over a distance of only a few kilometres. The preferred method for sending messages over land was the use of beacons: lighting a fire on a hill, for example. The content of the message was severely restricted since the sender and receiver had to have previously agreed on the meaning of the signal. For example, the lighting of a beacon on a particular hill may inform one's allies that the enemy was approaching from the north, say. In 1792, the French Legislative Assembly approved funding for the demonstration of a 35 km visual telegraphic system. This was essentially

semaphore on land. By 1794, Lille was connected to Paris by a visual telegraph [3]. In England, in 1795, messages were being transmitted over a visual telegraph between London and Plymouth – a return distance of 800 km in 3 minutes [4].

North American Indians are reputed to have communicated by creating puffs of smoke using a blanket held over a smoking fire. Such a system would require clear daylight as well as the absence of wind, not to mention a number of highly skilled operators.

A method of telecommunication used in the rain forests of Africa was the “talking drum”. By beating on the drum, a skilled operator could send messages from one village to the next. This system of communication had the advantage of being operational in daylight and at night. However, it would be subject to operator error, especially when the message had to be relayed from village to village.

1.3 THE ELECTRIC TELEGRAPH

The first practical use of electricity for communication was in 1833 by two professors from the University of Goettingen, Carl Friedrich Gauss (1777–1855) and Wilhelm Weber (1804–1891). Their system connected the Physics Institute to the Astronomical Observatory, a distance of 1 km, and used an induction coil and a mirror galvanometer [4].

In 1837, Charles Wheatstone (1802–1875) (of Wheatstone Bridge fame) and William Cooke (1806–1879) patented a communication system which used five electrical circuits consisting of coils and magnetic needles which deflected to indicate a letter of the alphabet painted on a board [5]. The first practical use of this system was along the railway track between Euston and Chalk Farm stations in London, a distance of 2.5 km. Several improvements were later made, the major one being the use of a coding scheme which reduced the system to a single coil and a single needle. The improvement of the performance, reliability and cost of communication has since kept many generations of engineers busy.

At about the time when Wheatstone and Cooke were working on their system, Samuel Morse (1791–1872) was busy doing experiments on similar ideas. His major contribution to the hardware was the relay, also called a repeater. By connecting a series of relays as shown in Figure 1.1, it was possible to increase the distance over

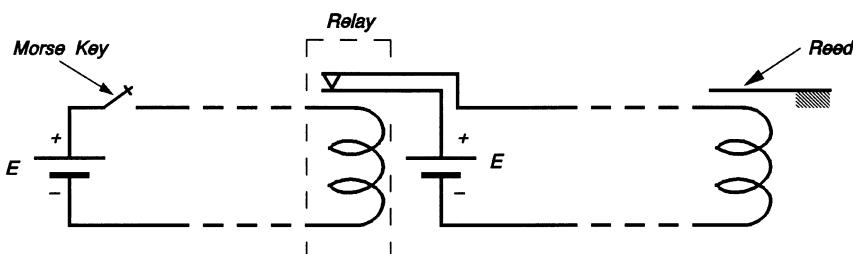


Figure 1.1. The use of Morse's relay to extend the range of the telegraph.

which the system could operate [5]. Morse also replaced the visual display of Wheatstone and Cooke with an audible signal which reduced the fatigue of the operators. However, he is better known for his efficient coding scheme which is based on the frequency of occurrence of the letters in the English language so that the most frequently used letter has the shortest code (E: dot) and the least frequently used character has the longest code ('-apostrophe: dot-dash-dash-dash-dot). This code was in general use until the 1950s and it is still used by amateur radio operators today.

In 1843, Morse persuaded the United States Congress to spend \$30,000 to build a telegraph line between Washington and Baltimore. The success of this enterprise made it attractive to private investors, and Morse and his partner Alfred Vail (1807–1859), were able to extend the line to Philadelphia and New York [6]. A number of companies were formed to provide telegraphic services in the east and mid-west of the United States. By 1851, most of these had joined together to form the Western Union Telegraph Company.

By 1847, several improvements had been made to the Wheatstone invention by the partnership of Werner Siemens (1816–1892) and Johann Halske (1814–1890) in Berlin. This was the foundation of the Siemens telecommunication company in Germany.

The next major advance came in 1855 when David Hughes (1831–1900) invented the printing telegraph, the ancestor of the modern teletype. This must have put a lot of telegraph operators out of work (a pattern which was to be repeated over and over again) since the machine could print messages much faster than a person could write. Another improvement which occurred at about this time was the simultaneous transmission of messages in two directions on the same circuit. Various schemes were used but the basic principle of all of them was the balanced bridge.

In 1851, the first marine telegraphic line between France and England was laid, followed in 1866 by the first transatlantic cable. The laying of this cable was a major feat of engineering and a monument to perseverance. A total of 3200 km of cable was made and stored on an old wooden British warship, the HMS Agamemnon. The laying of the cable started in Valentia Bay in western Ireland but in 2000 fathoms of water, the cable broke and the project had to be abandoned for that year. A second attempt the following year was also a failure. A third attempt in 1858 involved two ships and started in mid-ocean and it was a success. Telegraphic messages could then be sent across the Atlantic. The celebration of success lasted less than a month when the cable insulation broke down under excessively high voltage. Interest in transatlantic cables was temporarily suspended while the American Civil War was fought and it was not until 1865 that the next attempt was made. This time a new ship, the Great Eastern, started from Ireland but after 1900 km the cable broke. Several attempts were made to lift the cable from the ocean bed but the cable kept breaking off so the project was abandoned until the following year. At last in 1866, the Great Eastern succeeded in laying a sound cable and messages could once more traverse the Atlantic. By 1880, there were nine cables crossing the ocean [6].

The telegraph was and remained a communication system for business, and in most European countries it became a government monopoly. Even in its modernized

form (telex) it is essentially a cheap long-distance communication network for business.

1.4 THE FACSIMILE MACHINE

In 1843, the British Patent Office issued a patent with the title “Automatic electrochemical recording telegraph” to the Scottish inventor Alexander Bain (1810–1877). The essence of the invention is shown in Figure 1.2. Two identical pendulums are connected as shown by the telegraph line. For simplicity, we assume the “message” to be sent is the letter *H* and it is engraved on a metallic plate and shaped to the appropriate radius so that the “read” stylus makes contact with the raised parts of the plate as the pendulum sweeps across it. On the far end of the telegraph line, the stylus of the second pendulum maintains contact with the electrosensitive paper which rests on an electrode shaped to the same radius as before. The electrosensitive paper has been treated with a chemical which produces a dark spot when electric current flows through it.

To operate the system, both pendulums are released from their extreme left positions simultaneously. Since they are identical, it follows that they will travel at the same speed, one across the “message” plate and the other across the electro-sensitive paper. At first no current flows, but as the transmitter pendulum makes contact with the raised portion of the plate, the circuit is complete and the resulting current causes the electro-sensitive paper to produce a dark line of the same length as the raised metal segment. The original patent included the functions: (a) an electromagnetic device to keep the pendulums swinging at a constant amplitude

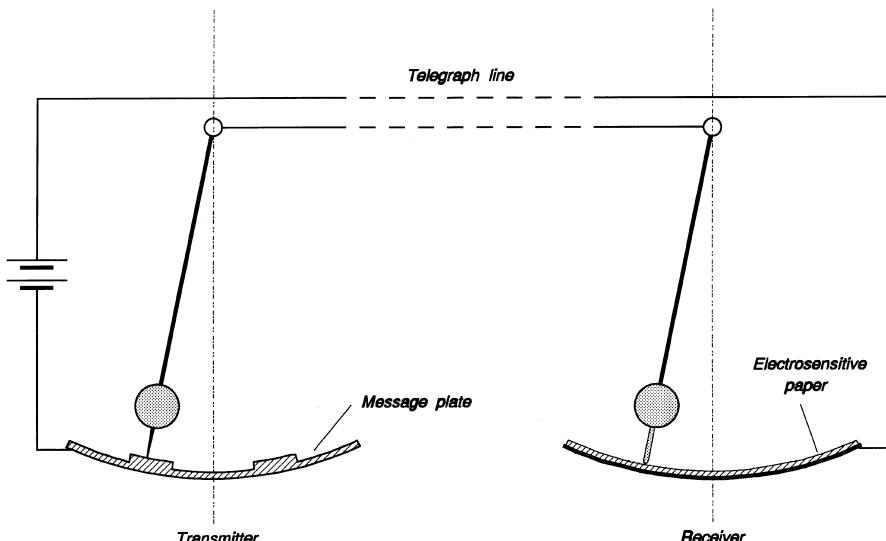


Figure 1.2. The configuration of the Bain “Automatic electrochemical recording telegraph.” To keep the diagram simple, additional circuits required for synchronization, phasing and scanning are not shown.

(synchronization), (b) a second electromagnetic arrangement to ensure that the two pendulums start their swings at the same instant (phasing), and (c) a mechanism to move the message plate and the electrosensitive paper simultaneously one step at a time after each sweep at right angles to the direction of the pendulum swing (scanning). When several sweeps have occurred, the lines produced will form an exact image of the raised metal parts of the “message” plate. Figure 1.3(a) shows the letter *H* scanned in 20 lines and Figure 1.3(b) shows the corresponding current waveforms. Figure 1.3(c) shows the reproduced image.

All the facsimile machines since the Bain patent have the three functions listed above. In modern facsimile machines, the first two functions have been replaced by electronic techniques which ensures that the transmitter and the receiver are “locked” to each other at all times. The mechanism for scanning the message is also largely electronic, although in most machines it is still necessary to move the page mechanically as it is scanned.

In 1848, Frederick Bakewell, an Englishman, produced a new version of the fax machine in which the “message plate” as well as the image were mounted on

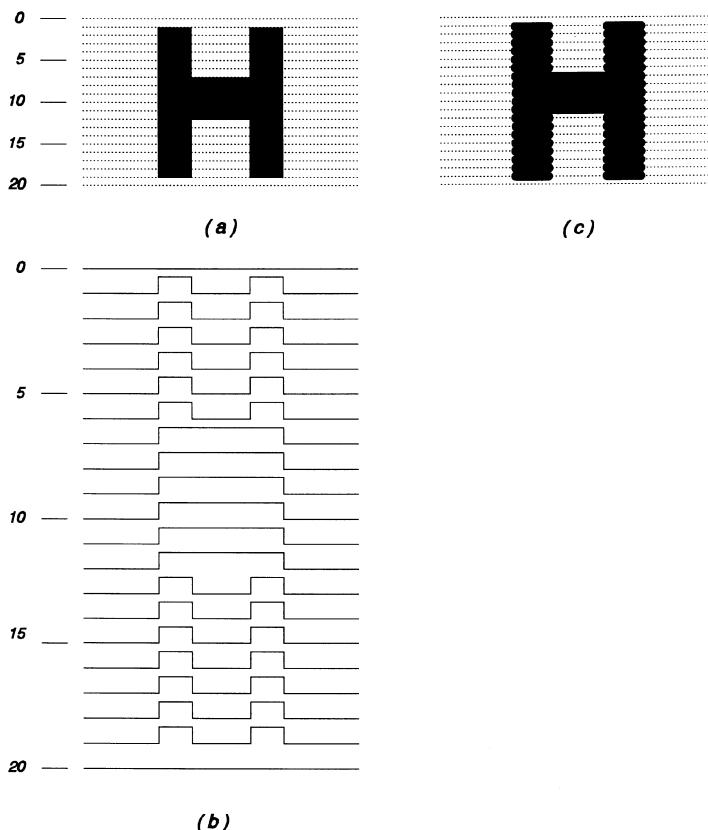


Figure 1.3. (a) Shows the letter *H* scanned in 20 lines, (b) shows the current waveforms for each line scanned and (c) shows the reproduced *H*. Note the effect of the finite width of the receiver stylus on the image.

cylinders which were turned by falling weights, similar to a grandfather clock. To ensure that the cylinders turned at the same speed he used a mechanical speed governor. The scanning head (stylus) was propelled on an axis parallel to that of the cylinder by a lead-screw. This was an example of “spiral scanning”. Unlike Bain, he used an insulating ink to write the message on a metallic surface. But, as before, the paper in the receiver was chemically treated to respond to the flow of electric current and it was mounted on an identical cylinder with the “write” head driven by an identical lead-screw. The main difficulty with this design was the necessity to keep the two clock motors in remote locations starting and running at the same speed during the transmission.

In 1865, Giovanni Caselli (1815–1891), an Italian living in France, patented an improved version of Bain’s machine which he called the “Pantelegraph”. He then established connections between Paris and a number of other French cities. His machine was a combination of the insulating ink message plate of Bakewell, the pendulum of Bain’s transmitter, and the Bakewell cylindrical receiver. The pantelegraph was a commercial success and it was used in Italy and Britain for many years.

By the end of the 1800s it was possible to send photographs by fax. The picture had to be etched on a metallic plate in the form of raised dots (similar to the technique used for printing pictures in newspapers). The size of the dots represented the different shades of gray; small for light and large for dark gray. The transmitter stylus traced lines across the picture making contact with the raised dots and thus producing corresponding large and small dots at the receiver.

In 1902, Arthur Korn demonstrated a scanning system which used light instead of physical contact with a metallic plate and the resultant flow of current. His method was far superior to all the previous techniques, especially in the transmission of photographs. He wrapped the photographic film negative of the picture on the outside of a glass cylinder which was turned at a constant rate by an electric motor. An electric lamp provided the light and a system of lenses were used to focus the light onto the negative. The light that passed through the film was reflected by a mirror onto a piece of selenium whose resistance varied according to how much light reached it. The selenium cell was used to control the current flowing in the receiver. The receiver recorded the image directly onto film. To ensure that the transmitter and receiver cylinders were in synchronism at all times, he used a central control system with a tuning fork generating the control signal.

In the 1920s the large American telecommunication companies, American Telephone and Telegraph (AT&T), Radio Corporation of America (RCA) and Western Union, became interested in fax machine development and they used new techniques, materials and devices such as the vacuum tube, phototubes and later semiconductors to produce the modern fax machine.

1.5 THE TELEPHONE

In 1876, Alexander Graham Bell (1847–1922) was conducting experiments on a “harmonic telegraph” system when he discovered that he could vary the electric

current flowing in a circuit by vibrating a magnetic reed held in close proximity to an electromagnet which formed part of the loop. By connecting a second electromagnet together with its own magnetic reed in the circuit, he could reproduce the vibration of the first reed. Using a human voice to excite the magnetic reed led to the first telephone for which he was granted a patent later that year. He went on to demonstrate his invention at the International Centennial Exhibition in Philadelphia and before the year ended, he transmitted messages between Boston, Massachusetts, and North Conway, New Hampshire, a distance of 230 kilometres. Few people realized the potential of the new invention and in 1878, when Bell tried to sell his patent to the Western Union Telegraph Company, he was turned down [7].

The early telephone system consisted of two of Bell's magnetic reed-electromagnet instruments in series with a battery and a bell. Bell's instrument worked very well as a receiver, in fact so well that it has survived almost unchanged to this day. As a transmitter, however, it left a lot to be desired. It was soon replaced by the carbon microphone (one of the many inventions of Thomas Edison (1847–1931)) which was, until recently, the most widely used microphone in the telephone system.

In the early telephone system, each subscriber was connected to a central office by a single wire with an earth return. This led to cross-talk between subscribers. At about this time, electric traction had become very popular which resulted in increased interference from the noise generated by the electric motors. The earth-return system was gradually replaced by two-wire circuits which are much less susceptible to cross-talk and electrical noise. The rapid growth of the telephone system was based almost entirely on the fact that the subscriber could use the system with the minimal amount of training. The ease of operation of the telephone outweighed the disadvantages of having no written record of conversations and the requirement that both parties have to be available for the call at the same time. The basic central office responds to a signal from the subscriber (calling party) indicating that he wants service. A buzzer excited by current from a hand-cranked magneto was the standard. The telephone operator answers and finds out whom (called party) the calling party wants to talk to. The operator then signals to the called party by connecting his own hand-cranked magneto to the line and cranking it to ring a bell on the called party's premises. When and if the called party responds, he connects the lines of the two parties together and withdraws until the conversation is over, at which point he disconnects the lines. In order to carry out his function, the operator had to have access to all the lines connected to the exchange. This was not a problem in an exchange with less than fifty lines but as the system grew, more operators were required for each group of fifty subscribers. If the calling party and called party belong to the same group of fifty, the above sequence was followed. If they belong to two different operators, it was necessary for the two operators to have a verbal consultation before the connection could be made. The errors, delays and misunderstandings in large central offices led to a re-organization whereby each operator responded to only fifty incoming lines but had access to all the outgoing lines. Another improvement in the system was to replace all the batteries on the subscribers' premises with one battery in the central office.

The motivation for the changeover from manual switching to the automatic telephone exchange was not, as one would expect, the inability of the central office to cope with the increasing volume of traffic. It was because the operators could listen to the conversations. The inventor of the automatic exchange, Almon B. Strowger (1839–1902), after whom the system was named, was an undertaker in Kansas City around the 1890s. There was another undertaker in the city whose wife worked in the local telephone exchange; whenever someone died in the city, the telephone operators were the first to know and the wife would pass a message to the husband, giving him a head-start on his competitor [8]. The automatic exchange certainly improved the security of telephone conversations; it was also one more example of machines replacing people.

The success of the telephone system led to a large number of small telephone companies being formed to service the local urban communities. Pressure to inter-connect the various urban centres soon grew and techniques for transmission over longer distances had to be developed. These included amplification and inductive loading. Since these transmission lines (trunks) were expensive to construct and maintain, techniques for transmission of more than one message (multiplex) over the trunk at any one time became a matter of great concern and an area of rapid advancement.

1.6 RADIO

In 1864, James Maxwell (1831–1879), a Scottish physicist, produced his theory of the electromagnetic field which predicted that electromagnetic waves can propagate in free space at a velocity equal to that of light [9]. Experimental confirmation of this theory had to wait until 1887 when Heinrich Hertz (1857–1894) constructed the first high-frequency oscillator. When a voltage was induced in an induction coil connected across a spark gap, a discharge would occur across the gap setting up a damped sinusoidal high-frequency oscillation. The frequency of the oscillation could be changed by varying the capacitance of the gap by connecting metal plates to it. The detector that he used consisted of a second coil connected to a much shorter spark gap. The observation of sparks across the detector gap when the induction coil was excited showed that the electromagnetic energy from the first coil was reaching the second coil through space. These experiments were in many ways similar to those carried out in 1839 by Joseph Henry (1797–1878). Several scientists made valuable contributions to the subject, such as Edouard Branly (1844–1940) who invented the “coherer” for wave detection, Aleksandr Popov (1859–1906) and Oliver Lodge (1851–1940) who discovered the phenomenon of resonance.

In 1896, Guglielmo Marconi (1874–1937) left Italy for England where he worked in cooperation with the British Post Office on “wireless telegraph”. A year later, he registered his “Wireless Telegraphy and Signal Co. Ltd” in London, England to exploit the new technology of radio. On the 12th of December 1901, Marconi received the letter “S” in Morse code at St. Johns, Newfoundland on his receiver whose antenna was held up by a kite, the antenna which he had constructed for the

purpose having been destroyed by heavy winds. He had confounded the many skeptics who thought that the curvature of the earth would make radio transmission impossible [10].

Up to this point, no use had been made of “electronics” in telecommunication: high-frequency signals for radio were generated mechanically. The first electronic device, the diode, was invented by Sir John Ambrose Fleming (1849–1945) in 1904. He was investigating the “Edison effect” that is, the accumulation of dark deposits on the inside wall of the glass envelope of the electric light bulb. This phenomenon was evidently undesirable because it reduced the brightness of the lamp. He was convinced that the dark patches were formed by charge particles of carbon given off by the hot carbon filament. He inserted a probe into the bulb because he had the idea that he could prevent the charged particles from accumulating by applying a voltage to the probe. He soon realized that, when the probe was held at a positive potential with respect to the filament, there was a current in the probe but when it had a negative potential no current would flow: he had invented the diode. He was granted the first patent in electronics for his effort. Fleming went on to use his diode in the detection of radio signals – a practice which has survived to this day.

The next major contribution to the development of radio was made by Lee DeForest (1873–1961). He got into legal trouble with Marconi, the owner of the Fleming diode patent, when he obtained a patent of his own on a device very similar to Fleming’s. He went on to introduce a piece of platinum formed into a zig-zag around the filament and soon realised that, by applying a voltage to what he called the “grid”, he could control the current flowing through the diode. This was, of course, the triode – a vital element in the development of amplifiers and oscillators.

1.7 TELEVISION

Shortly after the establishment of the telegraph, the transmission of images by electrical means was attempted by Giovanni Caselli (1815–1891) in France. His technique was to break up the picture into little pieces and send a coded signal for each piece over a telegraph line. The picture was then reconstituted at the receiving end. The system was slow, even for static images, but it established the basic principles for image transmission; that is, the break up of the picture into some elemental form (scanning), the quantization of each element in terms of how bright it is (coding), and the need for some kind of synchronization between the transmitter and the receiver. Subsequent practical image transmission schemes, whether mechanical or electronic, had these basic units.

The discovery in 1873 by Joseph May, a telegraph operator at the Irish end of the transatlantic cable, that when a selenium resistor was exposed to sunlight its resistance decreased, led to the development of a light-to-current transducer. Subsequently, various schemes for image transmission based on this discovery were devised by George Carey, William Ayrton (1847–1908), John Perry and others. None of these was successful because they lacked an adequate scanning system and

each element of the picture had to be sent on a separate circuit, making them quite impractical.

In 1884, Paul Nipkow (1860–1940) was granted a patent in Germany for what became known as the Nipkow Disc. This consisted of a series of holes drilled in the form of spirals in a disc. When an image is viewed through a second disc with similar holes driven in synchronism with the first, the observed effect was scanning point-to-point to form a complete line and line-by-line to cover the complete picture. This was a practical scheme since the point-to-point brightness of the picture could be transmitted and received serially on a single circuit. The persistence of an image on the human eye could be relied on to create the impression of a complete scene when, in fact, the information is presented point-by-point. Nipkow's scheme could not be exploited until 1927 when photosensitive cells, photomultipliers, electron tube amplifiers and the cathode ray tube had been invented and had attained sufficient maturity to process the signals at an acceptable speed for television. Several people made significant contributions to the development of the components as well as to the system. However, two people, Charles Jenkins (1867–1934) and John Baird (1888–1946), are credited with the successful transmission of images at about the same time. They both used the Nipkow disc. Mechanical scanning methods of various forms were used with reasonable success until about 1930 when Vladimir Zworykin (1889–1982) invented the “iconoscope” and Philo Farnsworth (1906–1971) the electronic camera tube, which he called the “image dissector”. These inventions finally removed all the moving parts from television scanning systems and replaced them with electronic scanning [11]. The application of very-high-frequency carriers and the use of coaxial cables have contributed significantly to the quality of the pictures. The use of color in television had been shown to be feasible in 1930 but would not be available to the general public until the mid-1960s. By the 1980s, satellite communication systems brought a large number of television programs to viewers who could afford the cost of the dish antenna. By the beginning of the 21st century, the dish antennas had shrunk in size from over 3 m to less than 70 cm and the signal had changed into digital form.

1.8 THE GROWTH OF BANDWIDTH AND THE DIGITAL REVOLUTION

Electrical telecommunication started with a single wire with a ground return, but, as the system grew, the common ground return had to be replaced with a return wire, hence the advent of the open-wire telephone line. The open-wire system with its forests of telegraph poles along city streets strung with an endless array of wires eventually gave way to the twisted pair cable. The twisted pair cable owes its existence to improved insulating materials, mainly plastics, which reduced the space requirements of the cable. The bandwidth of an unloaded twisted pair is approximately 4 kHz and it decreases rapidly with length. This can be improved by connecting inductors (loading coils) in series with the line at specific distances and by various equalization schemes to about 1 MHz. However, the twisted pair has found a niche in the modern telephone system where its bandwidth approximately

matches that required for analog audio communication. This is still the dominant mode of telephone communication up to the central office. Beyond the central office the network of inter-office trunks use a variety of conduits for the transmission of the signal.

Increased bandwidth alone was not an answer to the expanding telecommunication traffic. High-frequency carriers had to be developed in order to exploit fully the bandwidth capability of new telecommunication media such as coaxial cables, terrestrial microwave networks and fiber optics. The development of the coaxial cable, which confines the electromagnetic wave to the annular space between the two concentric conductors, reduced significantly the radiation losses that would otherwise occur. As a result the bandwidth was increased to approximately 1 GHz and attenuation was reduced. Terrestrial as well as satellite microwave communication systems have further expanded the bandwidth into the terahertz range and, for those who can afford the dish antenna and its associated equipment, it has increased the number of television channels available to over 800. The application of fiber optics to telecommunication has extended the channel bandwidth to that of visible light (1×10^{12} Hz). It is now possible for one optical fiber to carry as many as 300×10^9 telephone channels at the same time.

An increasingly dominant factor in telecommunication is the enormous popularity of digital techniques. The information is reduced to a train of pulses (binary digits; 1s and 0s) and sent over the channel. The limited bandwidth, phase change and the noise in the channel cause the signal to deteriorate so it is necessary to “refresh” or regenerate the signal at various points along the channel. This is accomplished by using repeaters whose function is to determine whether the digit sent was a 1 or a 0 and to generate the appropriate new digits and transmit them. At the receiving end, the digits are converted back into an analog signal. The compact disc music recording system is a common example of this technique. Although the need for information transfer between computers spurred on the development of digital communication, speech signals increasingly are being converted into digital form for telephone transmission.

1.9 THE INTERNET

The use of personal computers as a means of communication gained enormous popularity in the last decade of the 20th century. However, computer science experts have used the ARPANET (Advanced Research Projects Agency of the U.S. Department of Defense) for communication between computers since 1969. The basic idea was to enable scientists in different geographic locations to share their research results [12] and also, as a money saving scheme, their computing resources. The first four sites to be connected were the Stanford Research Institute, the University of California at Los Angeles, the University of California at Santa Barbara, and the University of Utah in Salt Lake City. The messages traveling between these centers were over 50 kbps telephone lines. In 1962 when the ARPANET was being designed, the Cold War was in full swing and so one of the

specifications for the design was that the network should survive a nuclear attack in which parts of it were knocked out [13]. Needless to say, this feature of the design was never tested! The ideal structure was a network with every node connected to every other node (high redundancy) so that, if a part of the network went down for whatever reason, the traffic could be routed around the trouble spot. Moreover, in such a design all nodes are of equal importance, hence there is no one node the destruction of which would cripple the network. The configuration of the network is shown in Figure 1.4. This is similar to the electric power grid which was designed to provide electric power to consumers with a maximum reliability service.

Another design feature of the ARPANET which further improved its robust nature was the use of “packet” switching. In packet switching, the incoming message is first divided into smaller packets of binary code. Each packet is labeled with a number and the address of its destination and then transmitted to the next node when the local router can accommodate the packet. Each packet, in theory, can travel from the source to the destination by a different route and arrive at different times. At its destination the packets are re-assembled in the proper order ready for the recipient. The strength of packet switching is the fact that, if a number of nodes are put out of operation, the packets will still find their way to their proper destination by way of the remaining operational nodes, in perhaps a longer time. Moreover, error-detection codes can be included with each packet and, when errors are present in a packet, that packet can be re-sent.

The ARPANET grew so that by 1983 there were 562 sites connected. By 1992, the number of “host” or “gateway” computers connected to it had reached one million. Four years later, the number was 12 million. It has been estimated that by the year 2000 the number with access to the Internet worldwide will be 100 million [14]. The term “Internet” came into use in 1984, and this was also the time when the

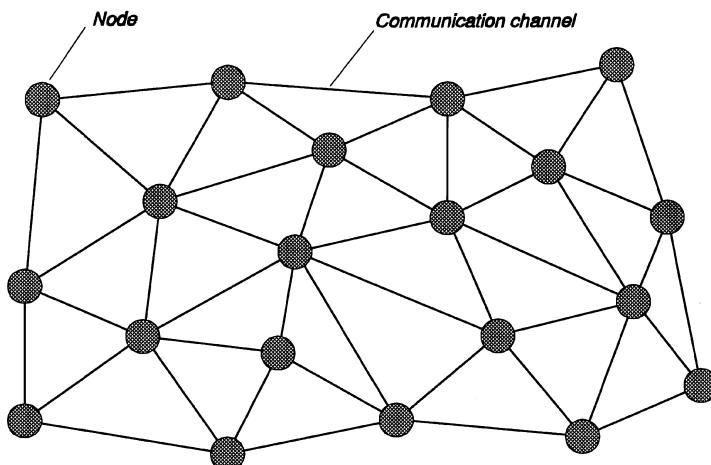


Figure 1.4. Each node of the Internet is connected to all the neighboring nodes. The increased redundancy implies a high level of reliability. The design of the electric power grid follows the same principle for the same reason.

United States Department of Defense handed over the oversight of the network to the National Science Foundation. The Internet is currently run in a very loose fashion by a number of volunteer organizations whose membership is open to the public. Their main activity is centered around the registration of names, numbers and addresses of the users of the system.

The Internet is a collection of a large number of computers connected together by telephone lines, coaxial cables, optical fiber cables and communication satellites with set protocols to enable communication between them and also to control the flow of traffic.

1.10 THE WORLD WIDE WEB

What factors have contributed to the unprecedented growth of the Internet?

Personal computers have been in common use in scientific laboratories and in universities since the mid-1980s but they were mostly used for calculation, information storage and retrieval. Many businesses acquired desk-top computers for preparing invoices, word-processing and general book-keeping. Some enthusiasts owned their own personal computers and some belonged to clubs for the exchange of computer software which they had developed. The growth of computing power of the personal computer was one of the pivotal developments that made the Internet possible. In 1971, the Intel Corporation produced its first microprocessor, the Intel 4004. It was used in a calculator and its clock frequency (an indication of how fast it operates) was 108 kHz. The following year the Corporation produced the Intel 8008 which was twice as fast (200 kHz) as the 4004 and it was used in 1974 in a predecessor of the first personal computer. Also in 1974, Intel produced the 8080 which was clocked at 2 MHz. The 8080 was marketed to computer enthusiasts as part of a kit and it very quickly became the “brains” of the modern personal computer. By the year 2000, the Intel Pentium III processor had achieved a clocking speed of 1.13 GHz (over four orders of magnitude faster than the original 4004). This phenomenal increase in speed was coupled to an equally incredible decrease in price which made the personal computer affordable to the general public. A hypothetical comparison with the automobile industry in 1983 was as follows:

“If the automobile business had developed like the computer business, a Rolls-Royce would now cost \$2.75 and run 3 million miles on a gallon of gas” [15].

Even this comparison was considered conservative fifteen years later. In the *Spirit of the Web*, Wade Rowland amends the statement as follows:

“That Rolls-Royce would now cost twenty-seven cents and run 300 million miles on a gallon of gas”.

The telephone system was already in place, although its capacity would have to be expanded to carry the digital data in addition to the voice signals for which it was designed. Most of the main communication lines carrying Internet data have been

updated, for example the copper cables (coaxial) laid across continents and on the sea-bed are capable of 2.5 Mbps and fiber optic cables can go as high as 40 Gbps. The building of the throughways for large volumes of data was the easy part of the problem. The more difficult part was to get the data to its destination in the workplaces and into the living quarters of the owners of personal computers. Unfortunately, the cost of wiring houses with optical fiber cable or even coaxial cable cannot be justified on economic grounds. Currently, the speed limit to data flow is determined by the analog telephone line (sometimes referred to as a “twisted pair”) between the central office (or its equivalent) and the wall socket to which the most personal computers are connected. This is popularly known as *the-last-mile* problem. New circuits have been developed to speed up data transfer on the existing twisted pair cable. These include the T1, the Integrated Services Digital Network (ISDN), the High-bit-rate Digital Subscriber Line (HDSL) and the Asymmetrical Digital Subscriber Line (ADSL).

In the early 1990s, cable television service had reached into a large number of homes in North America and Europe and there was talk of them providing high-bandwidth (mainly coaxial cable) conduits to subscribers for access to the Internet. Unfortunately, the television cable network had millions of amplifiers and one-way traps installed which restricted signal flow in only one direction. Connection to the Internet required a bilateral flow of information and the cost of the conversion was considered prohibitive. At the time this book was going to press, the cable television companies were in the process of converting their networks for bilateral flow of information (cable modems). The speed of transmission was predicted to be from 10 Mbps to 400 Mbps [16].

Another serious impediment to the free flow of data from one computer to another was the almost incomprehensible commands required to effect computer communication. The “spoken language” of the computers was UNIX and this was quite unfriendly to the uninitiated. The “point-and-click” feature of the computer “mouse” and the development of browsers, such as Netscape Navigator and the Internet Explorer, finally lowered the threshold to a level where even computer neophytes could successfully access information from the Internet. These “facilitators” were all available by 1989 and they would have a very profound effect on the popularity of the Internet. The Internet can be seen as a network connecting various sites where information is stored. The stored information and the technology for transferring the information back and forth is the World Wide Web.

The World Wide Web in its infancy carried only text. Later the transmission of graphics, in color, was added. With the increasing speed and sophistication of the personal computer, sound and video have been added subsequently. What made the Web particularly useful is the ability of the Internet browsers and search engines to provide a list of Internet sites where the requested information may be located. It is necessary to prompt the system with a set of keywords for the search to begin. At the Web site there are “links” to other sites so the search can “fan out” in very many directions.

An important factor that stimulated the growth of the Internet was the decision of the United States government to turn over the running of the Internet to commercial

interests. This happened in stages. The National Science Foundation's NSFnet was superceded by the Advanced Networks and Services network, ANSnet, a non-profit organization. However, in 1998 the Federal Communications Commission (FCC) insisted that subscribers to the telephone system be billed at the same rate for voice services as for data. The way was now open for commercial organizations to offer their services on the Internet. The Internet Service Providers (ISP) collect money from their subscribers for access to their servers. They in turn have to pay for access to the long-distance, high-speed Internet backbone.

The decentralized nature of the Web and its ability to transfer information bilaterally meant that its users could add their own contribution to the vast amount already present in the form of their "personal home page". Commercial organizations, special-interest groups and even governments would take advantage of the possibilities offered by the Web. But so would people and groups with hidden and not-so-hidden agendas to propagate their own distortions. As there is no authority to monitor the content of the Web sites, only the criminal laws of the country in which the Web site is registered can be used to control information on the sites.

A number of services are available to people who have access to the Internet. Newsgroups can be found on practically any topic. These newsgroups run round-the-clock and anyone can join in the discussion from his or her keyboard. The participants are free to use assumed names and identities. The number of people in any given newsgroup can vary from zero to several thousand, so it is possible to reach a very large audience from one's keyboard. Chat rooms are similar to newsgroups except that the number of people "in" a chat room is likely to be much smaller. It is essentially a conversational mode of communication from one's keyboard. One of the more popular features of the Internet is electronic mail (e-mail). It is the nearest thing to mailing a letter to a correspondent, and although it is not as secure as the service provided by the Post Office, it is much faster. To send e-mail, one has to register a unique e-mail address and choose a password.

REFERENCES

1. *Statistical Yearbook 1999*, UNESCO, Paris.
2. Berto, C., *Telegraphes et Telephones de Valmy au Microprocesseurs*, Le Livre de Poche, 1981.
3. Stumpers, F. L. H. M., "The History, Development and Future of Telecommunications in Europe", *IEEE Comm. Magazine*, 22(5), 1984.
4. Fraser, W., *Telecommunications*, MacDonald & Co, London, 1957.
5. Tebo, Julian D., "The Early History of Telecommunications", *IEEE Comm. Soc. Digest*, 14(4), pp. 12–21, 1976.
6. Osborne, H. S., *Alexander Graham Bell*, Biographical Memoirs, *Nat. Acad. Sci.*, Vol. 23, pp 1–29, 1945.

7. Smith, S. F., *Telephony and Telegraphy A*, 2nd Ed., Oxford University Press, New York, 1974.
8. Bernal, J.D., *Science in History, Vol. 2*, Penguin Books Ltd, Middlesex, 1965.
9. Carassa, F., “On the 80th Anniversary of the First Transatlantic Radio Signal”, *IEEE Antennas Propagat. Newslett.*, pp. 11–19, Dec., 1982.
10. Knapp, J. G. and Tebo, J. D., “The History of Television”, *IEEE Comm. Soc. Digest*, 16(3), pp 8–21, May 1978.
11. Licklider, J. C. R. and Clark, W., “On-line Man-Computer Communications”, Massachusetts Institute of Technology, Aug., 1962.
12. Baran, P., “On Distributed Communications Networks”, RAND Corporation, Washington DC, Sept., 1962.
13. Zakon, R. H., “Hobbes’ Internet Timeline v5.1” at www.isoc.org/zakon/internet, Oct., 2000.
14. “The Computer Moves In”, *Time*, January 3, 1983, 10.
15. Rowland, W., *Spirit of the Web*, Somerville House Pub., Toronto, 1997.
16. www.wired.com/news

BIBLIOGRAPHY

- Jones, C. R., *Facsimile*, Murray Hill Books, New York, 1949.
- Costigan, D. M., *Fax: The Principles and Practice of Facsimile Communication*, Chilton Book Co., Philadelphia, 1971.
- McConnell, K., Bodson, D., and Urban, S., *Fax: Facsimile Technology and Systems*, 3rd Ed., Artech House, Boston, 1999.
- Dodd, Annabel, Z. *The Essential Guide to Telecommunications*, 2nd Ed., Prentice-Hall, Englewood Cliffs, NJ, 2000.
- Lehnert, Wendy, G., *Internet 101: A Beginner’s Guide to the Internet and the World Wide Web*, Addison-Wesley, Reading, MA, 1998.

2

AMPLITUDE MODULATED RADIO TRANSMITTER

2.1 INTRODUCTION

A radio signal can be generated by causing an electromagnetic disturbance and making suitable arrangements for this disturbance to be propagated in free space. The equipment normally used for creating the disturbance is the transmitter, and the transmitter antenna ensures the efficient propagation of the disturbance in free space. To detect the disturbance, one needs to capture some finite portion of the electromagnetic energy and convert it into a form which is meaningful to one of the human senses. The equipment used for this purpose is, of course, a receiver. The energy of the disturbance is captured using an antenna and an electrical circuit then converts the disturbance into an audible signal.

Assume for a moment that our transmitter propagated a completely arbitrary signal (that is, the signal contained all frequencies and all amplitudes). Then no other transmitter can operate in free space without severe interference because free space is a common medium for the propagation of all electromagnetic waves. However, if we restrict each transmitter to one specific frequency (that is, continuous sinusoidal waveforms) then interference can be avoided by incorporating a narrow-band filter at the receiver to eliminate all other frequencies except the desired one. Such a communication channel would work quite well except that its signal cannot convey information since a sinusoid is completely predictable and information, by definition, must be unpredictable.

Human beings communicate primarily through speech and hearing. Normal speech contains frequencies from approximately 100 Hz to approximately 5 kHz and a range of amplitudes starting from a whisper to very loud shouting. An attempt to propagate speech in free space comes up against two very severe obstacles. The first is similar to that of the transmitters discussed earlier, in which they interfere with each other because they share the same medium of propagation. The second obstacle is due to the fact that low frequencies, such as speech, cannot be propagated

efficiently in free space whereas high frequencies can. Unfortunately, human beings cannot hear frequencies above 20 kHz which is, in fact, not high enough for free space transmission. However, if we can arrange to change some property of a continuous sinusoidal high-frequency source in accordance with speech, then the prospects for effective communication through free space become a distinct possibility. Changing some property of a (high-frequency) sinusoid in accordance with another signal, for example speech, is called *modulation*. It is possible to change the amplitude of the high-frequency signal, called the *carrier*, in accordance with speech and/or music. The modulation is then called *amplitude modulation* or AM for short. It is also possible to change the phase angle of the carrier, in which case we have *phase modulation* (PM), or the frequency, in which case we have *frequency modulation* (FM).

2.2 AMPLITUDE MODULATION THEORY

In order to simplify the derivation of the equation for an amplitude modulated wave, we make the simplification that the modulating signal is a sinusoid of angular frequency ω_s and that the carrier signal to be modulated (also sinusoidal) has an angular frequency ω_c .

Let the instantaneous carrier current be

$$i = A \sin \omega_c t \quad (2.2.1)$$

where A is the amplitude.

The amplitude modulated carrier must have the form

$$i = [A + g(t)] \sin \omega_c t \quad (2.2.2)$$

where

$$g(t) = B \sin \omega_s t \quad (2.2.3)$$

is the modulating signal. Then

$$i = (A + B \sin \omega_s t) \sin \omega_c t \quad (2.2.4)$$

The waveform is shown in Figure 2.1.

The current may then be expressed as

$$i = (A + kA \sin \omega_s t) \sin \omega_c t \quad (2.2.5)$$

where

$$k = \frac{B}{A}. \quad (2.2.6)$$

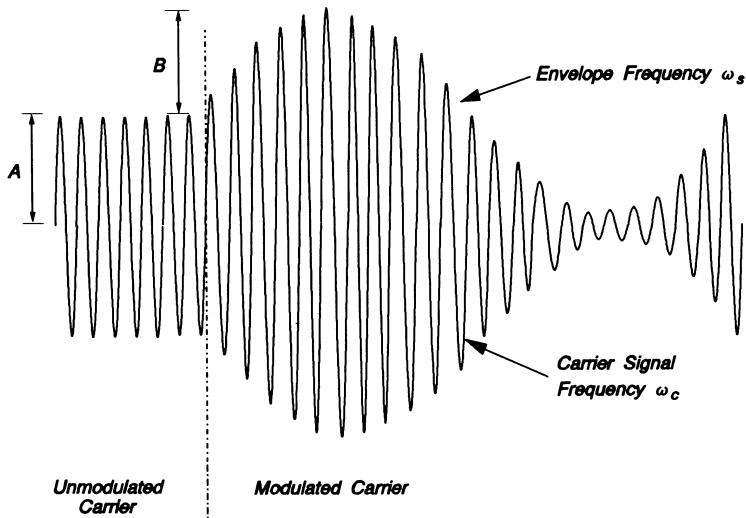


Figure 2.1. Amplitude modulated wave: the carrier frequency remains sinusoidal at ω_c while the envelope varies at frequency ω_s .

The factor k is called the *depth of modulation* and may be expressed as a percentage. Simplification of Equation (2.2.5) gives

$$i = A \sin \omega_c t + \frac{kA}{2} [\cos(\omega_c - \omega_s)t - \cos(\omega_c + \omega_s)t] \quad (2.2.7)$$

The frequency spectrum is shown in Figure 2.2.

From Equation (2.2.7) it is evident that modulated carrier current has three distinct frequencies present: the carrier frequency ω_c , the frequency equal to the difference between the carrier frequency and the modulating signal frequency

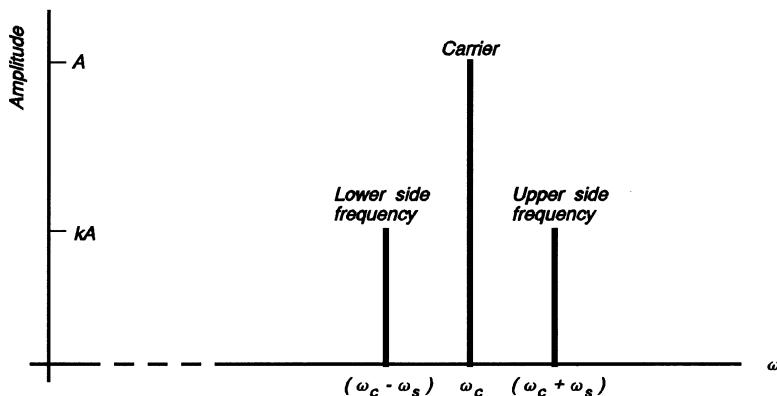


Figure 2.2. Frequency spectrum of the AM wave of Figure 2.1. Note that there are three distinct frequencies present.

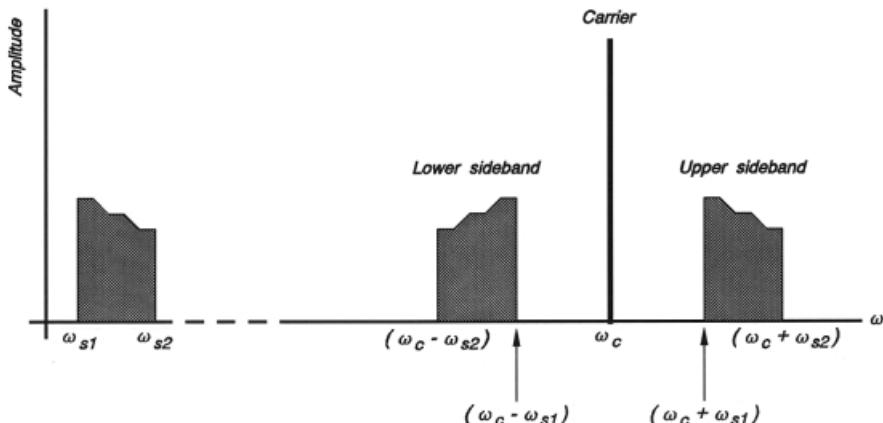


Figure 2.3. Frequency spectrum of the AM wave when the single frequency modulating signal is replaced by a band of audio frequencies. Note that the information in the signal resides only in the sidebands.

($\omega_c - \omega_s$), and the frequency equal to the sum of the carrier frequency and the modulating signal frequency ($\omega_c + \omega_s$). The difference and sum frequencies are called the “lower” and “upper” sidebands, respectively.

To make the situation more realistic, let us assume that the modulating signal is speech which contains frequencies between ω_{s1} and ω_{s2} . Then it follows from Equation (2.2.7) that the sum and difference terms will yield a band of frequencies symmetrical about the carrier frequency, as shown in Figure 2.3.

Figure 2.4 shows how two audio signals which would normally interfere with each other, when transmitted simultaneously through the same medium, can be kept separate by choosing suitable carrier frequencies in a modulating scheme. This method of transmitting two or more signals through the same medium simultaneously is referred to as *frequency-division multiplex* and will be discussed in detail in Chapter 9.

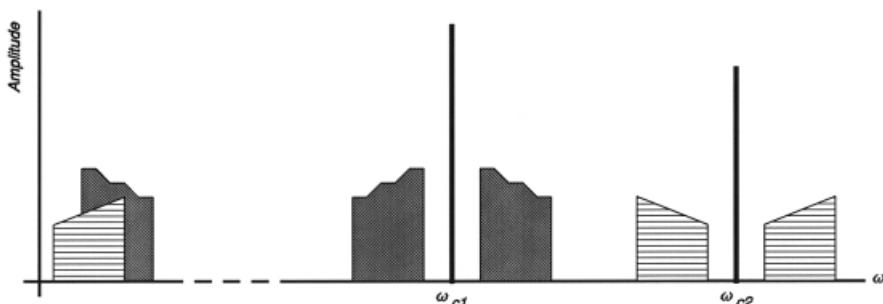


Figure 2.4. The diagram illustrates how two audio-frequency sources, which would normally interfere with each other, can be transmitted over the same channel with no interaction.

2.3 SYSTEM DESIGN

The choice of carrier frequency for a radio transmitter is largely determined by government regulations and international agreements. It is evident from Figure 2.4 that, in spite of frequency division multiplexing, two stations can interfere with each other if their carrier frequencies are so close that their sidebands overlap. In theory, every transmitter must have a unique frequency of operation and sufficient bandwidth to ensure no interference with others. However, bandwidth is limited by considerations such as cost and the sophistication of the transmission technique to be used so that, in practice, two radio transmitters may operate on frequencies which would normally cause interference so long as they propagate their signals within specified limits of power and are located (geographically) sufficiently far apart. The location as well as the power transmitted by each transmitter is monitored and controlled by the government.

Once the carrier frequency is assigned to a radio station, it is very important that it maintains that frequency as constant as possible. There are two reasons for this: (1) if the carrier frequency were allowed to drift then the listeners would have to re-tune their radios from time to time to keep listening to that station, which would be unacceptable to most listeners; (2) if a station drifts (in frequency) towards the next station, their sidebands would overlap and cause interference. The carrier signal is usually generated by an oscillator, but to meet the required precision of the frequency it is common practice to use a crystal-controlled oscillator. At the heart of the crystal-controlled oscillator is a quartz crystal cut and polished to very tight specifications which maintains the frequency of oscillation to within a few hertz of its nominal value. The design of such an oscillator can be found in Section 2.4.6. Figure 2.5 is a block diagram of a typical transmitter.

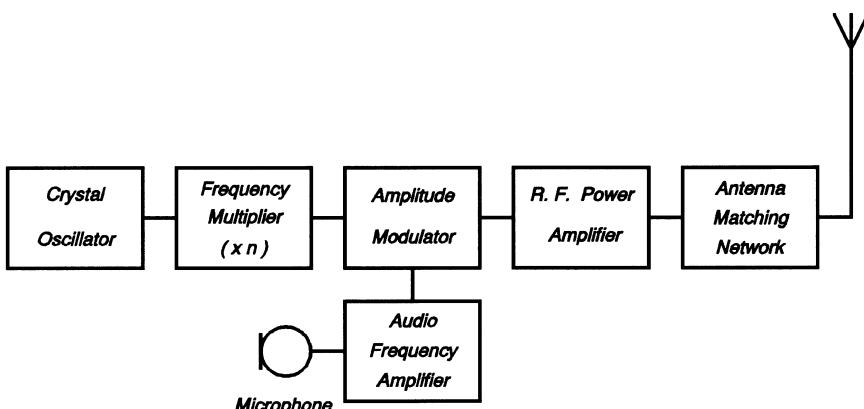


Figure 2.5. Block diagram showing the components which make up the AM transmitter.

2.3.1 Crystal-Controlled Oscillator

The purpose of the crystal oscillator is to generate the carrier signal. To minimize interference with other transmitters, this signal must have extremely low levels of distortion so that the transmitter operates at only one frequency. As discussed earlier, the frequency must be kept within very tight limits, usually within a few hertz in 10^7 Hz. It is difficult to design an ordinary oscillator to satisfy these conditions, so it is common practice to use a quartz crystal to enhance the frequency stability and to reduce the harmonic distortion products.

The quartz crystal undergoes a change in its physical dimensions when a potential difference is applied across two corresponding faces of the crystal. If the potential difference is an alternating one, the crystal will vibrate and exhibit the phenomenon of resonance. For a crystal, the range of frequency over which resonance is possible is very narrow, hence the frequency stability of the crystal-controlled oscillators is very high. In general, the larger the physical size of the crystal, the lower the frequency at which it resonates. Thus a high-frequency crystal is necessarily small, fragile, and has low reliability. To generate a high-frequency carrier, it is common practice to use a low-frequency crystal to obtain a signal at a subharmonic of the required frequency and to use a *frequency multiplier* to increase the frequency. Figure 2.5 shows that the crystal-controlled oscillator is followed by a frequency multiplier.

2.3.2 Frequency Multiplier

The purpose of the frequency multiplier is to accept an incoming signal of frequency f_c/n , where n is an integer, and to produce an output at a frequency f_c . A frequency multiplier can have a single stage of multiplication or it can have several stages. The output of the frequency multiplier goes to the carrier input of the *amplitude modulator*.

2.3.3 Amplitude Modulator

The amplitude modulator has two inputs, the first being the carrier signal generated by the crystal oscillator and multiplied by a suitable factor, and the second being the modulating signal (voice or music) which is represented in Figure 2.5 by the single frequency f_s . In reality, the frequencies present in the modulating signal are in the audio range 20–20,000 Hz. The output from the amplitude modulator consists of the carrier, the lower and upper sidebands.

2.3.4 Audio Amplifier

The audio amplifier accepts its input from a microphone and supplies the necessary gain to bring the signal level to that required by the amplitude modulator.

2.3.5 Radio-Frequency Power Amplifier

The power level at the output of the modulator is usually in the range of watts and the power required to broadcast the signal effectively is in the range of tens of kilowatts. The radio-frequency amplifier provides the power gain as well as the necessary impedance matching to the *antenna*.

2.3.6 Antenna

The antenna is the circuit element that is responsible for converting the output power from the transmitter amplifier into an electromagnetic wave suitable for efficient radiation in free space. Antennae take many different physical forms determined by the frequency of operation and the radiation pattern desired. For broadcasting purposes, an antenna that radiates its power uniformly to its listeners is desirable, whereas in the transmission of signals where security is important (e.g. telephony), the antenna has to be as directive as possible to reduce the possibility of its reception by unauthorized persons.

2.4 RADIO TRANSMITTER OSCILLATOR

Perhaps the simplest way to introduce the phenomenon of oscillation is to describe a common experience of a public address system going unstable and producing an unpleasantly loud whistle. The system consists of a microphone, an amplifier and a loudspeaker (or loudspeakers) as shown in Figure 2.6. The amplified sound from the

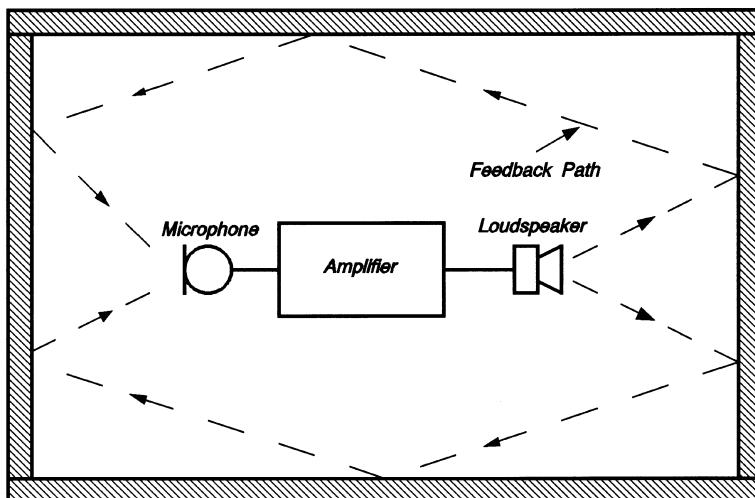


Figure 2.6. The diagram illustrates how acoustic feedback can cause a public address system to go unstable, turning the system into an oscillator.

loudspeaker may be reflected from walls and other surfaces and reach the microphone. If the reflected sound is louder than the original then it will in turn produce a louder output at the loudspeaker which will in turn produce an even louder signal at the microphone. It is fairly clear that this state of affairs cannot continue indefinitely; the system reaches a limit and produces the characteristic loud whistle. Immediate steps have to be taken to ensure that the sound level reaching the microphone is less than that required to reach the self-sustained value. If, on the other hand, we are interested in the generation of an oscillation, then the study of the characteristics of the amplifying element, the conditions under which the feedback takes place, the frequencies present in the signal and the optimization of the system to achieve specified performance goals are in order.

The electronic oscillator is a particular example of a more general phenomenon of systems which exhibit a periodic behavior. A mechanical example is the pendulum which will perform simple harmonic motion at a frequency determined by its length and the acceleration constant due to gravity, g , if the energy it loses per cycle is replaced from an outside source. In the case of the pendulum used in clocks, the source of energy may be a wound-up spring or a weight whose potential energy is transferred to the pendulum. The solar system with planets performing cyclical motion around the sun is another example of an oscillator, although this time there is no periodic input of energy because the system is virtually lossless.

Three theoretical approaches to oscillator design are presented below. The first is based on the idea of setting up a “lossless” system by canceling the losses in an LC circuit due to the presence of (positive) resistance by using a negative resistance. The second is based on feedback theory. The third is based on the concept of *embedding* an active device and the optimization of the power output from the oscillator.

2.4.1 Negative Conductance Oscillator

Consider the circuit shown in Figure 2.7. The externally applied current and the corresponding voltage are related to each other by

$$I = \left(G_0 + G_n + sC + \frac{1}{sL} \right) (V) \quad (2.4.1)$$

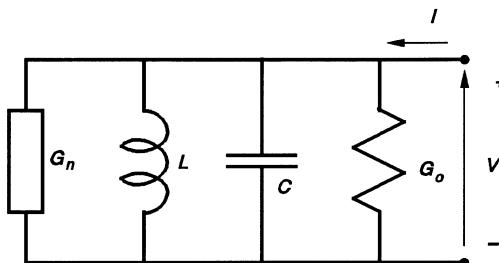


Figure 2.7. The negative conductance oscillator has a negative conductance generating signal power which is dissipated in the (positive) conductance. The components L and C determine the frequency of the signal. An alternate statement is that the negative conductance cancels all the losses in the circuit. It then oscillates losslessly at a frequency determined by L and C .

where G_0 is the load conductance, G_n is the negative conductance, I is current, V is voltage, s is the complex frequency, C is capacitance, and L is inductance. If the circuit is that of an oscillator, the external excitation current must be zero since an oscillator does not require an excitation current. Hence

$$0 = \left(G_0 + G_n + sC + \frac{1}{sL} \right) (V). \quad (2.4.2)$$

For a non-trivial solution, V is non-zero, therefore

$$G_0 + G_n + sC + \frac{1}{sL} = 0 \quad (2.4.3)$$

which gives the quadratic equation

$$s^2 CL + sL(G_0 + G_n) + 1 = 0. \quad (2.4.4)$$

The solution is then

$$s_1, s_2 = -\frac{(G_0 + G_n)}{2C} \pm \sqrt{\frac{(G_0 + G_n)^2}{4C^2} - \frac{1}{LC}} \quad (2.4.5)$$

when

$$|G_n| = G_0 \quad (2.4.6)$$

that is, the system is lossless. Equation (2.4.5) becomes

$$s_1, s_2 = \sqrt{-\frac{1}{LC}} = j\omega \frac{1}{\sqrt{LC}} \quad (2.4.7)$$

which is the resonant frequency for the tuned circuit. The circuit will continue to oscillate at this frequency as if it were in perpetual motion.

A number of devices exhibit negative conductance under appropriate bias conditions and may be used in the design of practical oscillators of this type. These include tunnel diodes, pentodes (N-type negative conductance), uni-junction transistors and silicon-controlled rectifier (S-type).

The voltage-current characteristics of N- and S-type negative conductances are shown in Figures 2.8(a) and (b), respectively.

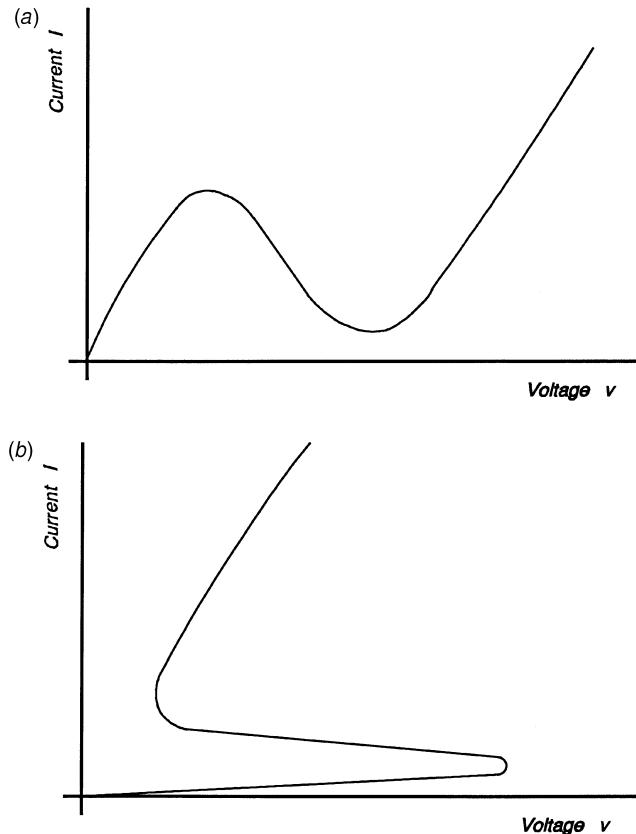


Figure 2.8. (a) Characteristics of an *N*-type negative conductance device. The device has a negative conductance in the region where the slope of the curve is negative. Examples of practical devices which have such characteristics are the tunnel diode and the tetrode. (b) Characteristics of an *S*-type negative conductance device. The device has a negative conductance in the region where the slope of the curve is negative. Examples of practical devices which have such characteristics are the four-layer diode and the silicon controlled rectifier.

2.4.2 Classical Feedback Theory

Consider the system shown in Figure 2.9 where A is the gain of an amplifier and β represents the transfer function of the feedback path. E_s is the signal applied to the input and E_o is the output of the system [1].

In the derivation that follows, it is necessary to make the following assumptions:

- (1) the input impedances of both the amplifier and the feedback network are infinite and their output impedances are zero,
- (2) both A and β are complex quantities.

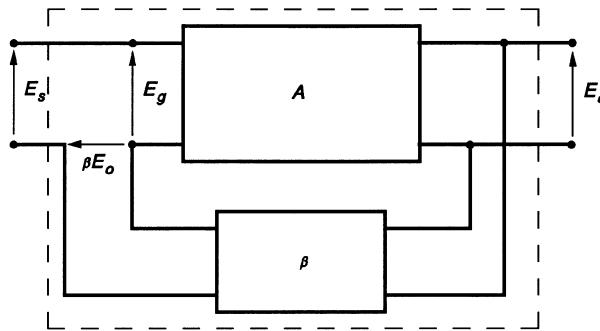


Figure 2.9. Classical feedback system with gain A and feedback factor β .

The gain of the amplifier alone is

$$A = \frac{E_o}{E_g}. \quad (2.4.8)$$

Application of Kirchhoff's Voltage Law (KVL) at the input gives

$$E_g = E_s + \beta E_o. \quad (2.4.9)$$

Substituting Equation (2.4.8) into Equation (2.4.9) gives

$$E_o = A(E_s + \beta E_o) \quad (2.4.10)$$

from which we obtain

$$\frac{E_o}{E_s} = \frac{A}{1 - \beta A}. \quad (2.4.11)$$

Since the E_s and E_o are the input and output, respectively, of the system as a whole, we can define this as A' where

$$A' = \frac{E_o}{E_s} = \frac{A}{(1 - \beta A)}. \quad (2.4.12)$$

Three separate conditions must be considered that depend on the value of the denominator of Equation (2.4.12)

- (1) *Positive feedback.* If the modulus of $(1 - \beta A)$ is less than unity, then the gain of the system A' is greater than the gain of the amplifier A and therefore the effect of the feedback is said to be positive.
- (2) *Negative feedback.* If the modulus of $(1 - \beta A) > 1$, then $A' < A$.

- (3) *Oscillation.* If the modulus of $(1 - \beta A) = 0$ then the gain A' is infinite because with no input ($E_s = 0$) there is still an output. In fact the system is supplying its own input and

$$\beta A = 1. \quad (2.4.13)$$

It must be noted that the waveform of the signal need not be sinusoidal and in fact it can take any form so long as the waveform of the signal that is fed back, βE_o , is identical to the signal E_o . However, the object of this exercise is to generate a carrier for a telecommunication system and therefore only sinusoidal signals are acceptable – any other waveform will generate other carriers (harmonics of the fundamental) and cause interference with transmissions of other stations.

2.4.3 Sinusoidal Oscillators

Since both A and β are complex quantities, condition (3) implies

$$|\beta A| = 1. \quad (2.4.14)$$

Stated in words, the magnitude of the *loop gain* must equal unity, and

$$\angle \beta A = 0, 2\pi, 4\pi, \text{etc.} \quad (2.4.15)$$

Again, in words, the *loop-gain phase shift* must be zero or an integral multiple of 2π radians. The condition given in Equation (2.4.13), which implies Equations (2.4.14) and (2.4.15), is known as the *Barkhausen Criterion*.

These two conditions must exist simultaneously for sinusoidal oscillation to occur.

2.4.4 General Form of the Oscillator

An oscillator circuit shown in Figure 2.10 [2] and an equivalent circuit is as shown in Figure 2.11, where the amplifying element is replaced by a voltage-controlled voltage source in series with a resistance R_o to simulate the output resistance of the element. The amplifying element may be a tube, a transistor or an operational amplifier.

The load seen by the amplifier is

$$Z_L = \frac{Z_2(Z_1 + Z_3)}{(Z_1 + Z_2 + Z_3)}. \quad (2.4.16)$$

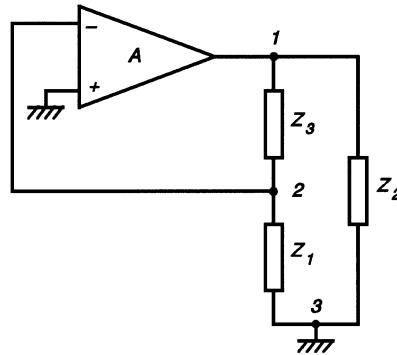


Figure 2.10. Circuit diagram for a more generalized form of the oscillator.

The amplifier gain without feedback is

$$A = \frac{V_o}{V_{32}} = -\frac{A_v Z_L}{(R_o + Z_L)} \quad (2.4.17)$$

and the feedback constant is

$$\beta = \frac{Z_1}{(Z_1 + Z_3)}. \quad (2.4.18)$$

The loop gain is

$$\beta A = -\frac{A_v Z_1 Z_L}{(R_o + Z_L)(Z_1 + Z_3)}. \quad (2.4.19)$$

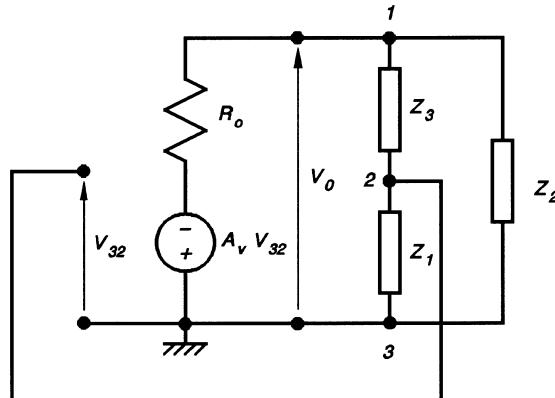


Figure 2.11. The equivalent circuit of the generalized form of the oscillator. R_o represents the output resistance of the amplifier.

Substituting for Z_L as defined in Equation (2.4.16), Equation (2.4.19) becomes

$$\beta A = -\frac{A_v Z_1 Z_2}{[R_o(Z_1 + Z_2 + Z_3) + Z_2(Z_1 + Z_3)]}. \quad (2.4.20)$$

For simplicity, we may assume that the impedances are lossless; hence

$$Z_1 = jX_1, \quad Z_2 = jX_2 \quad \text{and} \quad Z_3 = jX_3 \quad (2.4.21)$$

Then Equation (2.4.20) becomes

$$\beta A = -\frac{A_v X_1 X_2}{[jR_o(X_1 + X_2 + X_3) - X_2(X_1 + X_3)]}. \quad (2.4.22)$$

Recall that for oscillation to occur

$$1 - \beta A = 0. \quad (2.4.23)$$

This means that βA must be real and hence,

$$X_1 + X_2 + X_3 = 0 \quad (2.4.24)$$

that is,

$$X_2 = -(X_1 + X_3). \quad (2.4.25)$$

The expression for the loop gain becomes

$$\beta A = -A_v \frac{X_1}{X_2}. \quad (2.4.26)$$

Since $\beta A = 1$, it follows that X_1 and X_2 must have opposite signs; that is, if one of them is inductive, the other must be capacitive and X_3 can be capacitive or inductive, depending on the sign of $(X_1 + X_2)$. The two possibilities are shown in Figures 2.12 and 2.13, respectively.

The circuit shown in Figure 2.12 is better known as a *Colpitts oscillator*. The circuit is redrawn in Figure 2.12(b) to emphasize the symmetrical structure of the circuit.

The circuit shown in Figure 2.13 is better known as a *Hartley oscillator*.

From the point of view of the structure of the circuits, it can be seen that they are the same. It should be noted that the operational amplifier can be replaced by a tube or a transistor.

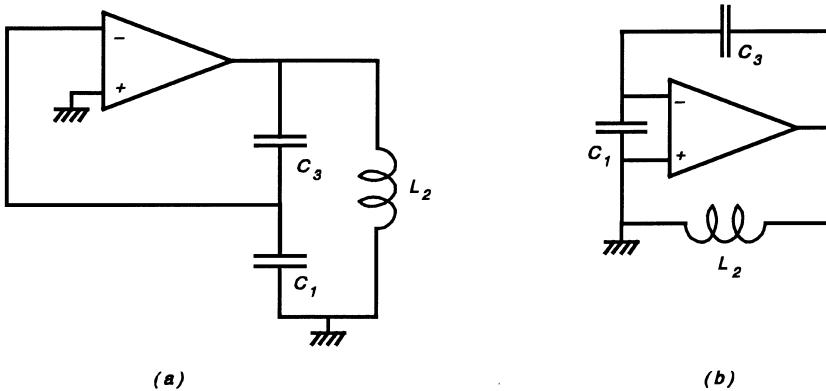


Figure 2.12. (a) The generalized form of the oscillator with two of the impedances replaced by capacitors and the third by an inductor to form a Colpitts oscillator. (b) The diagram in (a) has been redrawn to emphasize the symmetry of the circuit.

2.4.5 Oscillator Design for Maximum Power Output

A major flaw in the two previous designs is that they do not anticipate the necessity for the oscillator to supply power to a load. The theory of the design for maximum power output from an oscillator [3] is based on the characterization of the amplifying element (“active device”) as a two-port. A discussion of two-ports is beyond the scope of this book but may be found in any standard text on circuit theory.

A two-port can be described in terms of its terminal voltages and currents by four parameters: impedances, admittances, voltage ratios, and current ratios under constraints of open or short-circuit. Without limiting the generality, assume that the active device has been characterized in terms of the short-circuit admittance

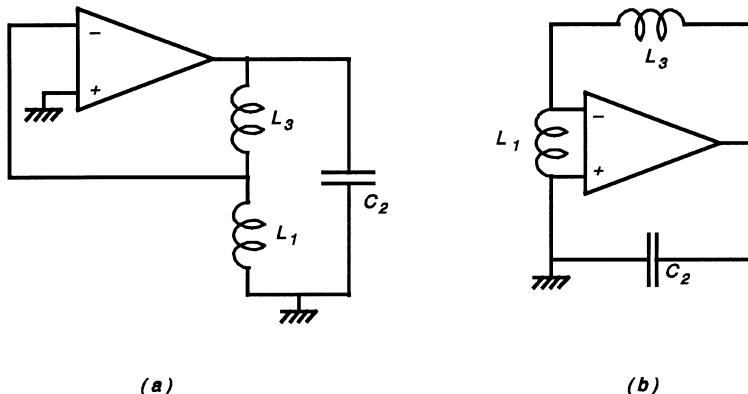


Figure 2.13. (a) The generalized form of the oscillator with two of the impedances replaced by inductors and the third by a capacitor to form a Hartley oscillator. (b) The diagram in (a) has been redrawn to emphasize the symmetry of the circuit.

parameters, or Y parameters, for short. Figure 2.14 shows the two-port and its terminal voltages and currents, which are assumed to be sinusoidal.

The Y parameters are functions of frequency and bias conditions, and in general, complex so that

$$Y_{11} = g_{11} + jb_{11}. \quad (2.4.27)$$

The total power entering the two-port is

$$P = V_1^* I_1 + V_2^* I_2. \quad (2.4.28)$$

The Y parameters and the terminal voltages and currents are related by

$$I_1 = Y_{11} V_1 + Y_{12} V_2 \quad (2.4.29)$$

and

$$I_2 = Y_{21} V_1 + Y_{22} V_2. \quad (2.4.30)$$

Substituting for I_1 and I_2 in Equation (2.4.28) gives

$$P = Y_{11}|V_1|^2 + Y_{22}|V_2|^2 + Y_{12}V_1^*V_2 + Y_{21}V_1V_2^*. \quad (2.4.31)$$

The ratio of the output voltage, V_2 , to the input voltage, V_1 , can be defined as

$$\frac{V_2}{V_1} = \mathbf{A} = A_R + jA_I. \quad (2.4.32)$$

The real power entering the two-port is

$$P_R = |V_1|^2 [g_{11} + g_{22}(A_R^2 + A_I^2) + (g_{12} + g_{21})A_R - (b_{12} + b_{21})A_I]. \quad (2.4.33)$$

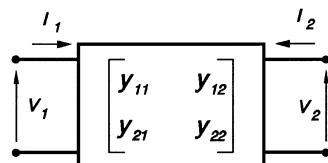


Figure 2.14. A two-port representation of an active device to be used in the design of an oscillator. Short-circuit admittance (Y) parameters are used in the design for convenience. Other parameters could be used in the description.

This can be rearranged as follows:

$$\frac{P_R}{g_{22}|V_1|^2} = \left[A_R + \frac{(g_{21} + g_{12})}{2g_{22}} \right]^2 + \left[A_I + \frac{(b_{21} - b_{12})}{2g_{22}} \right]^2 + \frac{4g_{11}g_{22} - (g_{21} + g_{12})^2 - (b_{21} - b_{12})^2}{4g_{22}^2}. \quad (2.4.34)$$

This equation is of the form:

$$z = (x - a)^2 + (y - b)^2 + c \quad (2.4.35)$$

and therefore it is that of a paraboloid in space with axes $P_R(g_{22}/|V_1|^2)$, A_R and A_I as shown in Figure 2.15.

It was assumed that real, positive power was supplied and dissipated in the two-port; therefore, it follows that negative values of power, as shown in Figure 2.15, must represent power generated by the two-port and dissipated in the surrounding or embedding circuit; that is, above the A plane, real power is supplied to the two-port, and below it the device supplies real power to the embedding circuit. Because the object of the exercise is to generate and supply real power to an external circuit, the most interesting part of Figure 2.15 is the part below the A plane. It is clear that movement towards the apex of the paraboloid represents increasing levels of power supplied by the “active” two-port and that the maximum power supplied occurs at the apex. We shall return to this remark when we consider the optimization of the power output.

The most general embedding circuit for the two-port is as shown in Figure 2.16 with each branch made up of a conductance in parallel with a susceptance. The susceptances can be considered as the tuned circuit which will determine the

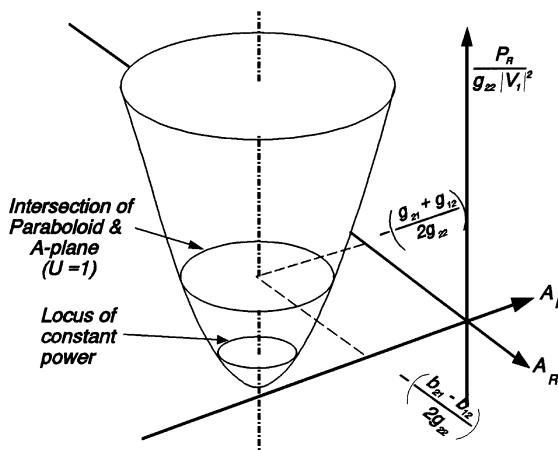


Figure 2.15. Three-dimensional representation of the output power of the oscillator as a function of the complex parameter A .

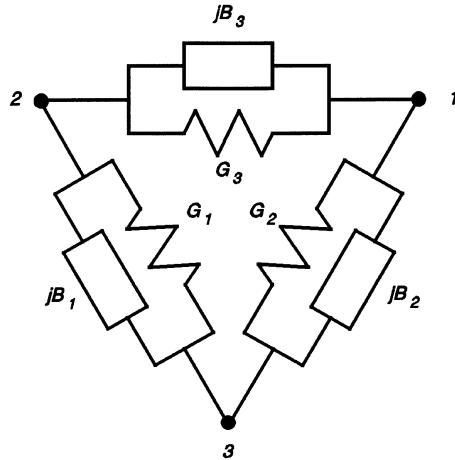


Figure 2.16. The general passive embedding circuit for a two-port.

frequency of oscillation and the conductances as the destination of the power generated by the active two-port.

The embedding network can also be described in terms of a two-port as follows:

$$I'_1 = (Y_2 + Y_3)V_1 - Y_3V_2 \quad (2.4.36)$$

$$I'_2 = -Y_3V_1 + (Y_1 + Y_3)V_2. \quad (2.4.37)$$

When the active device and the embedding are connected as shown in Figure 2.17, the composite circuit can be described by the two-port equations which are [Equations (2.4.29) + (2.4.36) and Equations (2.4.30) + (2.4.37)]:

$$I_1 + I'_1 = (Y_{11} + Y_2 + Y_3)V_1 + (Y_{12} - Y_3)V_2 \quad (2.4.38)$$

$$I_2 + I'_2 = (Y_{21} - Y_3)V_1 + (Y_1 + Y_3 + Y_{22})V_2. \quad (2.4.39)$$

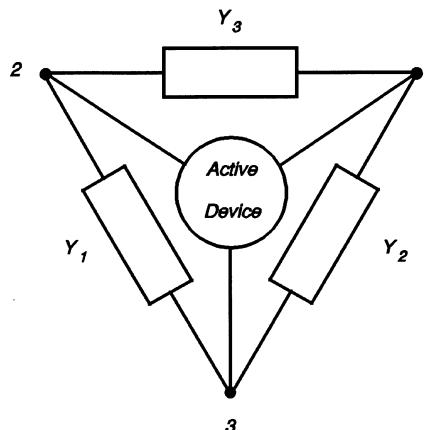


Figure 2.17. The active two-port is shown with the passive embedding connected.

For an oscillator, no external signal current is supplied at port 1 and therefore $I_1 + I'_1 = 0$. Similarly $I_2 + I'_2 = 0$. From Equation (2.4.32) we have

$$V_2 = V_1(A_R + jA_I) \quad (2.4.40)$$

From Equation (2.4.38) we have

$$V_1[Y_{11} + Y_2 + Y_3 + (A_R + jA_I)(Y_{12} - Y_3)] = 0 \quad (2.4.41)$$

and from Equation (2.4.39) we have

$$V_1[Y_{21} - Y_3 + (A_R + jA_I)(Y_1 + Y_3 + Y_{22})] = 0. \quad (2.4.42)$$

For non-trivial values of V_1 , real and imaginary values of Equations (2.4.41) and (2.4.42) are separately equal to zero; that is,

$$g_{11} + G_2 + G_3 + A_R(g_{12} - G_3) - A_I(b_{12} - B_3) = 0 \quad (2.4.43)$$

$$b_{11} + B_2 + B_3 + A_R(b_{12} - B_3) + A_I(g_{12} - G_3) = 0 \quad (2.4.44)$$

$$g_{21} - G_3 + A_R(G_1 + G_3 + g_{22}) - A_I(B_1 + B_3 + b_{22}) = 0 \quad (2.4.45)$$

and

$$b_{21} - B_3 + A_R(B_1 + B_3 + b_{22}) + A_I(G_1 + G_3 + g_{22}) = 0. \quad (2.4.46)$$

Equations (2.4.43) to (2.4.46) can be written in the form of a matrix as follows:

$$\begin{bmatrix} A_R & 0 & (A_R - 1) & -A_I & 0 & -A_I \\ A_I & 0 & A_I & A_R & 0 & (A_R - 1) \\ 0 & 1 & (1 - A_R) & 0 & 0 & A_I \\ 0 & 0 & -A_I & 0 & 1 & (1 - A_R) \end{bmatrix} \begin{bmatrix} G_1 \\ G_2 \\ G_3 \\ B_1 \\ B_2 \\ B_3 \end{bmatrix} = \begin{bmatrix} -g_{21} & -\text{Re}(Ay_{22}) \\ -b_{21} & -\text{Im}(Ay_{22}) \\ -g_{11} & -\text{Re}(Ay_{12}) \\ -b_{11} & -\text{Im}(Ay_{12}) \end{bmatrix} \quad (2.4.47)$$

All the terms in the matrix are known except G_1 , G_2 , G_3 , B_1 , B_2 and B_3 ; that is, there are six unknowns but only four equations so a unique solution cannot be found unless arbitrary values are chosen for at least two of the unknowns. Fortunately, an oscillator normally has only one conductive load and therefore two of the three conductances can be set to zero. The matrix equation can then be solved for one conductance and three susceptances.

2.4.6 Crystal-Controlled Oscillator

The oscillator used in a transmitter has to have a very tight tolerance on the stability of its frequency. This is necessary if interference between radio stations is to be

avoided. The drift of the frequency of an ordinary *LC* oscillator, for example, makes it unsuitable for this purpose. Greater frequency stability can be achieved by using a crystal as a part of the oscillator circuit [4]. In Section 2.3.1, the behavior of the crystal when it is excited by an ac signal was discussed. It is evident that, since the crystal reacts to electrical excitation, it must be possible to devise an electrical circuit made up of inductors, resistors and capacitors whose frequency characteristics are approximately those of the crystal. Such a circuit is shown in Figure 2.18. The approximate circuit is reasonably accurate at frequencies close to the resonant frequency. Over a larger frequency range a more complicated equivalent circuit has to be used.

Typical values of the components of the equivalent circuit are $C = 0.0154 \text{ pF}$, $R = 8 \Omega$, $L = 0.0165 \text{ H}$, $C_o = 4.55 \text{ pF}$. The capacitance C_o is due largely to the electrodes which are attached to the crystal. The crystal will therefore resonate in the series mode at a frequency ω_s where

$$\omega_s^2 = \frac{1}{LC} \quad (2.4.48)$$

which gives $f_s = 9.984 \times 10^6 \text{ Hz}$. It will resonate in the parallel mode at an angular frequency given approximately by

$$\omega_p^2 = \frac{1}{L \left[\frac{CC_o}{(C + C_o)} \right]} \quad (2.4.49)$$

which gives a resonant frequency, $f_p = 10.001 \times 10^6 \text{ Hz}$ – a change of less than 0.2%. The corresponding quality factor of the crystal is then $Q_o = 130,000$.

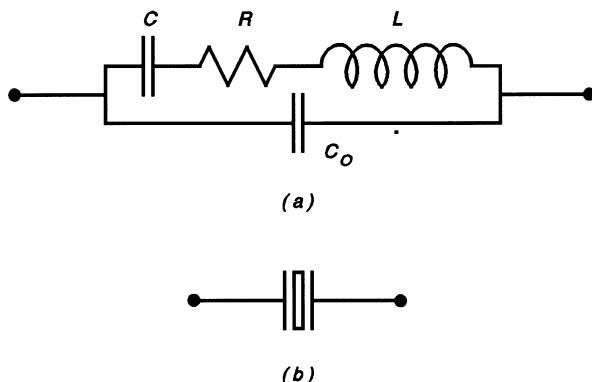


Figure 2.18. (a) The equivalent circuit of the crystal and its package. (b) The electrical symbol for the crystal.

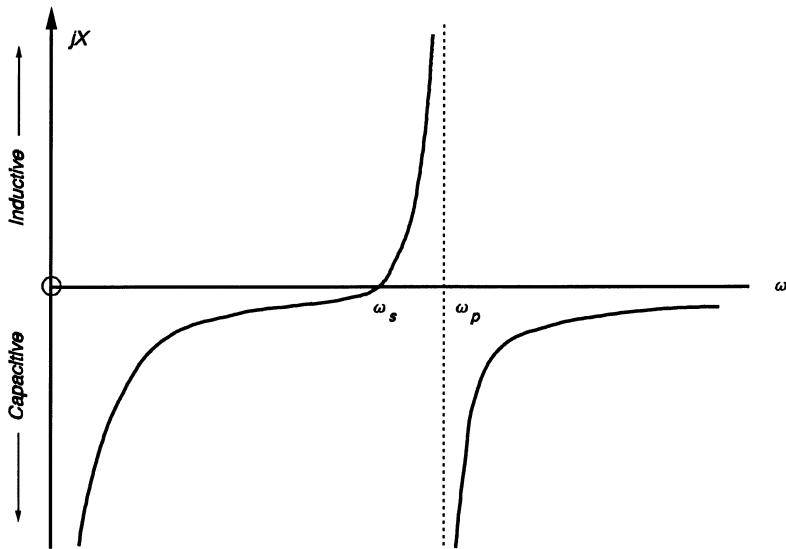


Figure 2.19. The reactance characteristics of the crystal. Note that this is not to scale.

Figure 2.19 shows the reactance of the crystal plotted against frequency. It should be noted that the reactance of the crystal is inductive over a narrow band of frequency and also that both reactance and frequency are not to scale.

Figure 2.20 shows a typical crystal-controlled oscillator. The crystal is substituted for one of the inductors in what would otherwise be classified as a Hartley oscillator. This type of crystal-controlled oscillator is called a *Pierce oscillator*. Similarly, the crystal-controlled oscillator corresponding to the Colpitts variety is called a *Miller oscillator*. In the circuit shown in Figure 2.20, the active element is a field-effect transistor whose gate-to-drain capacitance plus stray capacitance constitute C_3 .

The very high Q_o of the crystal ensures that the oscillator has an extremely limited range of frequencies in which it can continue to oscillate. Various other measures may be taken to improve the frequency stability, such as placing the crystal in a temperature-controlled environment and the Q factor can be enhanced by evacuating the glass envelope which protects it.

High precision oscillators are invariably connected to their load through a buffer amplifier. This ensures that variations in the load do not affect the operation of the oscillator.

2.5 FREQUENCY MULTIPLIER

The purpose of the frequency multiplier is to raise the frequency generated by the crystal-controlled oscillator to the value required for the transmitter carrier. As explained earlier, it is not possible to obtain physically robust crystals at high

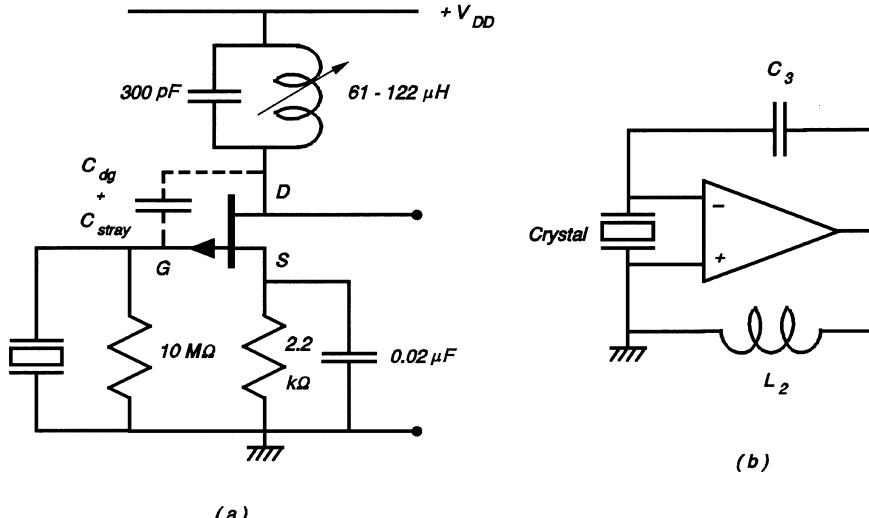


Figure 2.20. (a) A Hartley oscillator with one of the inductors replaced by a crystal. This circuit is called a *Pierce oscillator*. The field-effect transistor may be replaced by any other suitable active device. (b) The equivalent circuit of the Pierce oscillator demonstrating its symmetrical structure.

frequency since their physical size gets smaller as the frequency of oscillation gets higher. The standard technique is therefore to use a crystal to generate a signal at a frequency which is a subharmonic of the required carrier frequency and then to raise the frequency up to the required value using a cascade of frequency multipliers.

A useful analog of a frequency multiplier is a child's swing. With a child on the swing, the adult must give it a push to get the swing into operation. Subsequent to that, further supplies of energy must take place at a frequency determined by the length of the swing and the gravitational constant of acceleration, g . It is also necessary to supply the energy at a point in time when it enhances the swinging action rather than oppose it; on average, the adult will have to supply energy equal to that lost during the cycle to maintain a constant amplitude. If the energy supplied per cycle is less than the energy lost, the amplitude of the swing will decrease to a smaller value so as to restore the energy balance. If the energy supplied per cycle is greater than that lost per cycle, the amplitude will grow to a new steady-state value. The motion of the child will be very nearly a simple harmonic one if the total energy stored in the system is large compared to the energy supplied by the adult, that is, the system Q has to be large if the child is to execute a near-sinusoidal motion. The most important point of this analog is that the energy does not have to be supplied at the same frequency as the swing: it can be supplied at a subharmonic frequency, that is, the push can be given every other cycle of the swing or every third cycle or higher so long as enough energy is supplied to maintain the energy balance. When the push occurs every other cycle, it is clear that the output of the system is at twice the frequency of the input – this is a frequency multiplier with a multiplication

TABLE 2.1 Swing Analogy for a Frequency Multiplier

Swing, Including Child	Frequency Multiplier
Adult (timing)	Input signal source
Adult (energy transferred to child)	DC power supply
Length of swing, l , and gravitational constant, g	Inductance, L , and capacitance, C
Air resistance and bearing friction	Energy loss in R
Frequency $f = \frac{1}{2\pi\sqrt{\frac{l}{g}}}$	Frequency $f = \frac{1}{2\pi\sqrt{LC}}$
Amplitude of swing	Amplitude of voltage or current in tank circuit

factor of two. When the energy is supplied every third cycle, the multiplication factor of three is obtained, and so on. Evidently, there is a limit on how high the multiplication factor can be and it is determined by the amount of variation in the amplitude of the swing which can be tolerated. Table 2.1 shows a comparison of the swing and the frequency multiplier.

Figure 2.21 shows a typical frequency multiplier. Energy is fed into it by applying a suitable positive pulse to the base of the transistor. This causes the transistor to conduct momentarily, that is, current flows from the direct current (dc) power supply through the inductor and a finite amount of energy is stored in the inductor. The current flow is shut off when the input pulse ends and the transistor is essentially an open-circuit. The energy stored in the magnetic field of the inductor is transformed into energy stored in the electric field of the capacitor. The transformation of energy from one form to another and back again would continue indefinitely in a sinusoidal form if the system were lossless and this would take place at a frequency determined by the values of the inductance and capacitance. The resistance R_L represents the losses in the system – the amplitude of the sinusoid will decay with time. The steady-state amplitude of the voltage or current will be determined by the equaliza-

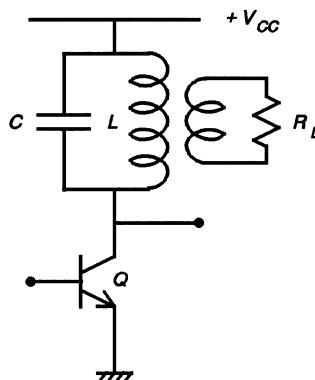


Figure 2.21. A class-C amplifier to be used, with minor modifications, as a frequency multiplier.

tion of the energy input and the energy output (loss). Subsequent to the initial input pulse, all input pulses must be timed to enhance rather than oppose the stored energy of the system. It is, of course, not necessary to have an input pulse for every cycle of the output; the input pulse can be supplied once for every two, three or greater number of cycles. When the input frequency is the same as the output, the frequency multiplier (multiplication factor = 1) is simply a class-C amplifier. Since a class-C amplifier represents the simplest frequency multiplier, the design of a class-C amplifier will be discussed next.

2.5.1 Class-C Amplifier

In a class-C amplifier, the current in the active device flows for a period much less than π radians of the output waveform. The active device current waveform is therefore highly non-sinusoidal. In a class-A or -B amplifier this would give a correspondingly non-sinusoidal output. However, in the class-C amplifier the collector load consists of a parallel LC circuit which is tuned to the frequency of the input signal. The tuned circuit (sometimes referred to as a *tank* circuit) presents a very high impedance at the resonant frequency to the collector of the transistor and hence a high gain is obtained at this frequency. Other frequencies, such as harmonics of the input frequency, are attenuated. Therefore, the output is very nearly sinusoidal.

In a practical class-C amplifier, the base of the transistor is held at a voltage such that the transistor is in the off state. The input signal brings the transistor into conduction at its positive peaks and causes enough current to flow in the inductor to store energy equal to that dissipated in the load resistor R_L and at other sites in the circuit. When the input signal drops below the threshold of the transistor, it switches off and the LC tank oscillates freely with a sinusoidal waveform at the resonant frequency of the tank circuit. Because of the dissipation in the circuit, the waveform is actually an exponentially damped sinusoid as shown in Figure 2.22.

The collector voltage, the base voltage and collector current waveforms are shown in Figure 2.23. Note that the quiescent value of the collector voltage waveform is V_{cc} and that its maximum amplitude is $2V_{cc}$.

The conversion efficiency of the amplifier is given by:

$$\eta = (\text{ac power output}) / (\text{dc power input}).$$

A class-C amplifier can have a relatively high conversion efficiency, usually about 85%, because the dc current flows for a very short part of the cycle and this happens when the collector voltage is at its lowest value. Thus the power lost in the transistor is minimal.

2.5.2 Converting the Class-C Amplifier into a Frequency Multiplier

To convert a class-C amplifier into a frequency multiplier with a multiplication factor of 2, the L and C of the tank circuit are chosen to resonate at $2\omega_0$ when the input signal frequency is ω_0 . Successful operation of the system demands that the Q factor

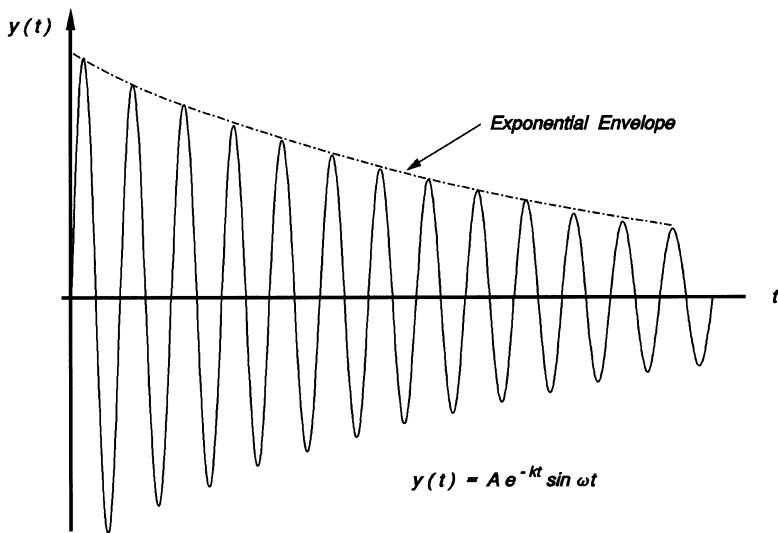


Figure 2.22. The output waveform of a class-C amplifier after a single pulse excitation. Note the sinusoidal waveform and the exponential decay of the envelope.

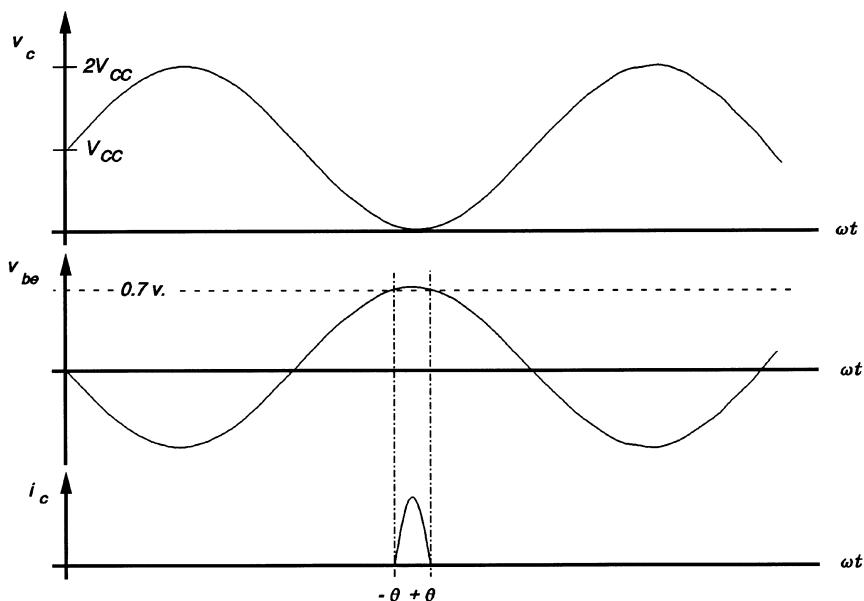


Figure 2.23. The collector voltage (v_c), base voltage (v_{be}) and collector current (i_c) of the class-C amplifier. Note that the base voltage need not be sinusoidal for the collector voltage to be sinusoidal.

of the circuit is sufficiently high so that the amount of damping at the end of the second cycle is negligible. Higher multiplication factors are possible with the damping problem progressively getting worse as the multiplication factor increases. Figure 2.24 shows the input current and output voltage waveforms of a frequency multiplier with a multiplication factor of 2. Note the effect of damping on the amplitude of the output voltage.

Example 2.5.1 Frequency Multiplier. A frequency multiplier with a multiplication factor of 2 is driven by an input current whose waveform can be assumed to be a half-sinusoid with a peak value of $200 \mu\text{A}$ at a frequency of $2/\pi \text{ MHz}$. The input current flows for a period corresponding to 5° of one cycle of the output waveform and the average input impedance over this period is 750Ω resistive. The following data applies to the circuit:

- (1) multiplier load = 50Ω resistive
- (2) transformer turns ratio = $15 : 1$
- (3) transformer coupling, $k = 1$
- (4) dc voltage supply = 15 V
- (5) current gain of the transistor = 100
- (6) loaded Q factor of the transformer primary = 50 (min).

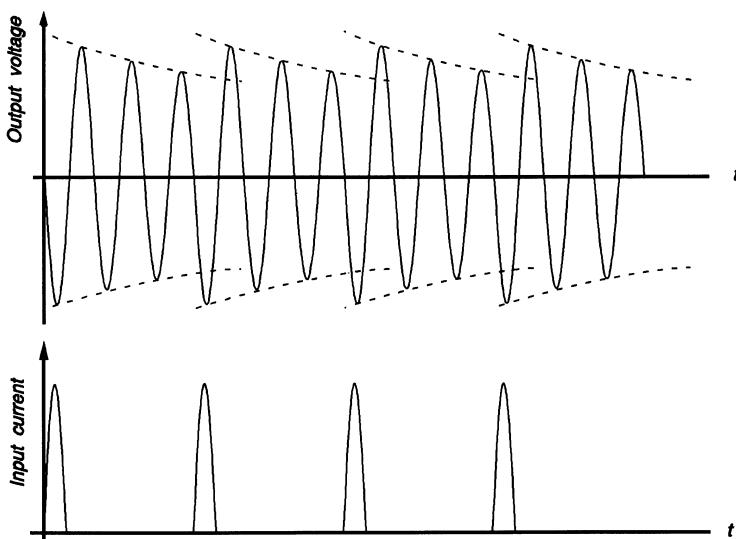


Figure 2.24. The output voltage waveform of a times-3 frequency multiplier and its driving current. The decay is exaggerated to show the effect of low Q factor.

Calculate the following:

- the value of the primary inductance
- the value of the tuning capacitance
- the impedance of the collector load at the output frequency
- the ac power dissipated in the load
- the power gain of the multiplier.

Solution. A frequency multiplier is essentially a class-C amplifier whose input is driven at a frequency which is a subharmonic of the output frequency. A suitable circuit is shown in Figure 2.25

$$\text{Input frequency} = 2\pi f = 4 \times 10^6 \text{ rad/s.}$$

$$\text{Output frequency} = 8 \times 10^6 \text{ rad/s.}$$

The input current and the output voltage waveforms are shown in Figure 2.26.

When the load is transferred to the primary the collector circuit is as shown in Figure 2.27 with the equivalent resistor having a value $n^2 R_L = (15)^2 \times 50 = 11.25 \text{ k}\Omega$.

- Assuming that the tank circuit is lossless, we can assign the loss in $n^2 R_L$ to the inductor and determine the series equivalent RL circuit (see Figure 2.28).

Equating the impedances,

$$\frac{j\omega L_p R_p}{R_p + j\omega L_p} = R_o + j\omega L_o. \quad (2.5.1)$$

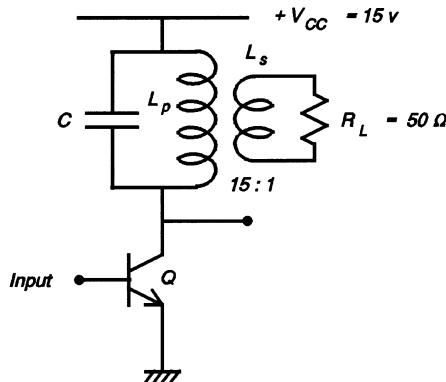


Figure 2.25. Circuit diagram for the frequency multiplier example.

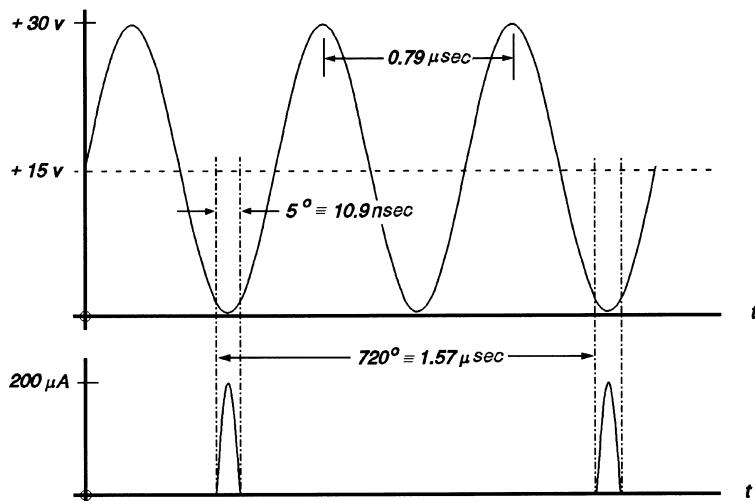


Figure 2.26. Collector and base current waveforms of the frequency multiplier example.

Rationalizing the left-hand side and equating real and imaginary parts gives

$$R_o = \frac{\omega^2 L_p^2 R_p}{R_p^2 + \omega^2 L_p^2} \quad (2.5.2)$$

and

$$L_o = \frac{L_p R_p^2}{R_p^2 + \omega^2 L_p^2}. \quad (2.5.3)$$

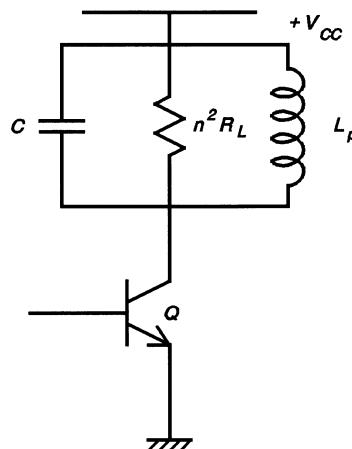


Figure 2.27. The frequency multiplier with the transformer load transferred to the primary.

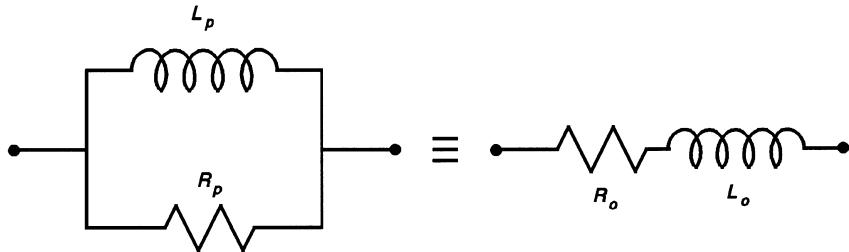


Figure 2.28. Transformation of a parallel RL to a series RL circuit.

The loaded Q factor of the tank circuit has a minimum value of 50; therefore

$$Q_o = 50 = \frac{\omega L_o}{R_o} = \frac{R_p}{\omega L_p} = \frac{n^2 R_L}{\omega L_p}. \quad (2.5.4)$$

The primary inductance is

$$L_p = \frac{n^2 R_L}{50\omega} = \frac{11.25 \times 10^3}{50 \times 8 \times 10^6} = 28.13 \mu\text{H}. \quad (2.5.5)$$

(b) For the tank circuit at resonance, then

$$\omega^2 = \frac{1}{L_p C} \quad (2.5.6)$$

and

$$C = \frac{1}{\omega^2 L_p} = 555 \text{ pF} \quad (2.5.7)$$

- (c) Because the parallel LC tank circuit is at resonance, it is actually an open-circuit. Therefore the load seen by the collector is $n^2 R_L = 11.25 \text{ k}\Omega$.
- (d) The base and collector currents are related by

$$i_c = \beta i_b = 20 \text{ mA peak}. \quad (2.5.8)$$

Current in the secondary = $n \times 20 \text{ mA peak}$. Since the system is in steady state, the energy dissipation associated with the current $n \times 20 \text{ mA peak}$ flowing in the load can be averaged over one cycle of the input signal (2 cycles of the output) and the ac output power calculated on that basis.

The collector current waveform is shown in Figure 2.29.

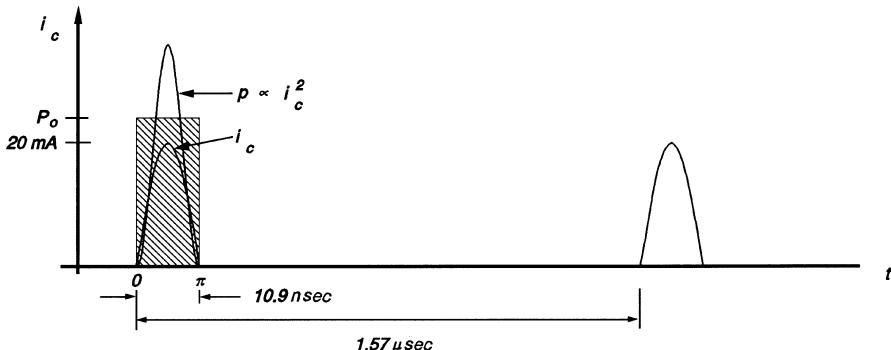


Figure 2.29. This diagram illustrates the various stages involved in the calculation of the output power of the frequency multiplier.

The instantaneous power in the collector circuit is

$$p(\theta) = i_c^2 R_L = (n \times 20 \times 10^{-3} \times \sin \theta)^2 R_L \quad (2.5.9)$$

and the average power is

$$P = \frac{1}{\pi} \int_0^\pi i_c^2 R_L d\theta. \quad (2.5.10)$$

The total energy stored in the tank circuit during \$n\$ radians (10.9 ns) of the input current is

$$W = (10.9 \text{ ns}) \times \frac{1}{\pi} \int_0^\pi i_c^2 R_L d\theta. \quad (2.5.11)$$

This amount of energy must be dissipated during a period corresponding to \$4\pi\$ radians (equivalent to \$720^\circ\$ or \$1.57 \mu\text{s}\$ in time) of the output waveform (two cycles) since the system is in equilibrium. Therefore,

$$P_o = \frac{10.9 \times 10^{-9}}{1.57 \times 10^{-6}} \times \frac{1}{\pi} \int_0^\pi i_c^2 R_L d\theta = 33.75 \text{ mW}. \quad (2.5.12)$$

(e) The input power is

$$P_{in} = \frac{1}{\pi} \int_0^\pi i_b R_{in} d\theta = 15 \mu\text{W}. \quad (2.5.13)$$

(f) The power gain is

$$G = 10 \log_{10} \frac{P_o}{P_{\text{in}}} = 33.5 \text{ dB.} \quad (2.5.14)$$

2.6 MODULATOR

In Section 2.3, the basic principles of modulation were discussed in general, and amplitude modulation was treated at some length. Briefly then, modulation is a means of achieving multiplexing, that is, providing a common conduit in which several messages can be transmitted at the same time without interference. In terms of amplitude modulation, the message (voice, code or music) is used to control the amplitude of a high-frequency sinusoidal signal (carrier). In this section, several circuits which can be used to generate amplitude modulation will be presented with the derivation of the appropriate design equations.

2.6.1 Square-law Modulator

From Equation (2.2.5) it can be seen that the modulated current has two components:

- (1) $A \sin \omega_c t$
- (2) $kA \sin \omega_s t \sin \omega_c t$.

These suggest that the circuit required to achieve amplitude modulation must generate the product of the carrier and the modulating signal frequency and add to this a suitably scaled carrier signal. The multiplication can be approximated by a non-linear element, such as a diode, and the addition by connecting the two sources in series. Since the diode has a voltage–current characteristic which is approximately a square-law relationship, it is commonly used in the square-law modulator. The modulator is sometimes called the diode modulator.

Consider the circuit shown in Figure 2.30.

Let

$$v_s = V_s \cos \omega_s t \quad (2.6.1)$$

and

$$v_c = V_c \cos \omega_c t \quad (2.6.2)$$

and assume the following:

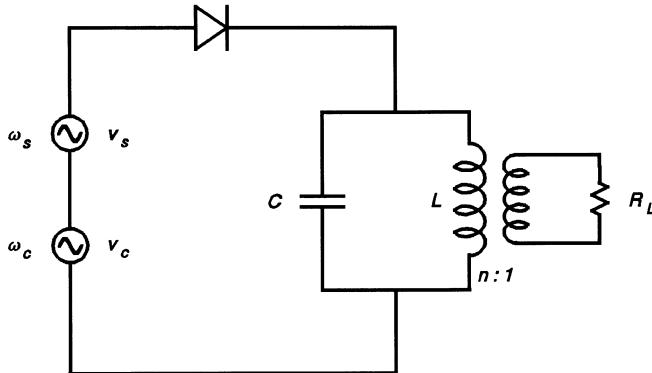


Figure 2.30. Using the nonlinearity of the diode to obtain amplitude modulation. The circuit is tuned to ω_c with sufficient bandwidth to select the carrier and sidebands and attenuate all spurious signals.

(1) Operation is at the resonant frequency of the circuit; hence

$$\omega^2 = \frac{1}{LC}. \quad (2.6.3)$$

(2) The diode current is given by

$$i = a_1 v + a_2 v^2 + \dots \quad (2.6.4)$$

Since the two voltage sources are connected in series,

$$v = V_c \cos \omega_c t + V_s \cos \omega_s t. \quad (2.6.5)$$

Substituting Equation (2.6.5) into Equation (2.6.4) gives

$$\begin{aligned} i = & a_1 V_c \cos \omega_c t + a_1 V_s \cos \omega_s t + a_2 V_c^2 \cos^2 \omega_c t + 2a_2 V_c V_s \cos \omega_c t \cos \omega_s t \\ & + a_2 V_s^2 \cos^2 \omega_s t + \dots \end{aligned} \quad (2.6.6)$$

Substituting

$$\cos^2 \phi = \frac{1}{2}(1 + \cos 2\phi) \quad (2.6.7)$$

into Equation (2.6.6) gives

$$\begin{aligned} i = & a_1 V_c \cos \omega_c t + a_1 V_s \cos \omega_s t + \frac{a_2 V_c^2}{2}(1 + \cos 2\omega_c t) + 2a_2 V_c V_s \cos \omega_c t \cos \omega_s t \\ & + \frac{a_2 V_s^2}{2}(1 + \cos 2\omega_s t). \end{aligned} \quad (2.6.8)$$

The LC tank circuit is tuned to ω_c and acts as a bandpass filter, and since $\omega_c \gg \omega_s$, signals of frequency ω_s , $2\omega_s$ and $2\omega_c$ will be filtered out, leaving

$$i = a_1 V_c \cos \omega_c t + 2a_2 V_s V_c \cos \omega_c t \cos \omega_s t. \quad (2.6.9)$$

When the load R_L is reflected into the primary, the voltage across the primary is then given by

$$v_p = a_1 n^2 R_L V_c \left(1 + \frac{2a_2 V_s}{a_1} \cos \omega_s t \right) \cos \omega_c t. \quad (2.6.10)$$

This is an amplitude modulated voltage wave, where the index of modulation is

$$k = \frac{2a_2 V_s}{a_1} \quad (2.6.11)$$

and

$$A = a_1 n^2 R_L V_c. \quad (2.6.12)$$

Thus

$$v = A(1 + k \cos \omega_s t) \cos \omega_c t. \quad (2.6.13)$$

2.6.2 Direct Amplitude Modulation Amplifier

Consider an amplifier into which a signal of frequency ω_c is fed [5, 6]. If the dc power supply of the amplifier is varied as a function of another signal of frequency ω_s , where $\omega_c \gg \omega_s$, then it is evident that the amplitude of the output voltage of the amplifier will be high when the dc supply voltage is high and low when the dc supply voltage is low. The basic circuit for a modulator operating on this principle is shown in Figure 2.31.

The circuit shows a class-C amplifier with its power supply connected through the secondary winding of a transformer. Node A is connected to ground by a capacitor whose value is chosen so that, at frequency ω_c , it is a short-circuit. From the point of view of the frequency ω_c , node A is at ground potential, as it should be for any amplifier. Assuming that T_1 is a 1 : 1 transformer, then with v_s applied to the primary, a voltage v_s will be developed in series with the dc supply voltage V_{cc} , such that

$$v_{cc} = V_{cc} + v_s. \quad (2.6.14)$$

Let

$$v_s = V_s \cos \omega_s t \quad (2.6.15)$$

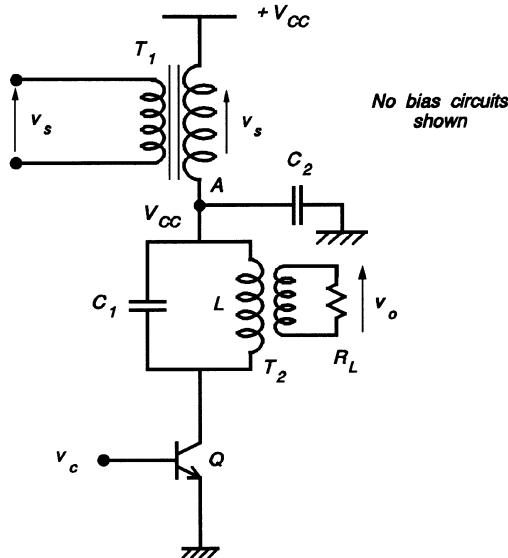


Figure 2.31. The circuit diagram of a direct AM amplifier. Note that, in this type of modulator, a high level of audio-frequency power is required to drive the circuit.

and

$$v_c = V_c \cos \omega_c t. \quad (2.6.16)$$

Then

$$v_{cc} = V_{cc}(1 + m \cos \omega_s t) \quad (2.6.17)$$

where $m = V_s/V_{cc}$.

Since the output voltage is proportional to the input and the supply voltage,

$$v_o = k v_{cc} V_c \cos \omega_c t \quad (2.6.18)$$

where k is a constant

$$v_o = k V_c V_{cc} (1 + m \cos \omega_s t) \cos \omega_c t \quad (2.6.19)$$

This is an amplitude modulated voltage. It is clear the C_2 must appear to be an open-circuit at the frequency ω_s because the dc supply source V_{cc} is, in fact, an ac short-circuit. The two constraints on C_2 namely that it must be a short-circuit at ω_c and an open-circuit at ω_s are easily satisfied because, in general $\omega_c \gg \omega_s$.

2.6.3 Four-Quadrant Analog Multiplier

An amplitude modulated signal can be obtained by adding the carrier signal to the product of the modulating (audio) signal and the carrier. Thus,

$$v_o = A \cos \omega_c t + kA \cos \omega_c t \times B \cos \omega_s t \quad (2.6.20)$$

There are a number of techniques which can be used to achieve this. One of the more practical of these is the variable transconductance method [7]. The output current of a common-emitter transistor amplifier depends on the input signal voltage at the base-emitter junction and the magnitude of the emitter resistance that can be controlled by varying the emitter current. The collector current is therefore proportional to the product of the input signal and the emitter current. Consider the circuit shown in Figure 2.32.

Assuming that the collector currents are approximately equal to the emitter currents, when V_k increases positively, the current I_1 increases at the expense of I_2 since the total current in the transistors Q_1 and Q_2 are fixed by the ideal current source, I .

Let

$$k = \frac{I_2}{I} \quad (2.6.21)$$

then

$$I_1 = (1 - k)I \quad (2.6.22)$$

where k is a current gain and $1 > k > 0$.

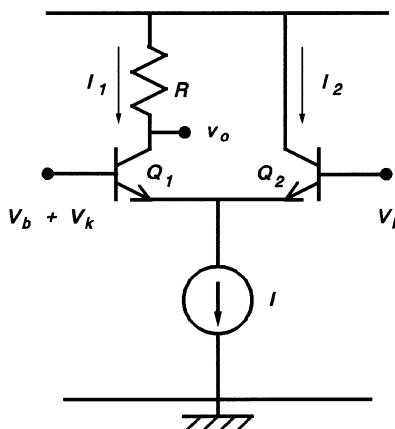


Figure 2.32. The long-tailed pair used as a variable transconductance element in an analog multiplier.

Clearly, the ideal current source, I , must have two components: a dc (bias) component and an ac (signal) component:

$$I = I_{dc} + I_c \cos \omega_c t \quad (2.6.23)$$

$$I_2 = kI = k(I_{dc} + I_c \cos \omega_c t). \quad (2.6.24)$$

Similarly, the voltage source V_k , when it is varied sinusoidally, must have two components: a dc (bias) and an ac (signal) component

$$k = V_{dc} + V_s \cos \omega_s t. \quad (2.6.25)$$

The output current is

$$I_2 = m_1 I_{dc} V_{dc} + m_2 V_{dc} I_c \cos \omega_c t + m_3 V_s I_{dc} \cos \omega_s t + m_4 I_c V_s \cos \omega_c t \cos \omega_s t \quad (2.6.26)$$

where the m s are constants. To convert I_2 to an amplitude modulated wave, the term

$$m_1 I_{dc} V_{dc} + m_3 I_{dc} V_s \cos \omega_s t \quad (2.6.27)$$

will have to be removed from Equation (2.6.26) to give

$$I_2 = A(1 + B \cos \omega_s t) \cos \omega_c t \quad (2.6.28)$$

where

$$A = m_2 I_c V_{dc} \quad (2.6.29)$$

and

$$B = \frac{m_4 V_s}{m_2 V_{dc}}. \quad (2.6.30)$$

Assuming that the circuit is to be realized in integrated circuit form, where the transistor characteristics can be very closely matched, the removal of Equation (2.6.27) is most easily achieved by using the concept of “balancing”, that is, by duplicating the circuit and interconnecting the two parts as shown in Figure 2.33.

Note that the ideal current source I' has a dc current equal to the dc component of I . A practical realization of the circuit with various bias arrangements, temperature compensation and a number of other features is shown in Figure 2.34 and it is manufactured by Motorola Semiconductor Products, Inc. and designated “MC1595 - Four-quadrant Multiplier”:

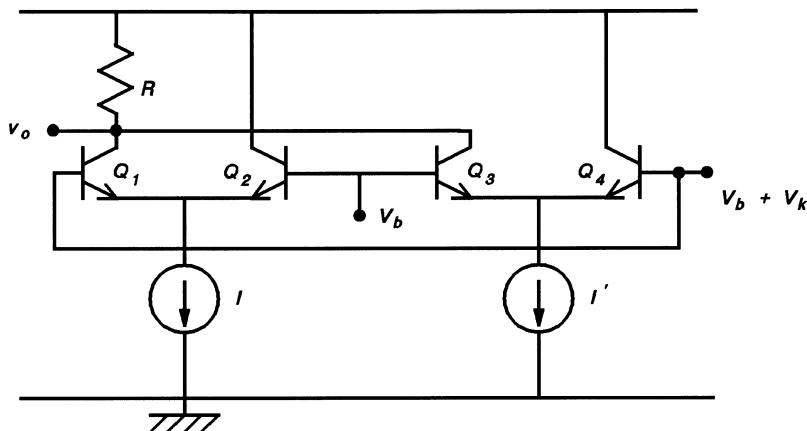


Figure 2.33. The connection of a second long-tailed pair as shown, removes the unwanted signals by using the concept of “balance”. This is possible when the circuit is integrated on a single chip because circuit integration produces matched transistors which can be relied on to track together with temperature.

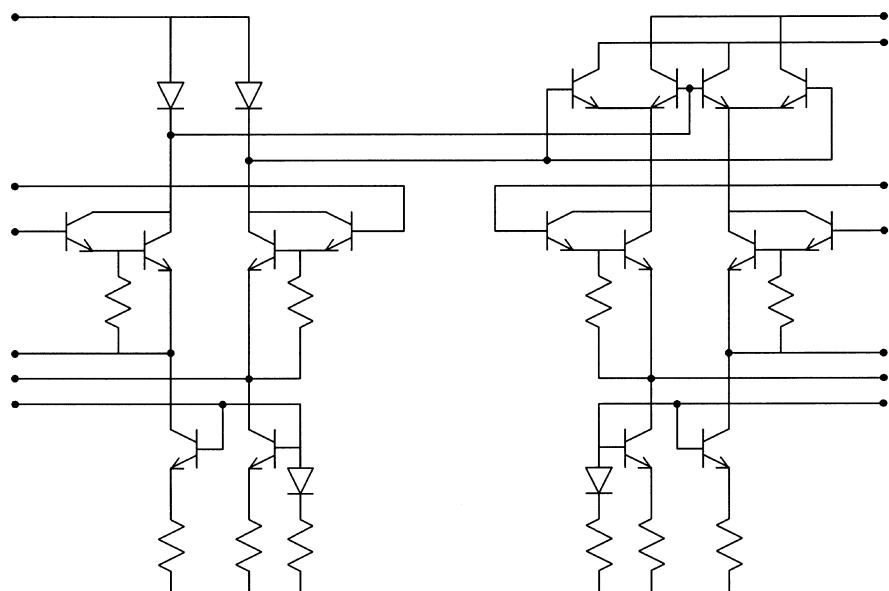


Figure 2.34. The circuit diagram of the four-quadrant multiplier (MC1595) manufactured by Motorola Semiconductor Products Inc. Reprinted with permission from the *Linear and Integrated Circuits Handbook*, 1983.

2.7 AUDIO-FREQUENCY AMPLIFIER

There are two types of amplifiers which can be used to boost the output of the microphone to the level required by the modulator – class-A and class-B. Both of these amplifiers are “linear” in the sense that the output is nominally an exact replica of the input. The difference is that a class-A amplifier has current flowing in the active device at all times, and generally its output power and conversion efficiency are modest. It is most often used with low-level signals. Class-B amplifiers have current flowing in the active device for only one-half of the cycle. This situation would normally lead to an unacceptable level of distortion so two active devices are used in a “push-pull” configuration. The advantage of this type of amplifier is that it has a higher power output level than the class-A amplifier and a higher conversion efficiency as well. The audio-frequency amplifier used in the transmitter will be determined by the output of the signal source (microphone) and the signal level required at the input of the modulator. It could be a combination of the two, that is, a class-A followed by a class-B amplifier. The two types of amplifier will be discussed in detail.

2.7.1 Basic Device Characteristics

Figure 2.35 shows the characteristics of a bipolar junction transistor (BJT) in the common emitter mode. This is used as the basis of the design and can be applied with slight modifications to other devices and other modes.

It is evident that the $V_{CE} - I_c$ characteristics are controlled by the base current, I_B , supplied to the transistor. When the transistor is connected as shown in Figure 2.36 and the value of I_B is reduced then it follows that the collector current I_c will decrease and so will the voltage drop across R_c until finally when $I_c = 0$, $V_{CE} = V_{cc}$.

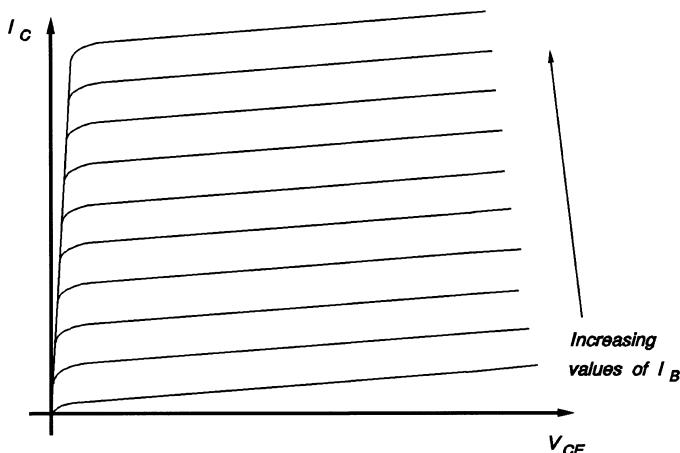


Figure 2.35. The collector characteristics of a junction bipolar transistor showing collector current (I_c) plotted against collector-emitter voltage (V_{CE}) for various values of base current (I_B).

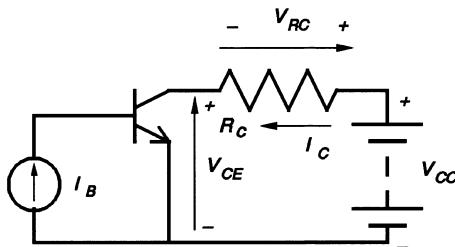


Figure 2.36. The basic common-emitter amplifier with load R_c .

The transistor is said to be “cut-off”. On the other hand, if I_B is increased, the collector current I_c will increase and so will the voltage drop across R_c until finally $V_{CE} = 0$ and $V_{RC} = V_{cc}$. The transistor is said to be “in saturation”. The collector current is no longer under the control of the base current but determined by the value of R_c . In practice, the value of V_{CE} is not quite zero and, for most discrete devices, V_{CE} has a value of about 0.5 V.

Applying KVL to the collector current path,

$$V_{cc} = V_{CE} + V_{R_c} = V_{CE} + I_c R_c \quad (2.7.1)$$

or

$$I_c = \left(\frac{1}{R_c} \right) V_{cc} + \left(-\frac{1}{R_c} \right) V_{CE}. \quad (2.7.2)$$

When Equation (2.7.3) is plotted on the characteristics of the device shown in Figure 2.37 it gives a straight line with a slope equal to $(-1/R_c)$ and intercepts on the y axis equal to V_{cc}/R_c and on the x axis equal to V_{cc} . This line is known as the *load line* and determines the behavior of the device when it is connected to the collector load, R_c , and the supply voltage V_{cc} .

2.7.2 Class-A Amplifier

Because the current in a class-A amplifier flows for the entire 2π radians of the input ac cycle, its analysis depends on a simultaneous consideration of the dc as well as the ac conditions of the circuit.

2.7.2.1 Base-Current-Biased Class-A Amplifier. The intersection of the load line and the characteristic representing the value of the base current, I_{bq} , establishes the *quiescent* or operating point of the transistor. Figure 2.37 shows how the collector voltage and collector current can be made to change sinusoidally by superimposing a sinusoidal current source, i_b , on the quiescent value.

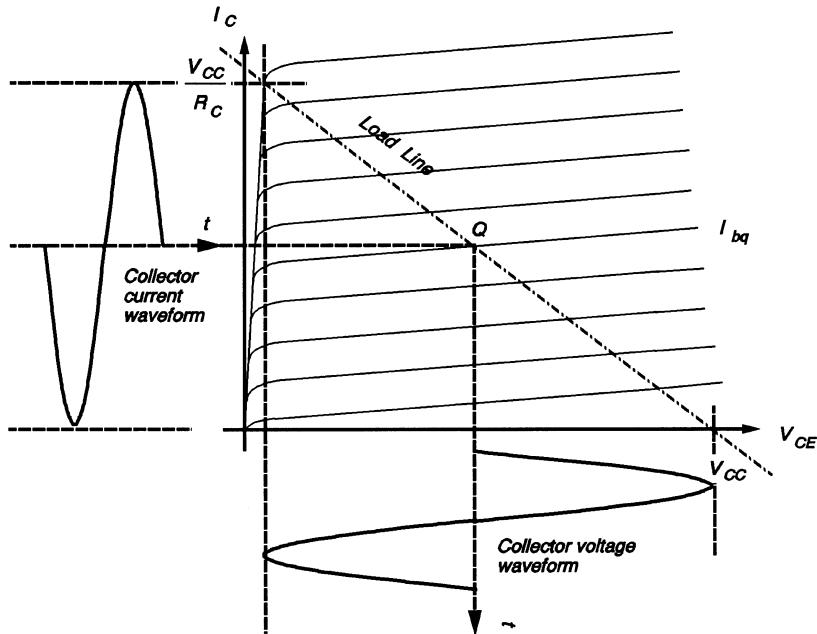


Figure 2.37. The operation of the amplifier is illustrated on the $I_c - V_{CE}$ characteristics by the load line with corresponding current and voltage waveforms.

The circuit can be made to provide amplification by a suitable choice of a collector resistor R_c , a base resistor R_b , the supply voltage V_{cc} and by connecting them as shown in Figure 2.38.

The design of an amplifier is a series of judicious choices and their modification, when necessary, in order to meet the design specification. The design is therefore not unique. The technique is best illustrated by a design example.

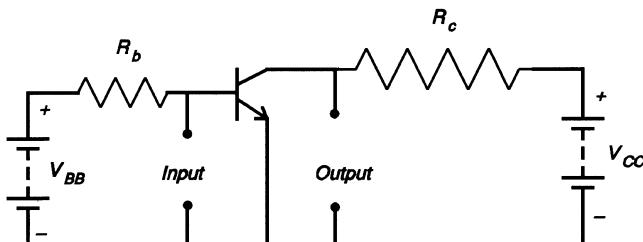


Figure 2.38. The common-emitter amplifier with the quiescent point determined by V_{BB} and R_b .

Example 2.7.1 Base-Current-Biased Class-A Amplifier. The following data apply to the circuit shown in Figure 2.38.

- Collector supply voltage, $V_{cc} = 10\text{ V}$
- Base supply voltage, $V_{BB} = 3\text{ V}$
- Current gain of transistor, $\beta = 100$
- Base-emitter voltage drop, $V_{BE} = 0.7\text{ V}$ (silicon)
- Input voltage = 0.5 volts peak-to-peak
- Required voltage gain = 10
- Amplifier load, $R_L = 10\text{ k}\Omega$.

Solution. The first step is to locate the point corresponding to V_{cc} on the V_{CE} axis of the transistor characteristics. The slope of the load line is determined by the value of R_c and the choice of R_c depends largely on the current required to drive the load of the amplifier. With an input voltage of 0.5 V peak-to-peak and a voltage gain of 10, the output voltage will be 5.0 V peak-to-peak. The current required to drive the load is therefore 0.5 mA peak-to-peak. To ensure that the operation of the transistor is not adversely affected when the load draws off 0.5 mA peak-to-peak from the transistor, we have to allow about 10 times the load current in the collector of the transistor *in the form of direct current*, that is, $I_c = 5\text{ mA}$ in the quiescent state.

The next step is to determine the value of R_c . For the amplifier to handle the maximum possible signal, it is necessary to place its quiescent point in such a position that the collector voltage can swing as far in the positive direction as it can swing in the negative direction. For the maximum signal, the transistor goes into saturation at the minimum, and it is cut off at the maximum value of the collector voltage. It then follows that the required point on the I_c axis is 10 mA. A straight line can now be drawn to join $V_{cc} = 10\text{ V}$ to $I_c = 10\text{ mA}$. From the slope of load-line, $R_c = 1\text{ k}\Omega$.

The next step is to choose the value of R_b to establish the appropriate value of I_b so that the quiescent point is in the middle of the load line. From the diagram, the required base current

$$\begin{aligned} I_b &= I_c/\beta \\ &= 50\text{ }\mu\text{A}. \end{aligned}$$

Neglecting the resistance of the base-emitter diode,

$$\begin{aligned} R_b &= (V_{BB} - V_{BE})/I_b \\ &= (3.0 - 0.7)/50 \times 10^6 \Omega \\ &= 46\text{ k}\Omega. \end{aligned}$$

The diode resistance is given approximately by the empirical formula:

$$\begin{aligned} R_d &= 26 \text{ mV}/(\text{diode current in amps}) \\ &= 26 \times 10^{-3} / 50 \times 10^{-6} \Omega \\ &= 520 \Omega. \end{aligned}$$

R_d is indeed negligible compared to R_b .

The input voltage source has to be adjusted to give the required input voltage at the base of the transistor and the coupling capacitor C_o is chosen so that, at the lowest frequency of operation, its reactance is negligibly small compared to the load R_L .

The amplifier, as designed, has the following disadvantages:

- (1) It has two dc voltage supplies.
- (2) Its input impedance is very low – that of a forward biased diode.
- (3) Its bias point is largely determined by V_{BE} which is sensitive to changes in temperature.
- (4) The voltage gain of the amplifier is sensitive to the current gain of the transistor. The current gain of two transistors fabricated next to each other on an integrated circuit chip can vary by as much as a ratio of three to one so it is necessary to design amplifiers which are essentially independent of the current gain of the device.

2.7.2.2 Resistive Divider-Biased Class-A Amplifier. The amplifier circuit can be modified to that shown in Figure 2.39. The base of the transistor is then held at a voltage determined by the resistors R_1 and R_2 . A resistor R_e is connected between the emitter and ground. The purpose of this is two-fold; it raises the input impedance of the amplifier and redefines the voltage gain as the ratio R_c/R_e .

The amplifier circuit is called “common-emitter” although the emitter is not connected directly to ground. A proper common-emitter amplifier has characteristics which are dependent on the parameters of the transistor, the absolute value of the collector load and other factors such as the temperature at which it is used. These disadvantages make the use of R_e attractive.

As before, the steps in the design are illustrated by the following example:

Example 2.7.2 Resistive Divider-Biased Class-A Amplifier

DC supply voltage, $V_{cc} = 10 \text{ V}$

Current gain of transistor, $\beta = 100$

Base-emitter diode voltage drop, $V_{BE} = 0.7 \text{ V}$ (silicon)

Input voltage = 0.5 V peak-to-peak

Required voltage gain = 10

Amplifier load resistance = 10 kΩ.

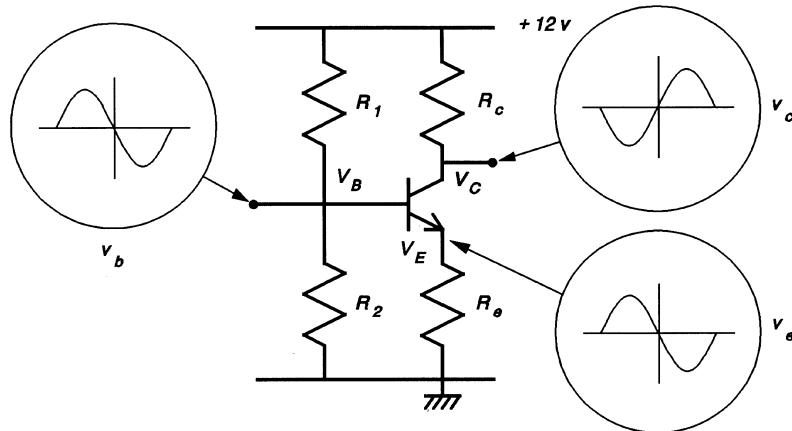


Figure 2.39. The resistive-divider biased “common-emitter” amplifier. The use of the emitter resistor R_e increases the input resistance and reduces the dependence of the amplifier gain on the transistor parameters.

Solution. To keep the transistor in conduction, the base voltage must be kept at 0.7 V above the emitter voltage. The point in the cycle when the transistor is most likely to be cut off is at the negative peak (0.5 V peak) of the input voltage. The minimum quiescent base voltage is therefore $(0.7 + 0.5 \text{ peak}) = 1.2 \text{ V}$.

The steps leading to the choice of a quiescent collector current remains the same as in Example 2.7.1 and gives the value 5 mA. Assuming that the current gain of the transistor is sufficiently high, then I_c is approximately equal to I_e . Under quiescent conditions, the emitter is held at 0.5 V and the current flowing in R_e is 5 mA. R_e has a value of 100Ω .

The ac voltage at the base of the transistor (0.5 V peak) is the same as what appears at the emitter, except for the dc shift of 0.7 V. Since the voltage gain of the amplifier is 10, it follows that the ac signal at the collector must be 10 times as large as what appears across the emitter. Keeping in mind that the emitter and collector currents are approximately equal, it follows that the collector resistance R_c must be $10 \times R_e = 1 \text{ k}\Omega$.

The base current required to maintain the collector current of 5 mA is given by the relationship:

$$I_c = \beta \times I_b$$

$$I_b = 50 \mu\text{A}.$$

The current I_b is drawn from the resistive chain R_1 and R_2 which also define the voltage at the base. To keep the base voltage reasonably constant for changes in I_b due to variations in β , the current in the resistive chain should be about $10 \times I_b$, that is, $500 \mu\text{A}$. The chain has 10 V across it so $(R_1 + R_2) = 20 \text{ k}\Omega$. Since the base voltage must be at 1.2 V, it follows that $R_1 = 17.6 \text{ k}\Omega$ and $R_2 = 2.4 \text{ k}\Omega$. The

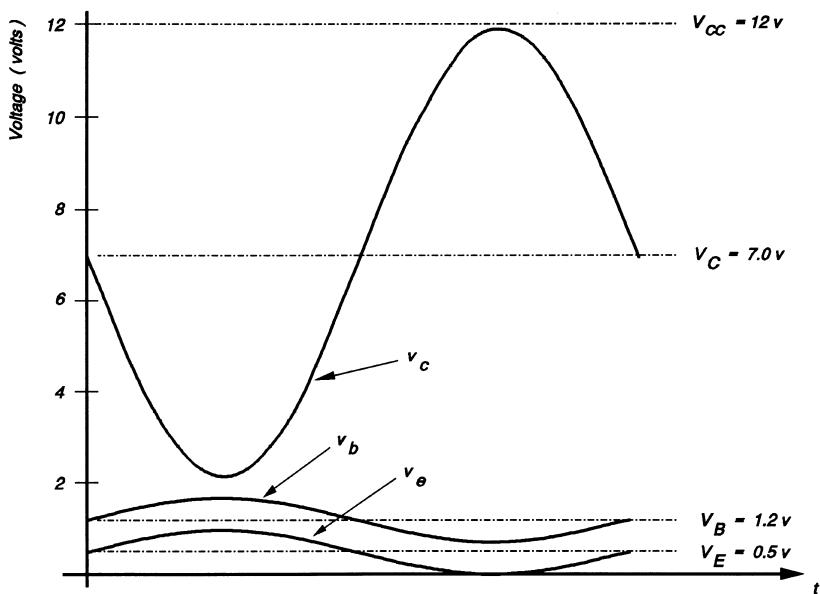


Figure 2.40. The emitter, base and collector voltage waveforms of the common-emitter amplifier with dc bias voltages shown.

waveforms at the emitter, base and collector are shown in Figure 2.40. Note that any increase in the amplitude of the input signal will cause the amplifier to clip the output signal in both the cut-off and saturation modes.

The input signal is coupled to the amplifier by a capacitor C_i and the output to the load by a capacitor C_o . The values of C_i and C_o are chosen so that at the lowest frequency of operation their reactances are less than 10% of the impedance of the load to which they are connected.

The advantages of this configuration over that presented in Example 2.7.1 are as follows:

- (1) Only one dc power supply is used.
- (2) The input impedance is given by the parallel combination of R_1 , R_2 and βR_e which gives $1.74 \text{ k}\Omega$.
- (3) The bias point is determined by the ratio of the resistive chain R_1 and R_2 which is independent of temperature.
- (4) The voltage gain of the amplifier is determined by the ratio of R_c/R_e and it is not a function of the current gain or any other parameter of the transistor.

2.7.3 Class-B Amplifier

2.7.3.1 Complementary Symmetry Amplifier. The basic circuit for a complementary symmetry amplifier is shown in Figure 2.41. To keep the circuit simple, no biasing circuits are shown.

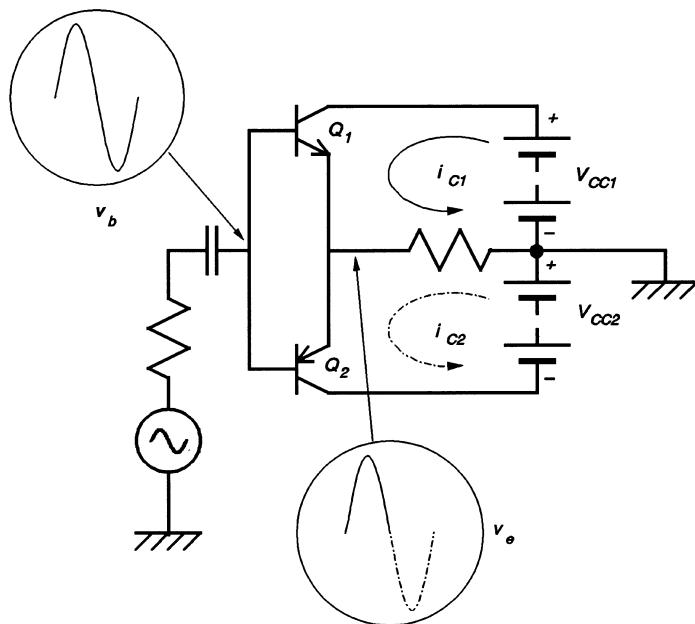


Figure 2.41. Basic class-B complementary symmetry amplifier. Note that the circuit is, in fact, two emitter followers (NPN and PNP) with a common load resistor.

It is necessary to make the following assumptions:

- (1) The amplitude of the input voltage is much larger than V_{BE} . This assumption ensures that *cross-over* distortion is kept at a minimum.
- (2) The transistors Q_1 and Q_2 are exact complements of each other. This condition (not strictly necessary since we are dealing with emitter followers) ensures that symmetry is maintained.
- (3) The dc power supply voltages, V_{cc1} and V_{cc2} , are equal.

When the input voltage goes positive, the base-emitter junction of Q_1 becomes forward biased and current i_{c1} flows as shown. The base-emitter junction of Q_2 is reverse-biased and Q_2 is cut off. Essentially, the circuit acts as an *emitter follower*.

When the input voltage goes negative, Q_2 conducts a current i_{c2} in the direction shown. Q_1 is cut off. If i_{c1} and i_{c2} are equal, the voltage developed across the load R_L will be sinusoidal.

The operation of the circuit, in terms of the device characteristics and the load line, can be seen in Figure 2.42. Note that the y axis of the PNP transistor, Q_2 , has been shifted to make the point V_{cc} coincide with the point $-V_{cc}$. The advantage gained is the simplicity of the diagram.

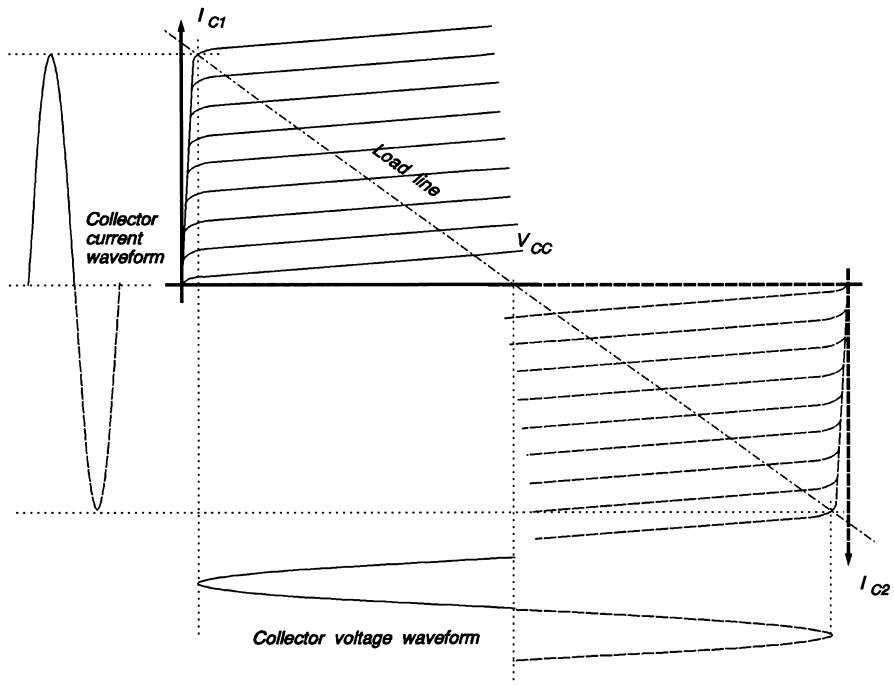


Figure 2.42. Composite characteristics of the complementary symmetry transistor pair with the load line, collector voltage and current shown. The solid lines refer to Q_1 , the dotted to Q_2 .

The ac power output is

$$P_{\text{out}} = \frac{\hat{I}_c}{\sqrt{2}} \frac{\hat{V}_{\text{cc}}}{\sqrt{2}} = \frac{\hat{I}_c \hat{V}_{\text{cc}}}{2}. \quad (2.7.3)$$

The current flowing in each transistor is a half-sinusoid whose average value (dc equivalent) is \hat{I}_c/π . The average current for the two transistors is $(2\hat{I}_c)/\pi$. Since this current is flowing against a potential difference of V_{cc} , the dc power input is

$$P_{\text{dc}} = \frac{2I_c V_{\text{cc}}}{\pi}. \quad (2.7.4)$$

The conversion efficiency is

$$\eta = \frac{\text{ac power}}{\text{dc power}} = \frac{\pi}{4} = 78.5\%. \quad (2.7.5)$$

The complementary symmetry class-B amplifier has the following disadvantages:

- (1) It requires two equal dc power supplies.
- (2) Complementary pairs of transistors are not easy to manufacture, especially at high power levels.
- (3) The circuit cannot have a voltage gain since it is basically two emitter followers with a common load.

To avoid the above disadvantages, the transformer-coupled class-B amplifier can be used.

2.7.3.2 Transformer-Coupled Class-B Amplifier. The transformer-coupled class-B amplifier uses two center-tapped transformers, as shown in Figure 2.43. The only assumption necessary is that the two transistors are matched. Since they are both NPN (or PNP) this is not, in general unduly restrictive.

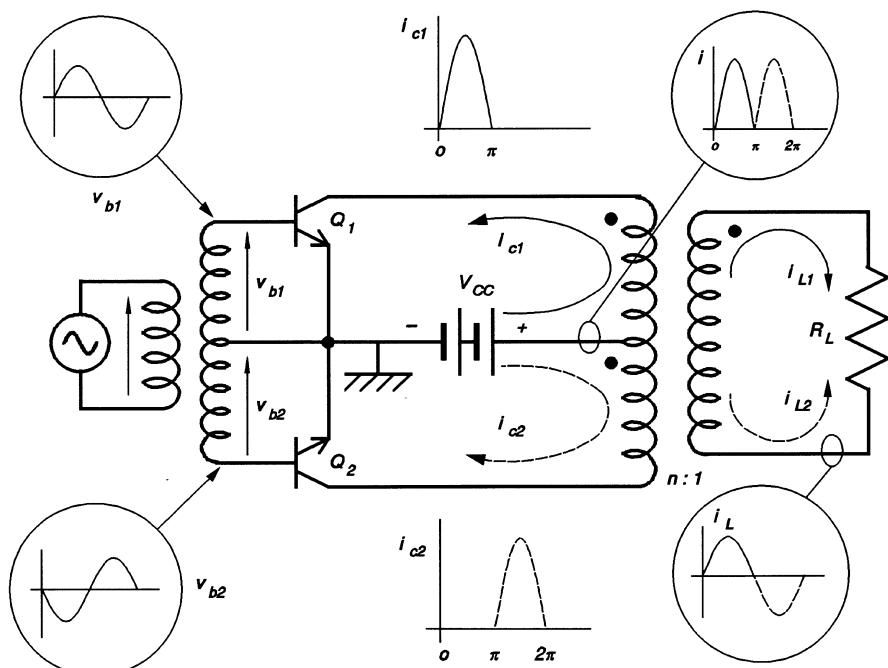


Figure 2.43. Circuit diagram of the transformer-coupled class-B amplifier. The waveforms of the currents and voltages at various points are indicated. The solid lines refer to Q_1 , the dotted line to Q_2 .

When the input voltage goes positive, two voltages v_{b1} and v_{b2} appear across the secondary of the input transformer as shown. The voltage v_{b1} biases Q_1 on and v_{b2} cuts off Q_2 . The current i_{c1} , which flows in Q_1 and one-half of the primary of the output transformer, is a half-sinusoid. The voltage developed across the secondary of the output transformer is therefore a half-sinusoid as shown.

When the input voltage goes negative, v_{b1} biases Q_1 off and v_{b2} brings Q_2 into conduction. The current i_{c2} , that flows in Q_2 and the other half of the primary of the output transformer, is a half-sinusoid but it flows in the opposite direction to i_{c1} . The voltage that is developed across the output transformer is then a negative-going half-sinusoid. The resultant current in the load i_L is a complete sinusoid.

It should be noted that at any point in time only one-half of the primary of the output transformer is coupled to the secondary.

Transformers are bulky, heavy and bandwidth limited so class-B amplifier designs which do not use transformers are available [8, 9].

Example 2.7.3 Class-B Amplifier. The idealized output characteristics of a transistor to be used in a class-B push-pull amplifier are given in Figure 2.44.

The following parameters are used:

- (1) Supply voltage $V_{cc} = 50$ V.
- (2) The amplifier load resistance = 8Ω .
- (3) The emitter-base diode characteristics can be approximated by a straight line as shown in Figure 2.45.

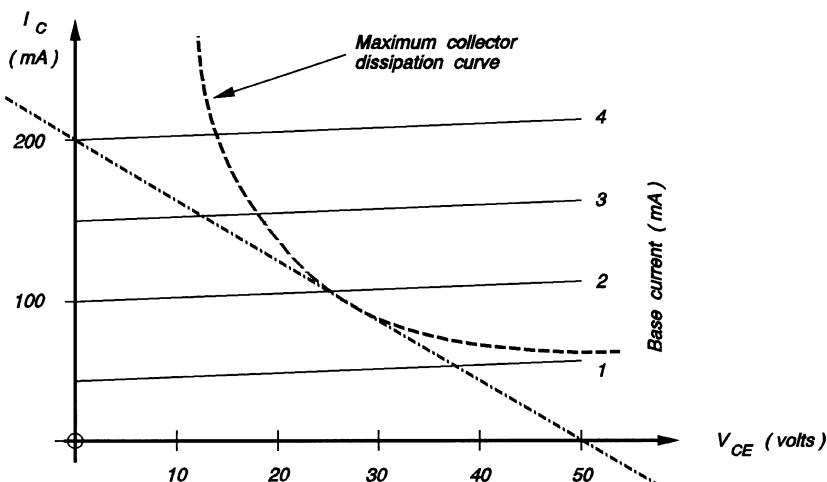


Figure 2.44. The transistor characteristics have been idealized and the maximum collector power dissipation curve superimposed on it.

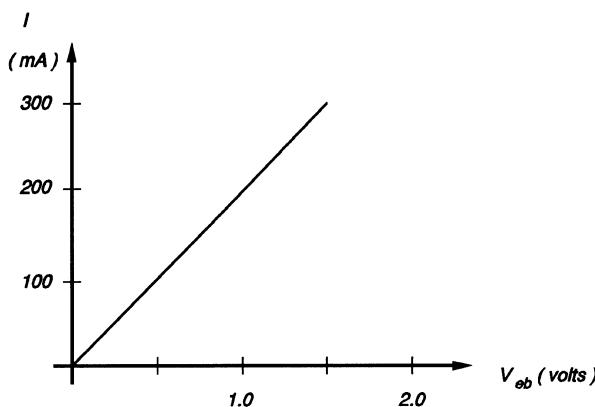


Figure 2.45. To keep the example simple, the nonlinear emitter-base diode has been made linear.

Calculate the following:

- (a) the turns ratio of the output transformer,
- (b) the conversion efficiency,
- (c) the input power,
- (d) the output power,
- (e) the power gain in decibels.

Solution. The collector characteristics has a rectangular hyperbola drawn on it to indicate the maximum dissipation capability of the device. The load line may then touch the locus of the maximum dissipation line but may not intersect it. In the absence of any requirement on the maximum power output of the amplifier, it may be assumed that it should provide the maximum power. The load line can then be drawn through $V_{CE} = 50$ V and tangential to the maximum dissipation curve. This gives the apparent collector load resistance value of 250Ω .

A suitable circuit for the class-B amplifier is shown in Figure 2.46. At any point in time only one-half of the transformer primary is linked to the secondary. The turns ratio $a : 1$ refers to this link and not to the whole of the primary which is, in fact, $2a : 1$. The load “seen” by the collector of the transistor is to be 250Ω but this represents a load R_L transformed through the transformer. Hence

$$250 = a^2 R_L$$

$$a = 5.59.$$

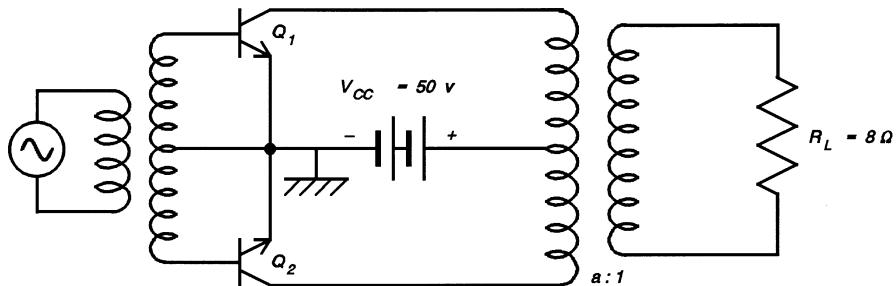


Figure 2.46. The circuit diagram for Example 2.7.3.

The collector characteristics of the “composite” transistor and the ac load line, together with the collector current and the corresponding collector voltage, are shown in Figure 2.47.

The following parameters can be derived:

- Current gain of the transistor is approximately equal to 50.
- Output voltage $= V_{cc}/\sqrt{2}$ V (rms).
- Output current $= \hat{I}_c/\sqrt{2}$ A (rms).
- Output power, $P_o = VI \cos \phi$ (where $\phi = \pi$ radians):

$$\begin{aligned} P_o &= -(50 \times 0.2)/2 \text{ W} \\ &= -5 \text{ W}. \end{aligned}$$

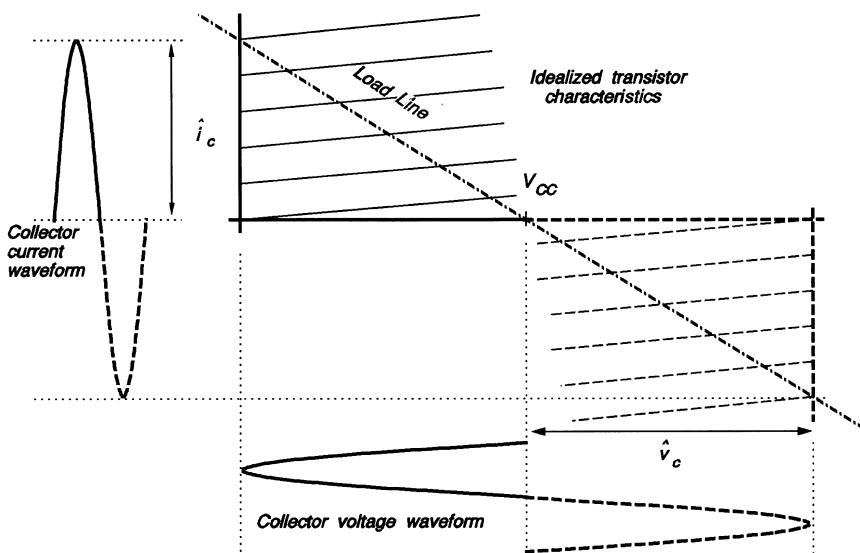


Figure 2.47. The composite transistor collector characteristics have been idealized with the load line, collector current and voltage waveforms shown. The solid lines refer to Q_1 , the dotted line to Q_2 .

The negative sign can be ignored since it simply indicates that the transistor is generating the power, not dissipating it.

- (e) The base current required to drive the maximum collector current $\hat{I}_b = 4 \text{ mA}$.

Using the approximation $I_c = I_E$ and the graph of $I_E - V_{eb}$, for $I_E = 200 \text{ mA}$,

$$\begin{aligned}\hat{V}_{eb} &= 1.0 \text{ V} \\ \text{Input power, } P_i &= (\hat{V}_{eb} \hat{I}_b)/2 \\ &= 0.002 \text{ W} \\ \text{Power gain} &= 10 \log_{10} P_o/P_i = 33.9 \text{ dB.}\end{aligned}$$

2.8 THE RADIO-FREQUENCY AMPLIFIER

The choice of an amplifier to boost the power coming out of the modulator to the level required to drive the antenna can be made when the following points have been considered:

- (1) The amplifier must be linear so as to preserve the nature of the modulated signal.
- (2) The output power of the amplifier can vary from a few watts to hundreds of kilowatts.
- (3) A high conversion efficiency is a necessity when high power is a requirement. The portion of the dc power supplied to the amplifier which is not converted into useable ac signal is dissipated in the amplifier components. Suitable arrangements have to be made to remove the heat. This can add to the complexity and the cost of operation of the transmitter.
- (4) The signal to be amplified consists of the carrier and the two sidebands which normally constitute a narrowband signal.

The usual choice is a class-B (actually a class-AB) amplifier with a tuned load. The class-B is chosen for its linearity, high output power and high conversion efficiency and the tuned load ensures that it is a narrowband amplifier. For relatively low power requirements, transistors may be used in the output stage of the amplifier. When the power output required is in excess of a few hundred watts, vacuum tubes are used.

A suitable circuit for the radio-frequency power amplifier is shown in Figure 2.48. Its operation is the same as that of the class-B audio-frequency amplifier discussed in Section 2.7.3.2.

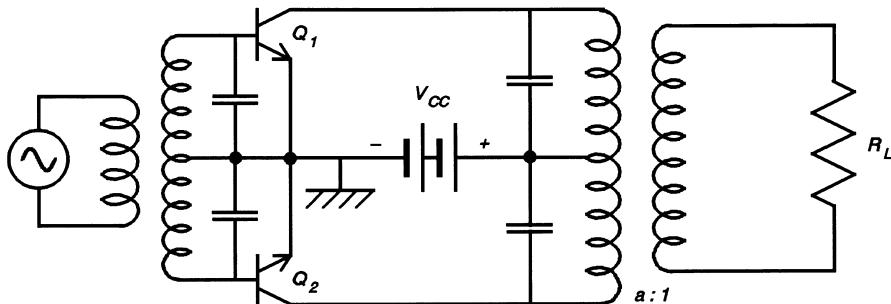


Figure 2.48. Class-B amplifier with a tuned input and output. The amplifier is designed to amplify the carrier and sidebands only.

Example 2.8.1 Class-B Tuned Amplifier. The final stage of a class-B radio-frequency power amplifier in a radio transmitter has a tuned circuit in which $C = 2500 \text{ pF}$. The carrier frequency $\omega_c = 5 \times 10^5 \text{ rad/s}$ and the modulating signal has frequencies ranging from $\omega_{s1} = 20 \text{ rad/s}$ to $\omega_{s2} = 10 \times 10^3 \text{ rad/s}$. The power is coupled through a transformer of turns ratio $10 : 1$ to an antenna whose impedance is assumed to be 250Ω resistive.

Calculate the following:

- The inductance in the tuned circuit.
- The Q factor of the tuned circuit if the extreme edges of the sidebands are to be within -3 dB of the carrier.
- What is the maximum winding resistance that the coil in the tuned circuit can have under the condition in (b)?
- What is the Q factor of the coil alone?

Solution

$$(a) \omega_c^2 = 1/(LC)$$

$$\begin{aligned} L &= 1/(25 \times 10^{10} \times 2500 \times 10^{-12}) \text{ H} \\ &= 1.6 \text{ mH}. \end{aligned}$$

(b) Now

$$Q_o = \frac{\omega_c}{\omega_2 - \omega_1}$$

where ω_1 and ω_2 are the two -3 dB frequencies

$$Q = (5 \times 10^5 / 20 \times 10^3) = 25.$$

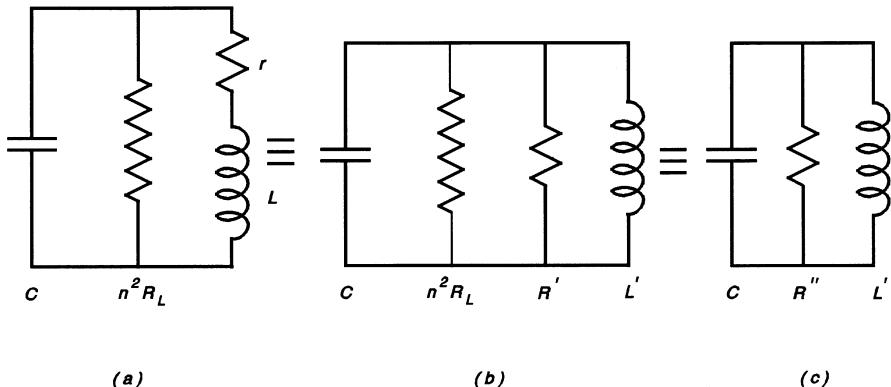


Figure 2.49. (a) The tuned collector circuit of the amplifier with the resistive load transferred to the primary of the transformer. (b) The collector circuit when the series RL has been replaced with the parallel equivalent. (c) The collector equivalent circuit when the two resistive elements have been combined.

- (c) There are two sources of loss in the tuned circuit which contribute to give a loaded Q of 25. They are:

 - (1) The antenna resistance¹ coupled into the tuned circuit by the transformer ($n^2 R_L$).
 - (2) The winding resistance of the coil, r .

The equivalent circuit of the load is shown in Figure 2.49(a–c) in three stages. Stage *a* shows the reflected resistance in parallel with the tuned load and the winding resistance in series with the coil. Stage *b* is the equivalent circuit when the winding resistance, r , is converted into a resistance R' in parallel with the inductance. Stage *c* shows the equivalent circuit when the resistances $n^2 R_L$ and R' are combined to form R'' .

Now, if the two circuits shown in Figure 2.50 are equivalent then

$$r + j\omega L = \frac{j\omega L_p R_p}{R_p + j\omega L_p}. \quad (2.8.1)$$

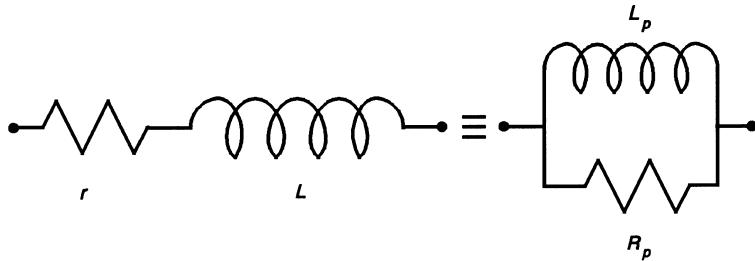


Figure 2.50. The conversion of the parallel $R_p L_p$ circuit to the series RL equivalent.

¹Antenna resistance is not the physical resistance of the antenna; it is an equivalent resistance in which the radiated power is dissipated.

Rationalizing and equating real and imaginary parts:

$$r = \frac{\omega^2 L_p^2 R_p}{R_p^2 + \omega^2 L_p^2} \quad (2.8.2)$$

and

$$L = \frac{L_p R_p^2}{R_p^2 + \omega^2 L_p^2}. \quad (2.8.3)$$

The Q factor is

$$Q = \frac{R_p}{\omega L_p} \quad (2.8.4)$$

When $Q \gg 1$, $R_p \gg \omega L_p$, then

$$r \simeq \frac{\omega^2 L_p^2}{R_p} \quad (2.8.5)$$

and

$$L \simeq L_p. \quad (2.8.6)$$

From equivalent circuit c ,

$$Q = \frac{R''}{\omega L'}. \quad (2.8.7)$$

Therefore

$$R'' = 25 \times \omega L' = 20 \text{ k}\Omega. \quad (2.8.8)$$

The resistance, R'' , has two components as shown in circuit b , $n^2 R_L$, in parallel with R' :

$$n^2 R_L = 10^2 \times 250 = 25 \text{ k}\Omega \quad (2.8.9)$$

Therefore

$$\frac{25R'}{(25 + R')} = 20 \quad (2.8.10)$$

which gives

$$R' = 100 \text{ k}\Omega. \quad (2.8.11)$$

Using Equation (2.8.1) when $R_p = R'$ gives

$$r = 6.4 \Omega. \quad (2.8.12)$$

(d) The Q factor of the coil only is

$$Q_0 = \frac{\omega L}{r} = 125. \quad (2.8.13)$$

2.9 THE ANTENNA

Antenna structures can take many physical forms, from the “whip” antenna used mostly for in-car radios, through “rabbit ears” for television to the microwave parabolic “dish” used in satellite communication. These are only the most common examples of antennas which are used for reception of broadcast signals. The antenna is the last processor of the signal at the transmitting end and the first at the receiving end. Antennae vary widely in shape, size and complexity, depending on the frequency of operation and the desired field pattern.

Radio communication in free space is possible because, when an alternating current flows in a conductor, part of the energy is lost in the form of electromagnetic radiation into free space. When the frequency of the current is low, the radiation “loss” is very small, but as the frequency increases, substantial losses can occur. In designing an antenna the object is to construct a structure which will maximize the radiated energy in a given direction or over a geographic area.

A useful analog of how an antenna radiates energy is given by a body which is floating in the middle of a pond on a windless day. For the purposes of this description we may assume that the body is a beach ball. If we can get a man suspended from a crane directly above the beach ball to push the ball very slowly into the water and release it equally slowly then the energy expended in pushing the ball into the water is recovered when it is released and very little will be lost. If the man increases the frequency at which he pushes and releases the ball, an increasing amount of energy will be lost in creating waves which will radiate from the ball outwards. The amount of radiated energy at any given point on the pond will bear an inverse relationship to the distance from the ball. It is necessary to bear in mind that the waves on the pond are surface waves (two dimensional) whereas those produced by an antenna are three dimensional.

The speed of propagation of the radio wave is the same as that of light ($c = 3 \times 10^8 \text{ m/s}$). The wavelength and the frequency are related by

$$\lambda f = c \quad (2.9.1)$$

where λ is the wavelength in meters and f is the frequency in hertz. Electromagnetic radiation has two vectors; the \mathbf{E} and \mathbf{H} which are orthogonal to each other and to the direction of propagation.

A detailed study of antenna design is beyond the scope of this text and the interested reader may consult a suitable textbook on the subject. The discussion will therefore be qualitative.

2.9.1 Radiation Pattern of an Isolated Dipole

A simple way to gain insight into the operation of a dipole is to start from an open-circuit balanced transmission line. Such a transmission line will have a standing-wave distribution of currents in opposite directions so that the net radiation from the structure will be quite small. Figure 2.51(a) shows the transmission line and the standing-wave pattern. It is assumed that the diameter of the conductors is infinitesimally small compared to the wavelength of the signal to be transmitted.

If the transmission line is bent at right angles as shown in Figure 2.51(b), the structure is called a *dipole* and it has the current distribution shown. Since the oppositely directed currents are now far apart, they do not counteract each other and hence the dipole is a good radiator of electromagnetic waves.

If it is assumed that $2l = \lambda/2$, the dipole is then called a *half-wavelength dipole*. The \mathbf{E} field of the half-wavelength dipole when plotted gives a circle tangential to the axis of the dipole touching the axis at the mid-point. In three dimensions, the pattern is a doughnut, as shown in Figure 2.52.

Now assume that $l \ll \lambda/2$: the dipole is said to be a *short dipole* and the field pattern is also a doughnut but the cross section is slightly distorted as if compressed vertically.

2.9.2 Monopole or Half-Dipole

The equations which describe the behavior of the isolated dipole can be used, with slight modifications, for the monopole or half-dipole by assuming that the Earth's

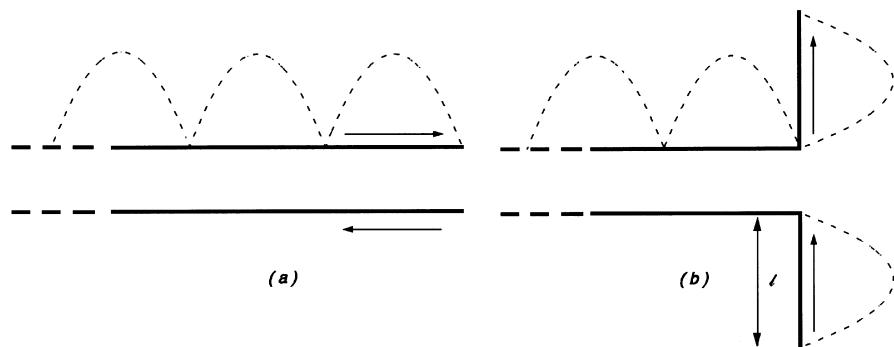


Figure 2.51. (a) The current standing wave pattern on an open-circuit balanced transmission line. (b) The current standing wave pattern when the line is bent at a point where $l = \lambda/2$ to form a dipole.

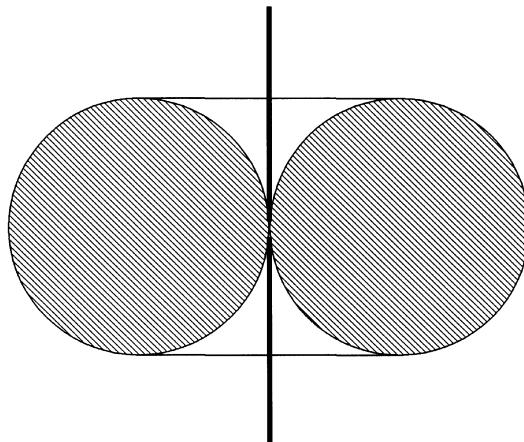


Figure 2.52. The radiation pattern of an isolated half-wavelength dipole.

surface, on which the antenna is to be placed, is a perfectly flat conductor. This permits the treatment of the monopole and its image on the ground plane as a dipole. Figure 2.53 shows the monopole and its image. This representation of the antenna is closer to reality than the isolated dipole.

2.9.3 Field Patterns for a Vertical Grounded Antenna

The **E** field pattern generated by a vertical grounded antenna resembles that of the monopole closely even though the Earth's surface is not exactly a perfectly flat

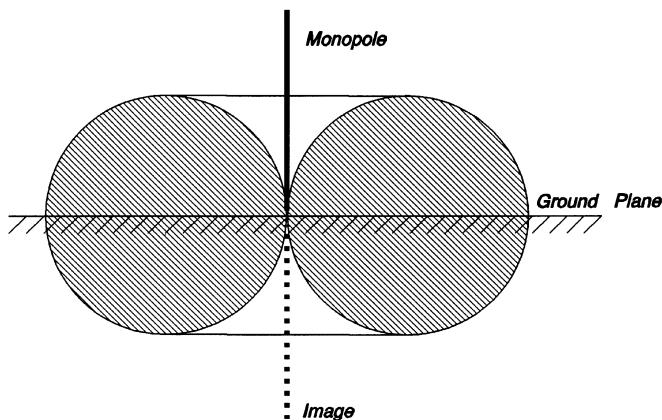


Figure 2.53. This diagram shows how the half-wavelength dipole can be replaced with a monopole and a flat conducting surface (ground plane).

conducting surface. Figures 2.54(a)–(e) are a series of diagrams which present the expected approximate field pattern as the height of the vertical grounded antenna changes. It is important to note that the height is measured in terms of the wavelength of the signal to be radiated.

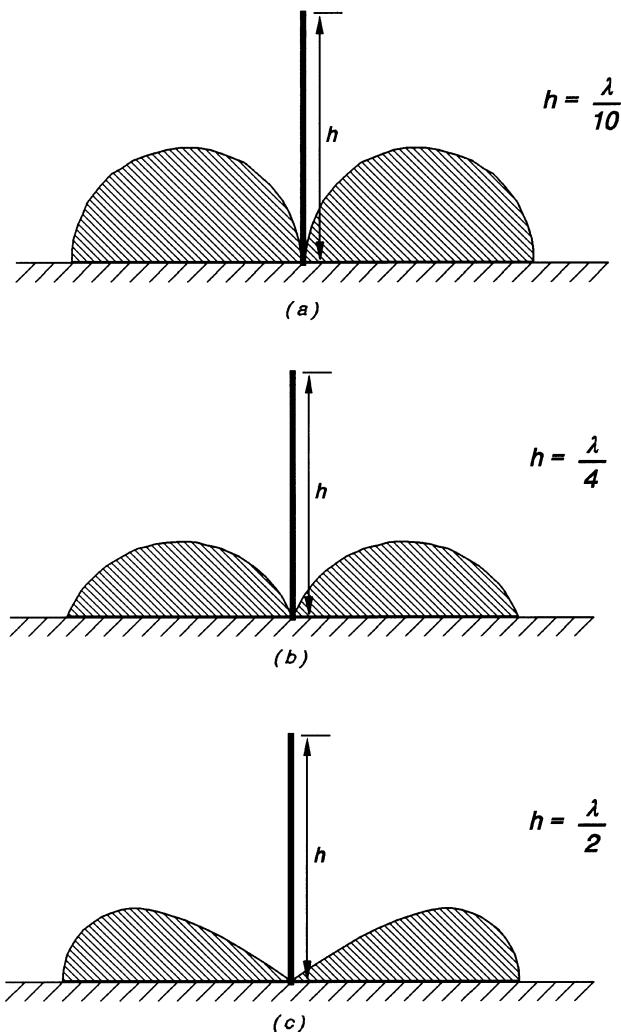


Figure 2.54. (a) Radiation pattern when h is approximately $\lambda/10$. (b) Radiation pattern when h is approximately $\lambda/4$. Note the slight elongation of the pattern along the ground plane. (c) The radiation pattern has undergone further distortion as h increases to approximately $\lambda/2$. (d) As h approaches $5\lambda/8$, the ground wave gets more elongated and the sky wave appears. (e) With h approximately $3\lambda/4$, the sky wave has grown considerably at the expense of the ground wave.

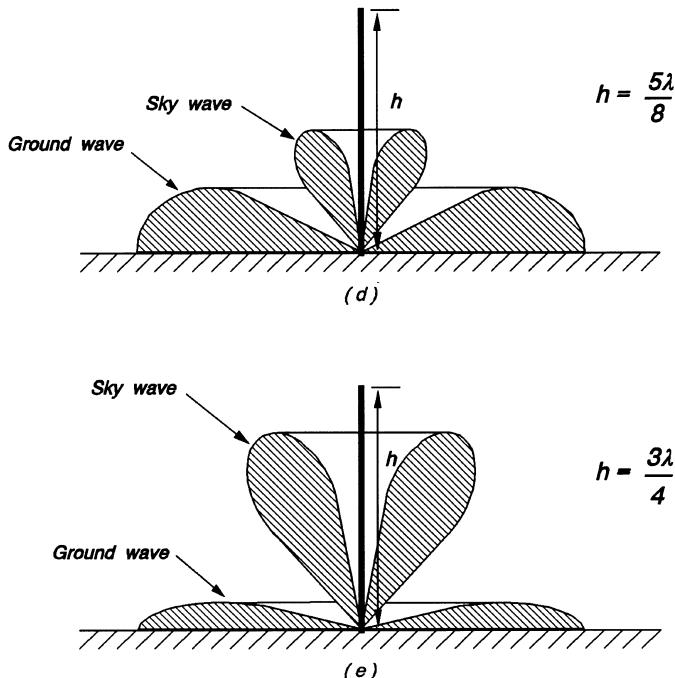


Figure 2.54. (continued)

2.10 CLASSIFICATION OF AMPLITUDE MODULATED RADIO-FREQUENCY BANDS

Amplitude modulated radio frequencies are grouped into three bands according to the wavelength of their carrier frequencies. The carrier frequency chosen depends to a large extent on the distance between the broadcasting station and the target listeners.

- 1. Long wave (low frequency).** All transmission whose carrier frequencies are less than 400 kHz are generally classified as long wave. At a frequency of 100 kHz, a quarter-wavelength antenna is 750 meters high. Such an antenna poses several problems such as vulnerability to high winds and danger to low flying aircraft. Long wave broadcasting stations therefore use an electromagnetically short antenna which necessarily limits their reach to a few tens of kilometers because the short antenna has only the ground wave.
- 2. Medium wave.** Carrier frequencies in the range 300 kHz to 3 MHz are regarded as medium wave. The height of the antenna becomes more manageable and the possibility of using the sky wave to reach distant audiences is a reality. Generally, it is used for local area broadcasting.

3. Short wave. Short wave generally refers to carrier frequencies between 3 MHz and 30 MHz. The wavelengths under consideration are between 100 meters and 1 meter. Antenna structures can be constructed to give specified directional properties. Most of the energy can be put into the sky wave and the signal can be bounced off the ionosphere (the layer of ionized gas that surrounds the Earth) to reach receivers halfway round the world. A very severe problem is encountered in short wave transmission, that is, the signal tends to fade from time to time. This phenomenon is caused by the multiple paths by which the signal can reach the receiver. It is clear that if two signals reach the receiver by different paths such that their phase angles are 180° apart they will cancel each other. The ionosphere sometimes experiences severe turbulence due mainly to radiation from the Sun. Short wave transmission is therefore at its best during the hours of darkness.

REFERENCES

1. Nyquist H., Regeneration Theory, *Bell System Tech. J.*, 11, 126–147, Jan. 1932.
2. Millman, J. and Halkias, C. C. *Integrated Electronics: Analog & Digital Circuits and Systems*, Chp. 14, McGraw-Hill, New York, 1972.
3. Vehovec, M., Houslander, L. and Spence, R., “On Oscillator Design for Maximum Power”, *Trans. IEEE*, CT-15, 281–283, 1968.
4. Buchanan, J. P. *Handbook of Piezoelectric Crystals for Radio Equipment Design*, WADC Tech. Report 56-156, U.S. Air Force, Oct. 1956.
5. Ryder J. D., *Electronic Fundamentals & Applications: Integrated and Discrete Systems*, 5th Ed., Prentice-Hall, Englewood Cliffs, NJ, 1976.
6. Fraser, W., *Telecommunications*, MacDonald, London, 1957.
7. Kaye, A. R., “A Solid-State Television Fader-Mixer Amplifier”, *J. SMPTE*, 74, 602–606, July 1965.
8. Bray, D. and Votipka, W., *25-Watt Audio Amplifier with Short-Circuit Protection*, Application Bulletin, Fairchild Semiconductor, August 1967.
9. Ruehs, R., *High Power Audio Amplifiers with Short-Circuit Protection*, AN-485, Motorola Semiconductors, 1968.

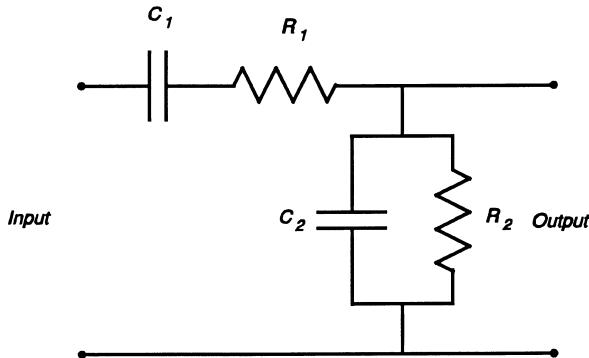
PROBLEMS

- 2.1** An ideal voltage source has a signal consisting of a sinusoidal carrier of frequency 1 MHz amplitude modulated by a sinusoidal signal of frequency 20 kHz to a depth of 100% and is applied to an *LC* series tuned circuit. The capacitor has a value 100 pF and may be assumed to be lossless. The *Q* factor of the circuit is 100. Calculate the following:
- (a) The value of the inductance required to tune in the signal.
 - (b) The value of the coil resistance.

- (c) The depth of the modulation of the current which flows in the circuit.
 (d) Discuss the reason(s) for the change of the depth of modulation in (c), if any.
- 2.2** In an AM transmitter, the current flowing in the antenna is 7.5 A when the carrier is not modulated. When a sinusoidal modulation is applied, the current changes to 8.25 A. Calculate:
 (a) The depth of modulation.
 (b) The antenna current when the depth of modulation is 65%.
- 2.3** A feedback circuit with a $\beta = (8.0 + j4.0) \times 10^{-3}$ is connected to an amplifier and the resulting system oscillates sinusoidally. Calculate the gain and phase shift of the amplifier. Design a circuit which can perform the feedback function and show how it can be connected to the amplifier. The frequency of the oscillation is $100/2\pi$ kHz and it may be assumed that the amplifier has a high input impedance and a low output impedance.
- 2.4** Show that when a source of emf E volts and internal impedance $Z_s = R_s + jX_s$ is connected to a load $Z_L + jX_L$, maximum power is transferred to the load when:
 (a) $R_s = R_L$
 (b) $X_s = -X_L$.
- A load resistance 1150Ω is connected across the secondary winding of a transformer. The inductance of the winding is 80 mH and the Q factor is 32. The primary winding has ideal characteristics and it is connected to a voltage source in series with a capacitor C_1 . The frequency of the voltage source is 20×10^3 Hz and the internal resistance is 142Ω . Determine the value of C_1 and the mutual inductance between the two coils for maximum power to be delivered to the load.
- 2.5** The wiper of a rheostate is moved linearly with respect to time between the 25% and 75% points repetitively at a frequency of 6 Hz. A voltage signal $V_{in} = 10 \cos \omega_o t$ is connected to the terminals of the rheostat. Sketch the waveform of the voltage across the wiper with respect to one of the terminals when $\omega_c = 2\pi \times 60$ rad/s. Derive a general expression for all the frequencies present and evaluate the amplitude of the largest component present other than ω_c .
- 2.6** A source of sinusoidal signal can be represented by an ideal voltage source of value 1 V peak in series with a 600Ω resistance. This source is to be used to drive a single transistor class-A amplifier coupled to a 10Ω resistive load by means of a transformer. Design a suitable amplifier given that a 12 V dc source is available and the maximum possible power output is required. The β of the transistor is 100 and the lowest frequency of interest is 20 Hz. The dc current in the collector of the transistor is not to exceed 2 mA. Calculate the power gain in

decibels, assuming that 0 dB refers to the situation when the source is connected directly to the load.

- 2.7** Show that an amplifier with a voltage gain A and feedback factor β will oscillate when $|\beta A| = 1$. What other condition(s) have to be met for sinusoidal oscillation to occur? The feedback circuit to be used in an oscillator is as shown in Figure P2.1.



Assuming an ideal voltage amplifier, derive an expression for:

- (a) the frequency of oscillation
- (b) the gain of the amplifier.

Evaluate (a) and (b) if $C_1 = C_2 = 1000 \text{ pF}$ and $R_1 = R_2 = 20 \text{ k}\Omega$.

- 2.8** The load of a class-C amplifier is coupled through a lossless transformer of turns ratio $10 : 1$, and it is tuned to a frequency $250/2\pi \text{ kHz}$. The average power dissipated in the load at this frequency is 250 mW . The amplifier load is 50Ω and the loaded Q factor of the primary with the load transferred to the secondary is 25. Calculate the following:

- (a) the primary inductance
- (b) the secondary inductance
- (c) the capacitance required to tune the primary
- (d) the frequency at which the output power will equal 125 mW
- (e) the -3 dB bandwidth of the amplifier
- (f) the minimum dc power supply voltage required.

Describe how you would convert the amplifier into a frequency multiplier with a multiplication factor of 2. What conditions have to be met in order to obtain a larger multiplication factor?

3

THE AMPLITUDE MODULATED RADIO RECEIVER

3.1 INTRODUCTION

The electromagnetic disturbance created by the transmitter is propagated by the transmitter antenna and travels at the speed of light as described in Chapter 2. It is evident that, if the electromagnetic wave encounters a conductor, a current will be induced in the conductor. How much current is induced will depend on the strength of the electromagnetic field, the size and shape of the conductor and its orientation to the direction of propagation of the wave. The conductor will then capture some of the power present in the wave and hence it will be acting as a receiver antenna. However, other electromagnetic waves emanating from all other radio transmitters will also induce some current in the antenna. The two basic functions of the radio receiver are:

- (1) to separate the signal induced in the antenna by the transmission which we wish to receive from all the other signals present,
- (2) to recover the “message” signal which was used to modulate the transmitter carrier.

3.2 THE BASIC RECEIVER: SYSTEM DESIGN

In order to separate the required signal from all the other signals captured by the antenna, we use a bandpass filter centered on the carrier frequency with sufficient bandwidth to accommodate the upper and lower sidebands but with a sufficiently high Q factor so that all other carriers and their sidebands are attenuated to a level where they will not cause interference. This is most easily achieved by using an LC tuned circuit whose resonant frequency is that of the carrier.

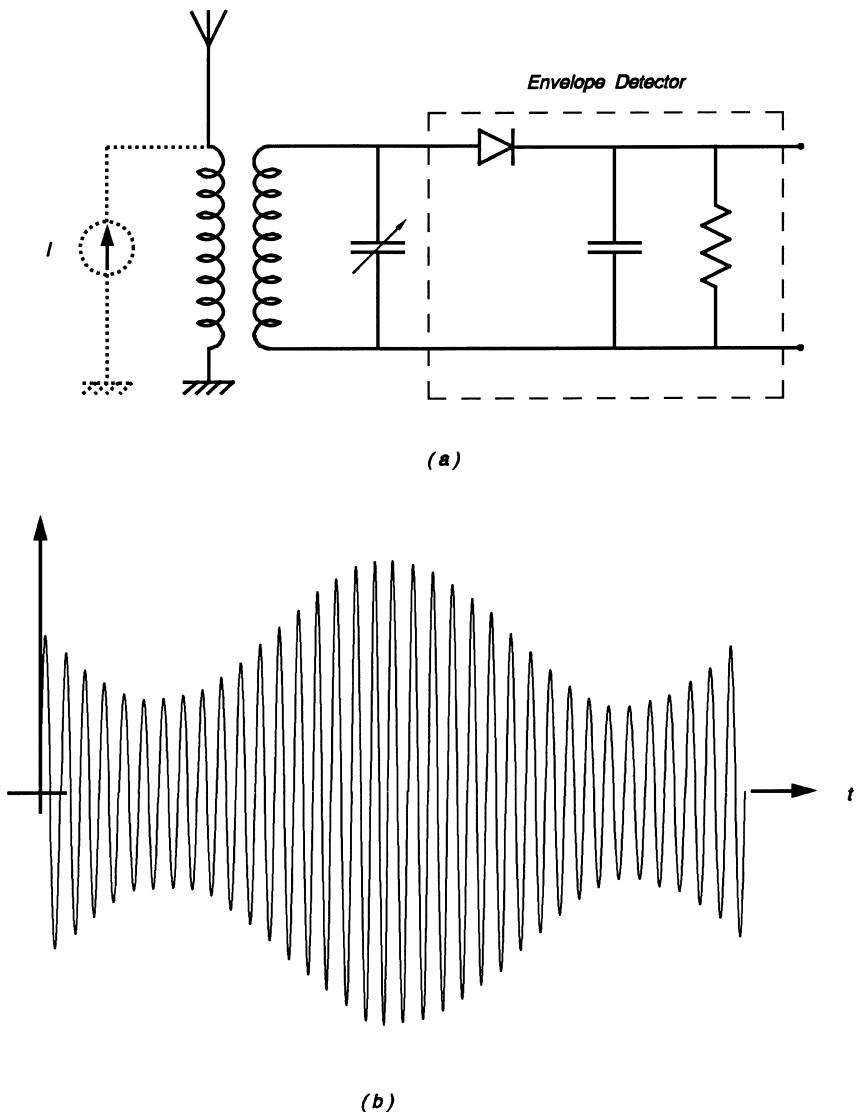


Figure 3.1. (a) The envelope detector circuit. The diode “half-wave” rectifies the AM wave and the RC time-constant “follows” the envelope with a slight ripple. (b) The input signal to the envelope detector. (c) The output signal of the envelope detector. Note that (1) when the voltage is rising the ripple is larger than when the voltage is falling. A longer time constant will help reduce the ripple; however, it will also increase the likelihood that the output voltage will not follow the envelope when the voltage is falling causing ‘diagonal clipping’. (2) In practice, the carrier frequency is much higher than the modulating frequency, hence the ripple is much smaller than shown.

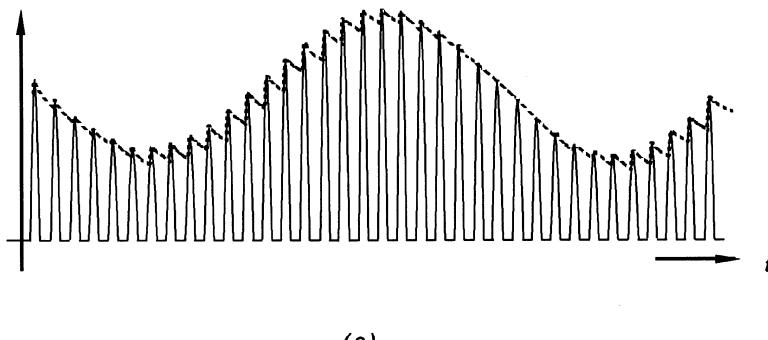


Figure 3.1. (continued)

To recover the “message” we require a circuit which will follow the envelope of the amplitude of the carrier. Such a circuit is called an envelope detector and it consists of a diode and a parallel RC circuit as shown in Figure 3.1(a).

The input signal to the circuit is most appropriately represented by an ideal current source connected to the primary of the transformer. This ideal current source represents all the currents induced in the antenna by all the radio stations broadcasting signals in free space. The signal is coupled to the parallel-tuned LC circuit which selectively enhances the amplitude of the signal whose carrier frequency is the same as the resonant frequency of the LC circuit. In Figure 3.1(b), only the enhanced modulated signal is shown at the input of the envelope detector. Because the diode conducts only when the anode has a positive potential compared to the cathode, only the positive half of the signal appears across the output resistor. Because the capacitor is connected in parallel with the resistor, when the diode conducts the capacitor must charge up to the peak value of the voltage. When the input voltage is less than the voltage across the capacitor, the conduction is cut off and the capacitor starts to discharge through the resistor with the voltage falling off exponentially. With the proper choice of time-constant RC , the output voltage waveform will have the form shown in Figure 3.1(c). This waveform is essentially the envelope of the carrier signal with a ripple at a frequency equal to the carrier frequency. A low-pass filter can be used to remove the ripple.

The circuit shown in Figure 3.1(a) has been used with success as a practical receiver with the resistor R replaced by a high impedance headphone. Needless to say, such a simple circuit has its limitations. The power in the circuit is supplied entirely by the transmitter and naturally it is at a very low level, especially as the distance between the transmitter and the receiver increases. Secondly, the ability of the LC tuned circuit to suppress the signals propagated by all the other transmitters is limited and therefore such a receiver will be subject to interference from other stations. These limitations can be overcome by using the *superheterodyne* configuration described below.

3.3 THE SUPERHETERODYNE RECEIVER: SYSTEM DESIGN

The superheterodyne receiver takes the incoming radio-frequency signal whose frequency varies from station to station and transforms it to a fixed frequency called the *intermediate frequency* (IF). It is then easier to do the necessary filtering to eliminate interference and, at the same time, to provide some power gain or amplification to the desired signal.

A normal AM superheterodyne receiver block diagram is shown in Figure 3.2.

The antenna has induced in it currents from all the transmitters whose electromagnetic propagation reach it. The first step is to use an *LC* tuned radio-frequency amplifier to enhance the desired carrier and its sidebands. The radio-frequency amplifier is tuneable over the frequency for which the receiver is designed by varying the capacitor in the tuned circuit. This capacitor is mechanically coupled or “ganged” to another capacitor which forms part of the local oscillator circuit. The local oscillator frequency and the frequency to which the radio-frequency amplifier is tuned are chosen in such a way that, as the value of the ganged capacitors change, they maintain a fixed frequency difference between them. The outputs from the local oscillator and the radio-frequency amplifier are used to drive the frequency changer or mixer. The frequency changer essentially multiplies the two inputs and produces a signal that contains the sum and difference of the input frequencies. Because of the fixed difference between the incoming radio-frequency and the local oscillator frequency, the difference frequency remains constant as the value of the ganged capacitor is changed. The output of the frequency changer is then fed into the intermediate-frequency amplifier. The intermediate-frequency amplifier is designed to select the difference frequency plus its sidebands and to attenuate all other frequencies present. Since the difference frequency is fixed (for domestic AM radios the intermediate frequency is 445 kHz) the filters required are relatively easy to

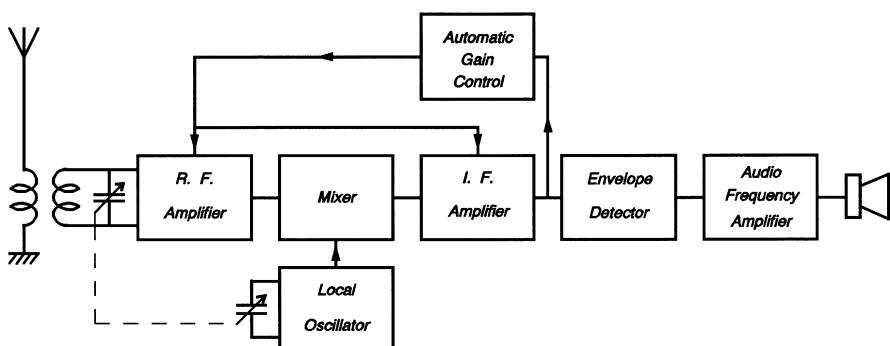


Figure 3.2. The block diagram of the superheterodyne receiver. The capacitor which tunes the radio-frequency amplifier is mechanically ganged to the capacitor which determines the frequency of the local oscillator. In the normal AM receiver, the oscillator frequency is always 455 kHz above the resonant frequency of the radio-frequency amplifier throughout the range of tuning.

design with sharp cut-off characteristics. The output of the intermediate-frequency amplifier which then goes to the envelope detector consists of the intermediate frequency and its two sidebands. The envelope detector removes the intermediate frequency, leaving the audio-frequency signal which is then amplified by the audio-frequency amplifier to a level capable of driving the loudspeaker. It is clear that there will be a very large difference between the signal from a powerful local radio station and a weak distant station. To help reduce the difference an automatic gain control (AGC) is used to adjust the signal reaching the envelope detector to stay within predetermined values.

The most interesting signal processing step in the system takes place in the frequency changer or frequency mixer or simply the mixer [1]. There are two basic types of mixers: the analog multiplier and the switching types. The analog multiplier frequency changer simply multiplies the radio-frequency signal and the local oscillator so that when the modulated carrier current is

$$i_m(t) = A(1 + k \sin \omega_S t) \sin \omega_C t \quad (3.3.1)$$

and the local oscillator signal is

$$i_o(t) = B \sin \omega_L t \quad (3.3.2)$$

the output of the mixer is

$$i(t) = A(1 + k \sin \omega_S t) \sin \omega_C t \times B \sin \omega_L t \quad (3.3.3)$$

$$i(t) = \frac{1}{2} AB(1 + k \sin \omega_S t)[\cos(\omega_L - \omega_C)t - \cos(\omega_L + \omega_C)t] \quad (3.3.4)$$

$$\begin{aligned} i(t) &= \frac{1}{2} AB[\cos(\omega_L - \omega_C)t - \cos(\omega_L + \omega_C)t + k \sin \omega_S t \cos(\omega_L - \omega_C)t \\ &\quad - k \sin \omega_S t \cos(\omega_L + \omega_C)t] \end{aligned} \quad (3.3.5)$$

$$\begin{aligned} i(t) &= \frac{1}{2} AB[\cos(\omega_L - \omega_C)t - \cos(\omega_L + \omega_C)t \\ &\quad + \frac{1}{2} k[\sin(\omega_L - \omega_C - \omega_S)t + \sin(\omega_L - \omega_C + \omega_S)t] \\ &\quad - \frac{1}{2} k[\sin(\omega_L + \omega_C - \omega_S)t + \sin(\omega_L + \omega_C + \omega_S)t]]. \end{aligned} \quad (3.3.6)$$

The spectrum of Equation (3.3.6) is shown in Figure 3.3. It should be noted that this has been simplified for clarity. The product formation in Equation (3.3.3) is not a precise process and tends to create a large number of frequencies due to sub- and higher harmonics present in both the radio-frequency and local oscillator signals. The radio-frequency and local oscillator signals are usually present in the output as well. It is important to keep all the unwanted signals outside the frequency band of the intermediate frequency and, failing that, to reduce their amplitude to a very low value.

It can be seen that the mixing operation gives two additional carriers and their sidebands at frequencies corresponding to the sum ($\omega_L + \omega_C$) and difference ($\omega_L - \omega_C$) of the local oscillator and carrier frequencies. The required signal at

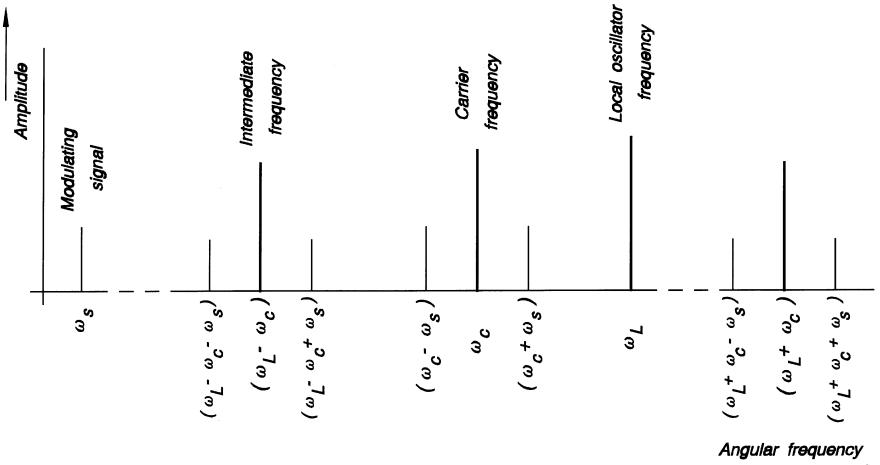


Figure 3.3. A simplified spectrum of the output from a frequency changer which uses a nonlinear device.

the difference frequency (intermediate frequency) can now be filtered out by the intermediate-frequency stage of the receiver. It should be noted that the mixing operation does not affect the sidebands. To clarify the changes in frequency that take place as the signal proceeds through the system, the AM broadcast band (600 kHz–1600 kHz) is used as an example in Table 3.1.

The frequency changer or mixer presents two immediate problems: the choice of the local oscillator frequency and the design strategy of the mixer itself.

- (1) It can be seen from Table 3.1 that the local oscillator frequency has been chosen to be higher than the incoming radio-frequency signal. There is a very good reason for this. The ratio of the maximum to the minimum capacitance

TABLE 3.1

	Radio frequency (kHz)	
	Low-frequency end	High-frequency end
Incoming signal, $f_c \pm f_s$	600 ± 5	1600 ± 5
Local oscillator, f_L	$600 + 455 = 1055$	$1600 + 455 = 2055$
Intermediate frequency, f_k	455	455
Image frequency ^a , f_{im}	$1055 + 455 = 1510$	$2055 + 455 = 2510$
Output, intermediate-frequency amplifier, $f_k \pm f_s$	455 ± 5	455 ± 5
Envelope detector, f_s	0–5	0–5

^aThe image frequency is the frequency of the unwanted signal which, when combined with the local oscillator frequency, will give the intermediate frequency. Normally the radio-frequency amplifier should suppress the image frequency but this may be difficult if the signal from the desired station is very weak and the image signal is very strong.

required to tune the local oscillator across the broadcast band is 3.79 when a higher local oscillator frequency is chosen. If the lower local oscillator frequency had been chosen, the ratio would have been 62.4. Such a variable capacitor would be difficult to manufacture with reasonable tolerance.

- (2) The mixing operation was treated earlier as an analog multiplication. However, the realization of a precise analog multiplier is a non-trivial problem. A crude analog multiplication can be achieved by using a device whose voltage-current characteristics are non-linear. An ordinary p-n junction diode can be used to perform the task. The derivation of the output signal is similar to that given in Section 2.6.1 and will therefore not be repeated here.

The switching type of mixer uses a device such as a diode or transistor carrying a current proportional to the radio-frequency signal and switches it from one state to another at the local oscillator frequency.

3.4 COMPONENTS OF THE SUPERHETERODYNE RECEIVER

3.4.1 Receiver Antenna

The AM receiver antenna can take many different forms such as the ferrite bar found in most portable receivers, the whip antenna found on automobiles, and the outdoor wire type consisting of several metres of wire strung between two towers. In general, the longer and higher off the ground the antenna is, the more likely it is that it will have a strong signal induced in it by the electromagnetic signals propagated by the transmitters. The level of signal induced in the antenna may vary from a few microvolts to a few volts, depending on the proximity of the transmitter, its radiated power output, the size of the receiver antenna, and its orientation to the transmitter. Because of the tremendous variation in the input signal a fixed gain amplifier will very often either not provide enough signal to the frequency changer or will overload it and consequently generate a large number of undesirable frequencies. To ensure a reasonable reception of the largest number of broadcasting stations, the gain of the amplifier is controlled automatically by the incoming signal – the weaker the signal, the higher the gain of the radio-frequency amplifier.

The antenna signal is coupled by a radio-frequency transformer to the input of the radio-frequency amplifier. The transformer is made up of two coils, each containing several turns of wire wound on a coil former which may or may not have a ferrite core. The major consideration in the design of the transformer is that the primary inductance be sufficiently high to ensure that signals at the lowest frequency of interest are not unduly attenuated. Since the signal frequency can vary from 600 to 1600 kHz, the transformer is not tuned.

3.4.2 Low-Power Radio-Frequency Amplifier

Since the input voltage of the amplifier is of the order of microvolts and the signal to be delivered to the *demodulator* is usually in volts, the amplifier must have a high gain. A multi-stage amplifier has to be used to realize the necessary gain. Some of the stages of gain can be placed before the frequency changer, in which case they are referred to as the radio-frequency amplifier stage, or after the frequency changer, in which case they are called the intermediate-frequency amplifier stage. It is usual to design the radio-frequency amplifier stage for a modest gain and the intermediate-frequency stage for high gain. Both the radio-frequency and intermediate-frequency amplifiers are narrow-band amplifiers. This is evident from calculating the Q factor for the two types of amplifiers. Considering that the normal bandwidth of the AM radio is 0–5 kHz, both radio-frequency and intermediate-frequency amplifiers have to have a bandwidth of at least 10 kHz.

The Q factor of the radio-frequency amplifier at the low end of the broadcast band (600 kHz) is 60 and at the high end (1600 kHz) is 160. The Q factor of the intermediate-frequency amplifier (centre frequency 455 kHz) is 45. However, operation of the radio-frequency amplifier with such a high Q factor will cause serious tracking problems with the local oscillator and will also lead to excessive attenuation at the edges of the sidebands. For practical purposes, a Q factor of about 10 is used in the radio-frequency amplifier, leaving the major part of the filtering problem to the intermediate-frequency stage. The design of the intermediate-frequency filter about a fixed center frequency is a much easier process and can be achieved with greater precision than in the radio-frequency stage, where the center frequency of the bandpass filter changes when the tuning capacitor is changed. In spite of the difference in Q factor of the radio-frequency and intermediate-frequency amplifiers, they have enough similarities for the same general principles to be used for their design.

The wide variation of the radio-frequency input signal level and the need for automatic gain control in the radio-frequency amplifier was discussed earlier. It is usual to amplify the incoming radio-frequency signal by a fixed amount in order to derive a control signal for the gain of a subsequent variable gain amplifier. A typical fixed gain radio-frequency amplifier is shown in Figure 3.4. Although a bipolar transistor is shown, a field-effect transistor can be used. The collector load is an *LC* tank circuit in which the capacitance is variable. The variable capacitance is mechanically ganged to the capacitance which controls the frequency of the local oscillator so that, as the capacitance is changed, the resonant frequency of the *LC* tank circuit tracks the local oscillator frequency with a constant difference equal to the intermediate frequency (455 kHz).

It can be seen that the above circuit bears a striking resemblance to the frequency multiplier circuit given in Figure 2.21. The difference is that the frequency multiplier operates in class-C while the radio-frequency amplifier operates in class-A.

The load driven by the amplifier may be coupled to the collector circuit by a transformer, in which case the inductor in the collector circuit becomes the transformer primary. The load may also be coupled by a capacitor. In both cases,

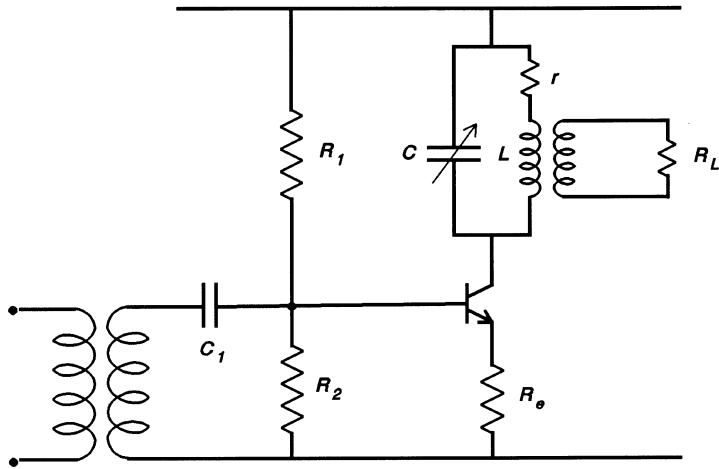


Figure 3.4. A typical radio-frequency amplifier. The load R_L represents the input resistance (impedance) of the circuits which are driven by the amplifier.

the load can be represented by a resistance R_L in parallel with the tuned circuit. To simplify the analysis of the circuit, the winding resistance r in series with the inductance is transformed into an equivalent shunt resistance (refer to Figure 2.50) R_p where

$$R_p = \frac{\omega^2 L_p^2}{r} \quad (3.4.1)$$

and

$$L_p = L \quad (3.4.2)$$

when $Q \gg 1$.

The amplifier load R_L combined in parallel with R_p is now the resistive part of the collector load. The new equivalent circuit is shown in Figure 3.5, where

$$R_{eq} = R_p \parallel n^2 R_L. \quad (3.4.3)$$

It can be seen from Figure 3.5 that:

- (1) The emitter resistor R_e has not been bypassed to ground with a capacitor.
- (2) At the frequency of resonance, the parallel LC circuit in the collector circuit will behave like an open circuit. The equivalent collector load is R_{eq} .
- (3) Because the inductor is connected directly between $+V_{cc}$ and the collector, the dc voltage on the collector is $+V_{cc}$.

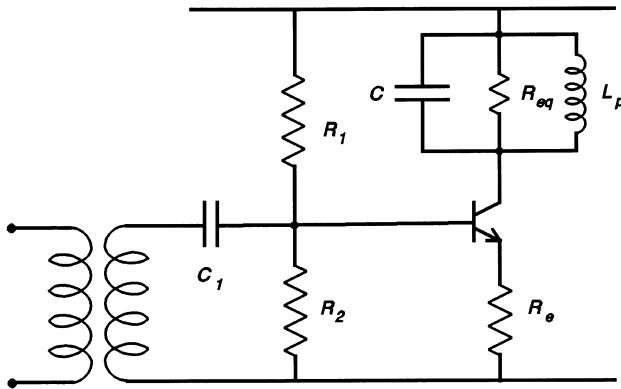


Figure 3.5. The amplifier shown in Figure 3.4 with the transformer load transferred to the primary and combined with the winding resistance r .

The major advantage of not bypassing R_e is that the gain of the amplifier is determined by the ratio of the collector-to-emitter load impedance, which, in this case, is R_{eq}/R_e and it is essentially independent of the transistor parameters such as current gain and transconductance. The design steps are illustrated in the following example.

Example 3.4.1 Low-Power Radio-Frequency Amplifier. The antenna of an AM radio receiver (600 to 1600 kHz) supplies 100 mV peak to the input of the radio-frequency amplifier when the modulation is a sinusoid, the modulation index is unity, and the radio-frequency is 600 kHz. The dc supply voltage is +6 V and the required gain is 20. The amplifier load represented by the input impedance of the automatic gain control circuit is 10 k Ω resistive and it is capacitively coupled. The variable capacitor used in the tuned circuit (and mechanically coupled to the capacitor used in the local oscillator) has a maximum value of 250 pF and a minimum value of 25 pF. The Q factor of the coil is expected to be about 50 and the current gain of the transistor is 100. Design a suitable amplifier.

Solution. The inductance of the tuning coil is given by

$$L = 1/(\omega^2 C).$$

When $\omega = 2\pi \times 600 \times 10^3$ and $C = 250 \text{ pF}$

$$L = 281 \mu\text{H}.$$

When $\omega = 2\pi \times 1600 \times 10^3$ the capacitance required to tune the amplifier

$$C = 35.2 \text{ pF}.$$

The combination of L and C can be used to tune the amplifier to any frequency in the AM broadcast band.

The winding resistance of the coil

$$\begin{aligned} r &= \omega L/Q \\ &= 21.2 \Omega. \end{aligned}$$

The equivalent parallel resistance

$$\begin{aligned} R_p &= \omega^2 L_p^2 / r \\ &= 52.9 \text{ k}\Omega. \end{aligned}$$

Combining R_p with the load resistance of $10 \text{ k}\Omega$ gives

$$R_{\text{eq}} = 8.41 \text{ k}\Omega.$$

The loaded Q of the collector circuit

$$\begin{aligned} Q_L &= R_{\text{eq}} / (\omega L) \\ &= 7.94. \end{aligned}$$

The relatively low Q should ensure that the sideband “edges” are not subject to severe attenuation.

At the resonant frequency, the parallel LC circuit in the collector behaves like an open circuit. The equivalent collector load is therefore R_{eq} .

The emitter resistance

$$\begin{aligned} Re &= R_{\text{eq}} / \text{gain} \\ &= 420 \Omega. \end{aligned}$$

The output voltage $= 100 \text{ mV} \times 20 = 2.0 \text{ V}$ (peak) and the current drawn by the $10 \text{ k}\Omega$ load is 0.2 mA (peak).

To ensure that the amplifier is capable of supplying 0.2 mA ac current to the load, the dc current in the collector may be set at ten times the load current, that is, $I_c = 2 \text{ mA}$. The dc voltages at the emitter and base are then $V_e = 0.84$ and $V_b = 1.54 \text{ V}$, respectively. The dc voltage on the collector is still 6 V . It is clear that the amplifier will go into saturation when the collector voltage drops to a minimum value of 1.34 V ($V_e + 0.5$) and to cut off when the collector voltage is 12 V . Since the collector signal is only $\pm 2 \text{ V}$, about a quiescent value of 6 V , there is no danger of clipping at the collector.

The dc base current

$$\begin{aligned} I_b &= I_c/\beta \\ &= 20 \mu\text{A}. \end{aligned}$$

The values of R_1 and R_2 are chosen so that a dc current of $10I_b$ will flow in the chain but with a voltage of 1.54 V at the base of the transistor. This gives

$$\begin{aligned} R_1 &= 22.3 \text{ k}\Omega \\ R_2 &= 7.7 \text{ k}\Omega. \end{aligned}$$

The coupling capacitor is chosen so that, at the lowest frequency of interest, its reactance is negligibly small compared to the load.

3.4.3 Frequency Changer or Mixer

Two distinct approaches can be used in the design of a mixer. The first is based on an analog multiplication of the radio-frequency and the local oscillator signals. The second uses the local oscillator signal to switch segments of the radio-frequency signal positive and negative. In this case the local oscillator must produce a square wave.

3.4.3.1 The Analog Mixer. As discussed earlier, a crude analog multiplication can be achieved by using a non-linear device such as a p–n junction diode which can be approximated by the equation

$$i = a_1 v + a_2 v^2 + \dots \quad (3.4.4)$$

A mixer using diodes produces an output signal with considerable loss. Various schemes exist, some employing several diodes with a single-ended or differential output.

If an “active” mixer is used, considerable gain can be obtained in the mixing process [5]. The preferred analog active mixer uses a dual-gate metal-oxide semiconductor field-effect transistor (MOSFET). The advantages of this design includes a lower power requirement from the local oscillator and improved isolation between the local oscillator and the receiver antenna. The isolation between the local oscillator and the antenna will ensure minimum radiation of the local oscillator signal and hence minimize interference with other electronic equipment.

To understand the design process, it is necessary to begin with the drain current (i_D) – drain-to-source voltage (v_{DS}) characteristics of the MOSFET. A typical n-channel depletion mode MOSFET is shown in Figure 3.6.

It can be seen that the characteristics are similar to those of a BJT except that the drain current is controlled by the gate-to-source voltage. An elementary common-source amplifier is shown in Figure 3.7.

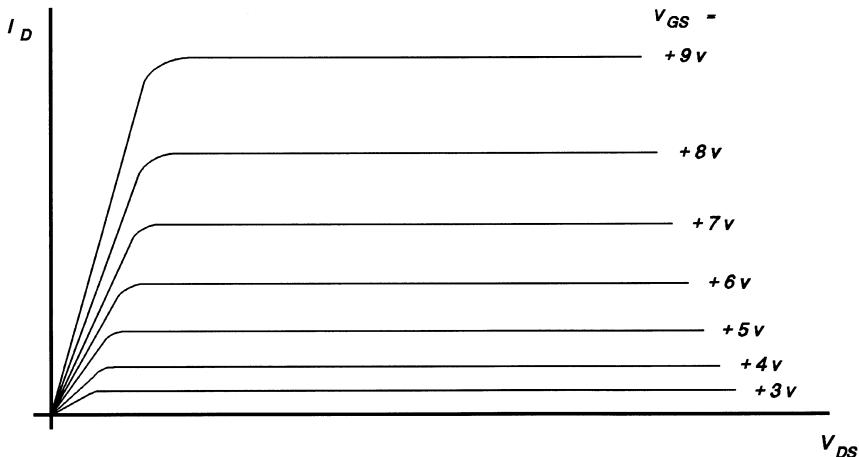


Figure 3.6. Typical characteristics of an n-channel, enhancement-mode metal-oxide semiconductor field effect transistor (MOSFET).

Applying KVL to the drain current path,

$$V_{DD} = i_D R_D + v_{DS} \quad (3.4.5)$$

or

$$i_D = \left(\frac{1}{R_D} \right) V_{DD} - \left(\frac{1}{R_D} \right) v_{DS}. \quad (3.4.6)$$

When Equation (3.4.6) is plotted on the FET characteristics shown in Figure 3.6, it gives a straight line with a slope of $(-1/R_D)$, an intercept on the x axis given by V_{DS} equal to V_{DD} and on the y axis $i_b = V_{DD}/R_D$. This is the load line which describes the behavior of the amplifier.

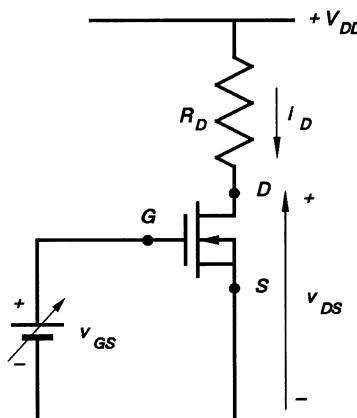


Figure 3.7. Typical biasing arrangement for the common-source MOSFET amplifier.

When designing an amplifier, it is necessary to select a bias point V_{GS} along the load line and the input signal v_{gs} so that the device will remain in the “active” region. This is achieved by ensuring that the device is biased above its threshold V_{th} . Then

$$v_{GS} = V_{GS} + v_{gs} \quad (3.4.7)$$

and the relationship between i_D and the v_{GS} can be approximated by

$$i_D = I_{DSS} \left(1 - \frac{v_{GS}}{V_{th}} \right)^2 \quad (3.4.8)$$

where the V_{th} and I_{DSS} are defined in Figure 3.8.

Substituting Equation (3.4.7) into (3.4.8) gives

$$i_D = I_{DSS} \left[\left(1 - \frac{V_{GS}}{V_{th}} \right)^2 - 2 \left(1 - \frac{V_{GS}}{V_{th}} \right) \left(\frac{v_{gs}}{V_{th}} \right) + \left(\frac{v_{gs}}{V_{th}} \right)^2 \right]. \quad (3.4.9)$$

The first term, $(1 - V_{GS}/V_{th})^2$, represents the dc component of the drain current. The second term, $2(1 - V_{GS}/V_{th})(v_{gs}/V_{th})$, is an ac current proportional to the input voltage and represents the normally desired output. The third term, $(v_{gs}/V_{th})^2$, represents a non-linearity which is normally undesirable. In terms of designing a mixer, however, this is the desired output. The relative value of this term can be increased by making the input signal v_{gs} large. However, since Equation (3.4.8) is an approximation, making v_{gs} too large can produce spurious signals which may interfere with the required signal.

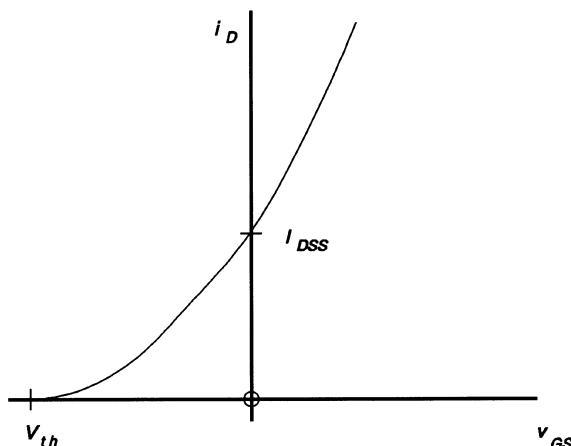


Figure 3.8. A typical $i_D - v_{GS}$ characteristic of an n-channel depletion-type MOSFET showing the threshold voltage, V_{th} , and the saturated drain-to-source current, I_{DSS} .

A more practical version of the MOSFET amplifier is shown in Figure 3.9.

The MOSFET used in the circuit is an n-channel enhancement device and the resistive chain R_1 and R_2 is chosen to hold the gate at a specific potential above that of the source. R_s is used partly to stabilize the dc bias point and partly to reduce the dependence of the gain on the parameters of the device. In general, semiconductor device parameters vary widely from one device to another and it is necessary to build some controls into the design. When semiconductor devices are used in the design of circuits whose specifications must be held to very tight tolerances, it is a good idea to use design strategies which rely on ratios of passive components, such as resistances, rather than on the values of the device parameters. In this case it can be shown that the gain of the amplifier is equal to the ratio of the drain impedance Z_D to the source resistance R_s .

Since the mixer has two input signals of different frequencies several problems, such as frequency “pulling” and local oscillator feed-through to the antenna, can be avoided by ensuring that the two sources are well isolated from each other. It is possible to achieve a high level of isolation by using a dual-gate MOSFET. The design process is best illustrated by an example.

Example 3.4.2 The Mixer. Design a mixer for an AM radio using the dual-gate n-channel depletion MOSFET whose characteristics are given in Figure 3.10. The following are specified:

- (1) Supply voltage, $V_{DD} = 12$ V.
- (2) Drain bias current, $I_D = 5$ mA.
- (3) Primary inductance of the drain transformer, $L_p = 250 \mu\text{H}$.
- (4) Centre frequency of the output (intermediate frequency) = 455 kHz.

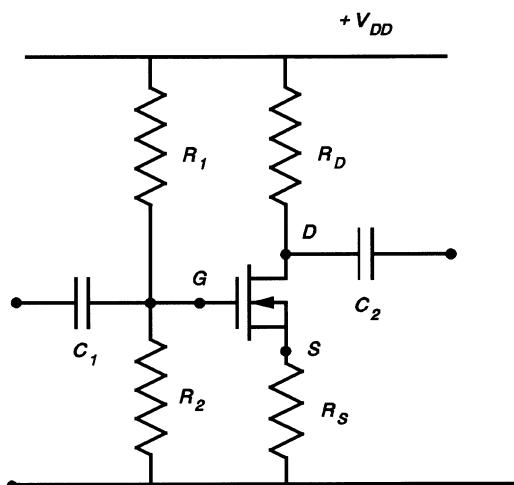


Figure 3.9. A more practical version of the MOSFET amplifier shown in Figure 3.7.

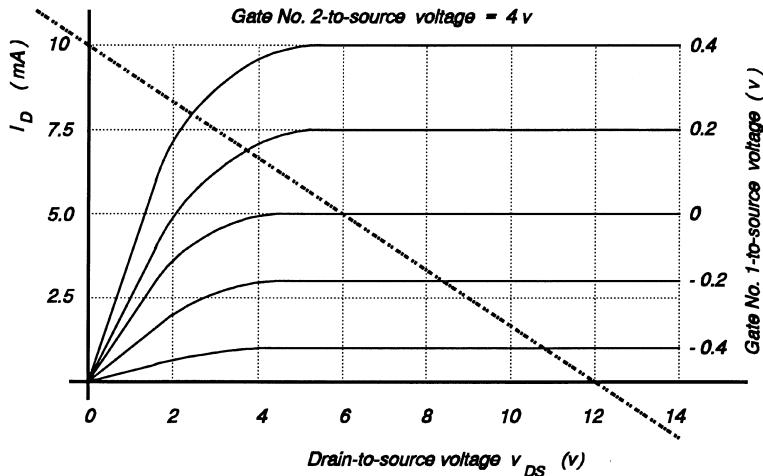


Figure 3.10. The drain characteristics of the dual-gate n-channel MOSFET used in Example 3.4.2.

- (5) -3 dB bandwidth = 20 kHz.
- (6) Transformer turns ratio is 10 : 1.

What is the value of the resistive load that the mixer must “see”, assuming that both the primary and secondary winding resistances are negligibly small?

Solution. A suitable circuit for the mixer is as shown in Figure 3.11.

The capacitance required to tune the drain to 455 kHz (intermediate frequency) is given by

$$\omega^2 = \frac{1}{L_p C_p} \quad (3.4.10)$$

$$C_p = \frac{1}{\omega^2 L_p} = \frac{1}{(2\pi \times 455 \times 10^3)^2 \times 250 \times 10^{-6}} = 489 \text{ pF}. \quad (3.4.11)$$

The bandwidth Δf is related to the center frequency f_0 by

$$Q_0 = \frac{f_0}{\Delta f} = \frac{455 \times 10^3}{20 \times 10^3} = 22.75. \quad (3.4.12)$$

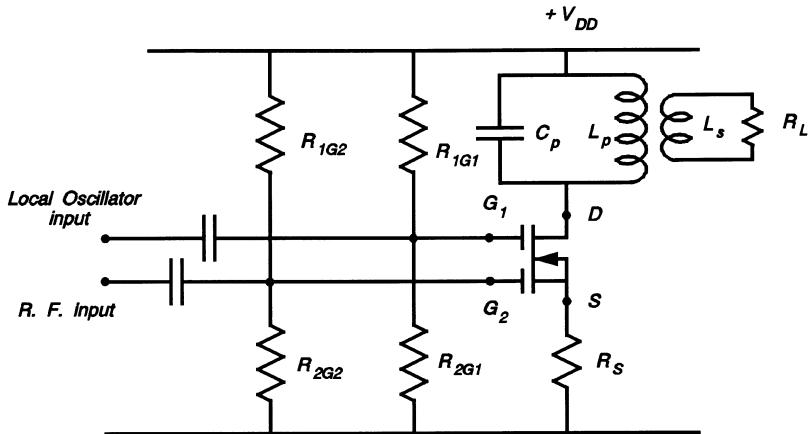


Figure 3.11. A typical MOSFET mixer using the dual-gate n-channel device.

The resistive load transferred to the primary $n^2 R_L$ will be in parallel with L_p . Q_0 for a parallel LR circuit is given by

$$Q_0 = \frac{n^2 R_L}{\omega L_p} = 22.75 \quad (3.4.13)$$

$$n^2 R_L = 2\pi \times 455 \times 10^3 \times 250 \times 10^6 \times 22.75 = 16.26 \times 10^3 \Omega. \quad (3.4.14)$$

Therefore

$$R_L = 162 \Omega. \quad (3.4.15)$$

From the device characteristics given in Figure 3.10, locate $V_{DS} = V_{DD} = 12$ V, and using a straight edge pivoted at this point, determine a point along the line given by $I_D = 5$ mA which will give a wide dynamic range for the drain current. A good point is given by the intersection of $V_{DS} = 6$ V and $I_D = 5$ mA. The load line can now be drawn in as shown. From the slope the required load is 1.2 k Ω . In a common-source amplifier, this load would normally be connected in series with the drain. However, from the dc point of view, the drain is connected to V_{DD} by a short-circuit (through the inductor L_p). The device can be correctly biased by connecting the 1.2 k Ω resistor in series with the source, that is, choose

$$R_s = 1.2 \text{ k}\Omega.$$

With a drain current of 5 mA, the voltage of the source will be

$$V_s = 1.2 \times 10^3 \times 5 \times 10^{-3} = 6.0 \text{ V}. \quad (3.4.16)$$

From Figure 3.10 it can be seen that the required gate-source voltage on gate 1 is 0 V. Gate 1 must therefore be biased at 6.0 V by the resistive chain so that

$$R_{1G1} = R_{2G1}$$

Since no current flows into the gate, the value of the two resistances is arbitrary and can be made as large as practicable. Let

$$R_{1G1} = R_{2G1} = 100 \text{ k}\Omega.$$

From Figure 3.10 it can be seen that the required gate-source voltage on gate 2 is 4.0 V. Because the source is biased at 6.0 V, gate 2 must be biased at 10 V. Again no current flows into gate 2 and therefore the resistance can be made as large as practicable. However

$$R_{1G2} : R_{2G2} = 2 : 10.$$

Choosing $R_{2G2} = 100 \text{ k}\Omega$ makes $R_{1G2} = 20 \text{ k}\Omega$. The coupling capacitors C_1 and C_2 are chosen so that they present negligible impedance to the radio-frequency and local oscillator, respectively.

The drain tank circuit is tuned to resonate at the intermediate frequency; therefore the drain load is

$$n^2 R_L = 16.26 \times 10^3 \Omega. \quad (3.4.17)$$

The gain of the stage is approximately equal to

$$\frac{n^2 R_L}{R_s} = 13.6 = 22.6 \text{ dB}. \quad (3.4.18)$$

This is not the same as the mixer gain which is defined as

$$10 \log_{10} \frac{(\text{intermediate frequency power})}{(\text{radio-frequency power})} \text{ dB}. \quad (3.4.19)$$

In a practical circuit, the relative signal levels of both the radio frequency and the local oscillator will be adjusted to optimize the intermediate-frequency signal.

3.4.3.2 Switching-Type Mixer. The circuit diagram of one of the simplest switching-type mixers is shown in Figure 3.12 [2]. The radio-frequency signal V_s is

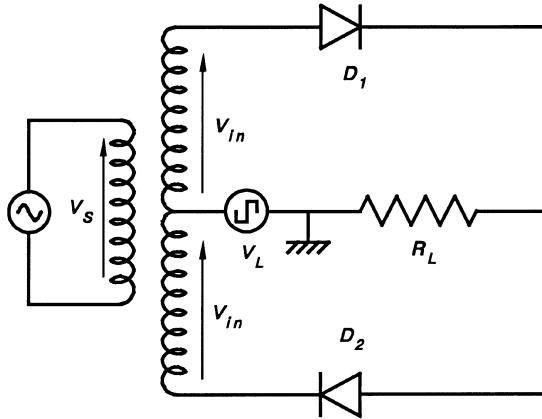


Figure 3.12. The circuit of a switching-type mixer.

a sinusoid. The local oscillator output V_L is a square wave at a frequency which is higher than the radio frequency. The square wave is defined as

$$g(t) = 1 \quad \text{for } 0 > t > \frac{T}{2} \quad (3.4.20)$$

$$g(t) = -1 \quad \text{for } \frac{T}{2} > t > T. \quad (3.4.21)$$

Assuming ideal diodes, and that V_L is larger than V_{in} , then when $V_L > 0$, D_1 conducts and D_2 is off:

$$V_o = V_L + V_{\text{in}} \quad (3.4.22)$$

and when $V_L < 0$, D_2 conducts and D_1 is off:

$$V_o = -(V_L + V_{\text{in}}). \quad (3.4.23)$$

The output voltage is then

$$V_o = V_L + V_{\text{in}} g(t). \quad (3.4.24)$$

This is evident from an examination of Figure 3.13.

The square wave $g(t)$ can be expressed in terms of its Fourier components as

$$g(t) = \frac{4}{\pi} \sum_{n=0}^{\infty} \frac{\sin(2n+1)\omega_L t}{(2n+1)}. \quad (3.4.25)$$

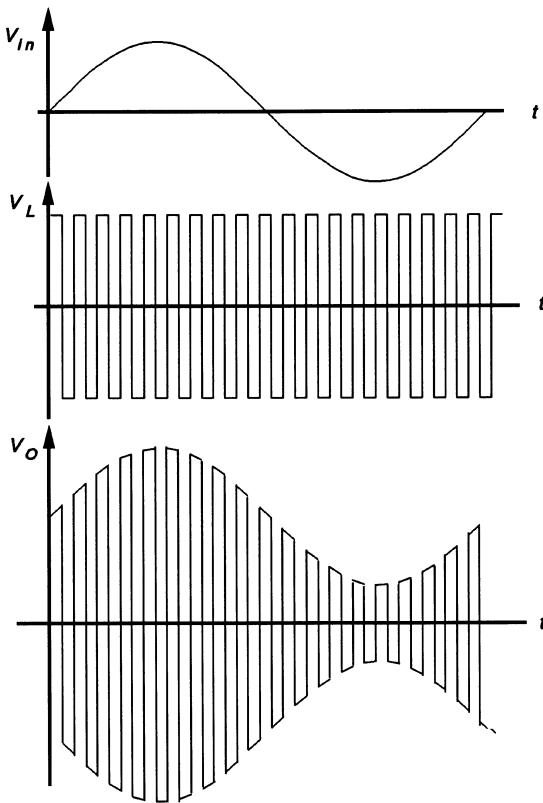


Figure 3.13. Waveforms of the inputs V_{in} and V_L and the output V_o .

But

$$V_{in} = A \sin \omega_C t. \quad (3.4.26)$$

Therefore

$$V_{in} g(t) = \frac{2A}{\pi} \sum_{n=0}^{\infty} \frac{\cos[(2n+1)\omega_L - \omega_C]t - \cos[(2n+1)\omega_L + \omega_C]t}{(2n+1)}. \quad (3.4.27)$$

The output of the mixer consists of the local oscillator frequency and an infinite number of sums and differences of the local oscillator harmonics and the radio frequency. The desired frequency components can be filtered in the intermediate-frequency stage that follows the mixer.

If the AM carrier equation is used in place of Equation (3.4.25), it can be demonstrated that the mixing operation maintains the relationship between the desired intermediate frequency and its sidebands.

The major disadvantages of this type of mixer are as follows:

- (1) The large signal required from the local oscillator to switch the diodes calls for considerable power output from the oscillator and this makes the design of the local oscillator difficult.
- (2) The large local oscillator signal is present in the output of the mixer and it can interfere with the filtering process, especially when the local oscillator frequency is much higher than the radio frequency. Hence the desired sum and different frequencies are close to each other and to the local oscillator. This part of the output can be removed by changing the basic circuit from a single ended to a differential output.

3.4.4 Intermediate-Frequency Stage

The output of the mixer contains a multitude of frequencies made up of the sums and differences of the local oscillator frequency and the radio-frequency signal and their various harmonics. The task at hand then is to select the frequency ($f_{lo} - f_{rf}$), together with its sidebands, and to amplify it if necessary before it is demodulated. A filter is required to achieve this. An ideal filter for this purpose would be one with rectangular characteristics – a flat response across the frequency band and infinitely steep “skirts” with infinite attenuation beyond. A practical filter will, of course, be much less exotic than that.

The only type of frequency selective circuit discussed so far has been the *LC* tuned circuit. A single tuned *LC* circuit can have steep skirts and high attenuation for out-of-band signals when the *Q* factor is high but a high *Q* factor also means that the in-band frequency is very narrow. The design of a filter which can select the intermediate frequency and its sidebands and suppress to an acceptable level all the other frequency components present in the output of the mixer is beyond the scope of this book. The interested reader will find a short list of sources for more information in the bibliography.

In general, a filter is placed between a resistive source and a resistive load which may or may not have the same value. For frequencies in the passband, the filter “matches” the source to the load so that the reflection of the signal from the load is minimal, that is, maximum power is transferred to the load. In the stopband, the filter input presents such a severe mismatch to the source that most of the signal power is reflected with very little reaching the load.

Filters can be classified as follows:

- 1. Passive *LC* Filters.** These filters are made up of only inductors and capacitors. They are considered to be lossless. In general, the closer the filter characteristics are to the ideal, the more *Ls* and *Cs* required. Discrete *LC* filters are used at frequencies from as low as 20 Hz to as high as 500 MHz. The low-frequency limit is set by the low *Q* factors of the inductors and at the high-frequency end by the circuit strays of both *L* and *C*. Passive *LC* filters can be used at frequencies as high as 40 GHz but the components have to be considered to have distributed parameters. Inductors and capacitors are then

“replaced” by open or short-circuited transmission lines and tuned *LC* circuits by resonators. The physical appearance of the components of these very high frequency filters bear no resemblance to ordinary *Ls* and *Cs*.

2. **Crystal Filters.** The very high *Q* factors which can be obtained from crystals offer the filter designer the possibility of high selectivity and steep skirt gradients. However, crystal parameters are, in general, not under the direct control of the filter design and consequently have not been very popular. Crystal filters have an upper frequency limit of a few megahertz.
3. **Active *LC* Filters.** These filters are made up *LC* sections separated by amplifiers. The advantages to be gained from this is that the filter can then have an overall gain instead of a loss and the buffering action of the amplifiers between various segments of *Ls* and *Cs* can reduce the amount of interaction between them, making the tuning of the filter considerably easier. The bandwidth of the amplifier used can limit the range of application of these filters to an upper range of about 300 MHz.
4. **Active *RC* Filters.** These filters are made up of resistors, capacitors and operational amplifiers. The introduction of integrated circuits prompted the development of these filters. Unfortunately, they can be used only at low frequencies with an upper limit of a few megahertz.
5. **Digital Filters.** These filters have become practical since the development of cheap and fast microcomputers with large memories. The signal is sampled by an analog-to-digital converter. The samples are converted into a digital code and multiplied by a function of the desired output characteristics. The resulting signal is fed into a digital-to-analog converter to give an analog output. Even with the fastest microcomputers digital filters are limited to an upper frequency of about 50 kHz.
6. **Mechanical Filters.** In these filters, the electrical signal is converted into a mechanical vibration by a magnetostrictive transducer. A number of mechanical resonators in the form of discs, plates, and rods determine which frequencies will be propagated through the device to reach the output and which will not. At the output another transducer converts the vibration back into an electrical signal. Mechanical filters normally operate in the frequency range 50–600 kHz. These filters have the attraction of being quite accurate and yet inexpensive.
7. **Surface-Acoustic-Wave Filters.** These filters belong to a more general class called transversal filters. In surface-acoustic-wave (SAW) filters, two transducers are fabricated at opposite ends of a highly polished piezoelectric material such as quartz or lithium niobate. When an electrical signal is applied to the input transducer, the material changes its physical shape and, at the appropriate frequency, will cause a travelling wave to be propagated on the surface of the material. At the output, the travelling wave is converted back into an electrical signal. The mechanical shape, size, and placing of the metallization on the piezoelectric material determine which frequencies are propagated and which are attenuated. SAW filters normally operate in the range 20–500 MHz.

8. Switched Capacitor Filters. These filters have been made possible by the ease with which MOS (metal-oxide semiconductor) technology can be used to realize capacitors, switches, and operational amplifiers on the same silicon chip. Two switches and a capacitor can be used to simulate a resistor. It is then possible to construct equivalent circuits for most of the filter structures used in active RC filters. Switched capacitor filters are generally used where accurate filtering is not required. However, the characteristics of switched capacitors continue to improve as they continue to be the subject of intense research interest. The need to switch the capacitors at a much higher frequency (clock frequencies between 30 and 40 MHz) than the highest frequency present in the signal limits these filters to operation below 3 MHz.

3.4.5 Automatic Gain Control

The function of the automatic gain control is to ensure that the signal reaching the demodulator is sufficiently high and within the limits for efficient demodulation. It does this by sensing the level of the signal at the input to the modulator and adjusting the gain of a variable gain amplifier to keep the level constant. In practice it is not possible to boost all signals to a constant level and, in any case, it is undesirable to amplify noise to the same level when that is all that is available. The desirable characteristics of an automatic gain control (AGC) circuit is as shown in Figure 3.14(a). Below the “knee”, the signal is amplified by a constant factor. Above the knee the output is kept constant. A practical characteristic is shown in Figure 3.14(b) where the knee is rounded and the output continues to rise but at a limited rate.

The AGC subsystem consists of a variable gain amplifier whose gain is controlled by a voltage or current derived from the output signal. A block diagram is given in Figure 3.15.

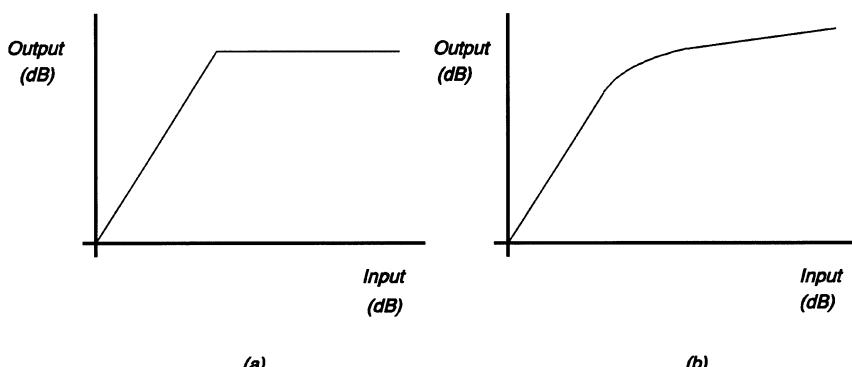


Figure 3.14. (a) The ideal input–output characteristics of an automatic gain control (AGC) circuit. (b) The input–output characteristics of a practical AGC circuit.

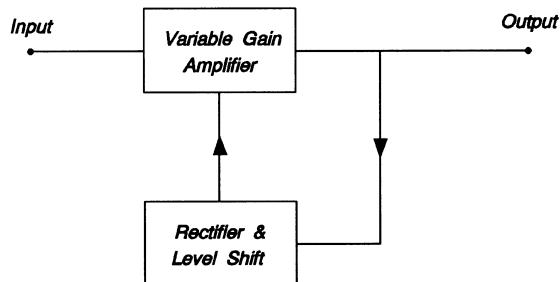


Figure 3.15. A block diagram of a feedback-type AGC circuit.

The rectifier produces a dc signal which is proportional to the signal which appears at the output of the variable gain amplifier. The dc signal is fed back to the variable gain amplifier and attempts to keep the output signal constant above a predetermined level. It is usual to place the variable gain amplifier before the mixer so that the radio-frequency signal reaching the mixer does not vary too widely.

A suitable amplifier for the AGC circuit is the four-quadrant analog multiplier discussed in Section 2.6.3. In the scheme shown in Figure 3.15, the four-quadrant amplifier will be part of the radio-frequency stage with the radio-frequency signal applied to one of its two inputs. The second input is a dc signal derived from the rectified and smoothed intermediate-frequency signal so that, when the intermediate-frequency signal is high, the constant of multiplication is low, and when the signal is low the constant is high.

Because the four-quadrant analog multiplier was discussed at length earlier, only the rectifier and its associated time-constant will be discussed.

A common characteristic of all AGC circuits is that they have a fast "attack" time and a somewhat slower "release" time. This means that the system can capture

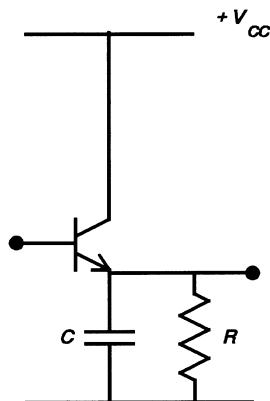


Figure 3.16. A circuit which provides a “fast attack” and “slow release” for the AGC control voltage.

sudden increases in signal level but take a longer time to adjust the gain upwards for low signals. A suitable rectifier and time-constant are shown in Figure 3.16.

The first positive-going signal that appears at the base of the BJT causes it to switch on and very rapidly charge the capacitor C with its emitter current which is much larger than the base current of the BJT. The dc signal required to control a sudden increase in signal is therefore produced very quickly. When the positive-going voltage starts to drop, the charged capacitor holds up the voltage of the emitter and the falling base voltage ensures that the BJT is cut off. The discharge of the capacitor is then controlled by the resistor in parallel with it. The attack time can be slowed down by connecting a small resistor in series with the capacitor. The correct attack and release times have to be determined by experiment.

3.4.6 Demodulator

The input to the demodulator is a carrier of frequency 455 kHz (intermediate frequency) with an amplitude envelope determined by the audio signal. The circuit used for demodulation is therefore the envelope detector which is shown in Figure 3.1. The operation of the circuit was discussed in Section 3.2. An example of an envelope detector follows.

Example 3.4.3 Envelope Detector. The amplitude of the signal applied to the input of an envelope detector is 4 V peak when the modulation index is zero and the intermediate frequency is 455 kHz. Choose a suitable time-constant for the detector to avoid diagonal clipping when the modulation index is 0.8 and the highest audio frequency component in the input signal is 10 kHz. Assume that the diode is ideal.

Solution. The output waveform with the capacitor removed will be as shown in Figure 3.17.

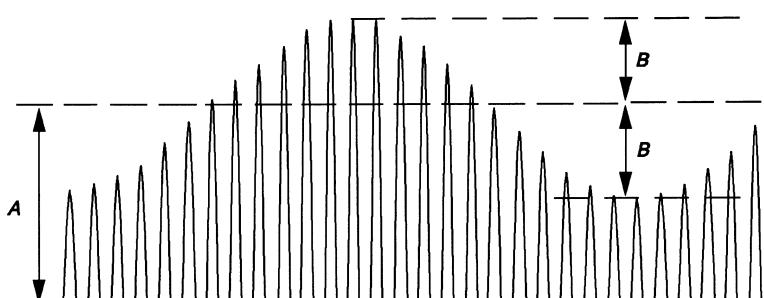


Figure 3.17. The output voltage waveform of the envelope detector used in Example 3.4.3.

Because $B/A = 0.8$, $B = 3.2$ V peak. The equation for the envelope is

$$v_s(t) = B \cos \omega_s t \quad (3.4.28)$$

where $\omega_s = 2\pi \times 10 \times 10^3$. The slope of the envelope is

$$\frac{dv_s}{dt} = -B\omega_s \sin \omega_s t. \quad (3.4.29)$$

From the diagram, it is clear that the maximum slope occurs at $t = T/4$, when $\sin \omega_s t = 1$. Hence

$$\left. \frac{dv_s}{dt} \right|_{\max} = -B\omega_s. \quad (3.4.30)$$

When the capacitor is connected to the circuit and the diode is not conducting, the output voltage decays according to the equation

$$v_c(t) = V e^{-t/\tau} \quad (3.4.31)$$

The slope of the decaying capacitor voltage is

$$\frac{dv_c}{dt} = -\left(\frac{1}{\tau}\right)V e^{-t/\tau}. \quad (3.4.32)$$

From the diagram it is clear that, when $t = T/4$ or when $\omega_s t = \pi/2$, $V = A$ and

$$\frac{dv_c}{dt} = -\left(\frac{1}{\tau}\right)A e^{-t/\tau}. \quad (3.4.33)$$

The maximum value occurs at $t = 0$:

$$\frac{dv_c}{dt} = -\frac{A}{\tau}. \quad (3.34)$$

To avoid diagonal clipping, the capacitor voltage must decay faster than the envelope voltage, that is,

$$\left| -\frac{A}{\tau} \right| \geq \left| -B\omega_s \right| \quad (3.4.35)$$

$$\tau = \frac{A}{B\omega_s} = \frac{4.0}{3.2 \times 2\pi \times 10 \times 10^3} = 19.89 \times 10^{-6} \text{ s.} \quad (3.4.36)$$

Let $R = 10 \text{ k}\Omega$, then $C = 1989 \text{ pF}$.

3.4.7 Audio-Frequency Amplifier

The audio-frequency amplifier was discussed in Section 2.7. A few minor changes may be necessary in the design of the final stage of the amplifier so that it can drive a loudspeaker efficiently.

3.4.8 Loudspeaker

So far, the audio-frequency signal is in an electrical form. To convert it to an acoustic signal, an electrical-to-acoustic transducer is required. This can take two basic forms: the loudspeaker and the headphone. The more common of the two devices is the loudspeaker [3, 4]. There are two basic types of loudspeakers: the direct radiation and the horn loudspeaker. The diaphragm of the direct radiation loudspeaker is coupled directly to the air whereas that of the horn loudspeaker (as the name suggests) is through a horn. The horn loudspeaker is generally more efficient due to the better matching between its diaphragm and the air provided by the horn. However, it is usually large, heavy, and expensive. A direct radiation loudspeaker, by contrast, has a simple construction, occupies a small space, and has a relatively uniform response over a moderate frequency band. The design of loudspeakers is a highly specialized field, and therefore outside the scope of this book. However, a qualitative description of the behavior of loudspeakers is a subject that should be of interest to radio engineers.

A cross-section of the direct radiation loudspeaker is given in Figure 3.18.

It consists of a paper cone, the apex of which is mechanically connected to a cylinder on which the voice coil is wound. The voice coil is located in a magnetic field provided by a permanent magnet. The outer rim of the cone is corrugated to permit the cone to move along its axis with the minimum of mechanical resistance. The whole cone may be viewed as the diaphragm of the loudspeaker. The

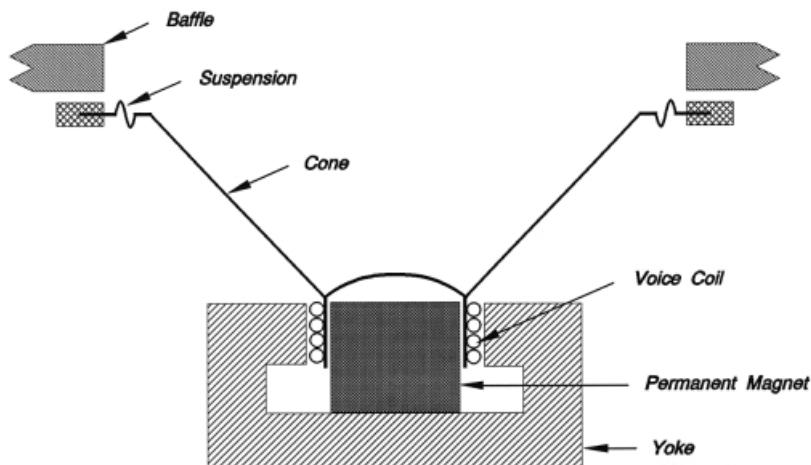


Figure 3.18. A cross-sectional view of a typical loudspeaker.

corrugation is referred to as the suspension of the cone. The outer edge of the suspension is fastened to the frame of the loudspeaker and the frame is secured to the baffle.

An ideal loudspeaker should produce a uniform sound pressure at all frequencies with non-directional characteristics, have a good dynamic range, high efficiency, and produce no distortion. The characteristics of the practical loudspeaker are quite different from the above and the reason is that the properties required to produce a low-frequency vibration are quite different from those required to produce a high-frequency sound. Furthermore, the environment in which the loudspeaker operates has a significant effect on its performance.

When a current is passed through the voice coil, a force is exerted on the voice coil by the interaction of the flux due to the permanent magnet and the flux generated by the current in the voice coil. The result is that the voice coil moves and so does the cone to which it is attached. When an ac signal is applied, the cone will vibrate in accordance with the signal, producing a sound. The cone, the voice coil, and the cardboard former on which it is wound have a finite mass, as has the air which the cone has to displace as it moves back and forth. Although the suspension is designed to offer the minimum mechanical resistance to motion, this resistance is not zero. When suitable assumptions have been made about the stiffness of the cone and the linearity of the suspension system and so on, a differential equation describing the motion of the cone can be written and used to study the effects of changing the parameters of the system on its performance. The following broad conclusions can be drawn from both theory and practice.

- (1) The force driving the cone is proportional to the magnetic flux density in the air gap (B), the length of the conductor in the voice coil (l) and the current flowing in the coil (i):

$$f = Bli. \quad (3.4.37)$$

A high flux density can be obtained by using a large powerful magnet in conjunction with a small air gap. The minimum air gap is dictated by the thickness of the voice coil former and the diameter of the conductor used to wind the coil and the need to keep the voice coil from touching the magnet assembly. The longer the wire in the coil the greater its mass and, since it is part of the cone, a larger force will be required to drive it. A large coil current makes demands on the amplifier required to drive it.

- (2) A good low-frequency response requires a large (area) cone. A large cone will, in general, have a relatively large mass. However, a good high-frequency response requires a cone of very small mass. It would appear that it is not possible to use one cone for the complete range of frequencies in the audio signal. A reasonable compromise, one which comes with a suitable price tag, is to use a filter to divert the low-frequency components of the signal to a large loudspeaker (woofer) and the high frequency to a small loudspeaker (tweeter).

- (3) When the frequency is low, the cone moves as one lumped mass, that is, all points on the cone remain in the same position relative to each other as the cone moves back and forth. As frequency increases, the cone starts to behave like a distributed mass and a phase difference between different parts occur. The result is that the response becomes irregular, with peaks and troughs depending on whether the radiated signals from the different parts of the cone add or subtract from each other at a given point within its field of radiation.
- (4) The sound wave from the front of the loudspeaker must be separated from the back wave. Otherwise, at low frequency the high pressure created by the forward movement of the cone is neutralized by the low pressure created at the back of the cone. The energy put into the loudspeaker simply moves the air around the edge of the loudspeaker instead of being converted into radiated energy. This can be prevented by mounting the loudspeaker on a large baffle whose dimensions are at least greater than one-quarter the wavelength of the lowest frequency. Similar results can be obtained by enclosing the back of the loudspeaker in a closed box.
- (5) The cone, as a mass supported on its suspension, is a system which can have a resonant response when suitably excited. It so happens that the resonant frequency is usually in the low-frequency end of the loudspeaker response. By adjusting the mass of the cone, the characteristics of the suspension and the damping, the peak of the resonant response can be moved to a frequency that improves the low-frequency performance of the loudspeaker.

The loudspeaker will generate harmonic distortion when the force-displacement characteristics of the cone suspension are non-linear. Another source of distortion is the non-uniformity of the flux density in the air gap, especially when the voice coil is at the extremities of its excursion.

3.5 SHORT-WAVE RADIO

Frequencies between 3 MHz and 30 MHz are set aside for use in long distance radio communication. This frequency band is generally referred to as the short-wave band. The wavelength that lies between 1 m and 100 m is short compared to that of the medium wave. It was pointed out in Section 2.9 that the increased frequency of the carrier used in short-wave communication gives its electromagnetic transmission wave increasingly directional properties. Transmission antennae can be designed to concentrate the power of the transmitter in a given direction. The ionosphere and the Earth's surface are used as reflectors to direct the signals from the transmitter to the receiver when very long distances have to be covered.

The superheterodyne arrangement is used in short-wave radio receivers. However, because the radio frequency is much higher the intermediate frequency used is also high. Intermediate frequencies in use with short-wave radio are usually between 1.5 and 28 MHz. In some special cases, a double superheterodyne system is used. The

radio frequency is translated to the first intermediate frequency, a high frequency, and then translated the second time to a lower frequency, between 455 kHz and 5 MHz. The advantage of this system is improved selectivity and image rejection.

REFERENCES

1. Lovering, W. F., *Radio Communication*, Longmans, Green & Co. Ltd., London, 1966.
2. Smith, J., *Modern Communication Circuits*, McGraw-Hill, Inc., New York, 1986.
3. Olson, H. F., *Elements of Acoustical Engineering*, Van Nostand, Inc., New York, 1947.
4. Hutchinson, C. L. (Ed), *The ARRL Handbook for the Radio Amateur*, The American Radio Relay League, Newington, CT, 1988.

BIBLIOGRAPHY

- Johnson, D. E., *Introduction to Filter Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1976.
 Herrero, J. L. and Willoner, G., *Synthesis of Filters*, Prentice-Hall, Englewood Cliffs, NJ, 1966.
 Humphreys, D. S., *The Analysis and Synthesis of Electrical Filters*, Prentice-Hall Inc., Englewood Cliffs, NJ, 1970.
 Zverev, A. I., *Handbook of Filter Synthesis*, John Wiley, New York, 1967.
 Van Valkenburg, M. E., *Analog Filter Design*, Holt, Rinehart and Winston, New York, 1982.
 Huelsman, L. P. (Ed). *Active Filters: Lumped, Distributed, Integrated, Digital and Parametric*, McGraw-Hill, New York, 1970.
 Haykin, S. S., *Synthesis of RC Active Filter Networks*, McGraw-Hill, London, 1969.
 Sedra, A. S. and Brackett, P. O., *Filter Theory and Design: Active and Passive*, Matrix, Inc., Champaign, IL, 1978.

PROBLEMS

- 3.1** An AM receiver is tuned to 1000 kHz but two other stations (the first operating at 1010 kHz and the second operating at 1200 kHz) are causing interference. The intermediate frequency of the receiver is set at 455 kHz. Explain with the aid of suitable calculations which of the two stations is most likely to cause the interference. The intermediate frequency is now set at 100 kHz. Will this change improve the situation? Explain your answer fully.
- 3.2** A diode which has characteristics described approximately by

$$i = 5 + v + 0.05v^2 \mu\text{A}$$

is to be used in a demodulator of an AM receiver. The carrier frequency is 1 MHz with a 100% modulation by a 5 kHz sinusoidal signal. Assuming that

both the carrier signal source and the local oscillator have a 5 V output on open circuit and a $1\text{k}\Omega$ internal resistance, design a suitable (single-tuned LC) circuit using the diode, transformers (inductors) and capacitors to select the desired intermediate-frequency signal and to supply maximum power to a resistive load 550Ω . Provide a suitable circuit diagram and specify the values of all the circuit elements used such that the sideband signals lie within the -3 dB bandwidth. Calculate the mixer gain in decibels.

- 3.3** A switching-type mixer, such as shown in Figure 3.12, is driven from an ideal sinusoidal voltage source of value 5 V peak. The local oscillator voltage (squarewave source) is assumed to be sufficient to cause instantaneous switching of the diodes. Calculate the amplitudes of the output signal for $n = 1, 2$, and 3 and their frequencies if the sinusoidal source has a frequency of 1 MHz and the squarewave is 8 MHz.
- 3.4** The secondary winding of a transformer has self-inductance $L = 500\mu\text{H}$ and resistance of 15Ω . The primary current is an amplitude modulated wave with a carrier frequency of 500 kHz modulated by a single-frequency sinusoidal signal of 7.5 kHz to a depth of 50%. Calculate the value of the capacitor required to tune it and the amplitude of the upper or lower sideband signal relative to the carrier in the secondary winding.
- 3.5** The collector load of a bipolar transistor amplifier consists of a coil of self-inductance $L_1 = 1\text{ mH}$ whose Q factor is equal to 100 when tuned to resonance at 500 kHz by a parallel capacitor which may be assumed to be lossless. The amplifier drives a resistive load through a perfectly coupled coil, the mutual inductance being equal to 0.5 mH . If the resistive load has a value $6.25\text{k}\Omega$, determine the equivalent impedance of the collector load. What is the -3 dB bandwidth of the circuit? Calculate the value of the emitter resistance for a voltage gain of 25.
- 3.6** Using the MOSFET whose characteristics are given in Figure P3.1, design a class-A amplifier with a voltage gain of 10. Allow for a drain current of 3 mA when the dc supply voltage is 12 V. What is the maximum sinusoidal input voltage that the amplifier can take without clipping? Calculate the input impedance of the amplifier.
- 3.7** Repeat the design using the device whose characteristics are given in Figure P3.2.
- 3.8** Using the device of Figure P3.1, design a transformer coupled class-A amplifier to operate at $500/2\pi\text{kHz}$ with a voltage gain of 10 and a -3 dB bandwidth of $20/2\pi\text{kHz}$. The load resistance is 100Ω and the dc supply voltage is 15 V. Specify the values of all components used. You may assume that the windings of the transformer have negligible resistances.

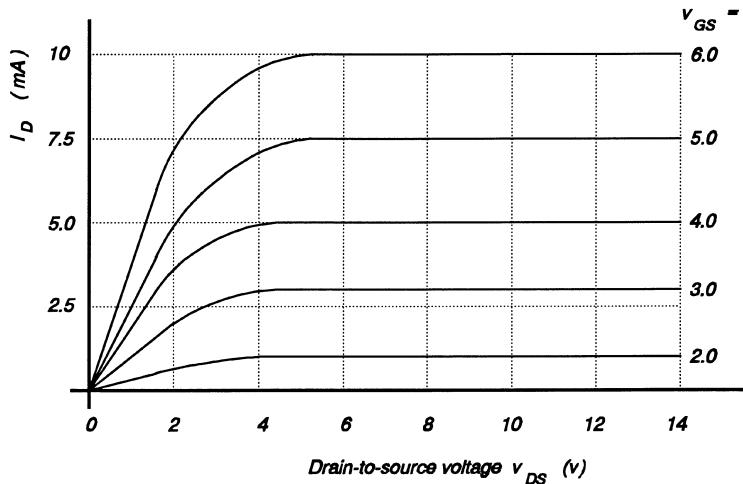


Figure P3.1.

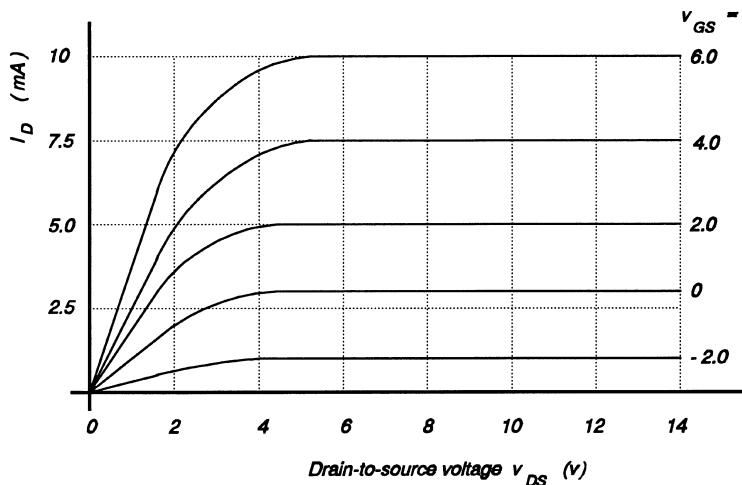


Figure P3.2.

- 3.9 An AM signal applied to an envelope detector has carrier frequency $\omega_c = 250$ krad/s. The upper envelope is a triangular waveform which goes from +10 V to +2 V and back up to +10 V at a constant rate in 0.2 ms (the lower envelope is similar!). Assuming that the diode is ideal, calculate the maximum time-constant of the detector load to avoid diagonal clipping. Justify any approximation that you make.

4

FREQUENCY MODULATED RADIO TRANSMITTER

4.1 INTRODUCTION

In Chapter 2, the amplitude of a high-frequency (carrier) sinusoidal signal was varied in accordance with the waveform of an audio-frequency (modulating) signal to give an amplitude modulated (AM) wave which could be transmitted, received, and demodulated to recover the original audio frequency signal.

In frequency modulated (FM) radio, the frequency of the carrier is varied about a fixed value in accordance with the amplitude of the audio frequency. The amplitude of the carrier is kept constant. The waveform of a sinusoidal carrier modulated by a saw-tooth wave is shown in Figure 4.1.

All signals carried on any transmission system will sooner or later be contaminated by noise so the susceptibility of the communication system to noise is an important consideration. The noise can be defined as a random variation superimposed on the signal. In AM systems, the information to be transmitted is contained in the envelope of the carrier signal. The noise therefore appears on the envelope and has a direct role in corrupting the signal. In FM systems the information to be transmitted is contained in the variation of the frequency of the carrier about a pre-set value. The amplitude of the FM signal is kept constant and, indeed, if there are changes in the amplitude of the FM signal, they are removed by clipping before demodulation. By comparison, FM systems are less susceptible to degradation by noise.

4.2 FREQUENCY MODULATION THEORY

While a saw-tooth modulating signal provides a simple picture of the FM signal, a sinusoidal modulating signal is the simplest for the derivation of the mathematical expressions to describe the FM signal.

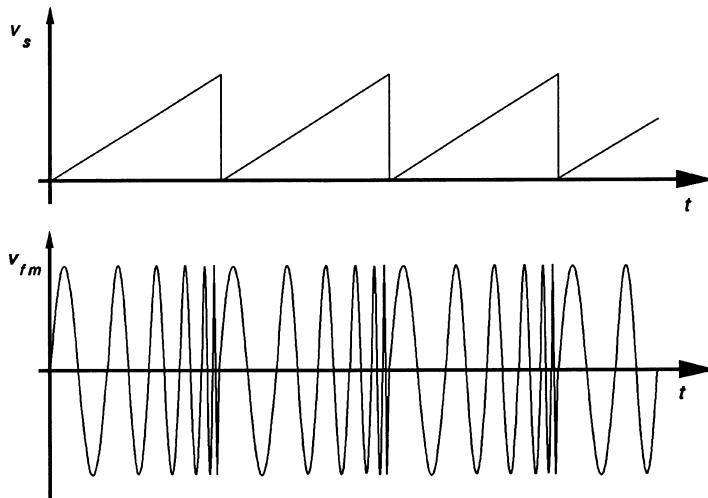


Figure 4.1. The sawtooth waveform v_s frequency modulates a carrier to give the output v_{fm} . Note that the relative change in frequency has been exaggerated for clarity. In normal FM radio, the change in frequency relative to the carrier is less than 0.15%

A sinusoidal voltage can be expressed as:

$$v = A \sin \omega t \quad (4.2.1)$$

$$v = A \sin \theta(t) \quad (4.2.2)$$

where ω is a constant representing the angular velocity of the sinusoid and θ is a phase angle with respect to an arbitrary datum. In general, the relationship between the phase angle and the angular velocity is given by

$$\frac{d\theta(t)}{dt} = \omega(t). \quad (4.2.3)$$

In a frequency modulated system, ω is varied about a fixed value ω_c , in accordance with the modulating signal which is assumed in this case also to be a sinusoid:

$$v_s = B \cos \omega_s t. \quad (4.2.4)$$

The instantaneous angular velocity

$$\omega_i = \omega_c + \Delta\omega_c \cos \omega_s t \quad (4.2.5)$$

where ω_c is the long-term mean angular velocity, $\Delta\omega \ll \omega$, and $\Delta\omega_c$ is the maximum deviation of the angular velocity about ω_c .

Substituting instantaneous values into Equation (4.2.3)

$$\omega_i(t) = \frac{d\theta_i(t)}{dt}. \quad (4.2.6)$$

Substituting Equation (4.2.5) into Equation (4.2.6) and integrating

$$\theta_i(t) = \phi + \int_0^t \omega_c dt + \int_0^t \Delta\omega_c \cos \omega_s t dt \quad (4.2.7)$$

where ϕ is the initial value of the phase angle which without loss of generality can be set equal to zero. Then substituting Equation (4.2.7) into Equation (4.2.2) gives

$$v_{fm}(t) = A \sin(\omega_c t + \frac{k\omega_c}{\omega_s} \sin \omega_s t) \quad (4.2.8)$$

where $k\Delta\omega_c = B$, the amplitude of the modulating signal.

The term $k\omega_c/\omega_s$ is called the *modulation index*, where

$$m_f = \frac{\text{max. frequency deviation of the carrier}}{\text{modulating signal frequency}} \quad (4.3.9)$$

$$v_{fm}(t) = A \sin(\omega_c t + m_f \sin \omega_s t). \quad (4.2.10)$$

Expanding,

$$v_{fm}(t) = A[\sin \omega_c t \cos(m_f \sin \omega_s t) + \cos \omega_c t \sin(m_f \sin \omega_s t)]. \quad (4.2.11)$$

The terms $\cos(m_f \sin \omega_s t)$ and $\sin(m_f \sin \omega_s t)$ can be expanded in the form of Fourier series with coefficients which are Bessel functions of the first kind $J_n(m_f)$ where n is the order and m_f is the argument, to give

$$\cos(m_f \sin \omega_s t) = J_0(m_f) + 2 \sum J_{2n}(m_f) \cos 2n\omega_s t \quad (4.2.12)$$

and

$$\sin(m_f \cos \omega_s t) = 2 \sum J_{2n+1}(m_f) \sin(2n+1)\omega_s t. \quad (4.2.13)$$

Substituting Equations (4.2.12) and (4.2.13) into Equation (4.2.11) and using

$$\cos x \sin y = \frac{1}{2}[\cos(x+y) + \cos(x-y)] \quad (4.2.14)$$

and

$$\sin x \sin y = -\frac{1}{2}[\cos(x+y) + \cos(x-y)] \quad (4.2.15)$$

the result is

$$\begin{aligned} v_{\text{fm}}(t) = & A \{ J_0(m_f) \sin \omega_c t \\ & + J_1(m_f) [\sin(\omega_c + \omega_s)t - \sin(\omega_c - \omega_s)t] \\ & + J_2(m_f) [\sin(\omega_c + 2\omega_s)t - \sin(\omega_c - 2\omega_s)t] \\ & + J_3(m_f) [\sin(\omega_c + 3\omega_s)t - \sin(\omega_c - 3\omega_s)t] + \dots \} \end{aligned} \quad (4.2.16)$$

where $J_n(m_f)$ is plotted against m_f for various values of n in Figure 4.2.

From Equation (4.2.16) it can be seen that:

- (1) The carrier frequency ω_c is present and its amplitude is determined by the modulation index m_f .
- (2) The next term represent two frequencies which are the sum $(\omega_c + \omega_s)$ and difference $(\omega_c - \omega_s)$ of the carrier and the modulating frequency with an amplitude $J_1(m_f)$.
- (3) The next two terms have an amplitude $J_2(m_f)$ and frequencies $(\omega_c + 2\omega_s)$ and $(\omega_c - 2\omega_s)$.
- (4) There are an infinite number of sums and difference of the carrier and integer multiples of the modulating signal.

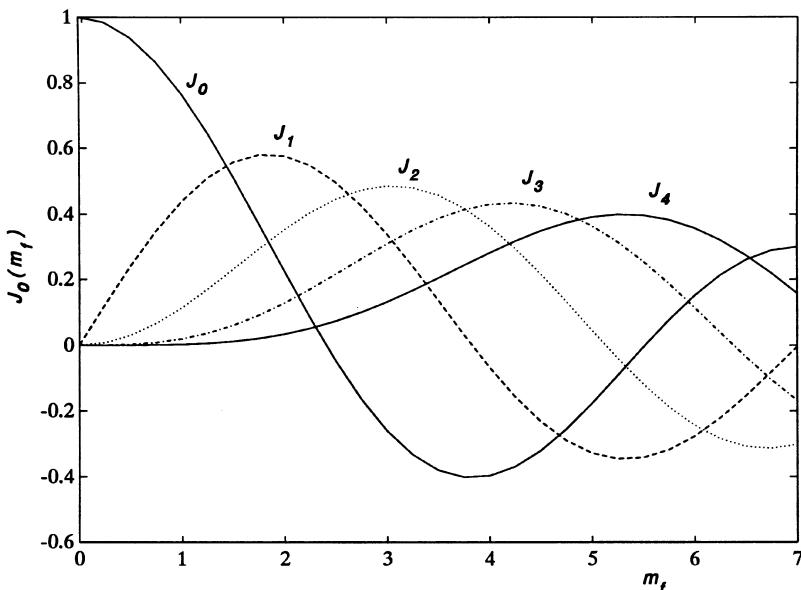


Figure 4.2. A plot of Bessel functions of the first kind, $J_n(m_f)$ against m_f for $n = 0, 1, 2$ and 3 . The values of $J_n(m_f)$ are used to calculate the amplitudes of the side-frequencies present in the FM signal.

It would appear that, in order to transmit a simple sine wave in an FM system, an infinite bandwidth is required to accommodate all the multiple sidebands. However, from Figure 4.2 it can be seen that, as n increases, the amplitudes of the sidebands decrease and their contribution to the signal power falls off rapidly. A second aspect of the bandwidth requirements of the FM system can be seen from Equation (4.2.8). Unlike the AM system, where a modulation index greater than unity causes severe distortion, the modulation index in FM does not appear to have an upper limit except that, for a fixed modulating signal frequency ω_s , increasing the modulation index m_f means a greater deviation of the FM signal from the carrier frequency. This implies a larger bandwidth. With no apparent technical limits to the bandwidth requirements and to permit the maximum number of FM stations to function with minimum distortion and interference, the maximum frequency deviation for commercial FM radio is set at ± 75 kHz. However, since there is substantial signal power beyond the ± 75 kHz limit, the actual bandwidth is set at 200 kHz, with the highest modulating frequency limited to 15 kHz.

Frequency modulated communication channels can be found from approximately 1600 kHz to 4000 MHz. Parts of this spectrum are reserved for the use of police, VHF and UHF television sound channels, VHF mobile communications, and point-to-point communication. The commercial FM radio broadcast band is from 88 to 108 MHz.

4.3 THE PARAMETER VARIATION METHOD

4.3.1 Basic System Design

A simple method for generating an FM signal is to start with any LC oscillator. The frequency of oscillation is determined by the values of C and L . If a variable capacitor ΔC is connected in parallel with C and the capacitance variation is proportional to the modulating signal then an FM signal will be obtained.

Consider a negative conductance oscillator as shown in Figure 4.3. The negative conductance may be obtained from a tunnel diode, a suitable biased pentode, or a bipolar junction transistor with a suitable feedback circuit.

The frequency of oscillation is given by

$$\omega_c^2 = \frac{1}{LC}. \quad (4.3.1)$$

When a variable capacitor ΔC is connected in parallel with C , the frequency of oscillation will be:

$$(\omega_c \pm \Delta\omega_c)^2 = \frac{1}{L(C \mp \Delta C)}. \quad (4.3.2)$$

Any p-n junction diode in reverse bias can be used to realise ΔC . However the relationship between ΔC and the bias voltage is non-linear and appropriate steps will

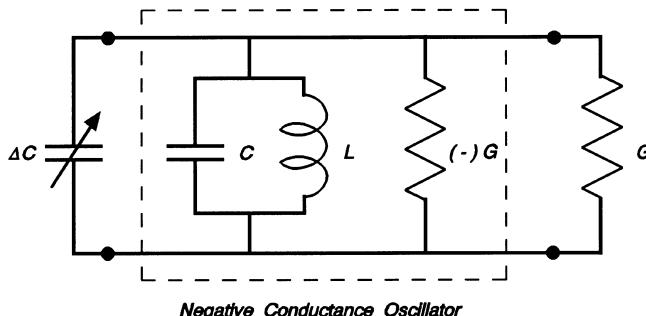


Figure 4.3. The basic FM generator; the value of the variable capacitor ΔC is controlled by the modulating signal which in turn determines the frequency of the oscillator.

have to be taken, such as using two diodes in push-pull or specially fabricated diodes (e.g. varactor diodes or diodes with hyper-abrupt junctions, both of which come under the classification *voltage-controlled capacitors*).

The direct generation of FM signals presents further difficulties. Assuming that the transmitter were to operate at 100 MHz, the required inductance and capacitance would be approximately 160 nH and 16 pF, respectively. Stray inductances and capacitances associated with the circuit will render it impractical. A less impractical idea is to generate the FM signal at a lower frequency, where larger values of L and C are required, and then to use a cascade of frequency multipliers to raise the frequency to the required value. Supposing that the low-frequency FM signal is generated at 200 kHz. Suitable values for L and C are 1.27 mH and 500 pF. Since the frequency has to be multiplied by 500 (actually $2^9 = 512$) to place it within the commercial FM frequency band, the maximum frequency deviation has to be $75 \text{ kHz}/500 = 150 \text{ Hz}$ so that the frequency deviation will not exceed the legal limit. The required change in capacitance $\Delta C = 2.5 \text{ pF}$. This is within the range of capacitance change that can be obtained from a p-n junction diode. A block diagram of the system is given in Figure 4.4.

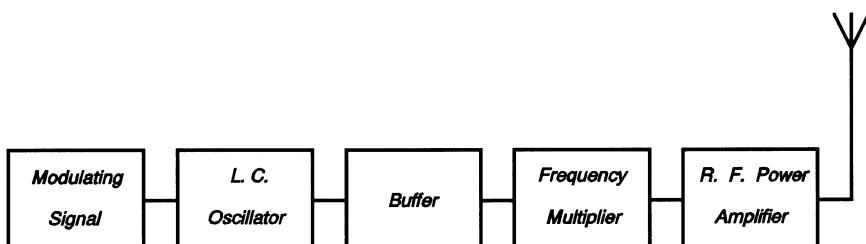


Figure 4.4. An improved FM generator design in which the FM signal is generated at a lower frequency and then multiplied by a suitable factor to give the required frequency.

The design of all the blocks shown in Figure 4.4 was discussed earlier. However, frequency stability requirements of the system makes the approach shown in Figure 4.4 impractical.

One of the requirements of a transmitter is that its frequency will remain at the assigned value at all times so that its frequency does not drift and cause interference with other channels of communication. In the scheme shown in Figure 4.4 there is no built-in mechanism to ensure that the long-term average frequency will be constant. A modified scheme is shown in Figure 4.5, in which the carrier frequency of the FM signal is compared with that of a crystal-controlled oscillator and the resulting error signal is used to correct the carrier frequency.

4.3.2 Automatic Frequency Control of the FM Generator

The main oscillator is an *LC*-type similar to the negative conductance oscillator discussed earlier. The variation ΔC generates the FM signal. The output is fed into a buffer amplifier which provides isolation between the oscillator and its load so that a changing load will have minimal effect on the oscillator operation. The signal proceeds to the *limiter* which removes any AM which may have occurred and also filters out the harmonics generated by the limiter. Part of the amplitude-limited FM signal is then fed to a mixer whose other input is from a crystal-controlled oscillator. The difference frequency is suitably amplified, filtered and then fed into a *discriminator* to produce a dc signal proportional to the difference between the required and the crystal-controlled oscillator frequency. The dc or error signal is

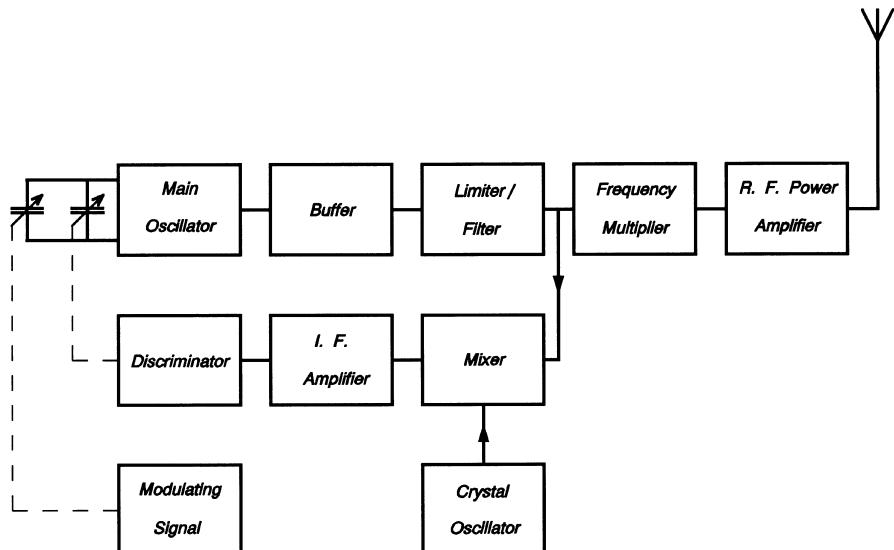


Figure 4.5. The circuit in Figure 4.4 has been improved by the addition of an automatic frequency control (AFC) circuit to stabilize the carrier frequency.

used to keep the main oscillator at the required frequency. Using this scheme, it is possible to impose the stability of the crystal-controlled oscillator on the main oscillator. It should be noted that the time-constant of the error signal should be chosen so it can correct for the long-term frequency deviation of the main oscillator without affecting the short-term frequency deviation due to the modulating signal. The output of the limiter/filter is fed into a chain of multipliers to bring the frequency to its operating value. A power amplifier then boosts the signal power to the desired value and drives the antenna which radiates the signal.

4.3.3 Component Design with Automatic Frequency Control

Two new component blocks were introduced into the system to realize the parameter variation method. These were the limiter and the discriminator. The circuit design of these component blocks follow.

4.3.3.1 The Amplitude Limiter. An FM signal, by definition, must have a constant amplitude. In practice, circuit non-linearities cause variations in the envelope of the FM signal related to the modulating signal, that is, some amplitude modulation takes place. It is evident that any AM present in an FM system will interfere with the signal during the demodulation process since most FM demodulators convert the variation of frequency to a variation of amplitude before detection. A second reason for limiting the amplitude of the FM signal is that the noise present in the communication channel generally rides on the envelope of the signal and therefore by clipping the amplitude of the signal some of the noise can be removed.

The ideal amplitude limiter can accept an input signal of any amplitude and convert it to one of constant amplitude. A practical approximation to the performance of the ideal limiter is obtained by amplifying the input by a large factor and clipping a small portion symmetrically about the time axis. The output waveform will then be a square wave. In order to get back the sinusoid, the signal must be passed through a bandpass filter which will remove all the harmonics leaving only the fundamental. The amplitude limiter has three identifiable parts: the pre-amplifier, the symmetrical clipper, and the bandpass filter. The sub-system block diagram is shown in Figure 4.6.

In Section 4.2, it was shown that an FM signal has an infinite number of sidebands but that it was not necessary to preserve all the sidebands in order to maintain a high level of fidelity. Working with the permitted modulation index, $m_f = 5$, and retaining sidebands with coefficients $J_n(m_f)$ greater than 0.01 (i.e., 1% of the unmodulated carrier) it can be shown that the first eight sidebands must be

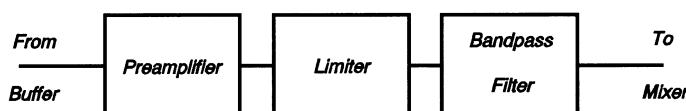


Figure 4.6. A block diagram showing the details of the limiter/filter shown in Figure 4.5.

preserved. Since the sidebands are spaced ω_s apart, and ω_s has a maximum value of 15 kHz, the required bandwidth is 240 kHz.

The (-3 dB) Q factor of the pre-amplifier with the required bandwidth will be just over 400. However, this would mean that the outermost sidebands will be subjected to a 3 dB attenuation compared to those nearer the carrier, causing further variation in the amplitude of the signal. In a practical situation, the modulating signal is made up of a large number of discrete frequencies and this causes further complications in the calculation of the required bandwidth. An amplifier with a flat frequency response could be used but it will have the disadvantage of amplifying the noise below and above the spectrum of interest. A compromise would be to use a tuned pre-amplifier with a much lower Q factor. It will have the advantage of a minimal variation in its response over the spectrum occupied by the significant sidebands and yet attenuate the noise present in the rest of the spectrum. A loaded Q factor of about 20 should be adequate.

It is accepted practice to generate the FM at a lower frequency and to use a cascade of frequency multipliers to bring it up to the operating value. If the FM is generated at 200 kHz, the appropriate frequency deviation will be about 150 Hz. The corresponding bandwidth to accommodate the first 8 sidebands will be 470 Hz.

Figure 4.7 shows a circuit diagram of the pre-amplifier and its load, the symmetrical clipper. The detailed design of a similar amplifier was presented in Section 3.4.2 and will not be repeated here. However, it is worth noting that the diodes can be considered as short-circuits and hence the resistance R will be in parallel with the tuned circuit.

4.3.3.2 Symmetrical Clipper. A simple symmetrical clipper is shown in Figure 4.7. The resistance R is chosen so that the Q factor of the tuned circuit is

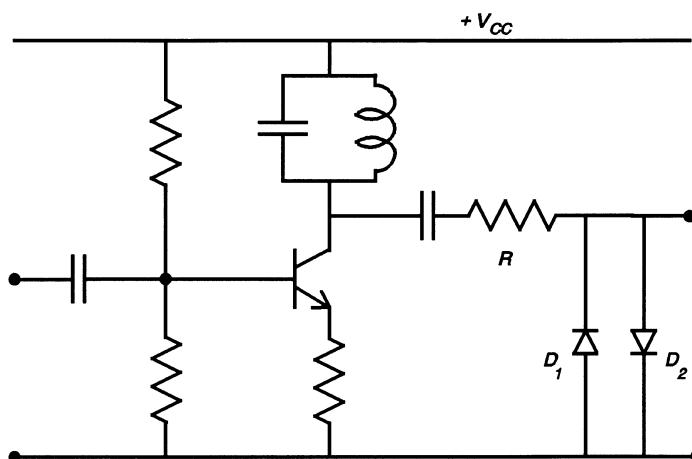


Figure 4.7. The circuit diagram of the pre-amplifier and the symmetrical clipper.

approximately equal to 20. The details of the design of a parallel tuned circuit with a specified Q factor was discussed in Section 3.4.2.

The output of the clipper will be about 0.7 V when single silicon diodes are used or multiples of 0.7 V, depending on how many diodes are connected in series to determine the level of clipping. It is evident that the more severe the clipping is, the more likely it is that the output of the clipper will be a square wave but then more gain will be required in the succeeding stages.

A more elegant design when the clipping level is high is to use back-to-back Zener diodes as shown in Figure 4.8. The Zener diode provides a much sharper cut-off than an ordinary diode when it is reverse biased. In the forward direction, it is an ordinary diode. With this arrangement, only one of the two Zener diodes is operating as a Zener diode while the other one is conducting current in the forward direction. They switch roles when the applied voltage changes polarity.

4.3.3.3 Bandpass Filter. It is reasonable to assume that the output of the clipper is a square wave. The frequency of the square wave varies but it is centered at the carrier frequency. Fourier analysis shows that the most significant harmonic in the square wave is the third. So with the sub-carrier at 200 kHz, the third harmonic will be at 600 kHz. A simple tank circuit tuned to the carrier frequency is all that is necessary to filter out the harmonics. But, as before, the choice of the Q factor of the tuned circuit must be made to ensure that all the significant sidebands are within the passband of the filter and are subjected to minimum attenuation. The factors that affect the choice of the bandwidth were discussed in Section 4.3.3.1.

4.3.3.4 Discriminator. The purpose of the discriminator is to convert the variation of frequency to a variation of amplitude. A frequency-to-amplitude convertor followed by an envelope detector is used to recover the message contained in the modulating signal. The transfer characteristics of two circuits that could be used for the frequency-to-amplitude conversion are shown in Figure 4.9.

The simplest circuit with the characteristics shown in Figure 4.9(a) is a simple RC high-pass circuit with its corner frequency much higher than the carrier frequency of the FM signal. Figure 4.9(b) shows the response of a simple RC lowpass circuit with its corner frequency chosen to be much lower than the carrier frequency. Both

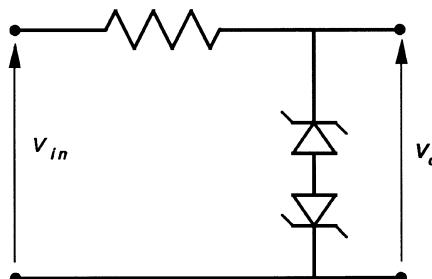


Figure 4.8. The use of two Zener diodes connected back-to-back is an improvement on the symmetrical clipper shown in Figure 4.7.

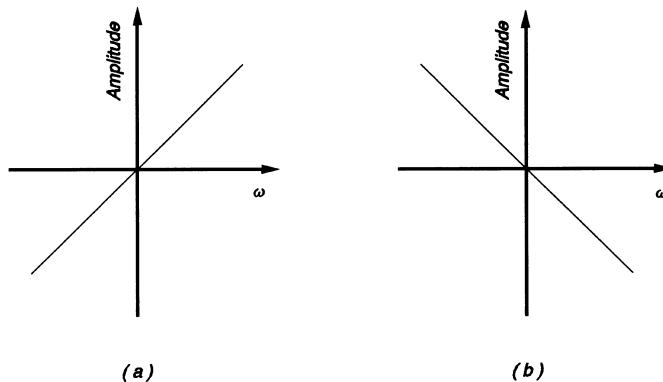


Figure 4.9. The required characteristics of a frequency-to-amplitude converter: (a) is a high-pass circuit and (b) is the low-pass version.

circuits should, in principle, convert the FM signal into an AM signal. However, in practice the sub-carrier frequency of the system shown in Figure 4.5 will be about 200 kHz, while the variation will be limited to a maximum of ± 150 Hz. Since both of these circuits have a slope of 6 dB/octave, the variation in amplitude of these circuits will be exceedingly low. The use of a high gain amplifier following such a circuit will lead to increased noise. A better approach is to find a circuit which has a much greater amplitude-to-frequency slope.

A tuned RLC circuit has a rapid change of amplitude with frequency on both sides of the resonance frequency, especially when the Q factor of the circuit is high. The circuit and its response are shown in Figure 4.10. It is evident that this circuit will be

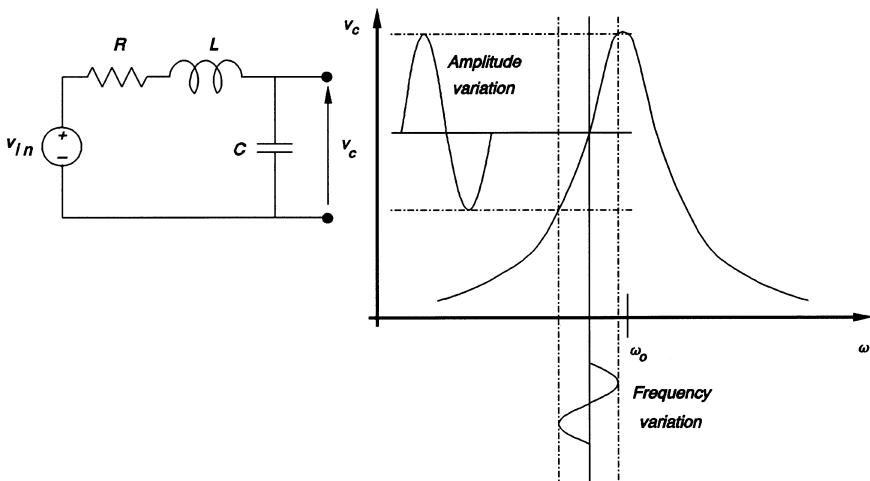


Figure 4.10. A single LC tuned circuit used as a frequency-to-amplitude converter. In this case, the skirt with the positive slope (high-pass) has been used. The skirt with the negative slope (low-pass) would be equally effective. Note that because the nonlinearity of both skirts is predominantly a second-order function, the distortion products of the circuit will be mainly even harmonics.

prone to produce distortion, especially of even harmonics (the non-linearity is unsymmetrical).

A variation on the circuit in Figure 4.10 is shown in Figure 4.11. The two tuned circuits made up of L_1-C_1 and L_2-C_2 are tuned to two different frequencies equally spaced from the carrier frequency. The individual and composite characteristics are shown in Figure 4.11. It can be seen that this circuit has a greater dynamic range than the previous circuit and, when properly tuned, should have no even harmonic distortion products in the output.

4.3.3.5 Envelope Detector. The basic envelope detector was discussed in Section 3.4.6. The importance of the choice of the time-constant in order to avoid distortion of the envelope was discussed at some length. The purpose of this envelope detector is to correct the long-term frequency deviation of the main oscillator while allowing the short-term frequency deviation caused by the modulating signal. Its time-constant is therefore chosen on the basis that the lowest frequency present in the modulating signal, $\omega_{s,\min}$, will not cause the frequency correction system to go into operation. This condition is satisfied when the time-constant of the detector, τ_{det} , is chosen so that

$$\omega_{s,\min}\tau_{\text{det}} \gg 1. \quad (4.3.3)$$

4.4 THE ARMSTRONG SYSTEM

The use of automatic frequency control in the generation of FM signals results in a practical scheme but one which depends on feedback. For proper operation of the system all of the circuits must be made to track each other. Furthermore, the voltage-controlled capacitor used in the automatic frequency control circuit has a non-linearity as well as a temperature characteristic which must be compensated for. Evidently, a different scheme with less complexity is needed. The Armstrong system, named after its inventor, is one such scheme. The theoretical basis of the Armstrong system is given by Equation (4.2.7). It is clear that, in commercial FM systems, the frequency deviation must be small compared to the carrier frequency, that is:

$$\omega_c(t) \gg k \frac{\omega_c}{\omega_s} |\sin \omega_s t| = m_f |\sin \omega_s t| \quad (4.4.1)$$

The following approximations can be applied to Equation (4.2.11):

$$\cos(m_f \sin \omega_s t) \approx 1 \quad (4.4.2)$$

$$\sin(m_f \sin \omega_s t) \approx m_f \sin \omega_s t \quad (4.4.3)$$

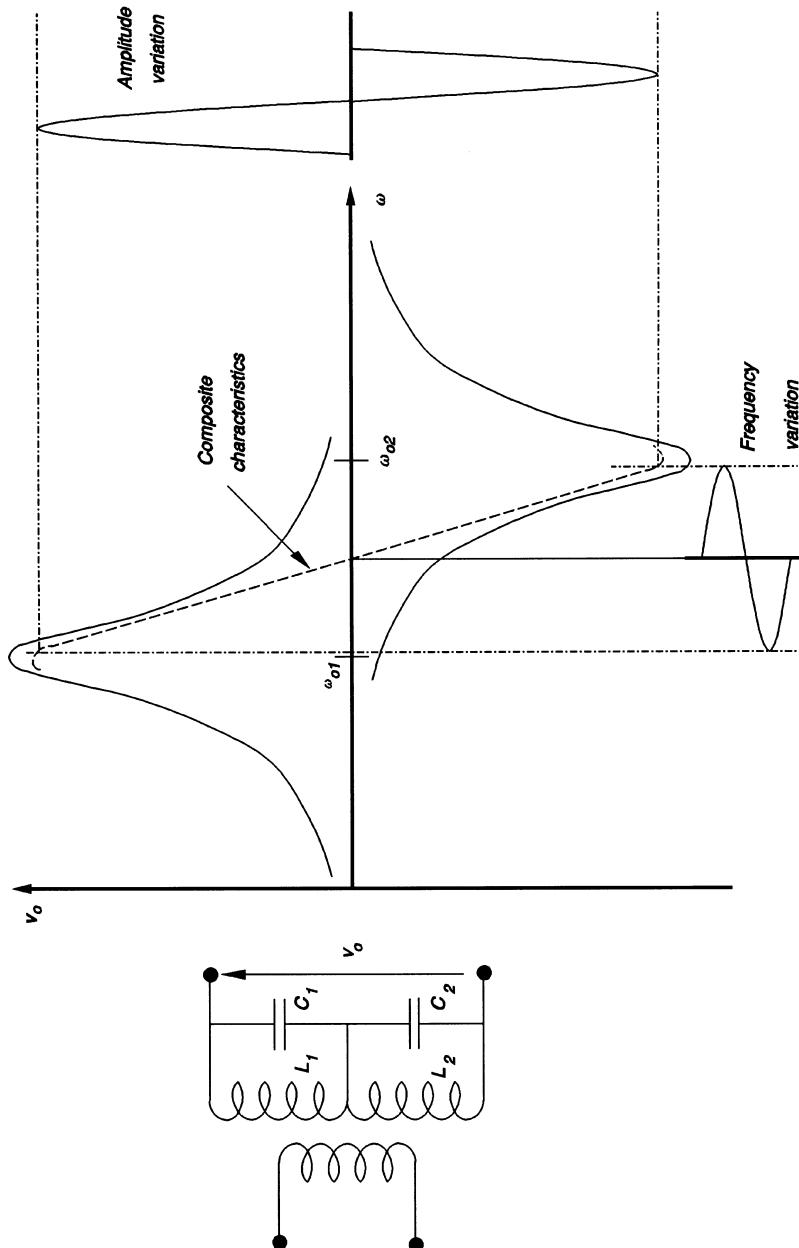


Figure 4.11. The use of a double-tuned circuit provides increased dynamic range and improves the linearity by balancing the non-linearity of one LC circuit with the other. Note that the “push-pull” arrangement shown here will generate predominantly odd harmonics.

to give

$$v_{\text{fm}}(t) = A[\sin \omega_c t + (m_f \sin \omega_s t) \cos \omega_c t] \quad (4.4.4)$$

$$v_{\text{fm}}(t) = A\{\sin \omega_c t + \frac{1}{2}m_f[\sin(\omega_c - \omega_s)t + \sin(\omega_c + \omega_s)t]\}. \quad (4.4.5)$$

It should be noted that the above approximation eliminates all the sidebands except the two closest to the carrier. The modulation is then said to be a *narrow band frequency modulation (NBFM)*. The advantage it offers is a simple geometric representation of the FM signal as a phasor.

The term $[(m_f \sin \omega_s t) \cos \omega_c t]$ is a *double-sideband-suppressed carrier (DSB-SC)* signal. The addition of the carrier term ($\sin \omega_s t$) would normally produce an AM signal but, in this case, the carrier is 90° out of phase with the DSB-SC signal. The difference between AM and FM is shown in the phasor diagrams in Figure 4.12.

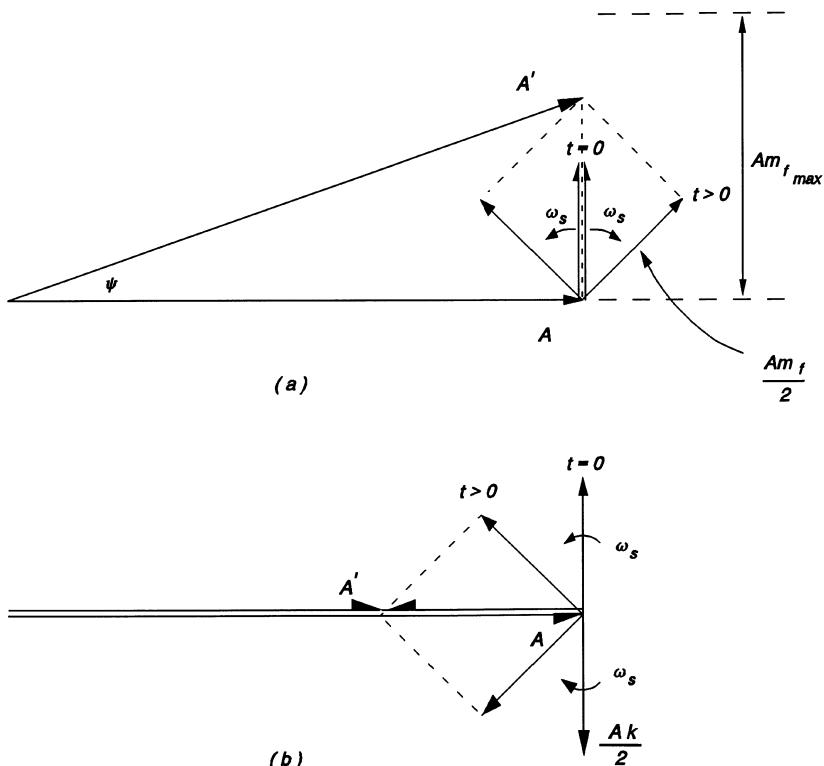


Figure 4.12. (a) A phasor diagram for an NBFM showing the two phasors of amplitude $\frac{1}{2}Am_f$ rotating in opposite directions at angular velocity ω_s in quadrature to the phasor A at $t = 0$. (b) A phasor diagram for an AM signal showing two counter-rotating phasors of amplitude $\frac{1}{2}kA$ and angular velocity ω_s . Note that the amplitude of the resultant phasor varies from $A(1 - k)$ to $A(1 + k)$.

Assume a coordinate system in which the phasor $A \sin \omega_c t$ rotates anticlockwise at an angular velocity ω_c . Suppose that the rotating phasor is used as a new reference system so that it is represented by a horizontal phasor of value A . In terms of this reference system, the term $(\frac{1}{2}Am_f) \sin(\omega_c + \omega_s)t$ is represented by a phasor of value $\frac{1}{2}Am_f$ rotating in an anticlockwise direction with an angular velocity of ω_s . Similarly, the term $(\frac{1}{2}Am_f) \sin(\omega_c - \omega_s)t$ is represented by a phasor of equal value rotating in a clockwise direction with an angular velocity of ω_s . At time $t = 0$, the component of each rotating phasor in the horizontal direction is zero. Therefore the two phasors must be parallel to each other but at right angles to the horizontal phasor A . This situation is depicted in Figure 4.12(a). The resultant is a phasor which apparently has a maximum value

$$A\sqrt{(1 + m_f^2)} \quad (4.4.6)$$

and a maximum angular displacement of

$$\psi = \tan^{-1}(m_f). \quad (4.4.7)$$

Since m_f is small, ψ is small and

$$\psi \approx m_f. \quad (4.4.8)$$

The apparent variation of the amplitude of the FM signal is due to the fact that this analysis has not taken into account the many sidebands present in FM.

For comparative purposes the AM system is treated similarly. An AM signal is represented by Equation (2.2.7):

$$v_{am}(t) = A \sin \omega_c t + \frac{kA}{2} \cos(\omega_c - \omega_s)t - \frac{kA}{2} \cos(\omega_c + \omega_s)t. \quad (4.4.9)$$

Again the carrier $A \sin \omega_c t$ is represented by a horizontal phasor of value A and the two sidebands by two counter-rotating phasors of value $kA/2$ rotating at ω_s . At time $t = 0$, the components of both of the sideband in the horizontal direction are zero since they are cosine terms and A is a sine term. The phasors representing the two sidebands must be at right angles to the A phasor as shown in Figure 4.12(b). Note that one is positive and the other negative. At some other time $t > 0$, the phasor representing the combination of the sidebands subtracts from or adds to the phasor A . The amplitude of the resultant therefore varies from $A(1 - k)$ to $A(1 + k)$ but at all times it is in the horizontal position.

The conclusion is that in a narrow-band ($\Delta\omega_c \ll \omega_c$) frequency modulated system, the phase difference between the carrier and the DSB-SC signal is 90° . This is the basis of the Armstrong system. A block diagram of the FM transmitter is shown in Figure 4.13.

4.4.1 Practical Realization

A crystal-controlled oscillator generates the carrier frequency $A \cos \omega_c t$. This signal is fed to two blocks. The first is a 90° phase shift circuit which, as the name suggests,

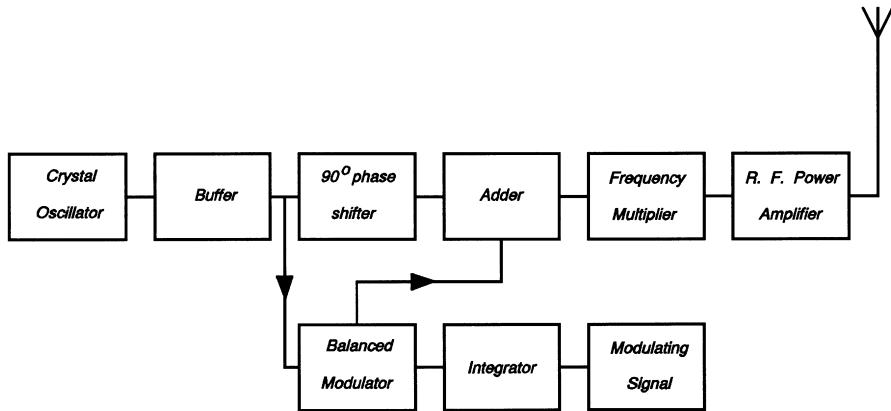


Figure 4.13. Block diagram of the basic Armstrong system FM transmitter. This is an example of an indirect generation of FM signal.

shifts the phase by 90° and therefore its output is $A \sin \omega_c t$. The second block is a *balanced modulator*. The other input to the balanced modulator is, of course, the modulating signal. But according to Equation (4.2.8), $m_f = \Delta\omega_c/\omega_s$. The modulating (audio) signal contains a band of frequencies from approximately 20 Hz to 15 kHz and therefore as ω_s increases, $\Delta\omega$ must increase or the modulation index m_f must decrease. However, in FM the frequency of the carrier changes proportionately to the *amplitude* of the modulating signal, not to the *frequency* of the modulating signal. In order to satisfy this condition, the amplitude of the modulating signal has to be reduced by one-half whenever the frequency doubles. This amounts to putting the signal through a low-pass filter with a slope of 6 dB/octave. An integrator with its corner frequency below the lowest modulating frequency is required. The amplitude of the modulating signal frequency band is thereby “predistorted” by the integrator to conform to the above condition. The need for the integrator is evident from Equation (4.2.7).

If the modulating signal is $B \sin \omega_s t$ the output of the integrator will be $(B/\omega_s) \cos \omega_s t$. This signal is fed to the balanced modulator which produces as an output the product of its two inputs, namely, $(k'AB/\omega_s) \sin \omega_c t \cos \omega_s t$, where k' is a constant. This signal and the output of the 90° phase shifter go to the *adder* which gives an output

$$v_o(t) = A \sin \omega_c t + \frac{k'AB}{\omega_s} \sin \omega_c t \cos \omega_s t. \quad (4.4.10)$$

The signal goes through a series of multipliers to bring the frequency up to the operating value. A power amplifier raises the signal to the proper level for radiation by the antenna.

In a practical FM transmitter, the carrier is generated at a lower frequency and the modulation process carried out. The frequency is then multiplied up to the operating value. If it is assumed that the sub-carrier frequency is again 200 kHz – a frequency at which a suitably robust crystal can be found – the required multiplication factor to place it in the middle of the commercial FM frequency band is then about 500 ($2^9 = 512$). The maximum frequency deviation for a sub-carrier at 200 kHz was calculated in Section 4.3.4 to be 150 Hz.

The ratio of the sub-carrier frequency to the maximum frequency deviation may not be the most convenient in a practical system. In order to produce a high-fidelity signal, that is, maintain a high level of linearity, it is an advantage to make the carrier frequency independent of the frequency deviation. This can be done by splitting the multiplication operation in two and inserting a mixer between them so that the carrier frequency can be changed without affecting the frequency deviation. To help explain the process, an example is given in Figure 4.14.

Suppose the sub-carrier frequency is 200 kHz and the maximum deviation is 20 Hz. To get the sub-carrier frequency up to the normal operating frequency of about 100 MHz, a cascade of frequency multipliers equal to 2^9 (512) will be required. Multiplication of the carrier also multiplies the frequency deviation so that the corresponding frequency deviation will be $(512 \times 10 \text{ Hz}) = 10,240 \text{ Hz}$. This is much lower than the limit allowed, which is 75 kHz. On the other hand, in order to convert the 20 Hz frequency deviation into the required 75 kHz, the multiplication factor is 3750 (approximately $3 \times 2^{10} = 3072$). However, multiplying the sub-carrier by this factor will give a carrier frequency of 614.4 MHz, which is outside the commercial FM radio band. To correct this, the multiplication process is split into two stages and a mixer is inserted between them. The first stage of multiplication may be 96 (3×2^5). The carrier frequency is then 19.2 MHz and the corresponding frequency deviation is 1.92 kHz. The signal is now mixed with a “local oscillator” whose output frequency is 16.2 MHz to produce a carrier frequency of 3.0 MHz but the frequency deviation remains unchanged at 1.92 kHz. The second stage of multiplication ($32 = 2^5$) produces a carrier frequency of 96.0 MHz and a frequency deviation of 61.4 kHz. This is not quite at the allowed limit but close enough for practical purposes. The “local oscillator” signal may be derived from the 200 kHz crystal-controlled oscillator by using a suitable frequency multiplier ($81 = 3^4$).

4.4.2 Component Circuit Design

In the block diagram of the Armstrong system given in Figure 4.14, four new components were added, namely the 90° phase shifter, the balanced modulator, the integrator, and the adder. The design of these circuit components now follow.

4.4.2.1 The 90° Phase Shift Circuit. In the circuit shown in Figure 4.15(a), the resistances R_C and R_E are equal. Since the collector and emitter currents are, for all practical purposes, equal, it follows that the voltages v_C and v_E must be equal in magnitude. But the phase angle between them is 180° . The voltages v_C and v_E are

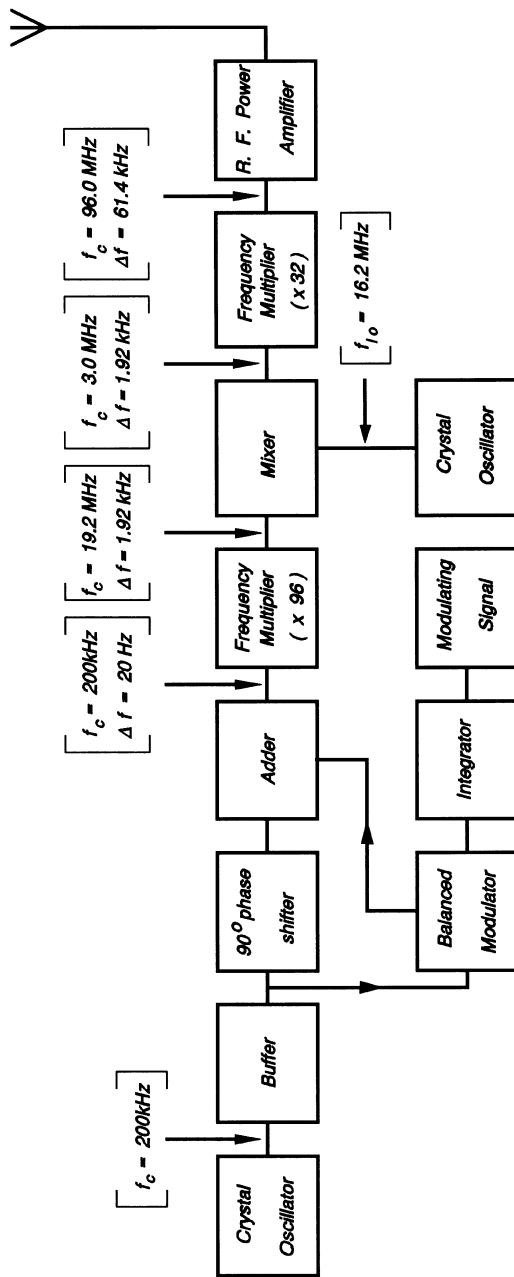


Figure 4.14. Block diagram of a practical Armstrong system FM transmitter.

represented on the phasor diagram shown in Figure 4.15(b). The voltage which appears across R and Z is $(v_C + v_E)$. If the impedance Z is purely reactive (capacitive or inductive), it follows the phasors v_R and v_Z will always be at right angles to each other. It follows then that the point A must lie on a circle whose radius is equal to the magnitude of v_C as shown. The output voltage is then v_o . As frequency goes from zero to infinity, point A will trace out a semicircular locus. The output voltage v_o changes its phase angle from zero to 180° but its magnitude remains constant.

It can be seen that, with the appropriate choice of R and Z at a given frequency, v_o can be made to lead or lag v_{in} by 90° . One major disadvantage of this circuit is that its load must be a very high impedance, preferably open-circuit, otherwise the magnitude of the output voltage v_o will not remain constant for all frequencies. This circuit is a simple example of a class of networks known as *all-pass filters*.

Example 4.4.1 90° phase shift circuit. Design an all-pass circuit with a phase shift of 90° at 200 kHz. The dc supply voltage is 12 V and it may be assumed that the circuit drives a load of very high impedance. The β of the transistor is 100.

Solution. A suitable circuit for the phase shifter is shown in Figure 4.16 with the impedance Z replaced by a single capacitance C .

Since the gain is unity $v_C = v_E$ and the stage will have the maximum dynamic range when the transistor is biased with V_E at $V_{cc}/4$ and V_c at $3V_{cc}/4$, that is, at 3 and 9 V, respectively. The maximum signal voltage that can be applied to the RC series circuit is then 12 V peak-to-peak. The equivalent circuit is then as shown in Figure 4.17.

Let the peak current in the RC circuit be approximately equal to 1 mA. Then $R = 4.2 \text{ k}\Omega$ and the reactance of the capacitor must also be $4.2 \text{ k}\Omega$ so that at 200 kHz, $C = 190 \text{ pF}$.

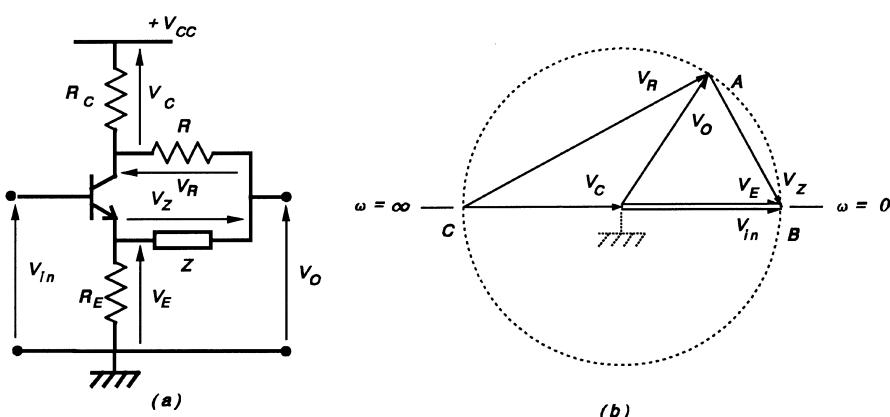


Figure 4.15. (a) An active 90° phase shift circuit. (b) Phasor diagram for the circuit in (a).

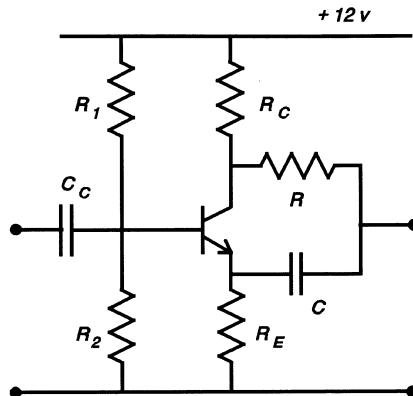


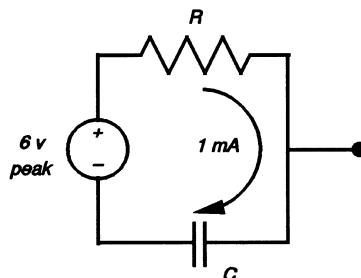
Figure 4.16. Circuit diagram used in Example 4.4.1.

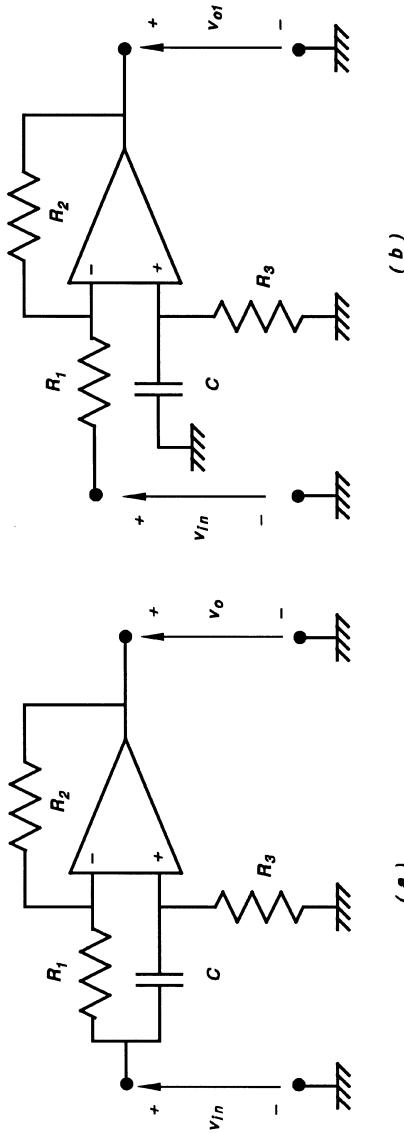
The dc collector current in the transistor must be chosen so that the 1 mA taken by the RC circuit will have negligible effect on the transistor operation. A collector current of 10 mA will be adequate. Under quiescent conditions, the emitter voltage is 3 V. Hence $R_E = 300 \Omega$ and $R_C = 300 \Omega$. Since the β of the transistor is 100, the base current is 100 μ A. Allowing a current 10 times the base current in the resistive chain (i.e. 1 mA) makes $(R_1 + R_2) = 12 \text{ k}\Omega$. But the voltage at the base is 3.7 V and therefore $R_2 = 3.7 \text{ k}\Omega$ and $R_1 = 8.3 \text{ k}\Omega$. The coupling capacitor has to be chosen so that it is essentially a short-circuit at 200 kHz; a 0.1 μ F capacitor is adequate.

4.4.2.2 The 90° Phase Shift Circuit: Operational Amplifier Version. When the frequency of operation is within the bandwidth of an operational amplifier, the circuit shown in Figure 4.18 can be used to realize a 90° phase shift.

The analysis of the circuit is simplified by the application of superposition. The first step is to separate the inverting input from the non-inverting input. With the non-inverting input connected to ground, and v_{in} connected to the inverting input, it can be shown that the output is

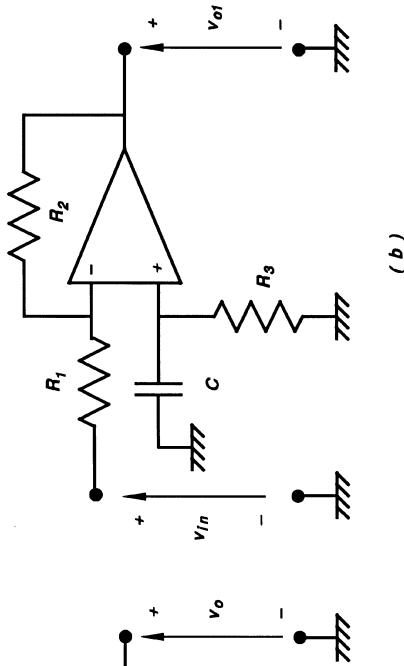
$$v_{01} = -\frac{R_2}{R_1} v_{\text{in}}. \quad (4.4.11)$$

Figure 4.17. An equivalent circuit of the 90° phase shift circuit shown in Figure 4.15(a).



(a)

Figure 4.18. (a) An operational amplifier version of the 90° phase shift circuit. The advantage of this circuit is that its output impedance is low. (b) The circuit in (a) when superposition is applied.



(b)

With the inverting input grounded and v_{in} applied to the non-inverting input, it can be shown that the output is

$$v_{02} = \frac{R_3}{R_3 + \frac{1}{j\omega C}} \left(1 + \frac{R_2}{R_1} \right) v_{\text{in}}. \quad (4.4.12)$$

The output when v_{in} is connected to both inputs is

$$v_0 = v_{01} + v_{02} \quad (4.4.13)$$

and

$$\frac{v_0}{v_{\text{in}}} = -\frac{R_2}{R_1} + \frac{j\omega CR_3}{1 + j\omega CR_3} \left(1 + \frac{R_2}{R_1} \right). \quad (4.4.14)$$

When $R_1 = R_2$ and $\omega CR_3 = 1$,

$$\frac{v_0}{v_{\text{in}}} = j1. \quad (4.4.15)$$

The gain is unity and the output leads the input by 90° . The main advantage of this approach is that the output impedance of the circuit is low enough to drive the balanced modulator directly.

4.4.2.3 Balanced Modulator. A balanced modulator is a circuit which accepts two input signals of different frequencies, a carrier $A \cos \omega_c t$ and a modulating signal $B \cos \omega_s t$, and produces an output proportional to the product of the two inputs

$$v_0(t) = kAB \cos \omega_C t \cos \omega_S t \quad (4.4.16)$$

The balanced modulator is therefore a two-input analog multiplier. It is called a balanced modulator because it produces an AM output with the carrier and modulating signals suppressed. It is also described as a *double-sideband-suppressed carrier* modulator (DSB-SC).

In Section 2.6.3, the four-quadrant analog multiplier was used to obtain an amplitude modulated wave. From Equation (2.6.26) the output current

$$I_2 = m_1 V_{\text{dc}} I_{\text{dc}} + m_2 V_{\text{dc}} I_C \cos \omega_C t + m_3 V_S I_{\text{dc}} \cos \omega_S t + m_4 V_S I_C \cos \omega_C t \cos \omega_S t \quad (4.4.17)$$

where the m_i are constants.

To obtain the required output, the first three terms of Equation (4.4.17) have to be removed. The four-quadrant analog multiplier is realized only in integrated circuit form and one of the basic properties of circuit integration is that the process

produces transistors and other components whose characteristics match very closely. Common-mode signals can be removed from the output by providing two identical circuits in "parallel" and taking the output as the difference between two corresponding points in the two circuits. The basic balanced modulator circuit is shown in Figure 4.19.

It should be noted that the major difference between the circuit of Figure 4.19 and that of Figure 2.33 is that in Figure 4.19, both transistor combinations Q_1-Q_3 and Q_2-Q_4 have a collector load R_L and the output is taken differentially. For specific biasing and detailed information the reader may refer to the specification of the MC 1595 (four-quadrant multiplier produced by Motorola Semiconductor Products, Inc.).

4.4.2.4 The Integrator. The integrator is, in fact, a simple low-pass filter with a 6 dB per octave slope (20 dB/decade) with its corner frequency chosen to be below the lowest frequency present in the modulating signal. The basic integrator circuit is shown in Figure 4.20(a) with the time-constant $\tau = RC$ chosen so that

$$\omega_{S, \min} \tau > 1. \quad (4.4.18)$$

Because the modulating signal contains frequencies spanning almost three decades, the highest frequency present will be attenuated by approximately 60 dB. Moreover the very long time-constant required for this application makes the simple RC low-pass circuit unattractive. A more attractive form of the low-pass filter is shown in Figure 4.20(b).

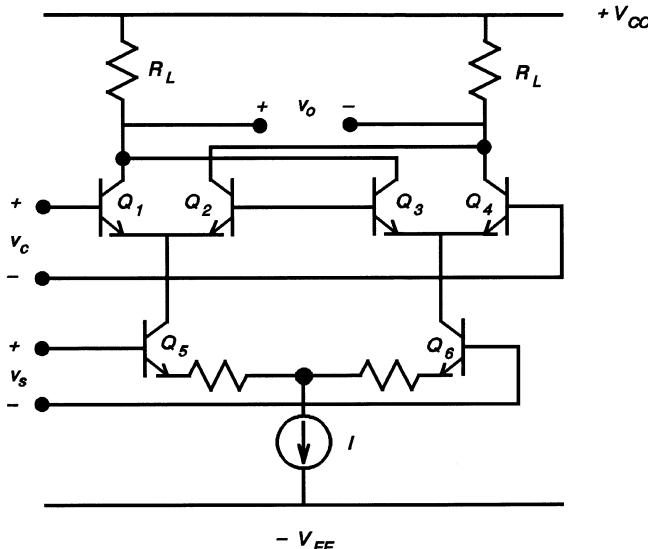


Figure 4.19. The four-quadrant multiplier used as a balanced modulator (DSB-SC). Note the difference between this and Figure 2.33.

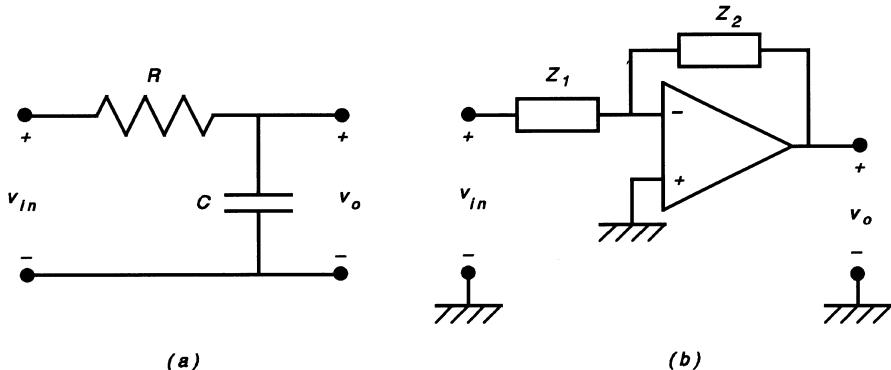


Figure 4.20. (a) A simple “integrator,” that is, a low-pass filter. (b) A practical integrator can be realized by connecting a suitable feedback RC circuit to an operational amplifier.

The output voltage of the operational amplifier is

$$v_0 = -\frac{Z_2}{Z_1} v_{in}. \quad (4.4.19)$$

When $Z_1 = R_1$ and $Z_2 = 1/(j\omega C)$,

$$v_0 = -\frac{1}{j\omega CR_1} v_{in} \quad (4.4.20)$$

It can be shown that in the time domain

$$v_0(t) = -\frac{1}{CR_1} \int_0^t v_{in}(t) dt. \quad (4.4.21)$$

The name “integrator” is therefore appropriate.

For an ideal operational amplifier (open-loop gain = infinity), the gain of the circuit shown in Figure 4.20(b) has to be infinite at dc since Z_2 is an open-circuit at dc. In a practical operational amplifier the open-loop gain, although large (approximately 10,000), is not precisely defined. In order to define the dc gain a resistance R_2 is connected in parallel with C so that the dc gain is equal to $-(R_2/R_1)$. The output voltage is then modified to

$$v_0 = -\frac{R_2}{R_1} \frac{1}{1 + j\omega CR_2} v_{in}. \quad (4.4.22)$$

Example 4.4.2 The Integrator. Design an integrator to satisfy the following specifications:

- (1) Zero dB gain at 1 kHz,
- (2) Corner frequency at 10 Hz,
- (3) What is the gain at 10 Hz and 15 kHz?

Solution. The required response is as shown in Figure 4.21.

The relationship between two frequencies f_1 and f_2 on the basis of decades of frequency is given by

$$f_1 \times 10^n = f_2$$

where n is the number of frequency decades. When $f_1 = 10 \text{ Hz}$ and $f_2 = 1 \text{ kHz}$, $n = 2$.

The gain at 10 Hz = $2 \times 20 \text{ dB} = 40 \text{ dB}$.

The voltage gain of 40 dB corresponds to a voltage ratio of 100. Therefore, $R_2/R_1 = 100$.

Choosing $R_2 = 1 \text{ M}\Omega$ makes $R_1 = 10 \text{ k}\Omega$. From Equation (4.4.22) it follows that at the corner frequency $\omega CR_2 = 1$ and C is therefore equal to $0.016 \mu\text{F}$. Using the results for 1 kHz, 15 kHz corresponds to 1.18 decades of frequency. Therefore, the gain at 15 kHz is

$$-20 \times 1.18 = -23.6 \text{ dB.}$$

4.4.2.5 The Adder. The addition of two signals is most easily accomplished by using a resistive T network as shown in Figure 4.22.

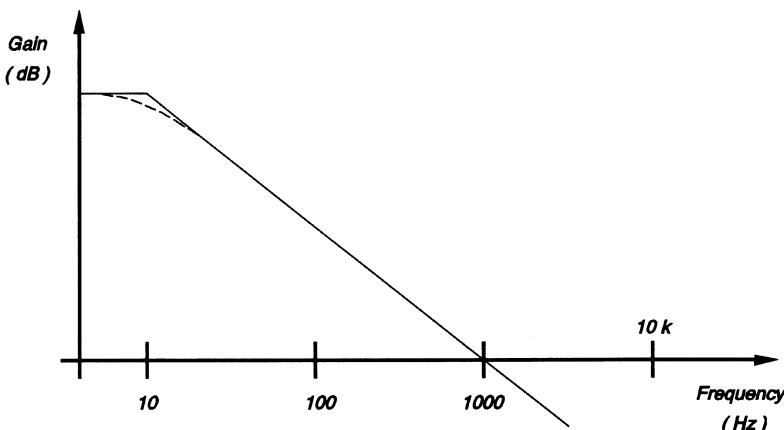


Figure 4.21. The required characteristics of the integrator specified in Example 4.4.2.

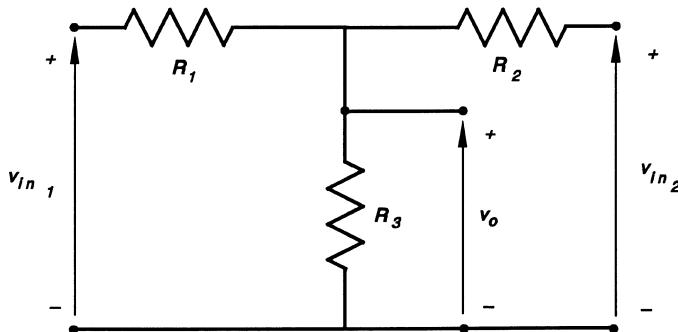


Figure 4.22. A simple resistive adder. The use of such a circuit would lead to severe signal attenuation.

This circuit has two major disadvantages: the output signal is a fraction of the input signal, and there will be a tendency for one signal to “pull” the other. To increase the isolation between the two signals, and hence reduce the possibility of pulling, R_1 and R_2 must be increased relative to R_3 . This aggravates the loss of signal power at the output of the circuit.

An adder circuit that can scale as well as amplify the signals to be added is shown in Figure 4.23.

Equation (4.4.19) can be extended to give

$$v_0 = -\frac{R_f}{R_1} v_1 - \frac{R_f}{R_2} v_2 - \frac{R_f}{R_3} v_3. \quad (4.4.23)$$

The relative values of R_1 , R_2 and R_3 determine the proportion of v_1 , v_2 and v_3 present in v_0 , respectively, while the ratio R_f to R_1 determines the voltage gain for the voltage v_1 and so on.

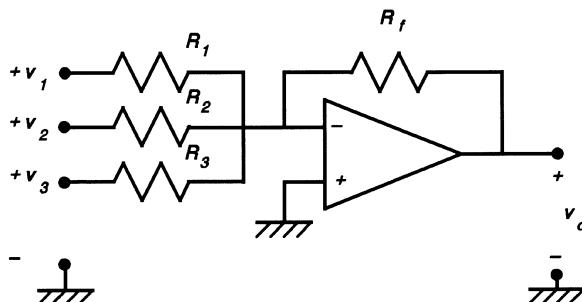


Figure 4.23. An operational amplifier type adder. This circuit enables the circuit designer to choose the gain required as well as the multiplication constants.

4.5 STEREOPHONIC FM TRANSMISSION

In an effort to create a realistic sound presentation from recorded music, two microphones are used, one to record the sound as it is perceived on the right side and the second on the left side. During playback the right- and left-hand side signals have to be fed to the right- and left-hand side loudspeakers, respectively. The listener sitting in front of the speakers can distinguish the sounds of the different instruments as coming from their proper positions when the music was recorded. When this technique is applied to the cinema and television, a car approaching from one side of the screen can be heard in the proper position presented in the picture and the sound appears to move with the picture as the car moves across the screen. In a normal recording process, two separate channels are required for a stereophonic system. So it is expected that a stereophonic FM system will require twice the bandwidth of a monophonic channel. In fact, it requires more bandwidth because the system had to be designed in such a way that listeners who own a monophonic FM receiver can tune in the stereophonic transmission and receive the same performance as they would if the original program had been monophonic.

4.5.1 System Design

The system diagram shown in Figure 4.24 is used to pre-process the modulating signal before it goes to frequency modulate a sub-carrier using the Armstrong technique discussed in Section 4.4.

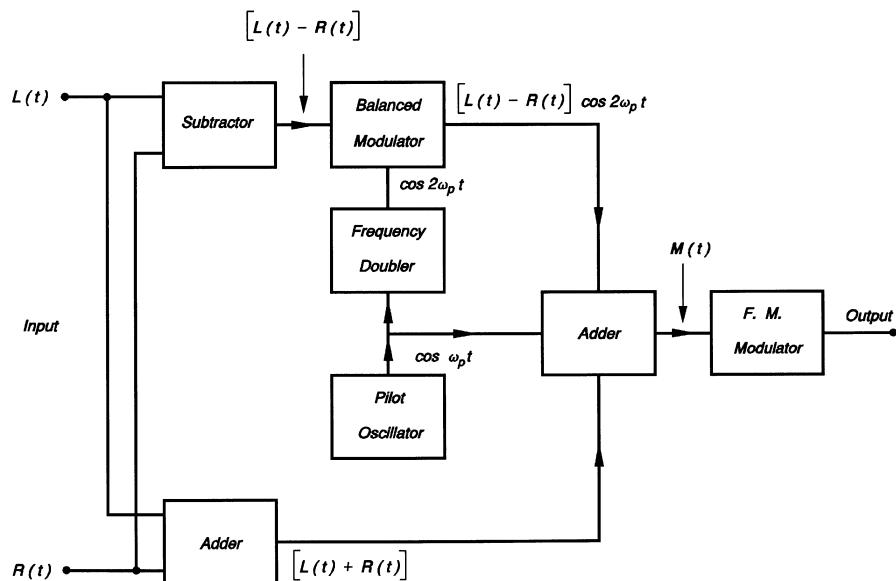


Figure 4.24. A block diagram showing the stages of processing of a stereophonic modulating signal for the Armstrong system FM transmitter.

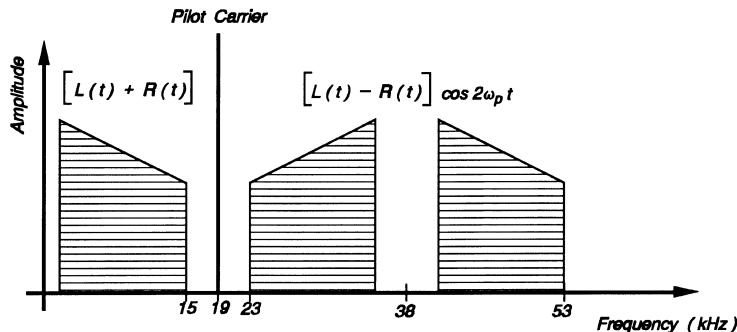


Figure 4.25. The spectrum of the composite modulating signal used in the stereophonic FM transmitter.

The left-hand side signal $L(t)$ and the right signal $R(t)$ are fed to an adder and a subtractor (the signal is first inverted and then added) to produce $[L(t) + R(t)]$ and $[L(t) - R(t)]$, respectively. The $[L(t) + R(t)]$ signal is fed into one of the inputs of a three-input adder as shown in Figure 4.24. The $[L(t) - R(t)]$ signal is one of the two inputs to the balanced modulator. The other input is a 38 kHz signal originally generated at 19 kHz (*pilot frequency*) and then doubled. The output of the balanced modulator is then the product

$$[L(t) - R(t)]A \cos 2\omega_p t \quad (4.5.1)$$

where $2\omega_p$ is the angular frequency of the doubled pilot oscillator frequency and A its amplitude. The other two inputs of the three-input adder are the pilot signal and the output of the balanced modulator. The output of the adder is then

$$M(t) = [L(t) + R(t)] + [L(t) - R(t)]A \cos 2\omega_p t + B \cos \omega_p t \quad (4.5.2)$$

where B is the amplitude of the pilot oscillator. The design of the circuits in all the component blocks shown in Figure 4.24 have been discussed in this and earlier chapters. The baseband spectrum of the stereophonic signal is shown in Figure 4.25.

The processing of the stereophonic signal and its separation into right and left signals will be discussed in Chapter 5.

BIBLIOGRAPHY

1. Taub, H. and Schilling, D. N., *Principles of Communication Systems*, McGraw-Hill, New York, 1971.
2. DeFrance, J. J., *Communications Electronics Circuits*, Holt, Rinehart and Winston, New York, 1966.
3. Stremler, F. G., *Introduction to Communication System*, 2nd Ed., Addison-Wesley, Reading, MA, 1982.

4. Clarke, K. K. and Hess, D. T., *Communication Circuits: Analysis and Design*, Addison-Wesley, Reading, MA, 1978.
5. Norwood, M. H. and Shatz, E., "Voltage Variable Capacitor Tuning: A Review", *Proc. IEEE*, 56(5), 788–798, May 1968.
6. Moline, R. A. and Foxhall, G. F., "Ion-Implanted Hyperabrupt Junction Voltage Variable Capacitors", *IEEE Trans. Electron Devices*, ED-19(2), 267–273, Feb., 1972.

PROBLEMS

- 4.1** The power output of an FM generator operating at $100/2\pi$ MHz is 10 W when no modulation is applied. What would you expect to see on a spectrum analyzer connected across its load? A sinusoidal modulating signal is applied and increased slowly until the amplitude of the carrier goes to zero. What is the average power in the remaining side frequencies?
Calculate:
- (1) the average power in the side frequencies next to the carrier (first-order side frequencies),
 - (2) the average power in the second-order side frequencies,
 - (3) the average power in the third-order side frequencies,
 - (4) the average power in the remaining side frequencies.
- 4.2** A direct FM generator uses a varactor diode for which

$$C = \frac{C_0}{\sqrt{1 + 2V}}$$

where V is the reverse-bias voltage across the p–n junction. Assuming that the oscillator has a parallel LC tuned circuit such as shown in Figure 4.3:

- (1) write an expression for the frequency of operation as a function of V ,
 - (2) write another expression for the frequency of oscillation f_0 when the dc voltage applied to the p–n junction is V_0 .
 - (3) Assuming that v is a small time-varying voltage superimposed on V_0 , calculate the slope (Hz/V) of the frequency–voltage characteristics of the circuit about the oscillating frequency f_0 when $f_0 = 25$ MHz and $V_0 = 5$ V.
- 4.3** A discriminator (frequency-to-voltage converter) is to be designed to operate at $10/2\pi$ MHz. You are provided with a center-tapped transformer whose secondary self-inductance is L and series resistance r for each half. Choose two capacitors of suitable values to tune the secondary for minimum second harmonic distortion, given that $L = 0.2$ mH and $r = 40 \Omega$ (some iteration may be necessary).

- 4.4** In an indirect FM transmitter, the carrier frequency f_c is 250 kHz with the maximum frequency deviation $\Delta f = 50$ Hz. This signal is to be frequency multiplied in two stages (with a mixer between them) to bring the carrier frequency up to 100 MHz. Stage one has a multiplication factor of 64. Choose a suitable local oscillator frequency and the multiplication factor of the second multiplier. Calculate the center frequencies and bandwidths as the signal progresses through the frequency multiplier/mixer. Will this transmitter meet the frequency deviation specifications for commercial FM stations?
- 4.5** In the circuit of Figure 4.18, C and R_3 are interchanged. Under what conditions will the circuit still behave like a 90° phase shifter?
- 4.6** If, in an FM transmitter operating at 100 MHz, a sinusoidal modulating signal of 15 kHz causes the carrier frequency to deviate by 3 kHz, derive an expression for the modulated wave, taking into consideration only the first two pairs of side frequencies. Calculate the amplitudes of the side frequencies relative to the unmodulated carrier.
- 4.7** In the circuit shown in Figure P4.1, a sinusoidal signal

$$v_s = 1.5 \sin \omega_s t$$

is applied to terminals 1 and 1' of the transformer T_1 , where $\omega_s = 5$ krad/s. A second signal

$$v_c = 10 \sin \omega_c t$$

is applied to the terminals A and B, where $\omega_c = 50$ krad/s. Assuming that the diodes have ideal characteristics:

- (1) Sketch the waveform of the signal which appears across terminals 2 and 2' of the transformer T_2 in relationship to v_s and v_c .
- (2) Derive an expression for the output voltage v_o in terms of v_s and v_c .

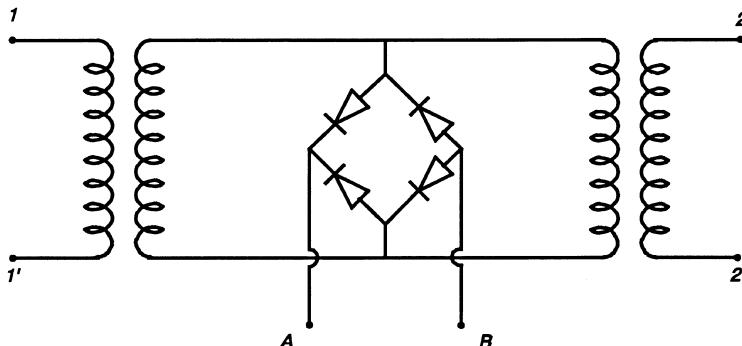


Figure P4.1.

- (3) Calculate the amplitude of the component of v_o at angular frequency $n\omega_c$ where n is an integer.
- (4) Comment on your results from (3).
- (5) Suggest an application for the circuit.

You may assume that for a square wave

$$f(t) = \frac{1}{2} + \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{1}{n} \sin n\omega t. \quad (P4.7)$$

5

THE FREQUENCY MODULATED RADIO RECEIVER

5.1 INTRODUCTION

In amplitude modulation, the frequency of the carrier is kept constant while its amplitude is changed in accordance with the amplitude of the modulating signal. In frequency modulation, the amplitude of the carrier is kept constant and its frequency is changed in accordance with the amplitude of the modulating signal. It is evident that, if a circuit could be found which will convert changes in frequency to changes in amplitude, the techniques used for detecting AM can be used for FM as well.

In Section 4.3.3.4, three frequency-to-amplitude conversion circuits were discussed and their performance in terms of linearity and dynamic range were examined. It therefore follows that the FM receiver must have the same basic form as the AM receiver. The structure of the FM receiver is as shown in Figure 5.1.

The superheterodyne technique is used in FM for the same reasons it is used in AM; it translates all incoming frequencies to a fixed intermediate frequency at which the filtering process can be carried out effectively.

The antenna is responsible for capturing part of the electromagnetic energy propagated by the transmitter. The basic rules of antenna design apply but, because in commercial FM radio the frequency of the electromagnetic energy is between 88 and 108 MHz, it is practical to have antennas whose physical dimensions are within tolerable limits.

The radio-frequency amplifier raises the power level to a point where it can be used in a mixer or frequency changer to change the center frequency to a lower frequency – the intermediate frequency (IF). The mixer in conjunction with the local oscillator translate the incoming radio frequency to an intermediate frequency of 10.7 MHz. There is nothing special about an intermediate frequency of 10.7 MHz except that it is a relatively low frequency at which the required values of L_s and C_s are large enough to reduce the effects of circuit strays. It is at this fixed frequency that filtering to remove the unwanted products of the mixing process and other

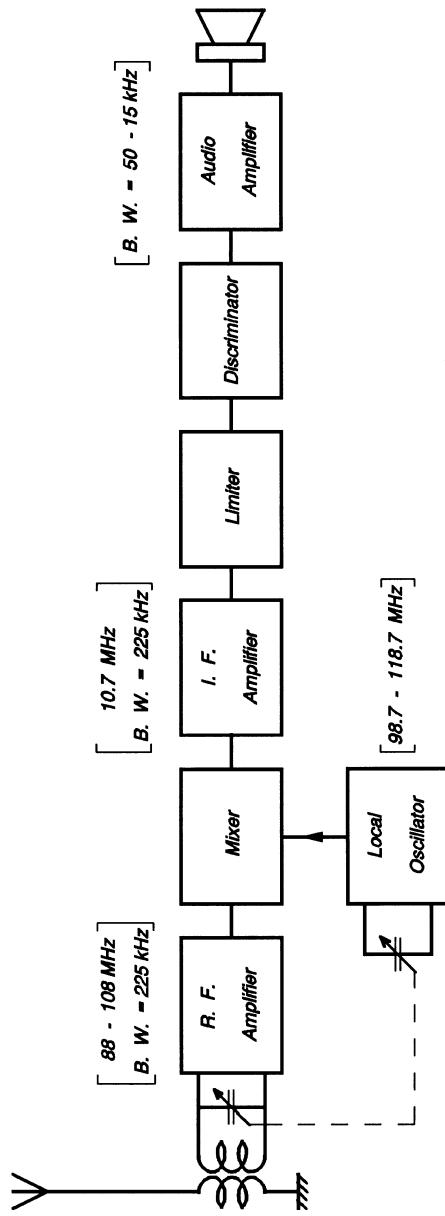


Figure 5.1. The block diagram of the domestic FM receiver showing frequency ranges and bandwidths.

interfering signals and noise takes place. The filtered signal then proceeds to the amplitude limiter. The need for the limiter becomes evident when one recalls that the FM signal is usually converted into an AM signal in the discriminator before it is detected. This means that any variation in the amplitude of the FM signal will be superimposed on the proper signal from the discriminator and hence will cause distortion. The amplitude limiter very severely clips the signal to a constant amplitude and also filters out the unwanted harmonics that are produced by the limiter. The signal then proceeds to the frequency discriminator (frequency-to-amplitude convertor) and onto the envelope detector. The audio-frequency amplifier raises the output of the envelope detector to a level suitable for driving a loudspeaker.

Although the structures of AM and FM receivers are similar, there are very important differences which require different design and construction approaches. These are the following:

1. The higher carrier frequencies (88–108 MHz) used in FM requires small values of both L and C in the tuned circuits used. This means that stray inductances and capacitances will constitute a larger percentage of the designed value and hence have a much greater effect on all tuned circuits. Although measures can be taken to incorporate the effects of fixed circuit strays in the design, there are other changes in element values, due to such factors as temperature and vibration, which can cause sufficient drift to necessitate retuning of the receiver during a program. The local oscillator is most vulnerable to stray elements, especially because it has to operate at a frequency 10.7 MHz above the carrier frequency. To ensure the stability of the oscillator, high circuit Q factors, negative temperature-coefficient capacitors and automatic frequency control (AFC) are used. Some radio-frequency amplifiers for FM front-ends use distributed parameter circuit components such as coaxial and transmission lines.
2. With the intermediate frequency set at 10.7 MHz, the band from which image interference can originate is from 109.4 to 129.4 MHz. This frequency band is reserved for aeronautical radionavigation systems. It follows that one FM station cannot cause image interference for another but the aeronautical radionavigation systems can. As in AM, image interference can be reduced by differentially amplifying the desired signal relative to the image signal. A high Q factor tuned radio-frequency amplifier is used for this purpose.
3. The use of a high Q factor tuned amplifier in the radio-frequency stage requires a very stable local oscillator frequency which will accurately track the incoming radio frequency and produce a minimum variation from the selected intermediate frequency. The local oscillator by itself is not capable of this but, used in conjunction with the AFC and other stabilizing measures, it can perform satisfactorily.
4. The ideal intermediate-frequency filter should have a bandpass characteristic which is flat with infinitely steep sides. The flat top is required to avoid any

frequency dependent amplitude variation, the steep sides to eliminate interference from the unwanted products of the mixing process and from adjacent channels. Two techniques, both involving a number of successive stages of tuned circuits, are used to obtain an approximation to the ideal filter characteristics. In the first, all the tuned circuits have the same resonant frequency. This is known as *synchronous tuning*. In the second the resonant frequencies are placed at different points in the passband. This is known as *stagger tuning*.

The above qualitative as well as quantitative changes from the AM system discussed in Chapter 3 make a separate consideration necessary.

5.2 COMPONENT DESIGN

5.2.1 Antenna

An important point to remember is that an antenna is a reciprocal device, that is, it can be used both for transmitting signals as well as for receiving them. An antenna structure that produces a good ground wave radiation pattern will have a good response to the same ground wave radiation when used in the receiving mode.

Commercial FM receivers commonly use two types of antennas: the vertical whip antenna, most commonly used with automobile radios, and the dipole and folded dipole antennas used with other types of portable and non-portable FM radios.

Assuming that the vertical whip antenna approximates a vertical grounded antenna, a half-wavelength antenna for the middle of the FM band will be about 1.5 m long. Such an antenna can be conveniently mounted on a vehicle. The field patterns of vertical grounded antennas are given in Figure 2.54. It can be seen that when the height h is less than $\lambda/2$, the response is limited to the ground wave. Commercial FM stations are designed to operate within a local area and therefore have antennas which ensure that most of the radiated power goes into the ground wave. An FM receiver with an antenna whose height is equal to or less than $\lambda/2$ will have a good response.

The dipole and its variation, the folded dipole, are commonly used with non-portable FM receivers. They can be used in conjunction with directors and/or reflectors to increase the gain of the antenna. Such an arrangement is called a Yagi–Uda antenna.

Antenna design is beyond the scope of this text. However the interested reader is encouraged to refer to any standard text on antennas.

5.2.2 Radio-Frequency Amplifier

The purpose of the radio-frequency amplifier is to boost the power of the incoming signal relative to all the other signals picked up by the antenna to a level which can be used in the frequency changer. A second function of the radio-frequency amplifier

is to act as a matched load to the antenna so that the antenna signal is not reflected at the interface leading to the loss of efficiency.

The bandwidth of an FM signal in commercial radio was calculated to be approximately 240 kHz in Section 4.3.3.1. With a carrier frequency of about 100 MHz, the required Q factor is therefore about 400. Such a high value of Q factor cannot normally be achieved in a simple tunable radio-frequency amplifier and the practical approach is to use a lower Q circuit and to correct for this in the intermediate-frequency stage that follows. An alternate technique uses a number of stages in cascade separated by buffer amplifiers. For the superheterodyne system to work, the local oscillator frequency must be set equal to the radio frequency plus the intermediate frequency; for the commercial FM band, the standard intermediate frequency is 10.7 MHz. The frequency of the local oscillator must be variable from 98.7 to 118.7 MHz. This is normally not a problem since the ratio of the high frequency to the low frequency is only 1.2 : 1. The more important point is that the center frequency of the radio-frequency amplifier and the frequency of the local oscillator must maintain the difference of 10.7 MHz throughout the FM frequency band. In the case of AM, the frequencies were low and some drift could be tolerated without serious deterioration of the signal. In the FM system, the frequencies are much higher and small percentage changes in one or both radio frequency and local oscillator frequency can cause large changes in the intermediate frequency. To overcome this problem, a system for automatic control of the frequency (AFC) of the local oscillator is used. The basic operation is similar to that described in Section 4.3.2.

5.2.3 Local Oscillator

The local oscillator can take any of the usual oscillator forms with a bipolar transistor as the active element. It must produce enough power to drive the mixer. The values of the inductors and capacitors have to be chosen to minimize the effects of circuit strays. As mentioned earlier, the local oscillator incorporates an AFC circuit for stable tracking with the (tunable) radio-frequency amplifier.

5.2.4 Frequency Changer

The basic frequency changer was discussed in Section 3.4.3. For this application the dual-gate FET mixer has the advantage of low leakage of the local oscillator signal to the antenna via the radio-frequency amplifier. Such a leakage and its subsequent radiation can cause interference with other communication and radionavigation equipment.

5.2.5 Intermediate-Frequency Stage

The intermediate frequency for commercial FM radio is 10.7 MHz. The required bandwidth of the filter is about 240 kHz centered at 10.7 MHz, giving a Q factor of

about 45. It is usual to realize the filter in two or more stagger-tuned stages with suitable buffer amplifiers between them.

5.2.6 Amplitude Limiter

The radio-frequency amplifier, the mixer, and the intermediate-frequency amplifier, in theory, should have a flat amplitude response in their pass bands. In practice, this is not so. The result is that the signal emerging from the intermediate-frequency amplifier has some variation of amplitude with respect to frequency. This is a form of AM and it must be removed if distortion is to be avoided.

The amplitude limiter was discussed in Sections 4.3.3.1 and 4.3.3.2. It is worth noting that sometimes the amplitude limiter is preceded by an automatic gain control. This reduces the severity of the clipping action and hence the required signal power and the spurious harmonics produced.

5.2.7 Frequency Discriminator

The purpose of the frequency discriminator is to convert relatively small changes of frequency (in a very high-frequency signal) to relatively large changes in amplitude with respect to time. The signal can then be demodulated using a simple envelope detector which was discussed in Section 3.4.6. Two basic frequency discriminators were discussed in Section 4.3.3.4 and these serve to illustrate the concepts. In practice, a number of more sophisticated discriminators are used. Some of these will now be discussed.

5.2.7.1 Foster–Seeley Discriminator. The Foster–Seeley discriminator [1,2] is similar to the balanced slope discriminator shown in Figure 4.11. The major difference is that it has two tuned circuits instead of three and both are tuned to the same frequency. This is a major advantage when the receiver is being aligned. A secondary advantage is that it has a larger linear range of operation than the slope discriminator.

The basic circuit of the Foster–Seeley discriminator is shown in Figure 5.2.

Connected to the collector of the transistor, L_1 and C_1 are tuned to resonate at the frequency f_o (the intermediate frequency of the receiver). The inductance L_1 is mutually coupled to a center-tapped inductor L_2 . The center tap is connected by a coupling capacitor C_c to the collector of the transistor. The inductor L_2 and C_2 are tuned to resonate at the frequency f_o . Two identical circuits consisting of a diode in series with a parallel combination of a resistance and a capacitance ($D_3-R_3-C_3$ and $D_4-R_4-C_4$) are connected across L_2 to form a symmetrical circuit. A radio-frequency choke (RFC – a high-valued inductance which can be considered to be an open-circuit at high frequency but a short-circuit at low frequency) connects the center tap to the “neutral” node of the $D_3-R_3-C_3$ and $D_4-R_4-C_4$ circuits.

The circuit can be divided up into two parts by the line connecting $X-X'$. The parts of the circuit to the left of $X-X'$ operate at a high frequency f_o with a relatively small deviation $\pm\Delta f$. The circuit to the right of $X-X'$ are two envelope detectors, as

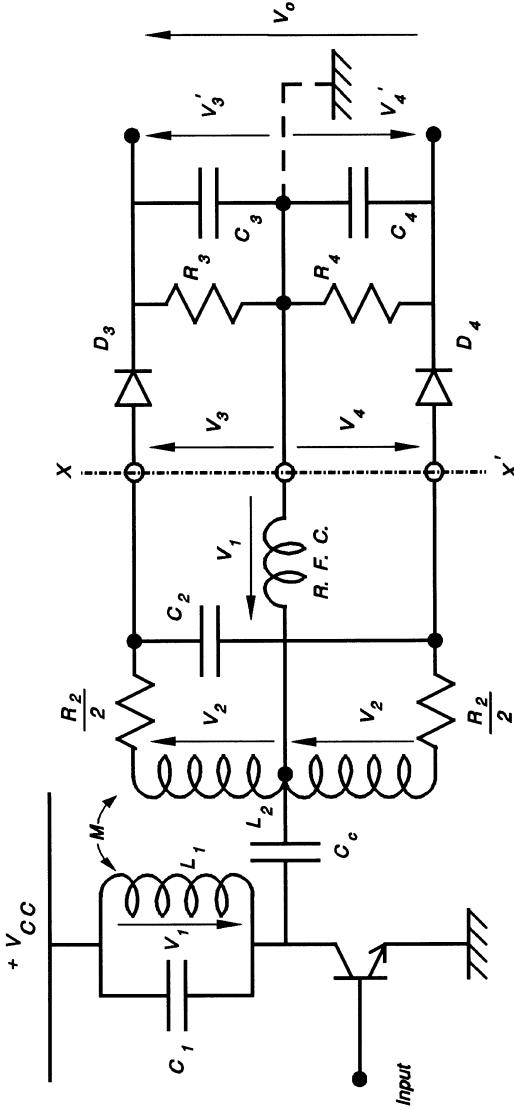


Figure 5.2. The Foster-Seeley discriminator. The line $X-X'$ divides the circuit into high (radio) and low (audio) frequency.

discussed in Section 3.2. The high-frequency input voltages to the envelope detectors are rectified by the diodes D_3 and D_4 and the time-constants R_3-C_3 and R_4-C_4 are chosen to smooth out the half-wave pulses but follow any slow changes in the envelope (amplitude) of the half-wave pulses. These slow changes represent low (audio) frequency.

Before proceeding to the analysis of the circuit, it is in the interest of simplicity, to make three assumptions:

- (1) The impedance of the coupling capacitance C_c is small enough to be considered as a short-circuit, at the frequency of operation.
- (2) The impedance of the RFC is an open-circuit at the high frequency f_o but a short-circuit at the low (audio) frequency.
- (3) The neutral node of the envelope discriminators can be considered to be grounded since the secondary circuit, including the envelope discriminators, is symmetric.

Now, all that needs to be demonstrated is that, when a signal of frequency $f_o \pm \Delta f$ is applied to the circuit, the amplitude of the voltage appearing across the inputs of the envelope detectors will vary proportionally to $\pm \Delta f$.

The tuned circuits L_1-C_1 and L_2-C_2 are both high Q factor circuits but their mutual coupling coefficient, M , is low. This means that the secondary load coupled into the primary circuit is negligible. The primary current is

$$I_1 \approx \frac{V_1}{j\omega L_1}. \quad (5.2.1)$$

The voltage induced in the secondary is

$$2V_2 = \pm j\omega M I_1 \quad (5.2.2)$$

where the \pm sign depends on the relative directions of the primary and secondary windings. Assuming the positive sign and substituting for I_1

$$2V_2 = \frac{MV_1}{L_1}. \quad (5.2.3)$$

Since the secondary circuit is tuned to resonance, the secondary current is

$$I_2 = \frac{MV_1}{R_2 L_1} \quad (5.2.4)$$

where R_2 is the series resistance of the secondary circuit.

The voltage across the capacitor C_2 is

$$2V_2 = \frac{I_2}{j\omega_0 C_2}. \quad (5.2.5)$$

Substituting for I_2 ,

$$2V_2 = \frac{MV_1}{j\omega_0 C_2 R_2 L_1}. \quad (5.2.6)$$

The secondary voltage applied to one envelope discriminator is given by

$$V_2 = \frac{M}{j2\omega_0 C_2 R_2 L_1} V_1. \quad (5.2.7)$$

It is now clear that, *at resonance*, the primary voltage V_1 is at right angles to the secondary voltage V_2 . The phasor diagram of the voltage applied to the inputs of the envelope discriminators is as shown in Figure 5.3(a). This can be modified by reversing the direction of one of the phasors representing V_2 as shown in Figure 5.3(b).

The discriminator input voltage phasors V_3 and V_4 are equal in magnitude and since the outputs of the envelope discriminators are proportional to the magnitude of the applied voltages, when the output is taken differentially, it is zero. This means that the Foster–Seeley discriminator has zero volts output at the *resonant frequency*.

The impedance of the secondary tuned circuit, at any given frequency ω , is

$$Z_2 = R_2 + j\left(\omega L_2 - \frac{1}{\omega C_2}\right). \quad (5.2.8)$$

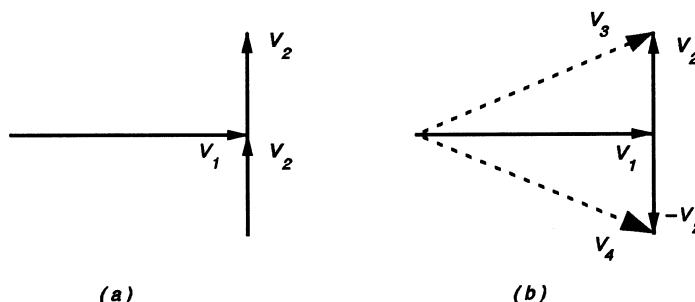


Figure 5.3. (a) The basic phasor diagram of the discriminator. (b) The phasor diagram when the FM carrier is unmodulated.

But at resonance

$$C_2 = \frac{1}{\omega_0^2 L_2}. \quad (5.2.9)$$

Eliminating C_2 from Equation (5.2.8) gives

$$Z_2 = R_2 + j \left(\omega L_2 - \frac{\omega_0^2 L_2}{\omega} \right). \quad (5.2.10)$$

If we now define the Q factor at resonance as

$$Q_0 = \frac{\omega_0 L_2}{R_2} \quad (5.2.11)$$

then Equation (5.2.10) can be written as

$$Z_2 = R_2 + j \omega_0 L_2 \left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \right). \quad (5.2.12)$$

Now consider relatively small changes of frequency about the resonant frequency ω_0 and define “fractional detuning” δ as

$$\delta = \frac{\omega - \omega_0}{\omega_0} \quad (5.2.13)$$

then

$$\frac{\omega}{\omega_0} = 1 + \delta \quad (5.2.14)$$

hence

$$\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} = 1 + \delta - \frac{1}{1 + \delta} = \delta \left(\frac{2 + \delta}{1 + \delta} \right). \quad (5.2.15)$$

For a high Q factor circuit at a frequency near the resonance, the fractional detuning δ is much smaller than 1, therefore

$$\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \approx 2\delta. \quad (5.2.16)$$

Substituting into Equation (5.2.12)

$$Z_2 = R_2(1 + j2Q\delta). \quad (5.2.17)$$

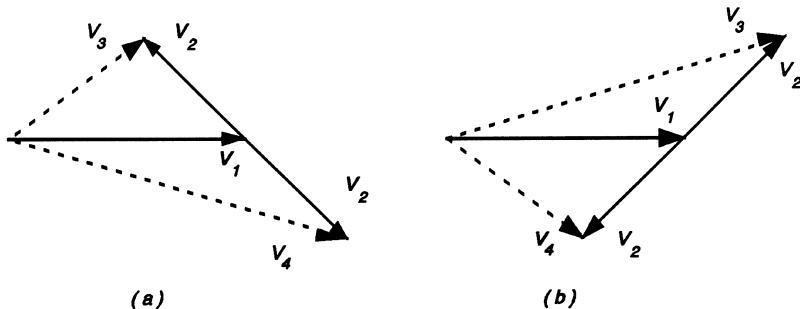


Figure 5.4. (a) The phasor diagram when signal frequency is lower than carrier frequency. (b) The phasor diagram when signal frequency is higher than carrier frequency.

Replacing R_2 in Equation (5.2.7) by Z_2 gives

$$V_2 = \frac{M}{j2\omega_0 C_2 L_1 R_2 (1 + j2Q_0 \delta)} V_1. \quad (5.2.18)$$

It is evident that V_1 and V_2 are no longer at right angles to each other. The angle between them depends on the magnitude and sign of δ . When δ is positive, the angle between V_1 and V_2 is less than 90° , and when δ is negative the angle is greater than 90° or vice versa. The phasor diagrams for positive and negative values of δ are shown in Figure 5.4(a) and (b), respectively.

It is clear from these that the magnitudes of the input voltages V_3 and V_4 are unequal when δ has any value other than zero. The characteristics of the Foster–Seeley discriminator are shown in Figure 5.5.

A variation on the Foster–Seeley discriminator which combines the functions of the amplitude limiter and frequency discriminator is called the *ratio detector* [3]. Its performance however leaves much to be desired.

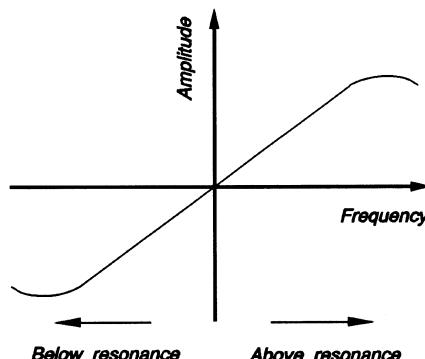


Figure 5.5. The amplitude–frequency characteristics of the Foster–Seeley discriminator.

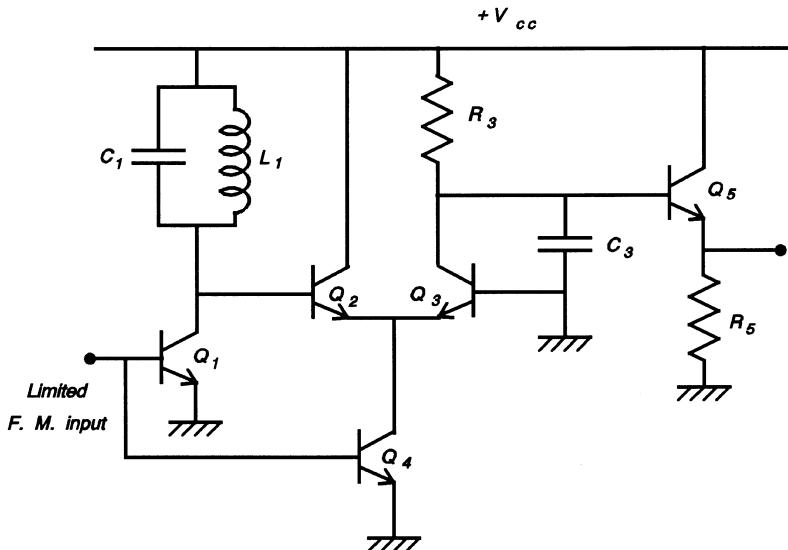


Figure 5.6. The circuit diagram of the quadrature detector with an emitter follower to give a low impedance output.

5.2.7.2 Quadrature Detector. The basic circuit diagram of the quadrature detector is shown in Figure 5.6. All biasing circuit components have been omitted for clarity of its operation.

The circuit consists of a tuned amplifier Q_1 , with a very high Q factor collector load. The input to the circuit is the output from the amplitude limiter which is a frequency modulated square wave (i.e. a square wave of fixed frequency with relatively small deviations in its zero crossings). Due to the high Q factor of the tuned circuit, the output from the amplifier is a sinusoid at the fixed frequency. The same square wave is fed to the base of Q_4 which is a constant current source for the differential pair Q_2-Q_3 . Therefore current flows in the differential pair only when Q_4 is switched on. The sinusoid applied to Q_2 determines what proportion of the constant current in Q_4 flows through Q_2 as opposed to Q_3 . It can be seen from Figure 5.7(a) that, when the input signal is unmodulated, that is, when the phase difference between the sinusoid and the square wave is fixed, the circuit can be adjusted so that Q_2 and Q_3 conduct equal currents, giving a constant voltage across C_3 . The time-constant $R_3 C_3$ hold the base of Q_5 at a dc value and the output remains constant.

When the input to the circuit is modulated, the sinusoid driving Q_2 is no longer coincident with the square wave and Q_3 now conducts for a period proportional to the "phase shift" between the two signals. This can be seen in Figure 5.7(b). The voltage across C_3 is a slowly varying direct current and the output is in fact the audio frequency which was used to frequency-modulate the radio-frequency.

It should be noted that:

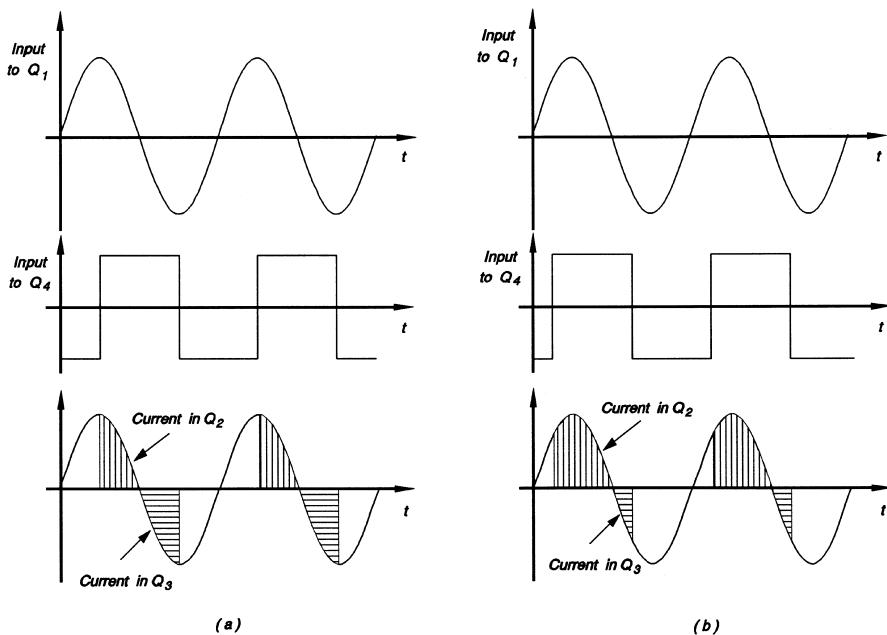


Figure 5.7. (a) The phase difference between the sinusoid and the square wave when the carrier is unmodulated is such that currents of equal magnitude flow through Q_2 and Q_3 giving a dc output. (b) When the carrier is modulated the relative phase between the sinusoid and the square shifts and the currents in Q_2 and Q_3 are no longer equal; the dc changes its value – the changing dc is the audio-frequency signal.

- (1) the amplifier Q_5 provides a low output impedance for the circuit,
- (2) the structure of the circuit is suitable for realization in integrated circuit form,
- (3) the time-constant R_3C_3 is chosen to “follow” the changes in the amplitude of the audio-frequency signal,
- (4) when the two signals are 90° out of phase (in quadrature), Q_2 and Q_3 conduct equal currents and this condition may be used as a datum.

5.2.7.3 Phase-Locked Loop FM Detector. The phase-locked loop [4] FM detector is the most complex of FM detectors but it has the advantage that it can be realized in integrated circuit form where complexity is not necessarily a disadvantage. The basic system is as shown in Figure 5.8.

It consists of a phase detector that generates an output signal which is proportional to the difference between the phases of the two input signals (“error” signal). The output signal is amplified and low-pass filtered and used to control a *voltage-controlled oscillator* (VCO) which usually operates at a higher frequency than the input signal. The output of the oscillator is divided by a suitable factor N to bring it to the same frequency as the input signal. This is the second input to the phase detector.

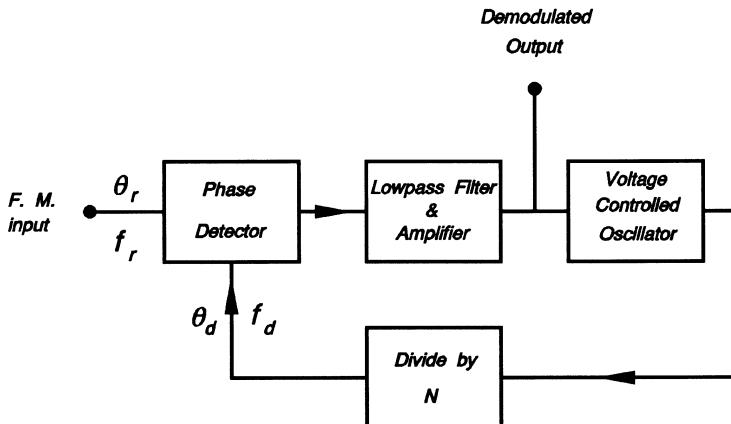


Figure 5.8. The block diagram of the phase-locked loop FM detector. Note that the relatively slow-varying dc required to keep the loop in lock is the audio-frequency signal.

The error signal fed to the VCO causes it to change frequency so that f_d moves closer to f_r . When the two frequencies are close to each other, the system *locks*, that is, the two frequencies become equal and their phase difference is zero. The control voltage from the low-pass filter is then dc. When the incoming signal changes its frequency, and hence phase, the control voltage will change its value to keep the system in lock. The excursions of the control voltage is, in fact, the required demodulated output.

The phase-locked loop FM detector has the advantage of having no *LC* tuned circuits. In its integrated-circuit form, it requires a number of external resistors and capacitors for its proper operation. This information is usually provided by the manufacturer.

5.3 STEREOPHONIC FREQUENCY MODULATED RECEPTION

The baseband frequency spectrum of the stereophonic FM signal was described in Section 4.5.2 and shown in Figure 4.25. The use of the signal to frequency modulate a suitable carrier remains essentially the same as in the case of monophonic transmission except for the fact that the stereophonic spectrum has a larger bandwidth – more than three times larger. At the receiver, the signal is demodulated and the baseband information is recovered. Figure 5.9 shows a scheme for separating the left, $L(t)$, and right, $R(t)$, information and passing them on to their respective loudspeakers.

The first bandpass filter has a passband 23–53 kHz and it is used to separate the double-sideband-suppressed carrier (DSB-SC) signal which contains the left-minus-right signal, $[L(t) - R(t)]$. The second bandpass filter is a narrow-band filter centered on 19 kHz and it separates the pilot carrier. The third filter is a low-pass filter with a

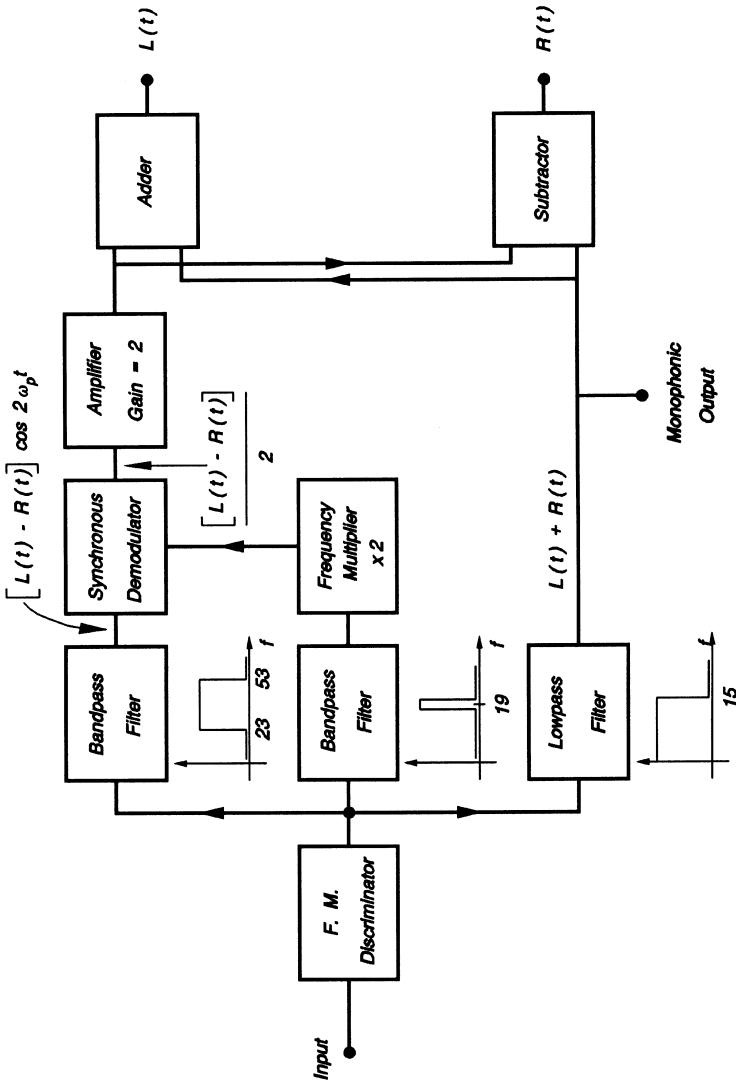


Figure 5.9. The block diagram of the stereophonic FM receiver showing the ideal characteristics of the various filters required and the access point for listeners with monophonic receivers.

cut-off frequency of 15 kHz and it separates the left-plus-right, $[L(t) + R(t)]$ (i.e. the monophonic) signal from the other two.

The DSB-SC signal cannot be demodulated unless the carrier is reinstated. Because the DSB-SC signal was obtained, in the transmitter, from the pilot oscillator followed by a frequency doubler, the process is repeated in the receiver to obtain the carrier at 38 kHz. The carrier is then fed into the synchronous demodulator to produce a baseband signal containing the $[L(t) - R(t)]$ information and, after an amplification factor of two, it is combined with the $[L(t) + R(t)]$ signal in the adder and subtractor, respectively, to produce $L(t)$ and $R(t)$.

5.3.1 Synchronous Demodulation

Synchronous demodulation of a DSB-SC signal can be achieved simply by multiplying the DSB-SC signal $[L(t) - R(t)] \cos 2\omega_p t$ by the synchronized signal $\cos 2\omega_p t$. The result is

$$v(t) = [L(t) - R(t)] \cos^2 2\omega_p t \quad (5.3.1)$$

$$v(t) = \frac{1}{2}[L(t) - R(t)][1 + \cos 4\omega_p t]. \quad (5.3.2)$$

The required baseband signal $\frac{1}{2}[L(t) - R(t)]$ is easily separated from the signal at $4\omega_p$.

5.3.2 Stereophonic Receiver Circuit

Filter design is beyond the scope of this book. A list of references on filter design can be found at the end of Chapter 3. The design of frequency multipliers was discussed in Section 2.5. The design of the synchronous demodulator (two-input analog multiplier) was discussed in Section 4.4.3.3. The adder/subtractor was the subject of Section 4.4.3.5.

REFERENCES

1. Foster, D. E. and Seeley, S. W., "Automatic Tuning Simplified Circuits and Design Practice", *Proc. IRE.*, 25, 289, 1937.
2. Seeley, S. W., *Radio Electronics*, McGraw-Hill, New York, 1956.
3. Seeley, S. W. and Avins, J., "The Ratio Detector", *RCA Review*, 8, 201, 1947.
4. Best, R. E., *Phase-Locked Loops*, McGraw-Hill, New York, 1984.
5. DeFrance, J. J., *Communication Electronics Circuits*, 2nd Ed., Rinehart Press, San Francisco, 1972.
6. Stark, H. and Tuteur, F. B., *Modern Electrical Communications*, Prentice-Hall, Englewood Cliffs, NJ, 1979.

PROBLEMS

- 5.1** A bipolar transistor operating at a frequency of 100 MHz has the following Y parameters: $y_{11} = 9.84 + j9.1$, $y_{12} = 0.01 - j0.68$, $y_{21} = 60.7 - j95.8$ and $y_{22} = 0.60 + j1.8$, measured in mSi in the common emitter configuration with 0.6 V peak-to-peak appearing across the base emitter junction. Using this device, design an oscillator to supply maximum power to a suitable conductive load. What modifications would you make to the oscillator to obtain a local oscillator for an FM receiver operating in the range 88–108 MHz?
- 5.2** An FM receiver is tuned to a transmitter at 100 MHz. The modulation is sinusoidal at a frequency of 15 kHz. The intermediate frequency is 10.7 MHz and the Q factor of the intermediate-frequency amplifier is 100. Estimate the maximum allowable drift in the local oscillator frequency, assuming that the $J_3(m_f)$ terms must lie within the -3 dB bandwidth of the intermediate-frequency amplifier.
- 5.3** Using the characteristics of the FET given in Figure P3.7, design an amplifier tuned to 100 MHz with a voltage gain of 15. The dc supply voltage is 15 V and you may allow 2 mA dc to flow through the drain. The -3 dB bandwidth should be no larger than 250 kHz when the amplifier is connected to a second stage whose input impedance is $250\ \Omega$ resistive.
- 5.4** Give two reasons why a double-sideband-suppressed carrier (DSB-SC) modulation system is not commonly used in communication systems. Name one communication system in which it is used, giving the relevant details.
An incoming signal in a communication receiver is given by

$$v(t) = \frac{1}{2}\beta[\cos(\omega_c - \omega_m)t - \cos(\omega_c + \omega_m)t] \quad (P5.1)$$

where ω_c is the carrier angular frequency and ω_m is the modulating signal angular frequency.

What do you need to demodulate the signal and how would you carry out the demodulation? Support your answer with suitable diagrams and derived formulae. Can you think of a second demodulation technique for the signal?

- 5.5** Describe with the aid of a block diagram the superheterodyne technique for the reception of commercial FM signals. You have an AM/FM receiver and you wish to make changes to the configuration of the receiver by introducing a second mixer to translate the intermediate frequency of the FM system from the standard 10.7 MHz to the standard AM system frequency of 455 kHz. Redesign the block diagram, giving all the necessary details to achieve your wish. What would be the advantages and disadvantages of your new receiver?
- 5.6** The dual-gate FET (whose characteristics are given in Figure P5.1) is to be used in the design of a frequency changer in a commercial FM receiver whose

intermediate frequency is 10.7 MHz. The output of the radio-frequency amplifier is applied to gate 1 and it is 0.4 V peak-to-peak at a frequency of 100 MHz. The signal from the local oscillator applied to gate 2 is a square wave whose voltage varies from +4 V to a negative value sufficient to cut off the drain current. Design a suitable mixer with the following specifications:

- (1) direct current power supply, 12 V,
- (2) average drain current, 4 mA,
- (3) -3 dB bandwidth, 240 kHz.

Choose all the other values and in every case justify your choice. Calculate the ratio of the intermediate-frequency power output to the radio-frequency power input.

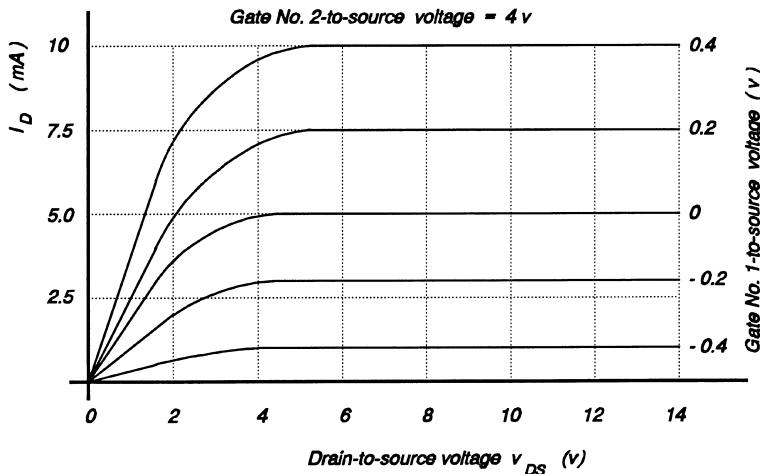


Figure P5-1.

- 5.7 The purpose of the discriminator in an FM receiver is to convert variation in frequency to variation in amplitude. Describe the operation of the Foster-Seeley discriminator and demonstrate its frequency-to-amplitude conversion ability. What measures should be taken prior to the signal entering the discriminator to reduce distortion and why?

6

THE TELEVISION TRANSMITTER

6.1 INTRODUCTION

The transmission of video images depends on a scanning device that can break up the image into a grid and measure the brightness of each element of the grid. This information can be sent serially or in parallel to a distant point and used to reproduce the image. It is evident that the smaller the size of the grid element, the better the definition of the image.

One of the simplest devices which can measure the brightness of light is the phototube. It consists of a cathode which is coated with a material which gives off electrons when light is shone on it and an anode which can collect the emitted electrons when a suitable voltage is applied to it. The cathode and anode are enclosed in an evacuated glass envelope. The number of electrons emitted by the cathode is proportional to the intensity of the light impinging on it. Assuming complete collection of the electrons, the current in the resistor R shown in Figure 6.1 will be proportional to the light intensity and so will the voltage across R .

A primitive video signal can be generated by using a 3×3 matrix made up of phototubes as shown in Figure 6.2.

For simplicity we assume that the tree is black and its background is white. A suitable lens focuses the image of the tree onto the matrix of phototubes. It is clear that the voltage output from phototubes (1,1), (1,3), (3,1) and (3,3) will be high; all others will be low. The voltages so obtained can be transmitted and used to control the brightness of a corresponding 3×3 matrix of lights at a distant point giving a vague idea of what the tree looks like! The picture detail can be improved by increasing the number of elements in the matrix so that each element corresponds to the smallest area possible. The assumption of a black tree on a white background is no longer necessary since, with increasing detail, different shades of grey can be accommodated.

The information may be sent along individual wires linking the phototube to the light matrix (parallel transmission) but this would be very expensive and impractical

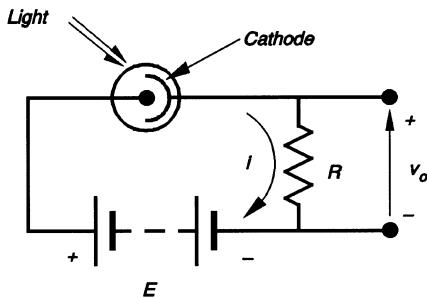


Figure 6.1. The phototube with its added circuitry to convert light intensity to voltage.

for any system other than the simple one described here. A better system would be one in which the voltage from each phototube is scanned in some given order and the voltage and position of each phototube are sent on a single wire to the receiving end for reconstruction (serial transmission). The price to be paid for reducing the number of wires is the increased complexity introduced by the scanner and a system for coding and decoding the voltage and position information at the transmitter and receiver, respectively.

6.2 SYSTEM DESIGN

Figure 6.3 shows the basic components of a television transmitter. A system of lenses focus the image onto a camera tube which collects and codes the information about the brightness and position of each element of the matrix forming the picture by scanning the matrix. A pulse generator supplies pulses to the camera to control the scanning process. The output from the camera goes to a video amplifier for amplification and the addition of extra pulses to be used at the receiver for decoding purposes.

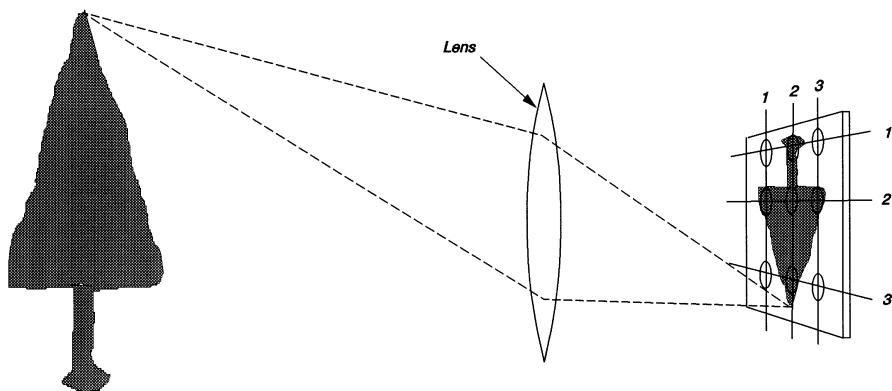


Figure 6.2. Generation of a primitive video signal using a 3×3 matrix of phototubes.

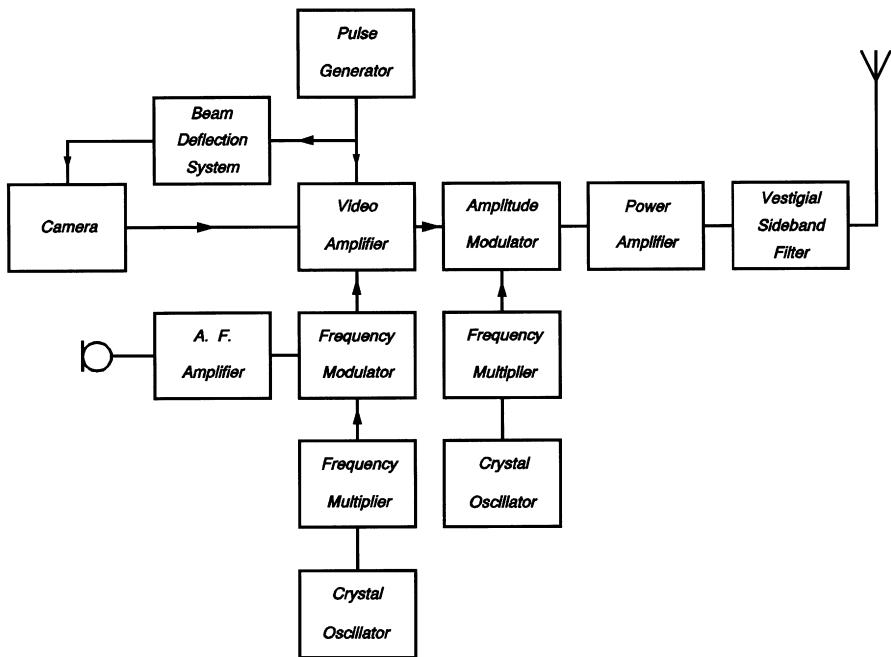


Figure 6.3. A block diagram of the television transmitter.

A microphone picks up the sound associated with the picture and after amplification the signal is fed to the audio terminal of a frequency modulator. The carrier signal supplied to the modulator is a 4.5 MHz signal, generated by a crystal-controlled oscillator at a lower frequency and multiplied by an appropriate factor. The FM signal carrying the audio information is added to the video signal. The output of the video amplifier consisting of the video signal, receiver control pulses, and the frequency modulated signal is fed to the amplitude modulator. The carrier of the amplitude modulator is supplied by a second crystal oscillator and associated multiplier which produce a signal of frequency which lies within the band 54–88 MHz (VHF). The radiofrequency power amplifier boosts the power to the legally determined value and the vestigial sideband filter removes most of the lower sideband signal before it goes to the antenna for radiation.

6.3 COMPONENT DESIGN

6.3.1 Camera Tube

6.3.1.1 Iconoscope. The first practical video camera tube invented by the American scientist Vladimir Zworykin [4,5] is best viewed as a progression from the primitive phototube arrangement discussed earlier. In Zworykin's iconoscope he

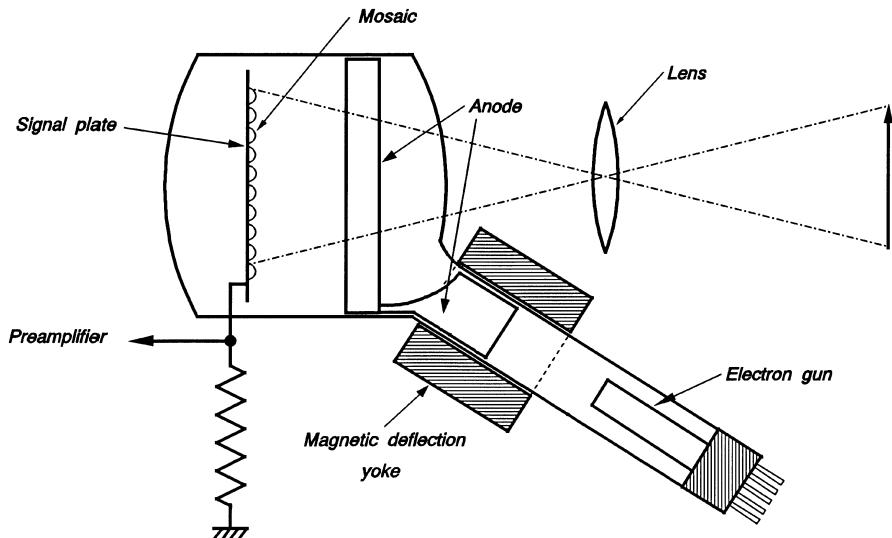


Figure 6.4. A cross-sectional view of the iconoscope. The two parts of the anode are cylinders formed on the inner wall of the tube and neck. Reprinted with permission from H. Pender and K. McIlwain (Eds) *Electrical Engineers Handbook*, 4th Ed., Wiley, 1967, pp. 15–21.

replaced the matrix of phototubes with what he called the *mosaic*. This was made up of a very large number of droplets of photosensitive material on one side of a sheet of mica. The other side of the mica sheet was covered with a very thin layer of graphite – the *signal plate*. A cross-section of the iconoscope is shown in Figure 6.4.

The mosaic and the signal plate constitute a large number of tiny capacitors which share one common plate. When the image is projected onto the mosaic, each individual droplet of photosensitive material emits electrons proportional to the light intensity. These electrons are collected by an anode placed close to the mosaic. The mosaic is now a picture “painted with electric charge”. The transformation of the charge into a voltage output is carried out by the electron gun and the associated circuitry.

The electron gun produces a very narrow beam of electrons focussed on the mosaic. The arrival of these new electrons have the effect of “discharging” the tiny capacitor on which they fall. Since the number of electrons required to discharge the capacitor is proportional to the charge induced by the light intensity, the electron gun current is a function of the charge present and hence of the light intensity. The electron gun current is then proportional to the voltage which appears across the resistor R .

In order to transform the picture into a video signal, it is necessary to make the beam of electrons sweep across the mosaic in a series of orderly lines. The electron gun has two deflection systems for scanning the picture. The first moves the electron beam at a constant speed along a horizontal straight line and returns it very quickly to the start ready for the next sweep (horizontal trace). The second system controls

the vertical position of the beam and ensures that each line is swept before returning the beam to the top of the picture ready for the next frame (vertical trace).

Two major disadvantages of the iconoscope are the high light intensity required to obtain acceptable quality images and the production of secondary electrons from the photosensitive material during the scanning process. The secondary electrons cause noise (false information) in the video signal. In due course, the iconoscope was replaced by the *image orthicon*, another invention of Zworykin.

6.3.1.2 The Image Orthicon. In the image orthicon [6] the two functions of the mosaic in the iconoscope, which are (a) the production of the image in terms of charge and (b) the target for the electron beam scan, are separated. Figure 6.5 shows a cross-section through the image orthicon.

The camera lens focusses the image onto the *photocathode*. Every small segment of the photocathode emits electrons proportional to the amount of light falling on it. The electrons are attracted to the *target* by the positive voltage applied to it. The *focus coil* creates a magnetic field which ensures that the electrons travel to the target in straight parallel paths. The target is a very thin sheet of glass which has a high conductivity through its thickness but low conductivity across points on its surface. The electrons striking the target dislodge secondary electrons from the photocathode side of it and are immediately captured by the very fine wire mesh, called the *target screen*, which has a relatively low positive voltage on it. This leaves a positive charge on the photocathode side of the target. Due to the high conductivity through the thickness of the target an identical pattern of positive charge is produced on the other side of it.

The *electron gun* positioned at the other end of the evacuated tube produces a narrow stream of electrons which are accelerated towards the target by the high positive voltage on the *accelerator anode*. The accelerator anode is a graphite coating on the inside of the neck of the tube. If the electrons were allowed to strike the target without any control of their speed, it is clear that they will produce secondary electrons. To control the speed of the electrons on arrival at the target, a second ring of graphite coating, called the *decelerator grid*, is provided. By adjusting the voltages on the accelerator anode and the decelerator grid, the speed of the electron stream at the target can be controlled to ensure that no secondary electrons are emitted. The electron stream striking an elemental area of the target uses up some of the electrons to neutralize the positive charge left there by the image. The remaining electrons are attracted in a backward direction by the accelerating anode but they take a different path. By placing an *electron collector* in the appropriate position in the neck of the tube, the returning electrons can be collected and amplified by an *electron multiplier*. The output of the electron multiplier is converted into a voltage inversely proportional to the amount of light impinging on the elemental area of the photocathode. By using a scanning system in conjunction with the electron gun (deflection coil), the complete image can be produced at the output in the form of a varying voltage which is a function of the brightness of all the elements of the image projected on the photocathode.

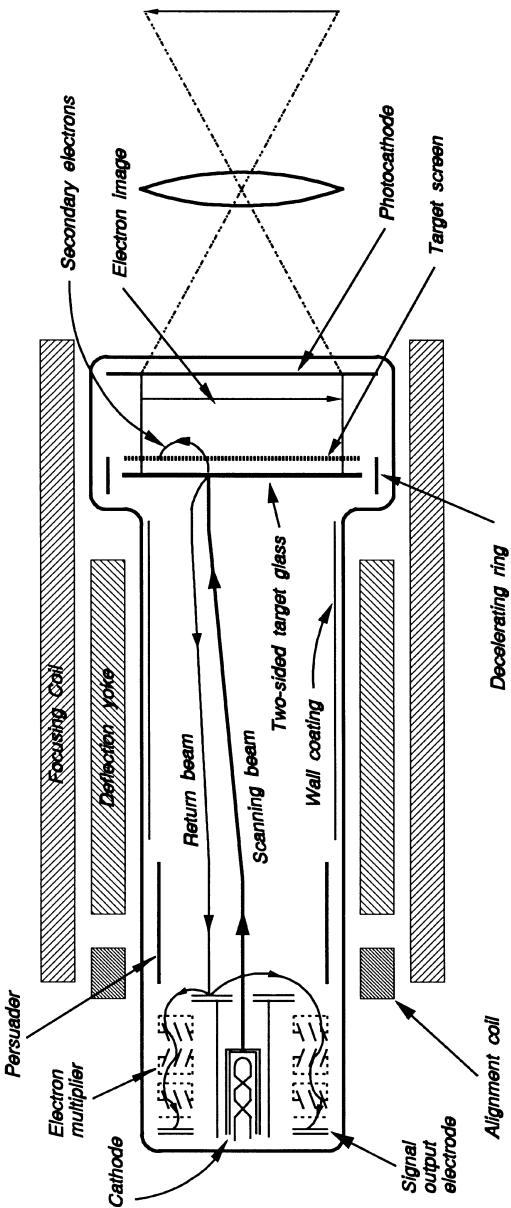


Figure 6.5. A cross-sectional view of the image orthicon. Reprinted with permission from V. K. Zworykin and G. Morton, *Television*, 2nd Ed., RCA, 1954, p. 367.

6.3.1.3 Vidicon. The vidicon is not as sensitive as the image orthicon but it is generally smaller in size and therefore well adapted to field applications such as surveillance, where high resolution is not critical. Its principle of operation is different from both the image orthicon and the earlier iconoscope since it relies on a change of photoconductivity as a function of light intensity. An electron scanning system is used to extract the video information.

6.3.2 Scanning System

In Section 6.3.1, the role of the electron beam scanning system in the production of the video signal was discussed. A series of pulses are generated which control the initiation of the horizontal sweep of the electron beam across the target (horizontal trace) from the left-hand side to the right. A blanking pulse is used to cut off the beam while it is returned to the left-hand side ready for the next sweep (harmonic retrace). During this period, the vertical trace circuit moves the beam down just the right distance for the second line to be swept. When the whole frame has been scanned, a reset pulse returns the beam to the top left-hand corner ready to repeat the process. The beam is blanked during the reset. All the control pulses are added to the video signal and used at the receiver to synchronize the receiver to the transmitter. All the pulses are derived from the 60 Hz power supply or from a 31.5 kHz crystal controlled oscillator and divided down to the appropriate frequency.

Black-and-white television in North America uses 525 horizontal lines of scan to cover the image. If the lines were scanned sequentially, the persistence of the phosphor on the first line would have faded before the last line would have been scanned. This would lead to an annoying flicker on the picture tube, especially at high levels of brightness. The scan is therefore *interlaced*, that is, all odd numbered lines are scanned first from 1 to 525 and then the beam is returned to line 2 and all even numbered lines are then scanned. So each frame has $525/2$ lines and they are scanned 60 times per second. This means that there are 262.5×60 lines scanned per second. The horizontal scan therefore operates at 15.75 kHz. It must be pointed out that not all the 525 lines are used for picture production; some are used for vertical retrace and other controls as well as equalization and synchronizing purposes.

The period for the horizontal scan is 63 μ s and it is divided up into 56 μ s for the trace and 7 μ s for the retrace. The beam deflection system is therefore required to move the beam from the left of the picture to the right in 56 μ s and return it in 7 μ s.

The mechanism for deflecting the electron beam relies on the fact that, when a current-carrying conductor is placed in a magnetic field, the conductor experiences a force mutually orthogonal to the field and the direction of the current. Motion will occur if the conductor is free to move. In the camera tube, the electron gun and its associated components produce a stream of electrons. Each of these electrons carries a charge and a moving charge is a current. The electron beam therefore experiences a force on it and, since it is free to move, it will be deflected from its original path. Figure 6.6(b) illustrates the situation.

Consider the current in the conductor (the electron beam) to be flowing into the plane of the page. The “right-hand rule” gives the direction of the magnetic flux B_i

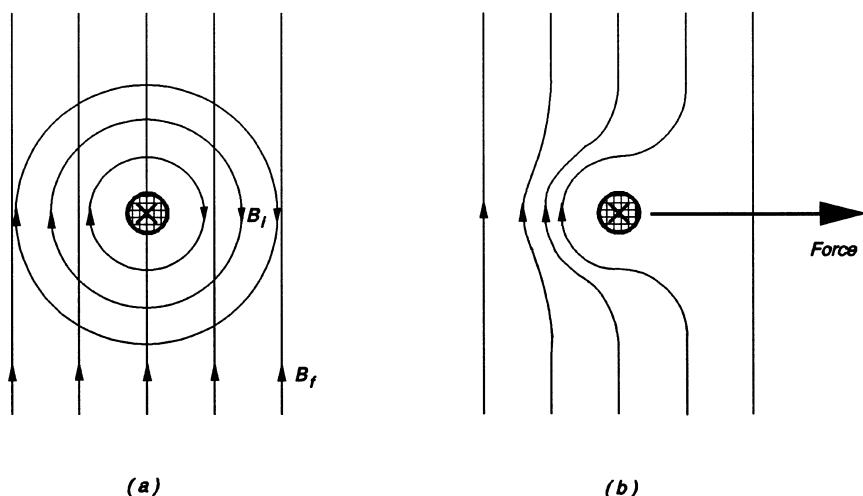


Figure 6.6. (a) The magnetic field, B_i due to a current, i flowing perpendicular to the page superimposed on a uniform magnetic field, B_f , with no interaction. (b) The combined effect of the two magnetic fields generates a force mutually at right angles to the conductor and magnetic field.

associated with the current in the conductor as concentric circles in a clockwise direction. Assuming the conductor is placed in a magnetic flux B_f which is in the vertical upward direction, the two fluxes will interact to produce the distorted field shown in Figure 6.6(b). The flux on the left of the conductor is strengthened while that on the right is weakened; the conductor therefore moves to the right. When trying to determine the direction of electron deflection, it is necessary to remember that electron flow is opposite to current flow.

The deflection of the electron beam therefore depends on designing a circuit to supply the appropriate current to the horizontal deflection coil so that it produces a magnetic field which is a linear ramp with respect to time. The vertical deflection system also has to be supplied with a current which produces a linear ramp magnetic field and its operation has to be synchronized to the horizontal trace so that it repeats the process after 262.5 cycles.

The video signal, the control pulses, and the output of the FM modulator are combined in the video amplifier. A typical composite video signal is shown in Figure 6.7. Note the following:

- (1) Only horizontal synchronizing pulses are shown ($\sim 5 \mu\text{s}$); vertical synchronization pulses ($\sim 190 \mu\text{s}$) are distinguished from the horizontal ones by the difference in their width.
 - (2) The blanking pulse starts before and ends after the horizontal synchronization pulse.
 - (3) The video signal contains the radio-frequency carrier of 4.5 MHz frequency modulated by the audio signal.

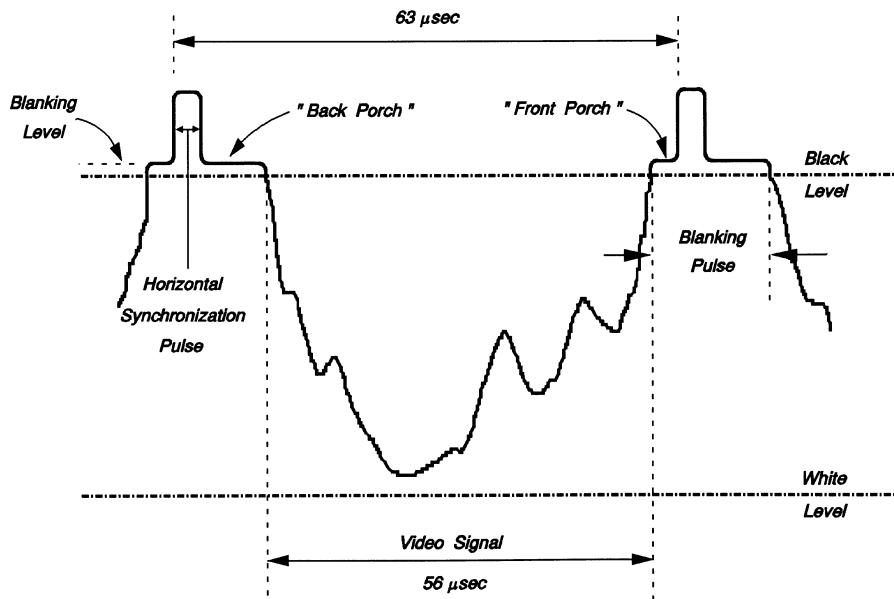


Figure 6.7. A typical composite video signal with blanking and horizontal synchronization pulses.

The spectrum of the composite signal at the output of the video amplifier is shown in Figure 6.8.

6.3.3 Audio Frequency and FM Circuits

Various audio frequency amplifiers were discussed in Section 2.7. The design of the crystal oscillator was discussed in Section 2.4.7 and the frequency multiplier in Section 2.5. Design details for the FM modulator can be found in Section 4.4.

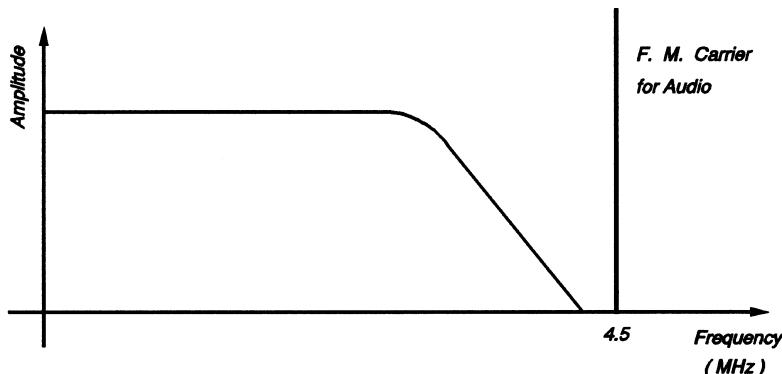


Figure 6.8. The spectrum of the composite video showing the FM carrier used for the transmission of the audio signal.

6.3.4 Video Amplifier

6.3.4.1 Calculation of Bandwidth. Ideally, a video signal consists of frequencies from zero (dc) to some high frequency. The dc response is required when large areas of black, white or other intermediate shades have to be transmitted. The limit of the high-frequency response is determined by the resolution required for a single vertical black line on a white background or vice versa. The minimum visible horizontal line has a height (width) equal to the width of the horizontal trace. To keep the same resolution for the horizontal as for the vertical line, the video signal must go from white to black in

$$56/525 \mu\text{s} = 0.11 \mu\text{s}.$$

The lowest frequency that can change from peak positive to peak negative is a sinusoid as depicted in Figure 6.9. Note that the period of the sinusoid is $2 \times 0.11 \mu\text{s}$ or $0.22 \mu\text{s}$.

It has been determined from psychological experiments that a picture that has a ratio of horizontal-to-vertical dimension less than $4 : 3$ is not pleasing to look at. This ratio has been adopted as standard for all video equipment. It is referred to as the *aspect ratio*. Taking the aspect ratio into account, one cycle of the sinusoid shown in Figure 6.9 must take $0.22 \times 3/4 \mu\text{s}$, (i.e., the period $T = 0.16 \mu\text{s}$). The lowest frequency that must be present in the video signal for it to be able to reproduce the vertical black line on a white background with the same resolution as a horizontal black line on a white background is given by

$$f = \frac{1}{T} = 6.25 \text{ MHz.} \quad (6.3.1)$$

This means that all the video circuits in the system must have bandwidths equal to or greater than 6.25 MHz. Note that the bandwidths of the sub-circuits must be greater than 6.25 MHz because cascading reduces the system bandwidth compared to that of the individual sub-circuit.

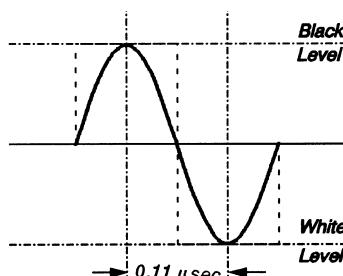


Figure 6.9. The sine wave shown represents the highest frequency required to follow the transition from black to white level in $0.11 \mu\text{s}$.

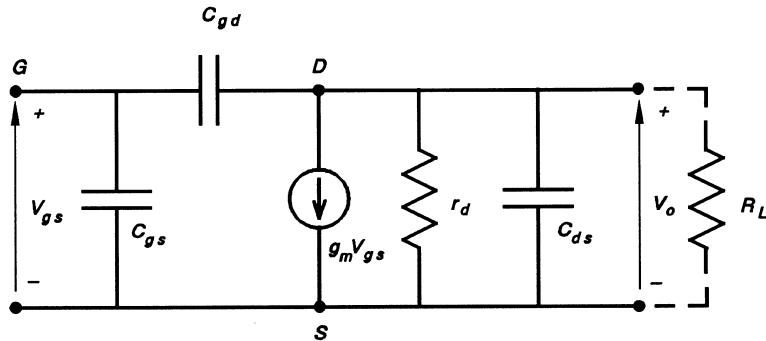


Figure 6.10. The equivalent circuit of a common-source FET.

6.3.4.2 Video Amplifier Design. Because the video signal contains frequencies from zero to some high frequency, a video amplifier must be designed to have a flat response from dc to the high frequency in question. Ideally, all circuits in the video chain must be directly coupled. In practice, an amplifier with a bandwidth from approximately 30 Hz to approximately 4.0 MHz is considered to be a video amplifier.

Consider a FET video amplifier [1] in which the FET is modelled in the common-source mode as shown in Figure 6.10.

Since $r_d \gg R_L$ and $|X_{C_{ds}}| \gg r_d$, both r_d and C_{ds} can be neglected as shown in Figure 6.11.

From Figure 6.11

$$I_1 = j\omega C_{gs} V_{gs} \quad (6.3.2)$$

$$I_2 = j\omega C_{gd}(V_{gs} - V_o) \quad (6.3.3)$$

$$I_3 = I_1 + I_2 = j\omega V_{gs}[C_{gs} + (1 - \frac{V_o}{V_{gs}})C_{gd}] \quad (6.3.4)$$

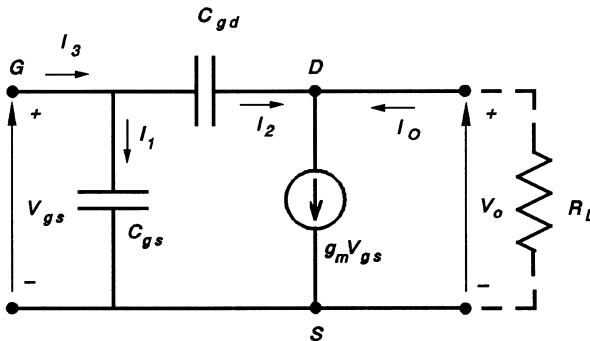


Figure 6.11. A simplified equivalent circuit of the FET with r_d and C_{ds} excluded.

Now $I_1 \gg I_2$, therefore

$$V_o = -g_m V_{gs} R_L. \quad (6.3.5)$$

Substituting into Equation (6.1.4),

$$I_3 = j\omega V_{gs} [C_{gs} + (1 + g_m R_L) C_{gd}] \quad (6.3.6)$$

$$\frac{V_{gs}}{I_3} = \frac{1}{j\omega [C_{gs} + (1 + g_m R_L) C_{gd}]} \quad (6.3.7)$$

The FET can now be represented by the equivalent circuit shown in Figure 6.12 in which

$$C_g = [C_{gs} + (1 + g_m R_L) C_{gd}]. \quad (6.3.8)$$

Consider the two-stage video amplifier shown in Figure 6.13.

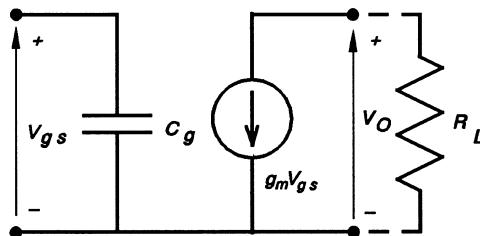


Figure 6.12. A modified form of the FET shown in Figure 6.11.

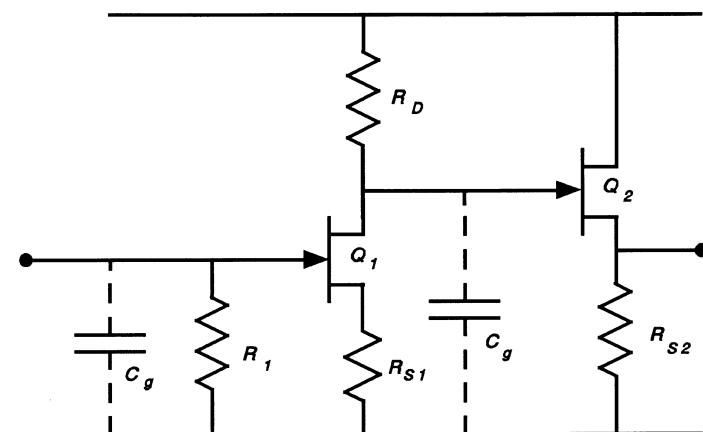


Figure 6.13. A two-stage video FET amplifier showing C_g for both devices.

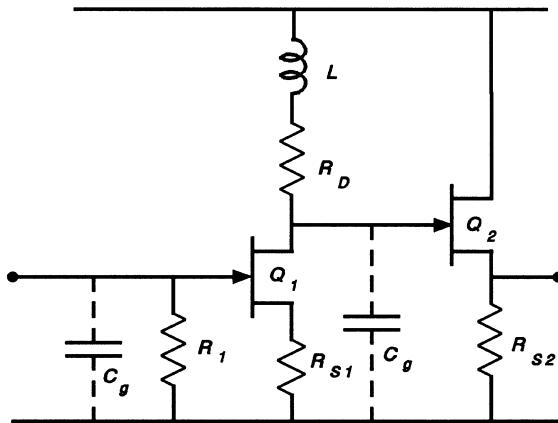


Figure 6.14. The amplifier of Figure 6.13 with the inductor L connected in series with R_D .

The resistor R_1 is used to bias the gate of Q_1 and it is usually very large ($M\Omega$ s). The time-constant $C_g R_1$ is therefore very large compared to $C_g R_D$. To a first approximation, the time-constant $C_g R_D$ determines the high-frequency response of the amplifier. The -3 dB frequency is

$$\omega_2 \approx \frac{1}{R_D C_g}. \quad (6.3.9)$$

The bandwidth ω_2 can be increased by connecting an inductance in series with the drain resistance R_D as shown in Figure 6.14.

The equivalent circuit is shown in Figure 6.15.

The load is

$$Z_L = \frac{1}{(R_D + sL)^{-1} + sC_g} \quad (6.3.10)$$

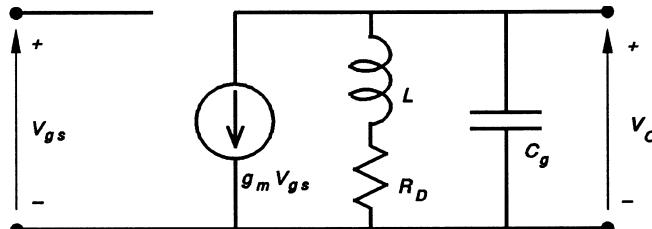


Figure 6.15. A simplified equivalent circuit of the first stage of the FET amplifier with the C_g of the second stage connected across the output.

where

$$s = \sigma + j\omega \quad (6.3.11)$$

and

$$V_o = -g_m V_{gs} Z_L. \quad (6.3.12)$$

The gain is

$$A = \frac{V_o}{V_{gs}} = -g_m Z_L \quad (6.3.13)$$

$$A = \frac{V_o}{V_{gs}} = -\frac{g_m}{C_g} \frac{s + \left(\frac{R_D}{L}\right)}{s^2 + s\left(\frac{R_D}{L}\right) + \omega_0^2} \quad (6.3.14)$$

where

$$\omega_0^2 = \frac{1}{LC_g}. \quad (6.3.15)$$

The gain function has two poles at

$$s_1, s_2 = -\frac{R_D}{2L} \pm \sqrt{\frac{R_D^2}{4L^2} - \omega_0^2}. \quad (6.3.16)$$

When

$$\omega_0^2 = \frac{R_D^2}{4L^2} \quad (6.3.17)$$

the gain function poles are conjugate on the imaginary axis of the s plane and these increase the gain at the resonant frequency ω_0 .

The critical value of L is

$$L_{crit} = \frac{R_D^2 C_g}{4}. \quad (6.3.18)$$

The response is then as shown in Figure 6.16.

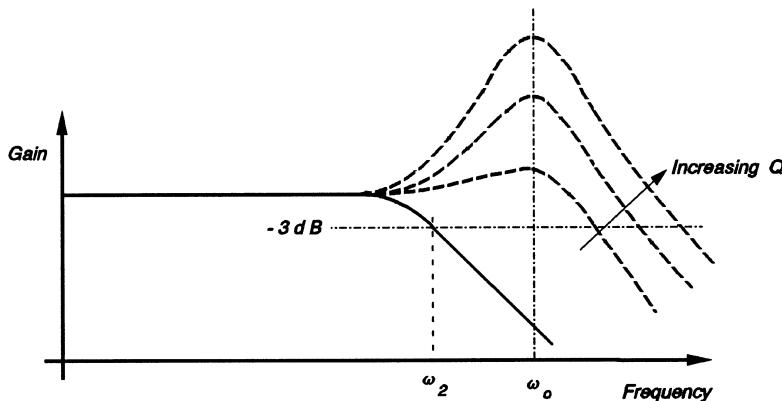


Figure 6.16. The frequency response of the amplifier after the addition of the inductor. Note that the resonant effect increases the gain at high frequency, thereby increasing the bandwidth.

The shape of the gain response can be controlled by varying the Q factor of the circuit. Let

$$q = \frac{\omega_2 L}{R_D} = \frac{L}{R_D^2 C_g} = \frac{L}{4L_{\text{crit}}}. \quad (6.3.19)$$

Substituting q and $s = j\omega$ into Equation (6.1.14) gives

$$|A| = \bar{A} \sqrt{\frac{1 + q^2 \left(\frac{\omega}{\omega_2}\right)^2}{1 + (1 - 2q) \left(\frac{\omega}{\omega_2}\right)^2 + q^2 \left(\frac{\omega}{\omega_2}\right)^4}}. \quad (6.3.20)$$

The gain response can, in general, be represented by the ratio of two polynomials in ω^2 :

$$H(\omega) = \frac{1 + a_1\omega^2 + a_2\omega^4 + \dots}{1 + b_1\omega^2 + b_2\omega^4 + \dots} \quad (6.3.21)$$

which gives a flat gain–frequency response when $a_1 = b_1$ and $a_2 = b_2$ and so on.

Equating the coefficients of $(\omega/\omega_2)^2$ in Equation (6.3.21) gives

$$q = \pm\sqrt{2} - 1 = 0.414 \quad \text{or} \quad -2.414. \quad (6.3.22)$$

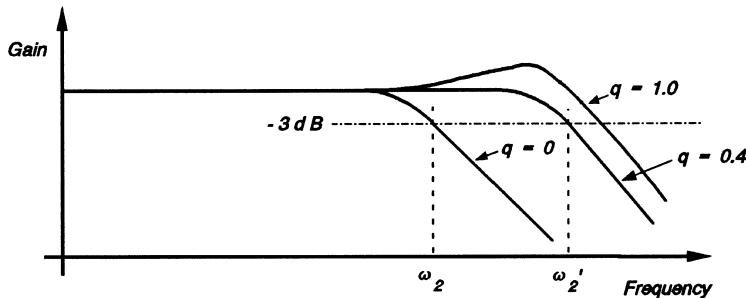


Figure 6.17. The frequency response of the amplifier for various values of q .

The negative value has no physical meaning and it is neglected. Substituting into Equation (6.3.19), we get

$$q = 0.414 = \frac{L}{4L_{\text{crit}}} = \frac{L}{R_D^2 C_g} \quad (6.3.23)$$

and

$$L = 1.656L_{\text{crit}} = 0.414R_D^2 C_g. \quad (6.3.24)$$

The frequency response for three values of q are shown in Figure 6.17.

The new -3 dB frequency is

$$\omega'_2 = 1.72 \omega_2. \quad (6.3.25)$$

Example 6.3.1 Video Amplifier. A video amplifier uses FETs and has two stages as shown in Figure 6.18.

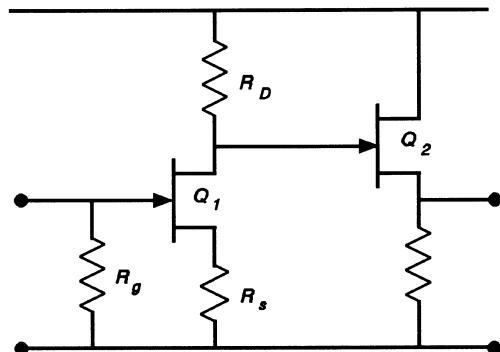


Figure 6.18. The video amplifier used in Example 6.3.1.

The FETs have the following parameters:

$$g_m = 0.004 \text{ S}, \quad C_{gs} = 1.0 \text{ pF}, \quad C_{gd} = 1.5 \text{ pF}.$$

Assuming that only the Miller effect capacitance affects the high-frequency response, calculate the -3 dB frequency of the amplifier when the gain is 15. Calculate the value of an inductance which, when connected in series with R_D , will extend the -3 dB cut-off frequency as high as possible while keeping the gain response flat. Calculate the new cut-off frequency.

Solution. Neglecting all capacitance effects, the equivalent circuit of the amplifier is shown in Figure 6.19, where

$$V_o = -g_m V_{gs} R_D \quad (6.3.26)$$

and the voltage gain is

$$A = \frac{V_o}{V_{gs}} = -g_m R_D = 15. \quad (6.3.27)$$

Therefore

$$R_D = \frac{15}{0.004} = 3750 \Omega. \quad (6.3.28)$$

The voltage gain

$$A = \frac{V_o}{V_{gs}} \approx \frac{R_D}{R_s} = 15. \quad (6.3.29)$$

Therefore

$$R_s = 250 \Omega. \quad (6.3.30)$$

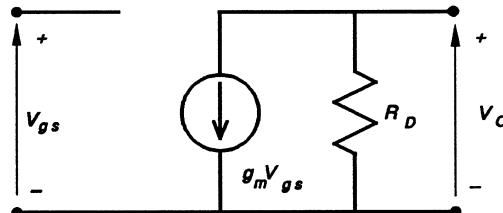


Figure 6.19. The equivalent circuit of stage 1 of the amplifier when all capacitances have been neglected.

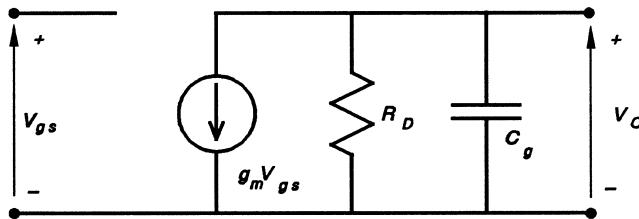


Figure 6.20. The equivalent circuit of stage 1 of the amplifier with the C_g of stage 2 included.

R_g can be made arbitrarily large since no current flows in it. The Miller effect capacitance of Q_1 will have negligible effect on the frequency response, especially if the internal resistance of the driving source is small. The Miller effect capacitance of Q_2 will appear across R_D as shown in Figure 6.20 where

$$C_g = C_{gs} + (1 + g_m R_D) C_{gd} = 1.0 + (1 + 15)1.5 = 25 \text{ pF}. \quad (6.3.31)$$

The -3 dB frequency ω_2 is determined by

$$\omega_2 \tau = 1 \quad (6.3.32)$$

where

$$\tau = R_D C_g \quad (6.3.33)$$

and

$$f_2 = \frac{1}{2\pi R_D C_g} = 1.67 \text{ MHz}. \quad (6.3.34)$$

When the inductance L is connected in series with R_D and L has the critical value,

$$L = 4qL_{\text{crit}} = qR_D^2 C_g. \quad (6.3.35)$$

When $q = 0.414$,

$$L = 145.5 \mu\text{H}. \quad (6.3.36)$$

The new -3 dB cut-off frequency is

$$f'_2 = 1.72f_2 = 2.87 \text{ MHz}. \quad (6.3.37)$$

6.3.5 Radio-Frequency Circuits

The composite video signal from the output of the video amplifier is used to amplitude modulate the radio-frequency carrier obtained at a lower frequency from a crystal oscillator and multiplied by a suitable factor by a frequency multiplier. Amplitude modulators are discussed in Section 2.6, frequency multipliers in Section 2.5, and crystal-controlled oscillators in Sections 2.3 and 2.4.

6.3.6 Vestigial Sideband Filter

From Figure 6.8, it can be seen that the composite signal occupies a bandwidth of approximately 0–4.5 MHz. If both sidebands were transmitted, it would require a bandwidth of just over 9.0 MHz. Not only would that consume scarce bandwidth but it is also unnecessary because the information in the upper and lower sidebands are the same. Single-sideband (SSB) transmission could be used but the filter needed to remove one of the sidebands is fairly sophisticated and the demodulation equipment is complex and difficult to maintain. A compromise is to remove most of the lower sideband using a bandpass filter. As shown in Section 7.2.4, a simple envelope detector is adequate for the demodulation of such an AM signal if the index of modulation is low. The frequency spectrum at the output of the vestigial sideband filter is shown in Figure 6.21.

6.3.7 Antenna

Television broadcast frequencies are either in the very-high-frequency (VHF) band which is from 30 to 300 MHz or in the ultra-high-frequency (UHF) band which is from 300 to 3000 MHz. At these frequencies, antennas have highly directional properties. To get the circular radiation pattern in the horizontal plane normally used for broadcasting television signals, several arrangements of antenna arrays can be

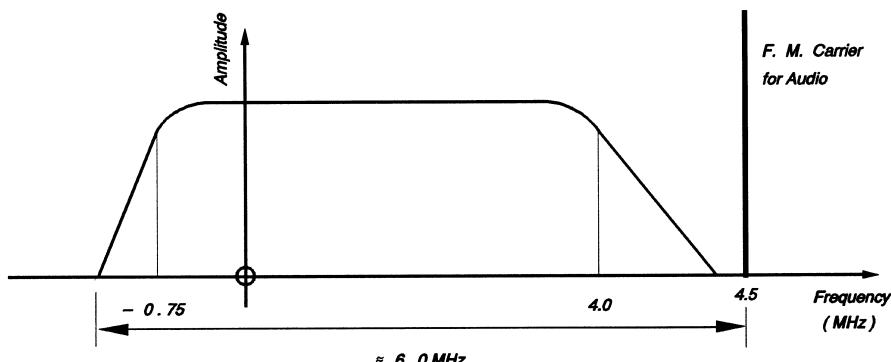


Figure 6.21. The spectrum of the video signal after amplitude modulation and vestigial sideband filtering. Note that the AM carrier signal occupies the zero frequency position after modulation.

used. The most popular is the *turnstile* array. The basic principle is that, when two or more radiating elements such as dipoles are placed in close proximity to each other, they interact to produce a radiation pattern that is the vector addition of the individual elements. By varying the relative physical positions of the elements and the phase angle of the signal, it is possible to use the interactive properties to create a radiation pattern which is approximately circular in the horizontal plane.

6.3.8 Color Television

The transmission of video signals in color is a subject which can take up several volumes. However, because color television is so common, a simplified explanation of how it works is now offered.

The first step is to discuss some of the properties of color and the results of mixing them. There are three primary colors: red, blue and green, and by using appropriate proportions of these, all other colors perceived by the human eye can be obtained. A simple framework which makes it easy to understand the properties of color, the *color triangle* [2,3], is shown in Figure 6.22.

The primary colors occupy the apices of the triangle. A mixture of equal proportions of red and blue produces the color *magenta*; that of red and green produces *yellow*, and finally blue and green produce *cyan*. These fit into the scheme as shown. The center of the triangle represents *white* since equal proportions of red, blue, and green produces white. Two pieces of information are required in order to code a video signal in color. The first is brightness; the better technical term is

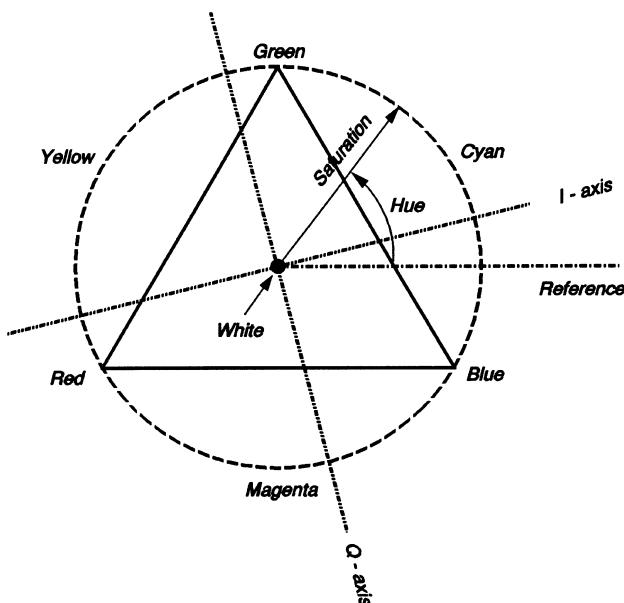


Figure 6.22. The color triangle with the *I* and *Q* axes.

luminance, which is a measure of the energy in the light. Luminance is the only information required for monochromatic (black-and-white) television. The second is *chrominance*, or simply color. Chrominance is best represented on the diagram by a polar coordinate scheme whose origin is at the center (white) of the triangle. When one moves along a straight line from the origin towards the red apex one sees at first white light followed by a slowly increasing presence of red (pink). At the apex (pure red light) the color red is said to have reached *saturation*. The distance from the center of the triangle (the magnitude of the complex number) is then a measure of the *saturation* of the particular color. The angle of the complex number, with respect to some arbitrary reference line, is called the *hue* of the light. The combination of chrominance and luminance will produce a cylindrical coordinate system in which luminance will be on the axis of the cylinder. In this scheme, the coding of color must have three coordinates: luminance, saturation, and hue.

The above scheme does not take into account the sensitivity of the human eye across the color spectrum and it needs to be modified since the human eye is the final arbiter (detector) of the video signals transmitted. The response of the human eye is shown in Figure 6.23.

A mixture of equal proportions of the primary colors is therefore perceived by the human eye as a greenish-yellow color. In fact, to obtain white light the mixture has to contain 59% green, 30% red, and only 11% blue.

The simplest system for coding video information for television transmission is simply to scan the picture for the three primary colors and process each one as if it were monochromatic. Some multiplex system could then be worked out to accommodate all three signals. Such a scheme was tried in the United States in the late 1940s and early 1950s but was found to be incompatible with the black-and-white system then in existence.

In 1953, a scheme proposed by the National Television System Committee (NTSC) was accepted by the Federal Communications Commission (FCC) and it has since been adopted as a standard by television broadcasting authorities in many

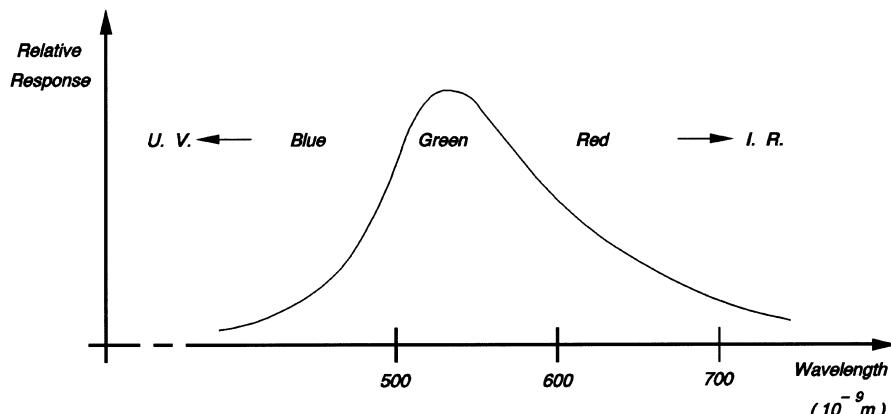


Figure 6.23. The relative sensitivity of the human eye to the wavelength of light.

countries. The NTSC scheme separates the signal into chrominance and luminance (the chrominance information is further separated indirectly into I and Q components, which correspond approximately to hue and saturation). This will be explained later. The actual system acquires the luminance information by taking the weighted sum of the three primary colors to obtain white light as perceived by the human eye as:

$$L_w = 0.59L_g + 0.30L_r + 0.11L_b. \quad (6.4.1)$$

L_w then contains all the information for black-and-white picture transmission. For color pictures, the three signals L_w , $(L_r - L_w)$ and $(L_b - L_w)$ are sent. The third possible signal $(L_g - L_w)$ is not used because it is the smallest of the three difference signals and therefore more likely to be corrupted by noise. The two difference signals $(L_r - L_w)$ and $(L_b - L_w)$ are quadrature multiplexed, a form of DSB-SC modulation, with the color sub-carrier of frequency ω_o . The two difference signals have the form:

$$(L_b - L_w) \cos \omega_o t + (L_r - L_w) \sin \omega_o t. \quad (6.4.2)$$

The signal

$$L_w + (L_b - L_w) \cos \omega_o t + (L_r - L_w) \sin \omega_o t \quad (6.4.3)$$

is used to amplitude modulate a suitable carrier in the usual vestigial sideband scheme.

At the receiving end, the two difference signals are de-multiplexed by multiplying them by the output of a local oscillator which is synchronized to the color sub-carrier of frequency ω_o in the transmitter. The color sub-carrier signal is sent in “bursts” during the trailing portion of the horizontal blanking pulse called the *back porch*. This is shown in Figure 6.24.

The demodulation system is shown in Figure 6.25.

The recovered signals L_w , $(L_b - L_w)$ and $(L_r - L_w)$ are used to reconstruct the missing difference signal $(L_g - L_w)$ as follows:

$$(L_g - L_w) = -\frac{0.30}{0.59}(L_r - L_w) - \frac{0.11}{0.59}(L_b - L_w). \quad (6.4.4)$$

The four signals are used to control the appropriate grids on the picture tube to reproduce the original scene.

In the above discussion, the bandwidth requirements of the three signals L_w , $(L_b - L_w)$ and $(L_r - L_w)$ were assumed to be the same. This is, in fact, not true. The human eye has a higher resolution in black-and-white than it does in color. It is adequate to present the uniform, large areas of the picture (low frequency) in color; the fine details (high frequency) can be presented in black-and-white. The luminance information L_w must have the whole bandwidth available. The difference signal containing the chrominance information requires much less bandwidth.

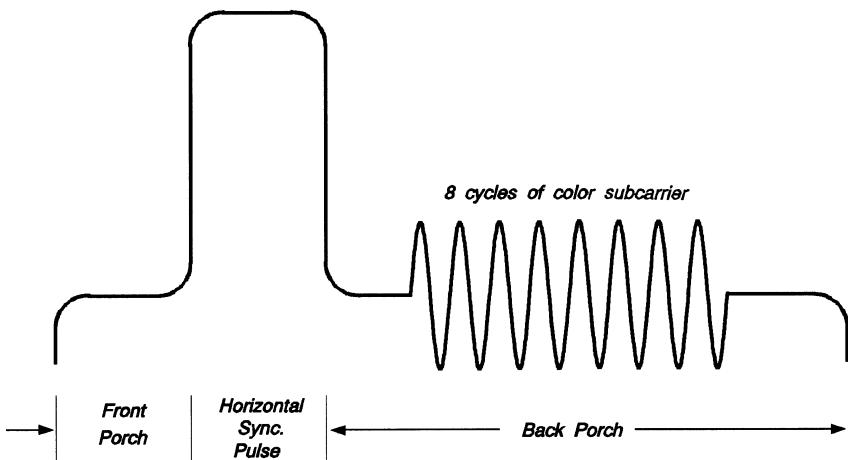


Figure 6.24. A modified back porch of the blanking pulse used to transmit the color subcarrier.

Further advantage can be taken of the greater sensitivity of the human eye to green-yellow light by de-emphasizing the green-yellow component by using a smaller bandwidth for transmitting that part of the spectrum. This is called the *Q* signal. The blues and the reds, to which the human eye is less sensitive, are given a relatively larger bandwidth. This is called the *I* signal. The comparative spectra for the luminance, *Q* and *I* signals are shown in Figure 6.26.

Because the *Q* signal occupies a total bandwidth of 1 MHz, and it fits within the overall bandwidth limits of the signal, both sidebands are transmitted. The upper sideband of the *I* signal lies outside the overall bandwidth and part of the upper sideband is removed by a vestigial sideband filter.

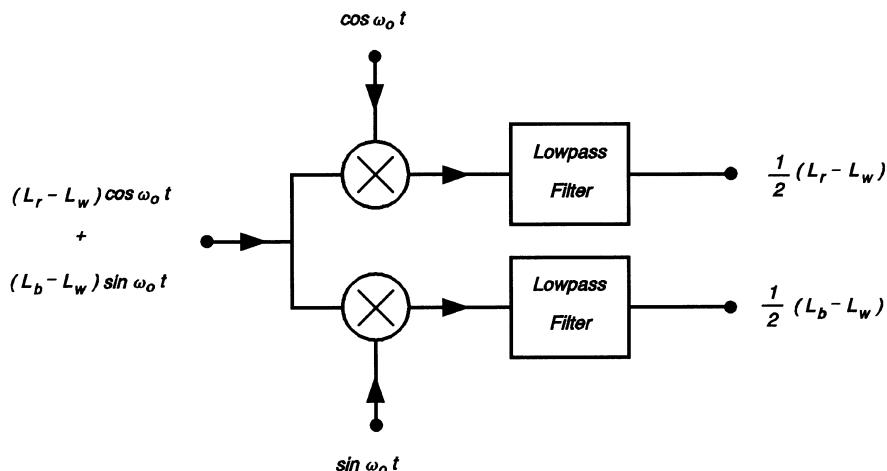


Figure 6.25. A scheme for demultiplexing the color information.

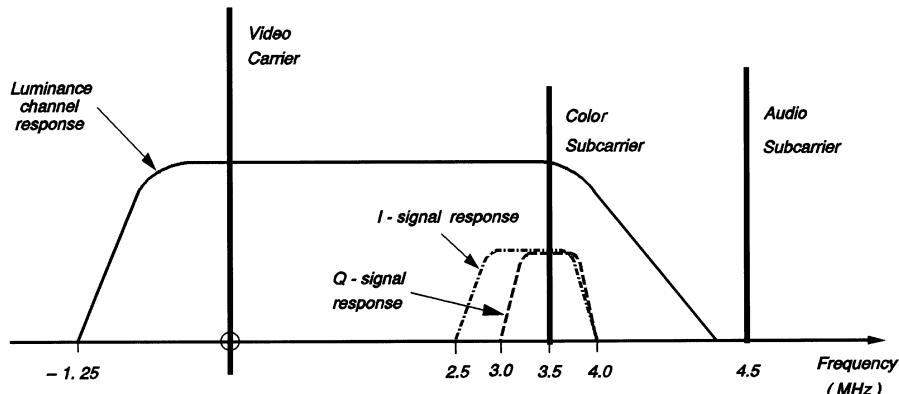


Figure 6.26. The full spectrum of the color TV signal. Note that zero frequency corresponds to the AM (video) carrier frequency.

REFERENCES

1. Ryder, J. D., *Electronic Fundamentals and Applications: Integrated and Discrete Systems*, 5th Ed., Prentice-Hall, Englewood Cliffs, NJ, 1976.
2. Lambert, J. H., (a) *Beschreibungen einer Farbenpyramide*, Berlin, 1772. (b) "Mémoire sur la Partie Photometrique de l'Art de Peinture", Preussische Akademie der Wissenschaften, Berlin; *Histoire de l'Acad. Royal des Science et Belles Lettres*, 24, 80–108, 1768.
3. Maxwell, J. C., "On the Theory of Three Primary Colours" in Niven, W. D. (Ed), *The Scientific Papers of James Clerk Maxwell*, Cambridge University Press, Cambridge, England, 1890.
4. Zworykin, V. K., Morton, G. A. and Flory, L. E., "Theory and Performance of the Iconoscope", *Proc. IRE*, 25(8), 1071–1092, 1937.
5. Zworykin, V. K. and Morton, G., *Television*, 2nd Ed., John Wiley, New York, 1954.
6. Rose, A., Weimer, P. K. and Law, H. B., "The Image Orthicon – A Sensitive Television Pickup Tube", *Proc. IRE*, 34(7), 424–432, 1946.

BIBLIOGRAPHY

1. Shure, A., *Basic Television*, J. F. Rider Publisher Inc., New York, 1958.
2. Fink, D. G. and Christiansen, D., *Electronics Engineers' Handbook*, 2nd Ed., McGraw-Hill, New York, 1982.
3. Stover, W. A. (Ed), *Circuit Design for Audio, AM/FM and TV*, McGraw-Hill, New York, 1967.
4. Fink, D. G., *Television Engineering Handbook*, McGraw-Hill, New York, 1957.
5. Stremler, F. G., *Introduction to Communication System*, 3rd Ed., Addison-Wesley, Reading, MA, 1990.

PROBLEMS

- 6.1** What is a video signal? Explain why a video amplifier has to be used in conjunction with a television camera tube.

Explain with the help of a block diagram the configuration and function of each block of a typical television transmitter. Indicate the type of signal you expect at the output of each block in terms of waveform, frequency, and bandwidth.

- 6.2** Explain with the help of a suitable diagram the operation of the iconoscope. What conditions have to be met for it to operate optimally?

- 6.3** Explain with the help of a suitable diagram the operation of the image orthicon. Discuss its performance relative to the iconoscope.

- 6.4** A television system has the following features:

- (1) 625 lines per frame with interlacing,
- (2) 25 frames per second,
- (3) 5% of the trace period is used for the retrace,
- (4) the aspect ratio is 4 : 3.

Assuming equal resolution on the vertical and horizontal axes of the picture tube, calculate the bandwidth required to handle the signal.

- 6.5** The voltage applied to the anode of a picture tube accelerates an electron to a velocity of 10×10^6 m/s along the axis of the tube (x axis). The electron then passes between two parallel (deflection) plates of length 2.5 cm (in the x direction). The distance between the plates in the y direction is 2.0 cm and a dc voltage of 270 V is maintained between them. Assuming that the electrostatic field is uniform and confined to the space between the plates, calculate the following:

- (1) the velocity of the electron in the y direction as it emerges from the electrostatic field,
- (2) the displacement of the electron in the y direction when it hits a target placed 5 cm away from the end of the deflection plates.

You may assume the following:

- (a) the force (in newtons) on the electron due to the electrostatic field is given by

$$F = eE$$

where e is the charge (coulombs) on the electron and E is the electric field intensity (in volts per meter),

(b) the mass of the electron is 9×10^{-31} kg and the charge is 1.6×10^{-19} C.

- 6.6** The electrostatic field described in Problem 6.5 is replaced by a uniform magnetic field of width 2.5 cm in the x direction. The magnetic field is oriented in the z direction. The flux density of the field is 1.44×10^{-3} T (Wb/m²). Assuming no fringing effects, calculate the following:

- (1) the velocity of the electron as it emerges from the magnetic field,
- (2) the displacement of the electron when it hits the target placed 5 cm away from the edge of the magnetic field.

You may assume that:

- (a) the force F (newtons) on a conductor of length l (in meters) carrying a current i (in amperes) in a magnetic field of flux density B (in teslas = Wb/m²) is given by

$$F = Bli$$

- (b) an electron carrying a charge e (in coulombs) at a velocity of v (in meters per second) is, in fact, a current $i = e$ (amperes = coulombs/s) flowing in a conductor of length l (meters).

- 6.7** A video amplifier with a voltage gain of 10 is to be designed in two stages such that the first stage provides the required voltage gain and the second provides a current gain with a low output impedance. Design such a circuit and estimate its bandwidth using the hybrid- π model given in Figure P6.1.

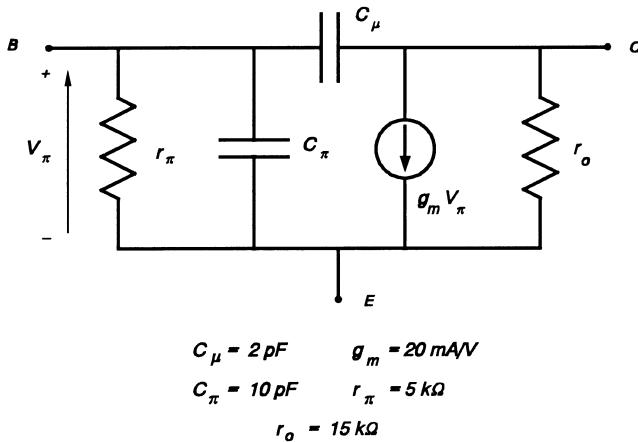


Figure P6.1.

Calculate the value of an inductor that, when connected in series with the collector load of the voltage-gain stage, will optimize the bandwidth of the amplifier.

7

THE TELEVISION RECEIVER

7.1 INTRODUCTION

In Chapter 6, the coding of video signals in a form suitable for transmission over a telecommunication channel was discussed. In this chapter, the techniques for decoding the signals and their presentation on a cathode ray tube will be examined.

The television receiver is almost identical to the AM radio receiver in its use of the superheterodyne principle. There are a few differences in the details of the signal processing due to the greater complexity of the system. Figure 7.1 shows a block diagram of a typical television receiver.

The antenna picks up the electromagnetic radiation from the transmitter and feeds it to the radio-frequency amplifier. After amplification and filtering to attenuate other incoming signals from other transmitters, the signal goes to the mixer where it is mixed with the output from the local oscillator. As before, the local oscillator and the radio-frequency amplifier are tuned to track each other with a constant frequency difference equal to the intermediate frequency. The intermediate frequency for the television receiver is usually 45.75 MHz. The signal is subjected to further filtering before it proceeds to the video demodulator for the recovery of the baseband information in the signal. The next stage is to separate the composite video signal into its three components, namely the video proper, the FM sound subcarrier and its sidebands, and the vertical and horizontal control pulses. The video signal is amplified to the level required to drive the picture tube by the video amplifier and the vertical and horizontal control pulses are suitably conditioned and used in the deflection systems of the receiver to synchronize it to the transmitter – a condition that must be met for proper reproduction of the images sent. The FM signal is amplified, amplitude limited and detected and after some amplification it is used to drive the loudspeaker.

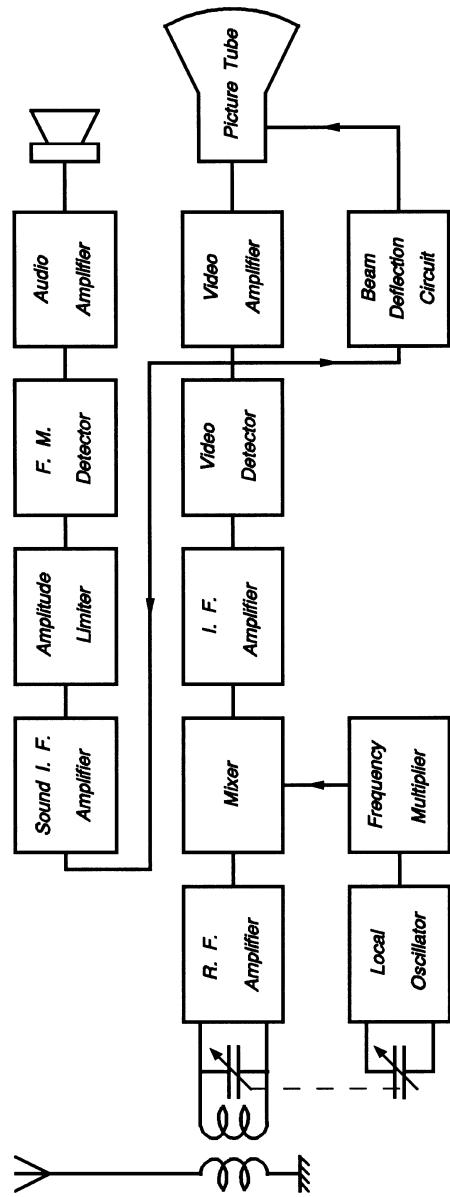


Figure 7.1. The block diagram of the television receiver.

7.2 COMPONENT DESIGN

7.2.1 Antenna

Antenna design is outside the scope of this book. However, a brief qualitative discussion can be found in Section 2.9. Further discussion of antennas for commercial FM reception is presented in Section 5.2.1. Frequencies for commercial FM (88–108 MHz) occupy the spectrum between channels 6 and 7 of the VHF television frequencies (54–88 MHz and 174–216 MHz, respectively). Except for slight differences in the physical dimensions, the antennas tend to take the same form. These are frequency ranges in which a half-wavelength dipole antenna has reasonable physical dimensions (0.7–2.7 m).

7.2.2 Superheterodyne Section

The radio-frequency amplifier is tunable over the VHF frequency range. This is accomplished with a variable capacitor which is mechanically ganged to the variable capacitor which tunes the local oscillator. The objective is to generate a local oscillator frequency which is equal to the radio-frequency amplifier center frequency plus the intermediate frequency. In this case the range of the radio frequency is from approximately 57 MHz (channel 2) to approximately 85 MHz (channel 6) and from approximately 177 MHz (channel 7) to approximately 213 MHz (channel 13) for VHF television. The bandwidth is nominally 6 MHz and the radio-frequency gain is between 20 and 50 times. Since the video intermediate frequency is normally 45.75 MHz, the local oscillator has to be tunable from 102.75 to 258.75 MHz. From Figure 6.21, it can be seen that there are two carrier frequencies in the composite video signal: the video carrier and the voice carrier. The voice carrier is 4.5 MHz above the video carrier, after mixing the video intermediate frequency is at 45.75 MHz and the voice intermediate frequency is at $(45.75 - 4.50)$ or 41.25 MHz. In general parlance, the radio-frequency amplifier, the local oscillator and the mixer are called the *television front end* or simply *television tuner*. These components are generally housed in a separate shielded container in an attempt to control the effects of stray electromagnetic fields and stray capacitive and inductive elements. The principles to be followed in the design of the circuits in the television front end are the same as discussed earlier in connection with radio. The only things that have changed are the frequency of operation and the bandwidth requirements. Greater attention must be paid to the physical layout of the practical circuits. Radio-frequency amplifiers were discussed in Section 2.8. Oscillator design can be found in Section 2.4 and mixer design in Section 3.4.3.

7.2.3 Intermediate-Frequency Amplifier

Like all superheterodyne receiver systems, the detailed selection of the desirable and the rejection of the undesirable frequencies take place at the intermediate-frequency stage. At the same time, some parts of the spectrum may be emphasized to equalize

the quality of the low-frequency video (large uniform areas) and the high-frequency video (areas with fine details). The exact frequency response of the video intermediate-frequency amplifier is of no importance at this point except to point out that it is designed to compensate for the frequency response of the vestigial sideband filter in the transmitter. To understand the techniques used to achieve this objective requires a good understanding of the theory of filter design which is beyond the scope of this book. A list of books on filter design is provided in the bibliography at the end of Chapter 3.

The output of the intermediate-frequency amplifier must be of the order of several volts to drive the video detector that follows. Intermediate-frequency amplifiers were discussed in Section 3.4.4.

7.2.4 Video Detector

It will be recalled that the video signal is amplitude modulated and, in theory, it requires a simple envelope detector to demodulate it. However, the situation is complicated somewhat by the fact that the input signal to the detector is a vestigial sideband signal.

7.2.4.1 Demodulation of Vestigial Sideband Signals. When a carrier of frequency ω_c is amplitude modulated by a signal of frequency ω_m , the result is

$$\phi_1(t) = A[1 + m \cos \omega_m t] \cos \omega_c t \quad (7.2.1)$$

$$\phi_1(t) = A \cos \omega_c t + \frac{mA}{2} [\cos(\omega_c + \omega_m)t + \cos(\omega_c - \omega_m)t]. \quad (7.2.2)$$

In vestigial sideband modulation, one of the sidebands is removed; it could be either of them but in this case it is assumed that it is the lower sideband. Also complete removal is assumed. For simplicity, we have

$$\phi_2(t) = A \cos \omega_c t + \frac{mA}{2} \cos(\omega_c + \omega_m)t \quad (7.2.3)$$

$$\phi_2(t) = A \cos \omega_c t + \frac{mA}{2} \cos \omega_c t \cos \omega_m t - \frac{mA}{2} \sin \omega_c t \sin \omega_m t \quad (7.2.4)$$

$$\phi_2(t) = A(1 + \frac{m}{2} \cos \omega_m t) \cos \omega_c t - \frac{mA}{2} \sin \omega_c t \sin \omega_m t. \quad (7.2.5)$$

The amplitude of the carrier signal is

$$|\phi_2(t)| = \sqrt{A^2 \left(1 + \frac{m}{2} \cos \omega_m t\right)^2 + \left(\frac{mA}{2} \sin \omega_m t\right)^2} \quad (7.2.6)$$

$$|\phi_2(t)| = \sqrt{A^2 \left(1 + \frac{m^2}{4}\right) + A^2 m \cos \omega_m t}. \quad (7.3.7)$$

When the depth of the modulation is low, $m < 1$ and $m^2/4 \ll 1$,

$$|\phi_2(t)| \approx A(1 - m \cos \omega_m t)^{1/2}. \quad (7.2.8)$$

Using the binomial expansion, we get

$$|\phi_2(t)| \approx A \left(1 + \frac{m}{2} \cos \omega_m t \right). \quad (7.2.9)$$

This contains the modulating frequency ω_m as well as its higher harmonics whose amplitudes diminish very rapidly.

The conclusion is that a vestigial sideband signal can be demodulated using a simple envelope detector so long as the modulation index is much less than unity. It is worth noting that the amplitude of the output signal is one-half of what it would have been if both sidebands had been present. Envelope detectors were discussed in Section 3.2. An example is given in Section 3.4.6. Figure 7.2 shows the input and output waveforms of a typical television envelope detector.

7.2.5 The Video Amplifier

The design of video amplifiers was discussed in Section 6.3.4. In the television receiver, the load of the video amplifier is the grid of the cathode ray tube (usually called the picture tube). This requires voltages between approximately 50 V and 100 V and, in theory, no current flows in the grid circuit. However, the grid represents a capacitive load and a capacitance requires the movement of charge (current) to change the voltage across it. The output stage of the video amplifier must be capable of providing the necessary current and hence power. Another way of saying the same thing is that the output stage of the video amplifier must have a low output resistance so that the grid capacitance can be charged much faster than the fastest change in voltage present in the video signal.

7.2.6 The Audio Channel

From the output of the video detector, a bandpass amplifier selects and boosts the FM signal centered at 4.5 MHz. The limiter is described in Sections 4.3.3.1 and 4.3.3.2, and the FM detector is identical to that described in Section 4.3.3.4. The audio-frequency amplifier was described in Section 2.7. The loudspeaker was discussed in Section 3.4.8.

7.2.7 Electron Beam Control Subsystem

In the television transmitter, the pulse generator output was used to control the vertical and horizontal sweeps of the electron beam which scanned the mosaic in the camera tube. The same pulses were added to the video signal together with the 4.5 MHz FM voice carrier to make up the composite video. Figure 7.2 shows two

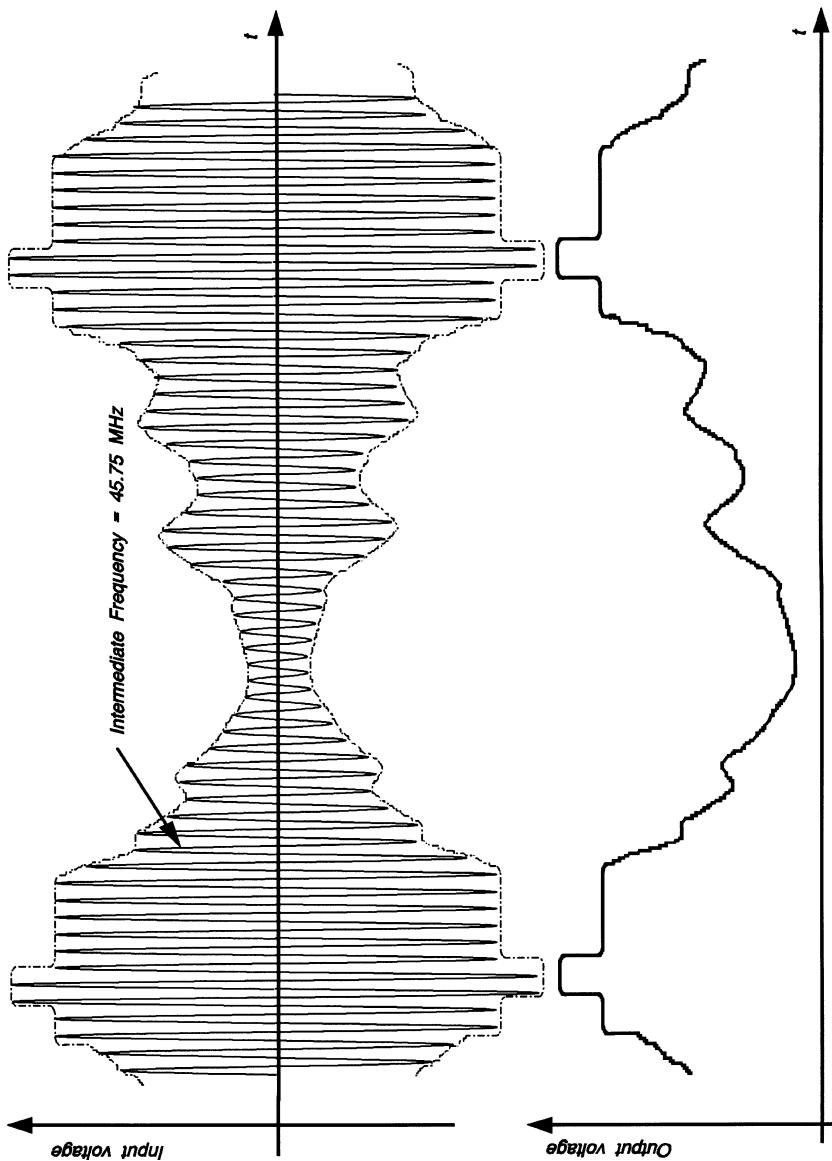


Figure 7.2. A typical intermediate-frequency television signal before and after detection.

such pulses (horizontal sync pulses only shown). The horizontal synchronization pulses are used to control the initiation of the horizontal sweep of the electron beam in the picture tube so that synchronism with the horizontal sweep of the electron beam in the camera tube is maintained. Similarly, the vertical synchronization pulses are used to keep the camera and picture tubes in step in the vertical direction. It is very important to keep the camera and picture tubes in synchronism in both directions, otherwise no meaningful image appears on the picture tube. The first step is to channel the timing information in the sync pulses into a separate circuit for further processing. Figure 7.3 is a block diagram of the electron beam control subsystem.

The composite video signal is fed into the *sync-pulse separator* which takes out both vertical and horizontal sync pulses. The output is used to drive the two separate branches of the system. The vertical branch has the vertical *sync separator* which is designed to produce an output only when a vertical pulse is present at the input. The two timing signals undergo essentially identical process steps, namely synchronization to the oscillator, generation of the sweep signal, and amplification to obtain enough power to drive the deflection coils. It is important to remember that the vertical sync pulses have a frequency of 60 Hz whereas the horizontal runs at 15.75 kHz.

The sync pulses have the following timing characteristics:

Vertical	Horizontal
Field period: 16.683 ms	Line period: 63.556 μ s
Blanking period: 1.335 ms	Blanking period: 10.5–11.4 μ s
Scan period: 15.348 ms	Scan period: 52.156–53.056 μ s

7.2.7.1 Sync Pulse Separator. The basic sync pulse separator is a simple transistor inverter such as shown in Figure 7.4.

The ratio of R_1 to R_2 is chosen so that the transistor remains in cut off until the applied voltage exceeds the blanking level (black level of the video signal). The transistor then conducts and, with an appropriate value for R_3 , it goes into saturation. The output of the circuit is then a series of rectangular pulses coincident with the sync pulses, but inverted. These pulses are used directly to control the horizontal deflection oscillator.

7.2.7.2 Vertical Sync Separator. It must be recalled that the horizontal sync pulses are 5 μ s long while the vertical are 190 μ s. They are easily separated by using a simple low-pass RC filter with a suitable time constant. Figure 7.5 shows a typical vertical sync separator circuit.

The *RC* low-pass filter shown in Figure 7.5 has three sections which should give it a higher rate of amplitude change with frequency. The difference between the two frequencies to be filtered (60 Hz and 15.75 kHz) make the filter design simple.

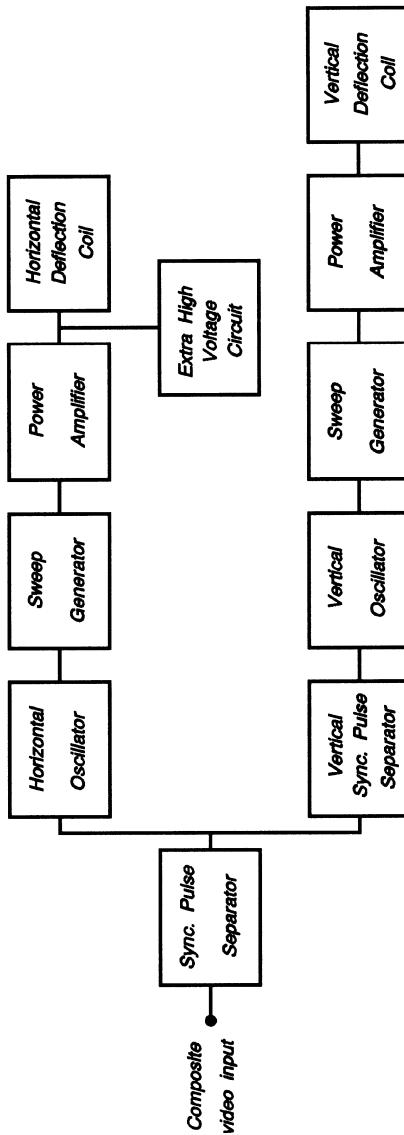


Figure 7.3. The block diagram of the electron beam scan control system.

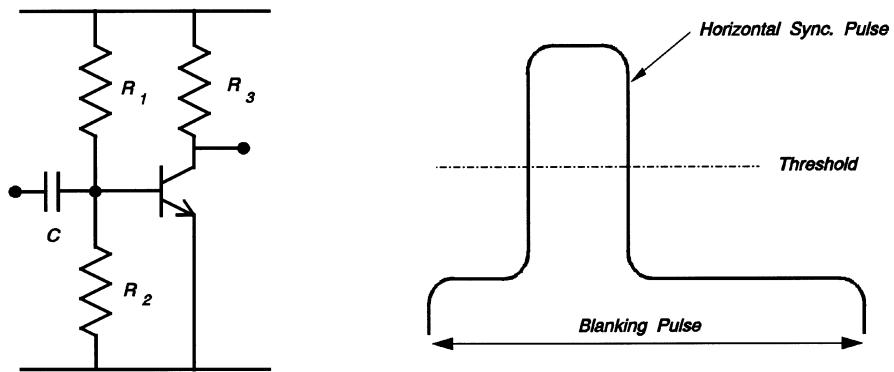


Figure 7.4. The circuit used for separating the horizontal sync pulse. The threshold is set by the relative values of R_1 and R_2 .

7.2.7.3 Vertical Deflection Oscillator. The vertical deflection oscillator is an astable multivibrator which is synchronized to the vertical sync pulses. Figure 7.6(a) shows the circuit diagram of the astable multivibrator.

When the dc power is first switched on, current is supplied to the bases of both transistors and they will both tend to conduct. In general, it can be assumed that one of the two transistors (say Q_1) will conduct a little bit better than the other. The voltage at the collector of Q_1 will therefore drop a little faster than that of Q_2 .

Because it is not possible to change the voltage across the capacitor C_2 instantaneously, the voltage at the base of Q_2 will be forced downwards. This will have the effect of reducing the forward bias on the base-emitter junction of Q_2 . The

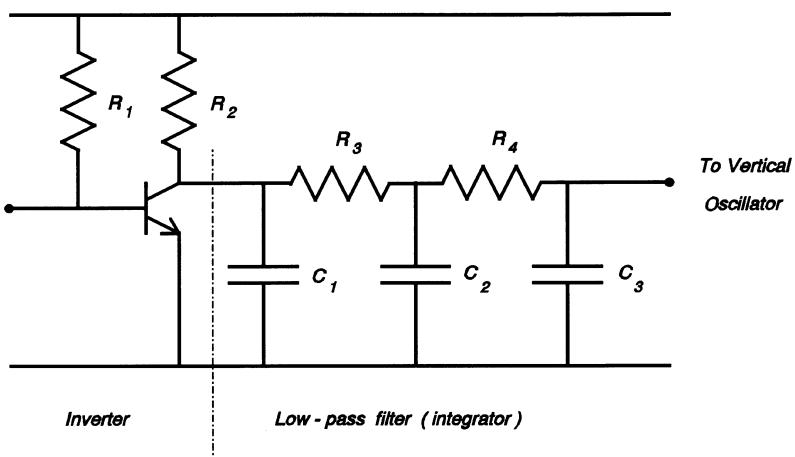
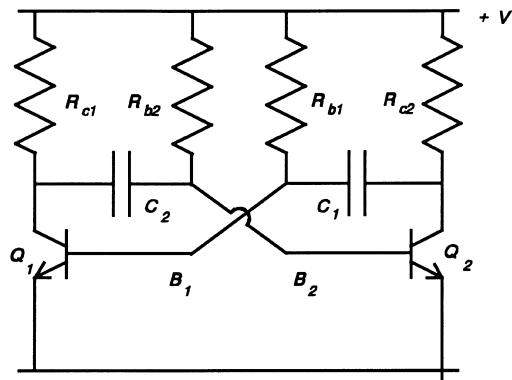
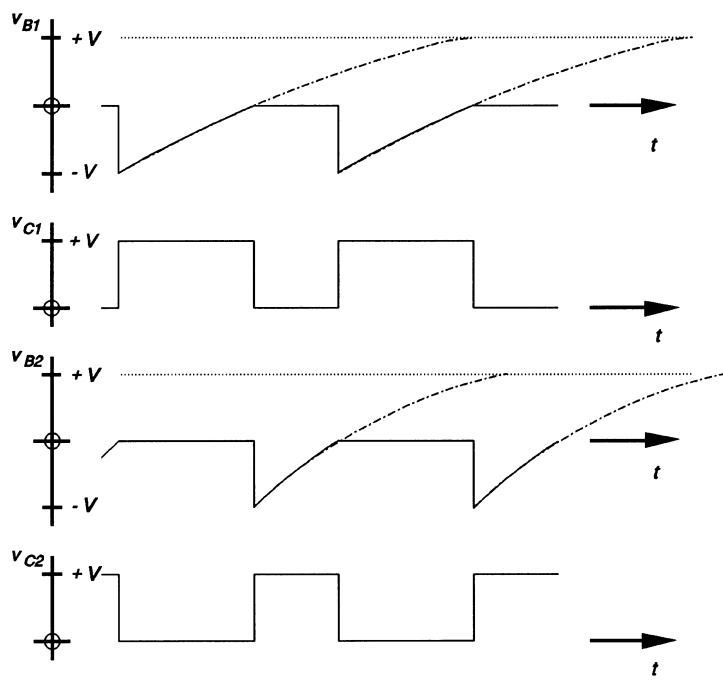


Figure 7.5. The vertical sync pulse separator. The low-pass filter ensures that it does not react to the horizontal sync pulse.



(a)



(b)

Figure 7.6. (a) The circuit diagram of the astable multivibrator. (b) The voltage waveforms for the bases and collectors.

collector current of Q_2 will be reduced and this will make the collector voltage of Q_2 go up. Again the voltage across C_1 cannot change instantly so the collector voltage of Q_2 will tend to force the base of Q_1 upwards. This will cause the base-emitter junction of Q_1 to become more forward biased than it was. More collector current will flow and the collector voltage of Q_1 will drop even more. We are now back to where we started and the downward change which started the chain of events is now magnified. This is evidently a regenerative (positive feedback) process which ends with Q_1 in full conduction and Q_2 cut-off. With the appropriate choice of R_{c1} and R_{b1} , Q_1 goes into saturation, essentially its collector voltage is zero. This phase of the operation is called the regenerative phase and, because regenerative phenomena are basically unstable, they tend to happen very fast.

The next phase of the operation, called the relaxation phase, starts as C_2 begins to charge up through R_{b2} . The voltage at the node B_2 will start to rise exponentially, and eventually it will reach a value sufficient to bias the base-emitter junction of Q_2 in the forward direction. A new regenerative phase starts. Q_2 starts to conduct and its collector voltage starts to drop. The voltage drop will be passed by C_1 to the base of Q_1 which will now conduct less, its collector will come out of saturation (that is, go positive) and will be passed by C_2 to the base of Q_2 . Q_2 will conduct more heavily and its collector voltage will drop even faster than before. Again the change that set the chain of events in motion has been magnified by the gain of the transistors. Q_2 will eventually go into full conduction (saturation) and Q_1 will be cut off. The appropriate choice of the values of R_{b2} and R_{c2} will ensure that Q_2 goes into saturation driving the base of Q_1 to a negative value equal to the dc supply voltage. C_1 starts to charge up exponentially headed for the positive dc supply voltage, but when it gets to the value required to forward bias the base-emitter junction of Q_1 , Q_1 will start to conduct again and the next regenerative phase is repeated followed by the next relaxation phase *ad infinitum*.

The voltage waveforms at the collector and bases are shown in Figure 7.6(b).

A simple way to synchronize the astable multivibrator to the vertical sync pulse is to couple the leading edge of the sync pulse to the base of one of the two transistors using a *differentiator circuit* to get a sharp clock pulse. In order to get a stable synchronization, the astable multivibrator should operate at a frequency slightly lower than the vertical sync pulse frequency when free-running. The design principles discussed above are best illustrated by an example.

Example 7.2.1 Vertical Deflection Oscillator Design. Design an astable multivibrator with a mark-to-space ratio equal to 15.35/1.34 (see scan and blanking periods in Section 7.2.7) and synchronize it to the leading edge of a 60 Hz square wave derived from the vertical sync pulse of a television receiver. The following are given:

- (1) supply voltage, 12 V dc,
- (2) two NPN silicon transistors, $\beta = 100$ and $V_{be} = 0.7$ V,
- (3) load current = 10 mA.

Solution. Assuming that:

- (a) the collector loads of the transistors are the loads of the oscillator,
- (b) the collector voltage is zero when the transistor is in saturation,

the collector resistors are

$$R_{c1} = R_{c2} = \frac{12}{10 \text{ mA}} = 1.2 \text{ k}\Omega. \quad (7.2.10)$$

The time-constant must be chosen so that one complete cycle takes slightly longer than 1/60 s, for example 17.0 ms. This means that one transistor will be in conduction for 15.6 ms and the other for 1.36 ms.

From Figure 7.6(b), it can be seen that the equation of the voltage on the base of Q_1 is

$$v_{b1} = 2V(1 - e^{-t/\tau}) - V. \quad (7.2.11)$$

The time taken by the base voltage of Q_1 to reach 0.7 V is $t = 15.6$ ms. The time-constant is

$$\tau = 20.71 \text{ ms}. \quad (7.2.12)$$

Because the transistor has a $\beta = 100$, the base current required to cause saturation in each transistor is $10 \text{ mA}/100 = 100 \mu\text{A}$. The base resistor is

$$R_{b1} = R_{b2} = \frac{(12 - 0.7)}{100 \mu\text{A}} = 113 \text{ k}\Omega. \quad (7.2.13)$$

The time-constant $\tau_1 = C_1 R_{b1}$; therefore,

$$C_1 = 0.183 \mu\text{F}. \quad (7.2.14)$$

Similarly, $\tau_2 = 1.79 \text{ ms} = C_2 R_{b2}$, where

$$C_2 = 0.016 \mu\text{F}. \quad (7.2.15)$$

To produce the clock pulse coincident with the leading edge of the square wave, the differentiation circuit shown in Figure 7.7 is used. The condition for approximate differentiation is that the time-constant $CR \ll T$, the period of the input voltage. The diode clips off the unwanted negative-going spike.

The base voltage of Q_1 is now an exponential curve with the positive clock pulse superimposed on it as shown in Figure 7.8. It is clear from the diagram that, when the clock pulse occurs at the time represented by the position a , the oscillator will not synchronize to the clock pulse. However, when the clock pulse occurs in the

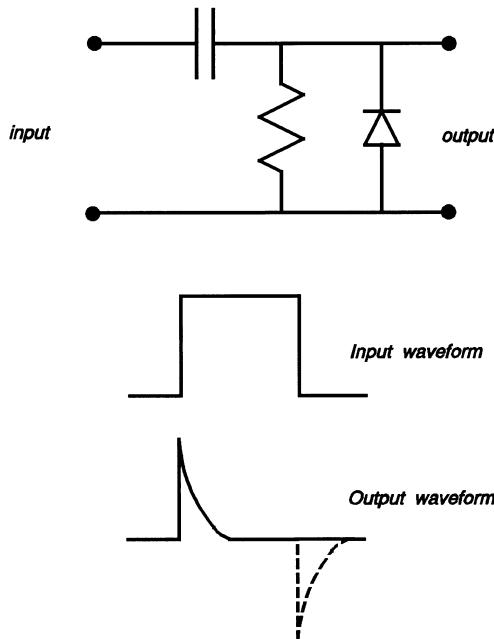


Figure 7.7. The differentiation circuit with its input and output waveforms.

position *b* the voltage at the base of Q_1 will exceed 0.7 V, Q_1 will go into the regenerative mode and the oscillator will be synchronized.

7.2.7.4 Vertical Sweep Current Generator. A very simple sweep current generator with a fairly linear output current with respect to time is shown in Figure 7.9. The circuit also act as a class-A amplifier and drives the yoke coil of the vertical deflection system.

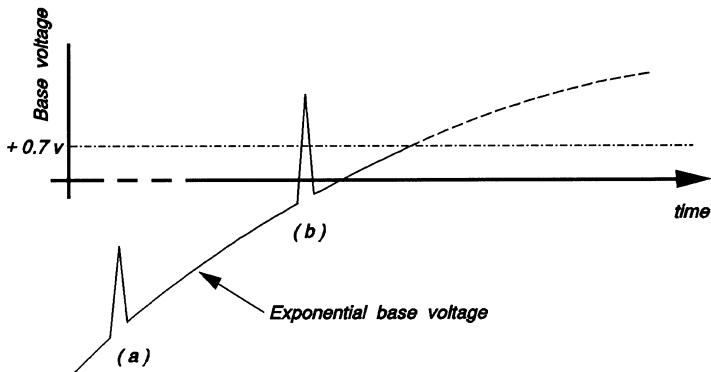


Figure 7.8. The base voltage of the transistor used in the astable multivibrator showing the synchronizing pulses.

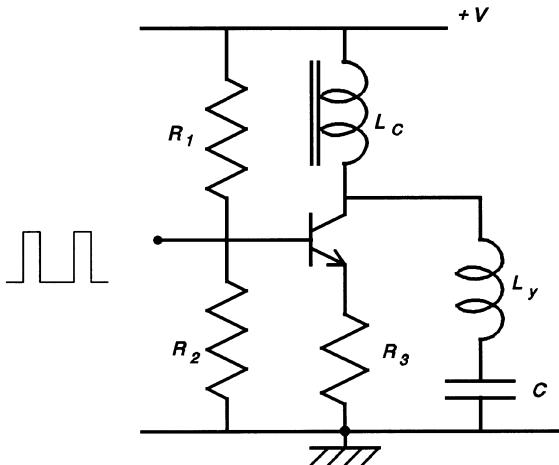


Figure 7.9. The circuit diagram of the driver for the vertical deflection yoke.

The transistor is biased for class-A operation. The collector is connected to the dc power supply by a large inductance with negligible resistance, that is, it is an open-circuit at the frequency of operation but a short-circuit to direct current. When the square wave from the multivibrator is applied to the base of the transistor, the collector current of the transistor will be a square wave. During current transitions a voltage will be developed across L_c ; otherwise; the yoke inductor L_y has the dc power supply connected directly across it. Because the dc voltage is a constant, di/dt in L_y is a constant. The capacitor C serves as a block to dc flowing to ground. Because of the low frequency of operation, this circuit does not consume much power, neither is there a large reactive power circulating in the circuit.

7.2.7.5 Horizontal Deflection System. From Figure 7.3, it can be seen that the horizontal deflection system has the same modules as the vertical. The description of the modules and their design will not be repeated. In practice, the vertical and horizontal deflection systems are different and the differences arise from:

- (1) The vertical system operates at 60 Hz while the horizontal runs at 15.75 kHz,
- (2) Deflection in the horizontal and vertical directions are in the ratio 4 : 3 (aspect ratio),
- (3) The horizontal deflection system is very sensitive to phase modulation and hence the horizontal oscillator has to have an automatic phase control loop to keep phase modulation as low as possible,
- (4) High power (~ 50 W) is required to run the horizontal deflection coil if the energy stored in it is dissipated at the end of each cycle. By using extra circuitry, it is possible to return most of the energy to the dc source by

resonating the inductance of the yoke with a suitable capacitor to cause ringing at approximately 60 kHz,

- (5) The high dc voltage ($\sim 10,000$ V) required to run the picture tube can be obtained by using a transformer coupling (auto-transformer) to step up the voltage followed by a rectifier (tube type). A simplified circuit diagram of the final stage of the horizontal deflection system is shown in Figure 7.10.

The horizontal oscillator is synchronized to the output of the sync separator. The oscillator then drives the base of Q_2 with a string of rectangular pulses at approximately 15.75 kHz. When Q_2 is on, it draws current through the primary of the transformer T_1 . This causes a current to flow in the secondary, biasing Q_1 on and putting it into saturation. The dc supply voltage V is therefore connected directly across the primary terminal P of the auto-transformer and ground. Q_1 and the winding of the autotransformer between node P and ground form an emitter follower. The horizontal deflection yoke inductance, L_x , which is connected to the auto-transformer at node N , has a constant voltage, higher than V , connected across it. The current that flows in L_x is the linear ramp required to produce a linear horizontal deflection. The capacitor C_c blocks dc from flowing in L_x , and it is a short-circuit at the frequency of operation. At the end of the linear ramp, Q_2 is cut off, a voltage is induced in the secondary of T_1 , and this is in the appropriate direction to shut off Q_1 abruptly. Normally, the sudden termination of the current flow in the auto-transformer and L_x will generate a huge voltage spike in the attempt to dissipate the energy stored in the inductances. However, the diode D_1 conducts

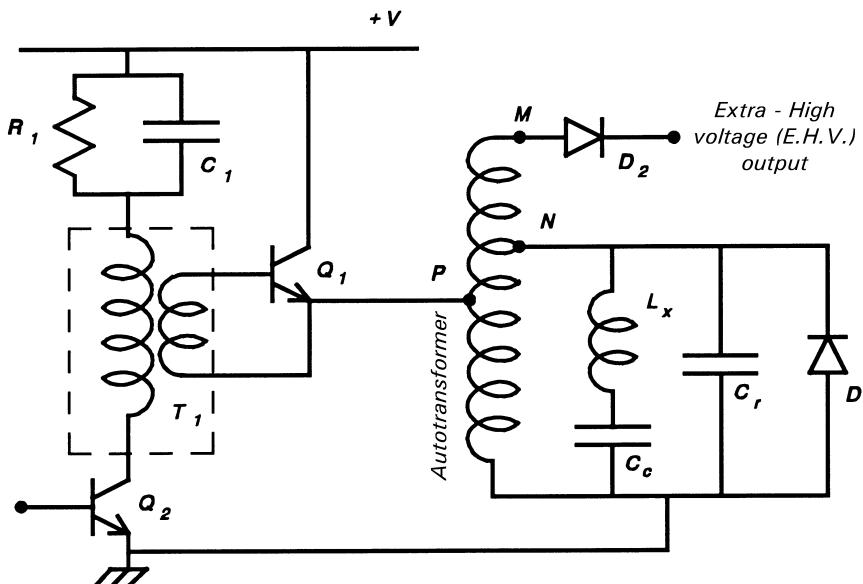


Figure 7.10. The driver of the horizontal deflection yoke with the extra-high voltage (EHV) generator required for the anode of the picture tube.

providing a path for the current to flow into the tuned circuit made up of L_x and C_r . L_x and C_r resonate ("ring") at approximately 60 kHz (four times the horizontal deflection frequency). If the timing is set up correctly, the energy stored in the resonant circuit will be flowing from the capacitor into the yoke just as the next current ramp is starting. This arrangement substantially reduces the amount of power taken from the dc supply to drive the yoke. As no diode is connected across the $M-N$ portion of the auto-transformer, a large voltage spike appears across it. This is rectified and used to bias the anode of the picture tube.

7.2.8 Picture Tube

The picture tube is an example of a cathode ray tube (CRT) used in a wide variety of display systems. It consists of a glass cylinder which flares into a cone with a nearly flat base. Figure 7.11 shows a cross-section of the typical CRT used for the display of television images.

At the end of the glass cylinder, there is an *electron gun*. The electron gun is made up of a cylinder, closed at one end and coated with oxides of barium and strontium which give off electrons when heated. The CRT has control grid(s), focussing and deflection electrodes or coils. An electric current is passed through the heater coil to bring the temperature of the *cathode* to the appropriate value, electrons are given off and, under the influence of a positive voltage placed on the anode, the electrons travel down the tube and strike the *screen* at high velocity. The grid(s) carry a negative voltage which can be varied to control the number of electrons that can leave the cathode and eventually reach the screen. The grid(s) therefore control the intensity of the light forming the image. The *electron lens* is either an electrostatic or a magnetostatic means of focusing the electrons, usually to produce a small spot on the screen.

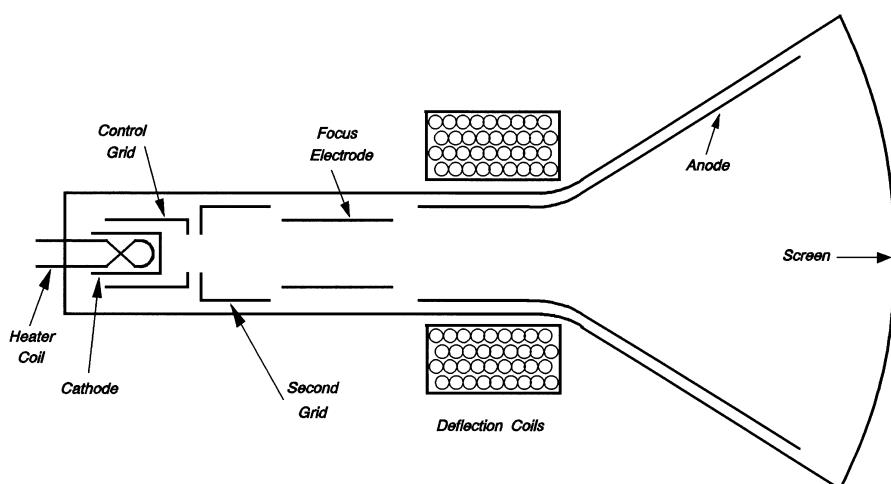


Figure 7.11. A cross-sectional view of the cathode ray tube (CRT). Reprinted with permission from T. Soller, M. A. Starr and G. E. Valley, *Cathode Ray Tube Displays*, Boston Tech. Pub. Lexington, MA, 1964.

The rest of the tube is made up of the *deflection electrodes* or *coils* which cause the electron beam to move on the screen on the application of the appropriate voltage or current. The anode is a conductive thin film on the inside wall of the flare as shown and requires several thousand volts positive for proper operation. The screen is coating with silicates, sulfides, fluorides and alkali halides which emit light when bombarded by electrons. The color of the light emitted is determined by the characteristic wavelength associated with each chemical element present in the screen material. For example, copper-activated 85% zinc sulfide and 15% cadmium sulfide gives off yellow light while 93% zinc sulfide and 7% cadmium sulfide produces a blue-green phosphor. The persistence (or afterglow) of the phosphor can be varied to suit different conditions.

Most modern CRTs used in television receivers have an electrostatic focussing system because it is simpler and less expensive to manufacture. The deflection system is, however, magnetostatic because of the tendency to have shorter tubes and larger screen areas.

7.3 COLOR TELEVISION RECEIVER

7.3.1 Demodulation and Matrixing

In Section 6.3.8, it was established that black-and-white and color television signals had many common features. Their basic differences are:

- (1) in addition to the luminance information (L_w), the chrominance information ($L_r - L_w$) and ($L_b - L_w$) is transmitted using a quadrature modulation (DSB-SC) scheme with a (color) subcarrier of frequency approximately 3.58 MHz,
- (2) on the back porch of the horizontal blanking pulse, an eight-cycle bursts of the color subcarrier is superimposed.

One scheme used for the recovery of the original red, blue, and green signals is shown in Figure 7.12. It is described as *pre-picture tube matrixing*.

From the output of the video detector, the color burst separator picks up the eight cycles of the 3.58 MHz subcarrier and passes them onto the color subcarrier regenerator. The regeneration circuit is essentially a very high Q -factor circuit tuned to 3.58 MHz. The eight-cycle burst of signal causes the high Q circuit to go into a free-running oscillation mode. The subcarrier must have limited variation in both amplitude and phase, for proper synchronous detection, and therefore the decay of the amplitude during the free-running period must be controlled. To maintain the amplitude variation to 90% of the initial value for the time required to sweep one horizontal line (63 µs) a Q -factor of 7000 is required. Crystals are used in the regenerator because only crystals have the necessary high Q -factor.

The regenerated 3.58 MHz signal is split into two branches. The first branch goes through a 90° phase shifter to produce the quadrature signal required for the synchronous demodulation of the ($L_r - L_w$) and ($L_b - L_w$) signals contained in the output of the color bandwidth filter. After demodulation, the matrix reproduces

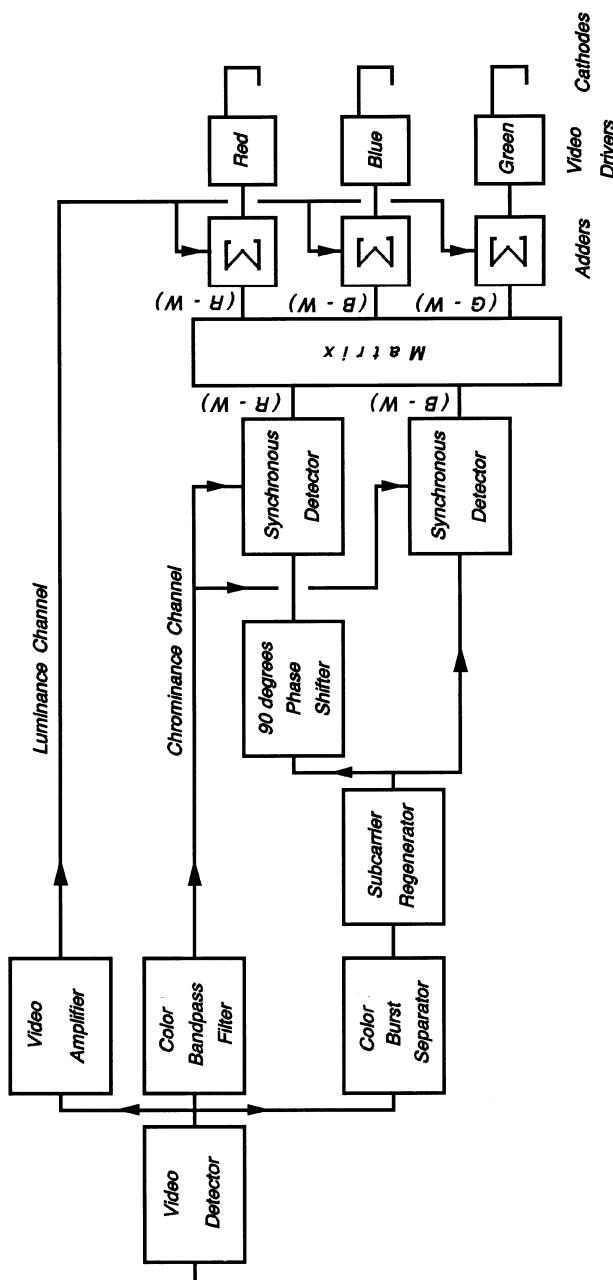


Figure 7.12. The schematic diagram for color separating in the receiver. This is described as pre-picture tube matrix because the colors are separated outside the tube.

the third difference signal, namely $(L_g - L_w)$. The composite video signal containing the luminance information L_w is now combined in the adders to give the original red, green, and blue information. After suitable amplification, they drive the picture tube cathodes. A variation on this scheme, called *picture tube matrixing*, produces the $(L_g - L_w)$ signal and the addition of the L_w is carried out in the picture tube by driving the grids and the cathodes differentially. This is shown in Figure 7.13.

When the I and Q signal scheme is used in the transmission, it must be decoded into $(L_r - L_w)$ and $(L_b - L_w)$ after the synchronous detectors and before the signal is applied to the matrix.

7.3.2 Component Circuit Design

The 90° phase shift circuit was discussed in Sections 4.4.3.1 and 4.4.3.2. The synchronous detector (four-quadrant multiplier) was discussed in Sections 2.6.4 and 4.4.3.3. The design of the adder was described in Section 4.4.3.5. Video amplifier design was discussed in Section 6.3.4.2.

7.3.2.1 Color Burst Separator. The eight cycles of the color burst subcarrier riding on the back of the horizontal blanking pulse are shown in Figure 6.24. Since this signal is required for the synchronous detection of the chrominance signal, no part of the video signal must be allowed to get through the separator. The most popular technique is to use the falling edge of the horizontal sync pulse to trigger a gate open and to close the gate immediately after the time it takes the eight cycles of 3.58 MHz to come through (2.23 μ s). The circuit used is shown in Figure 7.14.

Q_2 is an amplifier biased in the off state by the resistors R_3 , R_4 , R_5 , and R_6 . Its emitter is coupled through the tuned circuit LC to the collector of Q_1 . The base of Q_1 is driven by the output from the horizontal oscillator (see Figure 7.3). The resonant frequency of the LC circuit is adjusted so that the rise-time of the rectangular waveform from Q_1 is delayed long enough to let only the eight cycles of the 3.58 MHz signal through to the collector of Q_2 .

7.3.2.2 Color Subcarrier Regenerator. A typical circuit of the color subcarrier regenerator is shown in Figure 7.15.

The output from the transistor Q_1 is taken from both the collector and the emitter. C_2 is adjusted to cancel the shunt capacitance of the crystal (see Figure 2.18). The crystal is chosen so that its series resonant frequency is 3.58 MHz and it resonates when it is excited by the subcarrier. In the series mode resonance, the crystal impedance is essentially a short-circuit (ideal voltage source). The losses in R_e and R_3 in parallel with R_4 are therefore insignificant. The circuit has a high enough Q -factor to perform about 217 cycles of free-running oscillations between bursts of color subcarrier signals. Q_2 is used as a buffer amplifier. In some earlier models, the signal from the regenerator was used to synchronize an LC oscillator whose output was then used for the demodulation. This is an expensive way to solve the problem in view of the satisfactory performance of the free-running system.

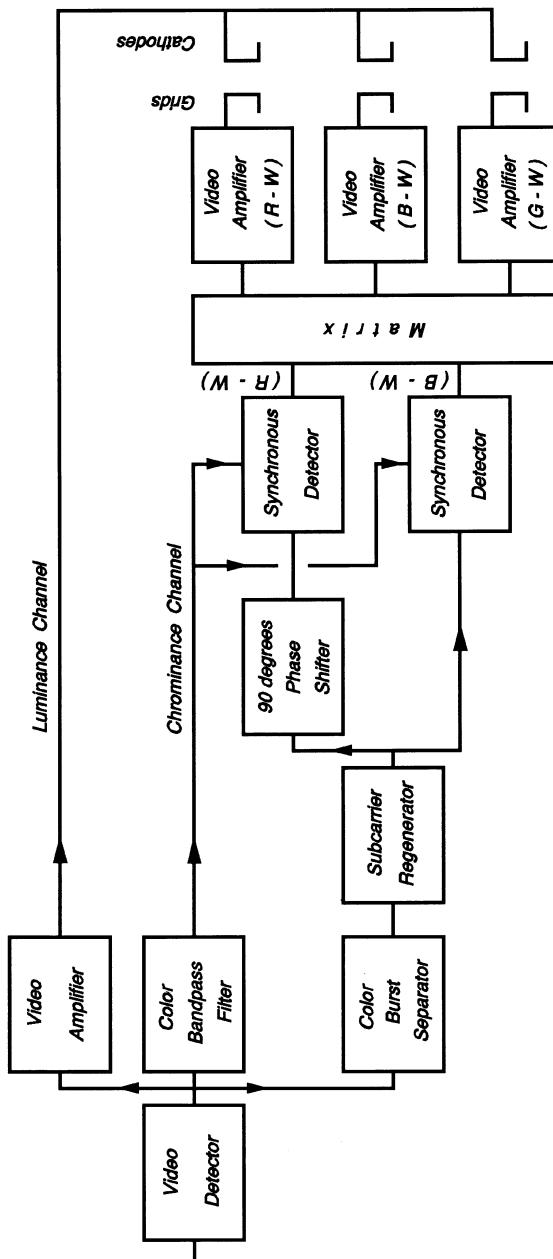


Figure 7.13. The block diagram for color separation in the tube. Note that the color differences are applied to the tube directly.

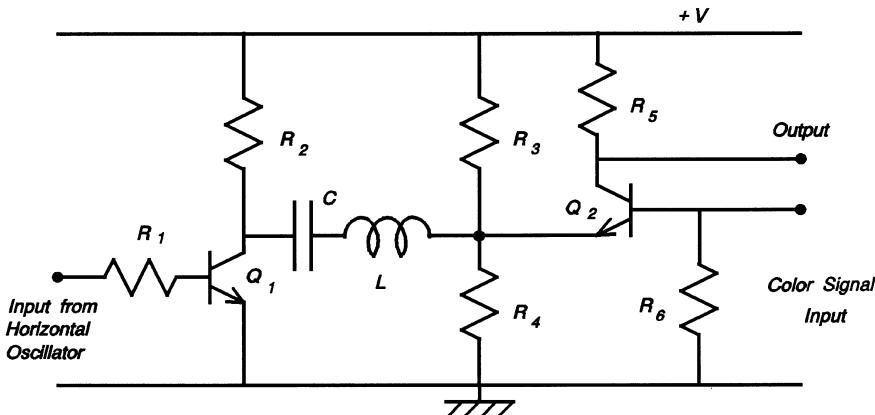


Figure 7.14. The color subcarrier burst separator. The output is used to synchronize the color subcarrier regenerator.

7.3.2.3 The Matrix. The function of the matrix is to reproduce the third difference signal ($L_g - L_w$). Equation (6.4.4) was derived for this operation and it is repeated here.

$$(L_g - L_w) = -\frac{0.30}{0.59}(L_r - L_w) - \frac{0.11}{0.59}(L_b - L_w). \quad (7.3.1)$$

It is clear that this can be achieved with a simple resistive network and a phase reversing buffer amplifier.

7.3.3 Color Picture Tube

In Section 7.2.8, it was stated that different phosphor materials coatings on the screen of the picture tube give different colors of light when bombarded by electrons. A number of different techniques for producing images in color on the CRT exist; all of them rely on this phenomenon. The most successful of these uses the *shadow mask* technique. Figure 7.16 illustrates the basic concept.

Three electron guns representing the three primary colors are mounted symmetrically on a plane perpendicular to the axis of the tube. Electrons from each gun will therefore bombard different areas of the screen after going through the holes in the *shadow mask*. If the area of the screen corresponding to the blue gun is coated with a phosphor that produces blue light and so on, then it is clear that, by depositing the different phosphor materials in dots grouped in threes at the appropriate points on the screen, the different colors can be reproduced by modulating the relative saturation of each component.

The shadow mask technique has the following disadvantages:

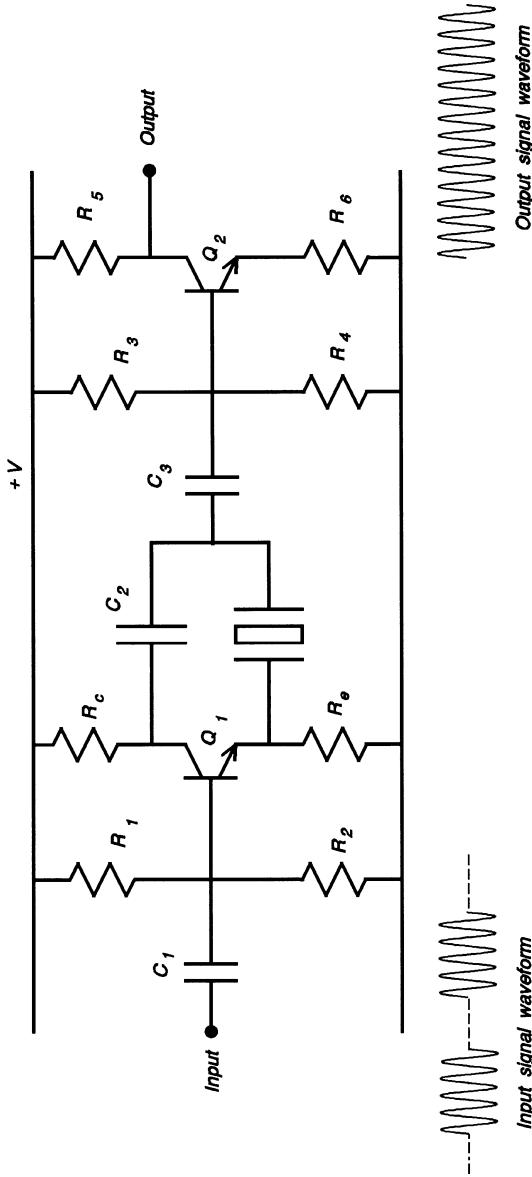


Figure 7.15. The color subcarrier regenerator. Note that, because of the high Q factor of the crystal in the circuit, it is virtually free-running between bursts of the subcarrier.

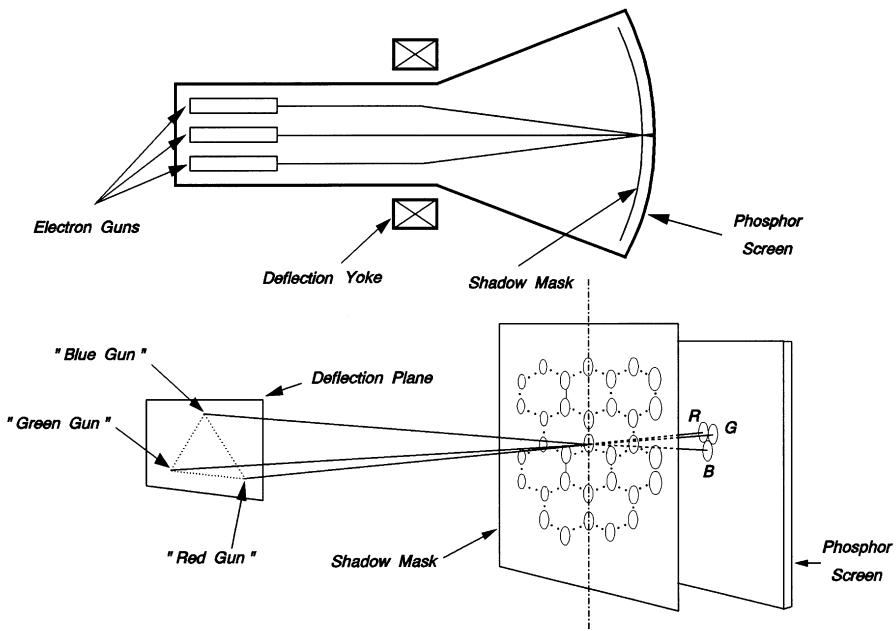


Figure 7.16. The color picture tube showing details of the shadow mask. Reprinted with permission from H. B. Law, "Shadow Mask", Proc. IRE (now IEEE), 39(10), 1951.

- (1) The shadow mask has to be very carefully aligned for good color resolution.
- (2) The presence of the mask reduces the average brightness (luminance) of the image but this can be compensated for by increasing the electron emission of the guns.

7.4 HIGH-DEFINITION TELEVISION (HDTV)

High-definition television is an attempt to produce images of quality comparable to that of a 35 mm motion picture film. The proposed system has 1125 lines of scan with 60 fields per second. Japanese HDTV uses analog techniques and requires a bandwidth of 20–25 MHz. In North America, HDTV uses digital technology which permits the compression of the data to fit into a 6 MHz bandwidth. The standard for digital television has the acronym MPEG-2 and it can achieve a bit compression ratio of 55 : 1. It does this by retaining most of the data and presenting only changes in a picture frame. Despite the improvement in picture definition achieved by HDTV the viewing public has been slow to switch over to the new format. The high price tag for the HDTV receiver means that only a very select number of viewers are willing to adopt the new system. Equally, the broadcasting organizations are unwilling to invest large amounts of money in HDTV to re-equip their studios

TABLE 7.1

	NTSC	HDTV
Technology	Analog	Digital
Bandwidth	6 MHz	6 MHz
Total no. of lines	525	1125
Actual no. of lines	486	1080
Aspect ratio	4×3	16×9
Max. resolution	720×486	1920×1080
Sound	2 channels (stereo)	5.1 ² channels

²The 0.1 channel refers to a subwoofer for bass “sounds you can feel”.

when they are not sure of reasonable returns on their investment. Table 7.1 compares the current standard (NTSC) to HDTV.

BIBLIOGRAPHY

1. Kennedy, G., *Electronic Communication Systems*, McGraw-Hill, New York, 1970.
2. Smith, J., *Modern Communication Circuits*, McGraw-Hill, New York, 1986.
3. Parker, N. W., “Television Image-Reproducing Equipment”, in *Electronics Engineers’ Handbook*, Fink, D. G. and Christiansen, D. (Eds), 2nd Ed., McGraw-Hill, New York, 1982.
4. Sherr, S., *Fundamentals of Display System Design*, Wiley-Interscience, New York, 1970.
5. Soller, T., Starr, M. A. and Valley, G. E., *Cathode Ray Tube Displays*, Boston Tech. Publishers, Inc., Lexington, MA, 1964.
6. Law, H.B., “Shadow Mask”, *Proc. IRE*, 39(10), 1187, 1951.

PROBLEMS

- 7.1** Give two advantages of vestigial sideband modulation systems. What are their disadvantages? A transmission consists of a carrier and its upper sideband. Show that, when the modulation index is small compared to unity, an ordinary envelope detector can be used to demodulate it. Calculate the signal power in a vestigial sideband signal if the peak value of the carrier voltage is 5 V when the load resistor is $2.5 \text{ k}\Omega$ and the modulation index is 0.15. You may assume that the diode is ideal.
- 7.2** A television receiver is tuned to receive a transmission with a carrier frequency 67.25 MHz (Channel 4). If the intermediate frequency is 45.75 MHz, what is the possible frequency(ies) at the local oscillator terminals of the mixer? If there are two possible frequencies, does one of them offer an advantage over the other? The local oscillator signal for the television receiver is generated at a lower frequency and multiplied by a factor of 512. Calculate the frequency of the oscillator and design a Hartley-type oscillator given that the tuning

capacitor has the value 144 pF. Discuss the precautions you would take to minimize changes in the frequency of oscillation.

- 7.3** Using suitable block diagrams and derived equations, explain how quadrature multiplex is used to transmit and receive the chrominance information in a color TV system. State the conditions which must be satisfied for demultiplexing to be possible. How are these conditions met in the domestic TV receiver?
- 7.4** Describe the construction and operation of a color TV picture tube. Explain why a color tube requires a higher (EHV) dc voltage than the monochrome for equal luminance.
- 7.5** Describe with the help of a suitable diagram the operation of a color subcarrier regenerator.

Show that for a series-tuned *RLC* circuit at resonance:

- (1) The total energy stored in the circuit is a constant,
- (2)

$$Q_0 = 2\pi \frac{\text{Energy stored}}{\text{Energy dissipated per cycle}}. \quad (\text{P7.5})$$

In a crystal-controlled color subcarrier regenerator operating at 3.58 MHz, the total energy remaining in the circuit after 218 cycles of free-running oscillation is 90% of its original value. Calculate the peak value of the voltage across the capacitor at the end of the free-running cycle and estimate the Q_0 of the circuit.

- 7.6** What is a video signal? Explain why a video amplifier has to be used in conjunction with a television camera tube.

A television system has the following features:

- (1) 625 lines per frame with interlacing,
- (2) 25 frames per second,
- (3) 5% of the trace period is used for the retrace,
- (4) the aspect ratio is 4 : 3.

Assuming equal resolution on the vertical and horizontal axes, calculate the bandwidth required to handle the signal.

8

THE TELEPHONE NETWORK

8.1 INTRODUCTION

The early history of the telephone system has been outlined in Chapter 1. The growth of the telephone system has been truly phenomenal and forecasts show a continuing growth as new services such as data transfer, facsimile and mobile telephone are added.

The telephone differs from the broadcasting system in two basic ways:

- (1) In broadcasting, a few people who, in theory, have information send it out to the many who are presumed to want the information; it is one-way traffic. The communication link provided by the telephone is two-way traffic.
- (2) The basic idea of broadcasting is to make the message available to anyone who has the equipment and the interest to tune in. This is in contrast to the norm in the telephone system where the privacy of the message is guaranteed by law.

Because of these differences, the two systems handle very different types of information – public versus private – and their patterns of development have been different.

8.2 TECHNICAL ORGANIZATION

For a telephone system to work, there must be a minimum of two people who wish to communicate. It is then possible to install the circuit shown in Figure 8.1.

This would be quite adequate except for the fact that these two people may not want to talk to each other all the time and therefore some additional system has to be set up for either person to indicate to the other that they wish to talk. What was

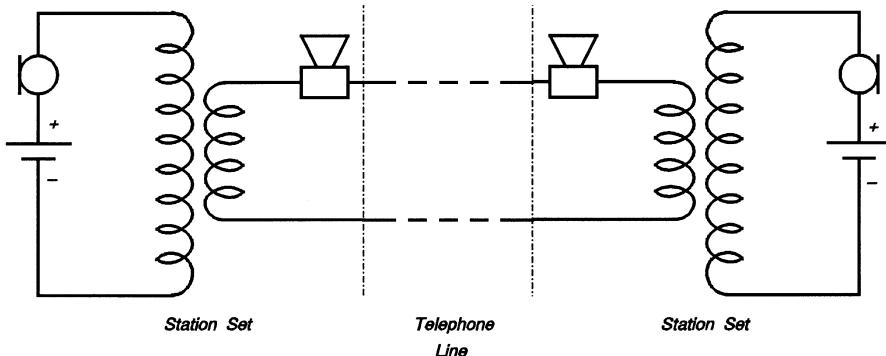


Figure 8.1. The basic elemental telephone system.

added was a bell on the called party's premises which can be rung from the calling party's premises.

Presumably the success of this prototypical communication system would soon attract the attention of other people who would want to set up similar systems. It is clear that soon the situation depicted in Figure 8.2 would develop where every subscriber would have to be wired up to every other subscriber. That would be prohibitively expensive and quite impractical.

Evidently, the way to deal with the situation is to connect every subscriber to a central location and arrange to have an attendant to interconnect the various subscribers in whatever combination that is required. That central location is, of course, the *central office* which has and continues to have a central role in the telephone system. The system configuration would then be as shown in Figure 8.3.

Assuming that the system has six subscribers then each subscriber has access to the other five subscribers. But in the meantime, a group, also of six, in the next town have heard of the success of the system and have set up a similar system of their

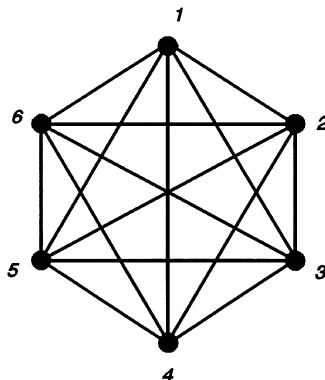


Figure 8.2. The connection diagram for six subscribers showing all the 15 possible telephone lines.

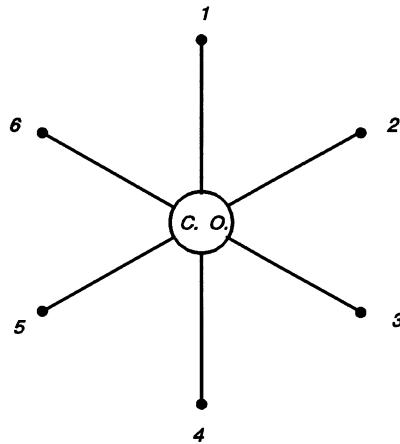


Figure 8.3. The concept of a central office reduces the number of lines to six.

own. Now there is a possibility of reaching eleven other subscribers if a connection can be made between the two systems. This brings up two very important points:

- (1) The greater the number of people on the communication network, the more attractive it is for other people to join.
- (2) There has to be some level of compatibility between the two systems.

The system would have evolved, as shown in Figure 8.4.

Continuing with the story, the distance between the two towns is quite long and the initial cost and upkeep are high but if this line can be made to carry more than one conversation simultaneously, the cost per conversation will be substantially

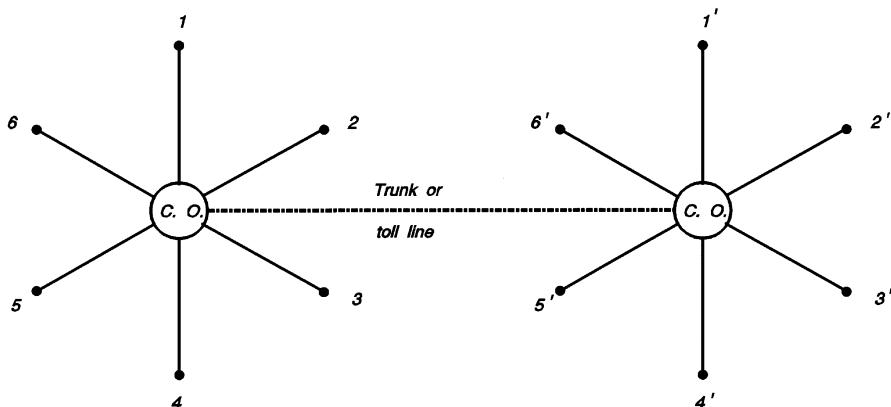


Figure 8.4. The connection of a trunk or toll line between two central offices increases the number of subscribers that can be reached from 5 to 11.

reduced. It is also very likely that because of the length of the line, the quality and the reliability of the system may be degraded. This brings up an important point:

The greater the number of communication channels that can be established over the same link, the less the cost per message. Multiplex and the conservation of bandwidth will become goals of several generations of communications engineers.

The telephone system that started with people talking to each other has acquired more than people for a clientele. Increasingly, the network is being used to supply services to machines such as computers, facsimile devices, security guard services and access to the Internet.

8.3 BASIC TELEPHONE EQUIPMENT

The basic telephone has surprisingly very few parts. These are shown in Figure 8.1. When the microphone is connected in series with the battery, it produces a current proportional to the pressure of the sound impinging on it. The transformer eliminates the dc and sends the ac portion of the current through the line. The earphone at the receiving end changes the variation of the current into sound. Obviously, the system works in the reverse direction.

8.3.1 Carbon Microphone

Figure 8.5 shows a cross-section of the carbon microphone [1].

It has a light-weight aluminum cone with a flexible support around the periphery so that it will deflect (vibrate) due to the changing sound pressure level. Attached to the apex is a disc which acts as a piston when the cone deflects. A plastic housing with an electrode attached to the bottom contains a loose pile of carbon granules. When the pressure on the cone is increased, the carbon granules become

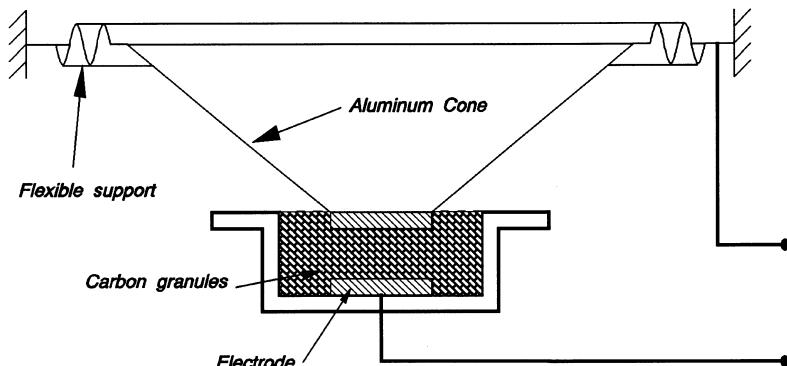


Figure 8.5. A cross-sectional view of the carbon microphone.

compressed, the resistance goes down and more current flows. The opposite happens when the pressure is released.

Assuming that the sound pressure level on the carbon microphone is a sinusoid then the resistance of the device is

$$r(t) = r_0(1 + k \sin \omega t) \quad (8.3.1)$$

where r_0 is the mean resistance, k is a coefficient less than unity, and ω is the frequency of the sound pressure.

When the microphone is connected to a battery of electromotive force E (volts) in series with a load R , as shown in Figure 8.6, we have

$$I = \frac{E}{R + r_0(1 + k \sin \omega t)} \quad (8.3.2)$$

$$I = \frac{E}{R + r_0} \frac{1}{1 + \frac{kr_0}{R + r_0} \sin \omega t} \quad (8.3.3)$$

$$I = \frac{E}{R + r_0} \left(1 + k \frac{r_0}{R + r_0} \sin \omega t\right)^{-1}. \quad (8.3.4)$$

Using the binomial expansion,

$$I = \frac{E}{R + r_0} [1 - \left(\frac{kr_0}{R + r_0}\right) \sin \omega t + \left(\frac{kr_0}{R + r_0}\right)^2 \sin^2 \omega t - \dots]. \quad (8.3.5)$$

Let

$$I_0 = \frac{E}{R + r_0} \quad (8.3.6)$$

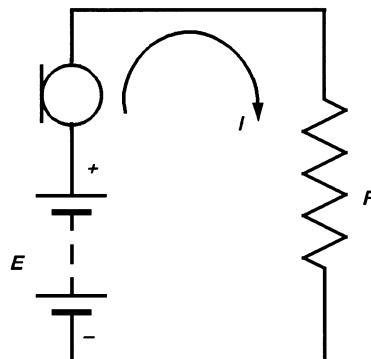


Figure 8.6. The carbon microphone with dc supply E and load resistance R .

then

$$I = I_0 - I_0 \left(\frac{kr_0}{R + r_0} \right) \sin \omega t + \frac{I_0}{2} \left(\frac{kr_0}{R + r_0} \right)^2 - \frac{I_0}{2} \left(\frac{kr_0}{R + r_0} \right)^2 \cos 2\omega t + \dots \quad (8.3.7)$$

Because $kr_0/(R + r_0)$ is smaller than unity higher order terms can be ignored. If it is desirable to reduce second harmonic distortion, $kr_0/(R + r_0)$ can be reduced, but in doing so the amplitude of the fundamental will be reduced as well. A compromise between distortion and signal amplitude has to be made.

The carbon microphone has the following attractive properties:

- (1) It is simple and therefore inexpensive to manufacture.
- (2) It is robust; it is not likely to need attention even in the hands of the public.
- (3) It acts as a power amplifier; under normal bias conditions, (the electrical power output far exceeds the acoustic power input. It does not normally require additional amplification.
- (4) Its input–output characteristics are shown in Figure 8.7. The non-linearity at low input levels helps to suppress background noise and that at high levels acts as an automatic gain control.

8.3.2 Moving-Iron Telephone Receiver

A cross-section of the moving-iron telephone receiver is shown in Figure 8.8. It consists of a U-shaped permanent magnet that carries a coil as shown. In front of the open face of the U, a thin cobalt iron diaphragm, is held by an annular ring support with a short distance between them. With no current in the coil, the diaphragm has a fixed deflection towards the magnet.

The signal current is passed through the coil and, assuming that it is sinusoidal, then for one-half of the cycle the flux generated by the current will aid the pull of the

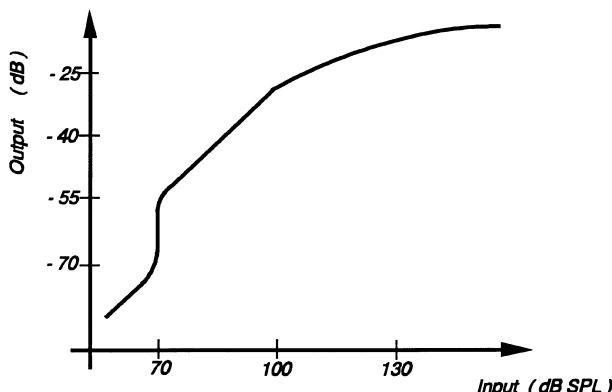


Figure 8.7. The input–output characteristics of the carbon microphone.

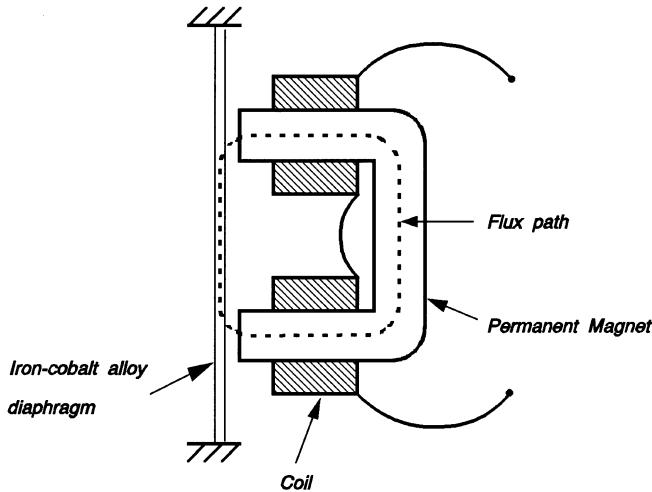


Figure 8.8. A cross-sectional view of the moving-iron telephone receiver.

permanent magnet on the diaphragm and it will deflect accordingly. During the other half of the cycle, the coil flux will oppose that of the magnet and the diaphragm will deflect much less.

The force between two magnetized surfaces is given by

$$F = \frac{B^2}{2\mu_0} \quad (\text{N/m}^2) \quad (8.3.8)$$

where B is the flux density in teslas (T) and μ_0 is the permeability of free space, that is, $4\pi \times 10^{-7}$. Let A be the area of the pole face (m^2), B_0 the flux density due to the permanent magnet (T), and $b_0 \sin \omega t$ the flux density due to the current (T).

The force in newtons is then

$$F = \frac{2A}{2\mu_0} (B_0 + b_0 \sin \omega t)^2 \quad (8.3.9)$$

$$\frac{A}{\mu_0} (B_0^2 + 2B_0 b_0 \sin \omega t + b_0^2 \sin^2 \omega t) \quad (8.3.10)$$

$$F = \frac{A}{\mu_0} (B_0^2 + \frac{1}{2} b_0^2 + 2B_0 b_0 \sin \omega t - \frac{1}{2} b_0^2 \cos 2\omega t) \quad (8.3.11)$$

The second harmonic component can be reduced by making B_0 large compared to b_0 . This will increase the direct component of the force, which is likely to cause the diaphragm to touch the magnet. Note that when B_0 is zero (no permanent magnet), the device produces only the second harmonic. This is to be expected since both the

positive and negative halves of the sinusoid will exert an equal force of attraction on the diaphragm.

8.3.3 Local Battery – Central Power Supply

The system as depicted in Figure 8.1 is powered from batteries that are located on the customer's premises. The batteries are of interest because they are a hazard to the customer and they pose a very difficult problem for the maintenance staff. Furthermore, their reliability is questionable because of their location among other considerations.

The solution to the problem is to have a common power supply located at the central office (out of the way of the telephone subscriber) and readily available to the maintenance personnel. The reliability of the service can then be improved by installing a backup power supply. The scheme for achieving this end is illustrated in Figure 8.9.

The central office battery in series with two inductors is connected to the lines of the calling and called party as shown.

The inductors have a high inductance and therefore appear to be open-circuits at audio frequency but short-circuits at dc. Every call requires two such inductors to complete the connection.

8.3.4 Signalling System

The signalling system consisted of a magneto and a bell which responded to high ac voltage input. The magneto was a hand-operated alternator whose flux was produced by a permanent magnet. The calling party turned the crank to produce about 100 V ac. The current travelled down the telephone line and caused the bell at the called party's end to ring. To avoid damage to the telephone receiver and to conserve

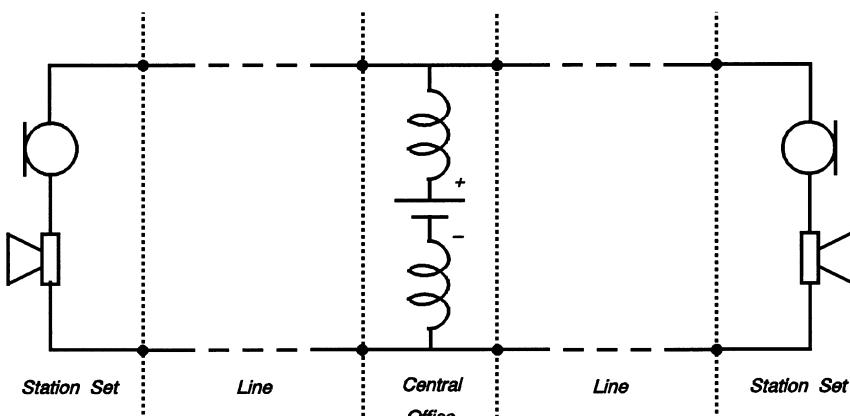


Figure 8.9. The local batteries are replaced by a central power supply.

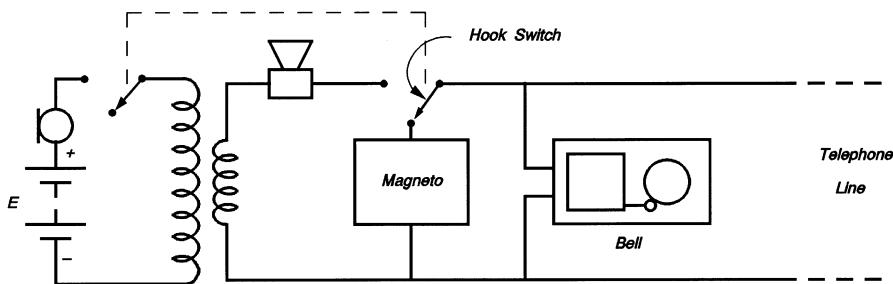


Figure 8.10. The elemental telephone with signalling devices (magneto and bell) shown.

battery power, the hook switch was disconnected them from the line when the telephone was not in use. A simplified diagram of the signalling system is shown in Figure 8.10.

8.3.5 The Telephone Line

Physically, the telephone line consists of a pair of copper wires supported on glass or porcelain insulators mounted on wooden poles. Electrically, an infinitesimally short piece of line can be modelled as shown in Figure 8.11 [2,3]. The elemental series resistance and inductance are represented by δR and δL and the elemental shunt capacitance and conductance are represented by δC and δG , respectively.

The analysis of the model is beyond the scope of this book. However, the analysis shows that the telephone line, at voice frequencies, can be approximated by an RC

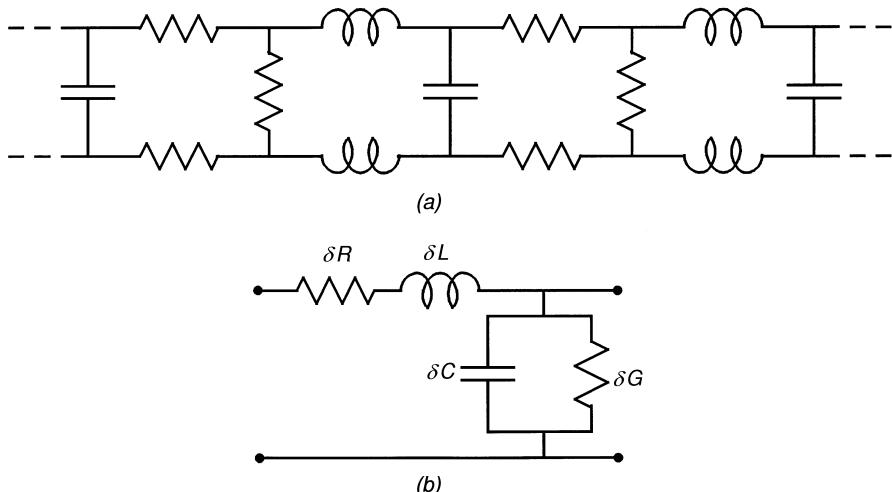


Figure 8.11. (a) The equivalent circuit of the telephone line showing series resistance R and inductance L and shunt capacitance C and conductance G . (b) An elemental equivalent circuit of the telephone line.

low-pass filter whose cut-off frequency is a function of its length. The longer the line is, the lower the cut-off frequency. The frequency response of a typical telephone line is shown in Figure 12.1.

8.3.6 Performance Improvements

From Figure 8.9 it can be seen that the dc required to power the carbon microphone has to flow through the receiver. This is not a good idea since it will make B_0 , the flux density of the permanent magnet (see Equation (8.3.11)), larger or smaller than it should be. A second disadvantage is that all the ac current generated by the carbon microphone has to flow through the receiver. This produces a very loud reproduction of the speaker's own voice in her receiver. The psychological effect is that the speaker lowers her voice, making it difficult for her listener to hear what she is saying. This phenomenon is called sidetone. The two problems can be solved by using the circuit shown in Figure 8.12. It is an example of a hybrid.

8.3.6.1 The Hybrid. The carbon microphone is connected to the centre-tap of the primary of the transformer. One end is connected to the telephone line and the other to an RC network which approximates the impedance of the line. The secondary is connected to the receiver. There is still a path for dc from the central office battery to flow through the carbon microphone.

In the transmit mode, the ac produced by the microphone divides up equally, with one half flowing through the telephone line and the other half in the line-matching impedance. Since these currents are in opposite directions in the primary of the transformer, no net voltage appears across the secondary. The speaker cannot hear himself. In the receive mode, the current I_R flows through the first half of the primary winding and then splits at node X with I_m flowing through the carbon microphone where the energy is safely dissipated. The remainder ($I_R - I_m$) flows through the second half of the primary into the line-matching impedance. This time, the

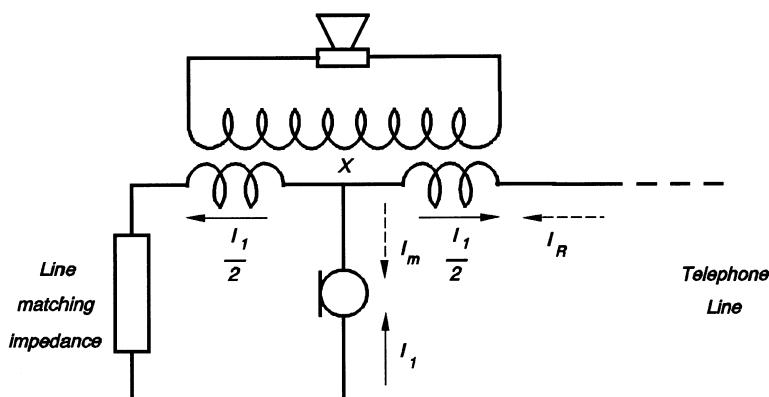


Figure 8.12. The use of a hybrid transformer to control sidetone.

directions of the two currents in the primary are the same; a net voltage appears across the receiver.

In practice, the level of sidetone fed back to the speaker has to be carefully controlled. When it is too low, the telephone appears dead to the speaker and her normal reaction is to raise her voice. When the sidetone is too high, it has the opposite effect.

8.3.6.2 The Rotary Dial. The rotary dial came with the invention of the automatic central office. Evidently, the automatic central office offered several advantages over the manual office. There was increased security of the messages since there was no human interface in setting up calls. The time for setting up and releasing a call was substantially reduced and the probability of operator errors decreased. It guaranteed 24-hour service with fewer more highly trained personnel.

The dial is simply a method of issuing instructions to the central office and it does this by producing a binary coded message by mechanically opening and closing a switch in series with the circuit. The basic dial is as shown in Figure 8.13. It has a finger wheel with ten finger holes and it is mechanically coupled by a shaft to a second wheel which has ten cam lobes as shown. The shaft is mounted so that both wheels can rotate about the axis. The wheel assembly is spring loaded so that, when it is rotated in a clockwise direction, it will return at a constant speed under the control of a mechanical governor.

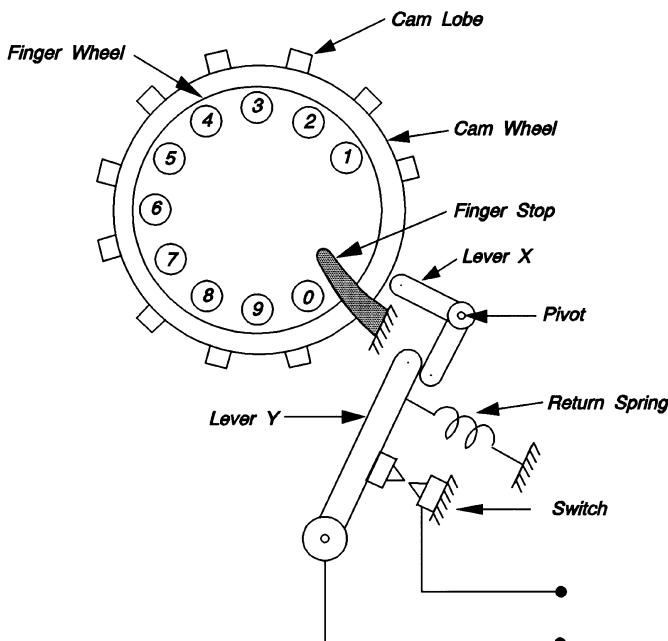


Figure 8.13. The essential features and operation of the rotary dial.

To operate the dial, the caller inserts his or her index finger into the hole corresponding to the number and pulls the finger wheel to the finger stop and then releases it. While the finger wheel is rotating in the clockwise direction, the lever *X* is free to move out of the way of the cam lobes without disturbing the switch lever *Y*. When the wheel assembly is rotating in the counter-clockwise direction, every cam lobe that passes *X* will cause the switch lever *Y* to open the switch. If current is flowing through the switch, the current flow will be disrupted the number of times corresponding to the number of the finger hole. The current pulses can be used to operate a device (to be discussed later) at the central office to effect the required connection. The return spring, cam lobes and mechanical governor are designed to produce 10 pulses per second with approximately equal mark-to-space ratio. Since it is bound to take the subscriber much longer than 1/10 seconds to rotate the finger wheel again, a pause longer than 1/10 seconds can be recognized by the central office as an inter-digit pause. It is then possible to send a second and subsequent string of pulses to effect a connection which requires a multi-digit code. Note that when the digit "0" is dialled, ten pulses are produced.

8.3.6.3 Telephone Bell. The telephone bell has two brass gongs with a clapper which is operated by an electromagnet. It is mechanically and electrically tuned to respond optimally (resonance) to current at 20 Hz. It is left connected to the telephone line at all times but the high impedance of its electromagnet coil ensures minimal effect at voice frequencies. Also the 10 Hz pulse from the rotary dial has no significant effect on it. Nominally, it operates on 88 V, 20 Hz ac supplied to it from the central office in the ring-mode.

8.3.7 Telephone Component Variation

The telephone components described in this section are meant to be a representative sample of what can be found within the territory of any telephone operating authority or company. For each component there are several possible variations, some made to get around patents rights granted to others, and some to lower cost and improve reliability.

The subscriber telephone instrument has changed in its physical appearance and electrical characteristics since it was first put into service. However, in broad terms, it remained basically the same until the introduction of electronics in the form of semiconductor devices. The availability of amplifiers at very low cost offered various options such as new microphones, electronic sidetone control, tone ringers and tone dialling. Some of these will now be discussed.

8.4 ELECTRONIC TELEPHONE

By the late 1960s a number of electronics research and development organizations were working on the development of electronic telephone sets. Manufacturing cost reduction, improved performance and the possibility of offering the subscriber a

number of new uses for their telephones were incentives for this development. The first truly electronic component to emerge was the tone dialler, more popularly known by its trade name TOUCH-TONETM. Telephone sets with other electronic components were not far behind so that by the early 1980s there were a number of telephones on the market with none of the well established components described above.

8.4.1 Microphones

The features of the carbon microphone that made it indispensable in the telephone set for a very long time were low cost, acoustic-to-electrical power amplification, suppression of low level background noise, and high level signal compression. Its drawbacks were distortion, high dc current requirements, and changes in its acoustic sensitivity due to dc current flowing in it. All of its advantages can be obtained with none of the disadvantages by using other microphones in conjunction with a suitable amplifier. The new microphones could be made considerably smaller than the carbon microphone. Examples of such microphones include the following:

- (1) *Electret Microphone*. This is a variation on the capacitance microphone. The incident sound causes the distance between two plates of a capacitance to change, resulting in a change of voltage. The output voltage and power are very low. The output impedance is very high (10 pF at audio frequency). The normal biasing of capacitor microphones is averted by a built-in charge that is placed on the capacitor during the manufacturing process. It has an excellent frequency response and is normally used in acoustical measuring instruments. [4]
- (2) *Ribbon Microphone*. A thin aluminum "ribbon" is suspended in the field of a small powerful magnet. The incident sound causes the ribbon to vibrate in the field, causing a voltage proportional to its velocity to be induced in it. It has a very good frequency response but a very low output power.
- (3) *Crystal Microphone*. A crystal of Rochelle salt (sodium potassium tartrate), quartz and other piezo-electric materials produce a voltage when subjected to mechanical deformation. The crystal is cut into a thin layer with suitable conductors connected to the faces. The incident sound pressure causes a voltage to appear across the conductors. The microphone is very often made up of several layers of crystals.

8.4.2 Receiver

The receiver is one of the few components that has successfully resisted change since Alexander Graham Bell patented it in 1876. The materials used for making the magnet and diaphragm and the actual mechanical construction have changed but the basic principle of operation remains the same.

8.4.3 Hybrid

The function of the ideal hybrid is to direct the signal from the microphone on to the telephone line without loss and to direct the incoming signal on the line to the receiver with no loss. The operation of the hybrid is therefore similar to the operation of a circulator – a well known device in microwave engineering. The two devices are compared in Figure 8.14.

There are two major differences:

- (1) There are two critical paths in the hybrid, (transmit and receive) but three critical paths in the circulator. In a normal circulator, the transmitter cannot have a path to the receiver.
- (2) Circulators are realizable in reasonable physical dimensions at microwave frequencies. A direct application of circulator theory at audio frequency predicts a device several kilometres in diameter.

The problem of realizing a circulator at audio frequency can be solved by using a *gyrator* [5]. Gyrators are better known for their ability to invert impedances [6,7]. They were the subject of intense interest at a time when the micro-electronics industry was looking for a micro-miniaturized version of the inductor.

Consider the two-port shown in Figure 8.15 with the port voltages and currents as indicated.

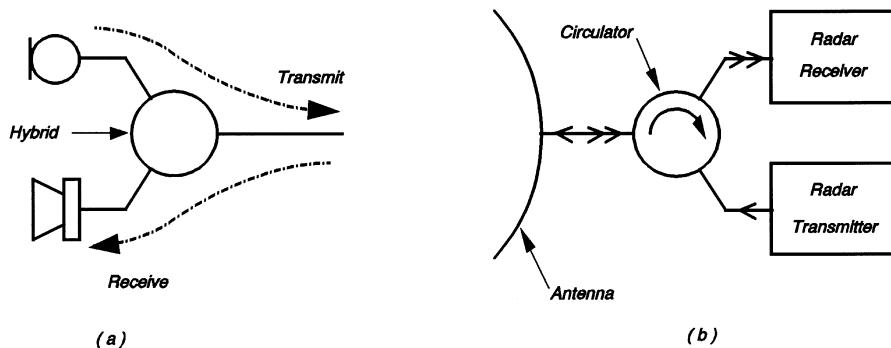


Figure 8.14. The telephone hybrid (a) compared to the circulator (b).

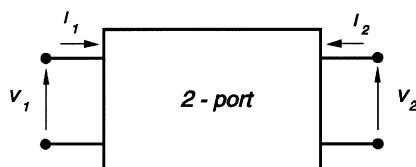


Figure 8.15. A two-port with its defining voltages and currents.

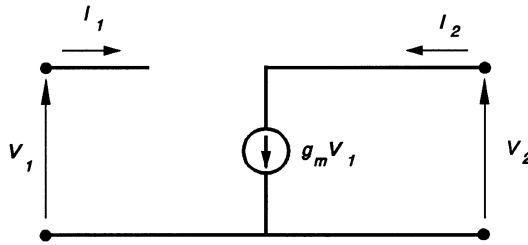


Figure 8.16. The ideal voltage-controlled current source.

The two-port is described by the matrix

$$\begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}. \quad (8.4.1)$$

Applying this description to the ideal voltage-controlled current source shown in Figure 8.16 gives

$$\begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ g_m & 0 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \quad (8.4.2)$$

where g_m is the transconductance.

Consider a second ideal voltage-controlled current source with a 180° phase shift. The matrix equation is

$$\begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ -g_m & 0 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \quad (8.4.3)$$

when the two ideal voltage-controlled current sources are connected back-to-back as shown in Figure 8.17.

The matrix equation is

$$\begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} 0 & -g_m \\ g_m & 0 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}. \quad (8.4.4)$$

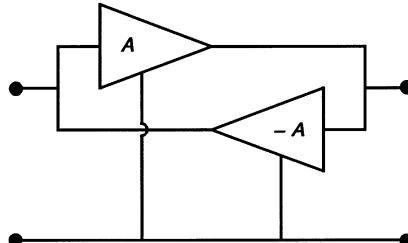


Figure 8.17. Two ideal voltage-controlled current sources connected back-to-back to form a gyrator.

The two-port can be converted into a three-terminal element by lifting the ground. Its matrix equation is then

$$\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \begin{bmatrix} 0 & g_m & -g_m \\ -g_m & 0 & g_m \\ g_m & -g_m & 0 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \\ V_3 \end{bmatrix}. \quad (8.4.5)$$

Figure 8.18 shows the three-terminal circuit in which the transconductance g_m has been replaced by the more general transadmittance Y_0 and port 1 has been terminated in an admittance Y_1 , port 2 in Y_2 and port 3 in Y_3 .

Consider that an ideal current source I_1 is connected across port 1 so that a voltage V_1 appears across it. This will induce voltages V_2 and V_3 across ports 2 and 3, respectively. The voltage ratio is

$$\frac{V_2}{V_1} = \frac{Y_1(Y_3 - Y_0)}{Y_2 Y_3 + Y_0^2}. \quad (8.4.6)$$

Similarly,

$$\frac{V_3}{V_1} = \frac{Y_1(Y_2 + Y_0)}{Y_3 Y_2 + Y_0^2}. \quad (8.4.7)$$

When $Y_1 = Y_2 = Y_3 = Y_0$, Equation (8.4.6) gives

$$\frac{V_2}{V_1} = 0 \quad (8.4.8)$$

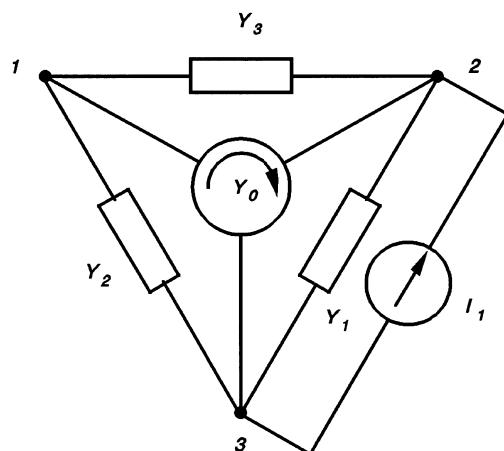


Figure 8.18. The gyrator when properly terminated behaves like a circulator.

and Equation (8.4.7) gives

$$\frac{V_3}{V_1} = 1. \quad (8.4.9)$$

This means that voltage across port 1 appears across port 2 but no voltage appears across port 3, that is,

$$V_2 = V_1 \quad \text{and} \quad V_3 = 0. \quad (8.4.10)$$

When the process is repeated with the ideal current source connected across port 2,

$$V_3 = V_2 \quad \text{and} \quad V_1 = 0. \quad (8.4.11)$$

Finally, with the ideal current source across port 3,

$$V_1 = V_3 \quad \text{and} \quad V_2 = 0. \quad (8.4.12)$$

It is clear that the circuit is behaving like a circulator oriented in a clockwise direction.

To get the desired effect, the transadmittance of the gyrator amplifiers have to be adjusted to fit the line admittance. The admittances of the microphone and receiver circuits have to be tailored so that the gyrator sees an admittance equal to its own admittance connected to each port. In practice, less than a perfect match can be achieved and therefore some sidetone is obtained. This property can be exploited to adjust the level of the sidetone to a comfortable level.

Several electronic telephones use the audio-frequency circulator concept and its various manifestations as hybrids.

8.4.4 Tone Ringer

In most electronic telephone sets, the bell has been replaced by some kind of tone ringer which usually emits an attractive musical note or notes to signal an incoming call. Quite often amplitude and frequency modulation are used to enhance the tone. The tone ringer must satisfy the following conditions:

- (1) The input impedance must be high so as not to interfere with the signal on the line to which it is permanently connected.
- (2) It has to operate on the 20 Hz, 88 V ac ringing signal that was used with the electromagnetic bell.

In terms of circuit design, what is required is one or two audio-frequency oscillators, a frequency and/or amplitude modulator, a power supply fed from the 20 Hz ring signal, an audio-frequency amplifier and a loudspeaker. The design of all these items has been discussed earlier. Oscillators were discussed in Section 2.4, modulators in

Section 2.6, audio-frequency amplifiers in Section 2.7, and loudspeakers in Section 3.4.8. The power supply for the ringer would require a rectifier and a capacitor to smooth the output of the rectifier to form a suitable dc supply for the ringer.

8.4.5 Tone Dial

Instead of producing current pulses to signal the number dialled to the central office, the tone dial produces a pair of audio-frequency tones. The frequencies of these tones are carefully chosen so they are not harmonically related. This reduces the probability of other tones or signals being recognized as dialled numbers. The dial pad and the corresponding frequencies produced are shown in Figure 8.19.

8.4.5.1 Touch-ToneTM – DigitoneTM Dial. This dial consists of two essentially identical oscillators, one of which produces the high-frequency and the other the low-frequency tones. The change in frequency is achieved by switching in resistors of appropriate values when the dial push button is depressed. Each button has a unique pair of tones associated with it. The central office equipment recognizes the number dialled by the two frequencies present.

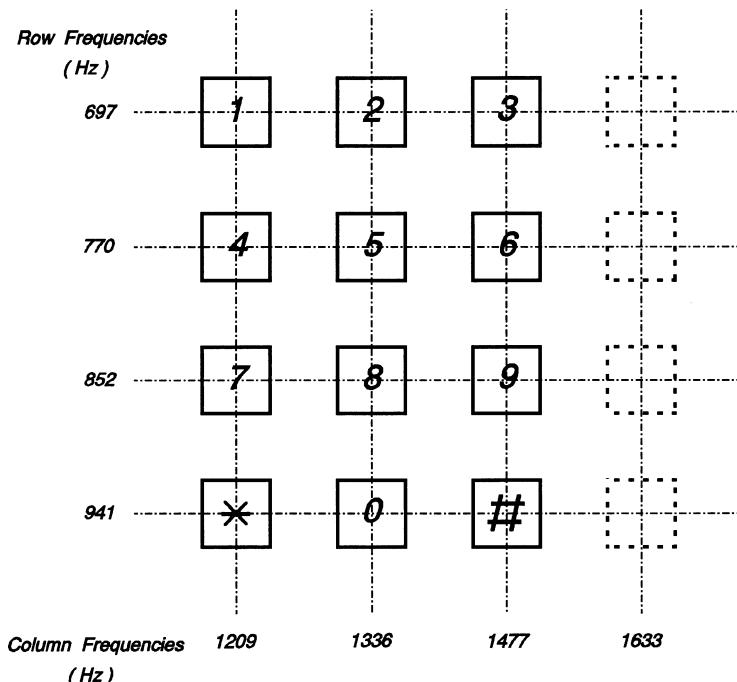


Figure 8.19. The push-button dial and its corresponding frequencies.

The early versions of the TOUCH-TONE™ dial used discrete bipolar transistors in conjunction with an inductor and capacitances in a Colpitts configuration. Both the low- and high-frequency groups were produced by a single oscillator. One of the capacitances in each group was changed as the various buttons were depressed to produce the required frequencies. Later versions used an integrated-circuit amplifier with an *RC twin-tee* feedback circuit to produce the tones. The basic configuration of the circuit is shown in Figure 8.20.

Two conditions have to be met for oscillations to occur, according to the Barkhausen criterion:

- (1) The closed loop gain must be equal to unity. In practice, the loop gain must be slightly larger than unity for sustained oscillation.
- (2) The change in phase around the loop must be an integer multiple of 2π radians.

The classical RC twin-tee filter has values of R_s and C_s as shown in Figure 8.21(a) and the amplitude and phase responses in Figure 8.21(b).

Its transfer function is given by

$$\frac{V_2}{V_1} = -\frac{1 - \omega^2 C^2 R^2}{1 - \omega^2 C^2 R^2 + j4\omega CR}. \quad (8.4.13)$$

Rationalizing and equating the imaginary part to zero gives the frequency at which the phase angle is zero or 180°

$$\omega_0 = \frac{1}{RC}. \quad (8.4.14)$$

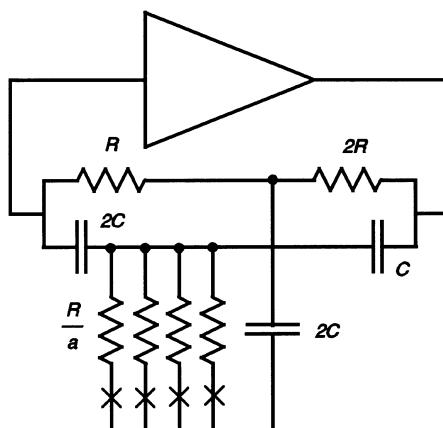


Figure 8.20. The modified twin-tee feedback oscillator. The closing of one of the four switches connects a different R/a to produce the tones. Two such circuits are used in the dial, one for the low frequencies and the second for the high frequencies.

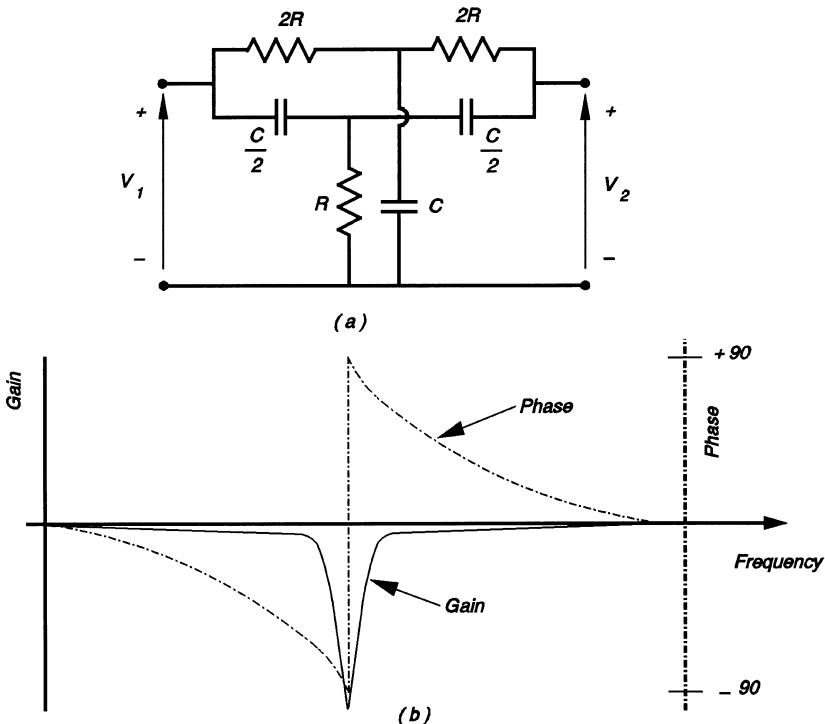


Figure 8.21. (a) The classical twin-tee notch filter showing its configuration and circuit element ratios. (b) The gain and phase characteristics of (a).

Under this condition the output voltage $v_2 = 0$; the circuit has a null in its frequency response with a very high Q factor. The high Q factor can be exploited for high stability of oscillating frequency if the oscillator is designed to operate at the frequency of the null. However, using the classical twin-tee values of R_s and C_s in oscillator design will be self-defeating since an amplifier with infinite gain will be required.

Departure from the standard ratios of R_s and C_s produces lower values of Q factor. The amplitude-frequency responses of the modified twin-tee filter (Figure 8.20) with different values of a are shown in Figure 8.22(a). The frequencies at which the phase shift of the filter is zero or 180° coincides with the frequency at which the output voltage is a minimum (null). It can be seen that the notch frequency for the particular configuration and circuit element ratios shown in Figure 8.20 changes with the parameter a . It can also be observed that the depth of the notch varies as a changes, with the “deepest” notch occurring when a is equal to approximately 2.25. For values of a less than 2.25, the modified twin-tee circuit has gain and phase characteristics as shown in Figure 8.22(b). When a is equal to or greater than 2.25, the gain and phase response is as shown in Figure 8.22(c). The

oscillator design technique described here therefore works only when a is greater than 2.25.

The oscillator design exercise consists of identifying the value of a which has its notch at the required frequency. The gain needed from the amplifier at the null or where the 180° phase shift occurs can be identified. The amplifier can then be designed to have the required gain and a phase shift of 180° .

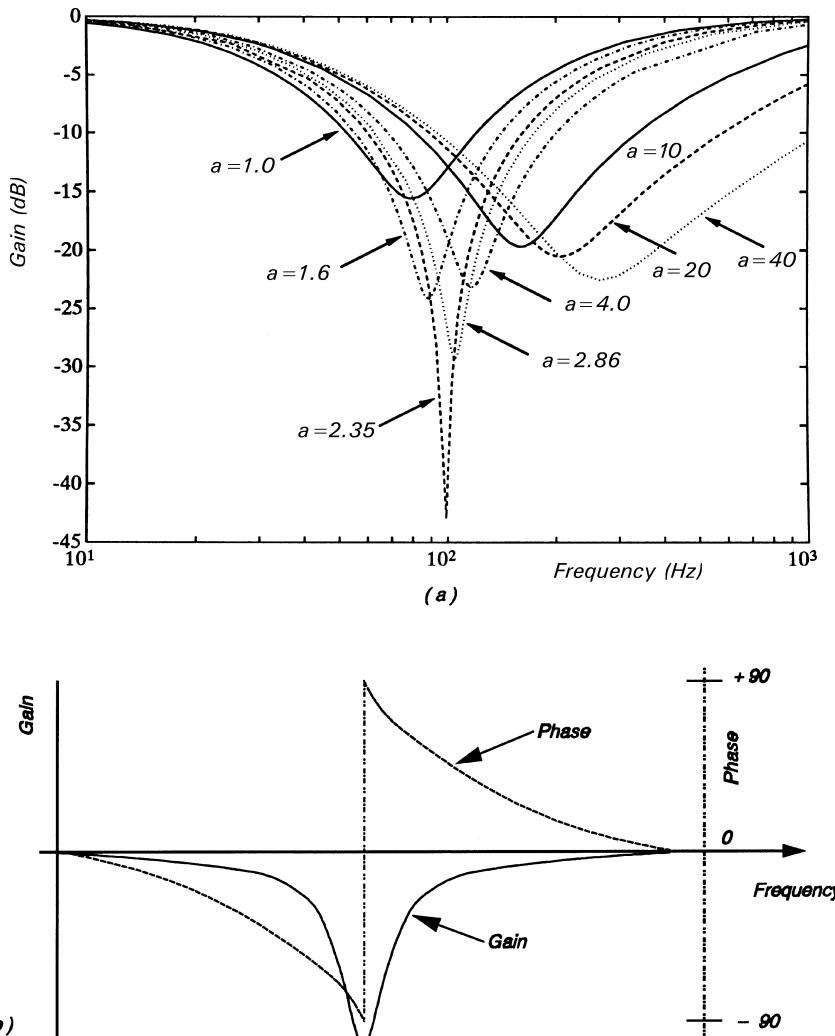


Figure 8.22. (a) The gain response of the modified twin-tee for various values of a . (b) The gain and phase characteristics of the modified twin-tee when a is less than 2.25. (c) The gain and phase characteristics of the modified twin-tee when a is greater than 2.25.

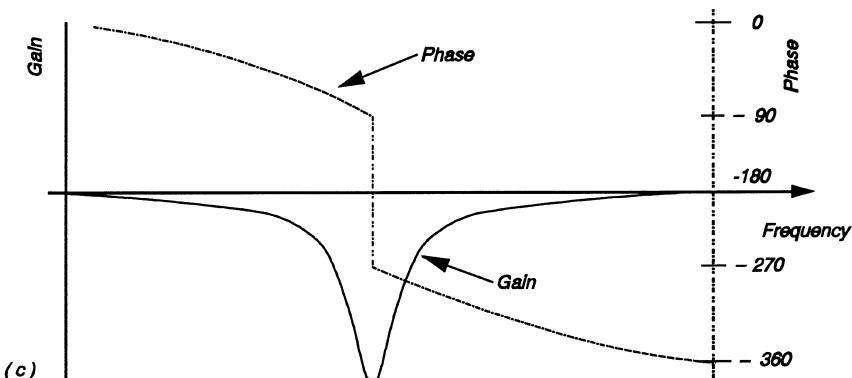


Figure 8.22. (continued)

8.4.5.2 Digital Tone Dial. The digital tone dial attempts to exploit the low cost of digital integrated circuits. The design of the system is shown in Figure 8.23.

The crystal-controlled oscillator generates a signal at 3.58 MHz. This frequency was chosen because a crystal designed to operate at that frequency was readily available and inexpensive (it is used in the color burst carrier of television sets; see Section 7.3.2.2). The actual frequency is not important so long as it is sufficiently high that division by an integer will produce frequencies which lie within the permitted error margin of the tone frequencies. The oscillator frequency is first divided by 16 to give a clock frequency of 223.75 kHz. The push-button pad has a number of contacts which are used to send a binary logic statement of 1s and 0s to the N coder. The function of the N coder is to generate its own set of 1s and 0s as input to the divide-by- N . The divide-by- N consists of a set of resetable binary counters with additional logic circuitry to reset the counters so that the number N can be changed according to the output of the N coder. The function of the eight-stage shift register is to produce eight sequential pulses at the output which have 1/8th the period of the required sinusoidal signal. These pulses are used to drive the digital-to-analog converter to produce a crude eight-step approximation to the sine wave. The low-pass filter attenuates the unwanted harmonics before the signal drives the telephone line through the line driver. Except for minor differences in the N coder, the two halves of the circuit are the same. The basic circuit blocks used are the NOR, NAND, NOT or inverter and the resetable bistable multivibrator, known collectively as logic gates. These gates are used in large quantities but each gate occupies such a small area on an integrated circuit chip that the cost is minimal. For example, this circuit had 10 NOR, 4 NAND, 44 NOT, and 23 resetable bi-stable multivibrators.

Logic gates can be realized in different forms, each with its own mode of operation. They are grouped into families such as *complementary metal oxide semiconductor* (CMOS), *transistor-transistor logic* (TTL), *resistor transistor logic*

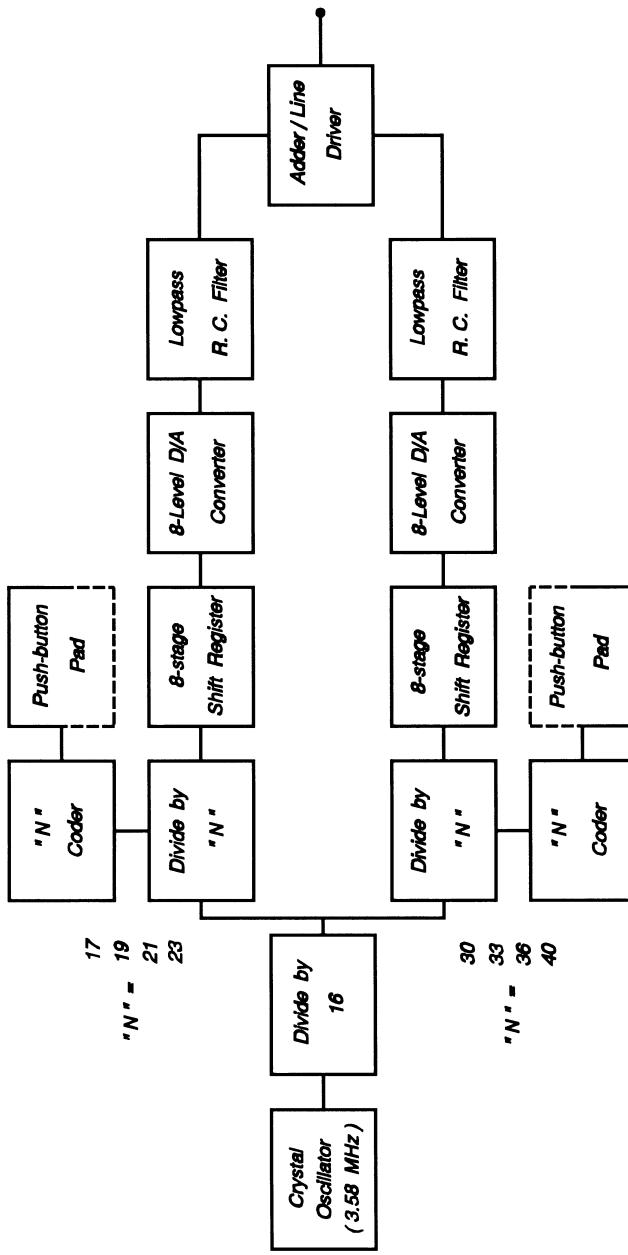


Figure 8.23. The block diagram of the digital tone dial.

(RTL), and so on. In explaining the design of logic gates, the logic family whose mode of operation appears to be the simplest is chosen.

Before discussing the design of the digital tone generator it is best to digress to explain the design of the basic building blocks shown in Figure 8.23.

Design of the NOT gate. The basic NOT gate is shown in Figure 8.24(a). When the input is a 0 (or essentially zero volts), the transistor is cut off and no current flows through the collector resistor R_c . The collector is then a 1 which is equivalent to the system voltage V_{cc} . When the input is a 1, current flows through the resistor R_1 and forward biases the base-emitter junction. Collector current flows and, given an appropriate value for R_c , the transistor goes into saturation, that is, the output is a 0.

Two components are added on as shown in Figure 8.24(b) to improve the performance of the gate. A capacitor is connected across R_1 . Since the voltage across a capacitor cannot change instantaneously, the leading edge of a positive pulse will cause the base voltage to rise immediately, causing the transistor to conduct with a minimum of delay. A capacitor used for this purpose is called a "speed-up" capacitor. The second component is the resistor R_2 which is connected to a dc source $-V_{BB}$. The purpose of this circuit is to ensure that the base-emitter voltage is normally kept at a slightly negative value so that the probability of the gate switching due to noise is reduced.

The design of this circuit is best illustrated by an example:

Example 8.4.1 The NOT-gate. A NOT gate drives a load which requires $I_L = 1 \text{ mA}$. The dc supply voltage $V_{cc} = 10$, $V_{BB} = 5 \text{ V}$ and the transistor is a silicon NPN bipolar with a $\beta = 100$. Determine suitable values for R_1 , R_2 and R_c .

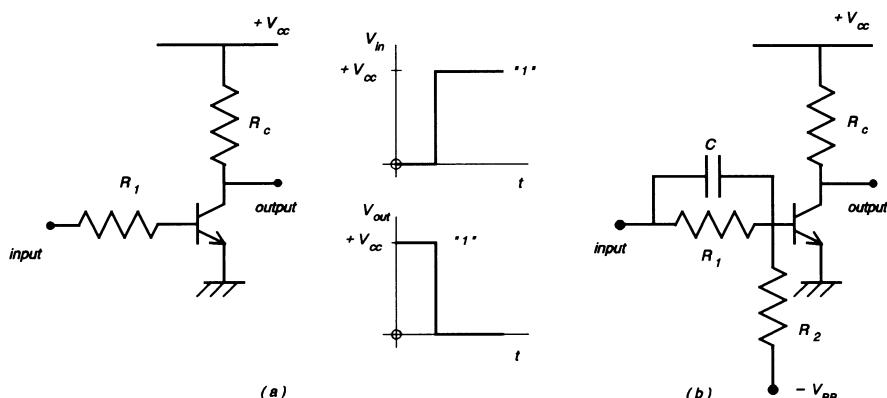


Figure 8.24. (a) The basic NOT gate and its transition diagram. (b) The NOT gate with "speed-up" capacitor and negative biasing to improve noise immunity.

Solution. To prevent the load current from interfering with the operation of the gate, it is necessary to make the collector current about 10 times the load current

$$I_C = 10 \times I_L = 10 \text{ mA}.$$

The transistor goes into saturation when the voltage drop across R_c is equal to $V_{CC} = 10\text{V}$:

$$R_C = V_{CC}/I_C = 10 \text{ V}/10 \text{ mA} = 1 \text{ k}\Omega.$$

Base current required for 10 mA collector current is

$$I_B = I_C/\beta = 10 \text{ mA}/100 = 100 \mu\text{A}$$

$$R_1 = (V_{CC} - V_{BE})/I_B = (10 - 0.7)/100 \mu\text{A} = 93 \text{ k}\Omega.$$

In practice, the value of R_1 may be reduced to half the calculated value to increase the margin of safety.

To improve the noise immunity, let the base voltage be reverse biassed by -3 V when the input is 0 V . This means that R_2 has to be chosen so that

$$R_2/R_1 = 2/3$$

$$R_2 = 62 \text{ k}\Omega.$$

It is necessary to verify that, when the input is at 10 V (V_{CC}), R_2 and V_{BB} will not hold the base voltage below 0.7 V . This step may be carried out by considering the transistor to be disconnected from the rest of the circuit. The voltage at the node of R_1 and R_2 can then be calculated. Any voltage greater than 0.7 V ensures that the transistor will indeed be biassed on when the input is a **1**. In this case the base voltage would have been 1.0 V but the base-emitter diode will hold it at 0.7 V .

The value of the speed-up capacitor depends on the frequency of operation of the gate. A reasonable choice is $50\text{--}100 \text{ pF}$.

Design of the NOR gate. Figure 8.25 shows a two-input NOR gate with its truth table.

When both inputs are **0**s no current flows in either transistor and the output is a **1**. When either A or B is a **1**, current flows in the corresponding transistor which goes into saturation and the output is a **0**. Finally when both inputs are **1**s both transistors conduct and the output is a **0**.

The NOR gate may be viewed as two NOT gates sharing a common collector resistor. The design of the NOR gate follows the same procedure as the NOT gate.

Design of the NAND gate. A modification of the NOT gate by the addition of an extra resistor and two diodes gives a NAND gate. This is shown in Figure 8.26 together with its truth table.

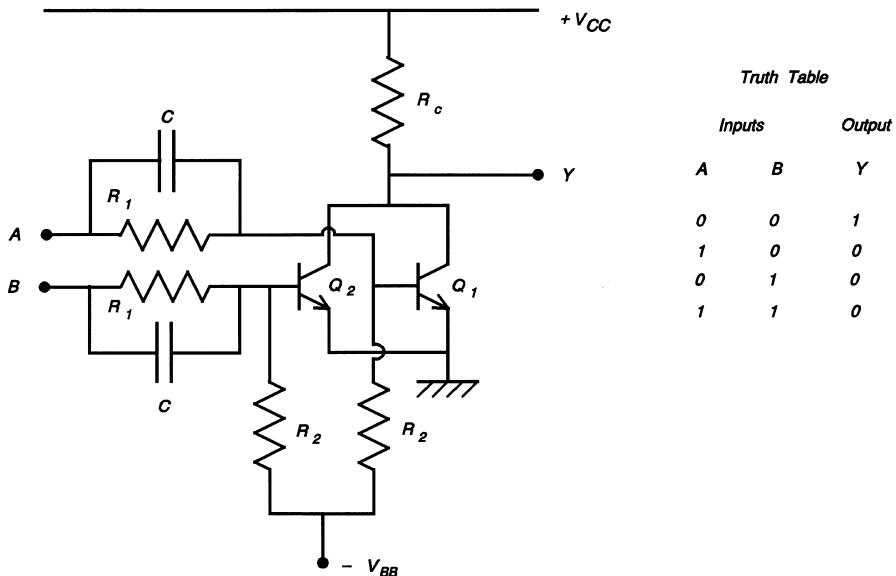


Figure 8.25. The NOR gate with its truth table.

When either *A* or *B* or both are **0**s, current flows from *V_{cc}* through *R₃* and the appropriate diode(s) to ground. The voltage at node *X* is held at 0.7 V and *Q* remains in cut-off. Only when both *A* and *B* are **1**s can base current flow through *R₃* and *R₁* and cause *Q* to go into saturation, producing a **0** at the output.

(The design of the NAND gate follows from that of the NOT gate. The procedure for calculating the values of *R_c*, *R₂* and *C* remain the same. *R₁* from the NOT gate

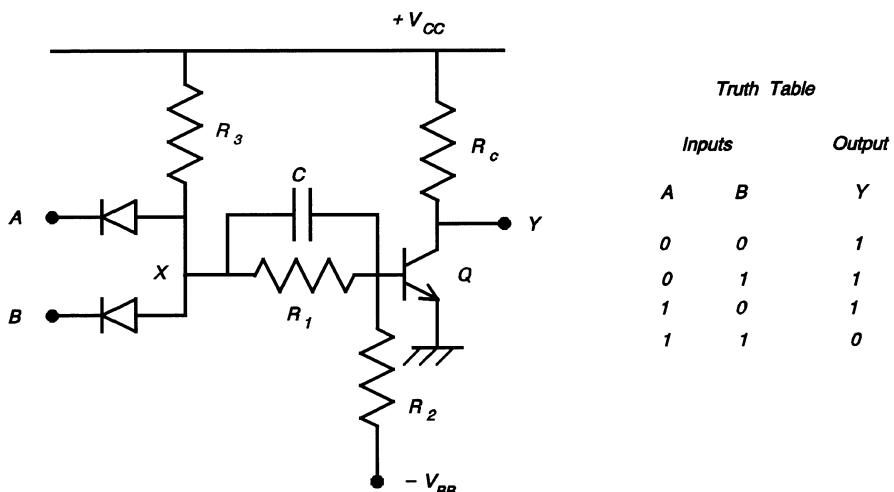


Figure 8.26. The NAND gate and its truth table.

example has to be split in two to form R_3 and R_1 as shown in Figure 8.26. A reasonable compromise might be to make $R_3 = 33\text{ k}\Omega$ and $R_1 = 60\text{ k}\Omega$. The circuit that drives the NAND gate will then not have to sink large currents through the diodes.

Design of the Bistable Multivibrator. The easiest way to understand the operation of the bistable multivibrator is to consider it to be two NOT gates connected back-to-back as shown in Figure 8.27.

When the dc power is first switched on, suppose Q_1 draws more current than Q_2 . The collector voltage of Q_1 will therefore start to drop faster than that of Q_2 . The base-emitter junction of Q_2 will be a little less forward biassed and its collector current will be reduced at a rate determined by the current gain of Q_2 . Its collector voltage will therefore tend to rise pulling up the base of Q_1 , thereby causing the base-emitter junction to become more forward biassed. More current will then flow through the collector of Q_1 due to the current multiplication in the transistor, causing the collector voltage to drop faster than before. The event that started the process has become magnified many times over. This is a regenerative process and comes to a stop when Q_1 goes into saturation and Q_2 is cut off. Unlike the astable multivibrator described in Section 7.2.7, there is no mechanism to cause the two transistors to switch back and forth. Q_1 stays in saturation and Q_2 is cut off until some external event causes them to change states. This event could be a small positive-going pulse or trigger applied to the base of Q_2 . If the trigger is sufficient to bring Q_2 into partial conduction momentarily, this will cause current to flow in the collector of Q_2 . Its

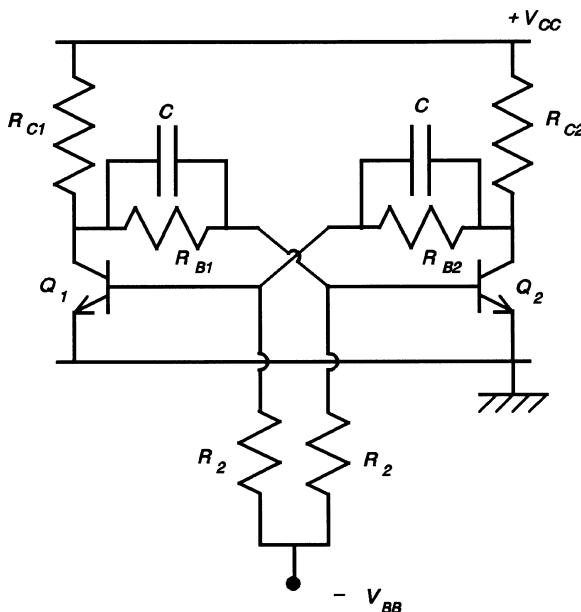


Figure 8.27. The basic bistable multivibrator.

collector voltage will drop, causing the base-emitter junction of Q_1 to be less positively biased. Less collector current will flow and the collector voltage of Q_1 will start to rise causing the base-emitter junction of Q_2 to become even more forward biassed. Eventually, Q_2 goes into saturation and Q_1 is cut off. This is the end of the regenerative cycle in the opposite direction and the circuit will remain in the current state until another external event causes it to change states.

The bistable multivibrator can be made much more interesting and useful by the addition of the “steering circuit” shown in Figure 8.28.

The diodes D_1 , D_2 and D_3 are connected and their common node is coupled to a source of negative-going clock pulses by the capacitor C_3 . Assuming that Q_1 is conducting, its collector voltage will be essentially at ground potential while that of Q_2 will be at V_{cc} . When the negative-going pulse arrives, D_2 will conduct momentarily causing the voltage at the collector of Q_2 to drop. The regenerative cycle is set in motion and the cycle ends with Q_1 off and Q_2 on. The next negative-going pulse will cause the transistors to change states yet again. This is illustrated in Figure 8.29.

It should be noted that the diode D_3 guarantees that static charge build-up on the capacitor C_3 cannot exceed $V_{cc} + 0.7$ V.

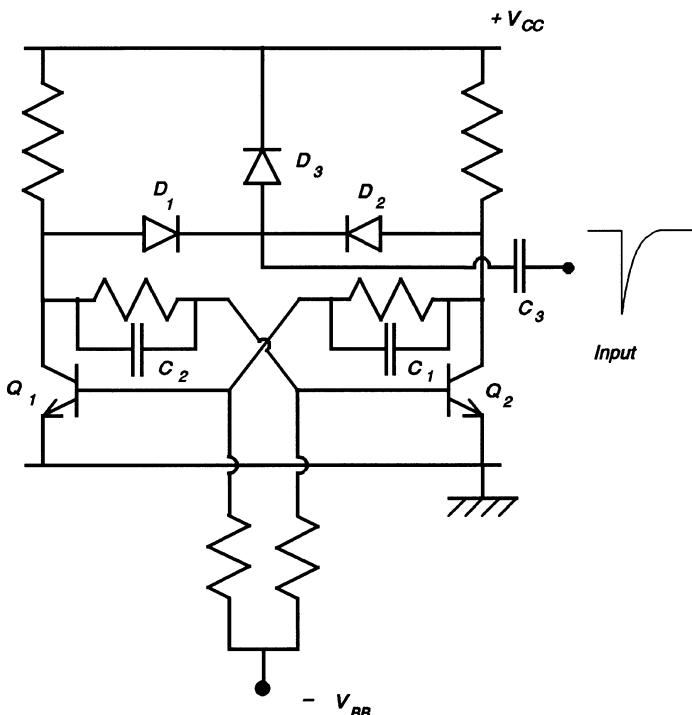


Figure 8.28. The bistable multivibrator with the steering diodes is called a flip-flop.

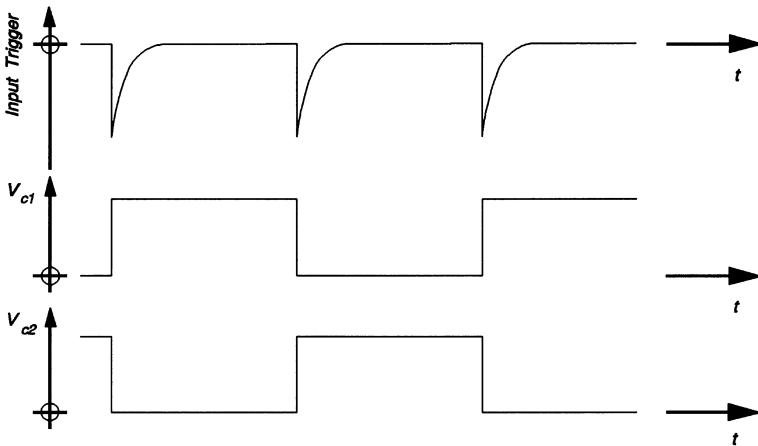


Figure 8.29. A timing diagram for the flip-flop showing the input trigger and the collector voltages.

Two very important conclusions can be drawn from Figure 8.29:

- (1) The frequency of the output signal taken from either transistor is one-half of the input frequency. This is a *divide-by-two* circuit.
- (2) The circuit has a *memory*; if the current state is known, one can deduce the previous state.

In terms of the current discussion, the first conclusion is the more relevant. By using two or more of these bistable multivibrators in cascade, it is possible to get a *divide-by-four* (2^2), three of them will perform a *divide-by-eight* (2^3) and so on. In Figure 8.23, a cascade of four bistables are used to realize the divide-by-sixteen. The bistable multivibrator with its steering circuit is popularly known as a *flip-flop*.

A *set* and a *reset* feature can be added to the circuit whereby the state of the flip-flop can be set before the arrival of a chain of pulses. For example, a **1** can be applied momentarily to the base of Q_1 to ensure that it is conducting current. It is then possible to predict the state it will be in after any given number of input pulses. A flip-flop with the set and reset features is called an *S-R* flip-flop.

Yet another useful feature can be added so that, when the clock pulse arrives, the flip-flop switches states only when a **1** or a **0** is present at a designated terminal. The flip-flop is then called a *J-K* flip-flop. Flip-flops are discussed in almost any textbook on digital circuits and the interested reader is advised to refer to such texts for more details.

The Crystal-controlled Oscillator. The crystal-controlled oscillator is shown in Figure 8.30.

With the two inputs to the NOR gate tied together, it becomes a NOT gate which is, in effect, a high gain amplifier with phase inversion. The crystal, together with the

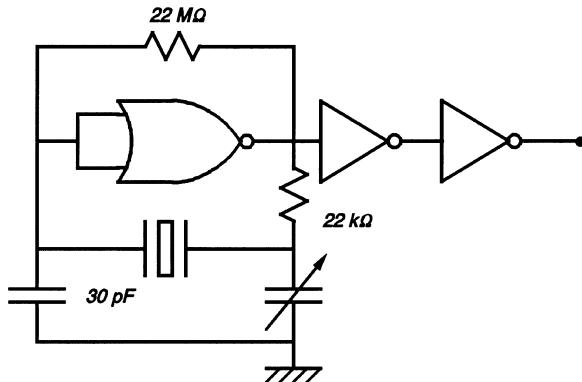


Figure 8.30. The complementary metal-oxide semiconductor (CMOS) transistor version of the crystal-controlled oscillator. The output is a square wave.

two capacitors provide the feedback path. The output waveform is not quite a square wave. The two inverters in cascade amplify and clip the signal to give a good square waveform.

The Divide-by-16 Circuit. As mentioned earlier, this is a cascade of four flip-flops connected in a binary “ripple counter” arrangement. It divides the crystal oscillator frequency by 16 to give an output of 223.75 kHz.

The N-coder Circuit. The N -coder uses the information from the dial pad to set up a pattern of **1s** and **0s** and sends these to the divide-by- N circuit. The particular pattern decides what the value of N has to be. The N -coder uses only NOR and NOT gates.

The Divide-by- N Circuit. A combination of the set and reset features and the exploitation of other logic circuits such as NOR, NAND, and NOT gates provide the possibility to reset a cascade of flip-flops after n pulses of input where n takes on values other than 2^m , where m is an integer. This is the scheme used in the divide-by- N circuit with the N -coder supplying the necessary **1s** and **0s** to the appropriate flip-flops to set and/or reset them.

Using the lowest frequency (697 Hz) as an example, the input frequency to the divide-by- N circuit is 223.75 kHz. The selected value of N is 40. This gives an output frequency of 5.594 kHz which is used to clock the eight-stage shift register.

The Eight-Stage Shift Register. The eight-stage shift register is made up of a cascade of flip-flops, NOR, and NOT-gates to set up a sequence of pulses which have a width equal to 1/8th the period of the required frequency. One-eighth the period is used because the required sine wave is to be approximated by an eight-step function as shown in Figure 8.31.

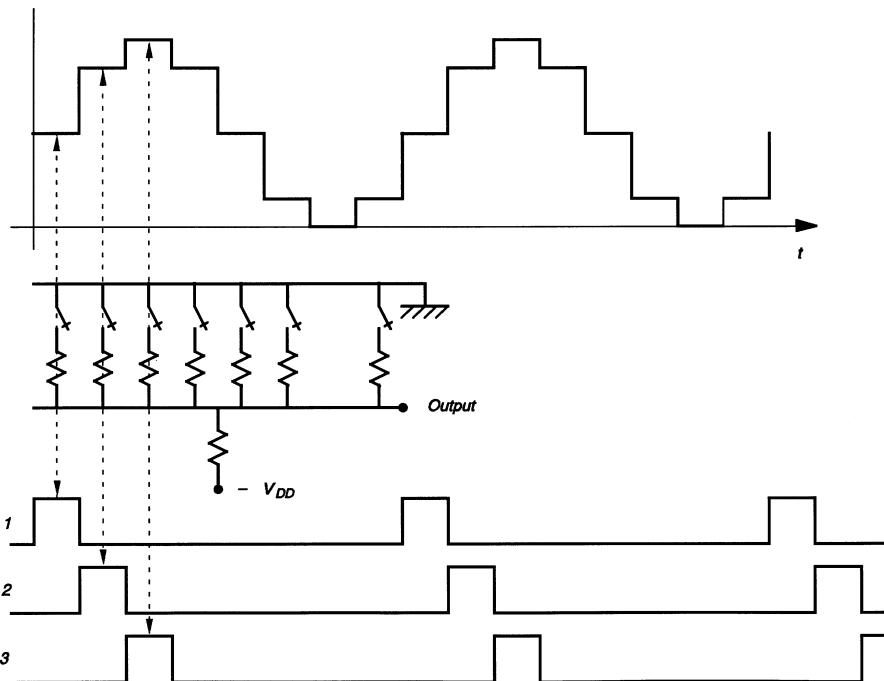


Figure 8.31. The eight-stage shift register closes each switch in turn. The resistor values are chosen to produce eight current levels which approximate a sinewave.

The Eight-Level Digital-to-Analog Converter. The sequential pulses from the shift register are used to drive a bank of eight field-effect transistors switches connected in parallel as shown in Figure 8.32.

The output is an eight-level approximation to a sinusoidal current. Going back to the example (under divide-by- N circuit), the fundamental frequency of the approximation to the sinusoid is $5.594/8 = 699.3$ Hz. The error is 0.3%.

The Low-Pass Filter. The low-pass filter attenuates the higher harmonics of the waveform before the signal reaches the line driver. With the exception of very short lines this filter is not required; the telephone line is a good enough low-pass filter. The harmonic content can be reduced by using more than eight steps to represent the sinewave.

Table 8.1 shows the target frequency f_0 , the required divisor n , the nearest integral divisor N , the actual frequency f'_0 , and the error for all 12 tone-dial signals. The errors indicated here are within the $\pm 1.5\%$ allowed.

8.5 DIGITAL TELEPHONE

Until the introduction of the digital telephone all signals from the station set were transmitted to the central office in analog form over the subscriber loop, which in

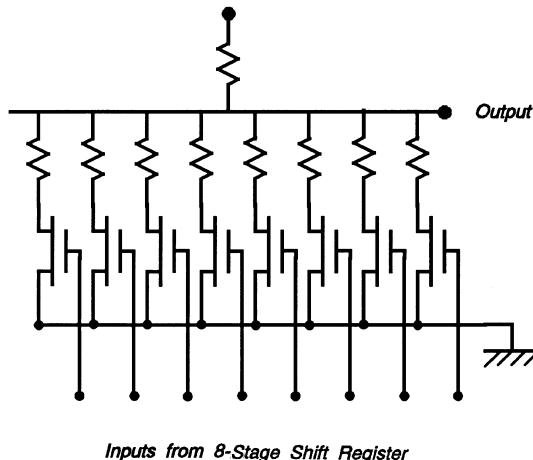


Figure 8.32. The switches in Figure 8.31 are replaced by FETs.

most cases, is the twisted pair of plastic-insulated copper wire. In the central office it may be subjected to a number of sophisticated signal processing techniques, depending on the routing of the message and the medium of transmission. In the digital telephone, the analog signal is converted into digital form by a *codec* (*coder-decoder*) using *pulse-code modulation* (PCM). The pulses are sent along the twisted pair to the central office where they may be further processed before transmission to their destination.

8.5.1 The Codec

The codec is available on an integrated circuit chip and it is installed in the station set. Figure 8.33 shows a possible configuration of the system.

In this configuration, the analog part of the telephone is left unchanged. The physical distance from the telephone hybrid to the 2-to-4 wire hybrid can vary from

TABLE 8.1

Target frequency f_0 (Hz)	$16 \times 8f_0$	Required divisor, n	Integer N	Actual frequency f_0 (Hz)	Error %
1633	209,024	17.12	17	1645.2	+0.7
1477	189,056	18.94	19	1472.0	-0.3
1336	171,008	20.93	21	1331.8	-0.3
1209	154,752	23.13	23	1216.0	+0.6
941	120,448	29.72	30	932.3	-0.9
952	109,056	32.83	33	847.5	-0.5
770	98,560	36.32	36	776.9	+0.9
697	89,216	40.13	40	699.2	+0.3

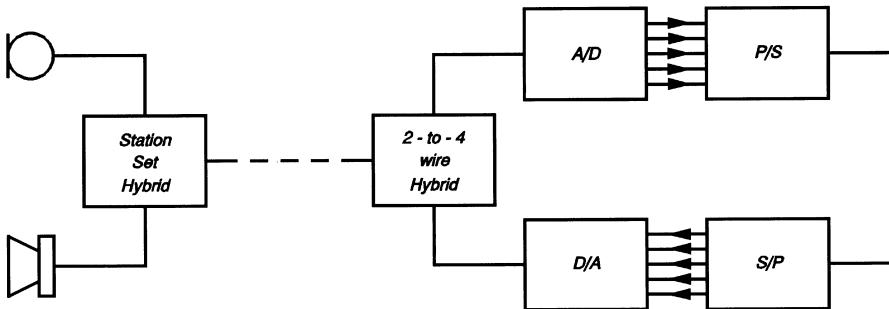


Figure 8.33. The block diagram of the codec used in the digital telephone.

essentially zero, when the codec is installed in the station set, to a few kilometres, when the codec is in the central office. In the transmit mode, the analog signal is fed to the input of the *analog-to-digital* (A/D) converter which gives a set of binary outputs in PCM in a parallel format. A *parallel-to-serial* (P/S) converter changes the parallel format to a serial format so it can be sent down the transmit leg of the digital line. In the receive mode, the input to the decoder is in PCM serial mode and the *serial-to-parallel* converter changes the signal into the parallel mode before going onto the *digital-to-analog* (D/A) converter. The analog output then goes to the receiver. The design of the circuits in the boxes shown in Figure 8.34 now follow. The order of presentation has been changed to take advantage of the fact that an A/D converter makes use of a D/A converter.

8.5.1.1 Digital-to-Analog Converter. As the name suggests, this circuit takes a digital input and converts it to the equivalent analog output. The digital input is presented in the form of **1s** and **0s** from the *least significant bit* (LSB) to the *most significant bit* (MSB). For example the number 26 is presented as:

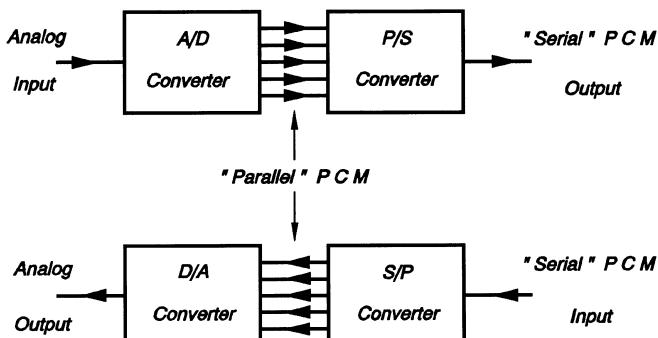


Figure 8.34. Details of the parallel-to-series PCM converter for both transmit and receive modes.

	2^0	2^1	2^2	2^3	2^4	
LSB	0	1	0	1	1	MSB
	0	+ 2	+ 0	+ 8	+ 16	= 26

Consider the operational amplifier connected as shown in Figure 8.35. This is a summing amplifier (Section 4.4.3.5) whose output is given by

$$v_0 = -\frac{V^+ R_2}{R_1} [a_0 2^0 + a_1 2^{-1} + a_2 2^{-2} + a_3 2^{-3} + \dots]. \quad (8.5.1)$$

The coefficients a_n are digital bits which are either **0** or **1** when the switch S_n is down or up, respectively. The voltage V^+ is the reference voltage and R_1 and R_2 are chosen to give a suitable level of signal at the output.

One obvious disadvantage of the circuit is that the larger the number of bits presented, the larger is the ratio of the largest resistor to the smallest. In integrated circuits, large resistors occupy large areas of the chip and are therefore undesirable. However, the function of the R_1 resistors is to scale the current flowing into the summing point of the amplifier according to the significance of the bit. Current scaling can be achieved by changing the circuit as shown in Figure 8.36.

The switches are either connected to the inverting input of the amplifier or they are connected to ground but the inverting input is a virtual ground so the resistive network can be redrawn as the ladder shown in Figure 8.37.

Looking to the right at point X , the resistance seen is R_1 . Again looking to the right at point Y , the resistance seen is still R_1 . It follows that looking at all points corresponding to X gives a resistance R_1 and hence the input resistance of the ladder

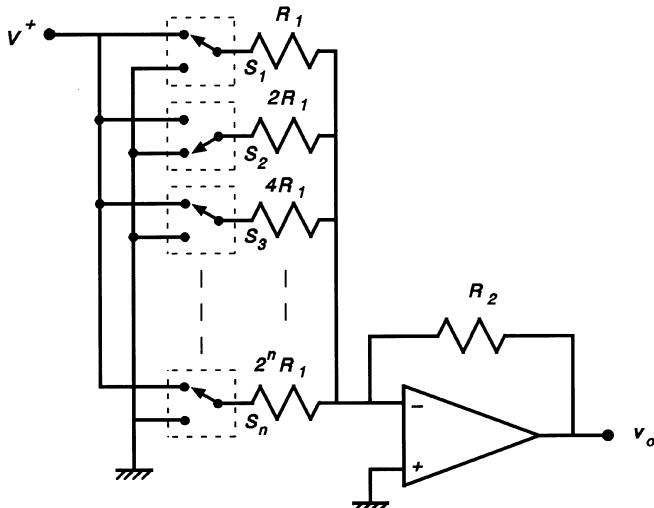


Figure 8.35. The circuit diagram of the digital-to-analog converter.

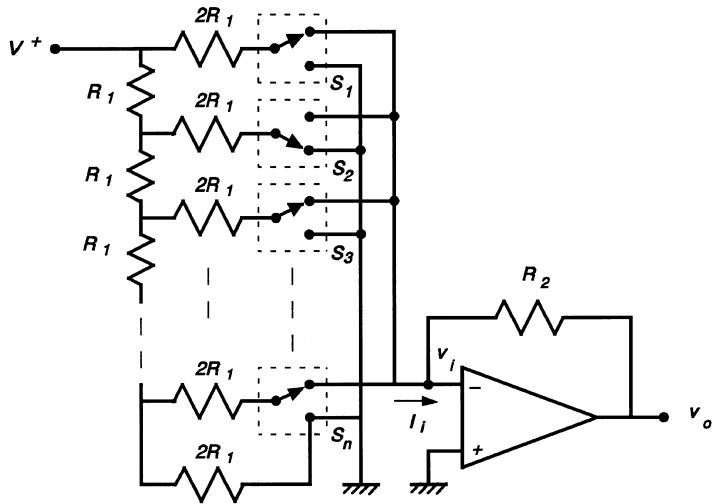


Figure 8.36. An improved version of the digital-to-analog converter which uses the $R-2R$ ladder.

is R_1 . This type of ladder is described as $R-2R$ ladder. Consider the current flowing into the ladder, I . It follows that I will divide equally at node 1 as shown. At node 2, the process is repeated with the current $I/4$ flowing in each branch so that at node n the current in each branch is $I/2^n$. All the currents are summed at the input of the operational amplifier to give

$$I_i = \frac{V^+ - v_i}{R_1} [a_1 2^{-1} + a_2 2^{-2} + a_3 2^{-3} + \dots]. \quad (8.5.2)$$

But

$$I_i = \frac{v_i - v_0}{R_2}. \quad (8.5.3)$$

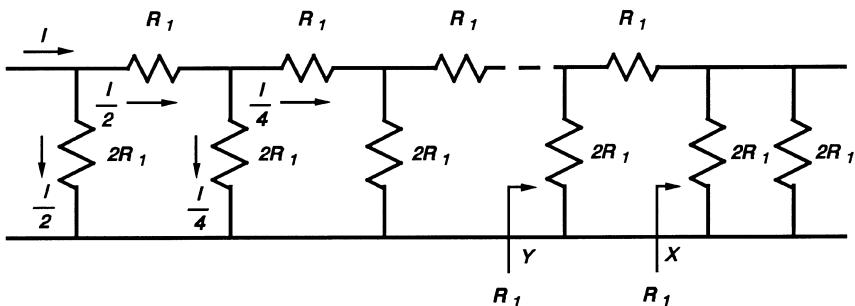


Figure 8.37. The $R-2R$ ladder.

When the operational amplifier gain is very large

$$v_0 = -\frac{V^+ R_2}{R_1} [a_1 2^{-1} + a_2 2^{-2} + a_3 2^{-3} + \dots] \quad (8.5.4)$$

where the a s determine whether a particular bit is a **1** or a **0**. Note that the error in the conversion is $\pm \frac{1}{2}$ LSB (Least Significant Bit).

8.5.1.2 Analog-to-Digital Converter. Several different strategies have been used in the design of A/D converters, each with its own advantages and disadvantages. For this discussion, the *Ramp Counter A/D* converter has been chosen because its operation is relatively easy to understand. The ramp counter A/D is shown in Figure 8.38.

Two inputs are applied to the *comparator*; the analog signal and a voltage from the *resistive network*. The comparator output is a **0** if the analog signal is greater than the voltage from the resistive network and a **1** if the opposite is true. The comparator output and a clock drive a two-input NAND gate so that the NAND gate permits the clock pulses to reach the *resetable binary counter* only when the comparator output is a **0**, that is, the NAND gate input is a **1** due to the inverter. The binary counter counts the clock pulses and simultaneously closes or opens the switches S_1 to S_n , depending on the bit presented (**1** or **0**) that it applies to the switches. Which switches are closed or open determine the output voltage of the resistive network. This is in fact a D/A converter! So long as the output voltage of the resistive network is lower than the input analog voltage, the binary counter continues to count and continues to increase the output voltage of the resistive network. Finally, the voltage from the resistive network exceeds the analog input voltage and the comparator output changes to a **0**. The NAND gate stops the clock pulses from reaching the binary counter and it therefore stops counting. The output from the binary counter can now be read. The *convert enable* clears the registers of the binary counter ready for the next sample of the analog input voltage to be converted.

The design of the NOT gate, the NAND gate, the binary counter, and the switches were discussed in Section 8.4.5. The design of the resistive network plus operational amplifier (D/A converter) was the subject of Section 8.5.1.1. The comparator, the parallel-to-serial convertor, and the 4-to-2 wire hybrid will now be discussed.

8.5.1.3 Comparator. The circuit of the comparator is shown in Figure 8.39(a). The ideal operational amplifier is assumed to have infinite gain so that, if the input varies infinitesimally about 0 V, a positive-to-negative change will cause the output to go from saturation at the negative dc supply voltage to the positive. The operational amplifier is therefore a comparator which compares its input to 0 V. To stop the output from saturating at the dc supply rails, a Zener diode may be connected as shown in Figure 8.39(a). For a very small positive voltage input, the output will go negative but it will be clamped when it drops to -0.7 V by the forward biassed Zener diode. Equally, a very small negative voltage input will cause the output to go positive but again it will be clamped, this time by the Zener voltage

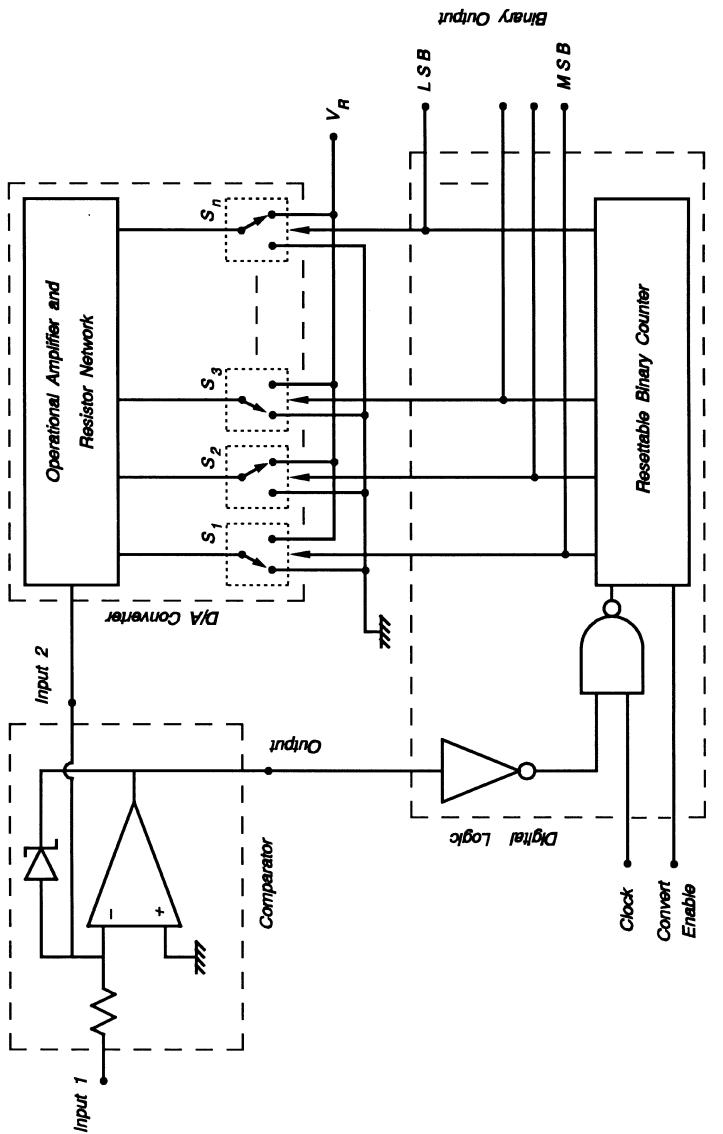


Figure 8.38. The block diagram of the analog-to-digital converter. Note that it contains within it the digital-to-analog converter.

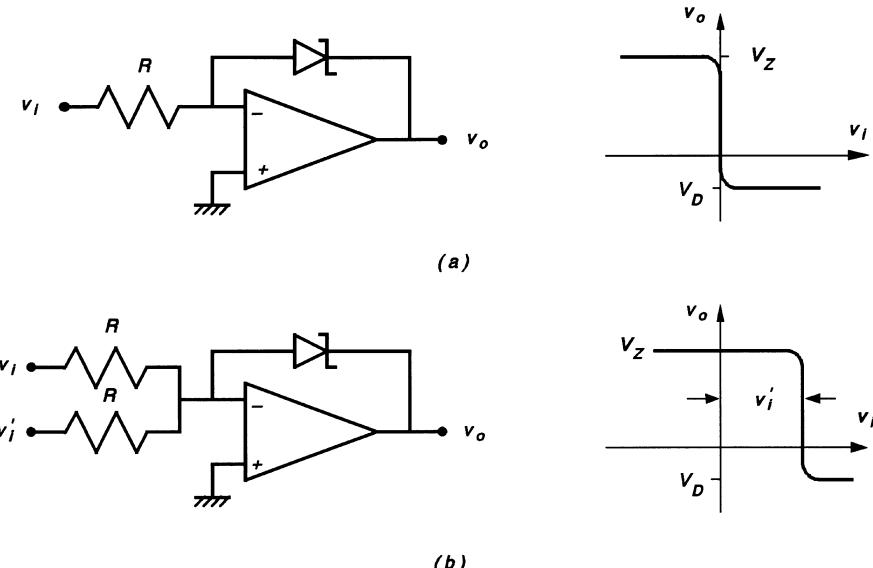


Figure 8.39. (a) The zero-voltage comparator and its output characteristics. (b) The comparator for v_i and v'_i and its output characteristics.

V_z of the diode. Evidently the clamping voltages were chosen so that a $\mathbf{0} \approx -0.7$ volts and $V_z \approx V^+ = 1$ – the system voltage. The input–output characteristics are given in Figure 8.39(a).

Now consider the circuit in Figure 8.39(b) where there are two input voltages v_i and v'_i connected by equal resistances R to the inverting input of the operational amplifier. Evidently, the only way to keep the output from going positive or negative (clamped condition) is that the current supplied by v_i must be equal to the current absorbed by v'_i . The two voltages must be equal in magnitude. When they are not equal, which they are not most of the time, the output is clamped at one extreme or the other ($\mathbf{1}$ or $\mathbf{0}$). The input–output characteristics show the effect of v'_i as an offset voltage. In the context of the A/D converter, the output of the comparator changes state when the output of the resistive network exceeds the analog input voltage by, at the most, the value of the LSB.

8.5.1.4 Parallel-to-Serial Converter. The output of the A/D converter, which is in pulse-code modulation (PCM parallel), has to be converted into serial form so that it can be sent along a single communication channel. A *universal asynchronous receiver/transmitter* (UART) performs this function. This is an elaborate, large-scale-integration (LSI) integrated circuit which can convert data from parallel to serial and vice versa. However, the basic operation of an eight-digit parallel-to-serial converter is illustrated in Figure 8.40. The parallel information is available at the nine (eight plus the start/stop) terminals. The moving switch sweeps past and “reads” the bits in serial form. Note that one or more of the bits can be used for the

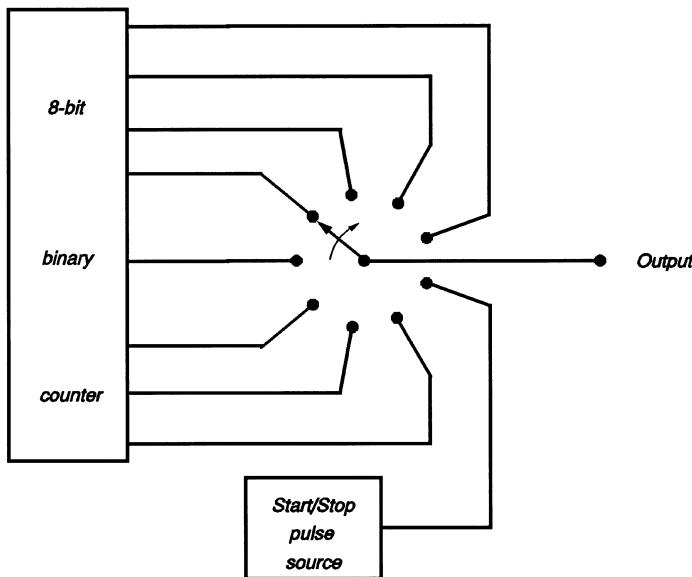


Figure 8.40. A mechanical analog of the parallel-to-serial converter.

“stop/start” code to tell the receiver when a segment of the code starts and stops. Other bits may be used as a parity code for correcting the signal code in the presence of noise. Examples of parallel-to-serial shift registers are the MC14014B (Motorola Semiconductor), the MM54HC165 (National Semiconductor Corp.) and the SN54166 (Texas Instruments) integrated circuits and are listed in most integrated circuit handbooks.

8.5.1.5 Serial-to-Parallel Converter. The operation of the serial-to-parallel converter is much simpler than its opposite number. Figure 8.41 shows an eight-stage serial-to-parallel converter.

The stop/start pulse is used to clear the register and to trigger the input enable circuit after a specified delay. The serial input data can then start to enter the shift register moved on by the clock. The arrival of the next stop/start pulse signals the end of the data stream. The binary outputs are read and the shift register is cleared for the next stream of data to arrive. An example of an integrated circuit serial-to-parallel converter is the SN54164 (Texas Instruments).

8.5.1.6 4-to-2 wire Hybrid. The basic 4-to-2 wire hybrid is shown in Figure 8.42.

There are two identical but separate transformers, T_1 and T_2 , connected to conform to the dot notation as indicated. When the number of turns in each coil

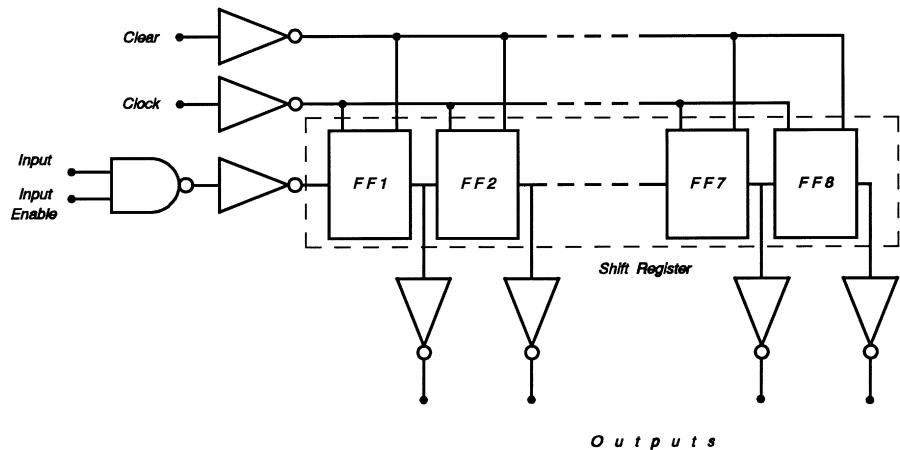


Figure 8.41. A block diagram of the serial-to-parallel converter.

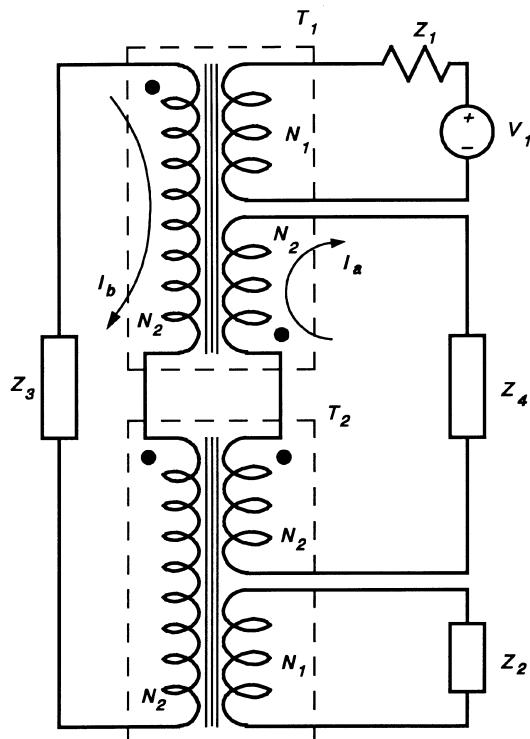


Figure 8.42. The 4-to-2 wire hybrid. Reprinted with permission from, *Transmission Systems for Communication*, 5th Ed., AT&T, Bell Labs, 1983.

is as shown, a voltage source V_1 in series with an impedance Z_1 connected to port 1 will cause a voltage equal to

$$\frac{N_2}{N_1} V_1 \quad (8.5.5)$$

to appear across the other two coils of T_1 , each of which has N_2 turns. Assuming that $Z_1 = Z_2$ and $Z_3 = Z_4$, it follows that the current I_a will be equal to I_b and they will flow in the directions indicated. In transformer T_2 , the magneto-motive force (mmf) of the two current will cancel each other and the net voltage across port 2 will be zero. The signal input power from port 1 is divided equally between ports 3 and 4. If the voltage source had been connected across port 2 instead of port 1, the same effect would be observed except there will be no voltage across port 1.

Consider the situation when a voltage source is connected across port 3. The voltages across the two halves of port 4 will cancel each other. However, voltages

$$\frac{N_1}{2N_2} V_3 \quad (8.5.6)$$

will appear across ports 1 and 2.

This arrangement can be used as a station set hybrid if the transmitter is connected across port 1, the receiver across port 2, and the line across port 3 so long as an impedance equal to that of the line is connected across port 4. Its normal application, however, is as a 4-to-2 wire hybrid, the four wires being connected to ports 1 and 2 and the two wires being connected to port 3.

8.6 THE CENTRAL OFFICE

8.6.1 Manual Office

The manual central office is largely historical although one sometimes hears stories of a few survivors in very remote places. A description of its operation was given in Chapter 1. It is evident that it was just a matter of time before automation would eliminate the telephone operator's job. The surprise was that it was not the human errors due to voice communication problems, delay in getting service, or rising labor costs that was the motivation for automation; it was the lack of security of the message (see Section 1.5).

8.6.2 Basics of Step-by-Step Switching

As described in Section 8.3.6.2, the request for service from the telephone subscriber to the central office comes in the form of a series of current pulses. These current pulses are at a frequency of 10 Hz with a variable, but certainly much longer, inter-

digit pause. The engineering problem was to use these current pulses to move a mechanical switch into the required position to effect the connection.

It is most convenient to start a description of the process with a small and fictitious central office with 10 subscribers, each with the pulse dial described in Section 8.3.6.2. Under these conditions, switching would take place when a single string of dial pulses arrived from the calling subscriber. Figure 8.43 shows an electromechanical contraption that would perform the switching function. The serrated wheel is attached to the switch wiper arm and it is spring loaded to keep it in its rest position as shown.

When a pulse is applied to the electromagnet *A*, the pawl pulls the serrated wheel the appropriate distance to cause the switch wiper arm to move from one contact to the next. The wheel is prevented from returning to its rest position after every pulse by the detent and the electromagnet *B*.

Consider the circuit shown in Figure 8.44. To initiate a call, the subscriber lifts the handset, thereby closing the hook switch. Direct current flows in the coil of electromagnet *B* and operates the detent to stop the serrated wheel from rotating in a clockwise direction; it is, however, free to rotate in an counter-clockwise direction. The subscriber now operates the dial and short-circuits the line seven times, say. The ac current generated flows through electromagnet *A* and causes it to operate seven times, moving the switch wiper arm onto the seventh fixed contact. There is no known telephone system that operates this way! The conversation can now proceed. When the call is over, the replacement of the handset opens the dc current loop, the

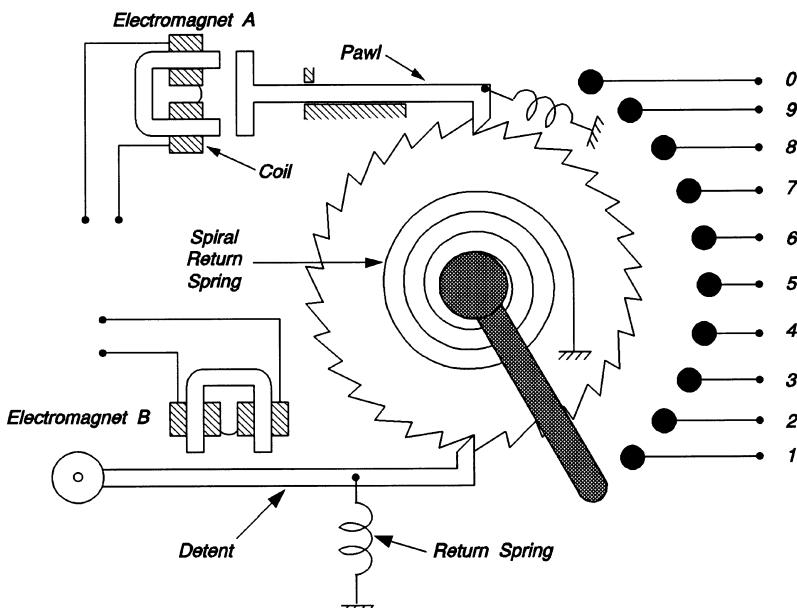


Figure 8.43. An electromechanical switch for a ten-subscriber central office.

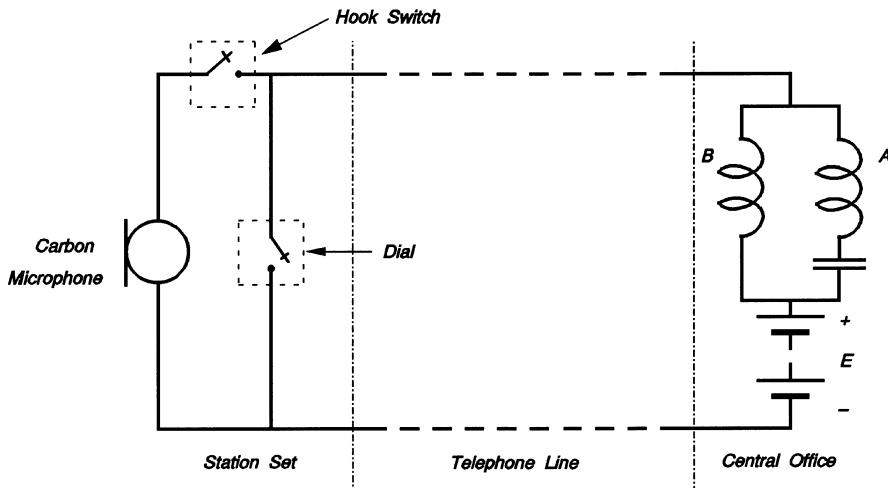


Figure 8.44. The switch shown in Figure 8.43 connected to the station set.

electromagnet B is de-energized, the detent moves away from the serrated wheel and, under the force of its return spring, the switch wiper arm returns to its rest position.

Each of the ten subscribers would be connected to such a switch and all ten switches connected as shown in Figure 8.45.

A switching system such as this could be extended to accommodate 100 subscribers by making the following changes:

- (1) Dialling will now consist of two trains of pulse with a suitable pause between pulse trains.
- (2) Each of the 10 selector switches will be connected to 10 other identical (secondary) switches.

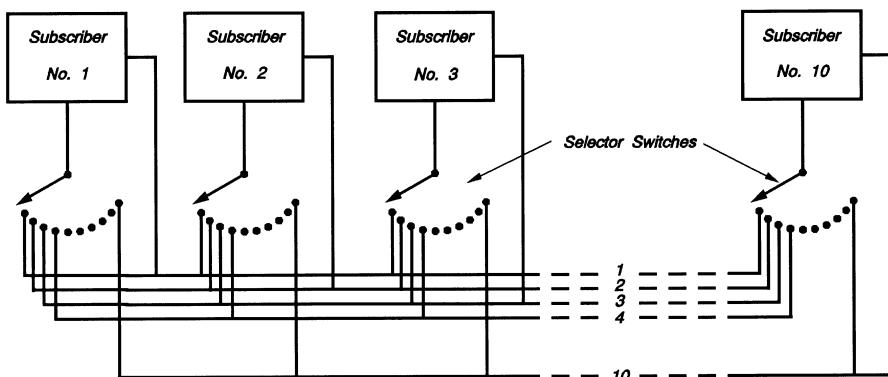


Figure 8.45. The complete ten-subscriber telephone system.

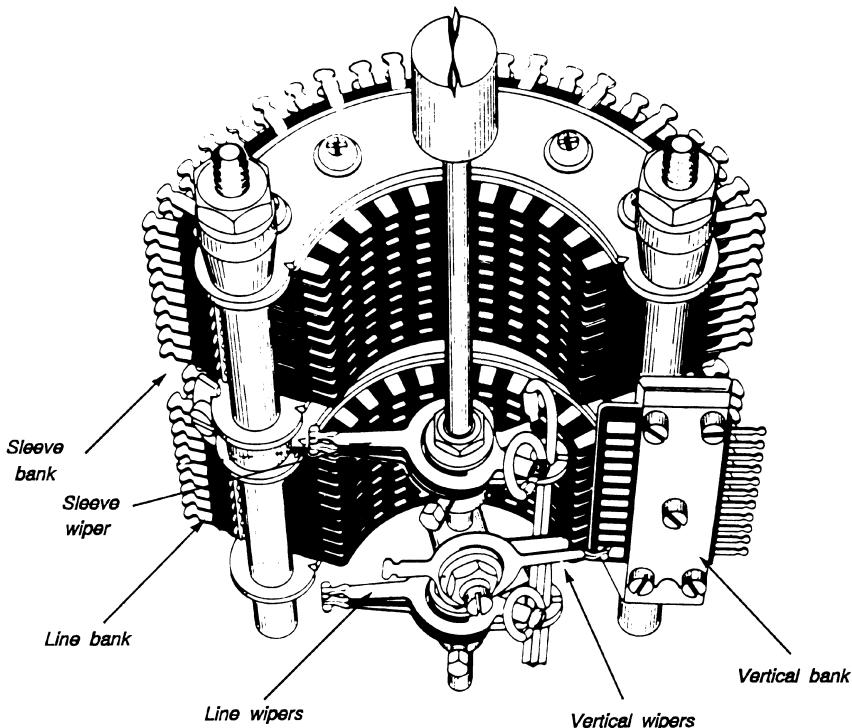


Figure 8.46. The Strowger switch showing horizontal and vertical banks as well as wipers. Reprinted with permission from B. E. Briley, *Introduction to Telephone Switching*, AT&T, Bell Labs, 1983.

- (3) Some electrical or mechanical arrangements will be made to transfer the second train of pulses to the bank of secondary switches.

Further extension to 1000 and 10,000 subscribers is possible by suitable changes in the central office and the introduction of three- and four-digit dialling codes respectively. Such a system is possible and would operate quite well, except for the fact that it would be very large, prohibitively expensive to manufacture, install, and operate.

8.6.3 The Strowger Switch

If you can imagine a new version of the fictitious electro-mechanical contraption shown in Figure 8.43 with movement in two directions, then you have a good idea of how the Strowger switch works. The actual Strowger switch is shown in Figure 8.46.

There are two sets of electromechanical drives: one vertical and the other horizontal. There are ten vertical positions and ten horizontal ones. Such a switch can accommodate up to 100 subscribers and therefore requires two trains of pulse.

The first moves the switch wiper upwards – to the third level if there were three pulses in the train. During the inter-digit pause, control is switched to the horizontal drive. The second train of pulse move the switch wiper horizontally to the seventh position – if it had seven pulses, to connect to the number 37. At the end of the conversation, the switch must be returned to its initial position.

Replacing the 10-position switches shown in Figure 8.45 with Strowger switches would convert a 10-line office into a 100-line office. It would appear at first that the two-dimensional switch would solve the economic problem; in fact, it is more complicated than the 10-position switch and therefore more expensive. The solution to the economic problem was discovered along with the fact that it is most unlikely that in a 100-line office 50 calls will be in progress at the same time. It was therefore not necessary to design the system for maximum capacity operation. The study of traffic statistics revealed that the system could be designed to handle a maximum number of calls (usually about 10% of subscribers) and if it happens that there is one more than this maximum then service would be denied. This led to a redesign of the central office so that the subscribers would share a considerable amount of common equipment. Every new incoming call initiates a *hunt* for an idle path through the system and, if no idle paths can be found, a busy signal is returned. The precise details of the inner workings of the central office are not necessary since the object of the exercise is to develop a general appreciation of how a call is connected. Figure 8.47 shows a simplified connection using line finders, selectors, and connectors.

8.6.4 Basics of Crossbar Switching

The basis of the crossbar switching system is a fallback to the manual central office and is shown in Figure 8.48.

In a hypothetical nine-subscriber central office, each subscriber is represented in the vertical and horizontal elements of the matrix. To connect any two of the subscribers it is necessary to have 81 contact points. In the crossbar system, it actually takes the operation of two relays to close one contact point – one for the vertical element and the other for the horizontal element. As in the other central office organization schemes, it is possible to reduce the number of contact points by:

- (1) Reorganizing the interconnection as shown in Figure 8.49(a). This shows that the number of contacts have been reduced to 54.
- (2) Since it is most unlikely that every subscriber will be using her telephone at the same time, let the maximum number of conversations passing through this office be three. By using three links as shown in Figure 8.49(b), it is still possible to connect every subscriber to every other subscriber so long as no more than three calls are in progress simultaneously.

The process of using these and similar schemes to reduce the number of switches is referred to as *concentration*.

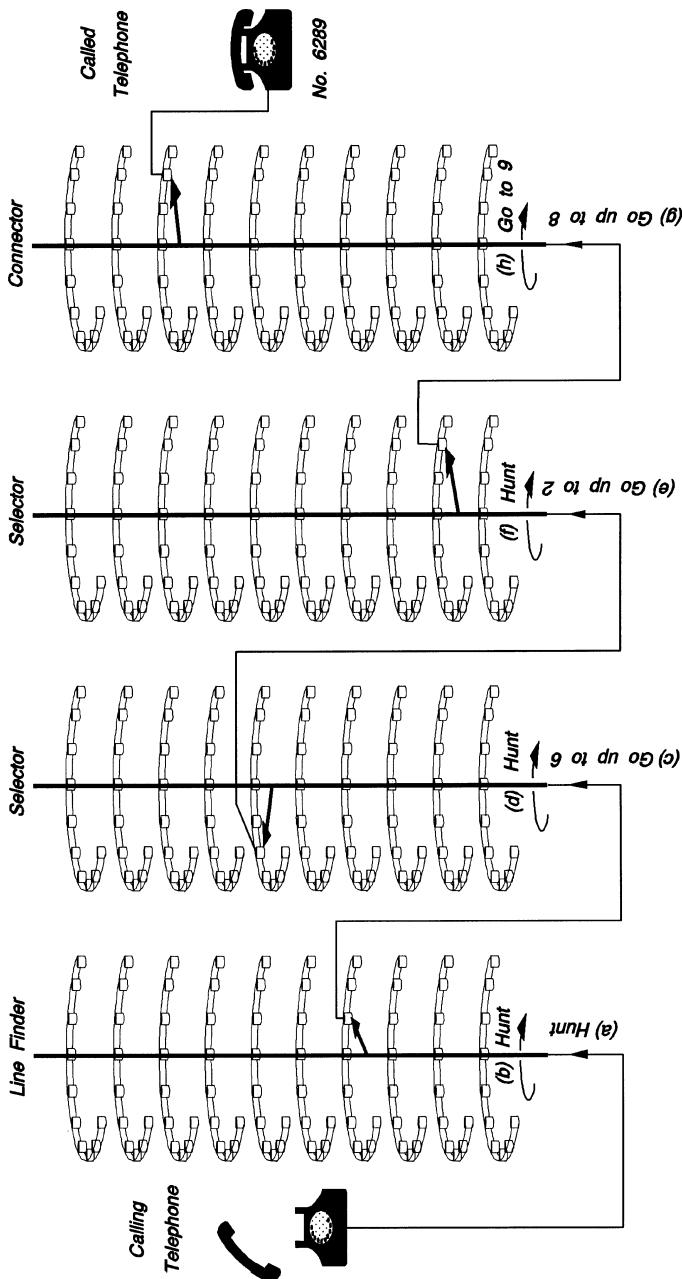


Figure 8.47. A simplified connection through the step-by-step central office. Reprinted with permission from B. E. Briley, *Introduction to Telephone Switching*, AT & T, Bell Labs, 1983.

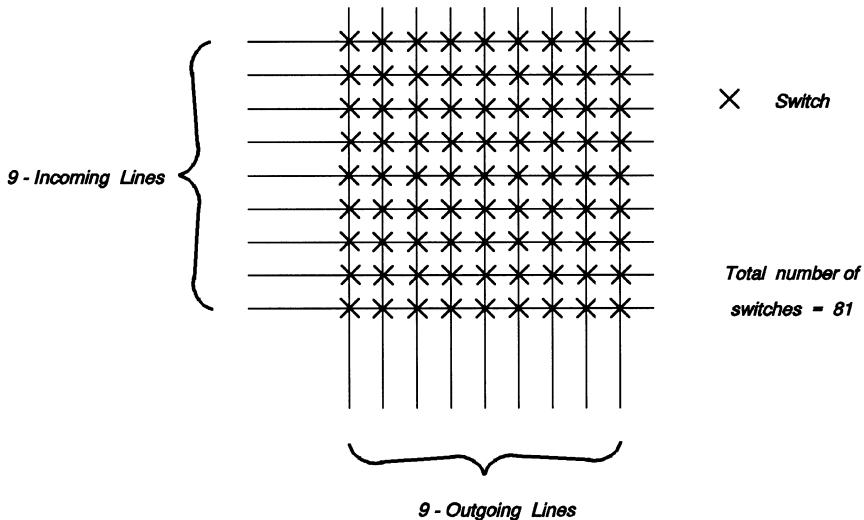


Figure 8.48. A representation of the cross-bar switch for a nine-subscriber central office showing all 81 possible connections. Reprinted with permission from S. F. Smith, *Telephony and Telegraphy A*, 2nd Ed., Oxford University Press, London, 1974.

8.6.5 Central Office Tone Receiver

So far the discussion of switching in the central office has been based on the 10 Hz pulse signal generated by the rotary dial or imitations of it. The use of tone dialling in the telephone system is becoming increasingly more important because it is faster and it offers the possibility to use the telephone system for purposes other than talking. There are already a number of household appliances and office equipment on the market which can be connected to the telephone system and controlled from a remote location by dialling special codes.

As discussed earlier, the tone dial generates two distinct frequencies when a single button is pressed. By identifying the two frequencies present the number of the button can be decoded. The system for decoding the signal is shown in Figure 8.50.

The tone signal goes through a band split filter which separates the low-frequency group from the high one. Each group is amplified and hard limited to produce square waves. The signal is filtered again by the channel filters which are tuned to the frequencies indicated. The output signals from the channel filters are rectified by the detectors to indicate the presence (**1**) or absence (**0**) of the frequencies produced by the dial. For an output to be valid, it has to have two frequencies present, one from each frequency group. The system has several built-in features which prevent mistaken identification of other audio signals as valid dial signals.

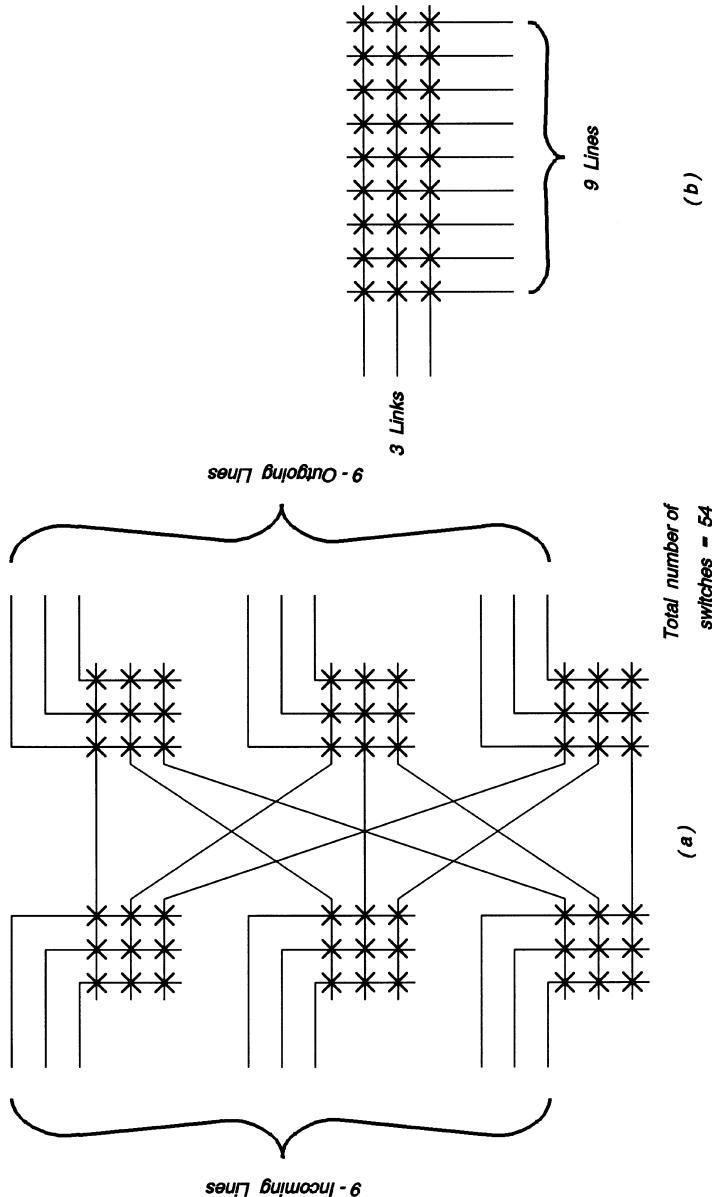


Figure 8.49. (a) A switching scheme for a nine-subscriber central office with the number of contacts reduced to 54. (b) The use of three links which reduces the number of conversations that can be carried out simultaneously to three also reduces the number of contacts to 27. Reprinted with permission from S. F. Smith, *Telephony and Telegraphy A*, 2nd Ed., Oxford University Press, London, 1974.

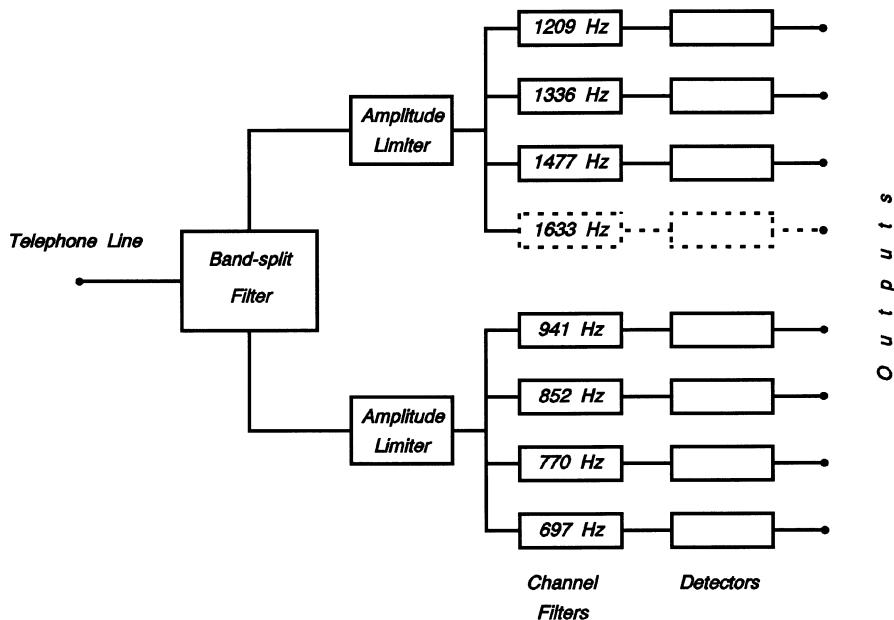


Figure 8.50. The block diagram of the tone receiver at the central office.

8.6.6 Elements of Electronic Switching

The basis of electronic switching is the same as the crossbar technique but with the electromechanical switches replaced by electronic components. The advantages gained were speed of operation, increased reliability and lower cost. During its development, several schemes were tried but none of them worked because the replacement of the metal-to-metal contact in the electromechanical switches with semiconductor switches proved to be unsuitable. The concept of electronic switching had to undergo a fundamental change before success could be achieved. The problem was that the crossbar technique was based on *space-division multiplex* (SDM) – an approach that relied on finding multiple paths for the different conversations passing through the central office by separating them in space, that is, by assigning each path to a space or position in the matrix. *Time-division multiplex* (TDM), which uses the same path for all conversations but separates them in time, was more successful. Time-division multiplex is discussed in Chapter 9.

So far, in the discussion of what takes place in the central office from the initiation of a call to its end, it has been assumed that the system was under the direct control of the pulses or tones as they arrived. It is certainly advantageous to have a memory in the system so that the information dialled in by the subscriber can be stored, analyzed and an optimal path determined for the call. In the language of the telephone engineer, the memory device is a register (see Appendix E). The pulses from the dial are fed to a set of shift registers or decade counters, one set for each decade of the number dialled. With the tone dial, the digits are first decoded and then

stored in binary form in a shift register. What we have then are banks of hard-wired switches and a memory in which instructions can be stored, retrieved, and acted upon in due course. To help drive this point home, consider the activities which take place in the central office in order to connect a call. This is shown in the form of a flow diagram in Figure 8.51.

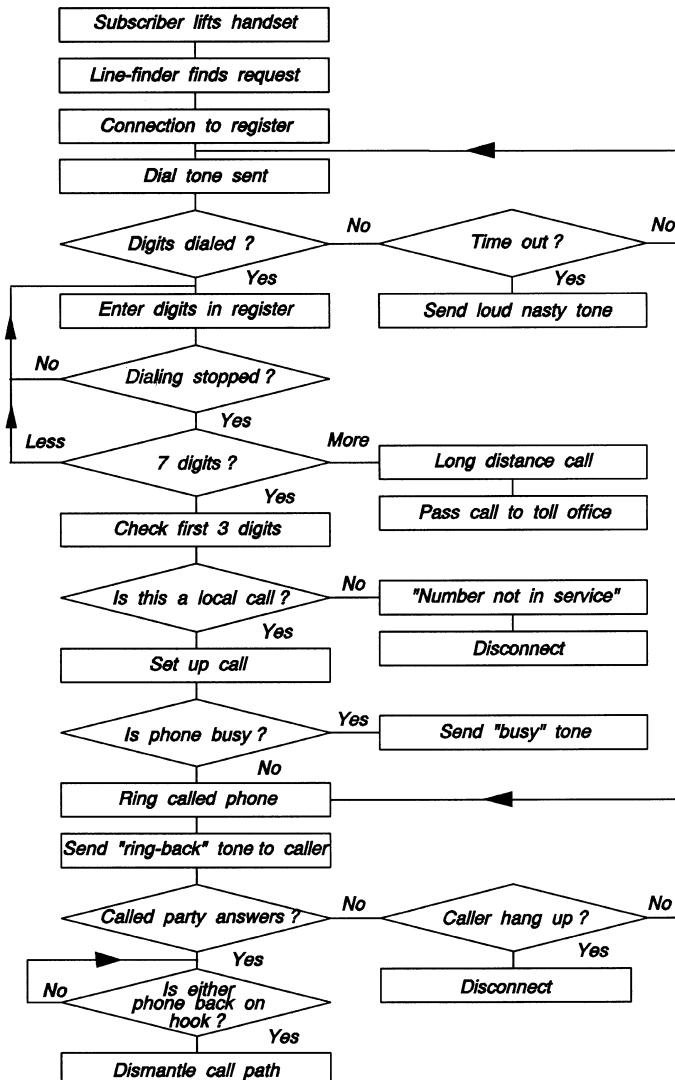


Figure 8.51. A simplified flow chart for setting up a telephone call. The similarity to a computer program flow chart is evident.

The flow diagram looks suspiciously like the flow diagram for a computer program. It is indeed a computer program and the instructions can be carried out by a general-purpose computer with the appropriate peripheral equipment. The modern central office has a dedicated computer which controls every aspect of placing a call. Humans simply supervise the machine and intervene only when something goes wrong or when they work out a better set of instructions for the machine.

REFERENCES

1. Fraser, W., *Telecommunications*, Macdonald & Co., London, 1957.
2. Plonus, M. A., *Applied Electromagnetics*, McGraw-Hill, New York, 1978.
3. Sadiku, M. N. O., *Elements of Electromagnetics*, Holt, Rinehart and Winston, New York, 1989.
4. Sessler, G. M. and West, J. E., "Electret Transducers: A Review", *J. Acoust. Soc. Am.*, 53, 1589–1599, 1973.
5. van der Puije, P. D., "Audio Frequency Circulator for use in Telephone Sets", *IEEE Trans. Comm. Tech.*, 1267–1271, Dec. 1971.
6. Tellegen B. D. H., "A General Network Theorem with Applications", *Philips Res. Rept.*, 7, 259–269, 1952.
7. Shenoi, B. A., "Practical Realization of a Gyrator Circuit and RC-gyrator Filters", *Trans. IEEE*, CT-12, 374, 1965.
8. Sheahan, D. F. and Orchard, H. J., "High Quality Transistorized Gyrators", *Elec. Lett.*, 2, 274, 1966.
9. Smith, S. F., *Telephony and Telegraphy A*, 2nd Ed., Oxford University Press, London, 1974.
10. Briley, B. E., *Introduction to Telephone Switching*, Bell Telephone Laboratories, Inc., 1983.
11. Members of the Technical Staff., *Transmission Systems for Communication*, 5th Ed., Bell Telephone Laboratories, Inc., 1983.

PROBLEMS

- 8.1** Describe the attractive features of the carbon microphone that make it well suited to the telephone system. A carbon microphone has a static resistance of $75\ \Omega$ and it is connected in series with a 12 V battery and a load resistor R . A sound wave impinging on the microphone causes a sinusoidal variation of the microphone resistance with a peak value equal to 15% of the static value. What is the value of the load resistor R if the second harmonic component of the current is 7.5% of the current at the fundamental frequency? Calculate the signal power delivered to R at the fundamental frequency and comment on it.
- 8.2** A successive approximation A/D converter has an output consisting of 8 bits. The input to the converter is $0.82564\ V_R$, where V_R is the reference voltage. The A/D converter is clocked at 1 MHz.

- (1) Determine the total error in the conversion.
- (2) Calculate the acquisition time.

If a change is made in the A/D converter so it counts from zero and advances by the value of the LSB for every clock pulse (ramp counter A/D converter), what would be the conversion error in the time given by (2) when the $\text{LSB} = 0.001 V_R$? How long will it take the ramp counter A/D to achieve the same conversion accuracy as the successive approximation A/D converter?

- 8.3 Design a bipolar transistor bistable multivibrator using a 12 V dc power supply. The multivibrator is to drive a load which requires 1 mA. Indicate how you would convert the multivibrator into a flip-flop. Show that the flip-flop performs a divide-by-two function.
- 8.4 Using standard digital gates, design a circuit which will divide a clock frequency by the following integer numbers: 7, 19, 25, 47, 77, and 92.
- 8.5 What is a shift register and how does it work? Illustrate your answer with a design of a four-stage shift register. Describe two applications where a shift register may be used.
- 8.6 What is sidetone and how can it be used to cultivate good speaking habits in telephone users? Does sidetone control offer any other advantages? With the aid of a suitable circuit diagram, describe the operation of a typical sidetone suppression circuit.

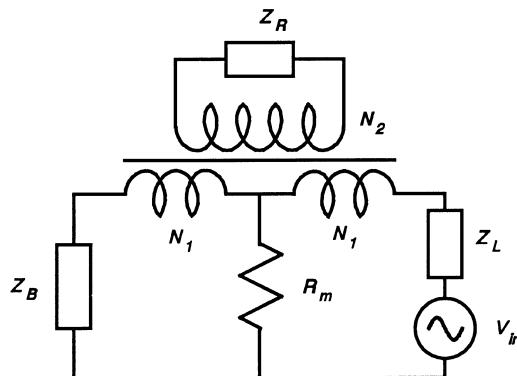


Figure P8.1.

The circuit of Figure P8.1 shows a transformer hybrid in the receive mode where R_m is the resistance of the microphone. Assuming that $Z_L = Z_B = 900 \Omega$, $Z_R = 1000 \Omega$, and $N_1 = N_2 = 200$ turns, calculate the power dissipated in Z_R when $V_{\text{in}} = 1 \text{ V rms}$ and the reactance of N_1 is 1200Ω at the frequency of operation. [Hint: transfer the load Z_R to the primary of the transformer].

- 8.7** In the design of digital circuits it is common to find two NOT gates connected in cascade. What is the purpose of such a connection?

Using only two-input NOR gates, show how you can realize a:

- (1) NOT gate,
- (2) two-input OR gate, or
- (3) two-input AND gate.

9

SIGNAL PROCESSING IN THE TELEPHONE SYSTEM

9.1 INTRODUCTION

Until the introduction of the digital telephone, there was virtually no signal processing on the subscriber loop. Indeed, there was no need for it. The majority of subscriber loops were able to transmit voice signals with no particular difficulty and in cases where the lines were longer than usual, line “loading” was used with success.

Signal processing has two major aims:

- (1) To improve the quality of signal transmission over the telephone communication channels.
- (2) To lower the cost of communication by improving the efficiency of channel use.

In general, the quality of a communication channel tends to deteriorate with distance. In addition, long distance channels are expensive to establish and maintain. It follows that the more messages that can be transmitted in a given time, the lower the cost per message. It is therefore on the long-distance channels (trunks or tolls) that signal processing techniques have proven to be most successful. In this chapter the common signal processing techniques used in the telephone system and some of the circuits employed will be examined.

9.2 FREQUENCY DIVISION MULTIPLEX (FDM)

Frequency division multiplex (FDM) is a technique in which a number of signals can be transmitted over the same channel by using them to modulate carrier signals with

different and appropriate frequency so that they do not interfere with each other. The assignment of specific carrier frequencies to radio stations for broadcasting and other purposes is, in fact, FDM. It can be used with amplitude modulation as well as other forms of modulation. In the context of the telephone, FDM is used in conjunction with amplitude modulation.

In normal amplitude modulation, the carrier, upper and lower sidebands are transmitted. When the depth of modulation is 100%, the amplitude of the carrier voltage is twice that of the sidebands. The power in the carrier is therefore $\frac{2}{3}$ of the total. Unfortunately, the carrier has no information content. Each of the sidebands contains $\frac{1}{6}$ of the total power. In radio, 100% modulation is almost never used so the power content of the sidebands is much less than described. It is noted that the information content is duplicated in the two sidebands. It is clear that one way to beat the corrupting influence of noise on the information content of the transmission is to put as much as possible, if not all, of the available power into one of the sidebands. An added advantage to this scheme is that the required bandwidth is reduced to one-half of its original value. Clearly, this would allow twice as many messages to be sent on the same channel as before. The transmission of only one sideband in an AM scheme is called *single-sideband* (SSB) modulation. The price to be paid for this advantage is that to demodulate an SSB signal, it is necessary to reinstate the carrier at the receiver. The reinstated carrier has to be in synchronism with the original carrier, otherwise demodulation yields an intolerably distorted signal. Providing a synchronized local oscillator requires complex equipment at the transmitter as well as at the receiver. In SSB radio, an attenuated form of the carrier is transmitted with the signal. This is used to synchronize a local oscillator in the receiver. In the telephone system, a centrally generated pilot signal is distributed to all offices for demodulation purposes. In some cases, a local oscillator without synchronization is used. If the frequency error is small (approximately ± 5 Hz), successful demodulation can be achieved [7].

9.2.1 Generation of Single-Sideband Signals

A block diagram of the SSB generator is shown in Figure 9.1.

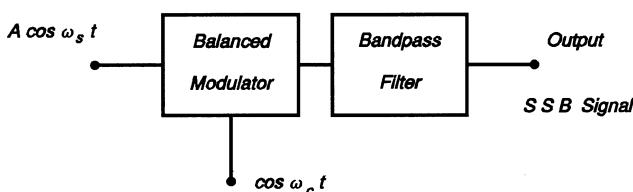


Figure 9.1. The block diagram for SSB modulation.

The signal and the carrier are essentially multiplied by the balanced modulator to give a DSB-SC output. The bandpass filter removes either the lower or upper sideband.

$$f(t) = A \cos \omega_s t \cos \omega_c t \quad (9.2.1)$$

$$f(t) = \frac{A}{2} [\cos(\omega_c + \omega_s)t + \cos(\omega_c - \omega_s)t]. \quad (9.2.2)$$

Assuming the upper sideband is eliminated, we get

$$f_1(t) = \frac{A}{2} \cos(\omega_c - \omega_s)t. \quad (9.2.3)$$

To eliminate the upper sideband, it is necessary to have a bandpass filter with a very sharp cut-off at the carrier frequency. This is not easy to achieve in practice, but the task is made simpler when the modulating signal ω_s has no low-frequency components. Under this condition, crystal and electromechanical filters can be designed to suppress the upper sideband. This is the case for a telephone voice channel which is nominally from 300 to 3000 Hz.

From Equation (9.2.3) only the lower sideband is transmitted. At the receiving end, the signal is demodulated (multiplied) by (the recovered) carrier, $\cos \omega t$. The result is

$$f_2(t) = \frac{A}{2} \cos(\omega_c - \omega_s)t \cos \omega_c t \quad (9.2.4)$$

$$f_2(t) = \frac{A}{4} \cos \omega_s t + \frac{A}{4} \cos(2\omega_c - \omega_s)t. \quad (9.2.5)$$

A low-pass filter is used to separate the required signal at frequency ω_s from that at $(2\omega_c - \omega_s)$.

9.2.2 Design of Circuit Components

The balanced modulator was discussed in Section 4.4.2.3. Filter design is outside the scope of this book but a representative list of books on filters is provided in the bibliography at the end of Chapter 3.

9.2.3 Formation of a Basic Group

In the trunk or toll system, 12 channels form a basic group. The basic group is formed by SSB modulation of 12 subcarriers at 64, 68, 72, ..., 108 kHz. These carriers are generated from a 4 kHz crystal-controlled oscillator and multiplied by the appropriate factor. The upper sidebands are removed and they are added together to form the group. Figure 9.2(a) shows a block diagram for channel 1. Figure 9.2(b) shows the spectrum of the basic group.

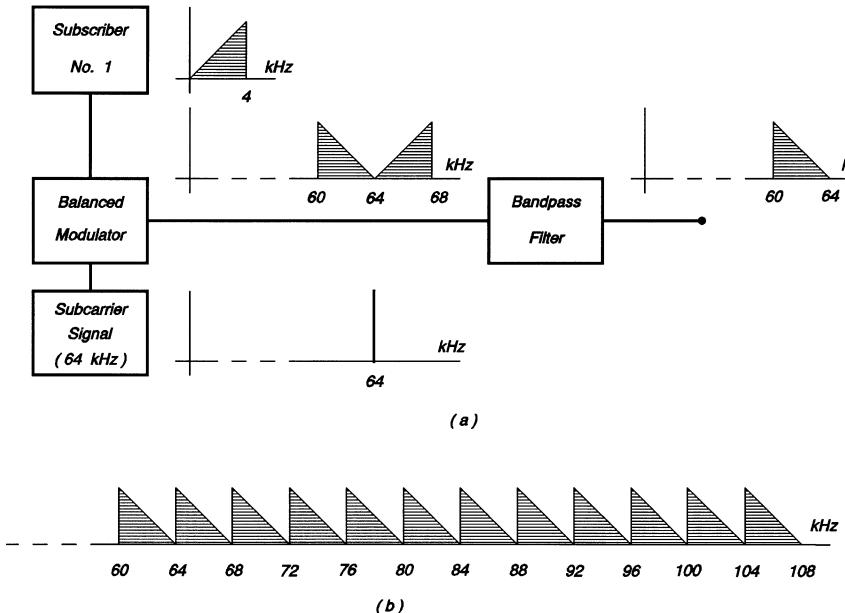


Figure 9.2. Formation of the basic group with spectra. Reprinted with permission from *Transmission Systems for Communications*, 4th Ed., AT&T, Bell Labs, 1970.

For a small-capacity trunk, the basic group may be transmitted without further processing. The transmission channel can be a twisted pair or coaxial cable.

9.2.4 Formation of a Basic Supergroup

For higher capacity channels, five basic groups are combined to form a basic supergroup. Figure 9.3(a) shows the block diagram of the basic supergroup 1. Note that to make the filtering problem easier, the carrier frequency is chosen to be 420 kHz. Figure 9.3(b) shows the frequency spectrum of the basic supergroup.

Table 9.1 shows the carrier frequencies and bandwidths for each basic supergroup.

For a 60-channel trunk, the signal can be transmitted in this form. Again a twisted pair with coil loading or amplification and coaxial cable may be the medium of transmission.

By organizing the 12 basic groups into a basic supergroup of 5 it clear that the subcarrier frequencies, the balanced modulators, and bandpass filters can all be duplicated five times over. If the basic group had been made larger, new subcarrier frequencies would have had to be generated and bandpass filters of different characteristics would have been necessary.

9.2.5 Formation of a Basic Mastergroup

To create a 600-channel trunk, 10 basic supergroups are combined to form a basic mastergroup. The frequency spectrum of the basic mastergroup is shown in Figure 9.4.

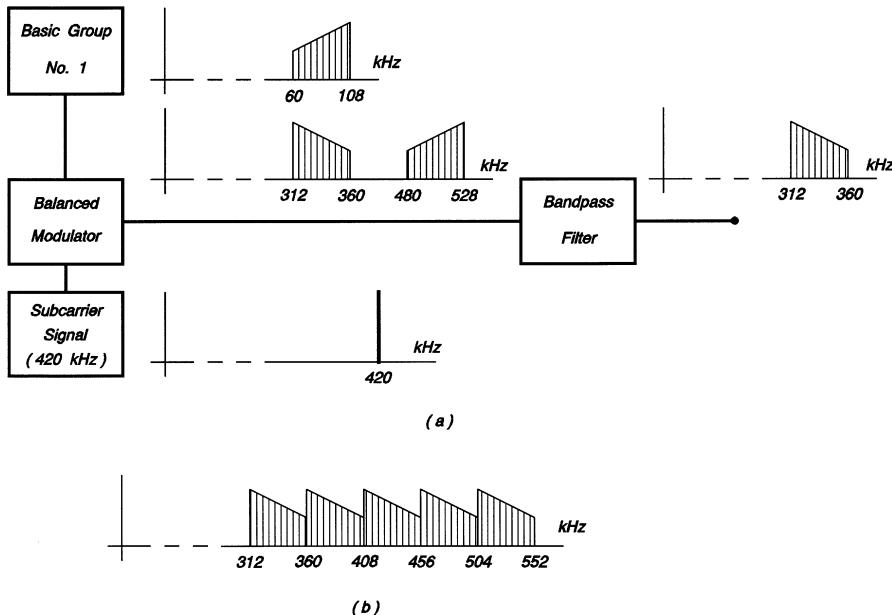


Figure 9.3. Formation of the basic supergroup with spectra. Reprinted with permission from *Transmission Systems for Communications*, 4th Ed., AT&T, Bell Labs, 1970.

Note that there are gaps of 8 kHz between each basic supergroup spectrum. These gaps are designed to make the filtering problem easier.

The carrier frequencies and bandwidths of the 10 basic supergroups are given in Table 9.2.

The basic supergroup can be transmitted over coaxial cable or it can be used to modulate a 4 GHz carrier for terrestrial microwave transmission or even sent over a satellite link.

TABLE 9.1

Supergroup number	Carrier frequency (kHz)	Bandwidth (kHz)
1	420	312–360
2	468	360–408
3	516	408–456
4	564	456–504
5	612	504–552

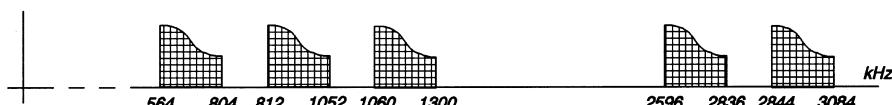


Figure 9.4. Formation of the basic mastergroup with spectra. Reprinted with permission from *Transmission Systems for Communications*, 4th Ed., AT&T, Bell Labs, 1970.

TABLE 9.2

Basic supergroup number	Carrier frequency (kHz)	Bandwidth (kHz)
1	1116	564–804
2	1364	812–1052
3	1612	1060–1300
4	1860	1308–1548
5	2108	1556–1796
6	2356	1804–2044
7	2652	2100–2340
8	2900	2348–2588
9	3148	2596–2836
10	3396	2844–3084

Other larger groups can be formed, for example 6 mastergroups may be combined to form a *jumbogroup* with 3600 voice channels.

To recover the original baseband signals from the various groups, the appropriate number of filtering/demodulation processes will have to be carried out. At each stage of the demodulation process, the correct carrier will have to be reinstated for this to be possible.

9.3 TIME-DIVISION MULTIPLEX (TDM)

In FDM, voice signals were “stacked” in the frequency spectrum so that many such signals could be transmitted over the same channel without interference. In time-division multiplex (TDM), each voice signal is assigned the use of the complete channel for a very short time on a periodic basis. The theoretical basis of this technique is the *Sampling Theorem*. An informal statement of the sampling theorem is:

If the highest frequency in a signal is B Hz, then the signal can be reconstructed from samples taken at a minimum rate of $2B$ samples per second (*Nyquist sampling rate or frequency*).

The proof of this theorem is beyond the scope of this book. However, there are a number of practical problems which arise in the application of the theorem:

- (1) The theorem assumes that the samples have infinitesimally narrow pulse widths. This is clearly not so in a practical circuit. The sampling rate is usually chosen to be higher than the Nyquist frequency since it is the minimum; it is discrete to avoid extreme conditions when dealing with an imperfect situation.
- (2) The theorem assumes that an ideal low-pass filter is used to remove all frequencies above B Hz ahead of the sampler. When using a practical filter, it

is necessary to sample the signal at a higher rate (oversampling) to avoid distortion due to *aliasing*.

A TDM system with two input signals is illustrated in Figure 9.5. The *samplers* or *commutators* are shown here as switches which are driven in synchronism.

The TDM system shown in Figure 9.5 is an example of *pulse amplitude modulation* (PAM). Practical TDM systems based on PAM have been built and used in the telephone system (No. 101 ESS-PBX) [8].

9.3.1 Pseudodigital Modulation

To code an analog signal in pulse form one can use the height of the pulse, the width (or duration), or the position of the pulse relative to standard position. When the height is used, it is called *pulse-amplitude modulation* (PAM). When the coding is in terms of the width it is called *pulse-width modulation* (PWM) and when the position is used it is called *pulse-position modulation* (PPM). Pulse height, width, and position are analog quantities which in turn can be quantized and represented by a binary code where the digits are present, **1**, or absent, **0**. When this has been done the modulation scheme is called *pulse-code modulation* (PCM). Although PCM is qualitatively different from the other modulating schemes, they are compared in Figure 9.6.

These schemes would work equally well in a noiseless environment. When noise is present, and it always is, PCM has a clear advantage over the others. In the case of PAM, PWM and PPM the receiver has to determine what the original amplitude, width, and position were respectively in order to reconstruct them. In PCM, the decision is simplified to whether the digit sent was a **1** or a **0**. In all cases, it is necessary to transmit timing information with the signal so that the receiver knows where the bit stream starts and stops.

9.3.2 Pulse-Amplitude Modulation Encoder

To illustrate the design principle of a PAM communication channel, a four-channel PAM system has been chosen. The coder or commutator is shown in Figure 9.7.

The master clock drives the four-phase ring counter. The ring counter drives four sampling gates on and off in the correct sequence. When one of the four outputs is on, **1**, all the others are off, **0**, so only the sampling gate with the **1** is connected to the adder. Note that the second input to the fourth sampling gate are connected to the master clock. This means that channel 4 will always produce a positive pulse. The amplitude of this pulse is adjusted to be higher than the most positive value of the analog input voltage. This is called the *synchronization pulse* or *sync pulse* for short. It is used to identify and time the other channels.

The design of the component circuits now follows.

9.3.2.1 Four-Phase Ring Counter. The four-phase ring counter and its timing diagram are shown in Figure 9.8.

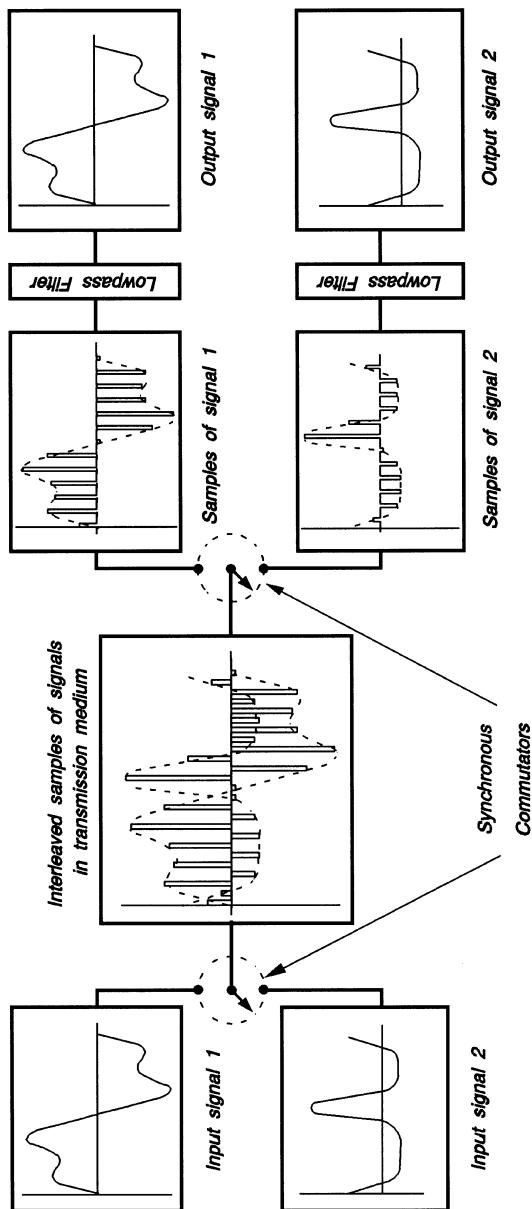


Figure 9.5. A mechanical illustration of time-division multiplex (TDM) with pulse amplitude modulation (PAM). Reprinted with permission from B. P. Lathi, *Modern Digital and Analog Communication Systems*, CBS College Pub., New York, 1983.

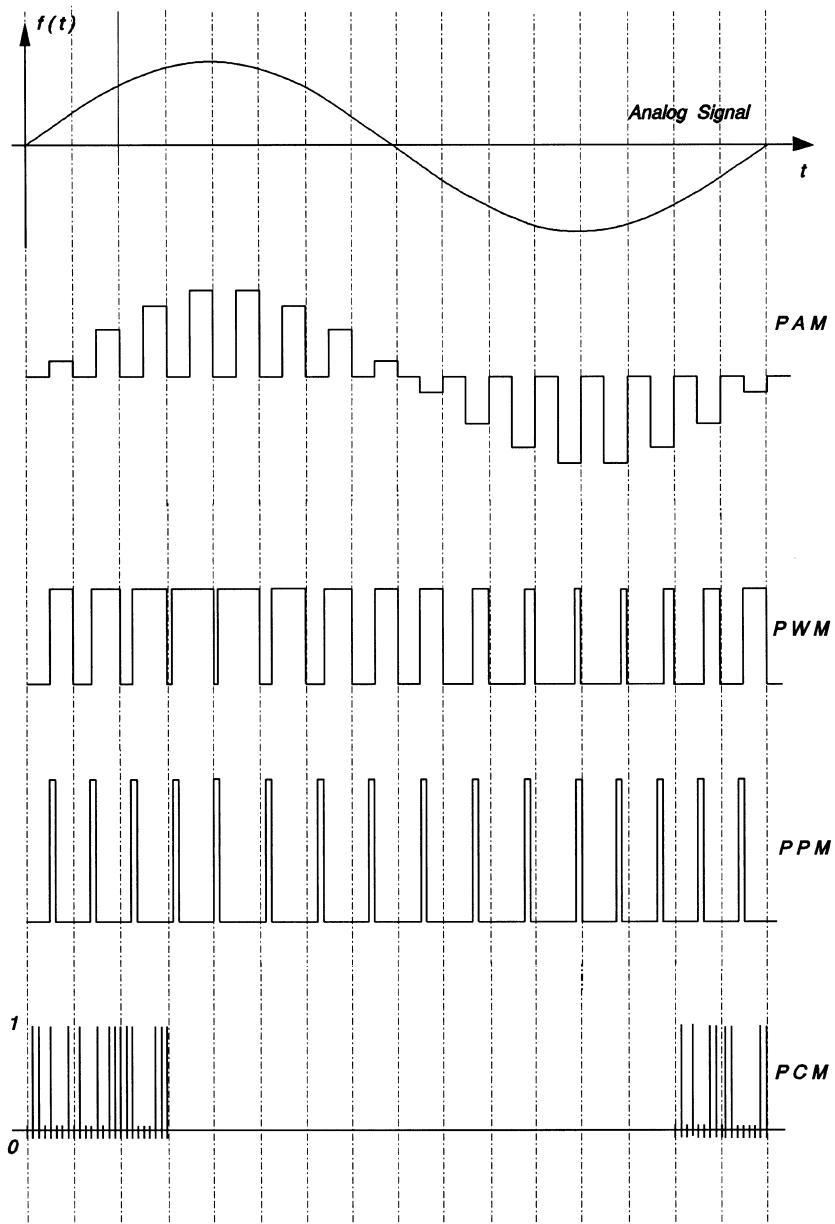


Figure 9.6. A comparison of PAM, PWM, PPM and PCM. Note that PAM, PWM and PPM are not truly digital since they convey information by the variation of analog quantities, that is, amplitude, duration and position in time. Reprinted with permission from B. P. Lathi, *Modern Digital and Analog Communication Systems*, CBS College Pub., New York, 1983.

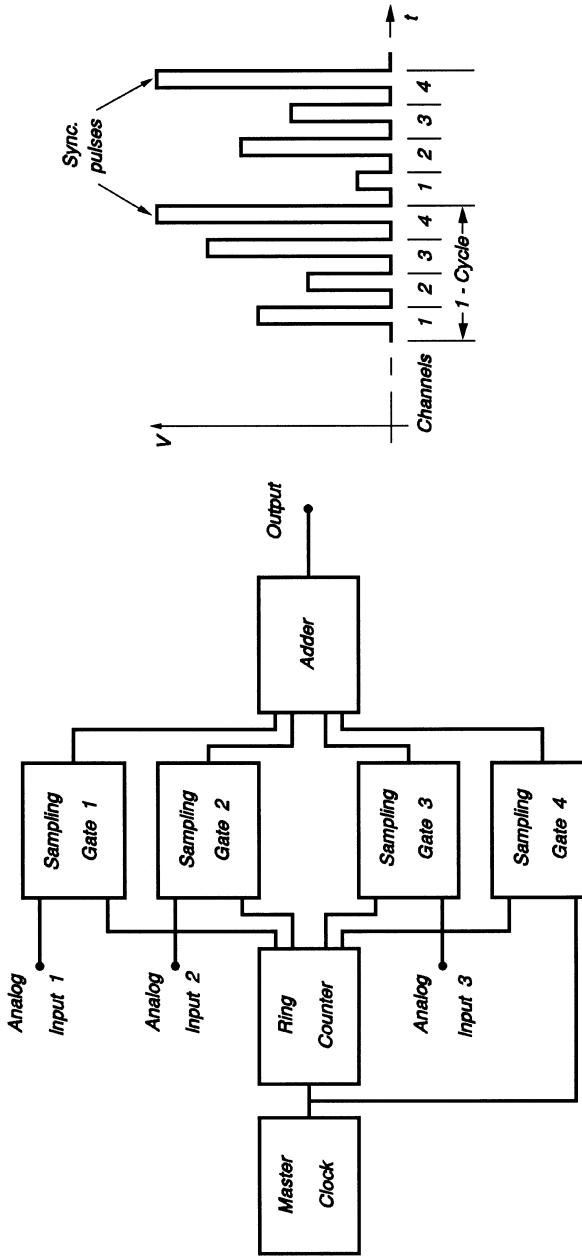


Figure 9.7. A block diagram for a four-channel PAM system. Note that channel 4 is used for timing purposes.

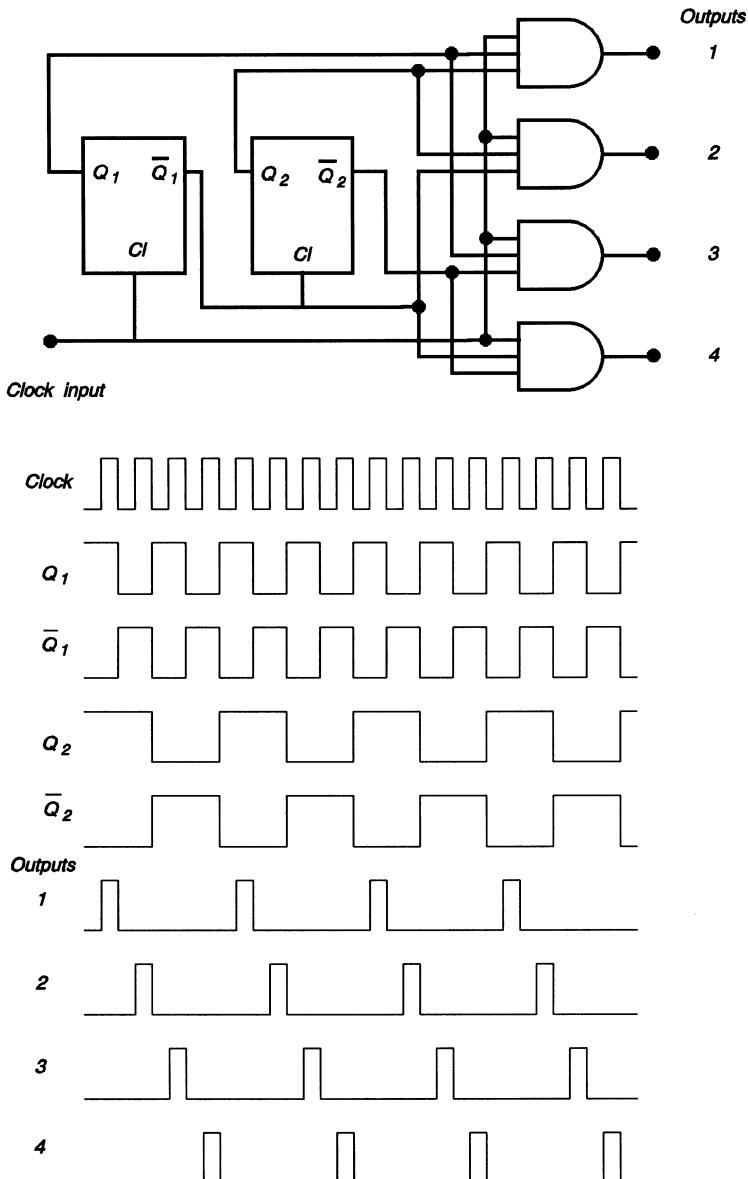


Figure 9.8. A block diagram of a four-phase ring counter with its timing diagram.

It can be seen from the diagram that, in the time taken by one frame, the output pulses go through one cycle. The outputs are used to drive the sampling gates.

9.3.2.2 Series Sampling Gate. The configuration of the series sampling gate is shown in Figure 9.9.

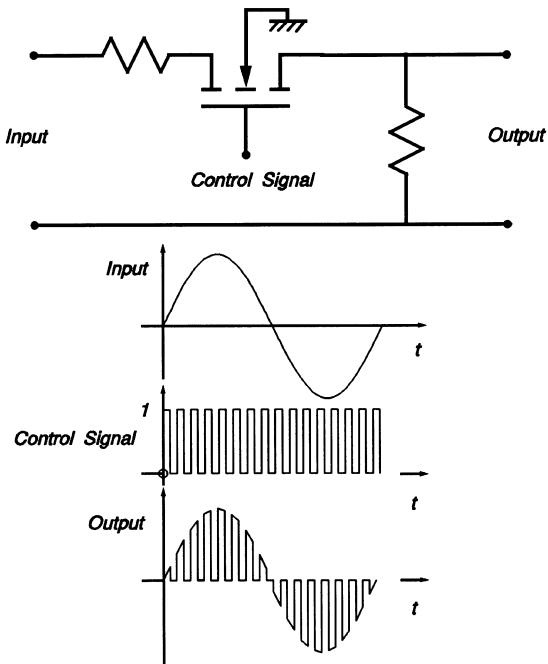


Figure 9.9. The series sampling gate using a FET.

The transistor is an open-circuit when the control signal is a 0 and a short-circuit when it is a 1. The output is as shown.

9.3.2.3 Shunt Sampling Gate.

The shunt sampling gate is shown in Figure 9.10.

The transistor acts as a switch and short-circuits the output when the gate voltage is a 1. When the gate voltage is a 0, it is an open-circuit and a path exists between the input and the output.

9.3.2.4 Series-Shunt Sampling Gate. The two circuits shown above have an inherent deficiency because, when the transistor is on, its source-to-drain impedance is low but not equal to zero. To improve the performance, the action of the two gates can be combined as shown in Figure 9.11.

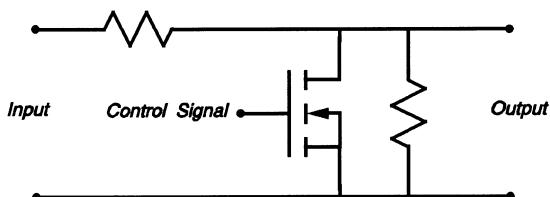


Figure 9.10. The shunt sampling gate.

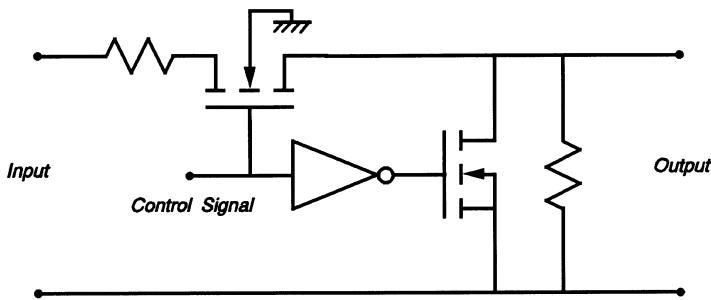


Figure 9.11. A combination of the series and shunt sampling gates improves performance.

9.3.2.5 Operational Amplifier Sampling Gate. The circuit is shown in Figure 9.12.

The operational amplifier is connected to give a gain of R_2/R_1 when the transistor is in the off state. When the transistor is on, R_2 is short-circuited and the gain is reduced to unity.

9.3.2.6 Multiplier Sampling Gate. A PAM sampler can be seen as a multiplication of the analog signal and a train of pulses. The process is illustrated in Figure 9.13.

One of the best methods for accomplishing analog multiplication is to use the four-quadrant analog multiplier. This circuit was described in Section 2.6.3. A practical integrated circuit realization of this is the MC1595 (four-quadrant multiplier manufactured by Motorola Semiconductor Products, Inc.).

9.3.2.7 The Adder. The adder was discussed in Section 4.4.3.5.

9.3.3 Pulse-Amplitude Modulation Decoder

The first step in the recovery of the original three signals is to reverse the action of the commutator by separating them into their respective channels. Low-pass filters

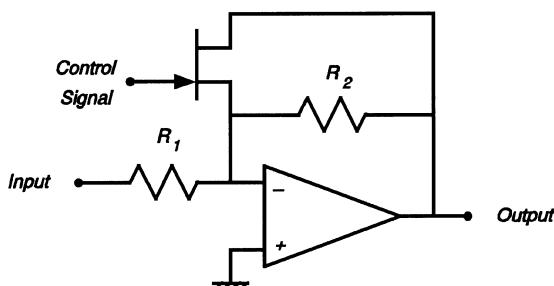


Figure 9.12. A sampling gate using an operational amplifier.

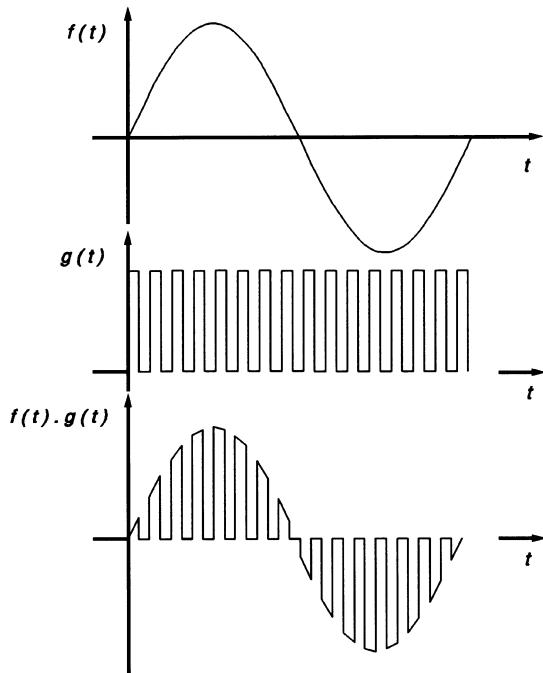


Figure 9.13. The multiplier sampling gate illustrates that sampling is equivalent to the multiplication of two signals.

are then used to reconstruct the analog waveform from the PAM pulses. The PAM decoder system is shown in Figure 9.14.

The incoming signal is fed into the *Schmitt trigger*. The trip level of the Schmitt trigger is set so that only the large sync pulse will trigger it. The output of the Schmitt trigger is then used to synchronize an *astable multivibrator*. The astable multivibrator then runs in synchronism with the master clock in the PAM coder. The output of the astable multivibrator drives a four-phase ring counter which produces four sequential output pulses in synchronism with the ring counter in the encoder. These pulses are used to drive the control gates of the *sample-and-hold* (S/H) circuits. The incoming signal is also fed to the analog inputs of the sample-and-hold circuits. Since the control gate of *only* the S/H-1 is open during the period allotted to channel 1, the pulse amplitude of the signal in channel 1 is passed onto S/H-1. The other channels follow in sequence. The low-pass filters remove the high-frequency components of the PAM signals, producing a replica of the original analog signals.

9.3.3.1 Schmitt Trigger Design. The circuit diagram of the Schmitt trigger is shown in Figure 9.15.

The circuit is designed so that, with no input, Q_1 has no base current and it is therefore off. Q_2 is supplied with base current from the resistive chain R_{c1} , R_1 and R_2 and it is therefore on. Current flows in Q_2 and with the correct choice of R_{c2} , Q_2 will

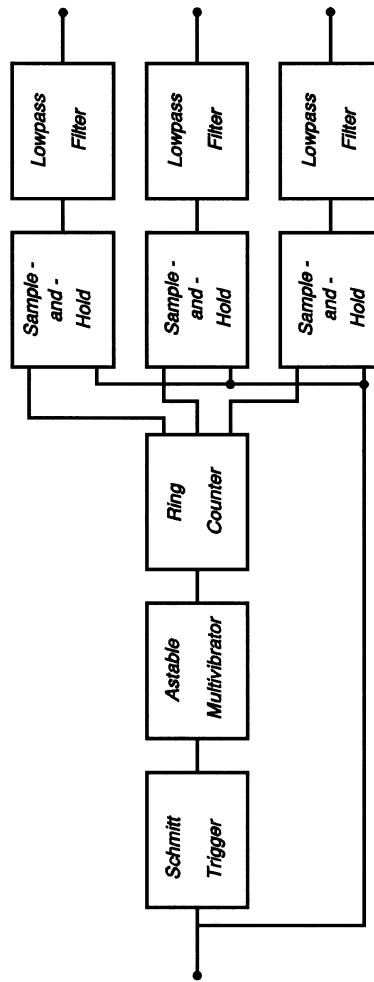


Figure 9.14. A block diagram of the PAM decoder. Note that in Figure 9.7, the first three channels had a signal and the fourth was used for timing purposes.

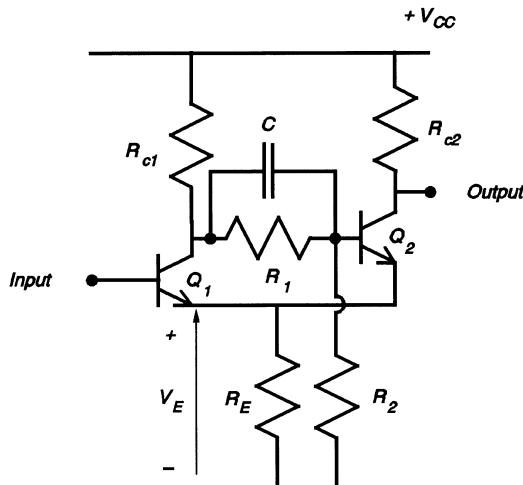


Figure 9.15. The circuit diagram of the Schmitt trigger.

be in saturation. The emitter current of Q_2 flows in R_E and sets up a voltage V_E . If an increasing positive voltage is applied to the input, when it reaches the value $V_E + V_{BE}$, Q_1 will start to conduct. Current is drawn through R_{c1} causing the voltage on the base of Q_2 to drop. Q_2 conducts less vigorously and the voltage across R_E tends to drop. But this drop in voltage at the emitter causes the base-emitter voltage of Q_1 to increase rapidly. This is a form of regeneration and proceeds very fast, ending with Q_1 conducting and in saturation and Q_2 cut off.

When the input voltage is decreasing, there comes the point when it is slightly below the value $V_E + V_{BE}$. Q_1 conducts less current, causing its collector voltage to tend to rise and its emitter voltage to tend to drop. The rising trend at the collector of Q_1 is passed onto the base of Q_2 by R_1 and C . The combined effect of a decreasing V_E and a rising base voltage causes Q_2 to switch on regeneratively and go into saturation. The Schmitt trigger reacts to a slowly changing input voltage by producing a voltage step when its trip level is exceeded. This happens for increasing as well as decreasing voltages. The design of the Schmitt trigger is best illustrated by an example.

Example 9.3.1 Schmitt Trigger Design. Design a Schmitt trigger circuit so that it triggers when a voltage in excess of 3.0 V is applied at the input. The dc supply voltage is 10 V and two NPN silicon bipolar transistors with $\beta = 100$ are provided. $V_{CE(sat)} = 0.5$ V and $V_{BE} = 0.7$ V. The load driven by the Schmitt trigger requires a current of 1 mA.

Solution. The transistors are made of silicon; therefore, $V_{BE} = 0.7$ V. For the circuit to trigger at 3.0 V, V_E must be designed to be equal to $(3.0 - 0.7) = 2.3$ V. Since the Schmitt trigger is to drive a load that requires 1.0 mA, it is good design practice to

allow about 10 times this current to flow in the collector of Q_2 . Collector current of Q_2 is then 10 mA; therefore, the emitter current is also 10 mA, and

$$R_E = 2.3/10 \text{ k}\Omega = 230 \text{ }\Omega.$$

Since Q_2 is in saturation, the collector voltage must be equal to $(V_E + 0.5) = (2.3 + 0.5) = 2.8 \text{ V}$. The voltage drop across R_{c2} is $(10.0 - 2.8) = 7.2 \text{ V}$. Since 10 mA flows through R_{c2} ,

$$R_{c2} = 7.2/10 \text{ k}\Omega = 720 \text{ }\Omega.$$

The base voltage of Q_2 must be at the voltage $(V_E + 0.7) = (2.3 + 0.7) = 3.0 \text{ V}$. The base current of Q_2 is

$$I_B = I_C/\beta = 10/100 \text{ mA} = 100 \mu\text{A}.$$

To maintain a reasonably stable trigger point, it is necessary to allow about 10 times the base current to flow in the resistive chain R_{c1} , R_1 and R_2 . The current in the resistive chain is therefore 1 mA and its total resistance is $10 \text{ V}/1 \text{ mA} = 10 \text{ k}\Omega$. Let R'_1 be equal to $(R_{c1} + R_1)$ so that

$$R'_1/R_2 = (10.0 - 3.0)/3.0 = 2.33.$$

But

$$R'_1 + R_2 = 10 \text{ k}\Omega.$$

Therefore

$$R'_1 = 7.0 \text{ k}\Omega \quad \text{and} \quad R_2 = 3.0 \text{ k}\Omega.$$

When Q_1 is in saturation, it must draw the same current as when Q_2 was in saturation,

$$R_{c1} = R_{c2} = 720 \text{ }\Omega.$$

Therefore

$$R_1 = (7.0 - 0.72) = 6.28 \text{ k}\Omega.$$

The capacitor C is a speed-up capacitor which helps the transition of Q_2 from the on state to the off state. A reasonable value is 50–100 pF.

9.3.3.2 Sample-and-Hold Circuit. The circuit diagram of the ideal S/H circuit is shown in Figure 9.16.

The switch S closes and the ideal voltage source charges capacitor C instantaneously. When S opens, the capacitor retains its charge indefinitely. In practice, S is

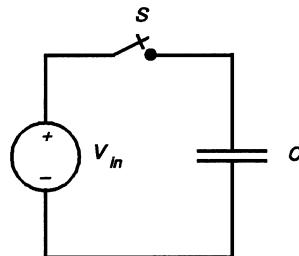


Figure 9.16. The ideal sample-and-hold circuit.

not an ideal short-circuit when closed and the voltage source has an internal resistance R_s . This means that C charges up with a time-constant $\tau = CR_s$. So long as the pulse width of the driving source is significantly longer than τ the error can be regarded as insignificant. Another practical consideration is that, when S is open, C can lose its charge due to leakage in the dielectric of the capacitor and the finite impedance of the load driven by the S/H.

A simple but practical circuit for the S/H is shown in Figure 9.17.

The operational amplifiers are connected as voltage followers with a gain of unity. The output impedance of A_1 is low enough for it to drive the required charge into the capacitor. The N -channel JFET is switched on by a pulse applied to the gate and the capacitor charges up to the value of the input voltage. When the JFET switch is turned off, the high input impedance of A_2 drains minimal current from C . The design of the S/H circuit is best illustrated by an example.

Example 9.3.2 Sample-and-Hold Circuit. The S/H circuit shown in Figure 9.17 uses a JFET as a series sampling gate. The voltage follower A_2 takes a current of 500 nA. The JFET has a source-to-drain resistance of 25 Ω when it is in the on state and may be considered to be an open-circuit when it is in the off state. The signal

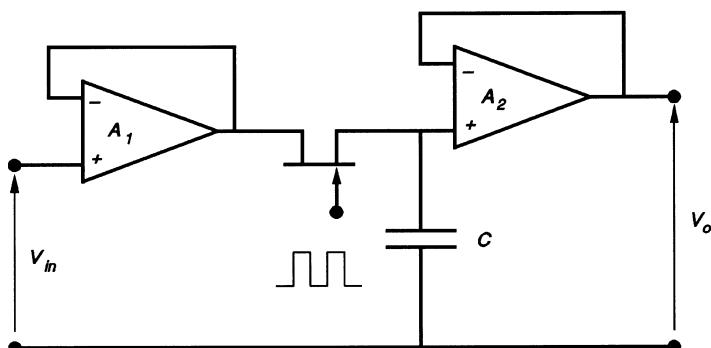


Figure 9.17. A practical sample-and-hold circuit.

amplitude is 2.0 V, the sample-time is 5 μ s and the hold time is 500 μ s. The capacitor has a value 0.4 μ F. Calculate the error in the output at the end of the hold time. Assume that the signal source has negligible resistance.

Solution. When the sampling gate is closed, assume that the current flowing into the capacitor $i \gg 550$ nA (the leakage current taken by A_2):

$$\tau = CR_s = 25 \times 0.4 \times 10^{-6} \text{ s.} \quad (9.3.1)$$

For the RC circuit shown in Figure 9.18,

$$v_c = V_s(1 - e^{-t/\tau}). \quad (9.3.2)$$

When $t = 5 \mu$ s, and $V_s = 2.0$ V, $v_c = 1.987$ V.

Charge on C at $t_1 = 5 \mu$ s,

$$Q = Cv_C = 0.4 \times 10^{-6} \times 1.987 = 0.795 \mu\text{C}. \quad (9.3.3)$$

Charge lost by C in time $t_2 = 500 \mu$ s,

$$\Delta Q = I\Delta t = 500 \times 10^{-9} \times 500 \times 10^{-6} = 25 \times 10^{-11} \text{ C.} \quad (9.3.4)$$

Charge on C at time t_2 ,

$$Q - \Delta Q = (0.795 \times 10^{-6} - 25 \times 10^{-11}) = 0.794 \times 10^{-6} \text{ C.} \quad (9.3.5)$$

Voltage across C at time t_2 ,

$$v_{c_2} = \frac{Q - \Delta Q}{C} = 1.986 \text{ V} \quad (9.3.6)$$

Percentage error in the output is 1.4%.

9.3.3.3 Other Circuit Blocks in the Decoder. The synchronized astable multivibrator was discussed in Section 7.2.7.3 and the ring counter in Section 9.3.2.1. The design of the filters is beyond the scope of this book.

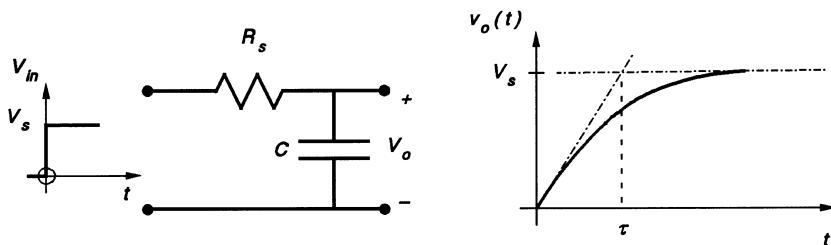


Figure 9.18. The RC circuit for the Example 9.3.2 showing the exponential response.

9.3.4 Pulse Code Modulation Encoder/Multiplexer

As mentioned earlier signals coded in PAM tend to be susceptible to corruption by noise and circuit non-linearities since the information is contained in the amplitude of the pulses. Greater noise immunity can be obtained if the amplitude of the pulse is coded in binary form. In the presence of noise and other forms of pulse degradation it is much easier for the receiver to distinguish between the presence (**1**) or absence (**0**) of a pulse as opposed to the height of a pulse. PCM is the preferred technique in all modern telephone systems.

To illustrate the basic principles of the design, a four-channel PCM carrier system has been chosen and it is shown in Figure 9.19(a). The four voice channels are sampled in sequence so the sampler output is a time-division multiplexed PAM signal. The PCM encoder converts the amplitude of each of the four samples into an eight-bit binary code. Figure 9.19(b) shows the sampled analog signals in the four-channel system, the interleaved PAM samples, and the PCM output. The 32 (4×8) bits are transmitted over a twisted-pair telephone wire. One bit (frame bit) is added for synchronization purposes; it may be wider than the other pulses or it may have a specially coded sequence of **1s** and **0s** which can be easily recognized by the receiver.

Comparing Figures 9.7 and 8.34 with Figure 9.19 shows that the PCM is a PAM system feeding into the *coder* of the *codec* (A/D followed by a P/S converter). The design of the ring counter and the sampling gates were discussed in Section 9.3.2. The A/D and P/S converters were discussed in Section 8.5.1.

9.3.5 Pulse-Code Modulation Decoder/Demultiplexer

The system diagram of the decoder/demultiplexer is shown in Figure 9.20. The *frame bit extractor* is designed to recognize the frame bit and send out a pulse to synchronize the local oscillator to the master oscillator in the PCM encoder. The output of the local oscillator is used to drive the S/P converter, the A/D converter, and the four-phase ring counter. The output from the D/A converter and the ring counter are fed to the S/H circuits. If the timing is correct, the decoder reconstructs the four samples from the eight-bit codes. The circuit assigns each sample to its correct channel (demultiplex) and the outputs of the low-pass filters follow the envelope of the PAM signal, producing a replica of the original audio-frequency signal.

Comparing Figures 8.34 and 9.14 with Figure 9.20 shows that the decoder/demultiplexer is, in fact, the *decoder* of the *codec* followed by the PAM demultiplexer. The design of the S/P and D/A converters was discussed in Sections 9.5.1.5 and 9.5.1.1, respectively. The design of the circuit components for the PAM demultiplexer was discussed in Section 9.3.3.

9.3.6 Bell System T-1 PCM Carrier

The Bell T-1 carrier system uses an eight-bit PCM in 24 voice-channel banks. The sampling rate is 8 kHz. The number of bits generated for one scan of the channels

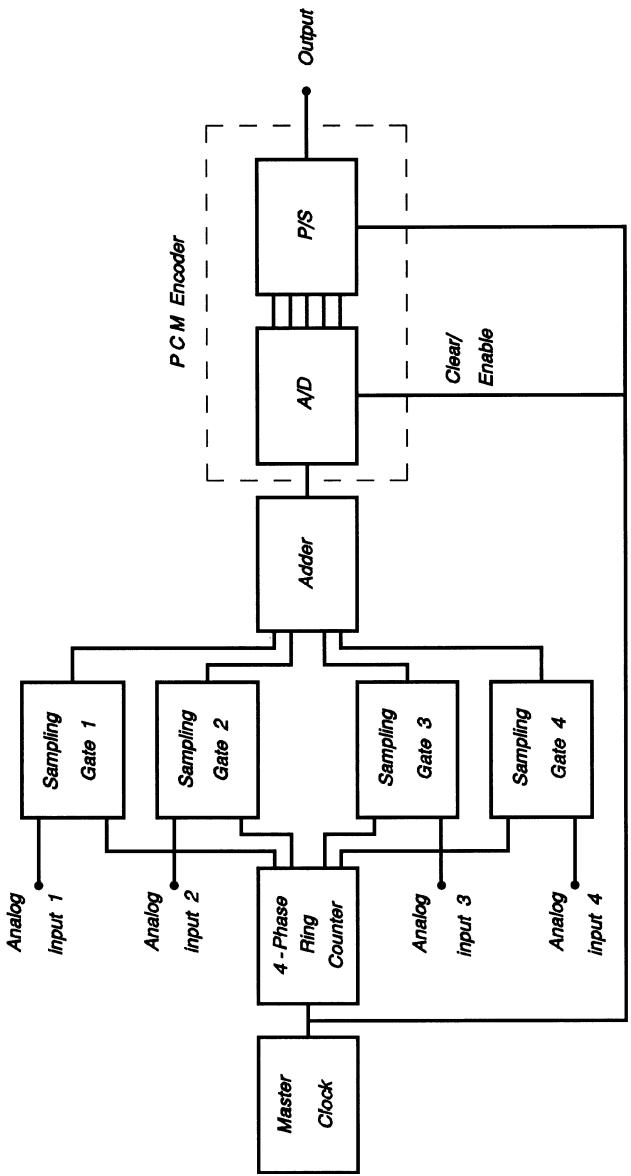


Figure 9.19. (a) The block diagram of a four-channel PCM system. Note that this is equivalent to a PAM circuit feeding into the coder of the codec. Reprinted with permission from *Transmission Systems for Communications*, 5th Ed., AT&T, Bell Labs, 1982.

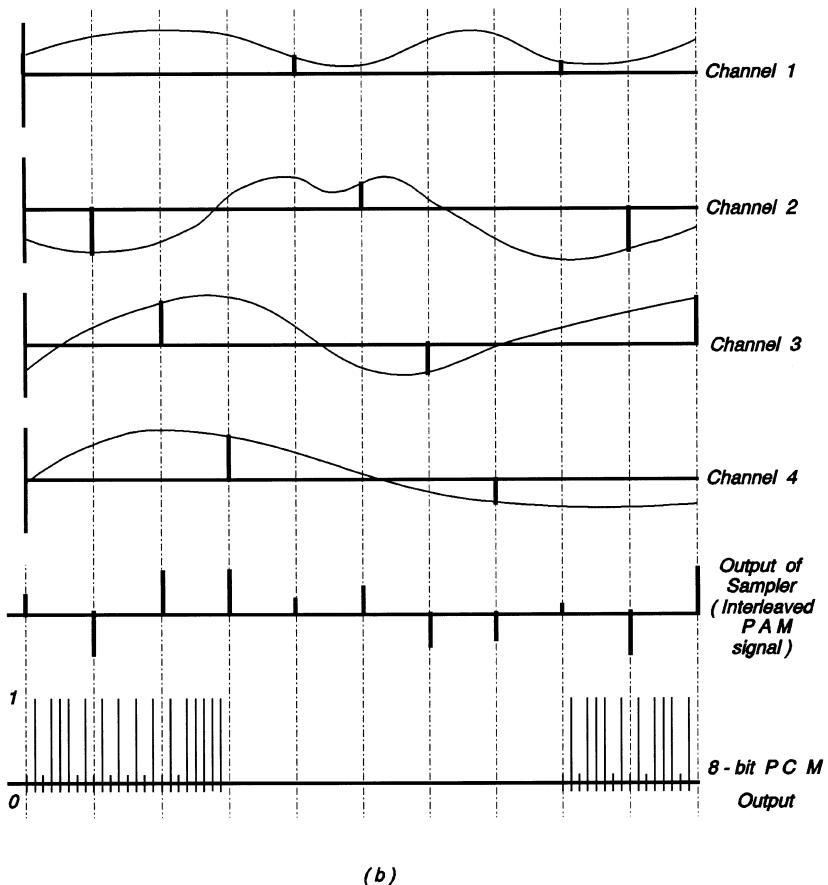


Figure 9.19. (b) An illustration of the signals in the four channels, the interleaved PAM and the 8-bit PCM equivalent. Reprinted with permission from *Transmission Systems for Communications*, 5th Ed., AT&T, Bell Labs, 1982.

(frame) is $24 \times 8 = 192$. One bit (frame bit) is required for synchronization so the total number of bits per frame is 193. The analog signal is band-limited to approximately 3.5 kHz and, because the sampling rate is 8 kHz, the bit rate is $(193 \times 8000) = 1.544$ Mbit/s. The minimum bandwidth required to transmit the signal is 1.5 MHz. The use of the eight-bit code means that the voice signals are quantized at 256 (2^8) levels. Some of the less significant bits may be robbed and used for signalling purposes such as dialling and detection of hook switch on/off. The quantization error resulting from this is considered to be tolerable, although several compression schemes are used to minimize its effect [9].

When the signal is sent over a twisted-pair telephone wire, it suffers considerable degradation from noise, bandwidth limitation, and phase delay. It is therefore necessary to place repeaters and equalization circuits at intervals of approximately

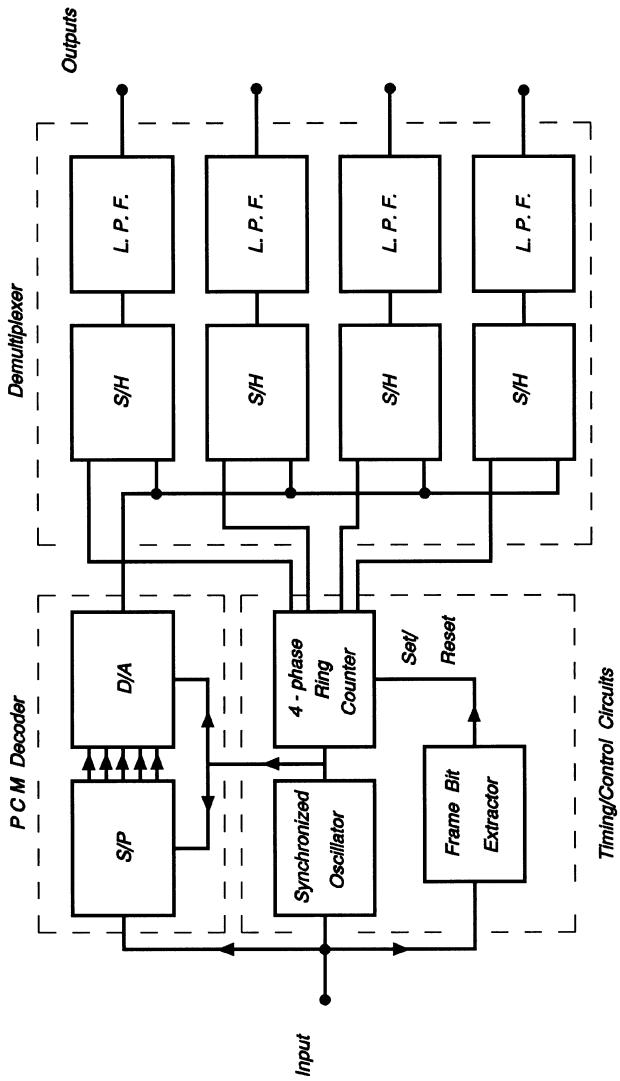


Figure 9.20. The block diagram of the PCM decoder/demultiplexer.

6000 ft to restore the pulses. 6000 feet (approximately 2000 m) of 22 gauge twisted-pair non-loaded wire has a 3 dB bandwidth of approximately 4 kHz.

9.3.7 Telecom Canada Digital Network

The Telecom Canada Digital Network (DS-system) is similar to the American Telephones and Telegraph (AT&T) *T*-system. In the Telecom Canada digital network [1], 24 voice channels each operating at 64 kbit/s are multiplexed into a 1.544 Mbit/s channel, designated DS-1. Four DS-1 channels are multiplexed into a DS-2 which operates at 6.312 Mbit/s. DS-3 has seven DS-2 feeding into it and it has a bit rate of 44.736 Mbit/s. Six DS-3 channels are multiplexed to form DS-4. Its bit rate is 274.176 Mbit/s and it can handle 4032 voice channels. The signals from DS-1 and DS-2 may be transmitted on number 22 shielded twisted pairs. DS-3 and DS-4 use coaxial cable. All DS-system output pulses are put into the *alternate mark inversion* (AMI) format. This format eliminates the dc content of the signal and it can be used to detect errors in transmission as well as provide substitution codes for time synchronization. The elimination of the dc content of the signal means that the dc power required to operate the repeater can be sent on the same line as the signal.

The output from these PCM switches can be used to modulate a suitable carrier for transmission over terrestrial microwave links, fiber optic cables or satellite links.

Newer and more elaborate digital systems have come on the market offering more channels, higher bit rates, greater flexibility, higher voice fidelity, and lower cost. The basic idea of a hierarchy of systems made up of modules which are interchangeable remains (see Section 9.2.4).

9.3.8 Synchronization Circuit

For large switches such as the DS-2 or DS-3, the simple timing circuit used in the example of the four-channel system is evidently inadequate. A centrally located cesium beam atomic clock (oscillator) generates the primary signal. The frequency stability of this oscillator is guaranteed to be less than one part in 10^{11} during its lifetime. The output signal is distributed to all nodes of the network where timing information is required according to a hierarchy determined by the importance of the node. At each node, the signal is used to synchronize a local crystal-controlled oscillator. The local oscillator (nodal clock) then has the accuracy of the primary cesium beam atomic clock. The nodal clock must be accurate on its own (one part in 10^{10} per day) so that, at times when the primary source is lost, the system can function satisfactorily. All DS-1 signals are therefore synchronized indirectly to the cesium clock. The synchronization of the nodal clocks is carried out as shown in Figure 9.21.

This is a phase-locked loop (see Section 5.2.7.3) in which the frequency of the local voltage-controlled crystal oscillator (VCXO) is adjusted by a voltage derived from the frequency difference between the two oscillators. When the difference is zero no adjustment is applied and the VCXO stays in synchronism.

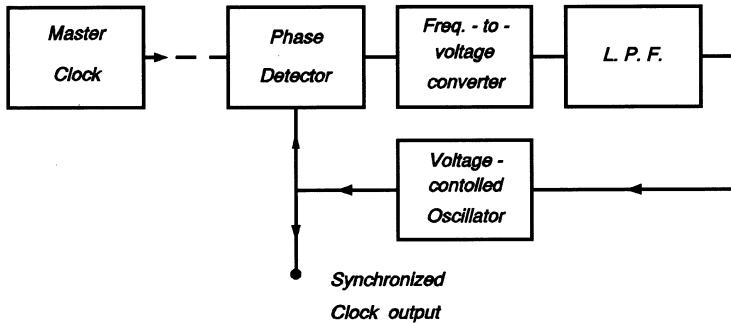


Figure 9.21. A block diagram of the system for the synchronization of the master clock to the nodal clock.

9.3.9 Regenerative Repeater

It was mentioned earlier that the output of the DS-1 switch was transmitted on the bandwidth-limited twisted-pair telephone line. The result of this is a rapid degradation of the signal with other factors such as noise contributing to it. It is therefore necessary to install regenerative repeaters approximately 1800 m apart to detect the degraded pulses and send new ones [3,4]. The action of a regenerative repeater includes amplification, equalization, detection, timing, and pulse generation. The regenerative repeater does not accumulate noise the way a repeater (amplifier) on an analog line would; it generates a new clean pulse. A block diagram of the basic regenerative repeater is shown in Figure 9.22.

The incoming signal is fed into a *pre-amplifier/equalizer* to boost the level of all frequencies present in the signal to a suitable value. The output of the amplifier then goes to the input of the frame *bit detector* (not shown) and the *full-wave clock rectifier*. The frame detector identifies the start/stop of each frame and the rectifier changes the bipolar form of the signal to unipolar, thus producing a discrete frequency component at the signalling rate. A high *Q* factor *LC*-tuned amplifier selects the sinusoidal frequency component at the clock rate. The resulting sinusoid is amplified and fed to a *phase shifter*. The phase shifter output is further amplified and then goes to the *amplitude limiter* which produces a square wave. The square wave is fed into a *differentiator* to generate positive- and negative-going clock pulses

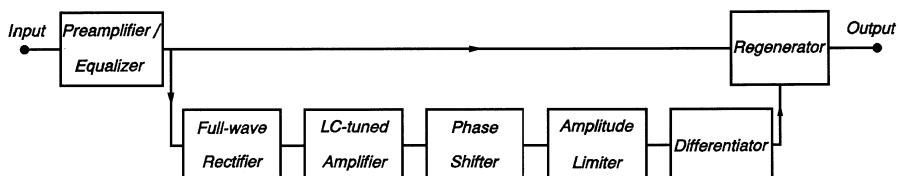


Figure 9.22. The block diagram of the regenerative repeater. The bit frame detector is not shown. Reprinted with permission from *Transmission Systems for Communications*, 5th Ed., AT&T, Bell Labs, 1982.

at the zero crossings. A phase shift network is used to adjust the phase of the clock pulses so that the positive clock pulses coincide with the maximum points (see Figure 9.26: this is referred to as “maximum eye opening”) on the incoming signal (amplified and equalized). The resulting pulse and the output of the preamplifier/equalizer go to the *regeneration repeater* where a decision has to be made, in each time slot, whether the received bit was a **1** or a **0**. If the decision is that the received bit was a **1**, the regenerator produces a pulse of appropriate dimensions.

9.3.9.1 Pre-amplifier/Equalizer. As mentioned in Section 9.3.6, the PCM signal transmitted over the twisted-pair telephone line degenerates quite rapidly due to electrical noise, both man-made and natural. The restricted bandwidth and delay characteristics of the line attenuate the higher frequency components of the signal, reducing the rise and fall times of the pulses. The amplifier/equalizer has an appropriate high-frequency response boost built into it to compensate approximately for 1800 m of 22-gauge twisted-pair telephone line. This circuit may include an automatic gain control to keep the signal amplitude within specified limits.

9.3.9.2 Frame Bit Detector. The frame consists of 192 bits (8-bit code \times 24 channels). Since the sampling frequency is 8 kHz, each frame must be scanned in 125 μ s. It is necessary, however, to have a marker to indicate where the frame starts or ends so that each 8-bit word can be identified and steered to the correct channel at the receiving end [6]. To provide such a marker, one extra bit, called the framing bit, is added at the end of each frame. By changing the framing bit from a **1** to a **0** and back again at the end of every frame, it is possible to identify it at the receiving end and to use it for synchronization of the channels. Positive identification is possible because the alternating value of the framing bit represents a frequency of 4 kHz but all the signals in the other channels were pre-filtered to remove all frequencies above 3.5 kHz. It follows that none of the other bits derived from signals in the channel can have a pattern of alternating bits from frame to frame.

The incoming pulses are scanned for the alternating bit in the 193rd position, and when it is found the system latches onto it. When a specified number of errors occur in the 193rd time slot within a given time period, the system reactivates itself and starts the scanning process over again. While this is going on information is lost and the telephone user might hear a click or two.

The circuit of the frame bit detector and the reframing process when an error occurs is quite extensive and will not be discussed in greater detail.

9.3.9.3 The Clock Rectifier. The full-wave rectifier restores the timing frequency to its original value and reverses the alternate mark inversion (AMI) process. This can be seen in Figure 9.23(a).

The rectification process is best achieved by a center-tapped transformer and two diodes as shown in Figure 9.23(b).

9.3.9.4 The LC-Tuned Amplifier. The Q factor of the amplifier is designed to be in the range 75–100. This ensures a narrow bandwidth for the precise recovery of

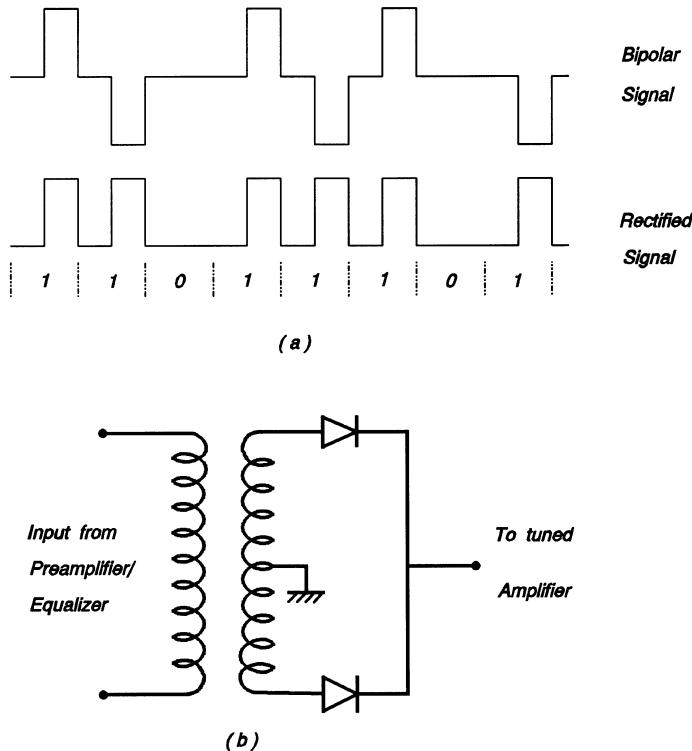


Figure 9.23. The bipolar signal is rectified by the center-tapped transformer and diodes.

the timing frequency and the so-called flywheel effect, that is, the circuit continues to operate even when some of the pulses are missing. The output is sinusoidal. Narrow bandwidth LC -tuned amplifiers were discussed in Section 2.5.2.

9.3.9.5 Phase Shifter, Limiter and Differentiator.

The phase shifter, limiter and differentiator are shown in Figure 9.24.

The phase shifter shown in the diagram provides an output which leads the input signal; R_1 and C_1 can be interchanged if a phase lag is required. The time-constant, R_1C_1 , determines the amount of phase shift.

The limiter is an operational amplifier with two back-to-back Zener diodes in its feedback path. The output of the operational amplifier is clamped in the positive direction by D_2 acting as an ordinary diode and D_1 acting as a Zener diode. Their roles are reversed when the output voltage goes negative. The resulting square wave output goes to the differentiator.

The choice of the time-constant for the “differentiator” is governed by the condition $R_2C_2 \ll T$, the period of the input frequency.

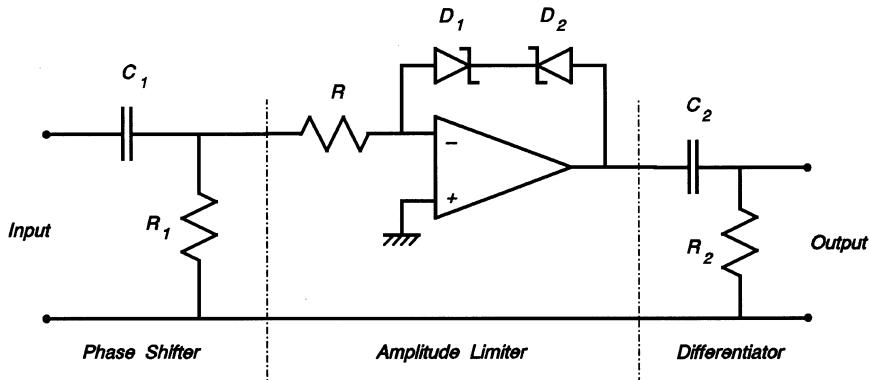


Figure 9.24. The phase shifter, amplitude limiter and differentiator.

9.3.9.6 Pulse Regenerator. The circuit of the pulse regenerator [4] is shown in Figure 9.25.

The pulse regenerator consists of a bistable multivibrator whose operation is controlled by the clock pulse and the received signal pulse (amplified and equalized) through the transformer T_1 and diodes D_1-D_6 .

The received signal pulse is applied to the primary of the transformer T_1 whose secondary center-tap is grounded. Assuming that the voltage at node X is positive and also that this happens to coincide with a positive clock pulse, the diodes D_6 and D_4 are therefore reverse biased. Assume further that Q_1 is on and in saturation and Q_2 is off; the base of Q_1 is then at +0.7 V. Diode D_2 will conduct current through R'_4 due to the dc supply, V_{BB} . There will be a tendency for Q_1 to conduct a little less current than before. Meanwhile, the voltage at node Y is negative and D_5 is forward biased and the voltage at node S is more positive than the voltage on the base of Q_2 . D_1 is therefore reverse biased; consequently the base voltage of Q_2 is free to change. Returning to Q_1 , its collector voltage will tend to rise which will cause the base voltage of Q_2 to rise too. The collector voltage of Q_2 will start to drop and this will cause the base voltage of Q_1 to fall even faster. This is the regenerative phase of the multivibrator action which ends with Q_1 cut off and Q_2 in saturation.

The signal from the pre-amplifier/equalizer is still in the AMI format so the next signal pulse input to the regenerator will cause node Y to go positive and if this happens to coincide with a positive clock pulse, the multivibrator will change states with Q_1 back on and Q_2 cut off.

Since the collectors of Q_1 and Q_2 are coupled through the primary winding of transformer T_2 to the positive dc supply, V_{cc} , whenever they change states, a pulse of current will flow in opposite directions in the primary winding of T_2 , causing a pulse of opposite polarity to occur across the secondary. The output pulses are therefore in the AMI format (bipolar).

The time-constant of the transformer inductance, its associated load and the current that flows in the transformer when the multivibrator changes states determine

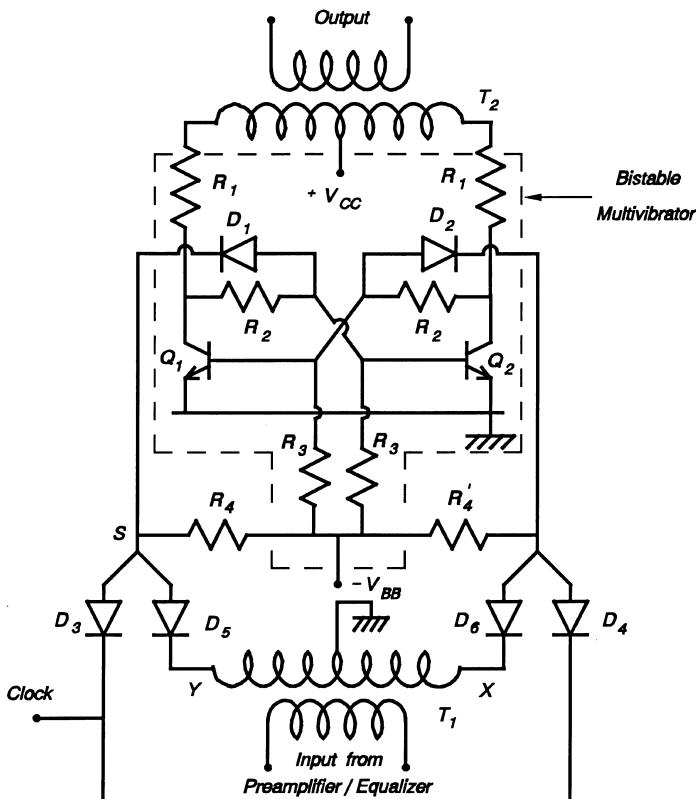


Figure 9.25. The bipolar pulse regenerator.

the width of the pulse. The pulse shape is primarily controlled by the $B-H$ (magnetic) characteristics of the transformer core material. A boxlike hysteresis loop produces a pulse with fast rise and fall times. This is because the core flux saturates early and any further increase in the current has no corresponding effect on the flux.

There are a number of different circuits which can be used in a regenerative repeater. Several versions use a blocking oscillator to restore the degraded PCM pulses.

To optimize the operation of the regenerative repeater, the phase shifter is adjusted so that the positive-going clock pulse coincides with the maximum value of the pre-amplifier output. This is shown in Figure 9.26.

9.4 DATA TRANSMISSION CIRCUITS

Data transmission over metallic loops predicated voice traffic by at least 30 years. In 1837, Wheatstone and Cooke in England, and Morse in the United States, established communication systems which used binary information exclusively.

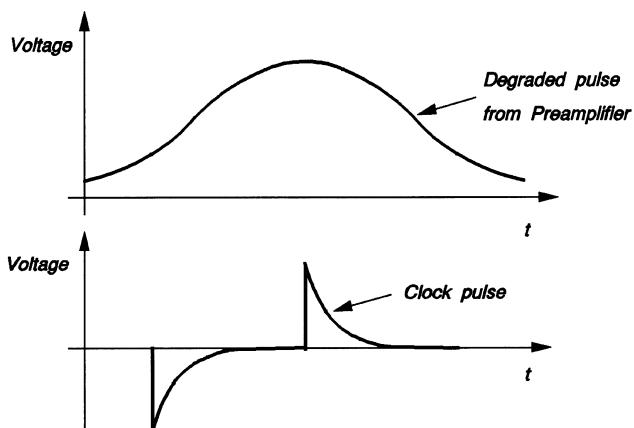


Figure 9.26. The clock pulse is adjusted to coincide with the peak of the degraded pulse. Reprinted with permission from J.S. Mayo "Bipolar Repeater for Pulse Code Modulation Signals", *Bell System Technical Journal*, 41, 25–97, 1962.

Later, much improved forms of the telegraph (Teletype or Telex) worked side by side with the telephone into the 1970s. One of the major reasons for its continued use was the need for the written word in business transactions. The advent of fast electronic computers capable of handling data at an ever faster rate hastened the development of transmission systems designed to cope with high speed data. The evolution of the teletype into the *integrated services digital network* (ISDN) will now be traced briefly.

The printing telegraph was invented in 1855 and it was the first attempt to transmit data automatically. In its practical form (Teletype) it had a keyboard based on the typewriter. In the transmit mode, the pressing of a key would cause a start code to be transmitted followed by a 5-bit message code followed in turn by a stop code. It was therefore not necessary to synchronize the transmitter to the receiver. This was all done by a set of mechanical cams which opened and closed contacts to produce the necessary pulses. In the receive mode, the current pulses were decoded and used to operate a suitable electromechanical system to print the required character. The information was transmitted over telegraph lines. This system was capable of speeds of 50–150 words per minute. It was usual practice to use a punched paper tape to compose and edit the message before transmission.

The next significant improvement came with the introduction of the Teletype-writer Exchange (TWX) [1]. This used a voice channel on the telephone system, including the switching system and represented an early example of service integration. It ran at about 100 words per minute.

The addition of "store-and-forward" capability to the TWX provided a new service called the Message Switched Data Service (MSDS). Store and forward, as the name suggests, was the ability to accept and store a message and to transmit it only when a channel was available and the receiver was ready. It also had the possibility to send the message to more than one recipient terminal.

The next advance made data transmission compatible with voice telephony; it used the analog telephone channel by interfacing through a *modem* (*modulator-demodulator*). It was called Dataphone. The digital signal was used to modulate an analog carrier in a *frequency-shift keying* (FSK) scheme. The digit **1** or *mark* is coded as one audio frequency and the digit **0** or *space* is coded as a different audio frequency. A typical modem generates 980 Hz for a mark and 1180 Hz for a space. These frequencies were chosen to be compatible with the bandwidth limitations of the channel and the speed of the transmission. At the receiving end the frequencies were decoded into binary form. Data transmission rates up to 9600 bit/s have been achieved using this technique.

The need to reduce the cost of data transmission led to the introduction of an essentially digital network called Dataroute. It operates at speeds up to 56 kbit/s and it can handle both synchronous and asynchronous data traffic at various speeds by using suitable conversion codes. The data may be inserted into the DS-1 PCM system and transmitted over the 1.544 Mbit/s channel.

The advantages offered by the use of computers in the telephone system are exploited in the Datapac system. The system consists of nodes interconnected by 56 kbit/s transmission trunks. The nodes have dedicated computers which receive, check for errors, package, address, and transmit data to other nodes in the network. This is an example of a packet switched network. It has facilities for retransmission of data when errors are detected.

The next significant improvement in data transmission was provided by Datalink. This has the capability of transmission rates of 2400, 4800 and 9600 bit/s. It is independent of code sets and protocols and compatible with the digital telephone network.

Integrated Services Digital Network (ISDN) is the result of international efforts to standardize the format to be used in the existing and growing number of new services offered by the telephone system and other organizations. It uses packet switching techniques in which the data stream is separated into packets of modest size and stored in a buffer until a channel becomes available for its onward transmission. It is quite likely that different packets of a message reach their destination at different times and by different routes. It is therefore imperative that every packet be labelled so the message can be reconstructed at the receiving end. Current international agreements assume that ISDN would carry a whole range of telecommunication services including digital video signals. The suggested classes of service are:

- (1) simple voice-channel service,
- (2) complex telephone service (smart telephone),
- (3) message storage and delivery,
- (4) telemetry,
- (5) low bit rate data,
- (6) 8–64 kbit/s data,
- (7) $N \times 64$ kbit/s data services, where N is large.

One of the more prominent data transfer systems under development is the *asymmetrical digital subscriber line* (ADSL). It belongs to a class of techniques generally referred to as xDSL which aim to create a high-speed data link between personal computers and the Internet using the existing twisted-pair telephone line. It is asymmetrical because, in general, the flow of data from the Internet to a personal computer far exceeds the data flowing in the opposite direction. ADSL is supposed to open up the bottleneck formed by the twisted pair between the telephone wall socket and the central office, often referred to as “the last mile”. The name “the last mile” is deceptive because the length of the twisted-pair telephone line can vary from a few meters to approximately 6 km. The variability of the line impedance and its frequency response are problems that all xDSLs have to deal with. The nominal bandwidth of the-last-mile is approximately 4 kHz. The bandwidth required for data from the central office to the subscriber is approximately 1.1 MHz for the transfer of acceptable video information for video conferences and the downloading of complex graphical displays from web pages. To achieve this objective high-gain, low-noise amplifiers with built-in frequency equalizers are required. However, amplifiers can generate gain for signals travelling in one direction only. It is therefore necessary to employ 2-to-4 wire hybrids so that signals can be transmitted in both directions.

9.4.1 Modem Circuits

The block diagram of a typical modem is shown in Figure 9.27. Typical frequencies are:

$$\begin{aligned} f_1 &= 980 \text{ Hz}, & f_2 &= 1180 \text{ Hz} \\ f_3 &= 1650 \text{ Hz}, & f_4 &= 1850 \text{ Hz}. \end{aligned}$$

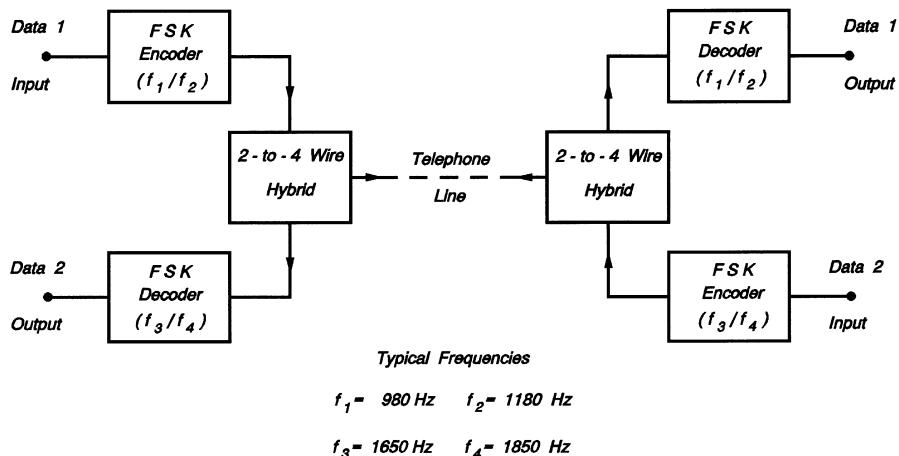


Figure 9.27. The block diagram of a typical modem.

Incoming Data 1 is fed into the input port of modem 1 where a FSK transmitter converts it into f_1 and f_2 for mark and space, respectively. It is then coupled to the telephone line through a 4-to-2 wire hybrid (in older models, the carbon microphone and the telephone receiver were used in conjunction with an acoustic coupler; the 4-to-2 wire hybrid would then be the station set hybrid). At the receiving end modem 2 hybrid directs Data 1 to the FSK decoder which converts f_1 and f_2 into mark and space, respectively. Modem 2 communicates with modem 1 in the same way except for the difference in the frequencies used.

The 4-to-2 wire hybrid was discussed in Section 8.5.1.6.

9.4.1.1 Frequency-Shift Keying Modulator (Transmitter). The FSK modulator of the modem has the block diagram shown in Figure 9.28. A crystal-controlled oscillator provides an accurate clock frequency which is fed into two divide-by- N circuits. The outputs are passed through narrow bandpass filters tuned to the required frequencies f_1 and f_2 . The outputs of the bandpass filters go to the inputs of the multipliers. The multiplicands are the binary signal to be FSK-coded and its complement. Note that the binary signal and its complement are used to control the flow of signal from the bandpass filters to the adder. When the binary input is a 1, a signal of frequency f_1 is fed to the adder; f_2 is cut off. When it is a 0, f_1 is cut off and f_2 appears at the input of the adder.

The crystal-controlled oscillator was discussed in Section 2.4.6. An inexpensive and readily available crystal designed to operate at 3.58 MHz (originally designed for the color burst carrier of television receivers; see Section 8.4.5.2) may be used in which case we have:

Target frequency, f	Required divisor	Integer N	Actual frequency, f'	Error %
980	3653.06	3653	980.02	+0.002
1180	3033.90	3034	1179.96	-0.003

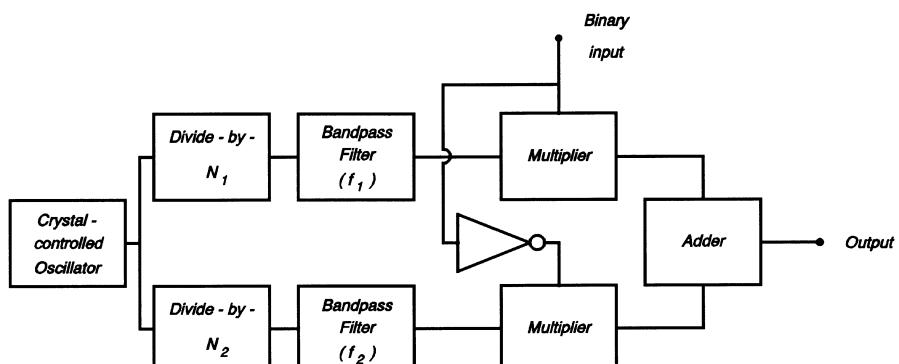


Figure 9.28. The FSK modulator with a crystal-controlled oscillator driver.

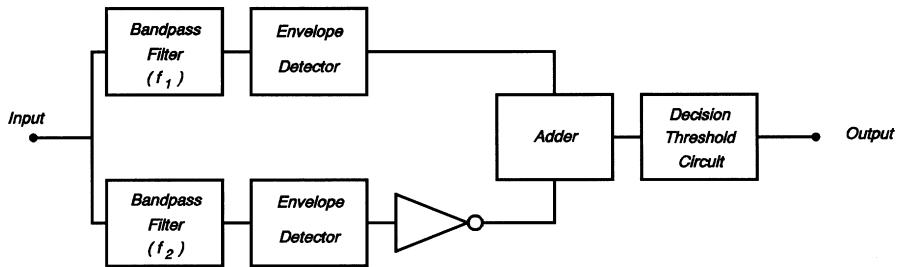


Figure 9.29. The FSK demodulator.

The divide-by- N circuits can be realized by using two 12-stage ripple counters ($2^{12} = 4096$) with suitable resets to get the required values of N . The bandpass filters can be simple LC -tank circuits with modest Q factors tuned to 980 Hz and 1180 Hz, respectively. The analog multiplier was discussed in Section 2.6.3, the inverter in Section 8.4.5.2, and the adder in Section 4.4.3.5.

There are several other schemes which can be used to realize the FSK coder.

9.4.1.2 Frequency-Shift Keying Demodulator (Receiver). A simple FSK demodulator of the modem is shown in Figure 9.29. Two narrow bandpass filters are tuned to the two frequencies present in the FSK signal. Their outputs are used to drive the envelope detectors. Note that the envelopes of the two frequencies f_1 and f_2 are both square waves. The output of one of the envelope detectors is inverted and added to the other to give the required binary output.

The narrow bandpass filters can be LC parallel-tuned amplifiers with modest Q factors. The envelope detector was discussed in Section 3.4.6, the inverter in Section 8.4.5.2, and the adder in Section 4.4.3.5. The decision threshold circuit is a Schmitt trigger which was discussed in Section 9.3.3.1.

There are several modems on the market which use sophisticated phase-locked loop techniques for both the modulator and demodulator; most use integrated circuits.

REFERENCES

1. Telecom Canada, *Digital Network Notes*, Ottawa, Canada, 1983.
2. Members of the Technical Staff, Bell Telephone Laboratories Inc., *Transmission Systems for Communications*, 4th Ed., Bell Telephone Labs Inc., 1970. Publisher: Western Electric Co. Inc., Winston-Salem, NC.
3. Mayo, J. S., "Bipolar Repeater for Pulse Code Modulation Signals", *Bell System Technical Journal*, 41, 25–97, 1962.
4. Andrews, F. T., "Bipolar Pulse Transmission and Regeneration", U.S. Patent 2,996,578., Aug. 15, 1961.

5. Gray, P. R. and Messerschmitt, D. G., "Integrated Circuits for Local Digital Switching Lines Interfaces", *IEEE Comm. Mag.*, 18(3), 12–23, 1980.
6. Cirillo, A. J. and Thovson, D. K., "D-2 Channel Bank: Digital Functions", *Bell System Technical Journal*, 51(8), 1972.
7. Stremler, F. G., *Introduction to Communication Systems*, 3rd Ed. Addison-Wesley, Reading, MA, 1990.
8. Briley, B. E., *Introduction to Telephone Switching*, Addison-Wesley, Reading, MA, 1983.
9. Dammann, C. L., McDaniel, L. D. and Maddox, C. L., "D-2 Channel Bank: Multiplexing and Coding", *Bell System Technical Journal*, 51(8), 1972.
10. Members of the Technical Staff, *Transmission Systems for Communications*, 4th Ed., Bell Telephone Labs. Inc., 1970. Publisher: Western Electric Co. Inc., Winston-Salem, NC.
11. Members of the Technical Staff, *Transmission Systems for Communications*, 5th Ed., Bell Telephone Labs. Inc., 1982. Publisher: Western Electric Co. Inc., Winston-Salem, NC.
12. Lathi, B. P., *Modern Digital and Analog Communication Systems*, CBS College Publishing, New York, 1983.

PROBLEMS

- 9.1** What are the advantages of single-sideband suppressed carrier transmission system? What are its disadvantages?

An AM radio transmitter delivers a maximum of 12 kW of radio-frequency power to its antenna when the modulating signal is sinusoidal and the index of modulation is 0.5. Calculate

- (1) the power in the carrier,
- (2) the power in the side frequencies.

Calculate the maximum power that can be radiated if the transmitter were used in a single-sideband-suppressed carrier (SSB-SC) system. What is the power gain offered by the SSB-SC system?

- 9.2** A telephone system uses frequency-division multiplex for 12 channels in a basic group configuration (see Figure 9.2), where two tones of frequencies $f_1 = 400$ Hz and $f_2 = 2.5$ kHz are applied to channel 5 (76–80 kHz). Unfortunately, the amplifier in channel 5 has a distorted characteristic which can be represented by

$$i_0 = av_i + bv_i^2 + cv_i^3.$$

Derive an equation showing all the frequencies present in the output current. Which channels in the basic group are likely to be disturbed by the distortion caused by the channel 5 amplifier?

- 9.3** What are the advantages and disadvantages of a pulse-code modulation system for the transmission of telephone signals?

A telephone system uses frequency-division multiplex for 600 channels in

a basic mastergroup configuration (see Figure 9.4). The output is to be pulse-code modulated using 128 levels. Calculate the minimum sampling frequency allowing for one guard space between PCM samples.

- 9.4** Twelve audio channels in a telephone system are each band-limited to 3.5 kHz and sampled at 8 kHz for transmission in a time-division multiplex PAM system.

- (1) Calculate the minimum clock frequency, allowing for a suitable synchronizing pulse. Where would you place the synchronizing pulse?
- (2) Describe how the output of the PAM system can be used to amplitude modulate a carrier and calculate the bandwidth required to transmit the signal, assuming that it is necessary to include up to the third harmonic of the pulse train.

- 9.5** A pulse amplitude modulated (PAM) wave has the waveform shown in Figure P9.1.

where the mark-to-space ratio is 1. Derive an expression for the PAM

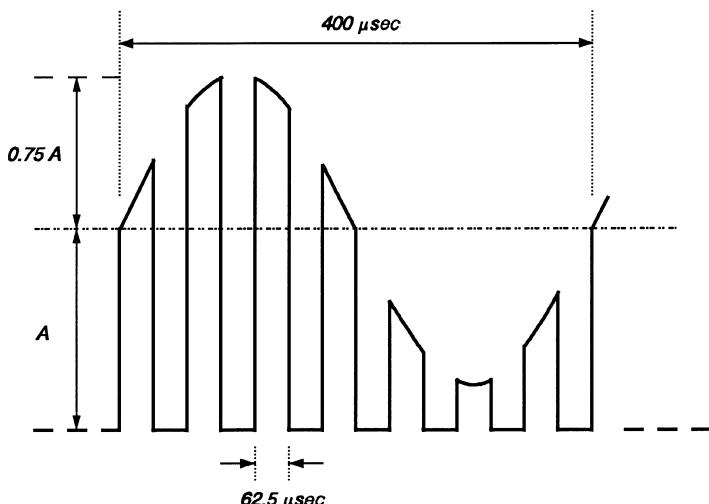


Figure P9.1.

waveform given that, for a square wave,

$$f(t) = \frac{1}{2} + \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{1}{n} \sin n\omega t. \quad (\text{P9.5})$$

- (1) The PAM waveform is applied to the input of an ideal low-pass filter with a cut-off frequency equal to 10 kHz. Find an expression for the output of the ideal low-pass filter. Sketch the waveform.
- (2) The PAM waveform is now applied to the input of an ideal bandpass filter of bandwidth 13 kHz to 19 kHz. Find an expression for the output of the filter. Sketch the waveform.
- (3) The bandpass filter in (2) is replaced by another ideal bandpass filter with bandwidth 45 kHz to 51 kHz. Find an expression for the output of the filter. Sketch the waveform.

Which of the above outputs would you choose in order to recover the information in the original PAM waveform and why?

- 9.6** In the telephone system, for both FDM and TDM, channels are grouped to form hierarchies for transmission. Discuss the advantages of this method of organizing the channels.
- 9.7** What is a ring counter? Illustrate your answer with the design of an eight-phase ring counter. Suggest an application for such a ring counter.
- 9.8** Design a Schmitt trigger to trigger when a voltage in excess of 4.5 V is applied to its input. The dc supply voltage is 12 V and the load driven by the Schmitt Trigger draws a current of 0.5 mA. Two enhancement-mode *N*-channel MOSFET with threshold voltage equal to 3 V are provided.
- 9.9** Design an astable multivibrator to drive a sampling circuit in which the sample time is 1 μ s and the hold time 15 μ s. The dc supply voltage is 6 V and the sampling gate draws a current of 1 mA when it samples the signal. What special precautions would you take in view of the large difference between the sample and hold times?
- 9.10** Design a series sampling gate using a bipolar transistor. The source of the signal to be sampled has a voltage of 5 V peak amplitude with an internal source resistance of 100 Ω . The load “seen” by the sampling gate is 10 k Ω . The transistor is NPN and has $\beta = 100$, $V_{CE(sat)} = 0.5$ V and a base leakage current (when the transistor is off) $I_{E(off)} = 50$ nA. Calculate the error in the output.

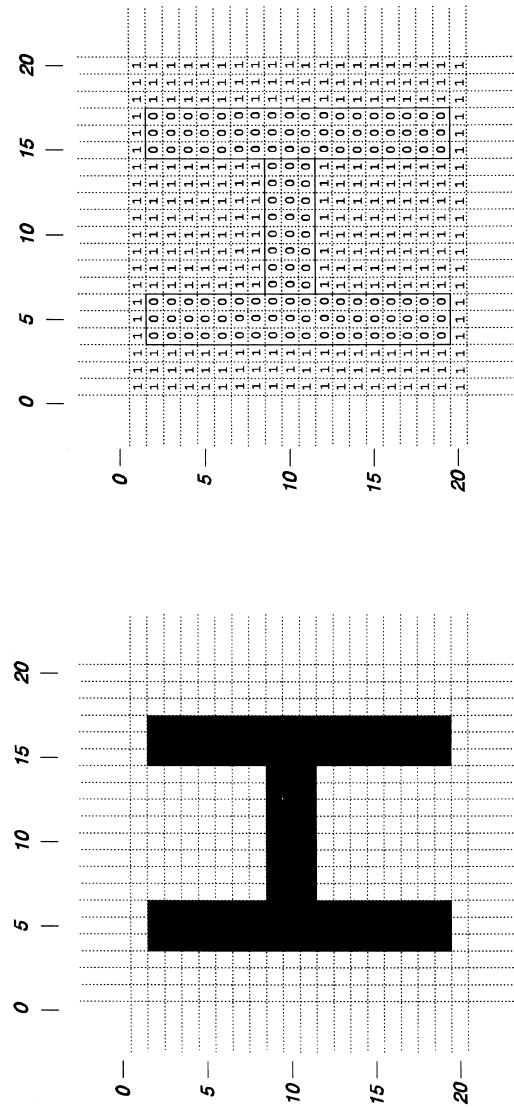
10

THE FACSIMILE MACHINE

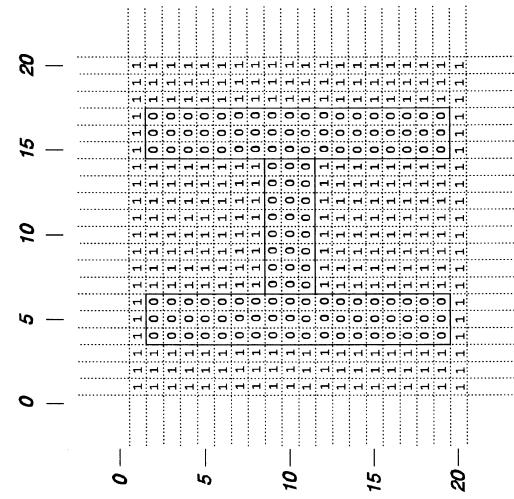
10.1 INTRODUCTION

Although the facsimile machine was invented in the 1840s, it remained largely a device used in the newspaper industry for the transmission of pictures until the mid-1980s. There were several reasons for this; some were technical and the others commercial.

The technical problems which held up the development of the fax machine are illustrated in Figure 10.1. For simplicity we use the letter *H* and assume that scanning is carried out horizontally from the top left side to the right. The scanning head then moves to the second line and the process is repeated. Again, for simplicity, the scanned field is divided into a matrix 20 by 20. Thus 400 pieces of information have to be sent to the receiver in order to reconstruct the *H*. The first task is to measure the level of light reflected or produced by each square and to assign a value of **1** or **0**; we assume here that a **1** is assigned when a square is white and a **0** when it is black (the opposite would work just as well). Figure 10.1(b) shows the result. Each one of these pieces of information is called a *pel* (which is a pixel with its gray scale or color information placed in two categories, black or white, depending on its relative brightness). For the transmission to be successful, the transmitter has to “tell” the receiver precisely which squares are to be left white and which are to be made black. In other words, the “read” head in the transmitter and the “write” head of the receiver must be exactly on their corresponding squares at the same time, that is, they must be in *synchronism* and in *phase*. To obtain synchronism, two pendulums of the same length (with a mechanism for keeping them in phase) were used [1]. The pendulum was not very practical because it had to be made quite large to store enough energy so that the losses during the scanning and printing processes would be negligible. Improved synchronization was obtained when the tuning fork replaced the pendulum but this new technique did not become accurate enough for the purpose until the 1940s. The problem of synchronization was never



(a)



(b)

Figure 10.1. (a) The letter H showing white and black pixels. (b) Binary representation of the pixels.

satisfactorily solved; indeed, the problem disappeared eventually when digital techniques were applied to fax machine development.

Another technical problem which accounts for the slow development of the fax machine was the speed at which the information could be transmitted. We recall from Chapter 1 that the initial attempts to construct fax machines took place before the telephone was invented. The telegraph lines at the time used single wires with ground returns. These were subject to electrical noise mostly generated by electric street vehicles, which were very popular at the time. To compound the problem, the telegraph lines used relays to extend their reach (Morse's relay) and these were inherently too slow to convey the volume of information required to make the fax machine a success. Note that at the minimum rate for scanning a 8.5×11 inch (21.6×28 cm) page (200 lines per inch) the number of pels generated is 2.86×10^6 . Even with modern coding schemes the telegraph lines could not have handled the sheer volume of information in a reasonable time to make this a success.

The next technical problem that had to be solved was the adoption of a suitable method of coding the information so as to reduce the high level of redundancy. We observe from Figure 10.1(b) that row 1 is completely white and hence it is represented by a row of twenty **1**s. A shorter code made up of a few **1**s and **0**s could be used to signal to the receiver to insert twenty **1**s in row 1. Similarly, in row 2, there are two transitions from **1** to **0** and two transitions from **0** to **1**. To convey this information it is possible to devise a code word with less than twenty bits to tell the receiver where the transitions occur and whether they are from **1** to **0** or vice versa. It can also be seen from Figure 10.1(b) that, in our example, row 3 is the same as row 2. A further reduction in the bits required can be achieved by sending a relatively short signal to the receiver to repeat row 2.

At least in the United States, commercial rivalry discouraged cooperation between the engineers who were working on the development of the fax machine. Each manufacturer developed their own standards, different from those of their rivals. The effect of this was to keep costs high; only the military, the police, and the large news organizations could afford to own and operate fax machines. It was not until the mid-1950s that the Institute of Radio Engineers came up with a set of standards which were accepted by the industry.

10.2 SYSTEMS DESIGN

The development of the modern fax machine starts with the adoption of the recommendations of the Consultative Committee of the International Telegraph and Telephones (CCITT) of the standards which came to be known as Group 3 (G3). Group 1 and 2 standards (which incidentally were analog) had been essentially ignored by the North American fax industry. The discussion of G3 standards started in 1976 during the General Assembly of the CCITT which had in the meantime transformed itself into the International Telecommunication Union (ITU), a special agency of the United Nations. By 1980 a full complement of recommendations which form the G3 standards had been adopted. These recommendations, with a lot

of options (to satisfy various special interest groups), covered the parameters for "handshaking" between send and receive machines, modem speeds, scan densities, coding schemes and a system that was completely digital. The decision to adopt digital techniques banished the problems associated with synchronization and phasing. Compromises were made and sometimes these were influenced by the possibility of royalty payments. Two coding schemes were accepted. The first because the patent had expired [2] and the second because the owners of the patent offered it free of charge to users.

10.2.1 The Transmit Mode

Figure 10.2 shows a block diagram of a typical G3 fax machine when it is operating in the transmit mode. The scanner has 1728 charge-coupled devices (CCD) arranged in straight line array. Each element of the array reads the brightness of the first row of pixels and converts these to an equivalent voltage. An amplifier strengthens the signal to a suitable level for the next stage of processing. A Schmitt trigger circuit assigns a value of 1 (white) or 0 (black) to each pixel, thus changing them into pels. This information is stored and the next row of pixels are read by the CCDs and likewise converted and stored. The signal is then coded to reduce the high level of redundancy that is present by using the Modified Huffman (MF), modified relative address, or simply modified READ (MR) or modified modified READ (MMR). The choice of which of these coding systems is used for the transmission depends on the capabilities of the fax machines involved. A table of the modified Huffman codes is given in Appendix D [3]. The coded information is accumulated in the memory and sent to the modem at a time determined by the microprocessor. The modem converts the signal into analog form for transmission along the telephone line to the receiving fax machine. The line adjuster may be used to modify the impedance of the line, the frequency response and/or to minimize echo on the telephone line. The stepper motor drives the mechanical system which advances the sheet of paper through the scanner.

10.2.2 The Receive Mode

Figure 10.3 shows a block diagram of the fax machine when it is in the receive mode. The modem receives its input from the telephone line, converts the analog signal into digital form and stores it in the memory ready for decoding. The decoder reconstructs the original message and it is suitably amplified. The output of the power amplifier drives a thermal printer. A roll of specially treated paper is drawn past a set of hot wires spaced at approximately 200 per inch (same as the resolution of the scanner). The roll is driven by the stepper motor. When the output of the decoder indicates that a pel is black, current flows through the corresponding hot wire and this causes the paper to produce a black spot. All the blocks represented in the diagram are under the control of the microprocessor. The current trend is away from thermal to xerographic and carbon film transfer printers which use ordinary paper.

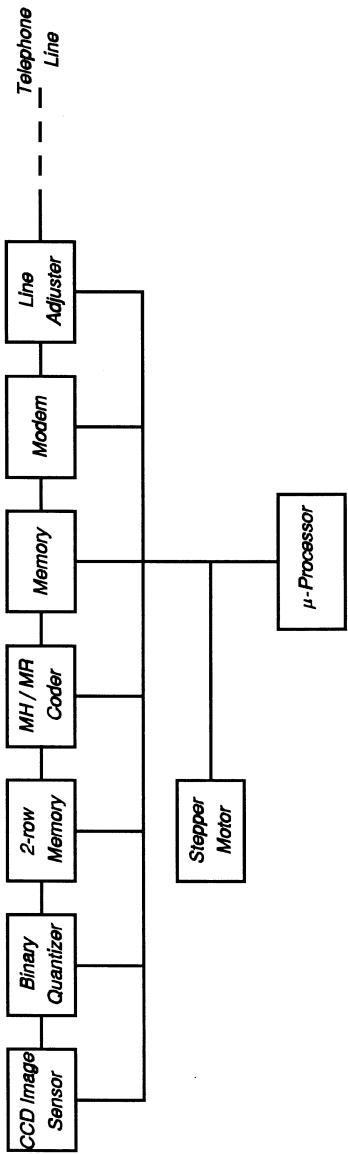


Figure 10.2. The block diagram of the "send" portion of the facsimile machine.

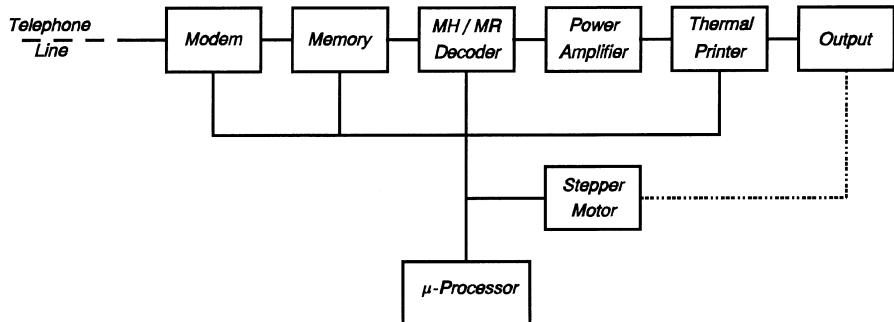


Figure 10.3. The block diagram of the “receive” portion of the facsimile machine.

It is clear that every stand-alone fax machine has within it all the components needed for both the transmit and receive modes.

10.3 OPERATION

10.3.1 “Handshake” Protocol

To send a fax message, one inserts the page into the send machine. The paper is caught between two rollers and it is immediately pulled part-way into the machine. The machine is ready to read the first line of the message. The next step is to dial the telephone number to which the receiving fax machine is connected. The number dialled is stored by the send machine. On pressing the “start” button, the following events take place:

- 1 The dial tone comes on.
2. The send machine dials the number stored and, if the number is not busy, the “ring-back” tone can be heard. Usually, the receive machine needs four rings before it responds.
3. The receive machine goes “off-hook” (connects itself to the line) and sends a 2.1 kHz signal lasting approximately 3 s to the send machine to identify itself as a fax machine.
4. The receive machine follows up by sending its identification code to the send machine. This code tells the sending machine what the capabilities of the receive machine are. The following information is vital and many other options may be included:
 - (a) the speed of the modem,
 - (b) the scan density (number of lines per inch or mm),
 - (c) the type of decoding (MH, MR or MMR) it is programmed to perform,
 - (d) the size of its memory.

5. The send machine then sends a command signal which locks the receive machine into conformity with the chosen attributes from the list in (4).
6. The send machine sends a standard test (training) signal to the receive machine.
7. The receive machine sends a confirmation signal that the test signal was correctly received.
8. The send machine sends the message.
9. If the test signal fails to arrive correctly, there may be options such as telephone line equalization, change of modem speed, etc., or the call may be terminated.
10. At the end of the message a special code is sent to indicate this to the received machine.
11. The receive machine then sends back a code indicating that the message was successfully received.
12. The send machine goes back “on-hook,” terminating the call.
13. The receive machine also goes “on-hook,” ready for the next message.

The handshake protocol is typical of an older version of the fax machine. Newer models may read a complete page at a time, process and store the information before sending it.

10.4 THE TRANSMIT MODE

In this section we examine the components that make up the circuit following the order in which they are encountered by the signal.

10.4.1 The CCD Image Sensor

10.4.1.1 Semiconductor Theory. To understand the operation of the CCD scanner, a short overview of semiconductor theory is necessary. There are a number of materials which can be made into semiconductors. The most widely used of these is silicon. Pure silicon does not conduct electricity under normal conditions because its electrons do not have enough energy to break away from the crystalline structure. When thermal energy, an electric field or light is incident on a piece of silicon, the electrons may acquire enough energy to escape the influence of the nucleus and become “free electrons” within the material. The creation of free electrons in silicon is facilitated by changing the material from a *nonconductor* to a *semiconductor*. In order to turn pure silicon into a semiconductor, it is necessary to introduce an “impurity”, that is, another chemical element. This process is called *doping*. A common element used for doping silicon is phosphorus. Silicon has a chemical valency of four and that of phosphorus is five. The phosphorus therefore introduces an extra electron into the structure of the silicon. The extra electron is then available

for conduction of electricity under the right set of conditions. Phosphorus-doped silicon is referred to as *n*-type (donor) silicon. In *n*-type silicon, the *majority* carriers are electrons and the *minority* carriers are *holes*.

A second element commonly used for doping silicon is boron. Boron has a chemical valency of three and it therefore produces a deficit of one electron. Instead of talking about a deficit of one electron, we call it a *hole*. The boron-doped silicon is referred to as *p*-type (acceptor) silicon. In *p*-type silicon, the *majority* carriers are holes and the *minority* carriers are electrons [4].

Figure 10.4(a) shows a cross section of *p*-type silicon overlaid with silicon oxide (an insulator) and then an aluminum electrode. With 0 V on the electrode, nothing happens, but as the voltage is increased, the positive voltage on the electrode repels the positive holes, creating a *depletion layer* as shown in Figure 10.4(b). As the electrode voltage is increased further, a point is reached where the effect of the electrode voltage has become so strong that it starts to accumulate electrons just below the surface of the silicon. An inversion of the *p*-type to *n*-type silicon, immediately under the electrode, has taken place. This layer is called the *inversion layer*. The value of the voltage at which inversion takes place is called the *threshold voltage*, V_{th} . Beyond the threshold voltage an increasing number of electrons accumulate under the positive electrode. Indeed, a capacitor has been formed between the electrode and the *p*-type silicon. This is shown in Figure 10.4(c) and (d). The structure is capable of generating (when the electrode voltage, V_E , is sufficient) and holding a charge similar to a capacitor.

When light is incident on a piece of semiconductor material, electron–hole pairs are generated. This happens when the energy of the photon is absorbed by the material causing the excitation of bound carriers into mobile states. Given suitable structures and the appropriate biasses, the minority carrier (that is, the electrons in *p*-type material in the mobile state) can be collected. The effect of the light is essentially the same as that of an electric field or a rise in temperature [5]. This is the basis for a large number of sensors such as phototransistors and infrared detectors.

10.4.1.2 Semiconductor Light Sensor. Figure 10.5 shows a typical structure of an image sensor [6]. The *p*-type silicon is formed into islands (wells). Each island is a cell and, using a suitable mask, the cells can be selectively exposed to light. Note that the light enters the cell from the back side of the silicon chip where the electrode is not in the way. Using the appropriate thickness of silicon, the correct level of doping in the silicon, and applying the correct voltage to the electrode, the electrons can be captured near the surface of the silicon just below the electrode. The number of electrons captured is almost linearly related to the intensity of the incident light. We have an *electronic image* of the mask “painted” in the form of charge.

10.4.1.3 The “Bucket Brigade”. The role of the “bucket brigade” is to transport the charge which forms the electronic image to an external circuit. The structure of the device shown in Figure 10.6 is similar to that in Figure 10.5. The difference is that there are four electrodes which are very closely spaced. We use the analog of a well filled with water to illustrate the action of the “bucket brigade” [6].

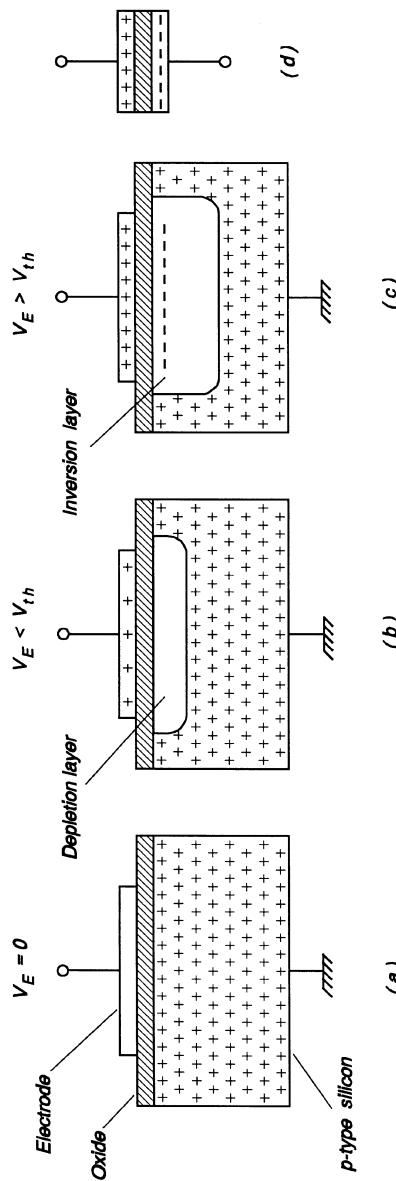


Figure 10.4. (a) The structure of the *p*-type silicon device with no voltage applied to the electrode. (b) The formation of the depletion layer when a positive voltage ($V_E < V_{th}$) is applied to the electrode. (c) With increasing value of V_E ($V_E > V_{th}$) the inversion layer is formed. (d) The comparison of the device to a capacitor.

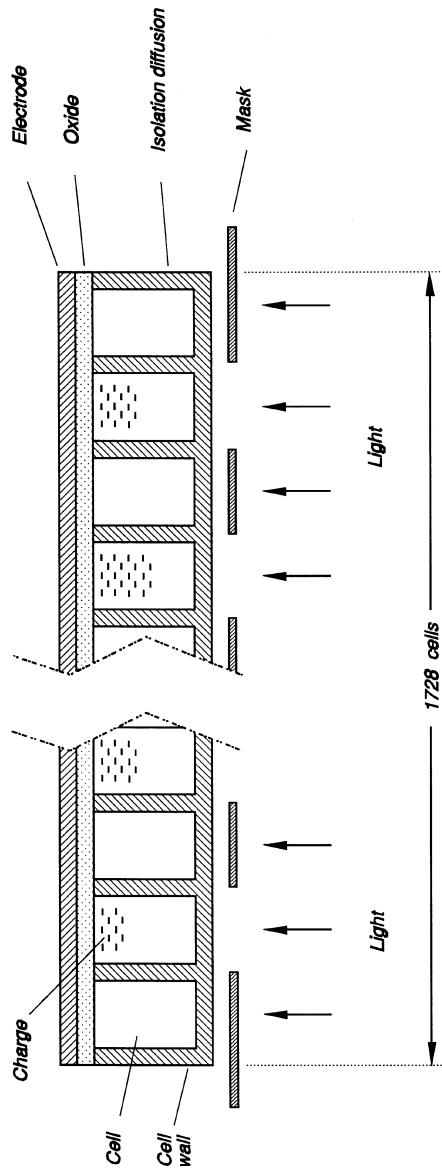


Figure 10.5. The image sensor showing the 1728 cells. The electrons shown in the cells were generated by the incident light.

When a positive voltage of sufficient magnitude is applied to one of the electrodes, a “well” (inversion layer) of charge is formed. The “well” is filled with water (negative charge, that is, electrons). We assume that there is water in the well under “electrode *b*” because of its higher voltage. The water stays there because there are no wells under *a* and *c*. Suppose we dig a well under electrode *c* (that is, change the voltage on electrode *c* from 2 V to 10 V). Clearly the water will flow into a well whose width has doubled, hence the level of the water will drop to one-half of the original (the quantity of water is assumed constant). Finally, we fill in the well under electrode *b*. The water level in the well under electrode *c* will go up to the original level. We have succeeded in moving the water one step to the right.

The charge stored under electrode *b* can be shifted to electrode *c* using the voltage waveform shown in Figure 10.6(*e*). Note that, to ensure that the charge is shifted unambiguously from left to right, the two electrodes on opposite sides of electrode *b* must be connected together. Similarly, every third electrode must be connected together as shown in Figure 10.7.

A combination of the light sensor and the “bucket brigade” (or analog shift register) makes it possible to “measure” the light incident on the image sensor and to transmit the information to an external circuit.

10.4.1.4 The CCD Image Sensor Unit. A combination of the light sensor and the “bucket brigade” is shown in Figure 10.7. To avoid the production of stray signals, a metal mask (not shown) shields all structures other than the light sensor.

To start the operation, a positive voltage is applied to the photogate and the transfer gate is “closed” by the application of a negative voltage. Charge proportional to the level of illumination and also to time of exposure accumulates in each cell. After the so-called *integration time*, the transfer gate is “opened” and the charges accumulated in the cells flow through to the electrode which is most positive of the set of three. This is a parallel process. The transfer gate is then closed and the three-phase clock drives the accumulated charges from left to right. As each packet of charge arrives at the terminal, it is converted into a voltage and suitably amplified. This is a serial operation.

10.4.1.5 Mechanical and Optical Systems. Figure 10.8 shows a typical optical and mechanical system used for scanning the page. A cylindrical lens concentrates the light from a fluorescent lamp onto the document. A spherical lens brings into focus a row of pixels formed from a line 0.005 inches (0.127 mm) wide (200 lines per inch). The document is moved one row at a time by the stepper motor until the page is completely scanned.

10.4.2 The Binary Quantizer

The voltage output from the CCD image sensor is a series of pulses whose amplitudes are proportional to the light incident on each cell. These are pixels and they must be converted into pels because the Group 3 fax standards are designed to operate with black and white levels only. How the G3 fax standard deals with

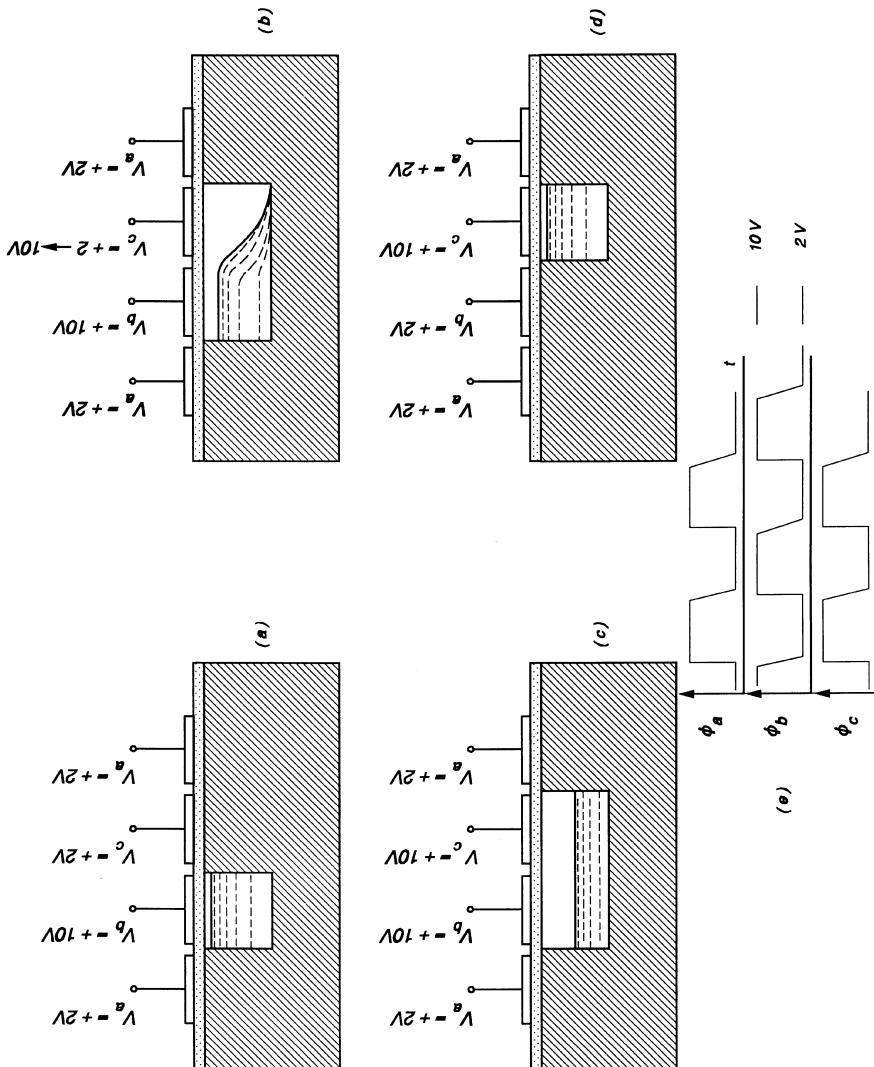


Figure 10.6. The hydraulic analog of the “bucket brigade” showing how the charge is transferred from left to right using the three-phase clock. Reprinted with permission from J. D. E. Beynon and D. R. Lamb, *Charge-Coupled Devices and their Applications*, McGraw-Hill, London, 1980.

Publisher's Note:
 Permission to reproduce
 this image online was not
 granted by the copyright
 holder. Readers are kindly
 requested to refer to the
 printed version of this article.

Figure 10.7. The CCD image sensor unit is the combination of the light sensor and the “bucket brigade”. Reprinted with permission from J. D. E. Beynon and D. R. Lamb, *Charge-Coupled Devices and their Applications*, McGraw-Hill, London, 1980.

shades of gray will be discussed in Section 10.9. The binary quantization is achieved by using a Schmitt trigger circuit. The design of Schmitt triggers is described in Section 9.3.3.1.

10.4.3 The Two-Row Memory

The two-row memory is required when a differential type coding system, such as the Modified READ, is used. There are many types of memories in which the pels can be stored. Because of the serial nature of the information coming from the binary

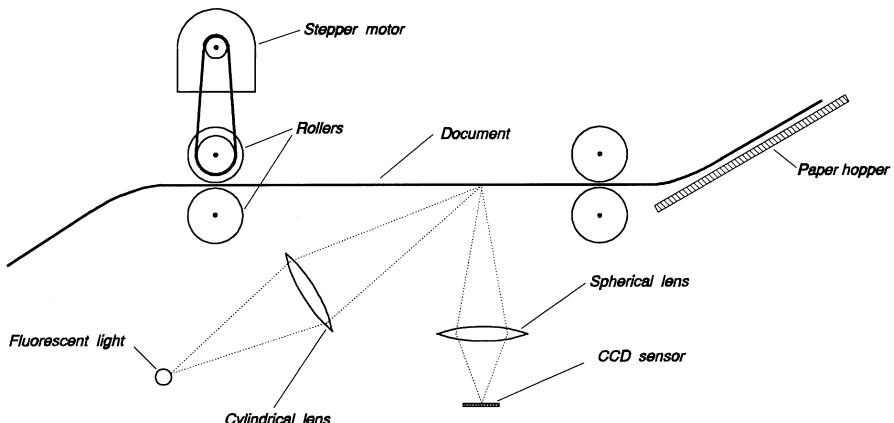


Figure 10.8. A typical optical and mechanical system used for scanning a page.

quantizer, we choose the shift register. A shift register is a string of flip-flops or bistable multivibrators which are connected such that the states of the individual flip-flops contain the information stored. In this type of shift register the data to be stored is entered one bit at a time in response to a clock pulse. Similarly, the information has to be “clocked out” one bit at a time.

The design of flip-flops or bistable multivibrators was described in Section 8.4.5.2. Briefly, a flip-flop is a digital circuit whose output(s) can take only two possible values, **1** or **0**. Whichever value it takes remains until the arrival of a clock pulse, at which point it switches to the other state. We can say that this circuit “remembers” its state before the clock pulse arrived. The flip-flop is the building block for memories of this type. In Section 8.4.5.2, the flip-flop was described as two NOT gates connected back-to-back. That was the non-resettable flip-flop. In order to build a shift register, we require a resettable flip-flop. The design of the resettable flip-flop and its application in the construction of a memory are presented in Appendix E.

10.5 DATA COMPRESSION

10.5.1 The Modified Huffman (MH) Code

As mentioned earlier, the data obtained from the scanner contains a huge amount of redundant information. The MH code is based on the lengths of the run of white and black pels [2]. It is relatively easy to count the number of white pels before a transition to black occurs and then to count the black ones to the next white pel. The number of white pels in the run is coded in the form of a n -bit word. Similarly, the number of black pels is given a different n -bit word. Further to this, a number of typical documents were scanned and the frequencies of the lengths of black and white runs plotted in the form of a histogram. The most frequently occurring run lengths were assigned shorter codes. This is similar to the approach used by Morse when he developed his code; he chose the most frequently occurring letter in the English language ($e = \text{dot}$) and assigned it the shortest code. The MH code is given in Appendix D [5]. Table D.1 gives the codes for run lengths from 0–63 and Table D2 gives the codes from 64–1728. If the fax machine scanned a row that was completely white (that is, 1728 white pels), it would transmit the code **010011011** (from Table D.2) followed by the terminating (that is, a run of length zero) code **00110101** (from Table D.1). Thus a white row with 1728 bits is reduced to a 17-bit code (plus a few more for “housekeeping” purposes such as when to start a new line and so on). Note that, because a page is likely to have more white spaces than black, in general the codes for the white runs are shorter than those for black.

10.5.2 The Modified READ Code

The Modified READ (MR) code takes advantage of the fact that in typescripts and manuscripts there are a lot of vertical or near-vertical lines. It therefore compares one

row of pels to the one immediately above it and codes the differences. Offsets of zero, ± 1 , ± 2 and ± 3 are allowed. This is described as two-dimensional coding. The modified modified READ (MMR) is an extension of the modified READ with an error correction option.

10.6 THE MODEM

The design of a modem was discussed in Section 9.4.1. G3 fax standards have several options for the speed (bit/s) of the modem. The two machines involved in a transmission “negotiate” the speed at which the data will be sent. The following are standard speeds used in fax transmission: 4.8 kbps, 9.6 kbps, 14.4 kbps, 21.6 kbps.

10.7 THE LINE ADJUSTER

As mentioned earlier, during the “handshake” between the two fax machines, a standard test (or training) signal is sent from the sender to the receiver. If errors occur during the training, one of several parameters can be adjusted such as the frequency characteristics of the telephone line. This is referred to as equalization. The telephone line has the frequency characteristics of a low-pass filter and the loss of gain at the higher end of the spectrum can cause errors during transmission. This can be corrected with an amplifier which boosts the high-frequency gain relative to the low end. Other telephone line characteristics such as echo can be modified to improve the quality of the transmission.

10.8 THE RECEIVE MODE

The design of the modem was described in Section 9.4.1, and electronic memory was discussed in Appendix E. The MH/MR decoder is a reversal “look-up” table of that given in Appendix D.

10.8.1 The Power Amplifier

The power amplifier has to produce enough power to heat the temperature-sensitive paper to about 110°C to make a dark mark. The print head is a resistor with very low thermal capacity. With a scanning resolution of 200 lines per inch (both vertical and horizontal), a page 8.5×11 inches has 1700 elements across (the actual number used is 1728) and 2200 elements from top to bottom. Assuming a print rate of a page per minute, each row of pels must be printed in less than 27 ms. The temperature of the resistors have to reach their operating value and drop down (low enough not to cause a smudge on the line below it) in much less time.

As the input signal to the amplifier is a pulse, the amplifier does not have to be linear. A simple emitter follower is adequate. Figure 10.9(a) shows an emitter

follower, also called a common collector amplifier. The resistor R_e is the resistance of the printer head, R_1 controls the base current of the transistor.

Example 10.8.1 The emitter follower. This is a special case design of the emitter follower. A normal linear emitter follower is biased to handle sinusoidal waveforms. In this case the signal is a pulse so it is quite adequate for the transistor to be in the off state until the pulse is applied to its base, then current flows in R_e . The capacitor C is called a “speed-up” capacitor because it helps to reduce the transition time from on to off and vice versa.

The power required to heat up the print head is usually about 0.5 W. Let $V_{cc} = 12$ V. The transistor is a silicon NPN device with current gain $\beta = 100$. Determine suitable values for R_e and R_1 .

Solution. Assume that the input pulse goes from 0 to 12V and also that output pulse goes from 0 to 10 V. Since the power is 0.5 W it follows that the current

$$I_e = \frac{10}{200} = 50 \text{ mA.}$$

The base current

$$I_b = \frac{I_e}{\beta} = \frac{50}{100} = 0.5 \text{ mA.}$$

Since the emitter voltage is 10 V, it follows that the base voltage must be 10.7 V (silicon transistor).

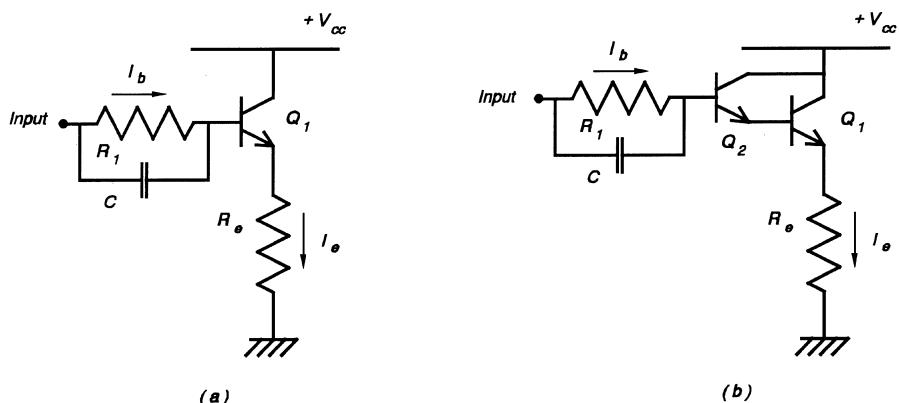


Figure 10.9. The power amplifier (emitter-follower) used for driving the thermal paper printer head.

The voltage drop across R_1 is therefore 1.3 V. The value of

$$R_1 = \frac{V_1}{I_b} = \frac{1.3}{0.5 \times 10^{-3}} = 2.6 \text{ k}\Omega.$$

A suitable value for the capacitor C is 100 pF.

The base current of 0.5 mA may be excessive for the output of the decoder to provide. In that case, the circuit shown in Figure 10.9(b) may be used. This combination of transistors is known as a *Darlington pair* and the effect of the second transistor is to increase the input resistance (decrease the current required to drive the amplifier by a factor of approximately β^2) of the amplifier. The design steps follow from the one above.

10.8.2 The Thermal Printer

The most significant part of the thermal printer is the print head which consists of 1728 heating elements spaced at 0.005 in (0.13 mm) apart. A cross section of the printer showing the essential parts is illustrated in Figure 10.10. Details of the print head are also shown.

The stepper motor advances the thermal sensitive paper a distance of 0.13 mm at a time. Current is supplied to heat the elements of the print head corresponding to the places where there are dark spots.

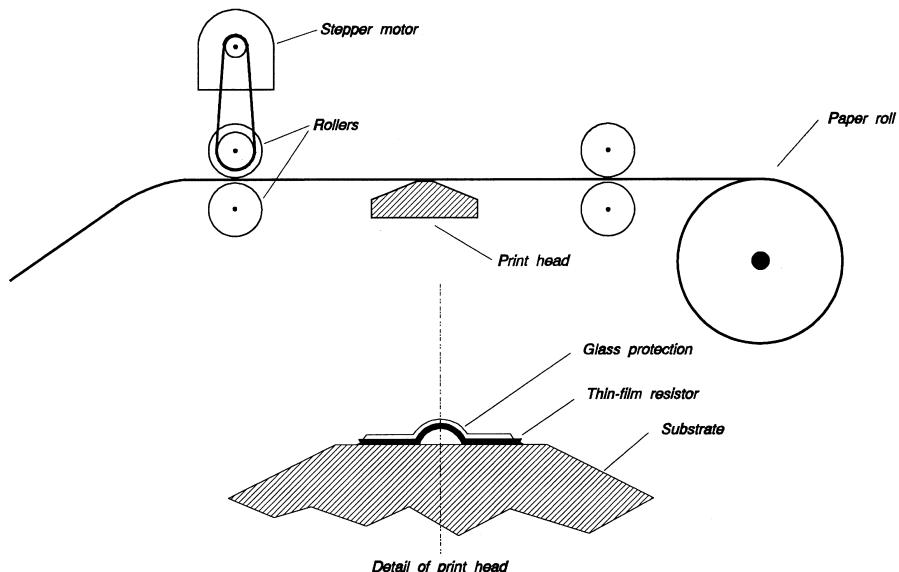


Figure 10.10. A cross section of the printer showing some of the essential parts. Details of the print head are also shown.

The print head heating elements can be made of thick or thin film. They are protected from abrasion by a layer of glass, as shown.

The impressions on a thermal fax paper tend to fade when exposed to sunlight and/or moderate temperatures. Higher quality models of fax machines have other types of printers. Some use xerography, carbon film transfer and ink-jet systems.

10.9 GRAY SCALE TRANSMISSION: DITHER TECHNIQUE

The G3 fax standards do not take into account the need to transmit images with gray scales. The result is that the transmission of documents with print and line drawings produce excellent results, but those with gray-scale information come out with high levels of distortion. To produce an acceptable gray-scale reproduction while using a binary form of coding, a technique known as *dither* was developed. This technique relies on the human eye to perform an “averaging” function when it looks at areas of varying densities of black pels; the more black pels per square, the darker the area

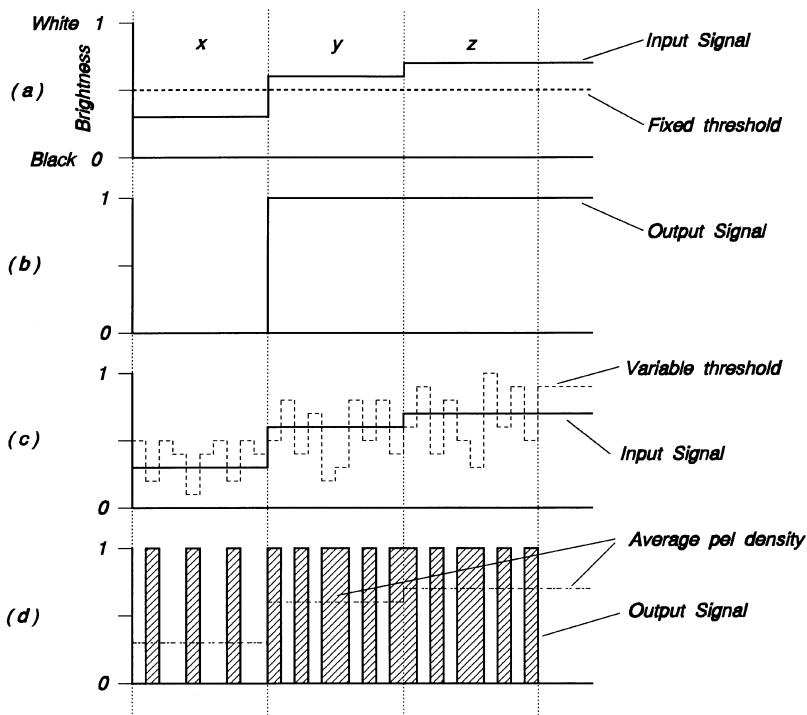


Figure 10.11. (a) Shows three discrete gray level signals and the constant threshold used to digitize them. (b) Shows the output signal from (a). (c) Shows the same three discrete gray level signals and the variable threshold used to digitize them. (d) Sector *x* has a lower density of black pels than sector *z* and therefore sector *x* appears to be a lighter shade of gray than sector *z*.

appears. This is similar to the use of dots of varying sizes to produce pictures in newspapers.

In a typical G3 fax machine, the scanner produces an analog output, with each pixel proportional to the brightness of the image. The signal is converted into a binary form by setting a threshold of brightness above which the signal is given the value **1** (white) and below which it is **0** (black). Figure 10.11(a) shows three discrete gray level signals and the threshold to be used to digitize them.

In the dither technique, the threshold for the binary coding system is varied so that its average value is proportional to the brightness (gray scale), as shown in Figure 10.11(c). The output signal is shown in Figure 10.11(d). For the gray scale represented by sector *x*, the black pel density is 70% while that for sector *z* is only 40%. Therefore, sector *x* appears darker than sector *z*.

REFERENCES

1. Costigan, D. M., Fax: *The Principles and Practice of Facsimile Communication*, Chilton Press, Philadelphia, 1971.
2. Huffman, D. A., "A Method for the Construction of Minimum Redundancy Codes" *Proc. IRE*, 40, Sept. 1952.
3. McConnell, K., Bodson, D. and Urban, S., *Fax: Facsimile Technology and Systems*, 3rd Ed., Artech House, Boston, MA, 1999.
4. Mauro, R., *Engineering Electronics: A Practical Approach*, Prentice-Hall, Englewood Cliffs, NJ, 1989.
5. Chilian, P. M., *Analysis and Design of Integrated Electronic Circuits*, Harper & Row, New York, 1981.
6. Reynon, J. D. E. And Lamb, D. R., *Charge-Coupled Devices and their Applications* McGraw-Hill, London, 1980.

BIBLIOGRAPHY

Dennis, P. N. J., *Photodetectors: An Introduction to Current Technology*, Plenum Press, New York, 1986.

Quinn, G. V., *The Fax Handbook*, TAB Books, Blue Ridge Summit, PA, 1989.

GLOSSARY

Pixel. A picture element that has more than two levels of gray-scale information or color.

Pel. A picture element which has only black and white information (that is, a pel is a pixel in which the average gray-scale level has been used to quantize it into a binary form; black or white).

Bit. A binary digit, either **1** or **0**.

Word. A set of bits; usually there are 8, 16, 32, 64 or more bits.

Read. Retrieve information stored in a memory.

Write. Store data in a memory.

Address. Location of data stored in a memory.

PROBLEMS

- 10.1** Compare and contrast the state of facsimile technology before and after the development of inexpensive electronic memory.
- 10.2** Compare and contrast the state of facsimile technology before and after the application of digital techniques.
- 10.3** Why is it necessary to code the output of the scanner in a fax machine before transmission?
- 10.4** What modifications would you make to the scanner of a fax machine so it can be used as a television camera?
- 10.5** Describe briefly what happens when two fax machines are involved in a “handshake”.
- 10.6** How do the G3 fax machine standards deal with different shades of gray?
- 10.7** What are the advantages and disadvantages of the thermal printer? Describe other technologies that have superior performance.

11

PERSONAL WIRELESS COMMUNICATION SYSTEMS

11.1 INTRODUCTION

The idea that a person could carry around with him a telephone booth of his own is a very attractive one. However, the technology required to make the telephone booth small and light enough for this to be possible, and furthermore convenient to carry, has been available only in the last 30 years. Mobile radio has been available in North America since the 1930s but they were exclusive in the hands of the police. Later taxicab operators installed radios in their vehicles. These mobile units were large, heavy and electrical power hungry. They used amplitude modulation which is notorious for poor performance in the presence of electrical noise and there was more than enough noise generated by the ignition systems of the vehicles in which they were installed. Moreover, they were operational in either the transmit or the receive mode at any given time; they were of the “push-to-talk” type.

The invention of the transistor and its progression to integrated circuits made it possible to reduce the weight and size of circuits and, at the same time, increase their capability and flexibility. These advances were accompanied by an enormous reduction in the amount of power required to operate transistor circuits. The stage was set for the introduction of the “personal telephone booth”.

There are many applications in which radio plays a vital part. These go from “remote keys” for the automobile, garage door openers, pagers, walkie-talkies, cordless telephones to cellular telephones with access to the Internet. In this chapter, we limit ourselves to a discussion of paging systems, cordless and cellular telephones. The paging system was designed to send information from a base station to a mobile terminal. The mobile terminal has no capability to transmit information in the opposite direction. A serviceman for home heating furnaces, for example, only needs to know the address of his next assignment; in general, he does not have to contact his base office. A communication system in which information travels in only one direction is described as *simplex* and the pager is an example.

Radio systems with a push-to-talk button use the same channel in both the forward and reverse directions. It is therefore necessary for each person to indicate when they have finished talking with the familiar word “roger”. They are described as *half-duplex*. A *full-duplex* system uses two channels simultaneously, the first for transmission and the second for reception. The cellular telephone is an example of a full-duplex system.

11.2 MODULATION AND DEMODULATION REVISITED

In Section 9.2.1 we discussed the generation of a single-sideband-suppressed carrier (SSB-SC) signal using a balanced modulator and a bandpass filter. Fig. 11.1(a) shows the circuit configuration of the SSB-SC as well as the frequency spectrum of the input and output. The output shows that the baseband signal has experienced an upward frequency shift equal to the carrier frequency, ω_c and an “inverted” version of it appears at a lower frequency and the two are symmetrically spaced about the position of the carrier. A bandpass filter is used to select the upper sideband. Clearly, the upper and lower sidebands contain the same information and only one of them should be required for the recovery of the original signal.

Figure 11.1(b) shows a circuit in which upper sideband is multiplied (balance modulated) with the carrier signal, $\cos \omega_c t$. The corresponding spectrum shows that the

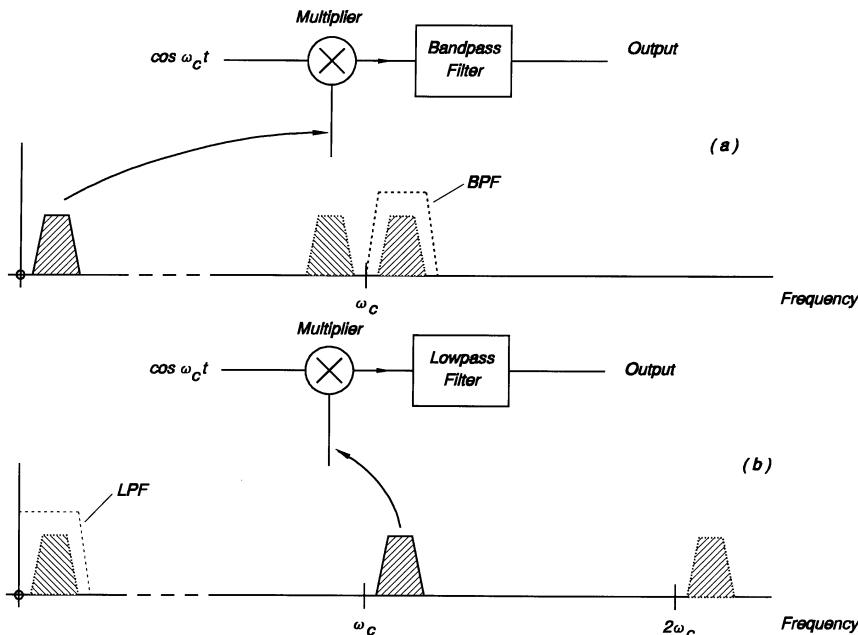


Figure 11.1. (a) The structure of the modulator and the spectrum of the corresponding frequency shift. (b) The structure of the demodulator and the spectrum of the corresponding frequency shift.

output has two “sidebands”. The first is at a frequency $2\omega_c$ and the second occupies the position of the original baseband signal in the spectrum. The original signal is recovered by using a lowpass filter. Equations (9.2.1 to 9.2.5) are the relevant equations.

This modulation and demodulation technique can be used with amplitude, frequency and phase or angle modulation schemes. When demodulation is carried out using the carrier signal in a balanced modulator as shown in Figure 11.1(b), it is referred to as *coherent demodulation* or *coherent detection*. The use of an envelope detector to demodulate an AM signal is known as *non-coherent demodulation*.

11.3 ACCESS TECHNIQUES

11.3.1 Multiplex and Demultiplex Revisited

When modulation is used to accommodate a number of signals on a single channel we refer to it as multiplexing. Figure 11.2 shows five baseband signals, each of which occupies the frequency band 300 Hz to 3 kHz.

By choosing suitable carrier frequencies for each one, they may be transmitted over the same cable or the airwaves by radio and subsequently demodulated with no interference between them. When different carrier frequencies are used to multiplex the baseband signals, it is referred to as *frequency-division multiplex* (FDM). Other methods of multiplexing are described below.

The success of personal wireless communication systems is in part due to the development of techniques which allowed a large number of signals to share a limited spectrum. One of the boundaries of the spectrum available for personal wireless communication is dictated by the size of the antenna for the radio interface. Efficient transmission of radio signal at low frequency requires antennas several thousand meters tall. Clearly, this is not possible as portability of the device is essential. The boundary on the other end of the spectrum is set by the character of high-frequency transmission which increasingly takes on the properties of visible light which requires line-of-sight. Clearly, the modern environment (cities) in which most of the potential subscribers live and work make line-of-sight communication

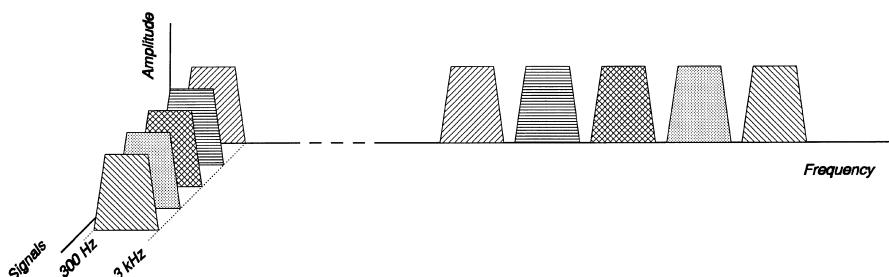


Figure 11.2. Five baseband signals occupying the same bandwidth can be separated by using frequency-division multiplex.

devices inadmissible. Between these two boundaries we have other systems in competition for the spectrum, such as air and sea navigation, satellite communication, radio and television broadcasting.

It so happens that national governments have arrogated to themselves the power to assign portions of the spectrum for specific purposes within their territories and to negotiate international treaties which govern their use. This, in short, brings us to the assigned frequency bands of 824–849 MHz and 869–894 MHz for personal wireless communication. For the large number of anticipated subscribers to be accommodated in such a restricted bandwidth it is necessary to develop techniques which reduce the possibility of interference with each other.

One major factor working in our favor is that, provided we keep the radiated power below a given level, and we are separated sufficiently by distance, we can reuse the spectrum over and over again. We shall now discuss the techniques which enable us to share the spectrum available.

11.3.2 Frequency-Division Multiple Access (FDMA)

Frequency-division multiple access is a fancy name for what is commonly done with AM and FM radio broadcasting and TV stations; they are assigned different carrier frequencies with suitable separation between them to ensure minimal interference with each other. They are required by law to keep their carrier frequencies constant. They also have a limited bandwidth and radiated power. The division of the spectrum according to frequency was discussed in Section 9.2 under the heading “Frequency-Division Multiplex” (FDM). Figure 11.3 shows a representation of the channels spaced by their assigned carrier frequencies and separated by limited bandwidth and appropriate guard bands.

11.3.3 Time-Division Multiple Access (TDMA)

In FDMA, a frequency band is dedicated to a particular channel for as long as it is required. In TDMA, several channels share the same bandwidth but each channel has the use of that bandwidth for a fraction of the time. TDMA was discussed in Section 9.3 under the other name used to describe this technique: “Time-Division

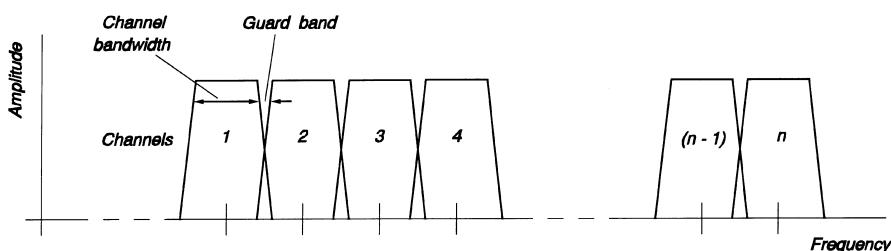


Figure 11.3. In frequency-division multiple access (FDMA), channels are spaced by their assigned carrier frequencies and separated by limitation on bandwidth and appropriate guard bands.

Multiplex" (TDM). The basis of this technique is the ability to reconstruct a signal from samples taken from it. Figure 11.4 shows how each channel is structured in time to form frames and the sequences of the content of each channel. In TDMA, it is necessary to synchronize the transmitter to the receiver so that bits from one channel do not end up in another channel, hence the synchronizing bits.

11.3.4 Spread Spectrum Techniques

In spread spectrum communication systems the radio-frequency carrier is changed very rapidly in a pseudo-random fashion over a bandwidth which is much wider than the minimum required to transmit the signal. Potentially it should cause interference with other users of the airwaves but, in fact, because the carrier operates for such a short time at any given frequency, its effect is almost imperceptible. The average perceived power on any given channel is very low and it therefore behaves like a low-power noise source spread across the bandwidth it uses. Many communication channels can operate in this fashion without interfering with each other. Spread spectrum technology has been of particular interest to the military because it is almost impossible to predict the next frequency of the transmission; they like to stay away from eavesdroppers and to avoid the jamming of their communication systems by the enemy. The real challenge in spread spectrum communication is to keep the receiver synchronized to the transmitter. We shall return to the problem of synchronization later.

There are two major types of spread spectrum techniques. They are *frequency hopped* and *direct sequence* spread spectrum technologies.

11.3.4.1 Frequency Hopped Multiple Access (FHMA). In FHMA transmission, the information is first digitized and then broken up into short passages. Each passage is transmitted on a different carrier frequency determined by a pseudo-random number generator. Because the modulation used is either narrow band FM or frequency-shift keying, at any instant, a frequency hopped signal occupies a single narrow channel. However, because the carrier frequency hops around, it makes use of a much wider bandwidth. Figure 11.5 shows a representation of a system that uses FHMA. Clearly, in an FHMA the receiver has to have prior access to the sequence of the carrier frequencies transmitted as well as the timing to be able to follow the hops (synchronize). It is quite likely that two or more transmitters will at some time try to use the same frequency.

11.3.4.2 Code Division Multiple Access (CDMA). In CDMA transmission, the information is first digitized and then multiplied by a binary pseudo-random sequence of bits (called *chips*) with a bit rate much higher than that of the digitized information [1]. Figure 11.6 shows the binary message signal, bits of a pseudo-random code, and the spread spectrum (coded) signal.

Note that, because the bit rate of the pseudo-random sequence is much higher than that of the message signal, it requires a much larger bandwidth for its transmission. The spread signal is used to modulate a carrier (usually FM or PM)

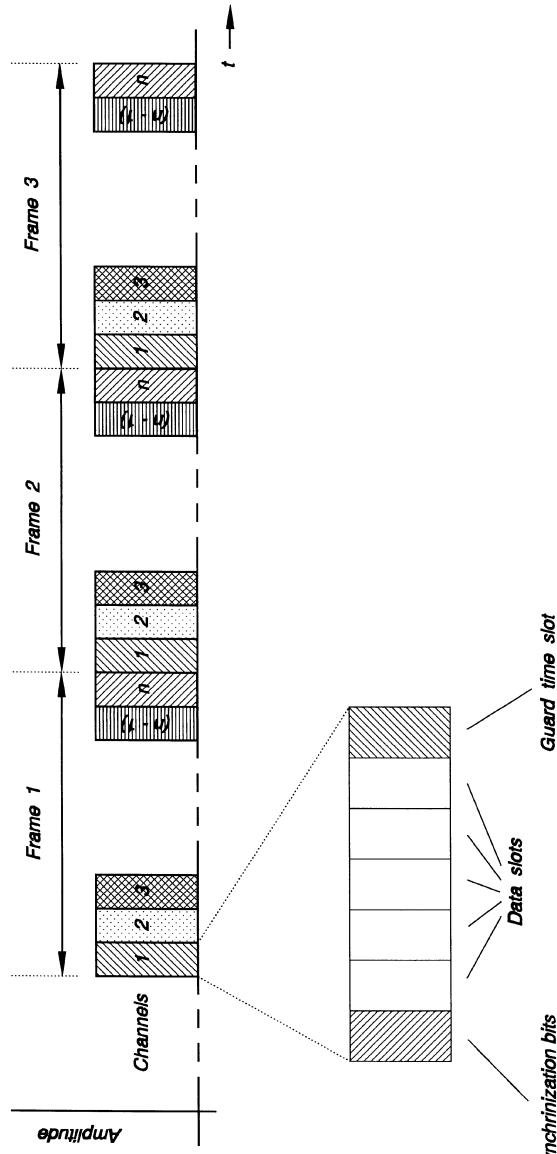


Figure 11.4. In time-division multiple access (TDMA), each channel is structured in time to form frames and the sequence of the content of each channel.

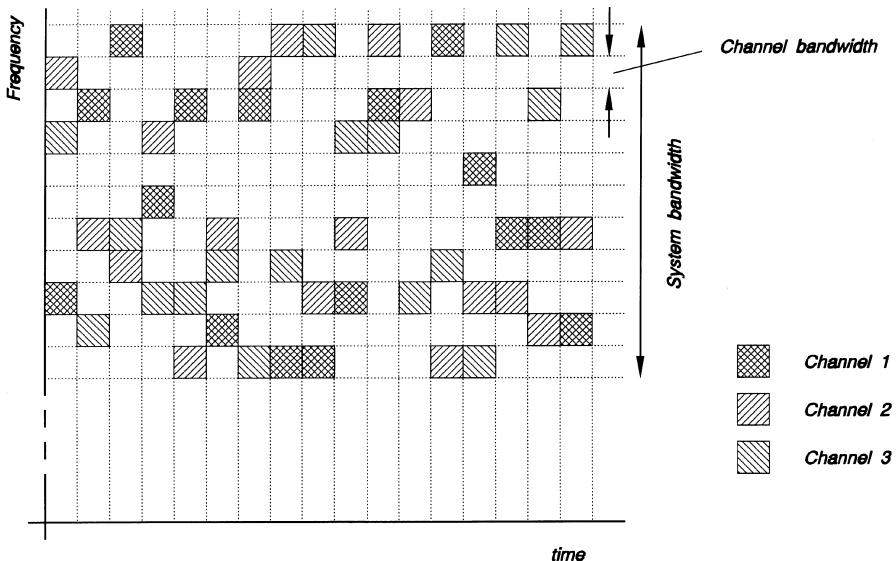


Figure 11.5. A representation of a frequency hopped multiple access (FHMA) system. Although no instances of two or more transmission on the same frequency and at the same time are shown, there is a clear possibility that this can happen. Note that for simplicity, all channels have equal bandwidth and occupy that bandwidth for the same length of time. Neither of these conditions apply in practice.

and then transmitted. At the receiving end the spread signal is demodulated then decoded using a locally generated pseudo-random bit sequence in a process called *correlation*. Because the number of chips representing a message bit (**1** or **0**) is large, the correlation does not have to be perfect; it has to correctly recognize the majority of the chips as representing that message bit (**1** or **0**). In a system which is subject to multipath fading, this is an advantage. It is clear that the receiver has to have prior knowledge of the pseudo-random code to be able to decode the message. To other receivers not using the identical code, the message appears to be just noise. The attraction of this technique is that it can be used to accommodate a large number of subscribers with different codes and they will not even know that they are sharing the same bandwidth. A by-product of CDMA is improved security of the message.

One disadvantage of CDMA is that the power of individual transmitters has to be controlled very carefully. A strong signal from one of the transmitters within the wideband can overwhelm the sensitive front-end of the system and prevent the reception of other signals. The transmit power control system for all the mobiles adds complexity and costs.

11.4 DIGITAL CARRIER SYSTEMS

So far, we have discussed carrier systems in which the message signal is in analog form. Increasingly, electronic systems are using a digital format. For example, it has

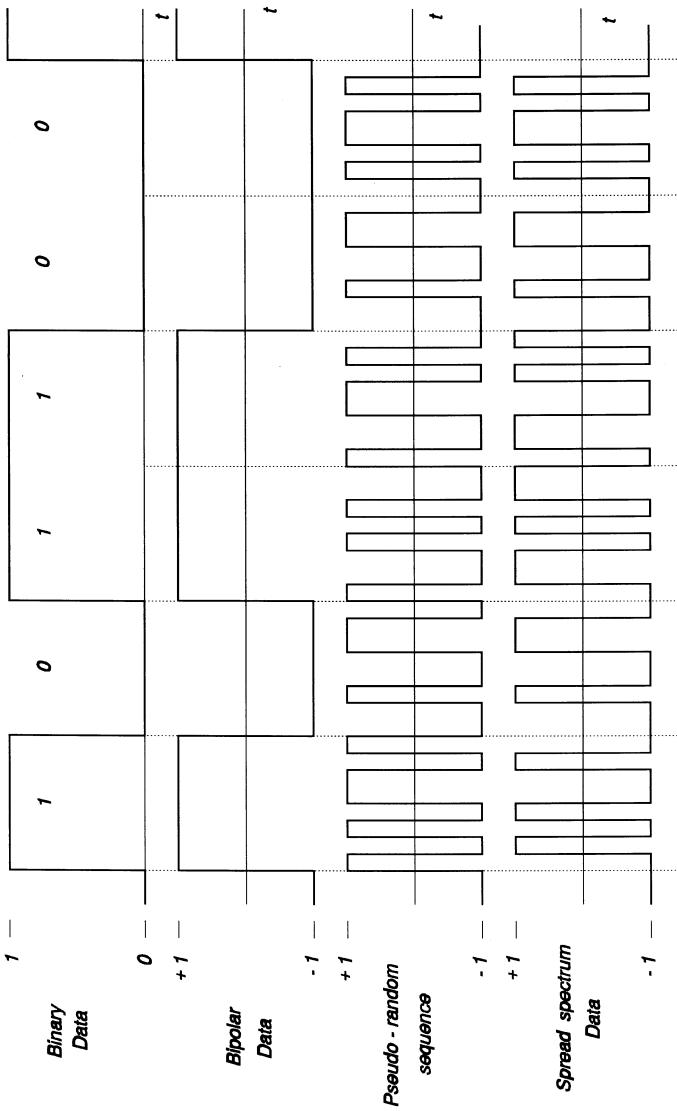


Figure 11-6. The binary data and its equivalent bipolar form, the pseudo-random sequence (chips) and the resulting spread spectrum data.

taken the music recording industry less than 15 years to replace the analog vinyl record with the digital compact disc. There are technical as well as economical advantages to be gained from this move. Moreover, the advent of integrated circuit technology with its ability to fabricate extremely large numbers of circuits on minuscule pieces of semiconductor has made the move to digital systems seem inevitable.

To transmit a baseband (message) signal over a radio channel, it is necessary to change some property of the radio-frequency signal using the baseband signal. We can change its amplitude, its frequency, or its phase angle. In Chapter 2 we discussed the modulation of a radio-frequency signal by a message signal in which the amplitude of the RF signal varied according to the amplitude of the message signal (amplitude modulation; AM). In Chapter 4 we discussed how to change the frequency of the RF about a fixed value using the message signal (frequency modulation; FM). It is now time to discuss the modulation scheme in which we vary the phase of the RF signal according to the message signal (phase modulation; PM). It must be pointed out that frequency and phase modulation are, in fact, the same. The only difference is that in PM the phase of the modulated waveform is proportional to the amplitude of the modulating waveform, while in FM it is proportional to the integral. Both schemes are sometimes referred to as angle modulation. Phase modulation, when the message signal is a continuous (analog or tone) function, does not appear to have any practical applications. When the message signal is digital, it has distinct advantages such as improved immunity to noise.

11.4.1 Binary Phase Shift Keying (BPSK)

When the modulating (message) signal is in binary form we refer to it as *keying*. This is a left-over from the days when telegraph operators opened and closed a circuit (presumably, using a “Morse key”) to generate Morse code.

Figure 11.7 shows a comparison of the waveforms of the three modulating schemes.

It should be noted that, for clarity, the RF has been chosen to be only four times the data rate. In practice, the RF is much higher than the data rate.

If we represent the digit **1** by the binary pulse $p(t) = 1$, the digit **0** by $p(t) = -1$ and the carrier by $\cos \omega_c t$, then after modulation we have

$$s(t) = p(t) \cos \omega_c t \quad (11.1)$$

for the digit **1** and

$$s(t) = -p(t) \cos \omega_c t = p(t) \cos(\omega_c t + \pi) \quad (11.2)$$

for the digit **0**.

Demodulation of a BPSK signal requires a balanced mixer and an exact replica of the carrier.

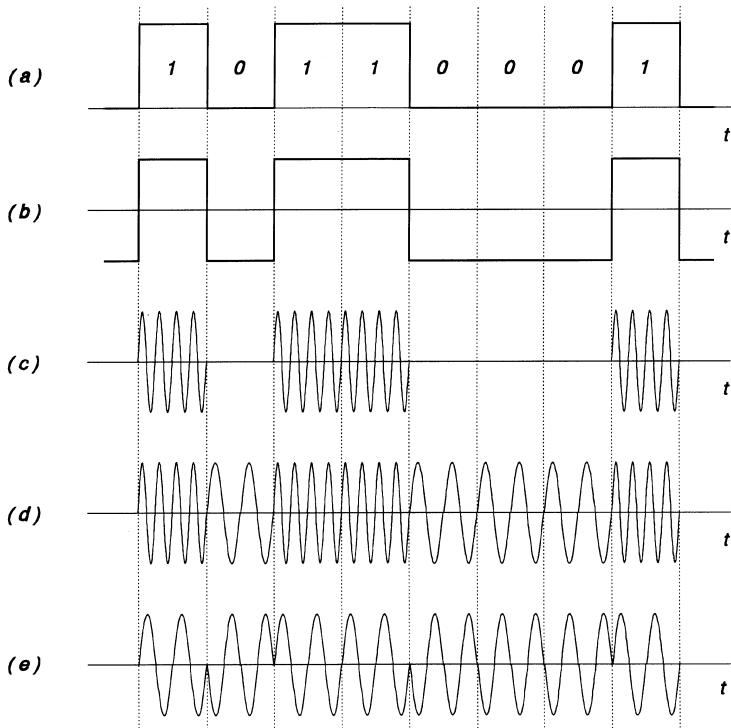


Figure 11.7. (a) Binary data, (b) its bipolar equivalent, (c) the amplitude-modulated waveform, (d) the frequency-modulated waveform, and (e) the phase-modulated waveform.

11.4.2 Quadrature Phase Shift Keying (QPSK)

In quadrature phase shift keying, the message signal is separated into *in-phase* (*I*) and *quadrature-phase* (*Q*) components and are then modulated separately by two carriers of the same frequency but with a phase difference of 90° . QPSK is used because twice the information can be carried in the same bandwidth as when BPSK is applied [2].

Figure 11.8 shows the structure of the QPSK modulator. It has been assumed that the pulses used for modulating the carrier are rectangular. In fact, rectangular pulses are quite undesirable since, in a limited bandwidth channel, they tend to smear into the time intervals of other pulses [3]. The pulse shaping filter is used at the baseband or at the IF stage to limit adjacent channel interference.

Figure 11.9 shows the waveforms of the original data, the *I* and *Q* components, the $I \cos \omega t$ and $Q \sin \omega t$ as well as the QPSK signal. Note that the QPSK signal is a combination of the waveforms of the $I \cos \omega t$ and $Q \sin \omega t$ components.

The demodulation of the QPSK signal is done coherently as shown in Figure 11.10. After down-conversion the received signal is split into two parts and each part is demodulated using a carrier signal derived from the received signal by a carrier

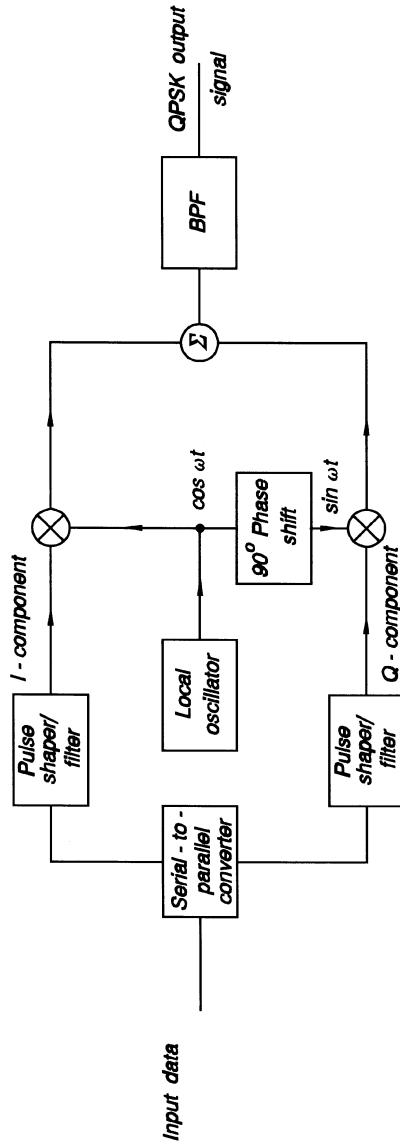


Figure 11.8. Block diagram of the QPSK modulator.

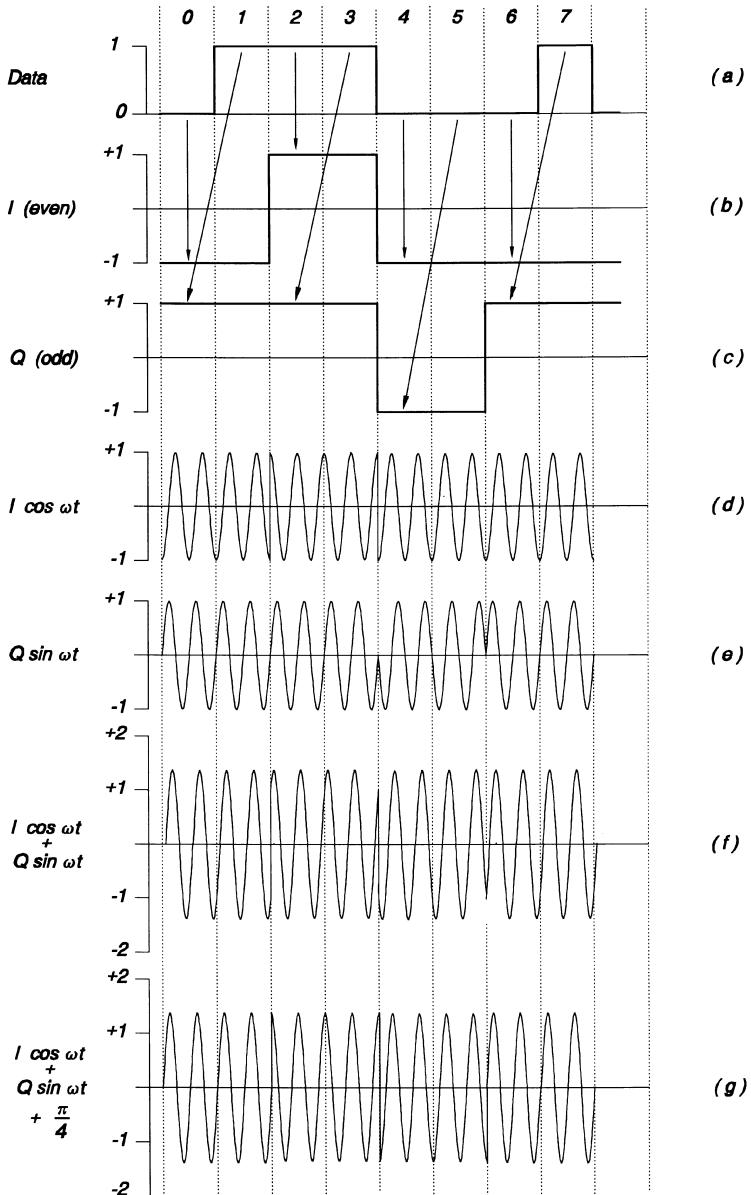


Figure 11.9. The waveforms of (a) the data, (b) the non-return-to-zero in-phase (I) component, (c) the non-return-to-zero quadrature (Q) component (note that the waveform shown in (a) is not coincident in time with those shown in (b) and (c)), (d) $I \cos \omega t$, (e) $Q \sin \omega t$, (f) the QPSK signal, and (g) the QPSK signal with the phase shifted by $+\pi/4$.

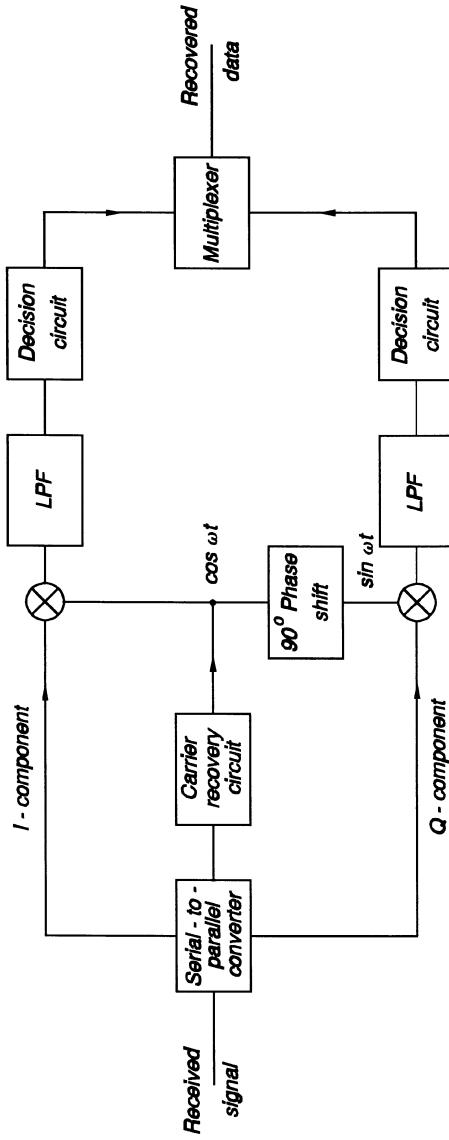


Figure 11.10. Block diagram of the QPSK demodulator.

recovery circuit. The low-pass filters remove the undesirable products of the multiplication process. Two circuits make decisions on whether the bit that was sent was a **1** or a **0**. The *I* and *Q* components are passed to a multiplexer which reconstitutes the original binary signal.

11.5 THE PAGING SYSTEM

There are a number of occupations in which the professional has to move around from one job to the next and essentially is almost never available at a wireline telephone. The paging system is designed to receive and store information until the professional is ready to read it. They are most commonly used by home appliance servicemen, office equipment servicemen, doctors, photographers, and in the last few years they have become very popular with teenagers. Different paging systems have varying capabilities. The message received and stored may be as simple as "Call the paging center to pick up your message" or it may give the number of the caller or an alphanumeric message, or in some of the more sophisticated systems the caller can leave a voice message. The important difference between paging and other systems of communication is that in paging, there is no need for an immediate response.

11.5.1 The POCSAG Paging System

In this section, we discuss the design and operation of one of the simpler paging systems currently in use. This is the Post Office Code Standardization Advisory Group (POCSAG) system. This system was introduced in the early 1980s as a standard for the manufacture of pagers for the British Post Office. It can handle up to 2 million addresses per carrier and supports tone only (alert only), numeric, and alphanumeric pagers. There are three speeds at which the POCSAG system transmits its messages; they are 512, 1200, and 2400 bps. These data rates would normally be considered to be slow but this is deliberate because, combined with high transmitter power (hundreds of watts to a few kilowatts), it improves reliability. The message is "broadcast" over the entire area of operation and it is supposed to reach the recipients whether they are in a building, on a highway, or in an airplane. Typical carrier frequency of operation of the transmitters is around 150 MHz. Each message is preceded by a CAP code which is a unique 7 or 8 digit code recognizable to only one paging receiver in the geographic area of operation.

11.5.1.1 The Paging Transmitter. Figure 11.11 shows a block diagram of the transmit portion of the paging system. Most of the messages come in over the telephone system. The source of the message can be from a Touch-tone® telephone whose keypad can be used to enter the information to be transmitted. It can be a voice message, in which case a human dispatcher in the paging center has to intervene and key in the appropriate message. The message can also come from a computer with the appropriate software and a modem. Whatever its source or form,

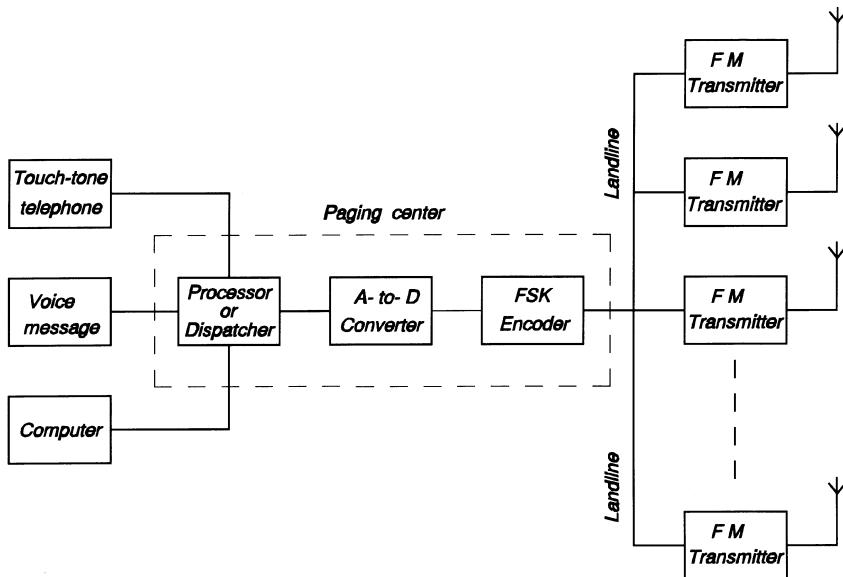


Figure 11.11. Block diagram of the transmit portion of the paging system.

the message goes into an A/D converter. The digital output is used to drive a frequency shift keying encoder in which the digit **0** is assigned a frequency of, say, 1200 Hz and the digit **1** is represented by a tone of frequency 2400 Hz. The message is placed in a queue with other messages. The appropriate CAP code is inserted ahead of each message frame. The dual tone signals may be sent over landlines or wireless systems to a large number of frequency modulated transmitters distributed over a geographic area.

11.5.1.2 Component Circuit Design. The function of the “processor” is to condition the analog signals coming over the telephone line into the paging center for the A/D converter. The human dispatcher plays the same role. The design of the A/D converter is described in Section 8.5.1.2. The frequency shift keying encoder is a form of modem. Modem circuits are described in Section 9.4.1. The frequency modulated (FM) radio transmitter was the subject of Chapter 4.

11.5.1.3 The Paging Receiver. The block diagram of the basic paging receiver is shown in Figure 11.12 [4]. The paging receiver is typically a small device which can be worn on a waist belt. It is basically an FM receiver with a fixed carrier frequency. It has an internal antenna typical of portable radios. Each receiver has a unique CAP code programmed into it and when a message arrives with the appropriate CAP code, the message is saved in the memory and the controller triggers the alert generator which sends a signal to the alert transducer. All other messages are ignored. The alert may be in the form of a sub-audio vibration, a

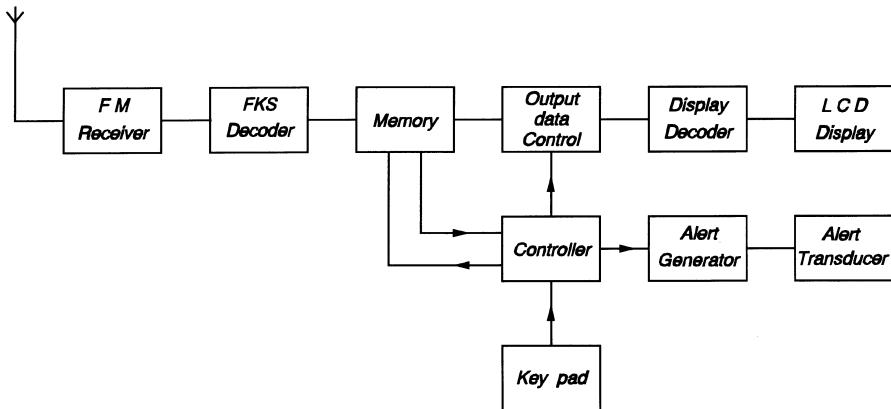


Figure 11.12. Block diagram of the paging receiver.

chime, a beep, or a short excerpt of a well known tune. For a “tone only” paging receiver, the wearer is simply alerted and has to place a call to a messaging center to get the message. For numeric and alphanumeric receivers, the stored message can be retrieved by pushing the appropriate buttons on the front of the device. On receiving the appropriate commands from the keypad, the controller causes the output data control to send the stored information to the decoder which converts it into a form suitable for display on the liquid crystal display (LCD). The messages remain in the memory until they are cleared.

11.5.1.4 Component Circuit Design. The frequency modulated radio receiver was the subject of Chapter 5. A frequency shift keying decoder was discussed in Section 9.4.1 (modems). Memory circuits, their control, coding, and decoding are discussed in Appendix E.

11.5.1.5 Liquid Crystal Display. Certain chemical compounds, such as the cyanobiphenols, have the property that causes the rotation of polarized light passing through them. These compounds are normally transparent to visible light but when seen in a container they appear to be translucent. This is because the axes of the molecules are normally randomly oriented and hence they scatter light in random directions. When an electric field is applied to the compound, the axes of the molecules line up and, depending on the orientation of the incident polarized light, they allow the light to go through or stop it [5]. It is possible in some of the most commonly used liquid crystals to make the light “twist” or gradually change its orientation as it travels through the liquid. This phenomenon is known as twisted nematic and the degree of twist can be set during manufacture. This is the basis of the common liquid crystal display which is used in watches, pagers, and many other electronic consumer goods.

Figure 11.13 shows two parallel transparent electrodes with backings of polarizing film, one vertically oriented and the other horizontally oriented. The space between the electrodes is filled with the liquid crystal compound.

When the light enters the vertically polarized film, only the vertical component will pass through and enter the liquid crystal. As it travels from the right-hand electrode to the left-hand electrode its orientation is changed from the vertical to the horizontal. If the left-hand side polarizing film is oriented horizontally, the light will pass through it.

When an electric field is applied across the electrodes the axes of the molecules are lined up such that they do not affect the orientation of the light. The vertically polarized light cannot go through the horizontally polarized left-hand side film. The contrast created by the presence or absence of the electric field is exploited in the application of the liquid crystal as a display transducer.

In the display of alphanumeric characters, the most commonly used units are the seven-segment and the dot-matrix displays [6]. The dot-matrix display is the more versatile of the two but its decoding system is more complex. The decoding of information in a binary format for display by the relatively simple seven-segment display is not as simple as might be expected and an explanation of how it is designed will be an unnecessary diversion at this point. The seven-segment LCD and its driver, the MC5400/7400 series integrated circuit, are presented in Appendix F.

11.5.2 Other Paging Systems

Since the early 1980s when the POCSAG system was introduced, a number of new systems have been developed. They have increased the speed of transmission, lowered the current drain on the battery, increased the number of addresses per carrier, improved reliability, and made the system more difficult to tamper with. Two of these systems are described very briefly below.

- (1) **ERMES (European Radio Message System).** This system was introduced in the early 1990s by the European Community. The data rate is fixed at 6250 bps and it is capable of operating over multiple radio-frequency channels. The pager can scan all the channels when the subscriber is away from his home base.
- (2) **FLEXTM (Flexible wide-area paging protocol).** This system was introduced in 1993 as a high performance multi-speed paging protocol (1600, 3200 and 6400 bps). FLEX can support over 5×10^9 addresses and conserves the pager battery life by sending data in specified time slots only [7].

At the end of the 20th century it was estimated that there were 192 million pages in use world-wide [8].

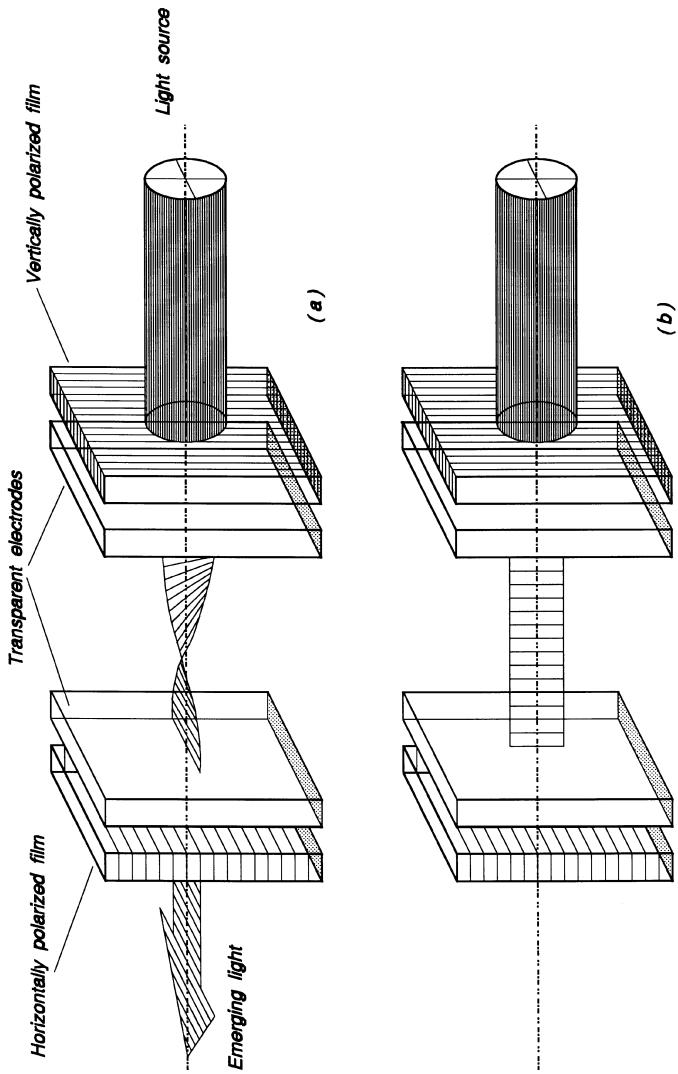


Figure 11.13. An illustration of the operation of the liquid crystal as a display transducer. Reprinted with permission from W. C. O'Mara, *Liquid Crystal Flat Panel Display*, Van Nostrand Reinhold, New York, 1993.

11.6 THE ANALOG CORDLESS TELEPHONE

The cordless telephone was designed to liberate the telephone user from the tether that the handset cord is. Before cordless telephones appeared on the market, long handset cords were used to increase the distance between the handset and the base but the longer the cord got the more clumsy it became. The cordless telephone not only increased the distance from the base, it made the handset completely portable.

11.6.1 System Design

Figure 11.14 shows the configuration of the cordless telephone. The base station is connected directly to the *Public Switched Telephone Network* (PSTN) and, from the point of view of the PSTN, it is just another telephone set. In fact, part of it is a transceiver which provides a two-way link to the handset. It can transmit signals to the handset and receive signals from the handset for onward transmission to the PSTN. The telephone part of the system is the same as any wireline telephone set (see Chapter 8) and the radio part of it uses frequency modulation (see Chapters 4 and 5) in the 900 MHz band. The base station transmitter operates at one frequency (say, 925.997 MHz) while the handset transmitter operates at another frequency (say, 902.052 MHz). This is an example of two simplex systems which form a *frequency-division duplex* (FDD). A device called a *duplexer* provides a coupling between the antenna and both the transmitter and the receiver. Other frequencies were used in the past and a new generation of cordless telephones, using digital technology, have been assigned spectra in the 2.4 GHz band in North America.

The handset antenna is coupled to both the FM transmitter and receiver by the duplexer. Separate amplifiers condition the signal from the microphone and the signal going to the speaker appropriately.

11.6.2 Component Design

The designs of all the components in Figure 11.14 were discussed earlier with the exception of the duplexer.

11.6.2.1 The Radio-Frequency Duplexer. The role of the RF duplexer is to couple the strong signal from the transmitter to the antenna with minimal loss but prevent the transmitter signal from reaching the input of the receiver. In many applications, such as radar and the wired telephone system, a hybrid performs this function (see Sections 8.3.6.1 and 8.4.3). The duplexers in both the base station and the handset have to ensure that the path from their transmitter to their receiver has the highest possible attenuation so that no significant local feedback is possible. At the same time they must ensure that the path from the transmitter to the antenna and that from the antenna to the receiver have minimum attenuation. There are two important factors to consider. The fact that the transmit and receive frequencies are separated by a fairly wide margin of approximately 25 MHz is a great help in the design. The fact that the signal power from the transmitter of the handset to its

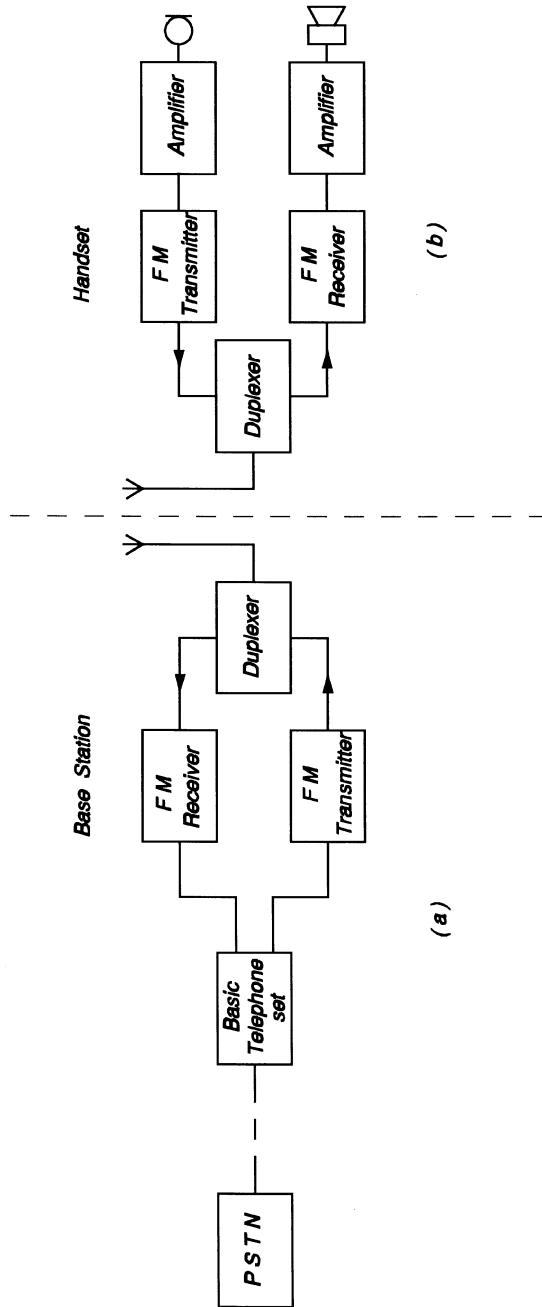


Figure 11.14. Block diagram of the cordless telephone: (a) the base station, (b) the handset.

antenna is several orders of magnitude greater than the signal power reaching the handset antenna from the base station makes the design of the duplex filters critical. The same is true for the base station duplexer.

The connection of the antenna to the duplex filters is shown in Figure 11.15(a). The attenuation characteristics of typical transmit and receive filters are shown in Figure 11.15(b). Note that at the center of the transmit filter passband (BPF(a); forward) there is a loss of about 2.5 dB but at this frequency the receive filter (BPF(b); reverse) has an attenuation in excess of 50 dB [9]. This means that the path of the transmitter signal to the antenna (double-headed arrow) has minimal loss while the path of the transmitter signal to the input of the receiver has very high attenuation. This is the critical path because, if the transmitter signal was allowed to leak into the input of the receiver, there would be a possibility of instability in the system. It is also critical that the weak signal from the antenna reaches the input of the receiver with minimal attenuation (single-headed arrow). Leakage of this signal into the output of the transmitter is not critical.

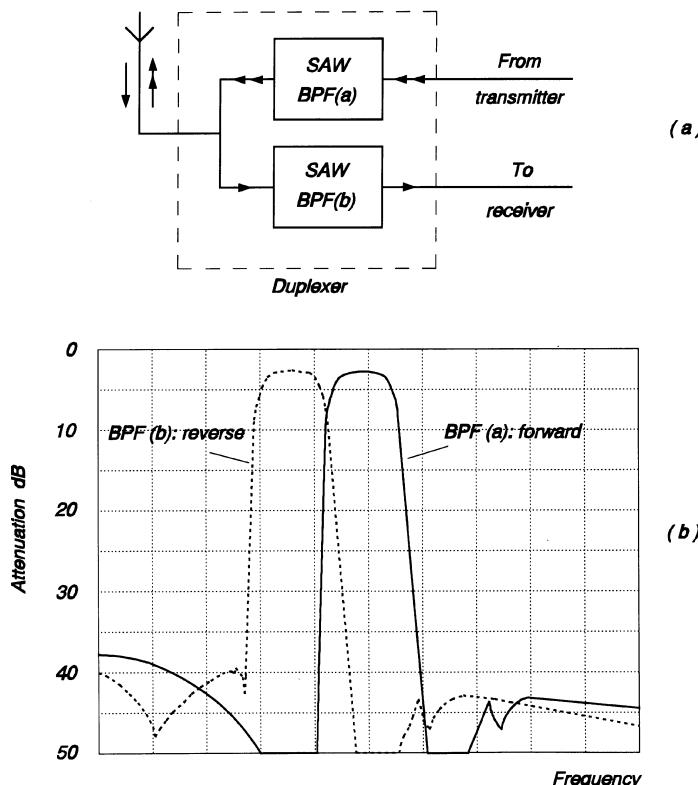


Figure 11.15. (a) A block diagram of the transmit and receive surface acoustic wave (SAW) filters, (b) the frequency characteristics of the two filters.

There are three types of RF duplexer technologies used in cordless telephones. They are ceramic, Surface Acoustic Wave (SAW), and Film Bulk Acoustic Resonator (FBAR). Ceramic RF duplexers are considered to have superior operating characteristics compared to SAW devices but they are, in general, bulkier. FBAR is a new technology that is supposed to match the performance while occupying only about 10% of the space of ceramic devices.

11.6.2.2 Surface Acoustic Wave Radio Frequency Duplex Filter. The design of filters is outside the scope of this book; however, the basics of a surface acoustic wave (SAW) device are easy to appreciate. A typical SAW device is shown in Figure 11.16. It consists of a piezoelectric substrate such as quartz, lithium niobate (LiNbO_3), lithium tantalate (LiTaO_3) and lithium borate ($\text{Li}_2\text{B}_4\text{O}_7$). These crystalline materials are cut at specific angles relative to the crystal axes and highly polished. Thin-film inter-digital electrodes are deposited on the surface as shown in Figure 11.16. Piezoelectric materials have the property of undergoing a mechanical deformation when an electric field is applied to them. Equally, they produce an electric field when they are deformed mechanically. When an alternating voltage is applied to the input electrode, it causes a mechanical deformation on the surface of the crystal which propagates in both directions, left and right. Part of the wave travelling to the left will be reflected at the end of the substrate and there will be losses associated with it. The wave travelling to the right will reach the output electrode where the mechanical wave will induce a voltage in the receive electrode. The amount of coupling between the two electrodes depends on a large number of factors such as the frequency of the signal, the crystal material, the angle of the cut of the crystal relative to its axes, and the geometry of the inter-digital transducers.

The most widely used SAW devices in cordless and cellular telephones are not considered to operate in a truly surface acoustic wave mode. They are generally described as *leaky surface acoustic wave* (LSAW) devices. This nomenclature is

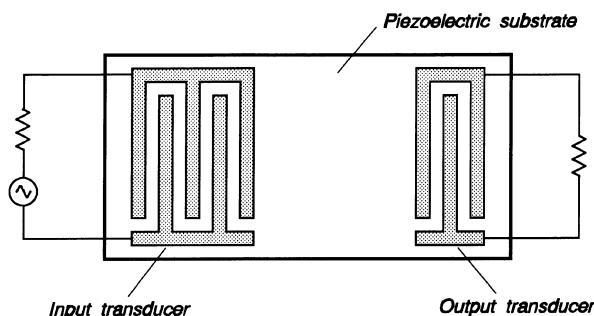


Figure 11.16. A typical surface acoustic wave (SAW) filter showing the input and output transducers. Reprinted with permission from C. K. Campbell, *Surface Acoustic Wave Devices for Mobile and Wireless Communication*, Academic Press, San Diego, CA, 1998.

appropriate because the mechanical deformation that takes place in LSAW devices does not limit itself to the surface of the crystal; it penetrates the material.

11.6.3 Disadvantages of the Analog Cordless Telephone

The range of the cordless telephone is limited to less than 200 m because its radiated RF power is limited to less than 0.75 mW. The distance limitation is not necessarily a disadvantage since the frequencies on which it operates can be reused by other cordless telephone users outside the 200 meter radius with minimal or no interference. The real disadvantage of the system is that, because it uses a radio link with no attempts at encryption, anyone with an FM radio capable of operating in the frequency bands assigned to cordless telephones can tune in and listen to the conversation. As mentioned earlier a new frequency band at 2.4 GHz has been assigned to cordless telephones in North America. These cordless telephones are digital, have a longer range (400 m), and use Spread Spectrum technology to ensure that private conversations remain private.

11.7 THE CELLULAR TELEPHONE

The cellular telephone system can be considered to be an “enlarged” form of the cordless telephone. A number of steps have to be taken to make such an expansion feasible.

- (1) To increase the range of the *mobile* or *portable* the number of base stations will have to be increased from one to many. This also establishes a need for the mobile/portable to be switched from one base station to the next seamlessly when it is on the move. This process is called *handoff*.
- (2) To make the system economically feasible, it will have to handle a large number of telephone conversations simultaneously. This means that the number of channels in the base station must be increased from one to many and interference between the channels must be avoided.
- (3) Due to the variability of the path between the base station and the mobile, fragments of the signal will find their way to the receiving antenna over different paths, resulting in different amplitudes and phases. The vector sum of the fragments may result in an augmentation or, with equal probability, a total cancellation of the signal. This process is called *fading*. It is necessary to take steps to keep the signal strength within acceptable limits.

A cellular telephone system which has the above attributes is the Advanced Mobile Phone System (AMPS).

11.7.1 System Overview

Figure 11.17 shows the configuration of the cellular telephone system. The hexagonal cell from which the system derives its name happens to be a convenient way to divide a geographic area without overlap. The radiation pattern of the antenna at the center of the cell is, in theory, a circle but using circles on a map will produce overlaps and/or voids. The effects of geographical features such as hills, buildings and other structures make the radiation pattern of the antenna difficult to predict. At the center of each cell there is a base station which is connected to the *Mobile Telephone Switching Office* (MTSO). The MTSO is connected to the PSTN.

The MTSO is the nerve center of the cellular system. It controls every parameter of the conversation from when a subscriber requests service to when she terminates the call.

11.7.2 Advanced Mobile Phone System (AMPS)

It has been estimated that at the end of 1997 there were 50 million cellular telephones in the United States [3]. In Canada the number was 7.5 million. With a growth rate of over 50% a year, the estimate for the year 2000 is close to 200 million cellular telephones in use in North America. Although there has been a shift towards digital cellular technology, the majority of cellular telephones in use in North America are based on AMPS. The AMPS cellular system, like the cordless telephone, uses FM radio as part of the interface between the subscribers and the PSTN. The AMPS uses *frequency-division multiple access* (FDMA). FDMA, when applied to AMPS, means that each mobile telephone is assigned two carrier frequencies for voice; one for communication from the base station to the mobile (*forward* or *downlink*) and the other from the mobile to the base station (*reverse* or *uplink*). In addition there are channels set aside for establishing calls and for monitoring and control of the system. These are called *traffic* or *control* channels and are shared by many mobiles.

AMPS uses two bands of frequencies: the forward channels from 869 to 894 MHz and the reverse channels from 824 to 849 MHz. The following are part of the system specification:

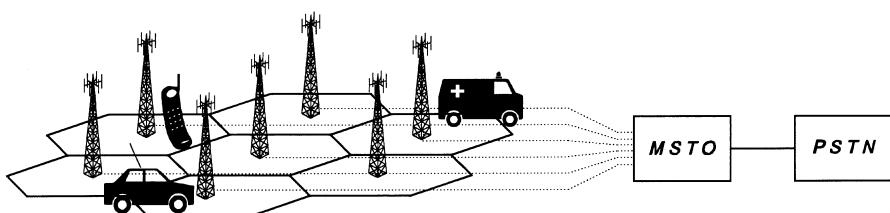


Figure 11.17. The configuration of the cellular telephone system.

- (a) There are two simplex channels combined to form a duplex.
- (b) For each mobile, the forward carrier frequency is 45 MHz higher than the reverse.
- (c) Each channel occupies a bandwidth of 30 kHz.
- (d) The total number of channels is 832.

11.7.2.1 Reuse of the Channels. Figure 11.18 shows the cells clustered in groups of seven [3]. With such an arrangement, it is possible to reuse the carrier frequencies in all cells marked (a), (b), (c) and so on, so long as the radiated power of each mobile in the cell is kept below the appropriate level considering the distance between the respective cells (a), (b), (c), etc. The minimum frequency reuse distance can vary quite widely. In densely populated centers, the cells are made smaller and there are more of them. The radiated power levels are much lower. In more sparsely populated areas, the cells are larger and the radiated power is higher.

11.7.2.2 Setting up the Call. The steps described below have been simplified considerably in the interest of brevity and clarity.

When a cellular telephone is turned off, the system has no way to reach the subscriber. If the telephone is on but not yet engaged in a conversation (idle), the

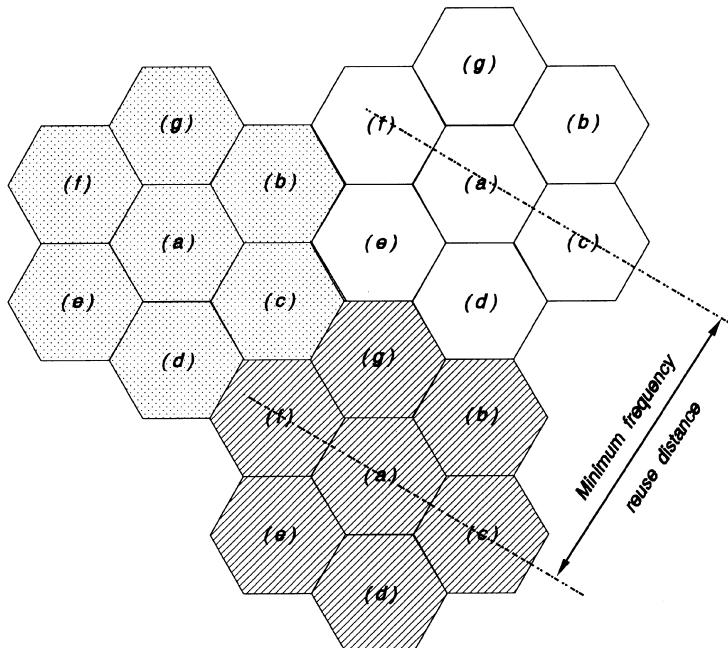


Figure 11.18. The formation of seven-cell clusters which permit frequency reuse under specified radiated power and minimum distance of separation.

mobile scans the forward control channels and determines the strongest of the group. It keeps checking the particular forward control channel until the signal strength falls below an acceptable level, at which point it starts to scan the other forward control channels again to find the strongest signal. The forward and reverse control channels are normally standard throughout the system so the mobile knows which frequencies to scan when it is in the idle mode.

When a call comes into the MTSO, it broadcasts the *mobile identification number* (MIN) of the called mobile (a 10-digit number derived from the telephone number assigned to the device) over all the forward control channels in every base station. The mobile recognizes its MIN and responds by identifying itself via the reverse control channel which it has been monitoring all along. The MTSO recognizes the MIN as a legitimate one and instructs the appropriate base station to move the mobile's center frequencies (forward and reverse channel pair) to an available pair of channels. On the basis of the signal received at the base station, the power level for the transmission from the mobile is set. A "ring" signal is sent to the mobile which results in an alert in the form of a sub-audio vibration, a beep, or a pleasant short tune. The subscriber answers the telephone and the conversation can start. The MTSO continues to monitor the call and will initiate a *handoff* if and when the quality of the channel falls below an acceptable level. A handoff may be necessary because the mobile has moved away from the original base station resulting in a weak signal. When the conversation is over the subscriber presses a button which puts the mobile into the idle state and sends a message to the MTSO to terminate the call.

When the call originates from the mobile (the subscriber has dialed the desired number and pressed the "talk" button) the mobile sends the following pieces of information:

- (a) its MIN,
- (b) a *station class mark* (SCM) (this is an indication of the power level of the transmitter),
- (c) its *electronic serial number* (ESN) (this is a unique 32-bit number programmed into the unit during manufacture),
- (d) the destination telephone number.

through a reverse control channel. The base station receives the information and passes it onto the MTSO for validation. If everything is in order, the MTSO instructs the base station and the mobile to move to an available pair of voice channels, dials the destination telephone number, and connects the system to the PSTN. When the call is terminated, the mobile sends an "end-of-call" code and goes into the idle state. The MTSO dismantles the connection and goes into standby mode waiting for the next call.

11.7.2.3 Handoff Process. Handoff occurs when the signal reaching the base station from the mobile gets too weak for adequate voice quality. Very often this

happens because the user of the mobile or portable is on the move away from the current base station. The MTSO detects this from a set of measurements on the received signal strength of the mobile and follows an algorithm which polls the base stations in adjacent cells for various parameters such as overload conditions. It then tests to determine whether the signal strength of the mobile in the intended new cell is adequate before initiating the handoff. There is a built-in hysteresis so that at the cell boundary the tendency to switch back and forth between the two base stations is eliminated. One has to remember that there is no frequency reuse in adjacent cells. The MTSO then sends a command to the mobile to change its frequency, select a new SAT, and set its initial transmitter power for the new cell. The mobile responds by turning its transmitter off and adjusting its parameters to the new values and switching its transmitter back on. All of these steps are carried out without the user being aware of what is going on.

11.7.2.4 System Design. Figure 11.19 shows the block diagram of a voice (downlink) channel in an AMPS cellular telephone mobile.

The Compander: The voice signal captured by the microphone is amplified and then fed into the *speech compressor*. The compressor reduces the difference of amplitude between the highest and lowest. This is necessary because the low amplitude sounds are likely to be swamped by the noise in the system while the high amplitudes are likely to cause the system to go into the non-linear portions of its characteristics and generate distortion (noise) [10, 11]. The input-output characteristics of the compressor is shown in Figure 11.20(a).

It can be seen that for low amplitude input signals the slope of the curve is quite high, that is, the voltage gain is high while for high amplitude input signals the gain is low. At the receiving end of the communication channel, it is necessary to reverse the effect of the compressor on the signal by passing it through a circuit with the opposite characteristics shown in Figure 11.20(b). The name of that circuit is the *expander*. The application of a compressor followed by an expander is called *companding*. The music recording industry uses this method to reduce noise; it is more commonly known as the Dolby system. (See Section 3.4.5 for circuit design details).

Frequency Pre-emphasis: The energy in the spectrum of human speech tends to be concentrated at the low-frequency end. This makes the high-frequency content of speech vulnerable to the corrupting influence of noise. It is advantageous to raise the level of the higher frequencies relative to the low frequencies before transmission and restore them at the receiving end.

The frequency characteristics of the pre-emphasis filter are shown in Figure 11.21(a) together with a suitable circuit. Note that the slope of the filter is 6 dB/octave. The frequency characteristics of the de-emphasis filter together with its circuit are shown in Figure 11.21(b).

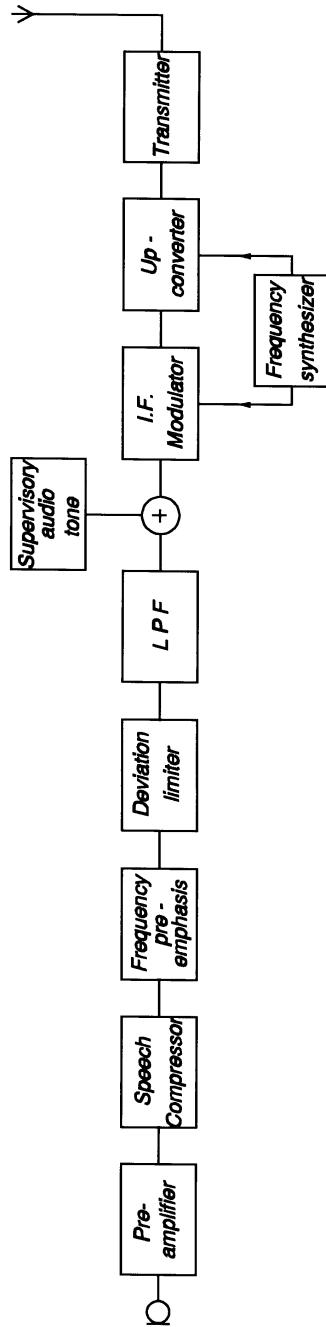


Figure 11.19. A simplified block diagram of the mobile cellular telephone.

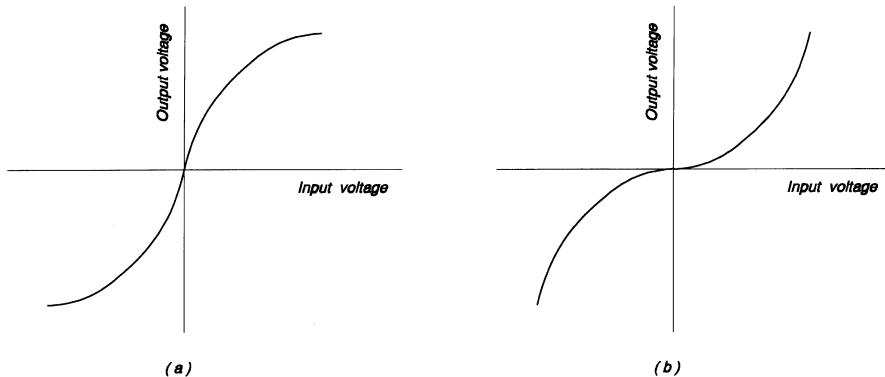


Figure 11.20. (a) The input–output characteristics of the compressor in the transmitter. (b) The input–output characteristics of the expander in the receiver.

Deviation Limiter: The purpose of the deviation limiter is to limit the maximum excursion of the incoming signal so that after frequency modulation the maximum deviation of the FM signal will not exceed ± 12 kHz. If this is not done, there will be adjacent channel interference.

Low-pass Filter: The deviation limiter restricts the signal by clipping it at its extremities, thus producing higher harmonics of the signal. The low-pass filter removes the harmonics.

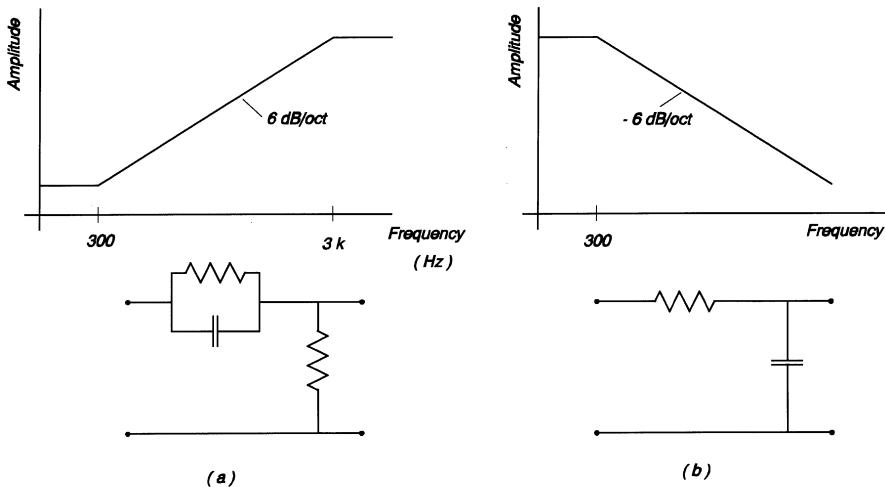


Figure 11.21. (a) The circuit and frequency characteristics of the pre-emphasis filter used in the transmitter. (b) The circuit and frequency characteristics of the de-emphasis filter used in the receiver.

The Supervisory Audio Tone Generator (SAT): The generator can produce three audio signals at 5970, 6000, and 6030 Hz. A base station superimposes one of the three tones on both the forward and reverse voice channels when it is in use. The choice of which one to use is determined by the MSTO. The purpose of this signal is to identify a particular mobile with a given base station during the conversation. The base station sends the SAT to the mobile as soon as the voice channel is established and the mobile has to re-transmit the same signal to the base station for the call to continue. The frequencies are chosen to be outside the 3 kHz bandwidth but well within the hearing range of the subscriber. The SAT is attenuated using a suitable bandpass filter so it is barely audible to the subscriber.

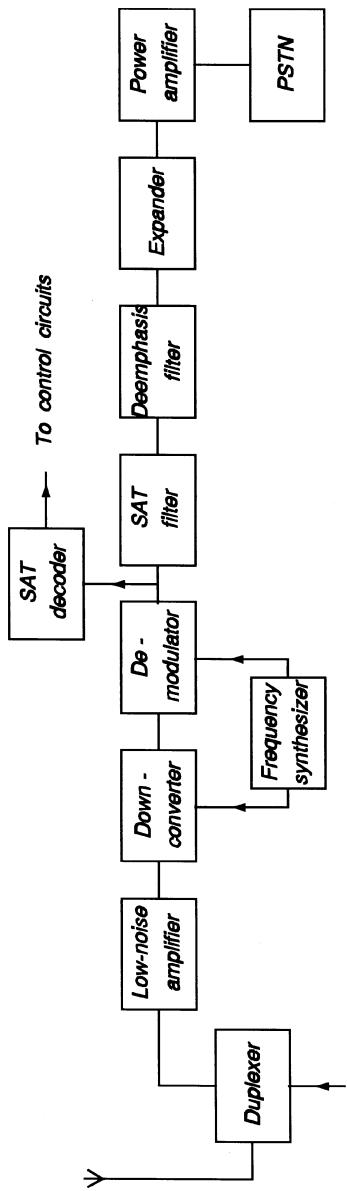
Frequency Modulator: The frequency modulator produces the modulated signal at an intermediate frequency (IF) using the output of the IF oscillator. The IF carrier is generated by a frequency synthesizer whose frequency is controlled over the forward control channel. The frequency modulator may have a frequency multiplier before and after it to bring it to the appropriate carrier frequency (see Figure 4.14).

Upconverter: The upconverter shifts the carrier frequency to its proper value using the output of the RF oscillator. The RF carrier is generated by a synthesizer whose frequency is controlled over the forward control channel. Here again, a frequency multiplier as well as a number of filters may be necessary.

Transmitter: The transmitter consists of an RF power amplifier which is coupled to the antenna. Note that the base station usually uses two different antennas, one for the forward and the other for the reverse channels. No duplexer is required in that case.

Figure 11.22 shows the block diagram of the voice (uplink) channel.

Low-noise Amplifier: The low-noise amplifier brings the antenna signal up the level required to operate the downconverter. In any communication system noise is a constant threat to the integrity of the information transmitted on it. This is especially true of analog systems such as the AMPS. The LNA typically must have a gain greater than 16 dB with a noise figure less than 2.5 dB. It has been demonstrated that the noise generated by the first stage of a chain of amplifiers has the most significant (damaging) effect on the system [12]. The low-noise amplifier is the first of a chain of amplification stages in the path of the signal coming from the radio antenna. Considerable effort has therefore gone into the development of low-noise amplifiers for this application. The LNA is usually a single chip integrated circuit using bipolar junction transistor (BJT), metal-semiconductor field-effect transistor (MESFET), complementary metal-oxide semiconductor (CMOS), and *N*-type metal-oxide semiconductor (NMOS) technologies. A detailed discussion of the design of low-noise amplifiers falls outside the scope of this book. The interested reader is encouraged to refer to the bibliography for further reading material on the subject.



From transmitter

Figure 11.22. A simplified block diagram of the voice (uplink) channel for the code-division multiple access (CDMA) cellular telephone.

The Downconverter: The downconverter is responsible for removing all the vestiges of the upconversion. The downconversion is typically carried out in two steps: the first to an IF and the second to the baseband. It uses the output of the frequency synthesizer.

The Demodulator: The demodulator recovers the baseband signal.

The SAT Filter: The SAT filter attenuates the audio tone to a level where it is almost imperceptible to the subscriber.

The Reverse-emphasis Filter: The 6 dB/octave slope that was added is now equalized by the simple *RC* low-pass circuit shown in Figure 11.21(b).

The Expander: This circuit reverses the effects of the compressor. The characteristics are shown in Figure 11.20(b).

The Power Amplifier: This is an audio amplifier designed to drive the circuits of the PSTN.

11.7.2.5 Component Circuit Design for the Transmit Path. The pre-amplifier is an audio-frequency amplifier designed to boost the signal from the millivolt level to about 1 V. The design of audio-frequency amplifiers was discussed in Section 2.7.

The compressor is an amplifier with a variable gain; high for low-level signals and low for high-level signals. This is the same as an automatic gain control (AGC) amplifier. This was the subject discussed in Section 3.4.5.

The design of the frequency pre-emphasis filter is given in Figure 11.21(a). Filter design is beyond the scope of this book but the interested reader may refer to [13] or any other basic circuit analysis book for a discussion of Bode plots.

The deviation limiter is essentially a voltage amplitude clipper. The amplitude clipper was discussed in Section 4.3.3.1.

The low-pass filter following the limiter is responsible for attenuating the higher harmonics of the signal which appear at the output of the clipper. The design of filters is outside the scope of this book. The interested reader is encouraged to refer to a specialized text on the subject.

The supervisory audio tone (SAT) is a sinusoidal audio frequency oscillator operating at approximately 6 kHz. The design of sinusoidal oscillators was discussed in Section 2.4.3.

The design of the balanced frequency modulator was discussed in Section 4.4.2.3.

The upconverter consists of balanced mixers, local oscillators, frequency multipliers, and various band-pass filters. The design of these components (with the exception of filters) have been discussed in Section 4.4. For a more general discussion of up- and downconverters, refer to Section 12.7.1.1.

The discussion of the design of the FM transmitter (power amplifier and filters) was in Chapter 4.

11.7.2.6 Component Circuit Design for the Receive Path. A description of the operation of the duplexer was given in Section 11.6.2.2. The low-noise amplifier is a highly specialized integrated circuit device whose design falls outside the scope of this book. All the other components are represented or have counterparts in the transmit path of the system described in Section 11.7.3.5.

11.8 OTHER ANALOG CELLULAR TELEPHONE SYSTEMS

The **ETACS** (European Total Access Communication System) is almost identical to AMPS. They both use FM transmissions for the air interface but there are differences in some of the specifications, such as the bandwidth of the channels and the system for the identification of subscribers. These differences are enough to stop AMPS-based units from operating in Europe and vice versa. Some of the specifications of the two systems are presented in Table 11.1

11.8.1 Disadvantages of Analog Cellular Telephone Systems

Early projections suggested that a digital cellular telephone system could yield a 20 : 1 improvement in spectrum efficiency relative to AMPS [10]. Later calculations produced less optimistic projections of 5–10 times improvement.

Messages sent on analog cellular systems are subject to interception by anyone with a frequency scanner and an elementary knowledge of electronics. The system can be further compromised by the unauthorized collection of the access codes by persons and/or organizations who use these to create “clones” with “free” access to the telephone system.

Conversion to digital format can improve privacy if only because the equipment necessary to intercept digitally coded messages is much more complex and expensive. Secondly, it is relatively easy to use encryption codes with digitized messages. Digital techniques also have the potential to identify fraudulent use of the cellular telephone system more quickly and to deal with it by denying service promptly.

TABLE 11.1

Parameter	AMPS	ETACS
Multiple access	FDMA	FDMA
Duplexing	FDD	FDD
Channel bandwidth	30 kHz	25 kHz
Traffic channel per RF channel	1	1
Reverse channel bandwidth	824–849 MHz	890–915 MHz
Forward channel bandwidth	869–894 MHz	935–960 MHz
Voice modulation	FM	FM
Peak deviation: voice channels control/wideband data	$\pm 12 \text{ kHz}/\pm 8 \text{ kHz}$	$\pm 10 \text{ kHz}/\pm 6.4 \text{ kHz}$
Data rate on control/wideband channel	10 kbps	8 kbps
Number of channels	832	1000

Source: Ref. [3]

11.9 THE CDMA CELLULAR TELEPHONE SYSTEM

The code-division multiple access cellular telephone system was designed to improve the capacity of the system and enhance the security of the messages as well as prevent fraudulent access to cellular telephone services [14]. The projected improvement in spectrum efficiency was a very powerful economic incentive for the service providers. Since no new bandwidth was assigned to the CDMA cellular system, it had to coexist with AMPS. CDMA is a wideband system requiring 1.25 MHz in each of the two cellular frequency bands (forward and reverse). Thus 10% of the available bandwidth used by AMPS or 41 adjacent channels had to be vacated to make room for the CDMA system. By the end of the 20th century it was estimated that there were about 200 million cellular telephone sets in use in North America. The majority of these operate under AMPS and it is unlikely that the subscribers will be willing to switch over to the CDMA system in a short time span. Switch over to a completely digital system implies a huge investment in new infrastructure by the cellular telephone service providers. In the meantime it became clear that new subscribers were mainly consumers who used the system during non-business hours and were very conscious of costs. It was therefore necessary for cellular telephone manufacturers to produce portables/mobiles and base stations capable of dual-mode operation.

11.9.1 System Design of the Transmit Path

In what follows, only the part of the system concerned with the path of the message signals is included. Even then much has been excluded in the interest of brevity and simplicity. The readers who need more detailed information may refer to the bibliography and the references at the end of the chapter.

A simplified block diagram of the path taken by the message through the CDMA cellular telephone system is shown in Figure 11.23. The voice is sampled at 8000 Hz and converted into digital form by the analog-to-digital converter (ADC) and, at 8 bits/sample, produces a 64 kbps data stream. It is then used to digitally modulate (multiply) the output of the pseudo-noise (PN) generator in a

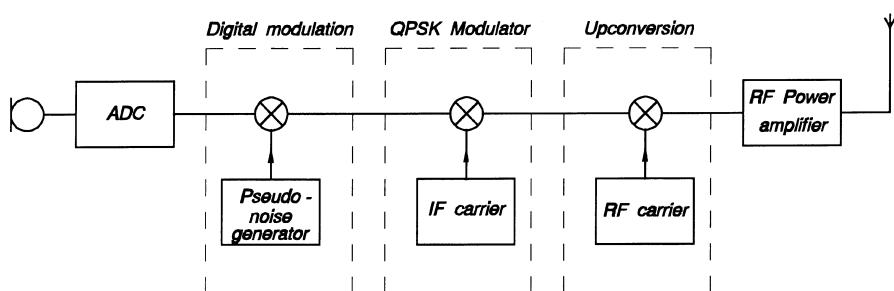


Figure 11.23. A simplified block diagram of the transmitter of a code-division multiple access (CDMA) cellular telephone system.

balanced mixer. The output of the mixer is then used to modulate the output of the IF carrier (sinusoidal). The resulting signal is upconverted by a second sinusoidal carrier, amplified and radiated by the antenna. Note that it is assumed that the appropriate filters are placed in the correct positions to make the process work correctly.

11.9.2 Component Circuit Design for the Transmit Path

Analog-to-Digital Converter: The design and operation of the ADC was described in Section 8.5.1.2.

Pseudo-noise Sequence Generator: A more accurate name for this circuit is Linear Recursive Sequence (LRS) generator. Consider a set of flip-flops connected as a shift register shown in Figure 11.24(a). Assume that there are four flip-flops and that each stage of the shift register is set arbitrarily as shown. If we connect the output to the input and apply the clock pulse, the effect will be to shift the data **1100** one stage for each clock pulse. We have a ring counter which will continue to circulate the data (**....001100110011....**) so long as the clock pulses are supplied. In this configuration the sequence repeats itself after four clock pulse, the same as the number of stages in the ring counter. Suppose we modify the circuit by adding the

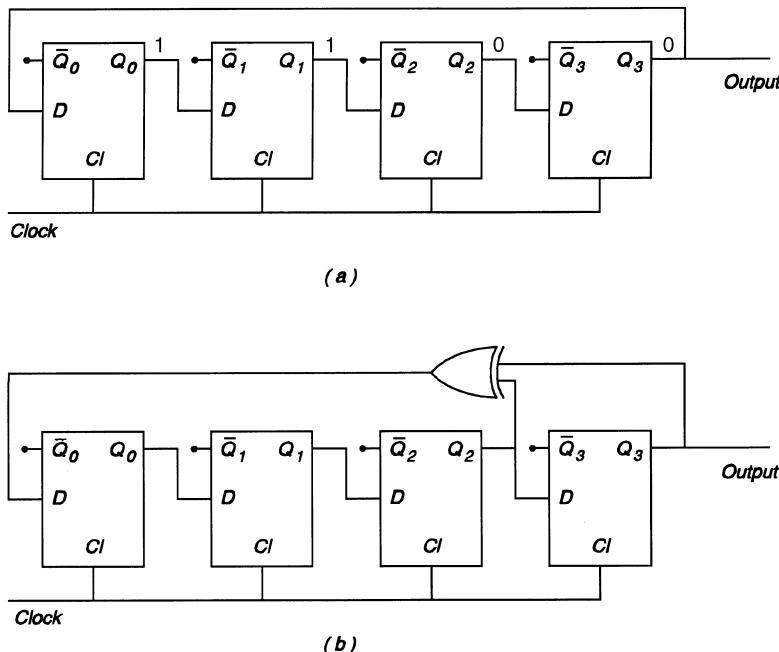


Figure 11.24. The four flip-flop linear recursive sequence (LRS) generator whose truth table is shown in Table 11.3.

TABLE 11.2

Inputs		Output
A	B	C
0	0	0
0	1	1
1	0	1
1	1	0

two-input *exclusive OR* (XOR) as shown in Figure 11.24(b). Now the digit fed back to FF_0 is determined by the outputs of FF_2 and FF_3 through the XOR. This circuit configuration is arbitrary, one of several that could have been used [15]. However, the object of this exercise is to generate the longest sequence for the minimum number of flip-flops in the ring counter. Using the truth table for the two-input XOR given in Table 11.2, we can construct the truth table of the four flip-flop LRS shown in Table 11.3.

Table 11.3 shows that the sequence repeats itself after the 15th clock pulse. In this case the initial states of the four flip-flops were carefully chosen to give the sequence its maximum length. Other initial states may end prematurely with 0s at the outputs of all the flip-flops. The LRS cannot recover from the all-zero state. It can be shown that, for a four-stage LRS, the maximum length of the sequence before it repeats itself is given by ($2^n - 1 = 15$) where $n = 4$. The CDMA system uses a 42-stage LRS and hence repeats itself after ($2^{42} - 1$) cycles! Anyone looking at the possibility

TABLE 11.3

Clock	Q_0	Q_1	Q_2	Q_3
1	1	1	0	0
2	0	1	1	0
3	1	0	1	1
4	0	1	0	1
5	1	0	1	0
6	1	1	0	1
7	1	1	1	0
8	1	1	1	1
9	0	1	1	1
10	0	0	1	1
11	0	0	0	1
12	1	0	0	0
13	0	1	0	0
14	0	0	1	0
15	1	0	0	1
16	1	1	0	0
17	0	1	1	0
18	1	0	1	1

of breaking the code by the observation of repeated patterns has a long wait. Code breaking is made more difficult because knowledge of the chip clock rate, the structure of the LRS (number of flip-flops and feedback circuit) and the initial settings of the shift registers are a few of the parameters required to reproduce the pseudo-noise sequence.

Other Components in the Transmit Path: The IF modulator using QPSK was discussed in Section 11.4.42. The upconverter was discussed in Section 10.7.1 and the RF power amplifier in Section 2.8.

11.9.3 System Design of the Receive Path

A block diagram of the receiver is given in Figure 11.25. The LNA receives its input from the antenna and conditions it for downconversion from the RF to the IF stage. After the second balanced mixer, the signal then goes to the QPSK demodulator which produces a replica of the message data with the chips from the transmitter PN present [16]. The PN must be extracted from the digitally modulated signal in order to recover the original (digitized voice) signal. The local LRS generates chips using identical settings used in the transmitter. It is then phase-aligned with the received signal by sliding (*sliding correlator*) the local code past the received code. The microprocessor keeps track of the number of bits that match and stops the sliding action when a peak occurs (*delay lock*). The locally generated LRS code is then mixed (multiplied) with the digitally modulated signal (demodulation). This step separates the LRS from the digitized voice signal. A low pass filter removes that unwanted products of the demodulation. The DAC converts the digitized voice to an analog signal which is amplified and fed to the speaker.

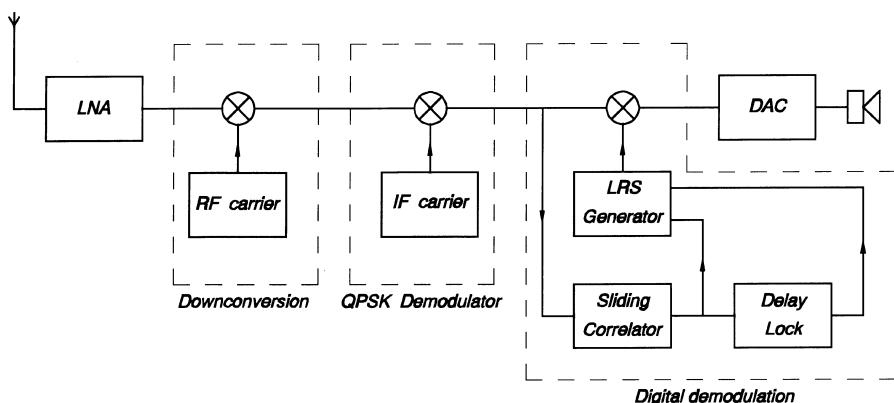


Figure 11.25. Shows a simplified block diagram of the voice (downlink) channel for the code-division multiple access (CDMA) cellular telephone.

11.9.4 Component Circuit Design for the Receive Path

All the components required for the downconversion and the QPSK demodulation are the same as those used in the upconversion and the QPSK modulation. The difference is in the type of filtering used following the two process. The correlation and the locking action are implemented in software and do not have an equivalent circuit.

11.10 OTHER DIGITAL CELLULAR SYSTEMS

USDC (United States Digital Cellular): This system was designed to increase the capacity of the cellular telephone service. In one AMPS channel, USDC supports three subscribers. USDC is a time-division multiple access (TDMA) system and it uses the same FDD for the downlink and uplink separated by 45 MHz. Like the system based on CDMA, conversion to USDC has to be carried out within the bandwidth reserved for cellular telephone services. It uses $\pi/4$ differential quadrature phase shift keying in a 30 kHz per channel FDD. Service providers are obliged to continue service in the AMPS format when they introduce the USDC. The USDC uses the same frequency reuse plan and operates from the same base stations as AMPS. Some of the latest mobiles are designed to operate with both USDC and AMPS. The plan is to fade out AMPS slowly and increase the presence of its digital counterparts. The problem facing USDC is that the CDMA system seems to have gained a wider acceptance and it may well become the dominant system.

GSM (Global System for Mobile Communications): The standards adopted for the development of the second-generation cellular telephone system was put together by the *Group Special Mobile* (GSM) a subsidiary of the Conference of European Postal and Telecommunications Administrations (CEPT). The system took on the French name of the group that proposed it. It primarily uses the frequency bands 890 to 915 MHz and 935 to 960 MHz but has also been assigned spectrum between 1.7 and 1.9 GHz. It is a digital system designed to be used throughout Europe and to replace the many incompatible systems in existence. GSM uses both FDMA and TDMA but it does not accommodate dual operation with any of the many analog systems currently in use in Europe. GSM is incompatible with all of the North America cellular systems and it is not likely that North Americans visiting Europe or Europeans in North America will be able to use their own cellular telephones any time soon. In the meantime, cellular telephone manufacturers have started to market a “triple mode” telephone which can work with the AMPS (analog), the USDC or CDMA (both digital) as well as the GSM (digital) systems. These telephones are essentially three different telephones housed in the same handset (mobile) with an automatic switch-over feature. Some of these cellular telephones can be used as pagers, they can receive voice mail and have wireless access to the Internet.

REFERENCES

1. Groe, J. B. and Larson, L. E., *CDMA Mobile Radio Design*, Artech House, Boston, MA, 2000.
2. Garg, V. K. and Wilkes, J. E., *Principles and Applications of GSM*, Prentice-Hall, Upper Saddle River, NJ, 1999.
3. Rappaport, T. S. *Wireless Communications: Principles and Practice*, Prentice-Hall, Upper Saddle River, NJ, 1996.
4. www.us.semiconductor.com/pip/pcf500lt
5. O'Mara, W. C., *Liquid Crystal Flat Panel Display*, Van Nostrand Reinhold, New York, 1993.
6. Diefenderfer, A. J., Holton, B. E., *Principles of Electronic Instrumentation*, Saunders College Publishing, Philadelphia, 1994.
7. www.motorola.com./FLEX
8. www.motorola.com/
9. Campbell, C. K., *Surface Acoustic Wave Devices for Mobile and Wireless Communication*, Academic Press, San Diego, CA, 1998.
10. Goodman, D. J., *Wireless Personal Communications Systems*, Addison-Wesley Longman, Reading, MA, 1997.
11. Spanias, A. S. "Speech Coding; A Tutorial Review", *IEEE Proc.* Oct. 1994.
12. Smith, J., *Modern Communication Circuits*, McGraw-Hill, New York, 1986.
13. Irwin, J. D., *Basic Engineering Circuit Analysis*, 5th Ed., Prentice-Hall, Upper Saddle River, NJ, 1996.
14. Telecommunications Industry Association/Electronic Industry Association Interim Standard-95, "Mobile Station—Base Station Compatibility Standard for Dual-Mode Wideband Spread Spectrum Cellular System", July 1993.
15. Dixon, R. C., *Spread Spectrum Systems*, John Wiley, New York, 1976.
16. Stremler, F. G., *Introduction to Communication Systems*, 3rd Ed., Addison-Wesley, Reading, MA 1990.

BIBLIOGRAPHY

- Gilberg, B., "The Design of Bipolar Si/SiGe LNAs from the Ground Up", BCTM97 Short Course Notes.
- Ghapurey, R., and Viswanathan, T., "Design of Front End RF Circuits", *Proc. 1998 Southwest Symp. On Mixed-Signal Design*, pp. 134–139.
- Shaefer, D. K., and Lee, T. H., "A 1.5 V, 1.5 GHz CMOS Low-Noise Amplifier", *IEEE J. Solid-State Circuits*, May 1997, pp. 745–759.
- Steele, R., *Mobile Radio Communications*, IEEE Press, New York, 1992.
- Tuttlebee, W. H. W., *Cordless Telecommunications Worldwide*, Springer-Verlag, Berlin, 1997.
- Brodsky, I., *Wireless: The Revolution in Personal Telecommunications*, Artech House, Boston, MA, 1995.

PROBLEMS

- 11.1** A pager is a one-way communication system. Why is it so popular when cellular telephones can provide two-way communication?
- 11.2** Compare the cordless telephone to the cellular telephone in terms of transmitted power, mobility, and privacy.
- 11.3** What are the differences between a conventional wireline telephone system and a personal wireless telephone system?
- 11.4** Why is it necessary to control the transmitted power of each mobile unit in a cellular telephone system?
- 11.5** Describe the procedures followed in a cellular telephone system when a mobile is handed over from one base station to another. What is the role of hysteresis in the handoff process?
- 11.6** In AMPS each mobile is assigned a bandwidth of 30 kHz but only approximately 3 kHz is required for a normal telephone conversation. Explain why?
- 11.7** Explain the following: frequency-division multiple access (FDMA), time-division multiple access (TDMA), frequency hopped multiple access (FHMA), and code-division multiple access (CDMA).

ABBREVIATIONS

ADC	Analog-to-Digital Convertor
AGC	Automatic Gain Control
AM	Amplitude Modulation
AMPS	Advanced Mobile Phone System
BFSK	Binary Frequency Shift Keying
BJT	Bipolar Junction Transistor
BPF	Band Pass Filter
BPSK	Binary Phase Shift Keying
CDMA	Code-Division Multiple Access
CEPT	Conference of European Postal and Telecommunications Administrations
CMOS	Complementary Metal-Oxide Semiconductor Transistor
CO	Central Office
DAC	Digital-to-Analog Convertor
DQPSK	Differential Quadrature Phase Shift Keying
DSP	Digital Signal Processing
EIA	Electronic Industry Association

ERMES	European Radio Message System
ESN	Electronic Serial Number
ETACS	European Total Access Communication System
FBAR	Film Bulk Acoustic Resonator
FDD	Frequency-Division Duplex
FDMA	Frequency-Division Multiple Access
FH	Frequency Hopping
FH-MA	Frequency Hopped Multiple Access
FH-SS	Frequency Hopped Spread Spectrum
FLEX	Frequency-Shift Keying based Paging System
FM	Frequency Modulation
FSK	Frequency Shift Keying
GSM	Global System for Mobile Communication
IEEE	Institute of Electrical and Electronics Engineers
IF	Intermediate Frequency
ISDN	Integrated Services Digital Network
ITU	International Telecommunications Union
LCD	Liquid Crystal Display
LNA	Low Noise Amplifier
LRS	Linear Recursive Sequence
LSAW	Leaky Surface Acoustic Wave
LSSB	Lower Single Side Band
MESFET	Metal-Semiconductor Field-Effect Transistor
MIN	Mobile Identification Number
MTSO	Mobile Telephone Switching Office
MUX	Multiplexer
NMOS	<i>N</i> -type Metal-Oxide Semiconductor Transistor
NRZ	Non-return to Zero
PCM	Pulse Code Modulation
PCS	Personal Communication System
PN	Pseudo-noise
POCSAG	Post Office Code Standardization Advisory Group
PSK	Phase Shift Keying
PSTN	Public Switched Telephone Network
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
RF	Radio Frequency
SAT	Supervisory Audio Tone
SAW	Surface Acoustic Wave

SCM	Station Class Mark
SDMA	Space Division Multiple Access
SNR	Signal-to-Noise Ratio
SS	Spread Spectrum
SSB	Single Sideband
ST	Signaling Tone
TACS	Total Access Communication System
TDD	Time-Division Duplex
TDMA	Time-Division Multiple Access
TIA	Telecommunications Industry Association
USSB	Upper Single Sideband
VCO	Voltage Controlled Oscillator

12

TELECOMMUNICATION TRANSMISSION MEDIA

12.1 INTRODUCTION

In this chapter the characteristics of the media in which the transmission of signals takes place will be discussed. It so happens that we humans basically communicate through speech/hearing and by sight. Human hearing is from 20 Hz to 20 kHz and we can see only the portion of radiation spectrum from about 4.3×10^{14} Hz (infrared; $\lambda = 7 \times 10^{-7}$ m) to approximately 7.5×10^{14} Hz (ultraviolet; $\lambda = 4 \times 10^{-7}$ m). These communication channels occupy only small portions of the detectable frequency spectrum which has no lower boundary but has an upper boundary of about 10^{22} Hz (gamma rays). Acoustic radiation in the frequency range 20–20 kHz is attenuated quite severely in our environment even when attempts are made to guide it along a conduit. It is therefore quite inefficient to transmit an acoustic signal over any distance which would qualify as *telecommunication*. The same observation can be made about visible light. To communicate over distances greater than what we can bridge by shouting, or see reliably, it is necessary to convert the signal into another form that can be guided (by wire, waveguide, or optical fiber) or which can be radiated efficiently in free space.

Wire, coaxial cables, waveguides, optical fiber, and free space transmission have characteristics which vary as frequency changes. A medium may be efficient in one frequency range but quite unsuitable for another frequency range. But efficiency is not the sole criterion for the choice of the frequency to which audio and video signals have to be translated for transmission. To help keep some order and to minimize interference among the various users of communication services, it is necessary to assign various frequency bands for specific uses and governments arrogate to themselves the right to demand a licensing fee for the use of these bands. For example, satellite communication has been assigned 4–6, 12–14, and 19–29 GHz but there is no technical reason why they cannot operate at frequencies in between these frequencies or indeed outside them.

12.2 TWISTED-PAIR CABLE

This consists of two insulated wires twisted together to form a pair. Several to many hundred pairs may be put together to form a cable. When this is done it is usual to use different pitches of twist in order to limit electromagnetic coupling between them and hence cross-talk. The conductor material is copper, usually numbers 19, 22, 24, and 26 American Wire Gauge (AWG), and the insulation is usually polyethylene. Wax-treated paper insulation was used in the past but the ingress of moisture into the cable was a problem in most applications; it is still a problem even with polyethylene insulated cables which are sometimes filled with grease-like substances to take up all the air spaces and thus discourage moisture from entering. Such cables may be suspended from poles where they are easy and inexpensive to service but are aesthetically undesirable, or buried which make them expensive and difficult to repair.

The frequency characteristics of a BST 26-gauge non-loaded cable terminated in $900\ \Omega$ are shown in Figure 12.1. It can be seen that the twisted pair has a low-pass characteristic. It should be noted that, contrary to expectation, the primary constants of the twisted pair (series resistance, shunt capacitance, series inductance and shunt conductance, all per unit length) change with frequency. The bandwidth of the twisted pair can be extended to a higher frequency by inductive loading of the line. Lumped inductances are connected in series with the line at specified distances. The best results are obtained when the interval is kept short and the value of the lumped inductance is kept low, thus minimizing the discontinuities introduced by loading. The frequency responses of a 12,000 ft (3.7 km) number 26-gauge with $900\ \Omega$ terminations for the loaded and unloaded cases are shown in Figure 12.2.

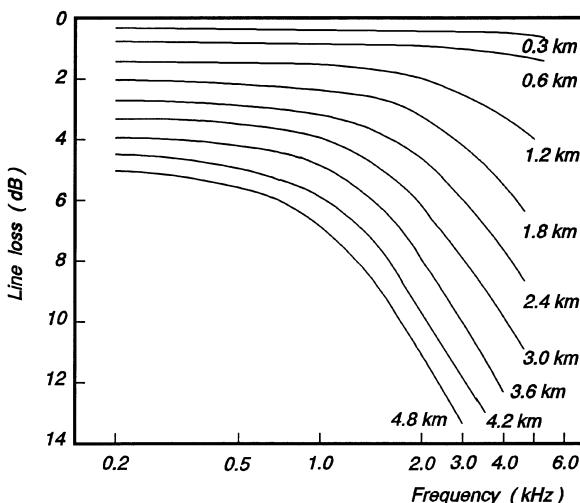


Figure 12.1. Frequency characteristics of 26 gauge BST non-loaded cable terminated in $900\ \Omega$.

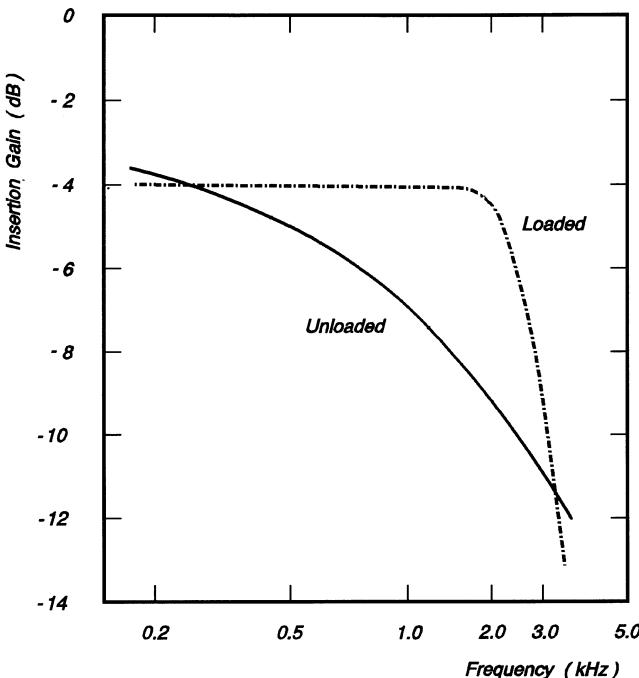


Figure 12.2. Comparison of loaded and unloaded 12,000 feet (3.7 km) number 26 gauge cable terminated in 900Ω and 2 mF . Reprinted with permission from *Transmission Systems for Communications*, 5th Ed., AT&T, Bell Labs, 1982.

It can be seen that, while loading solves the problem of limited bandwidth for the typical subscriber loop voice channel, it is quite inadequate for analog (the basic supergroup requires 552 kHz bandwidth) and digital (DS-1 requires 1.5 MHz bandwidth) carrier applications, for which it is used. In both these cases, the line has to be equalized by placing amplifiers or repeaters at specific distances along its length that emphasize the high-frequency response or regenerate the pulses. Lines used for digital transmission require phase equalization as well, otherwise pulse degradation due to dispersion takes place. Dispersion causes the rate of rise of the leading and trailing edges of the pulse to slow down and the base to spread out over a much longer time than the original pulse.

It can be seen from Figure 12.1 that there is a flat loss at lower frequencies, so it is usual to combine the equalizer with an amplifier. An amplifier used for this purpose is called a *repeater*. A repeater can take a number of forms. In a two-wire system where signals flow in both directions, a *negative impedance converter* is coupled in series and/or in shunt with the line through a transformer. The configuration of the negative-impedance converter and its connection to the line are shown in Figure 12.3.

Measures have to be taken to ensure that the negative impedance does not overwhelm the line impedance resulting in oscillation. The introduction of repeaters

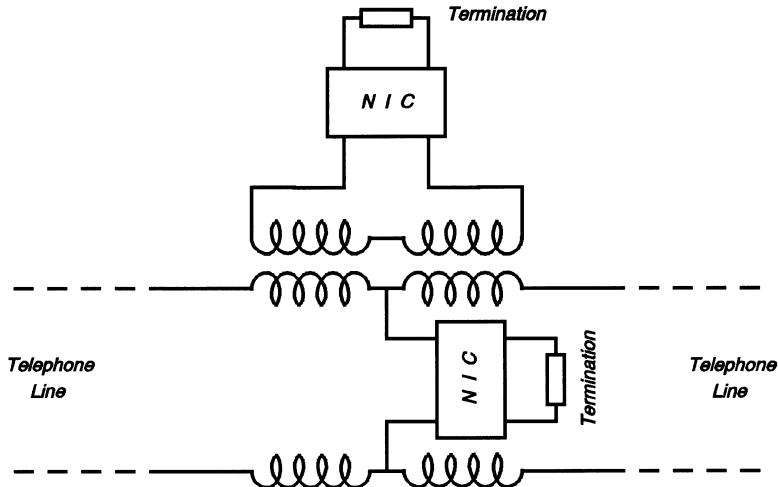


Figure 12.3. The connection of the negative impedance converter (NIC) to the telephone line. Reprinted with permission from *Transmission Systems for Communications*, 4th Ed., AT&T, Bell Labs, 1970.

into the cable causes an impedance mismatch at the point of connection and this can cause echo problems. Severe echo on the cable can impair the speech of most telephone users. There are circuits built into the cable or at the terminations to cancel the echo.

12.2.1 Negative-Impedance Converter

The negative-impedance converter is a two-port which converts an impedance connected to one port into the negative of the impedance at the other port. Consider the two-port shown in Figure 12.4 terminated at port 2 by an impedance Z_L .

If the two-port is a negative impedance converter (NIC) then

$$Z_{in} = -kZ_L \quad (12.2.1)$$

where k is a constant.

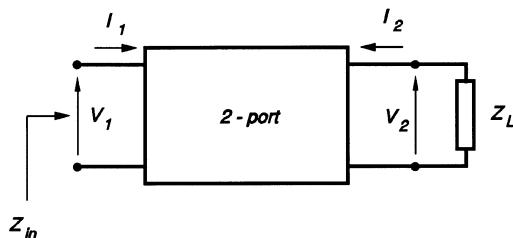


Figure 12.4. A two-port with its defining voltages and currents.

Such a two-port is best described by a chain matrix equation

$$\begin{bmatrix} V_1 \\ I_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} V_2 \\ -I_2 \end{bmatrix}. \quad (12.2.2)$$

This gives

$$V_1 = AV_2 - BI_2 \quad (12.2.3)$$

and

$$I_1 = CV_2 - DI_2. \quad (12.2.4)$$

From the termination we have

$$-I_2 Z_L = V_2 \quad (12.2.5)$$

Substituting into Equations (12.2.3) and (12.2.4) gives

$$V_1 = AV_2 + \frac{BV_2}{Z_L} \quad (12.2.6)$$

and

$$I_1 = CV_2 + \frac{DV_2}{Z_L}. \quad (12.2.7)$$

But for a two-port

$$Z_{\text{in}} = \frac{V_1}{I_1} = \frac{AZ_L + B}{CZ_L + D}. \quad (12.2.8)$$

For Z_L to be equal to $-kZ_L$, $B = C = 0$, then

$$Z_{\text{in}} = \frac{AZ_L}{D} \quad \text{so that} \quad k = -\frac{A}{D}. \quad (12.2.9)$$

There are two possibilities:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} -k_1 & 0 \\ 0 & k_2 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} k_1 & 0 \\ 0 & -k_2 \end{bmatrix}. \quad (12.2.10)$$

Both matrices satisfy the condition for a negative-impedance converter, namely

$$Z_{\text{in}} = \frac{-k_1}{k_2} Z_L = -k Z_L \quad (12.2.11)$$

where $k = k_1/k_2$

From the first matrix,

$$V_1 = -k_1 V_2. \quad (12.2.12)$$

This is called the voltage negative impedance converter or VNIC [5].

From the second matrix,

$$I_1 = k_2(-I_2) \quad (12.2.13)$$

This is called the current negative-impedance converter, INIC or CNIC.

Without loss of generality, we can make $k_1 = k_2 = 1$ so that

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad (12.2.14)$$

The NIC is an example of what is described as a degenerate two-port, that is, it cannot be described by open-circuit impedance [Z] nor short-circuit admittance [Y] parameters. However, it has chain and hybrid parameters (both [h] and [k]).

The VNIC may be described in terms of its hybrid k parameters as follows:

$$\begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \quad (12.2.15)$$

The basic form of the VNIC is shown in Figure 12.5.

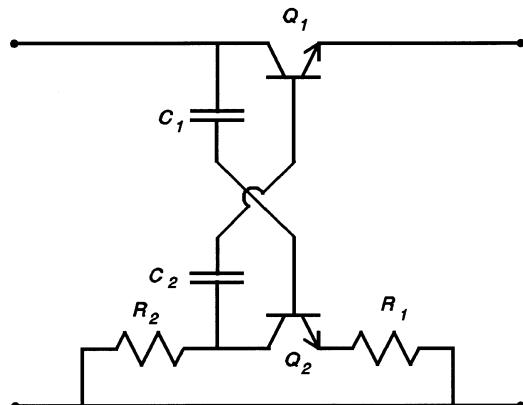


Figure 12.5. The basic form of the voltage negative-impedance converter.

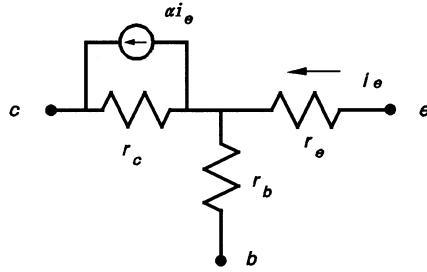


Figure 12.6. The T -equivalent model of the bipolar transistor.

The transistors Q_1 and Q_2 may be represented by the low-frequency T -equivalent model shown in Figure 12.6.

Assuming that C_1 and C_2 are short-circuits at the frequency of operation, the equivalent circuit of the VNIC is as shown in Figure 12.7.

The hybrid k parameters of the circuit in Figure 12.7 are

$$\begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{bmatrix} = \begin{bmatrix} \frac{1 - \alpha_2}{R_1 + r_{e2} + r_{b2}(1 - \alpha_2)} & -\alpha_1 \\ \frac{-\alpha_2 R_2}{R_1 + r_{e2} + r_{b2}(1 - \alpha_2)} & r_{e1} + (r_{b1} + R_2)(1 - \alpha_1) \end{bmatrix} \quad (12.2.16)$$

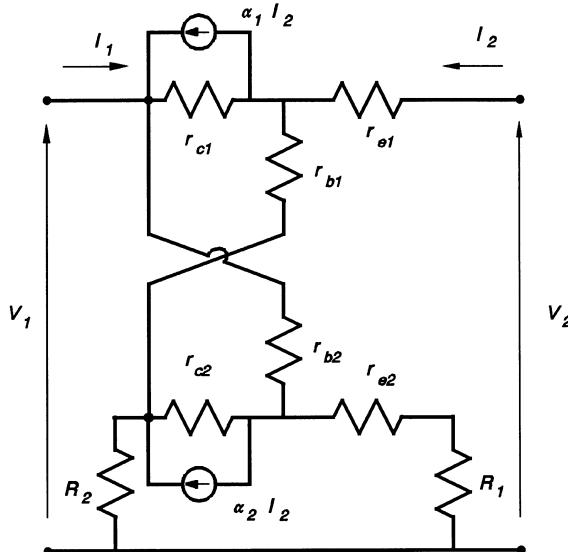


Figure 12.7. The basic VNIC when the transistors have been replaced with the T -equivalent circuit.

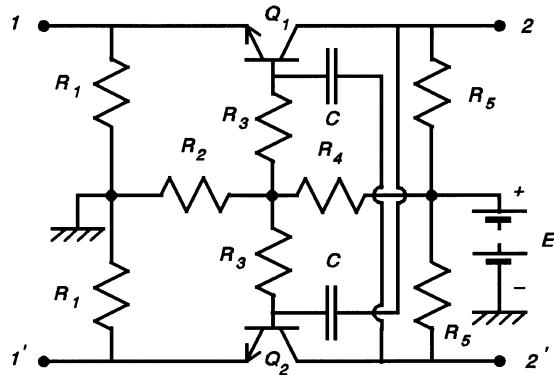


Figure 12.8. The circuit of the voltage negative-impedance converter.

When $R_1 = R_2 = R$ and r_e is small compared to R and $\alpha_1 = \alpha_2 \approx 1$, the circuit behaves like a VNIC. The practical version of the VNIC circuit is shown in Figure 12.8.

12.2.2 Four-wire Repeater

In a four-wire system, the forward and return paths are different and ordinary amplifiers may be used. This is shown in Figure 12.9. Again precautions have to be taken to counteract the possibility of instability through the hybrid-to-hybrid feedback path.

As frequency increases, the twisted pair has the tendency to lose signal power through radiation. Ultimately, its usefulness is limited by cross-talk between pairs.

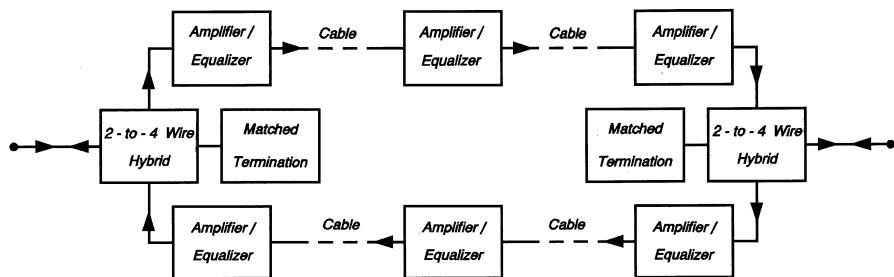


Figure 12.9. The use of ordinary amplifiers on the telephone line with 2-to-4 wire hybrid. Reprinted with permission from *Transmission Systems for Communications*, 4th Ed., AT&T, Bell Labs, 1970.

12.3 COAXIAL CABLE

In a coaxial cable, one conductor is in the form of a tube with the second running concentrically along the axis. The inner conductor is supported by a solid dielectric or by discs of dielectric material placed at regular intervals along its length. A number of these cable are usually combined together with twisted pairs to form a multi-pair cable.

The structure of the coaxial cable ensures that, at normal operating frequencies, the electromagnetic field generated by the current flowing in it is confined to the dielectric. Radiation is therefore severely limited. At the same time, the outer conductor (normally grounded) protects the cable from extraneous signals such as noise and cross-talk.

The primary constants of the coaxial cable are much better behaved than those of the twisted pair. The inductance, L , capacitance, C , and conductance, G , per unit length are, in general, independent of frequency. The resistance, R , per unit length is a function of frequency due to skin effect; it varies as a function of \sqrt{f} .

The frequency characteristics of a 0.375 inch (9.5 mm) coaxial cable are shown in Figure 12.10. As expected, the coaxial cable has a much larger bandwidth than the twisted pair. However, it still requires repeaters and frequency equalizers for analog lines and phase equalization for digital signal transmission. The characteristics of the repeaters are usually adaptively controlled to correct for changes in temperature and other operating conditions.

Coaxial cable is used for transmitting data at 274.176 Mbit/s in the LD-4 (Bell-Canada) and T4M (Bell System in the USA) systems. They have 4032 voice

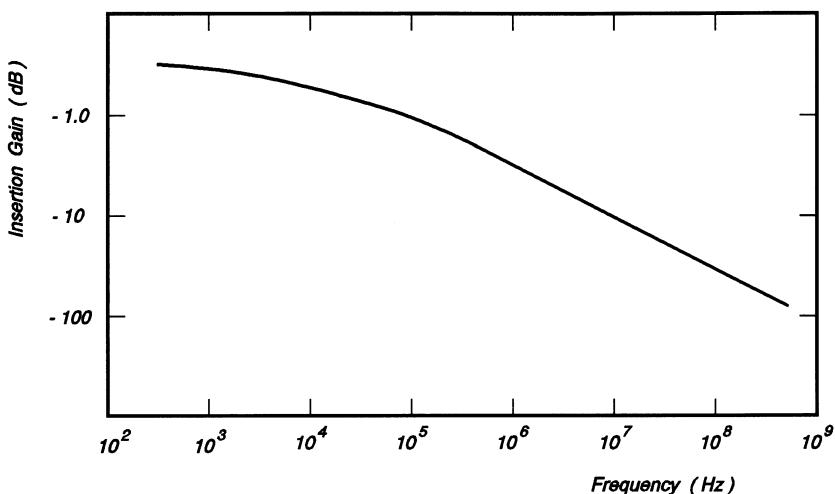


Figure 12.10. The insertion characteristics of a terminated 0.375 inch coaxial cable. Reprinted with permission from *Transmission Systems for Communications*, 4th Ed., AT&T, Bell Labs, 1970.

channels or the equivalent video or digital data traffic. Its regenerators are spaced at 1.8 km intervals and the total length of the line can be 6500 km [1].

Specially constructed coaxial cables with repeaters of very high reliability are used for submarine cable systems. Because of the very high cost of these cables, they are used to transmit messages in both directions by assigning separate frequency bands to each direction. In spite of the development of satellite communication channels, submarine cables are still viable for trans-Atlantic and trans-Pacific traffic. Because of the propagation delay involved in the signal travelling to the satellite and back, most trans-Atlantic telephone conversations use the satellite link in one direction only; cable is used in the opposite direction. In 1976, the TAT-6 (SG) trans-Atlantic cable system was installed. It used a 43 mm diameter coaxial cable with a 4200 voice channel capacity over a distance of 4000 km [2].

12.4 WAVEGUIDES

A waveguide may be viewed as a coaxial cable with the central conductor removed. The outer conductor guides the propagation of the electromagnetic wave. In its most common form it has a rectangular cross section with an aspect ratio of 2 : 1. The wider dimension must be about one-half the wavelength of the wave which it will transmit. Therefore the waveguide has a low-frequency cut-off. There are a number of modes in which the wave can propagate but in every case the electric and magnetic fields are orthogonal. When the electric field is at right angles to the axis of the waveguide, it is described as *transverse electric* (TE) mode. When the magnetic field is at right angles to the axis it is called *transverse magnetic* (TM) mode.

The mechanical structure of the waveguide disqualifies it from being used for long-haul transmission. Irregularities on the walls, such as projections, holes, lack of a perfect match at joints, bends, twists and imperfect impedance matching at the terminations, can cause reflection and spurious modes to be generated, all of which result in signal loss.

Waveguides are used mainly as feedlines to antennas in terrestrial microwave relay systems and for frequencies above 18 GHz they are superior to all other media in terms of loss, noise and power handling.

12.5 OPTICAL FIBER

The use of optical fiber as a medium for telecommunication was made possible by a coincidence of the development of a number of technologies.

- (1) The *laser* which is a coherent frequency source of the order of 10^{14} Hz and it can be modulated. A *light emitting diode* (LED) which produces non-coherent light can also be used.
- (2) A low-loss glass fiber which can be used as a waveguide for the light.
- (3) A detector for the signal at the receiving end.

The laser can be modulated at a rate in the range of 10^9 bit/s while the LED can operate at 10^8 bit/s. The information bearing capacity of the system is enormous. On-going research continues to increase the bit rate limits.

The optical fiber is essentially a high quality glass rod of about 50 μm diameter for multi-mode propagation and 8 μm for single-mode propagation. The mechanical properties of a glass rod that small will make the system impracticable. In practice, a second layer of glass concentric with the optical fiber proper is deposited in the outside, bringing the overall diameter to 125 μm . The outer glass sheathing, referred to as cladding, has a different refractive index and the signal is therefore confined to the core. The core may be of uniform refractive index or it may be graded. These techniques have resulted in optical fibers that have attenuation less than 0.2 dB/km. Various types of protective covering may be put on the fiber and several fibers put together to form a cable.

The optical receiver is a reverse-biassed semiconductor junction diode and it is coupled to the fiber so that the incoming light falls on the junction. The energy in the light is transferred to the electrons in the semiconductor lattice, causing them to break away and move into the conduction band. The high electric field sweeps the electrons out of the junction into the external circuit where they can be detected as a current. Silicon diodes have been used for 1 μm wavelength detectors. For longer wavelengths, such as 1.3 and 1.5 μm , germanium, InAs and InSb are used [3].

In 1988, a trans-Atlantic optical fibre communication system went into service (TAT-8). It spans a distance of 6500 km and provides the equivalent of 40,000 telephone channels. It operates on the 1.3 μm wavelength; repeater separation is 35 km and the bit rate is 274 Mbit/s.

12.6 FREE SPACE PROPAGATION

The transmission media discussed earlier had one thing in common; the propagation of the signal was guided by a twisted pair, a coaxial cable, a waveguide, or an optical fiber. We now consider transmission systems which rely on propagation through free space.

In 1873 Maxwell showed that electromagnetic waves can propagate through free space. It took three decades to demonstrate experimentally that this was possible, when Hertz constructed the first high-frequency oscillator – this was the famous spark-gap apparatus.

A large number of factors have to be taken into account when designing a free space propagation communication system including the following:

- (1) the distance between transmitter and receiver,
- (2) the carrier frequency of the transmission,
- (3) the physical size of the antenna to be used,
- (4) the power to be radiated
- (5) the effect of the transmission on other users of the same and adjacent channels.

Publisher's Note:

Permission to reproduce
this image online was not
granted by the copyright
holder. Readers are kindly
requested to refer to the
printed version of this article.

Figure 12.11. Free-space transmission showing various paths. Reprinted with permission from B. P. Lathi, *Modern Digital and Analog Communication Systems*, CBS College Publishing, New York, 1983.

Figure 12.11 shows different paths by which a signal radiated from a transmitter can reach a receiver.

12.6.1 Direct Wave

As the name suggests, the signal travels directly from the transmitter antenna to the receiver antenna. This requires that the two antennas be in line-of-sight. The direct wave is the major mode of propagation for medium wave AM radio (540–1600 kHz), commercial FM broadcasting (88–108 MHz), terrestrial microwave relay systems, used for long-distance telephone and television signals (2, 4, 6, 11, 18, and 30 GHz), and satellite transmission systems (4, 6, 8, 12, 14, 17–21, and 27–31 GHz).

12.6.2 Earth-Reflected Wave

Part of the propagated wave is reflected off the surface of the Earth and may arrive at the antenna with a different phase from the direct wave. Depending on the magnitude of the reflected wave, it can cause signal fluctuation and sometimes even complete cancellation of the direct wave. This phenomenon has the most noticeable effect on terrestrial microwave relay systems where automatic gain control, diversity protection, and adaptive equalization may be used to counteract it.

12.6.3 Troposphere-Reflected Wave

At a distance of approximately 10 km above the Earth's surface, there is an abrupt change in the dielectric constant of the atmosphere. This portion of the upper atmosphere is called the troposphere. High-frequency signals, such as those used in terrestrial microwave relay systems, may be reflected from the troposphere and have the same effect as the Earth-reflected wave at the receiver. On the other hand, the troposphere may be used as part of the communication channel in cases where direct line-of-sight conditions are not possible. This is called tropospheric scatter propagation and the best frequencies for this are 1, 2, and 5 GHz. It is most commonly used by the military for communication over difficult terrain, typically over a distance of 300–500 km.

12.6.4 Sky-Reflected Wave

Surrounding the Earth at an elevation of approximately 70–400 km is a layer of ionized air caused by constant bombardment of ultraviolet, α , β and γ radiation from the Sun as well as cosmic rays. It is called the ionosphere and it consists of several layers which have different reflective, refractive, and absorptive effects on radio waves.

When a radio signal reaches the ionosphere, a number of things can happen depending on the frequency and the angle of incidence. The wave may be reflected back to Earth, it may undergo refraction and eventually be returned to the Earth or it

may pass through the ionosphere and escape into outer space. With the correct choice of transmission frequency and angle of incidence it is possible to establish communication between two points on the Earth's surface where line-of-sight does not exist. This is the basis of short-wave (3–30 MHz) transmission.

The ionosphere changes its nature from day to night and with such phenomena as sunspot cycles and other cosmic events. The turbulent nature of the ionosphere makes communication by short-wave rather unreliable since the signal is subject to fading – both long and short term. This is due to cancellation and/or reinforcement of the different parts of the signal arriving at the receiver by diverse routes. To improve the reliability of communication via the ionosphere, automatic gain control and diversity protection techniques, such as the utilization of two or more carrier frequencies, are used.

The signal returning to the Earth after reflection from the ionosphere may be reflected from the Earth's surface up to the ionosphere once more. On its second return to Earth it may be received by a suitably placed receiver. This phenomenon is described as multiple hop and can be used to reach places beyond the single-hop distance.

12.6.5 Surface Wave

At very low frequencies (VLF: 3–30 kHz), the ionosphere and the Earth's surface form two parallel conducting planes and can act as a waveguide. VLF signals are used for worldwide communication and navigational aids. At higher frequencies, temperature inversions and other local phenomena can generate a surface wave which can add to the problem of fading.

12.7 TERRESTRIAL MICROWAVE RADIO

Terrestrial microwave radio is a relatively inexpensive medium for long-haul telephone and television signals. Its assigned frequencies are 2, 4, 6, 11, and 18 GHz, which make the system a line-of-sight operation. The distance between repeater stations is approximately 40 km at the lowest frequency and 3 km at the highest frequency, where rain can cause severe attenuation. It is well suited to difficult terrain where the cost of burying or stringing up on posts any form of cable would be prohibitively expensive. The repeater stations can be placed in strategic positions such as hill tops with easy access for maintenance personnel.

The block diagrams of a microwave transmitter and receiver are identical to that used in any other radio system. There are intermediate-frequency amplifiers, oscillators, modulators (upconverters), demodulators (downconverters) filters, equalizers, and automatic gain control (AGC). However, because of the higher frequencies involved, the hardware used to realize the various functions look very different. What appears to the untrained eye to be a piece of printed circuit board may be, in fact, tuned circuits, transmission lines, open and short-circuits and so on.

At microwave frequencies, the electromagnetic wave behaves increasingly like light. Therefore antennas for microwave radio take the form of parabolic dishes or the horn reflector (hog-horn); these structures focus the electromagnetic wave into a narrow beam for optimal transmission. Parabolic dishes are usually used for one frequency band only and the signals may be polarized in the vertical or horizontal direction. The horn reflector type are multi-band and may also be polarized.

A detailed discussion of the design of microwave, antennas, oscillators, amplifiers, modulators, frequency changes, and so on, is outside the scope of this book. A limited list of books on the subject are given in the bibliography.

12.7.1 Analog Radio

In analog radio, a signal made up of a large number of voice frequency telephone channels or its equivalent is formed into a basic group, supergroup, etc. The formation of the signal to be transmitted was discussed in Section 9.2. The appropriate subcarrier frequencies were given. The next step is to use the composite signal to modulate an intermediate-frequency carrier of 70 or 140 MHz before the final upconversion to microwave frequencies. Frequency modulation is the preferred method.

12.7.1.1 Terminal Transmitter and Receiver. The block diagram of the transmitter is shown in Figure 12.12(a). The modulating signal, 70 MHz, itself modulated by 3600 voice-frequency telephone channels (jumbo group) or its equivalent, is amplified and used to drive the modulator, mixer or upconverter. The other input to the upconverter is from the microwave oscillator. The output of the upconverter has to be bandpass filtered to remove undesirable products of the modulation process. After amplification by the radio-frequency amplifier, the signal goes to the transmit antenna for radiation.

The receiver is shown in Figure 12.12(b). The received signal is bandpass filtered to eliminate all but the required signal. It is then mixed with the output of the local oscillator to produce an intermediate-frequency signal. After amplification, equalization and the application of AGC, it goes to a demodulator where the original modulating signal (70 MHz) is recovered. It should be noted that, in the case of the 3600 voice-frequency telephone channels, several levels of demodulation have to be carried out before the voice frequency signals are recovered.

12.7.1.2 Repeater. A typical analog microwave radio repeater is shown in Figure 12.13. The antenna is very highly directive and it must be secured in a position where it faces the transmitter antenna as directly as possible; small deviation can cause significant loss of signal power. Antennas 1 and 2 are used simultaneously for transmission and reception. The received signal, RF1, goes to the receive circulator which directs it to the bandpass filter where all signals other than those desired are attenuated. The local oscillator and the mixer downconvert the signal to an intermediate frequency. After amplification, equalization, and the application of AGC, the signal is upconverted by the microwave oscillator and the upconvert mixer

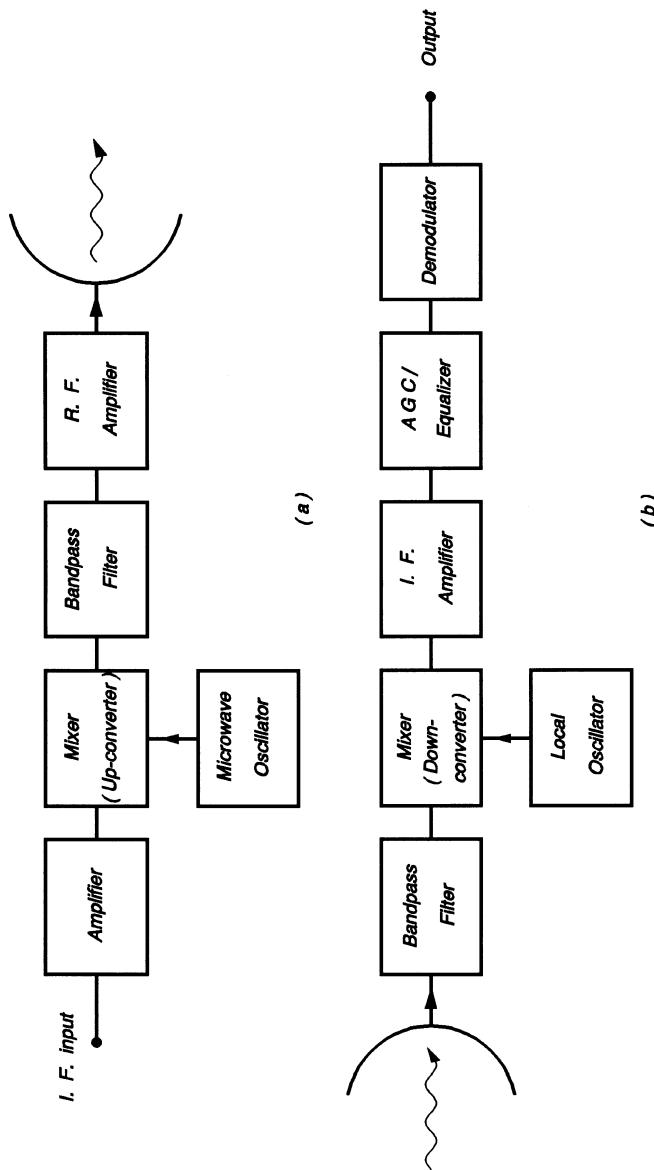


Figure 12.12. A typical analog microwave radio transmitter and receiver.

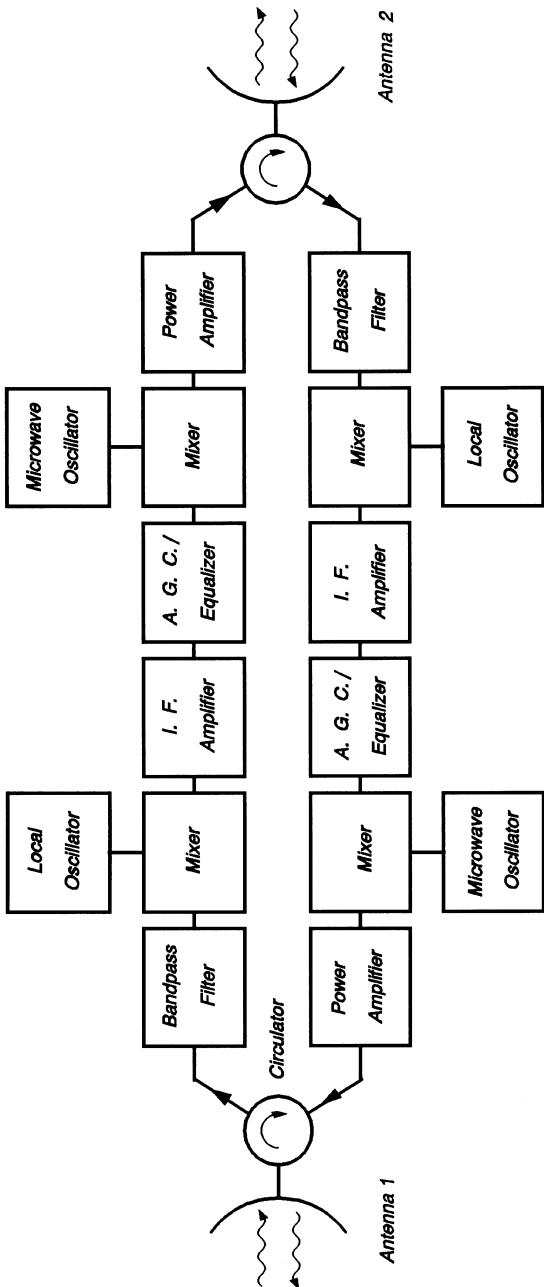


Figure 12.13. A typical analog microwave repeater. Note that the antennas are used for transmission and reception simultaneously.

to a different frequency, RF2 (changing the frequency reduces the possibility of instability in the system). The RF amplifier boosts the signal and feeds it to the transmit circulator which passes it on to the antenna for onward transmission.

The signal travelling in the opposite direction, RF3 (different from both RF1 and RF2), follows an identical path in the opposite direction. The intermediate frequency used in this path is, in general, different from the one used earlier. The signal leaves the system at a frequency RF4, different from all the others.

12.7.2 Digital Radio

All the functional blocks of the analog radio shown in Figure 12.13 are present in some form or another in the digital radio. The baseband signal to be transmitted may be the output of a digital switch operating at a rate of 1.544, 6.312, 44.736 or 274.176 Mbit/s. The binary output of the switch is used in a modulation scheme similar to that of the modem discussed in Section 9.4.1. The digit **1** is assigned a frequency f_1 and **0** is assigned a different frequency f_2 . Usually f_1 and f_2 are equally displaced from the carrier frequency f_0 , that is, $(f_0 - f_1) = (f_2 - f_0)$. At the receiving end, the demodulation process converts f_1 and f_2 to **1s** and **0s**.

12.7.2.1 Regenerative Repeater. The block diagram of the regenerative repeater is shown in Figure 12.14. The system is identical to that shown in Figure 12.13 except that the signal is demodulated to the baseband and then regenerated, modulated, filtered, upconverted, and amplified for onward transmission. The signal travelling in the opposite direction is subjected to the same processing steps, with the exception that the carrier radio frequencies are selected to minimize the possibility of instability. The regenerative repeater is described in Section 9.3.9.6.

12.8 SATELLITE TRANSMISSION SYSTEM

A satellite transmission system is just another microwave relay system with a single repeater located in outer space. Because of its height, it can cover a large area of the globe, making it possible to cover the entire surface with three geostationary satellites. It has applications in point-to-point communications as well as broadcasting and it can reach remote parts of the Earth where other systems cannot reach without large expenditures of money and effort. All that is required is the satellite and the terminal equipment in the Earth stations. Satellites are, however, not cheap. The cost of launching them and the difficulty of making repairs should anything go wrong dictate very high levels of reliability.

To avoid the complexity of tracking the satellite from rising to setting and to maintain continuous communication, it is best to “park” the satellite at a point directly above the equator and to choose the correct speed in the direction of rotation of the Earth so it stays in a relatively fixed position above a reference point on the Earth’s surface. In general, satellites rotate in an elliptical orbit with the Earth at one of the two foci of the ellipse (apogee and perigee). The two foci can be at the same

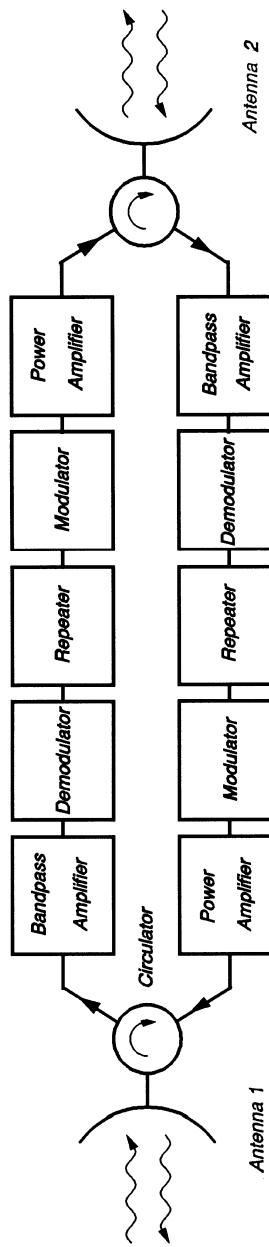


Figure 12.14. The regenerative repeater used in digital radio.

point, in which case the orbit is circular. There are distinct advantages to having a satellite in a circular orbit since this fixes the distance travelled by the message and hence the delay. To satisfy these conditions, an Earth satellite has to be 42,230 km from the center of the Earth. But it is not enough to park the satellite in this position, as it tends to drift away slowly due to the non-spherical shape of the Earth, the gravitational influence of the Sun, the Moon and the other planets. A small rocket is provided for the correction of minor deviations from the nominal position. The useful life of the satellite is therefore determined by how long the rocket fuel lasts. The satellite is made to spin or it has a wheel in it which spins. The gyroscopic effect of this helps to further stabilize the satellite. If the satellite itself spins, it is necessary to spin the antenna in the opposite direction to keep it pointing at the Earth at all times.

For about 30 minutes a day for several days around the equinoxes (March 21 and September 21) the Sun is directly behind the satellite and the electrical noise generated by the Sun makes operation impossible. It is necessary to switch to an alternate satellite or rely on more Earth-bound means of communication. Once a day, the satellite is eclipsed by the Earth and it not only loses its source of power from the solar panels, it experiences a drastic change in temperature. To overcome these problems, a battery is provided and all components are designed to operate in the extreme temperature conditions of outer space.

Placing the satellite above the equator means that Earth stations in the extreme north and south are not well illuminated by the satellite radiation. Furthermore, the low angle of elevation of the Earth station antenna makes it vulnerable to atmospheric fading and interference from terrestrial manmade noise. One solution to this problem is to use several satellites in non-equatorial, non-stationary orbits so that, as one satellite sets, the next one is rising. This means that tracking is necessary. The Russian domestic communication satellite system (Molniya) was designed on this basis since most of the territory it is designed to serve is in the northern part of the Northern Hemisphere.

The frequency bands assigned for satellite communication are 4–6, 12–14 and 19–29 GHz. In the higher frequency bands, there is increasing attenuation due to signal power dissipation in rain droplets and water vapor in the atmosphere. Frequency reuse is possible so long as the satellites are sufficiently far apart in orbit for an Earth station to focus on only one of them at a time.

The repeater on a satellite is usually referred to as a *transponder*. The basic structure of a typical transponder is shown in Figure 12.15. The antenna is used for both transmission and reception. The path for a signal arriving from Earth is through the circulator to the bandpass filter where all but the desired band of frequencies containing the carrier are eliminated. Some gain is provided by an amplifier and the mixer and the local oscillator change the carrier frequency to a different value. After further bandpass filtering to remove the unwanted products of the mixing process, the power amplifier boosts the signal power to the appropriate level for the return trip to Earth through the circulator and antenna. System instability is prevented by the correct choice of circulator characteristics and the change of the carrier frequency.

Publisher's Note:
Permission to reproduce
this image online was not
granted by the copyright
holder. Readers are kindly
requested to refer to the
printed version of this article.

Figure 12.15. The communication satellite transponder. Reprinted with permission from V. I. Johannes, "Transmission Systems for Telecommunications" in *Electronic Engineers' Handbook*, Fink, D. G. and Christiansen, D. (Eds), McGraw-Hill, Inc, New York, 1982.

The loss of signal power between an Earth station and the satellite is about 200 dB [10]. The transponder power output is usually about 10 W; higher power is not desirable because of the possibility of interference with other Earth-based communication systems. Besides, it is expensive to provide the dc power to run a high-power amplifier in outer space. To overcome these limitations, Earth stations use large antennas and high power transmitters and very low-noise amplifiers in the receiver. Sometimes the amplifiers are cooled with liquid nitrogen to achieve the necessary low-noise performance. Figure 12.16 shows the gains and losses of power in a television signal as it makes its way up to the satellite and back down to Earth.

The great distances traversed by the signal from one station to the other causes enough delay (240 ms) to make a telephone conversation over the satellite channel somewhat difficult; one person starts to talk when the other is already talking, without realizing it. This leads to a series of interruptions in the conversation. In general this is not intolerable but the situation certainly gets much worse when the delay is doubled for a two-way conversation. To get around this problem, the normal practice in a trans-Atlantic telephone conversation is that when one link is via satellite, the other must be via submarine cable. The satellite system also generates an echo that has the psychological effect of making people with no speech impediment stutter. The circuits used in satellite transmission usually have an echo cancellation feature.

To give some idea of how quickly satellite communication systems are changing, it is worth noting that, in early 1990, there were about 94 communication satellites in operation with another 194 launchings planned in the near future. By the end of the year 2000, there was a total of 2271 satellites in orbit (Russia had 1322, the United States 658, Japan 55 and 51 were international) [11]. It should be noted that not all of the satellites listed are for civilian communication. The first of a series of the International Telecommunication Satellites (INTELSAT I) was launched in 1965 and it had 240 two-way telephone voice channels. In 1980 INTELSAT V went into

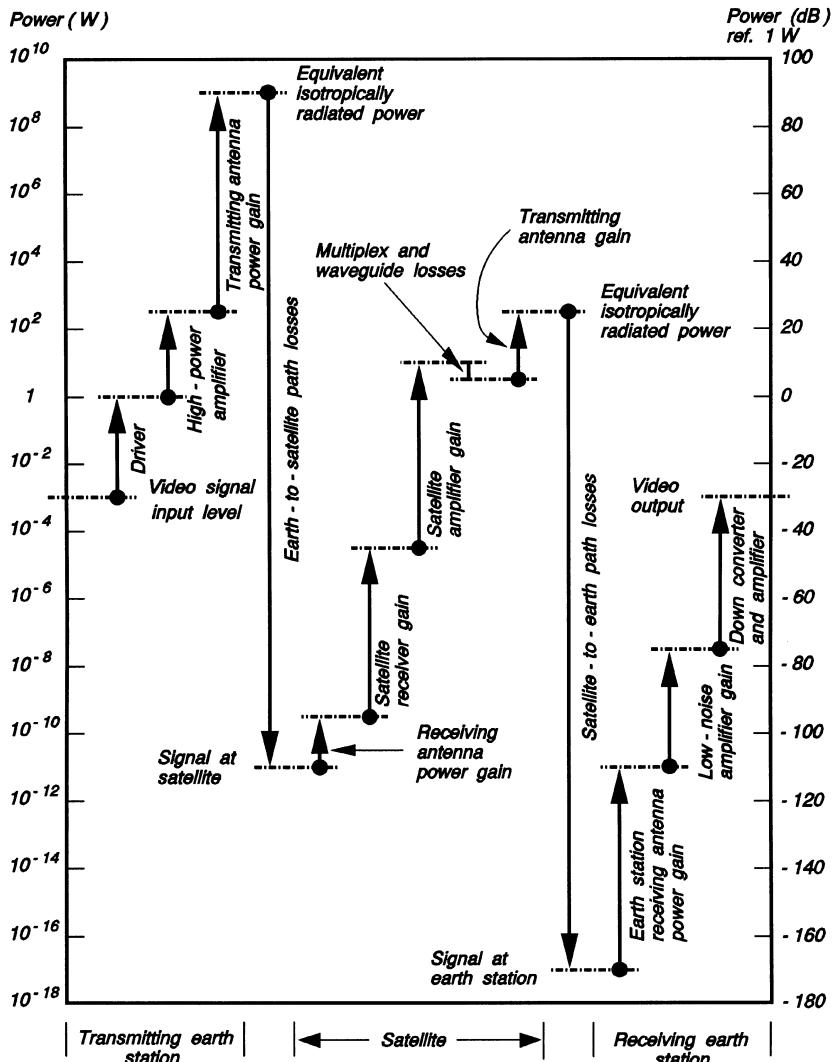


Figure 12.16. Power levels in the transmission of a TV signal via satellite. Reprinted with permission from B. I. Edelson, "Global Satellite Communications", *Scientific American*, February 1977.

service with 12,000 two-way voice and two television channels. By the end of the year 2000 there were 19 Intelsat satellites in orbit with plans for 22 new launches. The latest operational satellite, the INTELSAT VIII, has 22,500 two-way telephone voice and three television channels, and with the use of *digital circuit multiplication equipment* (DCME) the number of voice channels can be increased up to 112,500 [12].

REFERENCES

1. Telecom Canada, *Digital Network Notes*, Ottawa, Canada, 1983.
2. Easton, R. L., "Undersea Cable Systems – A Survey", *IEEE Comm. Magazine*, 13(5), 12–15, 1975.
3. Jones, W. E., *Introduction to Optical Fiber Communication Systems*, Holt, Rinehart and Winston, New York, 1988.
4. Members of the Technical Staff, *Transmission Systems for Communications*, 5th Ed., Bell Telephone Laboratories Inc., 1982. Publisher: Western Electric Co. Inc., Winston-Salem, NC.
5. Linvill, J. G., "Transistor Negative Impedance Converter", *Proc. IRE.*, 41, 725–729, 1953.
6. Li, T. "Advances in Optical Fibre Communication: an Historical Perspective", *IEEE J. Selected Areas in Comm.* SAC-1(3), 356–372. 1983.
7. Schwartz, M. I., "Optical Fiber Transmission – From Conception to Prominence in 20 Years", *IEEE Comm. Magazine*, 22(5), 38–48, 1984.
8. Pritchard W. L., "The History & Future of Commercial Satellite Communications", *IEEE Comm. Magazine*, 22(5), 22–37, 1984.
9. Johannes, V. I., *Electronics Engineers' Handbook*, Fink, D. G. and Christiansen, D. (Eds), McGraw-Hill, Inc, New York, 1982.
10. Edelson, B. I., "Global Satellite Communications", *Sci. Am.*, Feb., 1977.
11. www.atak.com/satellite/table.html
12. www.intelsat.int/tech/sats/satellite/html

BIBLIOGRAPHY

- Basawaptna, G. R. and Stancliff, R. B., "A Unified Approach to the Design of Wideband Microwave Solid-State Oscillators", *IEEE Trans. MTT-27(5)*, 379–385, 1979.
- Edwards, T. C., *Foundations for Microstrip Circuit Design*, John Wiley & Sons, New York, 1981.
- Ghandi, O. P., *Microwave Engineering Applications*, Pergamon Press, New York, 1981.
- Gonzalez, G., *Microwave Transistor Amplifiers: Analysis and Design*, Prentice-Hall, Englewood Cliffs, NJ, 1984.
- Ha, T. T., *Solid-State Microwave Amplifier Design*, John Wiley & Sons, New York, 1981.
- Ishii, T. K., *Microwave Engineering*, Harcourt Brace Jovanovich, Publishers, San Diego CA, 1989.
- Liao, S. Y., *Microwave Circuit Analysis and Amplifier Design*, Prentice-Hall, Englewood Cliffs, NJ, 1987.
- Midford, T. A. and Bernick, R. L., "Millimeter-wave CW-IMPATT Diodes and Oscillators", *IEEE, MTT-27(5)*, 483–492, 1979.
- Sadiku, M. N. O., *Elements of Electromagnetism*, Saunders College Publishing, New York, 1989.
- White, J., *Microwave Semiconductor Engineering*, Van Nostrand Reinhold, New York, 1982.

PROBLEMS

- 12.1** Describe how you would transmit a video signal of approximate bandwidth 1.5 MHz over a twisted-pair cable.
- 12.2** Derive an expression for the input impedance of the circuit shown in Figure P12.1, assuming that the operational amplifier is ideal. How can this circuit be used to improve the transmission characteristics of the twisted-pair or coaxial cable?

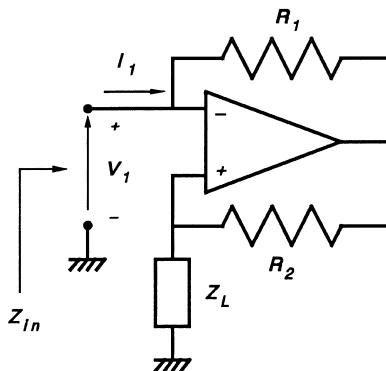


Figure P12.1.

- 12.3** Describe two techniques which can be used for the compensation of gain loss on a twisted-pair and/or coaxial cable. Discuss how these techniques can be modified to compensate for the relative loss of gain as a function of frequency.
- 12.4** Discuss the merits of optical fiber transmission relative to satellite communication.
- 12.5** Discuss the role of the ionosphere in the free space propagation of very low frequencies (VLF) and short-wave transmission between two points on the Earth's surface.
- 12.6** Terrestrial microwave radio transmission is subject to severe attenuation in the atmosphere. What elements in the atmosphere are responsible for this; can you suggest a simple explanation of this phenomenon?
- 12.7** Describe the processes that take place at a typical terrestrial microwave analog relay station using a suitable block diagram. What are "upconversion" and "downconversion" and why are they necessary?
- 12.8** Using suitable block diagrams, discuss the major differences between analog and digital microwave relay stations.
- 12.9** Compare and contrast the satellite transponder to the microwave relay station paying particular attention to signal processes that occur in them.

APPENDIX A

THE TRANSFORMER

A.1 INTRODUCTION

A transformer is a device which is used for coupling a signal or power from a source to a load in an efficient manner. It does this by changing the impedance of the source to match that of the load. A transformer can also be used for phase inversion and dc isolation. The transformer consists of two or more coils of wire wound on a common core so that the magnetic field of one coil links with the other coils. When a changing current is applied to one winding, it causes a changing magnetic field in the common core. The changing magnetic field induces a voltage in all the other coils and if a path exists, current will flow. The coils are said to be mutually coupled.

Consider the two coils connected as shown in Figure A.1. The application of KVL to loop 1 gives

$$j\omega L_{11}I_1 + j\omega M_{21}I_2 - E = 0 \quad (\text{A.1})$$

where the voltage $j\omega M_{21}I_2$ is induced in loop 1 as a result of the current I_2 flowing in loop 2 and it opposes the applied voltage E .

The application of KVL to loop 2 gives

$$j\omega M_{12}I_1 - (j\omega L_{22} + Z_L)I_2 = 0. \quad (\text{A.2})$$

where the voltage $j\omega M_{12}I_1$ is induced in loop 2 as a result of the current I_1 flowing in loop 1.

It is evident that

$$M_{12} = M_{21} = M \quad (\text{A.3})$$

where M is the mutual inductance between the two coils.

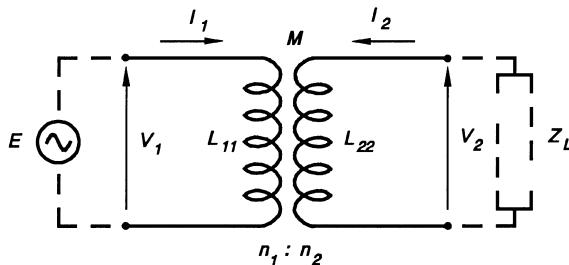


Figure A.1. An equivalent circuit of a two-winding transformer showing the applied voltages and currents.

Two coils may be closely or loosely coupled, depending on how much of the magnetic field generated by one links with the other. The coupling coefficient k is related to the mutual inductance M and the self-inductances of the two coils by

$$k = \frac{M}{\sqrt{L_{11}L_{22}}}. \quad (\text{A.4})$$

A.2 THE IDEAL TRANSFORMER

A transformer may be considered to be ideal if it satisfies all the conditions below:

(1)

$$a = \frac{n_1}{n_2} = \frac{V_1}{V_2} = \frac{I_2}{I_1} = \sqrt{\frac{L_{11}}{L_{22}}}. \quad (\text{A.5})$$

- (2) The primary inductance L_{11} and the secondary inductance L_{22} both approach infinity while the ratio L_{11}/L_{22} remains finite.
- (3) The coupling coefficient $k = 1$.
- (4) The windings have no resistance and the magnetic core has no losses.
- (5) Stray capacitances of the coils are zero and stray capacitance between the coils is zero.

To investigate the impedance transformation property of the transformer, consider the circuit shown in Figure A.2. Assuming that the resistance seen across the primary of the transformer is R_{in} then power flowing into the transformer is

$$P_{\text{in}} = \frac{V_1^2}{R_{\text{in}}}. \quad (\text{A.6})$$

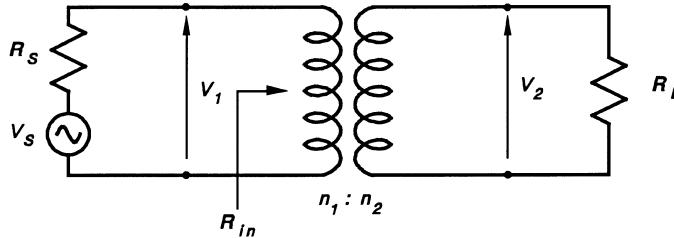


Figure A.2. The equivalent circuit of the ideal transformer used for matching the source resistance, R_S to the load resistance R_L .

and power flowing into the load R_L is

$$P_0 = \frac{V_2^2}{R_L}. \quad (\text{A.7})$$

Since the transformer is ideal, $P_{\text{in}} = P_{\text{out}}$. Therefore,

$$R_{\text{in}} = \left(\frac{V_1}{V_2} \right)^2 R_L. \quad (\text{A.8})$$

But

$$a = \frac{V_1}{V_2}. \quad (\text{A.9})$$

Therefore

$$R_{\text{in}} = a^2 R_L. \quad (\text{A.10})$$

The source then “sees” a load equal to $a^2 R_L$ instead of R_L if the load has been connected directly to the source. Since the turns ratio, a is a variable, it can be chosen so as to optimize the transfer of power to the load. This occurs when

$$a^2 R_L = R_S. \quad (\text{A.11})$$

The result can be generalized for the load impedance

$$Z_L = R_L + jX_L \quad (\text{A.12})$$

and the source impedance

$$Z_S = R_S + jX_S \quad (\text{A.13})$$

as

$$Z_S^* = a^2 Z_L. \quad (\text{A.14})$$

A.3 THE PRACTICAL TRANSFORMER

A transformer can be constructed in which the conditions in Equation (A.5) apply within a reasonable approximation. However, it is impossible to make the self-inductances L_{11} and L_{22} infinite. The coupling coefficient k can be made approximately equal to unity by using bifilar windings on a high permeability core. Winding resistance and core losses can be reduced considerably by a judicious choice of materials. Stray capacitance can be minimized by increasing the thickness of insulation between layers of the windings. By taking into account all the imperfections, an equivalent circuit of the practical transformer is as shown in Figure A.3, where

- R_1 is the winding resistance of the primary,
- R_2 is the winding resistance of the secondary,
- L_1 is the leakage inductance of the primary,
- L_2 is the leakage inductance of the secondary,
- C_1 is the total stray capacitance of the primary,
- C_2 is the total stray capacitance of the secondary,
- R_c represents an equivalent resistance in which the core losses are dissipated,
- L_m is the primary inductance in which the magnetizing current flows,
- C_m is the total stray capacitance between primary and secondary.

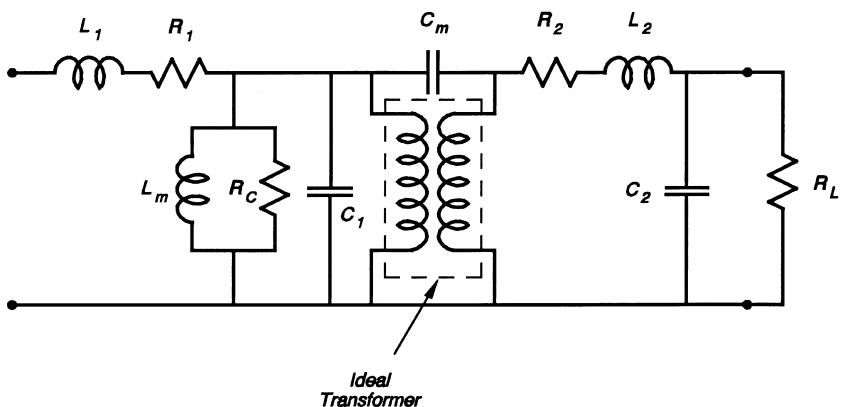


Figure A.3. The equivalent circuit of a practical transformer. The ideal transformer is used in conjunction with the non-ideal components which are shown outside the black-box containing the ideal transformer.

A number of the components shown in Figure A.3 can be made negligible by taking the following precautions:

- (1) Using low resistance wire for the windings will make R_1 and R_2 negligibly small compared to the source and load resistances.
- (2) Using bifilar windings in which the primary and secondary coils are wound side-by-side and turn-by-turn, on a highly permeable core, the leakage inductances L_1 and L_2 can be considered to have been eliminated. It is evident that the use of bifilar windings will increase the value of C_m .
- (3) The use of a thicker insulation between layers of winding will reduce C_1 and C_2 but will increase the leakage inductances.
- (4) By laminating the core, and thus increasing the resistance of the path for eddy currents to flow, the core losses can be reduced and therefore R_c can be left out of the equivalent circuit.

It should be noted that some of the steps described above are contradictory and suitable compromises have to be made to get the simplified equivalent circuit shown in Figure A.4.

The ideal transformer and its load R_L can be replaced by a resistance equal to $a^2 R_L$ as shown in Figure A.5. By applying Thevenin's theorem to the left of $X-X'$ the circuit can be simplified as shown in Figure A.6

$$V_{th} = \frac{a^2 R_L}{a^2 R_L + R_S} V_S. \quad (\text{A.15})$$

It can be seen that this is a high-pass filter with a slope of 6 dB per octave and a corner frequency

$$\omega_c = \frac{a^2 R_L R_S}{L_m (a^2 R_L + R_S)}. \quad (\text{A.16})$$

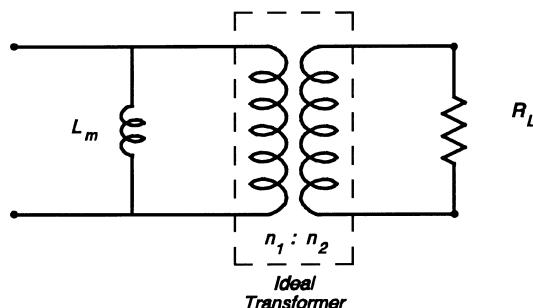


Figure A.4 The simplified equivalent circuit of the practical transformer when steps have been taken to minimize the effects of the non-ideal components.

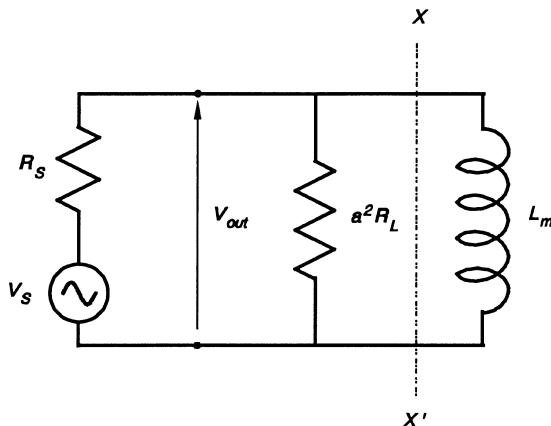


Figure A.5 The equivalent circuit for the analysis of the frequency characteristics of the transformer.

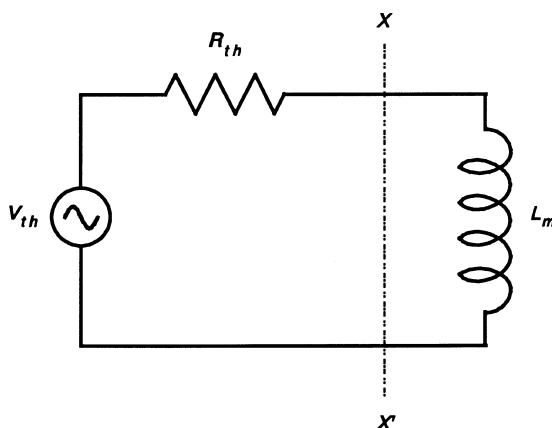


Figure A.6 The Thevenin equivalent circuit of the transformer showing that it is a high-pass circuit.

It is advantageous to make L_m (the primary inductance L_{11}) as large as possible in order to keep the corner frequency as low as possible.

The condition for maximum power transfer exists if

$$a^2 R_L = R_S \quad (\text{A.17})$$

then

$$\omega_c = \frac{R_S}{2L_m}. \quad (\text{A.18})$$

APPENDIX B

DESIGNATION OF FREQUENCIES

30–300 Hz	Extremely Low Frequency (ELF)
300–3 kHz	Voice Frequency (VF)
3–30 kHz	Very Low Frequency (VLF)
30–300 kHz	Low Frequency (LF)
300–3 MHz	Medium Wave Frequency (MW)
3–30 MHz	Short Wave Frequency (SW)
30–300 MHz	Very High Frequency (VHF)
300–3000 MHz	Ultra High Frequency (UHF)
3–30 GHz	Super High Frequency (SHF)
30–300 GHz	Extremely High Frequency (EHF)

VHF TELEVISION FREQUENCIES

Channel	Frequency band (MHz)	Video carrier (MHz)
1	(Not normally used)	
2	54–60	55.25
3	60–66	61.25
4	66–72	67.25
5	76–82	77.25
6	82–88	83.25
Commercial FM Broadcast Band (88–108 MHz)		
7	174–180	175.25
8	180–186	181.25
9	186–192	187.25
10	192–198	193.25
11	198–204	199.25
12	204–210	205.25
13	210–216	211.25

UHF TELEVISION FREQUENCIES

Channel	Frequency band (MHz)	Video carrier (MHz)
14	470–476	471.25
15	476–482	477.25
16	482–488	483.25
17	488–494	489.25
18	494–500	495.25
19	500–506	501.25
20	506–512	507.25
21	512–518	513.25
22	518–524	519.25
23	524–530	525.25
24	530–536	531.25
25	536–542	537.25 etc.

The channels follow in order, each one being 6 MHz in bandwidth with the carrier at 1.25 MHz above the lower cut-off frequency. The total number of channels is 69 and channel 69 has a bandwidth from 800–806 with the carrier at 801.25 MHz. It should be noted that in every case the FM carrier for the audio is 4.5 MHz above the video carrier.

APPENDIX C

THE ELECTROMAGNETIC SPECTRUM

Figure C.1 is a simplified electromagnetic spectrum with various uses and bandwidths shown.

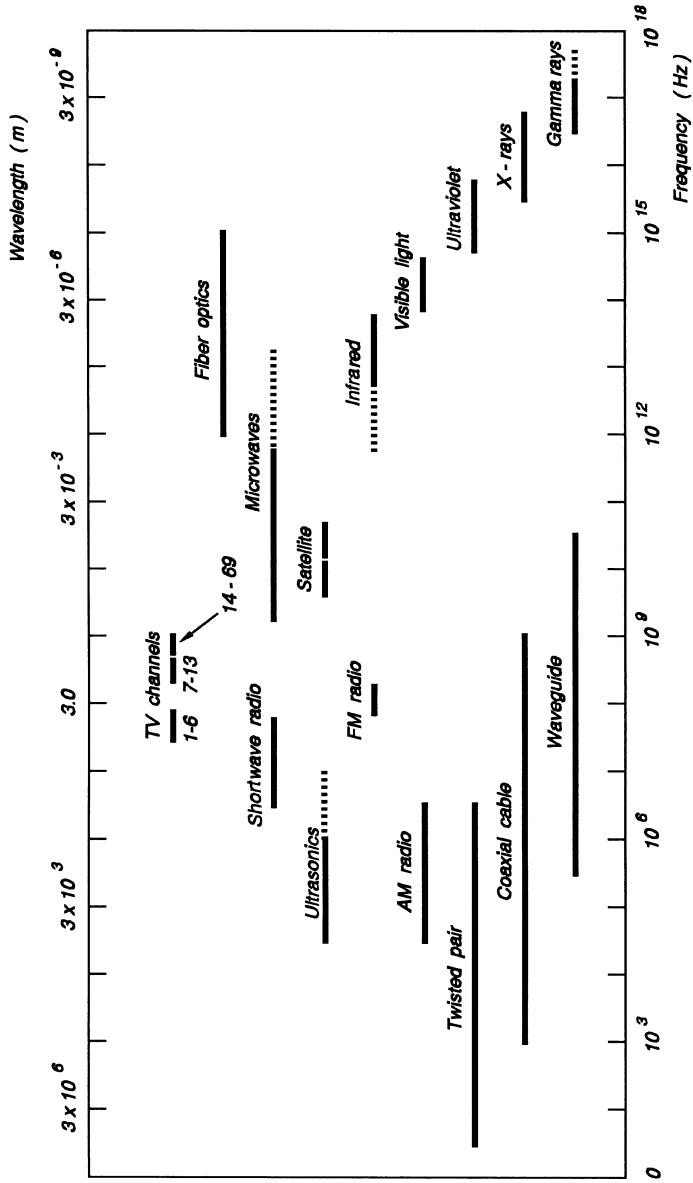


Figure C.1.

APPENDIX D

THE MODIFIED HUFFMAN DATA COMPRESSION CODE

Reprinted with permission from McConnell, Bodson and Urban, *Fax: Facsimile Technology and Systems*, 3rd Ed., Artech House, Boston, MA, 1999.

The Modified Huffman data compression code for facsimile machines is given in Tables D.1 and D.2. Table D.1 covers black and white runs of 0–63 and Table D.2 covers run lengths from 64 to the maximum of 1728. Run lengths not given in Table D.2 can be made up from both tables.

TABLE D.1

Run length	Code word (white)	Code word (black)
0	00110101	0000110111
1	000111	010
2	0111	11
3	1000	10
4	1011	011
5	1100	0011
6	1110	0010
7	1111	00011
8	10011	000101
9	10100	000100
10	00111	000100
11	01000	0000101
12	001000	0000111
13	000011	00000100
14	110100	00000111
15	110101	000011000
16	101010	0000010111
17	101011	0000011000

(continued)

TABLE D.1 (*continued*)

Run length	Code word (white)	Code word (black)
18	0100111	0000001000
19	0001100	00001100111
20	0001000	00001101000
21	0010111	00001101100
22	0000011	00000110111
23	0000100	00000101000
24	0101000	00000010111
25	0101011	00000011000
26	0010011	000011001010
27	0100100	000011001011
28	0011000	000011001100
29	00000010	000011001101
30	00000011	000001101000
31	00011010	000001101001
32	00011011	000001101010
33	00010010	000001101011
34	00010011	000011010010
35	00010100	000011010011
36	00010101	000011010100
37	00010110	000011010101
38	00010111	000011010110
39	00101001	000011010111
40	00101001	000001101100
41	00101010	000001101101
42	00101011	000011011010
43	00101100	000011011011
44	00101101	000001010100
45	00000100	000001010101
46	00000101	000001010110
47	00001010	000001010111
48	00001011	000001100100
49	01010010	000001100101
50	01010011	000001010010
51	01010100	000001010011
52	01010101	000000100100
53	00100100	000000110111
54	00100101	000000111000
55	01011000	000000100111
56	01011001	000000101000
57	01011010	000001011000
58	01011011	000001011001
59	01001010	000000101011
60	01001011	000000101100
61	00110010	000001011010
62	00110011	000001100110
63	00110100	000001100111

TABLE D.2

Run length	Code word (white)	Code word (black)
64	11011	0000001111
128	10010	000011001000
192	010111	000011001001
256	0110111	000001011011
320	00110110	000000110011
384	00110111	000000110100
448	01100100	000000110101
512	01100101	0000001101100
576	01101000	0000001101101
640	01100111	0000001001010
704	011001100	0000001001011
768	011001101	0000001001100
832	011010010	0000001001101
896	011010011	0000001110010
960	011010100	0000001110011
1024	011010101	0000001110100
1088	011010110	0000001110101
1152	011010111	0000001110110
1216	011011000	0000001110111
1280	011011001	0000001010010
1344	011011010	0000001010011
1408	011011011	0000001010100
1472	010011000	0000001010101
1536	010011001	0000001011010
1600	010011010	0000001011011
1664	011000	0000001100100
1728	010011011	0000001100101
EOL*	000000000001	000000000001

*EOL = End of Line.

APPENDIX E

ELECTRONIC MEMORY

E.1 INTRODUCTION

Modern electronic systems would be inconceivable without the possibility to store information. Information in the form of music has been stored on wax cylinders, vinyl discs, magnetic tapes and compact discs (CD). As electronic systems become increasingly digital, the need to store large quantities of data has grown. In digital electronic storage systems every bit (**1** or **0**) is stored in a device or circuit which can take only one of two states at any time. There has to be a mechanism for changing the state of the device or circuit according to the nature of the bit to be stored. It must be possible for the stored information to be retrieved when required. Magnetic tapes and discs are the preferred media for the storage of electronic data where the data is to be retained even after the system power is switched off. A second medium commonly used for data storage is the flip-flop which can take one of two states (high or low corresponding to **1** or **0**). In this case when the system power is shut off, the information stored is lost.

E.2 BASICS OF S-R FLIP-FLOP CIRCUITS

Figure E.1(a) shows the NOT gate (non-resettable) equivalent of the flip-flop, (b) shows the resettable equivalent using NOR gates and (c) shows a second realization using NAND gates.

In the flip-flop circuit shown in Figure E.1(b), the *S* and *R* terminals are the control inputs. *Q* and \bar{Q} are the outputs and they are complementary (that is when one of them is a **1** the other must be a **0**). There are four different combinations of **1**s and **0**s that can exist at the terminals *S* and *R*.

- (1a) If we assume that $S = R = 0$ then, since we are dealing with two-input NOR gates, it follows that the output, *Q* (and its complement), is determined by the logic levels present at the input terminals. If we assume that $\bar{Q} = 1$

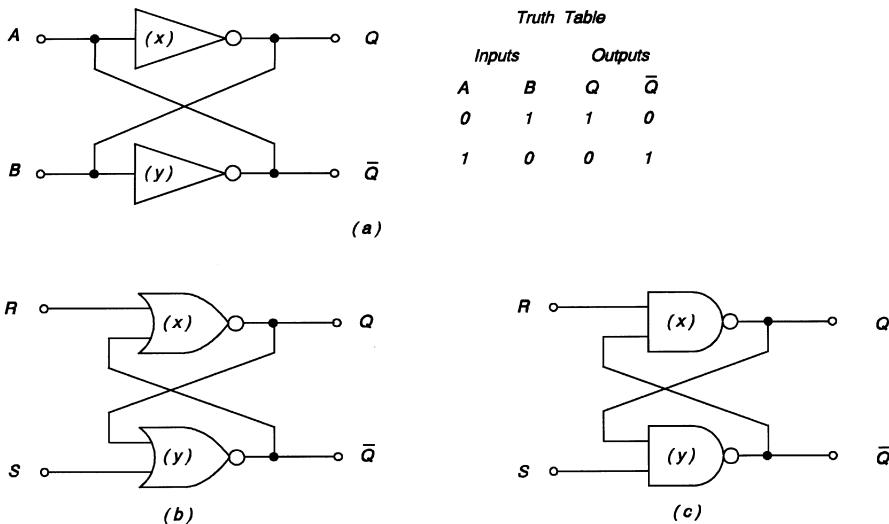


Figure E.1. (a) Two NOT gates connected to form a non-resettable flip-flop. (b) The connection of two two-input NOR gates to form an *S-R* flip-flop (resettable). (c) The connection of two NAND gates to form an *S-R* flip-flop.

then the lower input terminal of NOR gate, x is a **1** and hence the output of NOR gate x is a **0** ($Q = 0$). It follows that the upper input terminal of NOR gate y is a **0** and hence the output $\bar{Q} = 1$, as initially assumed. There is consistency throughout.

- (1b) We now repeat the process using the condition $S = R = 0$ and assuming this time that $\bar{Q} = 0$. A quick check shows that there is consistency throughout.

This means that there are two possible logic states, one with $Q = 0$ and $\bar{Q} = 1$ and the other with $Q = 1$ and $\bar{Q} = 0$. The first state is referred to as the *reset* or *clear state* and the second as the *set* or *preset state*. The reset state ($Q = 0$) is important because it usually represents the starting point of the logic operation that is to follow. However, in terms of storing and retrieving bits, the condition $S = R = 0$ is not a promising prospect because the output is independent of the input.

- (2) We now consider the situation when $S = 0$ and $R = 1$. Following the procedure used above, we find that $Q = 0$ ($\bar{Q} = 1$) is the only possibility (that is, the reset state). Consider the reset state ($S = 0$ and $R = 1$ and hence $Q = 0$). If we now go back to Figure E.1 and change R from **1** to **0**, we find that the flip-flop remains in the same (reset) state as before.
- (3) Similarly, with the opposite initial condition, that is, $S = 1$ and $R = 0$, the flip-flop will be in the set state with $Q = 1$ ($\bar{Q} = 0$) and a change in S from **1** to **0** so that $S = R = 0$ will not affect the state of the circuit.

TABLE E.1

<i>R</i>	<i>S</i>	<i>Q(t)</i>	<i>Q(t + τ)</i>	Comment
0	0	0	0	Condition in (1a)
0	1	0	1	The flip-flop goes from the <i>reset</i> to the <i>set</i> state
1	0	0	0	Condition in (2); the <i>reset</i> state is not affected
1	1	0	Not used	This can lead to an ambiguous state
0	0	1	1	This is the condition in (1b)
0	1	1	1	Condition (3); the <i>set</i> state is not affected
1	0	1	0	The flip-flop goes from the <i>set</i> to the <i>reset</i> state
1	1	1	Note used	This can lead to an ambiguous state

- (4) The final possibility is that $S = R = 1$. From Figure E.1(b) this is not possible since both outputs have to be 0s, thus violating the condition for them to be complementary to each other. The condition $S = R = 1$ is not used.

Table E.1 is a summary of the above.

The operation of the flip-flop circuit shown in Figure E.1(c) is left as an exercise to the reader.

E.3 THE CLOCKED S-R FLIP-FLOP

A flip-flop is normally a part of a much larger system of gates. Each gate takes a finite amount of time to change its state so there is no guarantee that the two inputs (*S* and *R*) of a flip-flop will arrive simultaneously or whether one will arrive before the other. For example, because of delays in their respective paths, suppose $S = 0$ and $R = 1$ and they are both supposed to change states simultaneously, then if *S* arrived a little ahead of *R*, we would momentarily have $S = 1$ and $R = 1$; an unacceptable condition which could lead to ambiguity and hence errors. It is clearly advantageous to wait for the proper values of *S* and *R* to be established before allowing the state of the flip-flop to change. This can be achieved with a third signal, namely, a *clock pulse*.

A clock pulse is used to *enable* a transition to occur in digital circuits. Because of delays (variable from gate to gate) it is necessary to enable the transition at the instant when the clock changes from enable-to-disable rather than from disable-to-enable. This allows all, the variables to reach their proper values before the clock pulse causes a transition to occur. It is therefore usual to trigger digital circuits on the negative-going edge of the clock pulse. This is described as edge-triggered logic (refer to Figure E.3).

Figure E.2(a) shows the NAND gate version of the *S-R* flip-flop with two two-input NAND gates added. We have chosen the NAND gate version because in the design of an integrated circuit one would normally choose an architecture that exploits the repetition of the same cellular building block; in this case, the NAND gate. The *S-R* flip-flop using NAND gates is shown in Figure E.1(c).

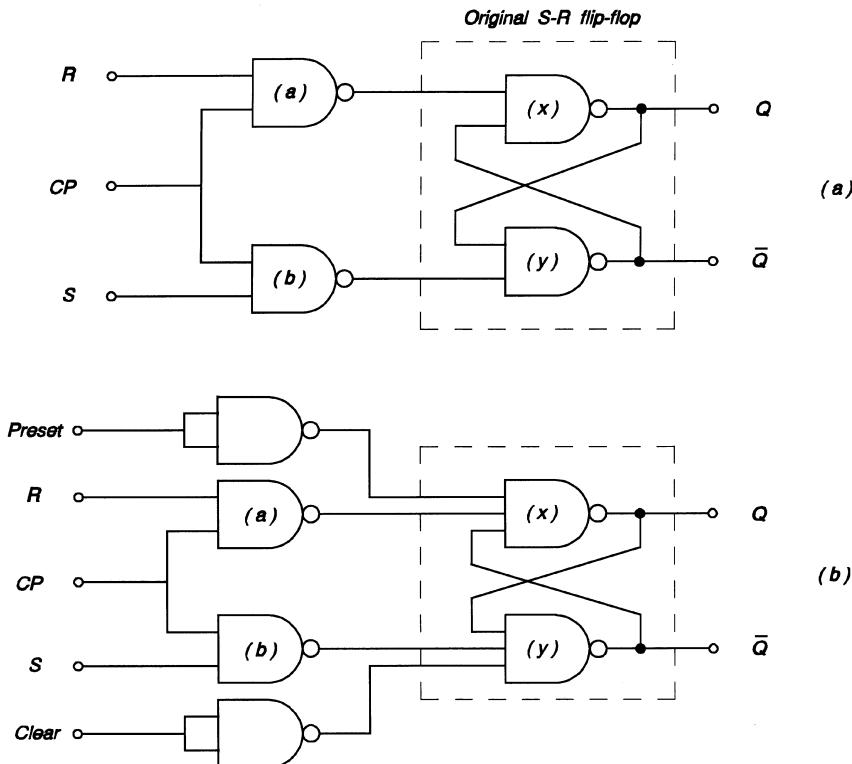


Figure E.2. (a) The NAND gate version of the *S-R* flip-flop [see Figure E.1(c)] modified to change states only when a clock pulse is supplied. (b) A further modification which provides terminals for "preset" and "clear".

In Figure E.2(a), NAND gate *a* will switch states only when both *R* and the clock pulse, *CP* are present. This will then provide an input to the (original) *S-R* flip-flop. Similarly, NAND gate *b* will switch states when both *S* and *CP* are present. It is now possible to introduce the clock pulse when the inputs *S* and *R* have had time to reach their proper binary values.

We revisit the basic characteristics of the *S-R* flip-flop by referring to the timing diagram shown in Figure E.3. It is assumed here that the negative-going edge of the clock pulse triggers the *S-R* flip-flop but whether it changes states or stays the same is determined by the values of *R* and *S* when the clock pulse arrives. The values of *S* and *R* are binary but arbitrary at the instants when the flip-flop is triggered. It can be seen that, at $t = t_0$, $S = 1$, $R = 0$ and $Q = 1$. The arrival of the clock pulse does not cause the flip-flop to change states. At $t = t_1$, $S = 0$ and $R = 0$; the flip-flop does not change states. At $t = t_2$, $S = 1$, $R = 0$; the flip-flop does not change states for the same reason as at $t = t_0$. At $t = t_3$, $S = 0$, $R = 1$, the arrival of the clock pulse causes the flip-flop to change states making $Q = 0$ and $\bar{Q} = 1$. At $t = t_4$, $S = 0$,

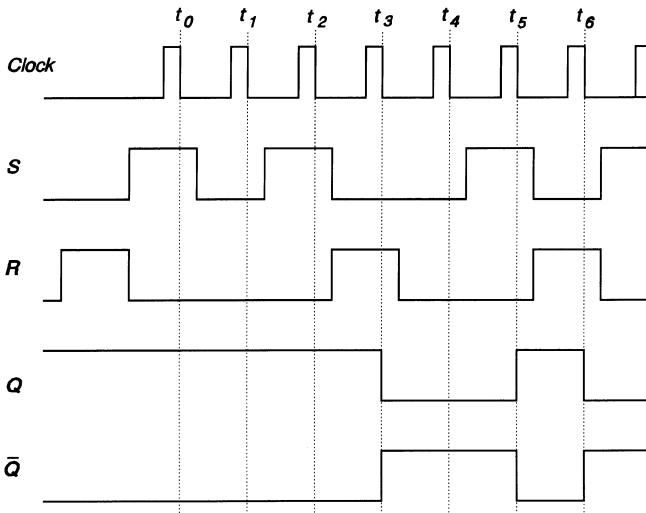


Figure E.3. The changes in the output(s) and the corresponding inputs on the arrival of a number of clock pulses.

$R = \mathbf{0}$; the flip-flop does not change states. At $t = t_5$, $S = \mathbf{1}$, $R = \mathbf{0}$; the arrival of the clock pulse causes the flip-flop to change states making $Q = \mathbf{1}$ and $\bar{Q} = \mathbf{0}$. Finally, at $t = t_6$, $S = \mathbf{0}$, $R = \mathbf{1}$; the arrival of the clock pulse causes the flip-flop to change states making $Q = \mathbf{0}$ and $\bar{Q} = \mathbf{1}$. These agrees with Table E.1.

E.4 INITIALIZATION OF THE S-R FLIP-FLOP

Before data can be stored in a flip-flop, it is necessary to ensure that it is in a known state, usually the *set* (or *clear*) state ($Q = \mathbf{0}$). Figure E.2(b) shows that NAND gates x and y in Figure E.2(a) have been changed from two-input to three-input and two NOT gates have been added. Note that when both inputs of a two-input NAND gate are connected together, we have a NOT gate.

When it is necessary to *clear* (or *preset*) the flip-flop, a $\mathbf{1}$ is applied to the appropriate terminal. Assume that the clock pulse is a $\mathbf{0}$ (it is undesirable to clear or preset the flip-flop when the clock pulse is present) so that the outputs of both NANDs a and b are $\mathbf{1}$ s. If a $\mathbf{1}$ is applied to the clear terminal ($\mathbf{0}$ to the preset terminal), the output of NAND y , namely \bar{Q} , will be a $\mathbf{1}$, regardless of the value of the signal on the third terminal. All three inputs to NAND x are $\mathbf{1}$ s and $Q = \mathbf{0}$. The flip-flop is ready (cleared) to accept data. Similarly, a $\mathbf{1}$ applied to the preset terminal will make $Q = \mathbf{1}$.

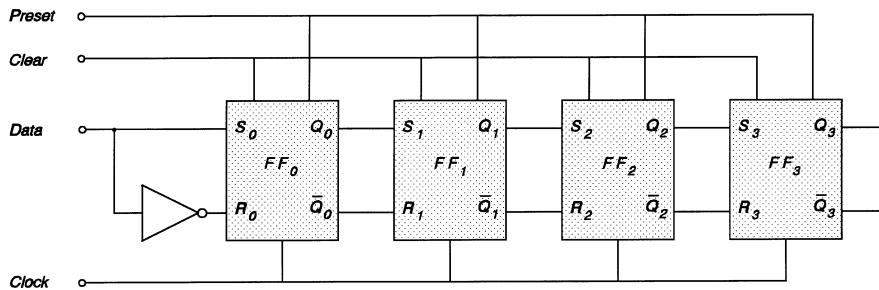


Figure E.4. The structure of a four-stage shift register formed from *S-R* flip-flops.

E.5 THE SHIFT REGISTER

A shift register constructed with *S-R* flip-flops is shown in Figure E.4. The NOT gate connected to FF_0 ensures that S_0 is the complement of R_0 . Thus FF_0 will flip into the state presented at the data terminal if it is not already in that state. The connection of Q_0 to S_1 and \bar{Q}_0 to R_1 , etc., assures complementarity at the interfaces between flip-flops and hence the acceptance of the data presented at the input of the next flip-flop.

A pulse can then be applied to the clear terminal to set $Q_0 = Q_1 = Q_2 = Q_3 = 0$. Assume that the data to be stored in the shift register has the following four bits: **1101**. The bits are presented at the data input and are clocked into the shift register

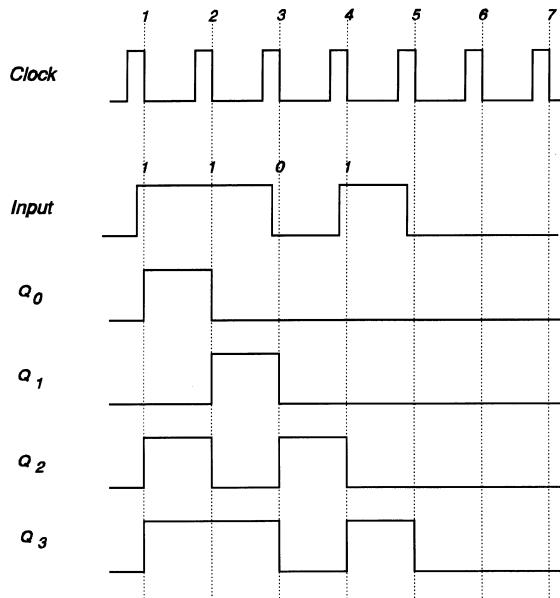


Figure E.5. The input data **1101** changes the states of the flip-flops on the arrival of each clock pulse to store the input bits.

one bit for every clock pulse. The first to be clocked in is the **1** on the right-hand side followed by the **0** and then the two **1**s. Before the first clock pulse $S_0 = \mathbf{1}$, $R_0 = \mathbf{1}$ and $Q_0 = \mathbf{0}$. So after the first clock pulse, FF_0 changes states and $Q_0 = \mathbf{1}$; all the other FF s are not affected because their inputs were **0** during the first clock pulse. The next bit presented at $S_0 = \mathbf{0}$ and $S_1 = \mathbf{1}$. So after the second clock pulse $Q_0 = \mathbf{0}$ and $Q_1 = \mathbf{1}$. Figure E.5 shows the waveforms of the clock, the input, and the outputs. After the fourth clock pulse we have $Q_0 = \mathbf{1}$, $Q_1 = \mathbf{1}$, $Q_2 = \mathbf{0}$ and $Q_3 = \mathbf{1}$. A fifth clock pulse will shift the data one position to the right, leading to the loss of the first bit.

E.6 ELECTRONIC MEMORY

A memory is required to store information in the form of bits. Each bit is stored in a flip-flop. There are various classifications of memories depending on the material (paper tape, magnetic medium or semiconductor) used to hold the data and the mode for storage and retrieval of the information.

When we store information in a memory, we say that it has been *written* into the memory. When we retrieve data stored we say that the memory has been *read*.

The data can be stored magnetically on tape, disc or drum and since the magnet retains its properties after the power is switched off, it is an example of a *non-volatile memory*. Semiconductor memories which rely on the states of a set of flip-flops lose the data stored in them when the power is switched off. They are examples of *volatile memories*.

If the information is stored on tape (magnetic or paper) it is described as *sequential access memory* because if we wish to retrieve data which is stored in a different location than where we are currently, we have to wind/rewind the tape to get there. Sequential access memories are, in general, cheap but the process of storing and retrieving information from them is slow. When the data is stored on a magnetic disc or drum, where it is possible for the read/write head to go directly to the location of the information, we describe it as *random access memory* (RAM). Accessing a RAM to read or write information into it is much faster than in a sequential access memory. Semiconductor RAMs are used extensively in computers to store large amounts of data.

Another form of memory, in which the information is only meant to be read, is the *read only memory* (ROM). The information can usually be accessed randomly but the data is placed there at the time of manufacture. A compact disc with recorded music is an example of a ROM.

E.7 RANDOM ACCESS MEMORY (RAM)

In Section E.5, we discussed a simple memory which was used to store the bits **1101**. It is advantageous to use this as a platform for the description of the semiconductor random access memory.

The information to be stored is in the form of bits (**1**s and **0**s). A group of bits form a *word* so that **1101** is described as a four-bit word. It may represent a number to be used in an arithmetic operation. Each bit has to be stored in a specific memory cell. When a word is stored, the specific location or address must be known so that the correct information can be retrieved when required.

Figure E.6 shows a single memory cell. It can be seen that there is a *read/write* terminal. When the read/write terminal has the value **1** data can be written into the cell and when it has the value **0**, the output can be read. When the address line is a **1**, the memory cell is activated and it can be written into or read. The bit to be stored is presented at the *input*. A suitable clock pulse (not shown) is used to drive the cell from the current state to the new state.

- (1) *Write*. The memory cell must be reset ($Q = 0$) before the operation can begin. To write data into the memory cell we set the address line to **1**, the read/write line also to **1**. Now if the data to be written is presented at the input as a **1** then a careful examination shows that the output of AND gate *c* is **0** ($R = 0$) and that of AND gate *d* is **1** ($S = 1$). According to Table E.1, Q goes from **0** to **1**. At the input to AND *e*, the output of NOT *a* is **0**, $Q = 1$ and the address line is **1**. Therefore the output of AND *e* is **0**.
- (2) *Read*. To read the data, the address line has to be **1** and the read/write line must be **0**. The input can be **1** or **0** but this is of no interest since the read/write line is **0** and the data cannot enter the cell. A careful examination of the circuit shows that *all* three inputs to AND *e* are **1**s and hence the output of AND *e* is **1**.

Note that when the address line is **0** (no address is specified), AND *e* is disabled and its output is **0**, independent of what happens in the rest of the circuit.

The read and write operations of the circuit in Figure E.6 when the input data is **0** are left as an exercise to the reader.

Figure E.7 shows a five-bit four-word memory. This means that the memory can store four words, each consisting of five bits. All the read/write lines are connected

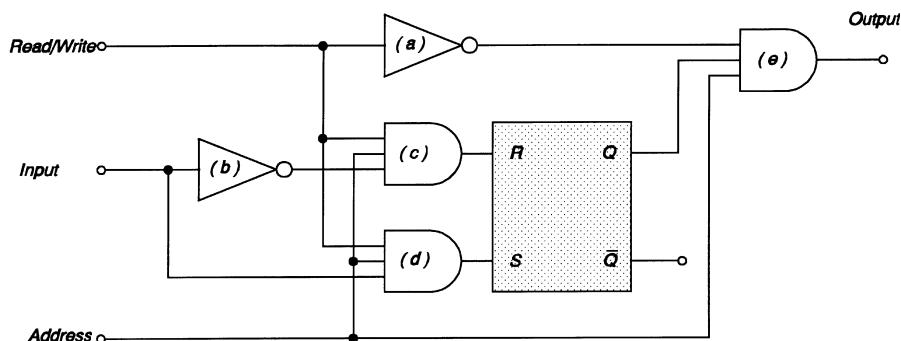


Figure E.6. A single memory cell in which the *S-R* flip-flop has been modified to provide a “read/write” and an “address” line.

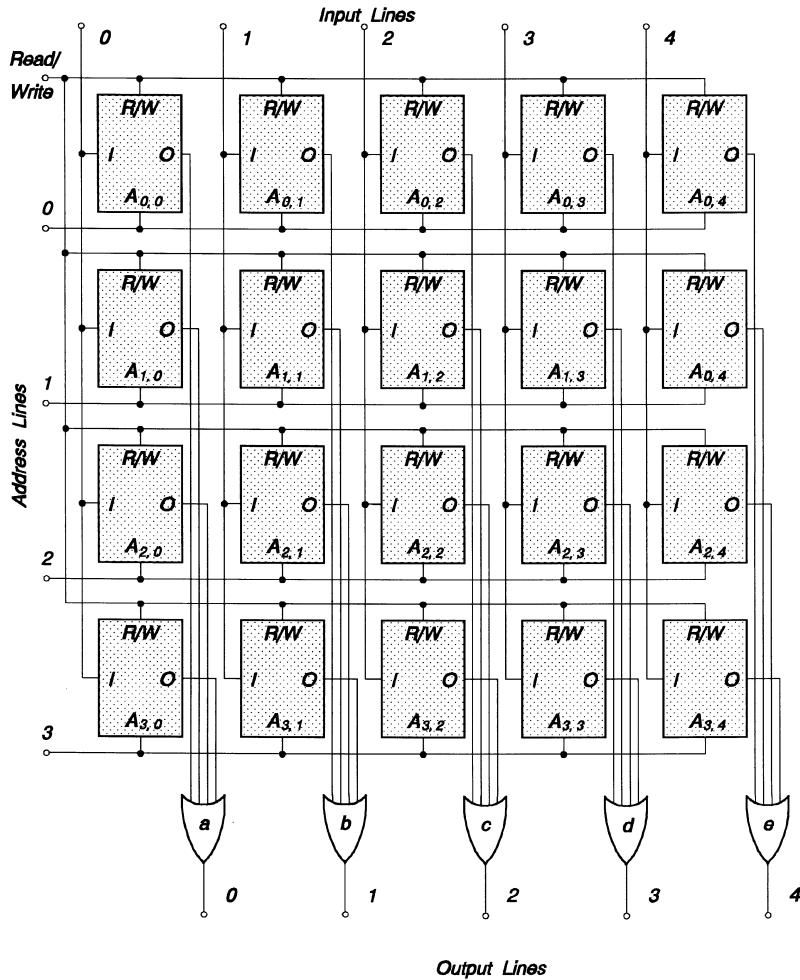


Figure E.7. A memory capable of storing four words each of which has five bits. Note that the input of the data is carried out serially but the output is presented in a *parallel* format. Circuits are available to change the format of the data from serial to parallel and *vice-versa* (see Section 8.5.1).

together and, depending on the binary signal applied, the memory can be read or data can be stored in it. The address lines of all the cells in each row are connected together so all the five bits which make up the word can be accessed together. The input lines connect all the cells in each column.

In the write mode, suppose we apply a 1 to address line 2 and data is applied to the input lines. Then it follows that the data will affect only the cells in row 2 and they will take on the values of the bits presented at the inputs. In the read mode, the read/write line is activated and if a 1 is placed on address line 3 then the information

at the outputs of all the cells in row 3 will appear on the output lines through the respective OR gates. Note that this is an example of a *parallel operation*; the retrieval of the five bits of the word happen simultaneously on separate lines, not one bit at a time. Note also that the process of reading data from the memory does not destroy the contents. The content remains until new data is written over it. This is an example of a *non-destructive* reading of the contents of a memory.

A practical memory is usually much larger than the 4×5 example shown in Figure E.7 and it is possible only in integrated circuit form. Clearly, as the size of the memory increases, the sheer number of leads that have to be brought out of the integrated circuit package becomes a very important consideration. Sophisticated digital techniques are used to reduce the number of leads required for a large memory. The interested reader is encouraged to refer to a suitable textbook on memories.

APPENDIX F

BINARY CODED DECIMAL TO SEVEN-SEGMENT DECODER

The MC5400 and MC7400 series (see Figure F.1) belong to a family of circuits for decoding binary coded decimal data to drive seven-segment liquid crystal displays (LCD) and light emitting diodes (LED). Reproduced with permission from Motorola Semiconductor Inc.

BCD-TO-SEVEN SEGMENT
DECODER/DRIVERS

MC5400/7400 series

MC5446L • MC7446L,P*
MC5447L • MC7447L,P*

Compatible with MC5400/7400 Series devices.

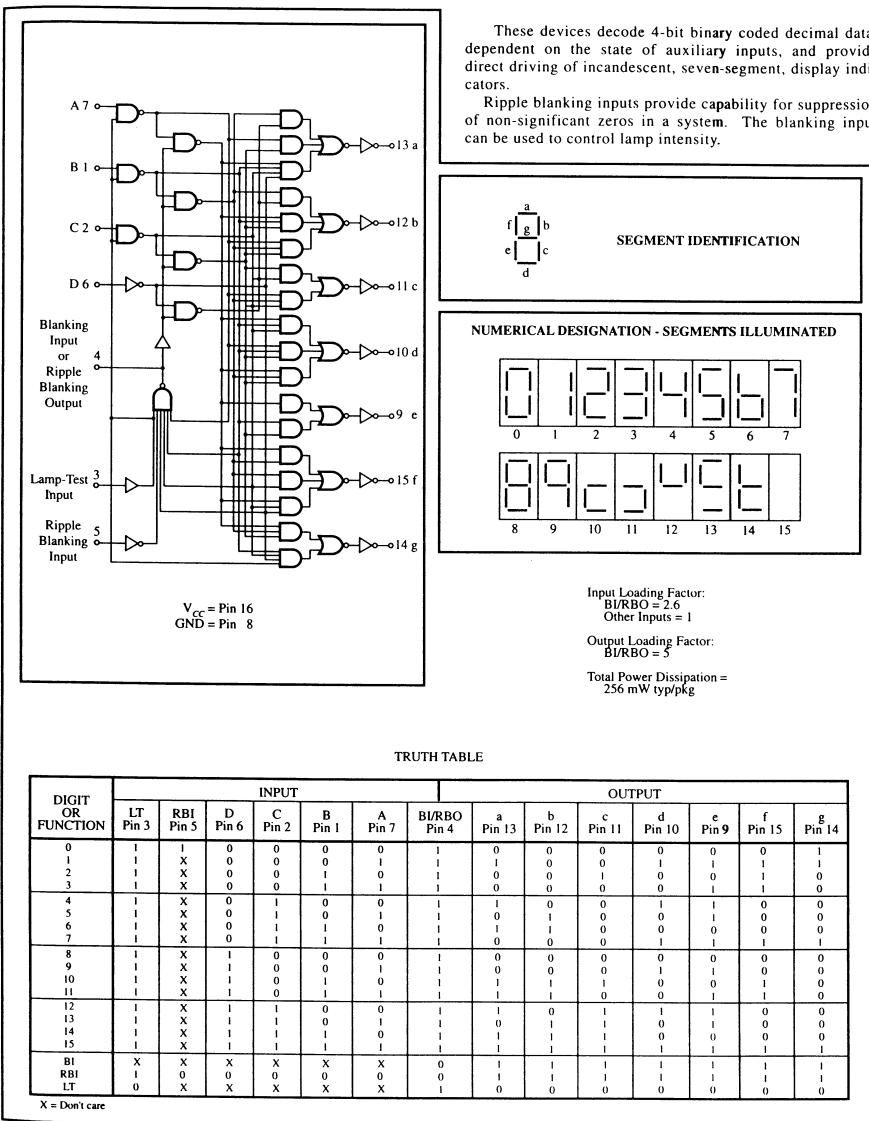


Figure F.1.

INDEX

- 4-to-2 wire hybrid 251
2-to-4 wire hybrid 244, 298
90° phase shift circuit 125, 127
- A-plane 33
A/D converter 339
Accelerator anode 165
Active LC filters 100
Active RC filters 100
Adder 126, 135
Address 412
Adjacent channel interference 334
Advanced Mobile Phone System (AMPS) 347
Alert generator 339
Alert transducer 339
Aliasing 273
All-zero state 360
Alternate mark inversion (AMI) 292
American Wire Gauge (AWG) 368
Amplifier 8
 class-A 54, 55
 class-B 54, 60
 class-C 40
 power 23, 319, 356
 radio frequency 67, 143, 146
Amplitude limiter 117, 118, 148, 291
Amplitude modulation (AM) 18, 333
Amplitude modulation theory 18
Amplitude modulator 22, 163
Analog mixer 90
Analog multiplier 85
Analog radio 381
Analog-to-digital (A/D) converter 245, 248, 358
Angle modulation 333
Anode 202
ANSnet 15
Antenna 23, 71
Apogee 384
ARPANET 11
- Aspect ratio 170, 376
Astable multivibrator 195, 280
Asymmetrical digital subscriber line (ADSL) 14, 298
Attack time 102
Audio frequency amplifier 22, 54
Auto-transformer 201
Automatic Frequency Control (AFC) 117, 122, 145
Automatic Gain Control (AGC) 83, 101
Automatic telephone exchange 8
Ayrton, William 9
- B-H characteristics 295
Back porch 182
Baffle 106
Bain, Alexander 4
Bakewell, Frederick 5
Balanced mixer 333, 359, 361
Balanced modulator 126, 132, 269, 326
Baird, John 10
Band split filter 259
Barkhausen Criterion 28, 231
Baseband signal 187, 327, 384
Basic group 269
Basic mastergroup 270
Basic supergroup 270
Bell, Alexander Graham 6
Bell System T-1 PCM Carrier 286
Bessel functions 113
Bias point 60
Bifilar winding 394
Binary coded decimal 415
Binary digits 11
Binary Phase Shift Keying (BPSK) 333
Binary quantization 317
Bipolar junction transistor (BJT) 54, 354
Bistable Multivibrator 239, 294
Boron 312

- Branly, Edouard 8
 Buffer amplifier 37, 205
 Bucket Brigade 312
- Cable modem 14
 Cable television 14
 Camera tube 163
 CAP code 338
 Carbon microphone 216
 Carbon transfer printer 308
 Carey, George 9
 Carrier, 18
 Caselli, Giovanni 6, 9
 Cathode 202
 Cathode ray tube (CRT) 10, 187, 202
 CCD image sensor 311
 CDMA Cellular Telephone System 358
 Cellular telephone 325, 347
 Central office 7, 214, 243, 253
 Central office tone receiver 259
 Central power supply 220
 Ceramic RF duplexers 346
 Cesium beam atomic clock 290
 Chain matrix 371
 Charge-coupled devices (CCD) 30
 Chat rooms 15
 Chemical valency 311
 Chip clock rate 361
 Chips 329
 Chrominance 181
 Circular orbit 386
 Circulator 226, 386
 Classical feedback theory 26
 Classical RC twin-tee filter 231
 Clock pulse 407
 Clock rectifier 292
 Clocked S-R flip-flop 407
 ‘Clonning’ 357
 Coaxial cable 11, 270, 375
 Code Division Multiple Access (CDMA) 329
 Codec 244, 286
 Coherent demodulation 327
 Coherent detection 327
 Coherer 8
 Color television subcarrier 182, 234, 299
 Color television 180
 Color triangle 180
 Colpitts 30, 231
 Common-emitter 58
 Common-source amplifier 90
 Commutator 273
 Compact disc (CD) 11, 333, 405, 411
 Compander 351
- Comparator 248
 Complementary metal oxide semiconductor (CMOS) 234, 354
 Complex frequency 25
 Composite video signal 168, 187
 Computer mouse 14
 Concentration 257
 Conductance 33
 Consultative Committee of the International Telegraph and Telephone 307
 Control channels 348
 Control grid 202
 Conversion efficiency 40, 54
 Convert enable 248
 Cooke, William 2
 Cordless telephone 325, 343
 Correlation 331
 Coupling coefficient 392
 Cross-over distortion 61
 Cross-talk 374
 Crossbar switching 257
 Crystal filters 100
 Crystal microphone 225
 Crystal-controlled oscillator 21, 22, 35, 117, 241
 Cut-off 55
- Damping 42
 Darlington pair 321
 Datalink 297
 Datapac 297
 Dataroute 297
 De-multiplexed 182
 Decelerator grid 165
 Degenerate 2-port 372
 DeForest, Lee 9
 Delay lock 361
 Demodulation 86, 103, 268
 Depletion layer 312
 Depth of modulation 19
 Depth of notch 232
 Detent 254
 Diagonal clipping 103
 Dielectric 375, 379
 Differentiator 197, 291, 293
 Digital filter 100
 Digital circuit multiplication equipment (DCME) 388
 Digital radio 384
 Digital revolution 10
 Digital telephone 243
 Digital tone dial 234
 Digital-to-analog (D/A) converter 234, 245
 DIGITONE™ dial 230

- Diode 9
- Diode modulator 47
- Dipole 72, 146
- Direct amplitude modulation 49
- Direct radiation loudspeaker 105
- Direct sequence spread spectrum technologies 329
- Direct wave 379
- Discriminator 117, 120
- Dispersion 369
- Dither technique 322
- Divide-by-N 234
- Doping 311
- Dot-matrix displays 341
- Double superheterodyne 107
- Double-sideband suppressed carrier modulator (DSB-SC) 124, 132
- Down-conversion 334, 356, 361, 381
- Dual-gate FET mixer 147
- Dual-mode cellular telephone 358
- Duplexer 343

- Earth return 7
- Earth-reflected wave 379
- Echo 308, 370, 387
- Eddy currents 395
- Edison, Thomas 7
- Edge-triggered logic 407
- Edison effect 9
- Eight-stage shift register 234
- Electret microphone 225
- Electric power grid 12
- Electromagnetic spectrum 399
- Electron beam control 191
- Electron collector 165
- Electron gun 164, 165, 202
- Electron lens 202
- Electron multiplier 165
- Electron tube amplifiers 10
- Electronic mail (e-mail) 15
- Electronic scanning 10
- Electronic serial number (ESN) 350
- Electronic sidetone control 224
- Electronic switching 261
- Electronic telephone 224
- Electrostatics 202
- Embedding 24
- Emitter follower 61, 319
- Enable 407
- Encryption 347, 357
- Envelope detector 81, 103, 122, 145, 148, 300
- Equalization 381
- ERMES (European Radio Message System) 341
- Error correction 319

- Error signal 155
- Error-detection codes 12
- ETACS (European Total Access Communication System) 357
- European Postal and Telecommunications Administrations (CEPT) 362
- Exclusive OR (XOR) 360
- Expander 351
- Extremely high frequency (EHF) 397
- Extremely low frequency (ELF) 397

- Facsimile machine 4
- Fading 347
- Farnsworth, Philo 10
- Federal Communications Commission (FCC) 15
- Feedback theory 24
- Fiber optics 11
- Film bulk acoustic resonator (FBAR) 346
- Finger wheel 223
- Fleming, Sir John Ambrose 9
- FLEX (Flexible wide-area paging protocol) 341
- Flip-flop 241, 405
- Flow diagram 262
- Focus coil 165
- Folded dipole 146
- Forward or downlink 348
- Foster-Seeley discriminator 148
- Four-phase ring counter 273
- Four-quadrant analog multiplier 51
- Four-wire repeater 374
- Fourier series 97, 113
- Fractional detuning 152
- Frame bit detector 292
- Frame bit extractor 286
- Free electrons 311
- Free space propagation 17, 71, 297, 367, 377
- Frequency changer 82, 83, 90, 147
- Frequency discriminator 145, 148
- Frequency division duplex (FDD) 343
- Frequency division multiple access (FDMA) 328, 348
- Frequency division multiplex (FDM) 20, 267, 327
- Frequency equalizers 375
- Frequency hopped multiple access (FHMA) 329
- Frequency mixer 83
- Frequency modulation (FM) 18, 111, 333, 354
- Frequency multiplier 22, 37
- Frequency-shift keying (FSK) 297, 299, 300
- Frequency-to-amplitude convertor 120, 145
- Full-duplex 326

- Gamma rays 367
- Ganged capacitor 82

- Gauss, Carl Friedrich 2
 Geostationary satellites 384
 Ground plane 73
 Ground returns 307
 Group 3 (G3) 307
 Group special mobile (GSM) 362
 GSM (Global system for mobile communications) 362
 Gyrorator 226
 Gyroscopic effect 386
- Half-wavelength dipole 72
 Half-duplex 326
 Halske, Johann 3
 Handoff 347, 350
 Handshake 308, 310
 Harmonic telegraph 6
 Hartley oscillator 30
 Henry, Joseph 8
 Hertz, Heinrich 8, 377
 High-bit-rate digital subscriber line (HDSL) 9, 14
 High-definition television (HDTV) 209
 Histogram 318
 Holes 312
 Hook switch 221, 254
 Horizontal Deflection System 200
 Horizontal trace 164
 Horn loudspeaker 105
 Horn reflector 381
 Hue 181
 Hunt 257
 Hybrid 222, 226
 Hybrid-to-hybrid feedback 374
 Hyper-abrupt junctions 116
 Hysteresis loop 295
- In-phase (I) signal 183, 334
 Iconoscope 10, 163
 Ideal transformer 392
 Image dissector 10
 Image Orthicon 165
 Inductive loading 10, 267, 368
 Infra red 312, 367
 Input impedance 26, 58
 Institute of Radio Engineers 307
 Integrated Services Digital Network (ISDN) 14, 296, 297
 Integration time 315
 Integrator 126, 133
 Intel 4004, 8080 13
 Intel Pentium III 13
 International telecommunication satellites (INTELSAT I) 387
- INTELSAT V 387
 INTELSAT VIII 388
 Inter-digit pause 224
 Interlacing 167
 Intermediate frequency (IF) 82, 98, 143, 147, 189
 International Telecommunication Union (ITU) 307
 Internet 11
 Internet browsers 14
 Internet service providers (ISP) 15
 Inversion layer 312
 Inverting input 246
 Ionosphere 107, 379
 Isolated dipole 72
- J-K flip-flop 241
 Jenkins, Charles 10
 Jumbogroup 271, 381
- Korn, Arthur 6
- Ladder network 246
 Large-scale-integration (LSI) 250
 Laser 376
 “Last mile” 14, 298
 Leakage inductance 394
 Leaky surface acoustic wave (LSAW) 346
 Least significant bit (LSB) 245
 Light emitting diode (LED) 376, 415
 Line-of-sight transmission 327, 380
 Liquid crystal display (LCD) 340
 Load line 55
 Local oscillator 82, 143, 147, 268, 290, 381
 Long wave 75
 Loop-gain 28
 Loop-gain phase shift 28
 Loudspeaker 105, 187
 Low frequency (LF) 397
 Low-noise amplifier 354, 387
 Luminance, 181
- Magnetic flux density 106, 167
 Magnetic tape 405
 Magneto 7, 220
 Magneto-motive force (MMF) 253
 Magnetostatics 202
 Majority carriers 312
 Manual central office 253, 257
 Marconi, Guglielmo 8
 Marine telegraphic cable 3
 Mark 297
 Maxwell, James 8, 377
 May, Joseph 9
 Mechanical filters 100

- Mechanical speed governor 6, 223
 Medium wave 75
 Medium wave frequency (MW) 397
 Memory 261, 308, 318, 340
 Message Switched Data Service (MSDS) 296
 Metal-oxide semiconductor field-effect transistor (MOSFET) 90
 Metal-semiconductor field-effect transistor (MESFET) 354
 Microprocessor 308
 Miller oscillator 37
 Miller-effect capacitance 177
 Minimum frequency reuse distance 349
 Minority carriers 312
 Mixer 82, 83, 90, 187, 381
 Mixer gain 96
 Mobile 331, 347
 Mobile identification number (MIN) 350
 Mobile radio 325
 Mobile Telephone Switching Office (MTSO) 348
 Modem 297, 298, 308
 Modified Huffman (MH) Code 308, 318, 401
 Modified modified READ (MMR) 308
 Modified READ (MR) 308
 Modified READ Code 318
 Modified twin-tee filter 232
 Modulation index 113
 Modulation 18, 47, 381
 Molniya 386
 Monochromatic 181
 Monopole 72
 Morse, Samuel 2
 Mosaic 164
 Most significant bit (MSB) 245
 Moving-iron telephone receiver 218
 MPEG-2 209
 Multipath fading 331
 Multiple-hop signal 380
 Multiplex system 181
 Multiplier sampling gate 279
 Mutual coupling coefficient 150, 391
 Mutual inductance 391
- N-coder 234
 N-channel JFET 284
 N-type metal-oxide semiconductor (NMOS) 354
 N-type negative conductance 25
 N-type silicon 312
 NAND gate 234, 237, 407
 Narrow band frequency modulation (NBFM) 124
 Negative conductance oscillator 24, 115
 Negative feedback 27
 Negative resistance 24
- Negative-impedance converter 369, 370
 Newsgroups 15
 Nipkow, Paul 10
 Nipkow disc 10
 Non-coherent demodulation 327
 Non-volatile memory 411
 Non-destructive reading of memory 414
 NOR gate 234, 237
 NOT-gate 234, 236
 Notch frequency 232
 NSFnet 15
 Nyquist sampling rate 271
- Operating point 55
 Operational amplifier 130
 Operational amplifier sampling gate 279
 Optical fiber 367, 376
 Oscillation 28
 Output impedance 26
- P-type silicon 312
 Packet switching 12, 297
 Paging system 325, 338
 Pantelegraph 6
 Parabolic dish antenna 381
 Parallel operation 414
 Parallel transmission 161
 Parallel-to-serial (P/S) converter 245, 250
 Passive LC filters 99
 Pawl 254
 Pel 305
 Pendulum 4, 305
 Pentodes 25, 115
 Perigee 384
 Perry, John 9
 Phase angle 305
 Phase equalization 369, 375
 Phase modulation (PM) 18, 333
 Phase shifter 291
 Phase-locked loop 290
 Phase-locked loop FM detector 155
 Phasing 5
 Phosphor 167
 Phosphorus 311
 Photo-transistors 312
 Photocathode 165
 Photoconductivity 167
 Photogate 315
 Photomultipliers 10
 Photon 312
 Photosensitive cells 10
 Phototube 161
 Picture tube 202

- Pierce oscillator 37
 Piezoelectric substrate 346
 Pilot signal 138, 268
 Pixel 305
 Polarized light 340
 Poles 174
 Popov, Aleksandr 8
 Portable 347
 Positive feedback 27
 Post Office Code Standardization Advisory Group (POCSAG) 338
 Pre-amplifier/equalizer 292
 Pre-picture tube matrixing. 203
 Printing telegraph 3
 Propagation delay 376
 Pseudo-noise (PN) generator 358, 359
 Pseudodigital modulation 273
 Pseudo-random sequence 329
 Public Switched Telephone Network (PSTN) 343
 Pulse code modulation encoder/multiplexer 286
 Pulse Regenerator 294
 Pulse shaping filter 334
 Pulse-amplitude modulation (PAM) 273, 279
 Pulse-code modulation (PCM) 244, 273, 286
 Pulse-position modulation (PPM) 273
 Pulse-width modulation (PWM) 273
 "Push-pull" 54
- Quadrature detector 154
 Quadrature multiplex 182
 Quadrature phase shift keying (QPSK) 334
 Quadrature-phase (Q-) signal 183, 334
 Quartz crystal 21, 22
 Quiescent 55, 59
- R-2R ladder 247
 Radio frequency choke (RFC) 148
 Radio frequency duplexer 343
 Ramp counter A/D converter 248
 Random access memory (RAM) 411
 Ratio detector 153
 RC twin-tee circuit 231
 Read memory 411
 Read only memory (ROM) 411
 Real power 32
 Recording telegraph 4
 Rectifier 230, 259
 Refraction 379
 Refractive index 377
 Regenerative repeater 288, 291, 369, 376, 381, 384
 Register 261
 Relay 2
 Release time 102
- Reset 241
 Reset or clear state 406
 Resetable binary counter 234, 248
 Resetable bistable multivibrator 234
 Resettable flip-flop 318
 Resistor transistor logic (RTL) 234
 Resonance 22
 Reverse or uplink 348
 Reverse-emphasis filter 356
 Ribbon microphone 225
 Right-hand rule 167
 Ring counter 359
 Rotary dial 223
- S-plane 174
 S-R flip-flop 241, 405
 S-type negative resistance 25
 Sample-and-hold (S/H) 280, 283
 Sampling 273
 - series gate 277
 - series-shunt gate 278
 - shunt gate 278
 Sampling theorem 271
 Satellite communication 11, 271, 367, 384
 Saturation 55, 181
 Scanner 162, 167, 305
 Schmitt trigger 280, 300, 308
 Search engines 14
 Siemens, Werner 3
 Selenium cell 6
 Semafore 1
 Sensors 312
 Sequential access memory 411
 Serial transmission 162
 Serial-to-parallel converter 251, 245
 Set or preset state 241, 406
 Seven-segment display 341, 415
 Shadow mask 207
 Shift register 318, 410
 Short dipole 72
 Short wave frequency (SW) 76, 397
 Short-circuit admittance 31
 Sideband 20, 98
 Sidetone 222, 229
 Siemens Telecommunication Company 3
 Signal plate 164
 Silicon-controlled rectifier 25
 Simplex 325
 Single sideband (SSB) 179, 268
 Single-hop distance 380
 Single-sideband-suppressed carrier (SSB-SC) 326
 Sinusoidal oscillator 28
 Sky-reflected wave 379

- Sliding correlator 361
- Space-division multiplex (SDM) 261
- Speech compressor 351
- Speed-up capacitor 320
- Spiral scanning 6
- Spread spectrum techniques 329, 347
- Square-law modulator 47
- Stagger tuning 146
- Station class mark (SCM) 350
- Step-by-step switching 253
- Stepper motor 308, 315
- Stereophonic FM signal 137, 156
- Stray capacitance 394
- Strowger, Almon B. 8
- Strowger switch 256
- Submarine cable 376, 387
- Sun spot cycle 380
- Super High Frequency (SHF) 397
- Superheterodyne 81, 82, 143, 187
- Supervisory audio tone generator (SAT) 354
- Surface acoustic wave (SAW) 100, 346
- Susceptance 33
- Switched capacitor filters 101
- Switching-type mixer 96
- Symmetrical clipper 119
- Sync-pulse separator 193
- Synchronism 5, 273, 305, 329
- Synchronous demodulation 158
- Synchronous detection 205
- Synchronous tuning 146

- T-equivalent model 373
- T1 PCM carrier 14
- Tank circuit 40
- Target 165
- TAT-6 (SG) trans-Atlantic cable 376
- TAT-8 (SG) trans-Atlantic cable 377
- Telecom Canada Digital Network 290
- Teletypewriter exchange (TWX) 296
- Television front end or tuner 189
- Terrestrial microwave network 11, 271, 380
- Thermal printer 308, 321
- Thevenin's Theorem 395
- Thin-film inter-digital electrodes 346
- Three-phase clock 315
- Threshold voltage 312
- Time division multiple access (TDMA) 328
- Time division multiplex (TDM) 261, 271
- Timing information 273
- Toll line 267
- Tone dial 224, 230
- Tone Ringer 224, 229
- TOUCH-TONETM 225

- Touch-tone 338
- Traffic channels 348
- Transadmittance 228
- Trans-Atlantic cable 3
- Transconductance 227
- Transfer gate 315
- Transistor-transistor logic (TTL) 234
- Transponder 386
- Transverse electric (TE) 376
- Transverse magnetic (TM) 376
- Triple-mode cellular telephone 362
- Troposphere-reflected wave 379
- Tropospheric scatter propagation 379
- Trunks 267
- Tuned circuit 40
- Tuning fork 305
- Tunnel diode 25, 115
- Turns ratio 393
- Turnstile array 180
- Tweeter 106
- Twisted nematic 340
- Twisted pair cable 270, 368
- Two-port 31

- Ultra high frequency (UHF) 397
- Ultra violet 367
- UNESCO 1
- Uni-junction transistors 25
- Universal asynchronous receiver/transmitter (UART) 250
- UNIX 14
- Up-converter 354, 381
- Upper sideband 326
- USDC (United States Digital Cellular) 362

- Vail, Alfred 3
- Varactor diode 116
- Variable gain amplifier 101
- Variable transconductance 51
- Vertical deflection oscillator 195
- Vertical sync-separator 193
- Vertical trace 165
- Very high frequency (VHF) 397
- Very low frequency (VLF) 397
- Vestigial sideband 163, 179, 190
- Video 161
 - amplifier 163, 170, 191
 - image 161
- Vidicon 167
- Vinyl discs 405
- Virtual ground 246
- Visual telegraphic system 1
- Voice coil 106

- Voice frequency (VF) 397
Volatile memories 411
Voltage-controlled capacitor 116, 122
Voltage-controlled crystal oscillator (VCXO) 290
Voltage-controlled current source 227
Voltage-controlled oscillator (VCO) 155

Waveform clipping 60
Waveguide 376
Wax cylinders 405
Weber, Wilhelm 2
Western Union Telegraph Company 3
Wheatstone, Charles 2
Winding resistance 69, 87, 394
Wireless telegraphy 8

Wireline telephone 338
Woofer 106
Word 412
World Wide Web 13
Write into memory 411

xDSL 298
Xerographic printer 308

Yagi-Uda antenna 146
Yoke coil 199

Zener diode 248, 293
Zener diodes 120
Zworykin, Vladimir 10, 163