Radio Receiver Design

Robert C. Dixon

MARCEL DEKKER, INC.

Radio Receiver Design

ELECTRICAL ENGINEERING AND ELECTRONICS

A Series of Reference Books and Textbooks

1. Rational Fault Analysis, *edited by Richard Saeks and S. R. Liberty*
2. Nonparametric Methods in Communications, *edited by P. Papantoni-Kazakos and Dimitri Kazakos*
3. Interactive Pattern Recognition, *Yi-tzuu Chien*
4. Solid-State Electronics, *Lawrence E. Murr*
5. Electronic, Magnetic, and Thermal Properties of Solid Materials, *Klaus Schröder*
6. Magnetic-Bubble Memory Technology, *Hsu Chang*
7. Transformer and Inductor Design Handbook, *Colonel Wm. T. McLyman*
8. Electromagnetics: Classical and Modern Theory and Applications, *Samuel Seely and Alexander D. Poularikas*
9. One-Dimensional Digital Signal Processing, *Chi-Tsong Chen*
10. Interconnected Dynamical Systems, *Raymond A. DeCarlo and Richard Saeks*
11. Modern Digital Control Systems, *Raymond G. Jacquot*

12. Hybrid Circuit Design and Manufacture, *Roydn D. Jones*
13. Magnetic Core Selection for Transformers and Inductors: A User's Guide to Practice and Specification, *Colonel Wm. T. McLyman*
14. Static and Rotating Electromagnetic Devices, *Richard H. Engelmann*
15. Energy-Efficient Electric Motors: Selection and Application, *John C. Andreas*
16. Electromagnetic Compossibility, *Heinz M. Schlicke*
17. Electronics: Models, Analysis, and Systems, *James G. Gottling*
18. Digital Filter Design Handbook, *Fred J. Taylor*
19. Multivariable Control: An Introduction, *P. K. Sinha*
20. Flexible Circuits: Design and Applications, *Steve Gurley, with contributions by Carl A. Edstrom, Jr., Ray D. Greenway, and William P. Kelly*
21. Circuit Interruption: Theory and Techniques, *Thomas E. Browne, Jr.*
22. Switch Mode Power Conversion: Basic Theory and Design, *K. Kit Sum*
23. Pattern Recognition: Applications to Large Data-Set Problems, *Sing-Tze Bow*
24. Custom-Specific Integrated Circuits: Design and Fabrication, *Stanley L. Hurst*
25. Digital Circuits: Logic and Design, *Ronald C. Emery*
26. Large-Scale Control Systems: Theories and Techniques, *Magdi S. Mahmoud, Mohamed F. Hassan, and Mohamed G. Darwish*
27. Microprocessor Software Project Management, *Eli T. Fathi and Cedric V. W. Armstrong (Sponsored by Ontario Centre for Microelectronics)*
28. Low Frequency Electromagnetic Design, *Michael P. Perry*
29. Multidimensional Systems: Techniques and Applications, *edited by Spyros G. Tzafestas*
30. AC Motors for High-Performance Applications: Analysis and Control, *Sakae Yamamura*
31. Ceramic Motors for Electronics: Processing, Properties, and Applications, *edited by Relva C. Buchanan*
32. Microcomputer Bus Structures and Bus Interface Design, *Arthur L. Dexter*
33. End User's Guide to Innovative Flexible Circuit Packaging, *Jay J. Miniet*
34. Reliability Engineering for Electronic Design, *Norman B. Fuqua*
35. Design Fundamentals for Low-Voltage Distribution and Control, *Frank W. Kussy and Jack L. Warren*
36. Encapsulation of Electronic Devices and Components, *Edward R. Salmon*
37. Protective Relaying: Principles and Applications, *J. Lewis Blackburn*
38. Testing Active and Passive Electronic Components, *Richard F. Powell*
39. Adaptive Control Systems: Techniques and Applications, *V. V. Chalam*
40. Computer-Aided Analysis of Power Electronic Systems, *Venkatachari Rajagopalan*

41. Integrated Circuit Quality and Reliability, *Eugene R. Hnatek*
42. Systolic Signal Processing Systems, *edited by Earl E. Swartzlander, Jr.*
43. Adaptive Digital Filters and Signal Analysis, *Maurice G. Bellanger*
44. Electronic Ceramics: Properties, Configuration, and Applications, *edited by Lionel M. Levinson*
45. Computer Systems Engineering Management, *Robert S. Alford*
46. Systems Modeling and Computer Simulation, *edited by Naim A. Kheir*
47. Rigid-Flex Printed Wiring Design for Production Readiness, *Walter S. Rigling*
48. Analog Methods for Computer-Aided Circuit Analysis and Diagnosis, *edited by Takao Ozawa*
49. Transformer and Inductor Design Handbook: Second Edition, Revised and Expanded, *Colonel Wm. T. McLyman*
50. Power System Grounding and Transients: An Introduction, *A. P. Sakis Meliopoulos*
51. Signal Processing Handbook, *edited by C. H. Chen*
52. Electronic Product Design for Automated Manufacturing, *H. Richard Stillwell*
53. Dynamic Models and Discrete Event Simulation, *William Delaney and Erminia Vaccari*
54. FET Technology and Application: An Introduction, *Edwin S. Oxner*
55. Digital Speech Processing, Synthesis, and Recognition, *Sadaoki Furui*
56. VLSI RISC Architecture and Organization, *Stephen B. Furber*
57. Surface Mount and Related Technologies, *Gerald Ginsberg*
58. Uninterruptible Power Supplies: Power Conditioners for Critical Equipment, *David C. Griffith*
59. Polyphase Induction Motors: Analysis, Design, and Application, *Paul L. Cochran*
60. Battery Technology Handbook, *edited by H. A. Kiehne*
61. Network Modeling, Simulation, and Analysis, *edited by Ricardo F. Garzia and Mario R. Garzia*
62. Linear Circuits, Systems, and Signal Processing: Advanced Theory and Applications, *edited by Nobuo Nagai*
63. High-Voltage Engineering: Theory and Practice, *edited by M. Khalifa*
64. Large-Scale Systems Control and Decision Making, *edited by Hiroyuki Tamura and Tsuneo Yoshikawa*
65. Industrial Power Distribution and Illuminating Systems, *Kao Chen*
66. Distributed Computer Control for Industrial Automation, *Dobrivoje Popovic and Vijay P. Bhatkar*
67. Computer-Aided Analysis of Active Circuits, *Adrian Ioinovici*
68. Designing with Analog Switches, *Steve Moore*
69. Contamination Effects on Electronic Products, *Carl J. Tautscher*
70. Computer-Operated Systems Control, *Magdi S. Mahmoud*
71. Integrated Microwave Circuits, *edited by Yoshihiro Konishi*
72. Ceramic Materials for Electronics: Processing, Properties, and Applications, Second Edition, Revised and Expanded, *edited by Relva C. Buchanan*

73. Electromagnetic Compatibility: Principles and Applications, *David A. Weston*
74. Intelligent Robotic Systems, *edited by Spyros G. Tzafestas*
75. Switching Phenomena in High-Voltage Circuit Breakers, *edited by Kunio Nakanishi*
76. Advances in Speech Signal Processing, *edited by Sadaoki Furui and M. Mohan Sondhi*
77. Pattern Recognition and Image Preprocessing, *Sing-Tze Bow*
78. Energy-Efficient Electric Motors: Selection and Application, Second Edition, *John C. Andreas*
79. Stochastic Large-Scale Engineering Systems, *edited by Spyros G. Tzafestas and Keigo Watanabe*
80. Two-Dimensional Digital Filters, *Wu-Sheng Lu and Andreas Antoniou*
81. Computer-Aided Analysis and Design of Switch-Mode Power Supplies, *Yim-Shu Lee*
82. Placement and Routing of Electronic Modules, *edited by Michael Pecht*
83. Applied Control: Current Trends and Modern Methodologies, *edited by Spyros G. Tzafestas*
84. Algorithms for Computer-Aided Design of Multivariable Control Systems, *Stanoje Bingulac and Hugh F. VanLandingham*
85. Symmetrical Components for Power Systems Engineering, *J. Lewis Blackburn*
86. Advanced Digital Signal Processing: Theory and Applications, *Glenn Zelniker and Fred J. Taylor*
87. Neural Networks and Simulation Methods, *Jian-Kang Wu*
88. Power Distribution Engineering: Fundamentals and Applications, *James J. Burke*
89. Modern Digital Control Systems: Second Edition, *Raymond G. Jacquot*
90. Adaptive IIR Filtering in Signal Processing and Control, *Phillip A. Regalia*
91. Integrated Circuit Quality and Reliability: Second Edition, Revised and Expanded, *Eugene R. Hnatek*
92. Handbook of Electric Motors, *edited by Richard H. Engelmann and William H. Middendorf*
93. Power-Switching Converters, *Simon S. Ang*
94. Systems Modeling and Computer Simulation: Second Edition, *Naim A. Kheir*
95. EMI Filter Design, *Richard Lee Ozenbaugh*
96. Power Hybrid Circuit Design and Manufacture, *Haim Taraseiskey*
97. Robust Control System Design: Advanced State Space Techniques, *Chia-Chi Tsui*
98. Spatial Electric Load Forecasting, *H. Lee Willis*
99. Permanent Magnet Motor Technology: Design and Applications, *Jacek F. Gieras and Mitchell Wing*
100. High Voltage Circuit Breakers: Design and Applications, *Ruben D. Garzon*

101. Integrating Electrical Heating Elements in Appliance Design, *Thor Hegbom*
102. Magnetic Core Selection for Transformers and Inductors: A User's Guide to Practice and Specification, Second Edition, *Colonel Wm. T. McLyman*
103. Statistical Methods in Control and Signal Processing, *edited by Tohru Katayama and Sueo Sugimoto*
104. Radio Receiver Design, *Robert C. Dixon*

Additional Volumes in Preparation

Radio Receiver Design

Robert C. Dixon

Omnipoint Corporation
Palmer Lake, Colorado

MARCEL DEKKER, INC. NEW YORK • BASEL • HONG KONG

Library of Congress Cataloging-in-Publication Data

Dixon, Robert C. (Robert Clyde), 1932–
Radio receiver design / Robert C. Dixon.
p. cm. -- (Electrical engineering and electronics : 104)
Includes bibliographical references and index.
ISBN 0–8247–0161–5 (alk. paper)
1. Radio -- Receivers and reception -- Design and construction.
I. Title. II. Series.
TK6563. D615 1998
621.384 ' 18 -- dc21 98–29981
CIP

This book is printed on acid-free paper.

Headquarters
Marcel Dekker, Inc.
270 Madison Avenue, New York, NY 10016
tel: 212-696-9000; fax: 212-685-4540

Eastern Hemisphere Distribution
Marcel Dekker AG
Hutgasse 4, Postfach 812, CH-4001 Basel, Switzerland
tel: 44-61-261-8482; fax: 44-61-261-8896

World Wide Web
http://www.dekker.com

The publisher offers discounts on this book when ordered in bulk quantities. For more information, write to Special Sales/Professional Marketing at the headquarters address above.

Current printing (last digit):
10 9 8 7 6 5 4 3 2 1

PRINTED IN THE UNITED STATES OF AMERICA

For Eugene P. Hoyt,
and as always,
for Nancy

This page intentionally left blank

Preface

Radio Receiver Design was prompted by the need for a text that takes into account the advances in technology that have been made in the communications field over the past few years, and considers their application to receiver design.

This book is intended for those who need an introduction to communications and the role of receivers in conveying information, both analog and digital; especially those who have been trained during a time when the world was so entranced with the promise of computers that it was forgotten that they would soon need ways to communicate from computer to computer, and that the communications needed would not always be practical or desirable over telephone lines. It is also intended for those working engineers who need to be brought up to speed on the state of current receiver design and what is practical in today's world.

Hundreds of working graduate engineers have been taught from the material found in this book; both those whose background is in computer science and those whose studies were completed when the world was analog have found it useful and up-to-date. The book should also be useful as an undergraduate text for communications systems.

Today's receivers typically employ the same architecture that has been employed in receivers for many years. That is, today's receivers are still superheterodyne in structure, but they employ very different components than receivers of just a few years ago. Even the most common components—resistors—are different, in that at present one-eighth-Watt resistors are commonly used, while half-Watt resistors were the standard ten or more years ago. Integrated circuits, employing silicon and gallium arsenide, are achieving higher performance in smaller packages, with less current consumption.

Technology has made the difference, and will continue to do so, even as receivers become implemented using more digital structure. As you read this, many engineers are engaged in development of components that will permit conversion of signals at higher and higher frequencies to formats that can be processed digitally for recovery of the information that they convey. For as long as I can remember, engineers have had the dream of being able to digitize at the output of an antenna and process the digital result with a computer, thereby doing away with all those "troublesome" analog circuits and components. That dream is rapidly becoming reality, but not without certain drawbacks. The greatest advantage of digital processing is flexibility, not simplicity or even elegance, and many of the things that are done with digital processing are done with simple iteration of brute force techniques. However, if the end result is a better filter (for example) and it can be implemented in less space, by a microprocessor in its "spare" time, then it must be used, and will always win out in the end.

This shift is already occurring. Cellular telephones that convert once to an IF and whose second conversion is through sampling are already in production, and when they don't work, they are simply thrown away because it is less expensive to do so than to build another one. GPS (global positioning system) receivers already sell for less than $100, primarily because of the degree of integration that has been achieved in these receivers, in both the analog and the digital circuits that are used. (GPS is, incidentally, the only spread spectrum receiver that must use a microprocessor. The others can function without one unless they are needed to perform calculations similar to those for which GPS receivers use their microprocessor.)

In a receiver, a microprocessor is simply a convenience, but its forte is flexibility. (Nevertheless, the death knell for many a project has been "all we have to do is modify the software.")

Marvelous results are being achieved today, and better results will be achieved tomorrow, in communications systems of all kinds. The primary reason is that modulation techniques are getting better, and receivers are keeping pace. They will continue to do so because the level of available technology continues to improve, and no end is in sight.

No matter what the type of modulation, or the medium, or the technology, we will still need something that functions as a receiver.

Robert C. Dixon

Contents

1

Introduction: Communication Links and the Receiver's Role

A receiver's job is to gather energy from the medium in which it exists and convert the energy to a form in which it may be assimilated by the receiver's user. Functionally, the receiver must perform a number of tasks to accomplish this miracle. These tasks include at least:

1. Transducer (antennas, etc.) matching
2. Selection of desired signals
3. Rejection of undesired signals
4. Amplification by very large factors
5. Demodulation
6. Error detection and/or correction
7. Received information conditioning and output

These receiver requirements are considered in detail in the following chapters.

A communication link consists of an information source, a signal conditioner, a transmitter, a transducer to drive the medium, the medium itself, a transducer to convert the energy from the form in which it exists in the medium to that which is useful to a receiver, and a receiver whose output goes to an information sink. Although the receiver may seem to be insignificant when an overall communication link is described in so many words, its job is arduous. Figure 1.1 illustrates the overall communication link and its components.

The communication link exists for the purpose of conveying information, with the transmitter serving to transform the information to a state in which it may be coupled into the propagation medium, and the coupling performed by a transducer (see Table 1.1). The receiver then performs the functions necessary to transform the signal, using a transducer at its input that transforms the energy in the medium to an electrical form, processing it, and often using another transducer at its output.

In a radio system, for example, information is input to the transmitter that

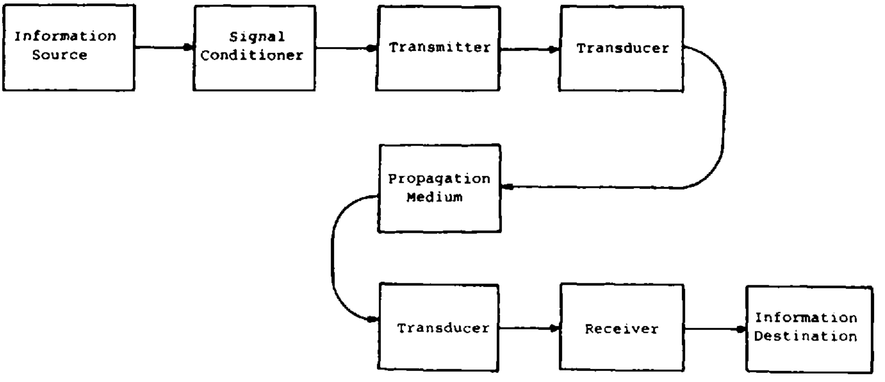


Figure 1.1 Illustration of a communications link.

modulates a "carrier" signal. This modulated carrier may then be coupled to the medium by an antenna (its transducer). Alternatively, the signal may be sent through coaxial cable—which may not require a transducer—or some other medium such as fiber optics, which would require converting the electrical signal to a light wave.

Propagation loss (Table 1.2) in wireless systems is primarily a function of distance and frequency. This loss may be calculated from the equation

$$\text{Loss} = 20 \log \left(\frac{4\pi D}{\lambda} \right) \text{ (in meters)}$$

where D is the distance of the path traveled, in meters,* λ is the wavelength of the carrier, in meters,* and $\lambda \approx 300/f$ (MHz) in meters. This loss is called the "free space" loss, because it is the loss that occurs for a straight line path without obstructions, interference, opacity, or multiple paths.

Table 1.1 Typical Transducers and Their Media

Transducer	Medium	Type of link
Antenna (Tx or Rx)	Electromagnetic field	Radio
IR diode (Tx or Rx)	Infrared rays	Infrared
Laser (Tx)	Light ray or optical fiber	Lightwave
Light-sensitive diode (Rx)	Light ray or optical fiber	Lightwave
Acoustic coupler (Tx or Rx)	Fluid	Sonar
Acoustic coupler (Tx or Rx)	Gas	Audio

*Or any other consistent measure.

Table 1.2 Propagation Loss in Transmission Media

Medium	Typical loss
Electromagnetic (wireless)	Moderate, depends on frequency and distance. 90 dB at 400 MHz and 1 mile.
Coaxial cable	High, depends on frequency and distance. 30 dB per 100 feet at 400 MHz.
Fiber optic	Low. Depends primiarily on distance.

An alternative method of calculating free space loss is through use of the equation

$$\text{Loss} = 36.6 + 20 \log D + 20 \log F$$

where D is distance in miles* and F is frequency in MHz. Either equation is valid.

The two most important characteristics of a receiver are unquestionably

1. *Sensitivity*, which is the minimum signal level that a receiver requires to produce an acceptable output signal-to-noise ratio when the input signal is modulated by a standard amount (for voice systems, "standard" modulation is typically 30 percent by a 1-kHz sine wave).
2. *Selectivity*, which is the ability of a receiver to reject signals outside a given band while accepting signals that are within that band.

There are other important properties to be sure, but a receiver with poor sensitivity or selectivity is usually of little value.

Sensitivity. Several measures of receiver *sensitivity* are commonly used. Some of these are $(S + N)/N$ sensitivity, SINAD sensitivity, "soft" and "hard" microvolts sensitivity, tangential sensitivity, and MDS sensitivity. Receivers for analog information and digital information are also differentiated by defining their output quality in terms of $(S + N)/N$ ratio for analog reception and BER for digital reception. Procedures for measurement are discussed in Chapter 10.

$(S + N)/N$ sensitivity: Receiver sensitivity on the basis of the signal-plus-noise/noise ratio is commonly employed as a measure of a receiver's reception quality. An $(S + N)/N$ ratio of 10 dB is usually considered to be the minimum that is acceptable.

*Some convenient conversion factors are:

1 mile (statute)	= 1609.344 meters
1 kilometer	= 0.621 miles (statute)
1 foot	= 0.3048 meters
1 yard	= 0.9144 meters

SINAD sensitivity: This is specified in terms of a signal-plus-noise-plus-distortion/noise and is most often considered to require 12 dB or more as the minimum acceptable ratio.

Soft versus hard microvolts: In military applications, it is not unusual to specify a sensitivity measurement that incorporates a 6 dB attenuator between the signal source and the receiver being tested. This, of course, increases the signal that must be output from the signal generator and causes the receiver sensitivity to appear to be worse. On specification sheets, "5 microvolts (*h*ard)" sensitivity is understood to mean that the measurement is made with a 6 dB attenuator in place and that the receiver should operate at 2.5 μV when the 6-dB attenuator is not used.

Bit error rate (BER) sensitivity: Receivers intended for use in reception of data usually employ bit error rate instead of signal-to-noise ratio, as a measure of performance. The level required is a function of the specific receiver application, but typical requirements are in the range 1×10^{-3} to 1×10^{-6} for uncorrected errors.

Tangential sensitivity: Tangential sensitivity is a measure of performance usually applied to receivers used to detect only that a signal is present (such as a radar or a radio astronomy receiver) in the presence of noise. Measurement of tangential sensitivity is considered to correspond to an 8-dB carrier-to-noise ratio.

Minimum discernible signal (MDS) sensitivity: Also (like tangential sensitivity) used primarily to detect whether signals are present. The level of carrier-to-noise ratio is considered to be 3 dB for MDS sensitivity.

Selectivity. A receiver operates in an electromagnetic environment that includes many signals that may be separated from one another in several ways:

Distance. This is manifest as a difference in signal level at a receiver, where the signal varies at least as the inverse square of the distance. Selectivity is limited by the receiver's dynamic range and the filter characteristics in its radio frequency (RF) and intermediate frequency (IF) stages.

Time of transmission. Selectivity defined by using time of transmission is limited by the accuracy with which a receiver can synchronize its timing to a transmitter or group of transmitters.

Frequency. This is the classic way of defining receiver selectivity. Signals are selected through the use of bandpass filters tuned to the frequency of a desired signal. Such filters are highly evolved but are limited by the Q (quality factor) of the components used in their construction.

Modulation. Signals may be distinguished from one another by their modulation. For example, FM and AM receivers are designed to be insensitive to the wrong kind of modulation.

1.1 MEDIA LOSS AND LINK BUDGETS

Loss in the signal transfer medium is the greatest loss seen in most communication links. It is not unusual to lose signal power to the extent of more than 100 dB (a factor of 10 billion) in everyday systems. Typical losses are as shown in Table 1.3. These losses are highly dependent on the frequency of operation and the distance between the transmitter and receiver, as well as the character of the medium.

A "link budget" consists of an analysis of the assets available (transmitter power, receiver sensitivity, etc.) and the use of those assets to provide reliable communications. Here, and throughout the remainder of this book, we will work with decibel (dB) measurements and dBm. The following section explains the relationship between these quantities for those whose acquaintance with them may be limited.

Returning to link budgets, let us consider the case of a 10-W transmitter and a receiver with a sensitivity of 5.0 μV, having antennas with zero gain, and a 3-km wireless transmission path operating at 500 MHz. Figure 1.2 illustrates this link. With a transmitter at +40 dBm and a receiver that can operate with signals as small as −93 dBm, the link budget allowance is

$$40 + 93 = 133 \text{ dB}.$$

Transmission loss (free space) at 3 km and 500 MHz is

$$20 \log\left(\frac{4\pi \times 3000}{\lambda}\right) = 20 \log\left(\frac{4\pi \times 3000}{0.6}\right) = 96 \text{ dB}$$

which means that this link has a margin of

$$133 - 96 = 37 \text{ dB}$$

more than the minimum required for operation. This 37 dB may be used to ensure a low bit error rate, to mitigate the effects of fading, or even to reduce the

Table 1.3 Typical Signal Loss

Type of link	Typical loss
Satellite	
Synchronous	190–200 dB
Low earth orbiting	140–160 dB
Aircraft	40–160 dB
Mobile	40–130 dB
Coaxial (short distance)	1–40 dB per 100 feet
Fiber optic	1–10 dB per mile

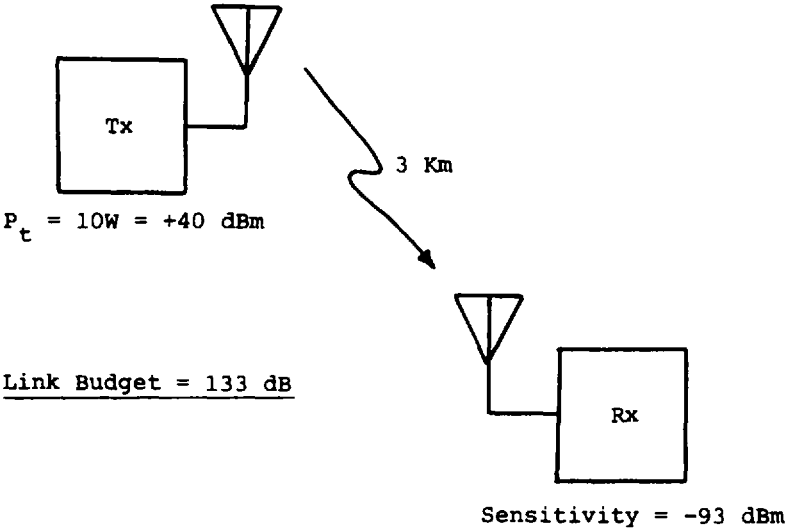


Figure 1.2 Link budget illustration.

transmitter's power. Of course, the receiver's required sensitivity might also be reduced. All of these are practical and common uses for an excess link margin.

Parts of any link budget are the available transmitted power, any antenna gain(s), receiver sensitivity, losses in the transmission medium, and any other factor that affects the level of the signal that reaches a receiver. The power level available at the transmit end of a link is often expressed in terms of the effective radiated power (ERP), which consists of the power amplifier's output, enhanced by any antenna gain and reduced by any losses. The term EIRP is also sometimes used to signify effective isotropic radiated power. In this case, any antenna gain is in comparison with a 0-dB gain, or isotropic, antenna. As previously stated, the free space loss for a signal is given in different forms, one of which is

$$L_{FS}\ (\mathrm{dB}) = 36.6 + 20\log f + 20\log R$$

where f is in MHz and R is in statute miles.

A second form is

$$L_{FS}\ (\mathrm{dB}) = 20\log\left(\frac{4\pi R}{\lambda}\right)$$

These two forms are equivalent and may be used to calculate overall loss interchangeably, although in instances in which one parameter is changing while others are constant, one or the other form may be easier to use.

If we begin with the second form

$$L_{FS}\ (\text{dB}) = 20\log\left(\frac{4\pi R}{\lambda}\right)$$

and convert

$$20\log\left(\frac{4\pi R}{\lambda}\right) = 20\log 4\pi + 20\log R + 20\log\frac{1}{\lambda}$$

$$= 21.98 + 20\log R + 20\log\frac{f}{v}$$

$$= 21.98 + 20\log\frac{1}{v} + 20\log R + 20\log f$$

where v and f are consistent terms.

Example. Given R in meters, f in MHz, and $v \cong 300$.* Assume $R = 10$ km and $f = 915$ MHz.

$$L_{FS} = 21.98 + 20\log\frac{1}{300} + 20\log 10{,}000 + 20\log 915$$

$$= 21.98 - 49.54 + 80 + 59.23$$

$$= 111.67\ \text{dB}$$

Using the alternative form,

$$L_{FS} = 20\log\left(\frac{4\pi R}{\lambda}\right)$$

$$\lambda = \frac{300}{915} = 0.328\ \text{m}$$

$$L_{FS} = 20\log\left(\frac{4\pi \times 10{,}000}{0.328}\right)$$

$$= 111.67\ \text{dB}$$

Note that both answers are the same.

Equations of the first form employ a different, constant, correction factor that depends on the units of distance with which one is working. For convenience, the four most commonly used factors are given in Table 1.4.

---

* $\lambda = \dfrac{f}{v} = \dfrac{\text{frequency in MHz } (10^6)}{\text{speed of light } (299.792 \times 10^6 \text{ m/sec})}$.

Therefore if f is in MHz, $\dfrac{1}{v} \cong \dfrac{1}{300}$.

Table 1.4 Equations for Free Space Loss

Dimension	Free space path loss equation
Miles (statute)	Loss = $36.6 + 20 \log f\,(\text{MHz}) + 20 \log R$ (mi)
Miles (nautical)	Loss = $37.8 + 20 \log f\,(\text{MHz}) + 20 \log R$(nmi)
Kilometers	Loss = $32.45 + 20 \log f\,(\text{MHz}) + 20 \log R$ (km)
Meters	Loss = $-27.56 + 20 \log f\,(\text{MHz}) + 20 \log R$ (m)

Table 1.5 lists typical losses that can be expected in coaxial cable, per hundred feet, versus frequency. Coaxial cable is often used in short-distance runs to avoid having to employ a wireless link. One example is its use in computer networks.

1.2 SIGNAL SELECTION AND REJECTION

A primary task performed by a receiver is that of selecting a desired signal while rejecting all of the signals that are present but not desired. The basic methods used by receivers to differentiate between signals are selection by frequency, selection by time, and selection by code. (Modulation may also be employed to differentiate between signals, but the ability of a receiver to differentiate on the basis of modulation alone is usually not sufficient to allow reliable signal selection. For example, AM broadcast stations and FM broadcast stations are readily selected by inexpensive receivers, but not on the basis of their modulation. The selection is actually made on the basis of their frequency, as AM and FM stations operate in frequency bands that are many MHz apart.)

Table 1.5 Loss in Typical 50-Ohm Coaxial Cable per 100 Feet as a Function of Frequency (Given in dB)

Frequency (MHz)	RG-58	RG-8	RG-213	RG-142	RG-393	¼-inch Hel.[a]	⅜-inch Hel.
150	5	2	1.5	4.6	2.7	2.2	1.5
450	11	4.3	2.8	8.4	4.9	4.9	2.7
960	16.5	9.0	4.4	13.0	7.6	5.9	4.0
2000	~40	~21	7.0	20.2	11.9	8.7	6.0

[a]Hel = Heliax. Heliax is a registered trademark of Andrew Corporation.

1.2.1 Frequency Selection

Frequency selection in receivers is accomplished through the use of bandpass filters and accurate* frequency sources. Both tunable filters, usually employed at the frequency used in the transmission medium, and fixed filters, used at intermediate frequencies, are found in common receivers. Historically, differentiation of users by frequency selection and rejection with filters has been the primary method employed.

When more than one transmitter is present and the transmitters are operating at separate frequencies, a properly designed receiver can reject signals at adjacent frequencies by 80 dB or more, using tuned bandpass filters. For many years, this (the use of filters for selection and/or rejection) was the only method used in radio communications. The entire radio industry was built upon it. So was the Federal Communications Commission, which exists for the purpose of regulating the use of the spectrum available, which it has carefully doled out, frequency by frequency. Recently, because of the need to consider other approaches to separating many users from one another, the overall method by which frequency of operation is employed to differentiate users has come to be called frequency division multiple access or FDMA.

When the most desirable frequencies were all given out, frequency division was expanded by using spatial separation of same-frequency users, but the mobility of the many users today has led to the need for sharing of frequencies. This means that users must be distinguishable by means other than frequency, although frequency is and always will be an important characteristic of a signal (we cannot do without frequency separation between signals). Other techniques have been developed to enhance our ability to differentiate between signals, as follows.

1.2.2 Time Selection

A very common technique used to differentiate between signals is to assign separate time intervals or "slots" to would-be users. This technique is commonly called time division multiple access, or TDMA, and has an everyday use in T1 carriers, which are a very good example.

In telephone applications, an analog voice signal is typically digitized by sampling at a rate of 8000 samples per second, using 8 bits per sample. This produces a data rate per voice signal of 8×8000 or 64 kbps. A T1 carrier is then used to carry 24 of these 64 kbps digitized voice signals, multiplexed together. The total transmission rate is then

*The degree of accuracy required is very application dependent.

$$
\begin{aligned}
24 \times 64000 &= 1{,}536{,}000 \text{ bps} \\
\text{Overhead} &= \underline{\quad\;\; 8{,}000 \text{ bps}} \\
\text{Total} &= 1{,}544{,}000 \text{ bps}
\end{aligned}
$$

where the 8 kbps of "overhead" is employed to keep track of the frame in which each user's signal is sent.

There are many other applications of time division as a tool to provide selectivity. Some of these will be covered in the chapters that follow.

Time division, like frequency division, depends on the accuracy of the receiver's oscillators and the ability of the receiver to synchronize to the frame in which the multiplexed signals are sent. Some frequency selection and rejection with filters are still required.

1.2.3 Code Selection

A technique that employs signals that are modulated with different codes to differentiate them from one another is also viable. This technique usually falls within the domain of the "spread spectrum" signaling methods, because the process of code modulation forces the use of a wider bandwidth than the information alone requires. Such systems are called code division multiple access or CDMA systems.

The primary limitation of such systems is the inability to construct codes that are sufficiently orthogonal to one another, to permit a receiver to tell the difference between one code and another reliably. A second limitation is that the ability to reject an unwanted signal is a function of the ratio of bandwidth to data rate. Therefore, the receiver's ability to reject undesired signals is often limited by the lack of sufficient bandwidth availability.

Code selection systems seldom, if ever, have as much selectivity (the ability to reject an unwanted signal) as a good receiver using frequency selectivity with bandpass filters. A good bandpass filter can readily provide 80 dB of unwanted signal rejection, whereas codes can rarely be relied on to give more than 40 dB of rejection.

1.3 DEMODULATION

The most critical point within a receiver is the input to the demodulator, because it is the signal level and the signal-to-noise ratio at the demodulator input that determine the quality of the recovered information that is presented to the user. Every type of modulation requires a specific class of demodulator, whether the information sent is analog or digital. (The term "digital" with regard to modulation is a misnomer; it should be called "discrete" modulation.)

Chapter 2, which follows, discusses modulation and lists recommended demodulators for particular types of modulation. Chapter 8 details the characteristics of these demodulators and their design.

1.4 ERROR DETECTION AND CORRECTION

Error detection and correction (EDAC) techniques may be employed to improve the bit error rate performance of a communication system. The receiver's role is to decode information in the form of digital data, making use of redundancy that is part of the received, encoded signal. The entire science of error correction coding is that of finding a set of symbols whose mutual Hamming distance is as large a number as possible, because this distance determines the number of bits in a symbol that can be detected and/or corrected.

In practice, up to approximately 15 percent of the bits in a symbol can be in error and the symbol can still be correctly identified. (This, of course, means that the bits in error can be identified and corrected.) A symbol must contain significantly more bits than are contained in the data it represents for this process to be carried out properly. Therefore, the encoded signal rate must be significantly greater than the unencoded rate, which means in turn that the receiver's bandwidth must be greater.

As a rule, for error detection and correction coding to be effective, the signal-to-noise ratio loss due to increasing the receiver's bandwidth must be less than the improvement achieved through encoding. We will expand on this and other aspects of error correction methods in Chapter 13.

1.5 OUTPUT REQUIREMENTS

Output requirements for receivers have been developed over decades and are highly application dependent. Analog voice receivers, for example, usually require an output signal-to-noise ratio of 10 dB or more, when the receiver input RF signal is the minimum that is acceptable. This signal-to-noise ratio requirement is the result of a great deal of testing to determine the ability of the average untrained listener to understand what is being said under typical operating conditions.

Data receiver output is usually specified in terms of bit error rate, although it is possible to define output in terms of signal-to-noise ratio or E_b/N_0 (the ratio of energy per bit to noise density). In most applications, the receiver output error rate is allowed to be in the range of one error per thousand bits to one error per million bits.

Output level and output impedance vary widely by application. Chapter 8 is concerned with receiver output requirements.

1.6 RECEIVER LIMITATIONS

A number of factors limit the ability of a receiver to select a particular signal and demodulate it. These factors include everything from inability to amplify a desired signal sufficiently, to rejection of unwanted signals. In some cases, the receiver itself generates signals that interfere with its operation. In others, the noise that exists in a particular application may limit what can be done with the receiver. Some of the factors that limit a receiver's performance are discussed in the following pages.

It is always well to bear in mind that when the receiver has been properly designed, with its noise figure and noise bandwidth minimized, and its demodulator is working as close to perfection as possible, there may be nothing more that can be done with the receiver itself. In that case, higher effective radiated power, lower propagation loss, or higher receiver antenna gain (or all of these) is necessary. The only other recourse is to change the specification.

1.6.1 Noise Figure

A receiver's noise figure is the amount of noise (in dB) that it adds to the input noise (KTB)* within its noise bandwidth. Today's receivers typically have noise figure in a range from a few tenths of a dB to a maximum of less than 10 dB at the highest frequencies. Satellite receivers routinely employ low-noise amplifiers (LNAs) with a noise temperature of less than 100 degrees.† This is a noise figure of

$$10 \log\left(1 + \frac{100}{290}\right) = 1.29 \text{ dB}$$

which is excellent. It also leads to the conclusion that current receivers have very little reason to be limited by noise figure, especially when one considers that reducing a receiver's noise figure to 0 dB would improve its performance by only a small amount. (In the case of the satellite receivers with a noise temperature of 100 degrees, reducing the noise temperature to zero would improve the receiver's per-

*K is Boltzman's constant = 1.38×10^{-23} W/Hz/°Kelvin.
T is temperature = 290° Kelvin.
B is Bn for the receiver.

†The relationship between noise temperature and noise figure is $T_n = 290(\text{NF} - 1)$, where NF is the noise factor.

formance by only 1.29 dB.) The point is that there is very little reason for today's receivers to be significantly limited by their noise figure or noise temperature.

The total noise figure in a receiver is not completely a function of the preamplifier, although the preamplifier is a major contributor. The total noise figure is

$$F_{\text{total}} = 10 \log \left(\text{NF}_1 + \frac{\text{NF}_2 - 1}{G_1} + \frac{\text{NF}_3 - 1}{G_1 G_2} \cdots \right)$$

where NF is noise factor, and the noise figure is $F = 10 \log \text{NF}$. It is readily seen from this equation that if G_1 is large compared with NF_2, $G_2 G_2$ is large compared with NF_3, and so on, then NF_1 (the noise factor of the first stage) is the major contributor to the receiver's noise. This is the reason for preoccupation with the preamplifier noise figure in most receiver designs.

Table 1.6 lists the overall noise figure for a preamplifier and second stage, as a function of preamplifier noise figure and gain, together with the second stage noise figure. From this table it is apparent that if a preamplifier has gain of 20 dB, a second stage such as an active mixer has little influence on the noise figure of a receiver unless the noise figure is greater than 15 dB.

1.6.2 Receiver Gain

The minuscule signal levels that routinely occur at the input of a receiver force the receiver to amplify the signal by very large factors. Amplification of 10^5 to 10^7

Table 1.6 Overall Noise Figure for a Pair of Receiver Stages as a Function of the First Stage Gain and Noise Figure and the Second Stage Noise Figure

Preamplifier		Overall noise figure			
		Second stage noise figure			
F (dB)	G (dB)	5 dB	10 dB	15 dB	20 dB
1.5	10	2.1	3.6	6.5	10.5
	15	1.7	2.3	3.8	6.6
	20	1.6	1.8	2.4	3.8
3.0	10	3.5	4.6	7.0	10.8
	15	3.2	3.6	4.7	7.1
	20	3.1	3.2	3.6	4.8
5.0	10	5.3	6.1	7.9	11.2
	15	5.1	5.4	6.2	8.0
	20	5.0	5.1	5.4	6.2

between the input signal and output signal ports of a typical receiver is common. Without gain in this range the output signal would not be usable, and this is true of receivers intended for application in either analog or digital signal reception.

The amount of gain required often results in the need to employ a superheterodyne receiver architecture with more than one intermediate frequency. This expedient is often employed to permit the receiver designer to keep the gain required at each frequency to a manageable amount (if the gain required at one frequency is too high, it is difficult to stabilize the amplifier).

It should be realized that engineers who work on radio frequency (and higher) circuit design spend the majority of their time doing one of two things: 1) trying to keep amplifiers from oscillating or 2) trying to get oscillators to oscillate. In the same vein, it is well to remember this adage from Charles R. Cahn:* "Nothing works like a little more signal."

1.6.3 Stability

The subjects of gain and stability cannot readily be separated. After all, the primary citerion for oscillation is

$$AB \geq 1$$

where A is gain and B† is the feedback factor. Therefore, the greater the gain, the smaller the amount of signal feedback needed for oscillation.

One of the most common causes of poor sensitivity in receivers is low-level oscillations within the receiver itself that cannot be seen when a strong signal is present but may be larger than the received signal when that signal is small. Often, such a problem is accentuated by automatic gain control (AGC), which adjusts the receiver's gain to compensate for smaller signal levels. Under such conditions, a receiver may very well oscillate only when its input signals are small or not present at all. In this case, the designer has reached a magic plateau: a receiver that does not require an input signal to produce an output.

The trick is to provide sufficient gain to meet the demodulator input signal level requirement and output signal requirement without having the receiver generate its own unwanted signals. A process by which a receiver may be designed to meet its requirements under widely varying conditions is described in Chapter 12.

Another form of stability is that property of a receiver to be set accurately to a particular frequency and remain set to that frequency (and bandwidth) over long intervals of time and over widely varying environmental conditions. Although it is quite practical to provide frequency accuracy to within a few parts in 10^{10} to 10^{12},

---

*In a private conversation with this author.

†$0 \leq B \leq 1$.

the cost of doing so, the size of the frequency sources required, and the manner in which receivers are actually used* produce accuracy that is usually no better than approximately 1 part in 10^6 and stability that is in the range of a few parts in 10^8 to 10^9 per day.

Oscillator and synthesizer accuracy and stability are discussed in Chapter 7.

1.6.4 Demodulator Performance

Specific types of demodulators (see Chapters 2 and 8) are employed for particular types of modulation. Their performance limits a receiver's ability to recover information:

1. To the extent that the demodulator requires a minimum input signal-to-noise ratio to perform properly and to produce a given minimum output signal-to-noise ratio. (The same is true of receivers whose performance is expressed in terms of the ratio of energy per bit to noise density and whose output is in terms of bit error rate.)
2. To the degree to which a demodulator reaches the performance of a perfect theoretical demodulator.
3. In relation to any "processing gain" that may be provided due to the modulation employed (see Chapter 2).

Almost all demodulators require an input signal-to-noise ratio of at least 6 to 10 dB to operate properly. They can usually be designed to perform to within a range of 1 to 3 dB worse than a perfect theoretical demodulator.

Curves showing the theoretical performance for various modulation techniques using a perfect demodulator are given in Chapter 8.

1.6.5 Bandwidth

A receiver's bandwidth is one of the primary factors that defines its sensitivity, because the bandwidth determines the effective input noise to the receiver, or KTB. Any receiver must have signal input that is greater than the input noise by an amount

$$\frac{S}{N}_{(\text{demod})} + F$$

where S/N (demod) is the input signal-to-noise ratio required by the demodulator to produce a useful output signal and F is the receiver's noise figure.

That is, a receiver with bandwidth B_N will have sensitivity that is not better than (in dBm)

*They are turned off most of the time and turned on sporadically.

$$-174 + \frac{S}{N}_{\mathrm{OUT}} + F + 10\log B_{\mathrm{N}} + \text{losses} - G_{\mathrm{p}}$$

where G_p is process gain—the improvement in S/N ratio provided by a demodulator, such as the FM gain in an FM demodulator.

A receiver with 0-dB noise figure, a perfect demodulator (no losses), a 10 dB signal-to-noise ratio, and no process gain can have sensitivity that is no better than that shown in Table 1.7.

1.6.6 Selectivity

Receiver selectivity is determined by the filters the receiver employs and is a measure of the receiver's ability to reject signals at frequencies outside the desired band. If a receiver's response to a signal is plotted from zero frequency to infinity the result is a "selectivity" curve, and the receiver's response to a signal R hertz away from the frequency to which the receiver is tuned is its selectivity. Typically, selectivity is specified as P_{dBc} at an offset of R hertz, as illustrated in Figure 1.3. A special case of selectivity, called "adjacent channel rejection," is simply the receiver's response to a signal in the channel next to the desired channel.

1.6.7 Spurious Responses

All receivers may have spurious (unwanted) responses to signals other than the specific ones they are intended to receive. The best general receiver configuration known is also the most vulnerable to spurious signal responses. The problem is that a receiver cannot properly demodulate a signal at the desired frequency or time, or with the desired code, when a spurious signal is present at the same time. Perhaps even worse is the case of a receiver that responds to a spurious signal when no desired signal is present. Spurious responses may be caused by a number of different mechanisms, which are discussed in Chapter 4.

Table 1.7 Sensitivity Possible from a Perfect Receiver as a Function of Its Bandwidth, Given T = 290 K and Output S/N Ratio = 10 dB

Bn	Best possible sensitivity (dBm)	*Bn*	Best possible sensitivity (dBm)
10 kHz	−124	750 kHZ	−105
50	−117	1.0 MHz	−104
100	−114	1.5	−102
250	−110	10.0	−92
500	−107		

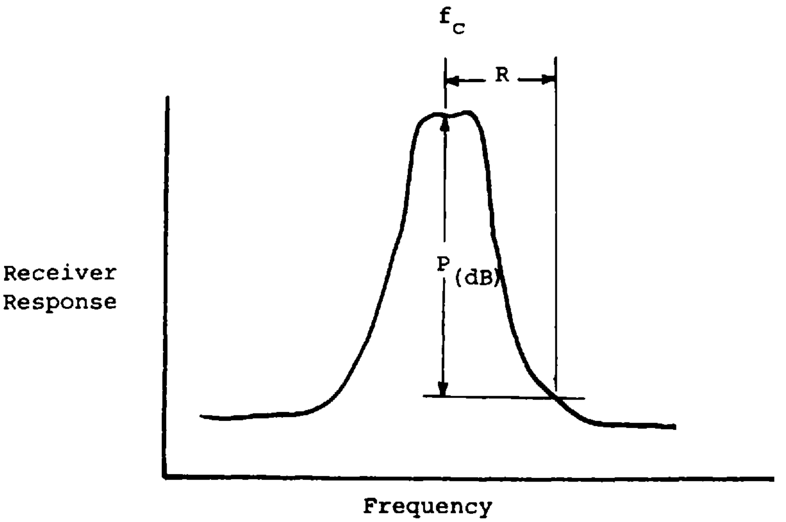


Figure 1.3 Receiver selectivity curve showing response −PdBc at offset *R*.

1.6.8 Linearity

In many cases, the linearity of a receiver determines its response to spurious signals. In other cases, the modulated signal must be linearly processed to avoid distortion of the information being conveyed. (AM is one example of such a signal.) On the other hand, FM receivers often employ limiting amplifiers that severely distort the signals they process. Because the information is conveyed by frequency, however, the information is not affected.

QUESTIONS

1. What is the free space propagation loss for a radio system operating at 900 MHz and a transmitter-to-receiver distance of 10 km?
2. Name the most important characteristics of any receiver.
3. What is the frequency range designation for a receiver that covers the band 225 to 400 MHz?
4. What is the wavelength in inches of a signal at 824 MHz?
5. Name the components that make up a link budget.
6. What is an impedance match and why is it important?
7. If a system has a link budget of 125 dB, can it be expected to transmit 99 miles at 2.3 GHz?

8. What can be done to improve the sensitivity of a receiver?
9. What determines a receiver's selectivity?
10. What is the effective input noise to a receiver whose noise bandwidth is 100 kHz? How is this related to receiver sensitivity?
11. What measures are normally specified for output performance in receivers with analog and digital output information?

REFERENCES

1. *Reference Data for Radio Engineers*, Howard W. Sams & Co., 1975.
2. Dixon, R. C., *Spread Spectrum Systems, With Commercial Applications*, John Wiley & Sons, New York, 1994.
3. Feher, K., *Digital Communications—Satellite/Earth Station Engineering*, Prentice-Hall, Englewood Cliffs, NJ, 1983.
4. Hamming, R. W., *Coding and Information Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1980.
5. Terman, F. E., *Electronic and Radio Engineering*, McGraw-Hill, New York, 1955.

2
Modulation and Bandwidth

Modulation and bandwidth are of very great importance to a receiver, because the receiver must accommodate the particular form of modulation to be received. Receiver bandwidth, in itself, is a function of the information rate and the form of modulation being employed. The receiver must recover the information being conveyed by the incoming signal, and to do this it demodulates the signal (a process that is, in itself, a modulation process). Receiver bandwidth, which must be wider as the information rate becomes higher, determines the amount of noise with which the receiver must deal.

We are concerned then, with modulation, because the receiver must have sufficient bandwidth to process the signal without significant distortion and because it has the job of recovering the information from the modulated carrier signal.

Why is modulation necessary? For the same reasons that tuning is necessary. It is quite possible to transmit a voice signal, for example, by amplifying it and then coupling it with an antenna into the electromagnetic medium. It would then propagate with very low loss through that medium. There are two difficult problems with doing this, however. The first is that the size of the antenna required would be completely impractical, as a quarter-wavelength antenna at the middle of the audio band* would be over 45 km long. The second problem is that in the entire world, very few persons at any one time could send a signal, as two or more of such signals could interfere with one another for many thousands of miles.

For these reasons it is necessary to use separate "carrier" signals at higher frequencies that are in turn modulated by the information to be sent (voice in this example). If the carriers, which are separated by frequency, are then selected by a receiver using filters to determine the carrier selected, the signal must still be

*300 to 3000 Hz.

Table 2.1 Analog Modulation Techniques, Bandwidths, and Demodulators

Analog modulation method	Modulated RF bandwidth	Modulated signal representation	Demodulator(s) typically employed
AM (DSB)	$2B_{\text{info}}$	$A\cos\omega_c t + \frac{Am}{2}[\cos(\omega_c + \omega_m)t + \cos(\omega_c - \omega_m)t]$	Envelope detector Product detector Phase lock loop
AM (DSBSC)	$2B_{\text{info}}$	$\frac{Bm}{2}[\cos(\omega_c + \omega_m)t + \cos(\omega_c - \omega_m)t]$	Product detector Costas loop Squaring loop
AM (SSB)	B_{info}	$Bm\cos(\omega_c + \omega_m)\,t$ or $Bm\cos(\omega_c - \omega_m)\,t$	Product detector (except vestigial carrier SSB)
FM	$\geqslant 2(f_m + f_d)$	$B\cos(\omega_c t + \cos\omega_m t)$	Discriminator Ratio detector Quadrature detector Phase lock loop

Table 2.2 "Digital" Modulation Techniques and Bandwidths

Digital modulation method	Modulated RF (null-null) bandwidth	Modulated signal representation	Demodulator(s) typically employed
M-ASK	$\frac{4R_{\text{data}}}{M}$	$\frac{DA}{M-1}\cos\omega_c t$	Envelope detector
NCFSK	$2R_{\text{data}} + \Delta f$	$f(0) = A\cos\omega_0 t$ $f(1) = A\cos\omega_1 t$	Any FM demodulator except phase lock loop Dual filter demodulator
CPFSK[a]	1.3 to $1.5R_{\text{data}} + \Delta f$	$A\cos\left(\frac{\pi t}{2T} - U\sin\frac{2\pi t}{T}\right)$ $0 < U < 1,\ -T < t < T$	Dual filter demodulator
M-FSK	$2R_{\text{data}} + (m-1)\Delta f$	$f(M) = A\cos\omega_m t$	Dual filter demodulator
BPSK	$2R_{\text{data}}$	$A\cos\omega_c t \pm 90°$	Squaring loop Costas loop Squaring demodulator
QPSK/OQPSK	R_{data}	$A\cos\omega_c t \pm 90° \pm 45°$	Multiplier loop Costas loop Multiplier demodulator
M-PSK	$\frac{4R_{\text{data}}}{M}$	$A\cos\omega_c t \pm 90° \pm 45° \pm 22.5° \cdots \pm 180°$	Multiplier demodulator

[a]Including MSK and other exotic waveforms.

demodulated to recover the information being conveyed. Here we will concentrate on the bandwidth required by a modulated signal and on the demodulation techniques available.

Modulation is divided into two general groups that are designated analog and digital. However, these designations do not refer to the modulation itself but to the type of information being sent.

All modulated carriers that are sent wirelessly* must be analog, without the discontinuities that are part (by definition) of a digital signal. Great effort is often expended in modulators and transmitters to reduce and/or suppress the effects of such discontinuities.

Summaries of analog and digital modulation techniques and their characteristics are given in Tables 2.1 and 2.2.

Please note that the purpose here is not to provide an exhaustive description of modulation and modulators but to describe the range of modulation methods, the bandwidth required, and the demodulation techniques employed. Demodulation itself will be discussed in detail in Chapter 8.

2.1 AMPLITUDE MODULATION

Amplitude modulation is well described by its name. That is, the information being conveyed causes the carrier signal level to vary as a function of the information. The amplitude modulated signal may be expressed as a carrier and two sidebands†

$$A \cos \omega_c t + \frac{Am}{2} \cos(\omega_c + \omega_m)t + \frac{Am}{2} \cos(\omega_c - \omega_m)t,\ 0 \le m \le 1$$

where $A \cos \omega_c t$ is the carrier signal, $(Am/2)\cos(\omega_c + \omega_m)t$ is the upper sideband, and $(Am/2)\cos(\omega_c - \omega_m)t$ is the lower sideband. The total bandwidth required is the difference between the upper and lower sideband frequencies, which are centered around the carrier frequency as illustrated in Figure 2.1.

Figure 2.1 illustrates the spectrum of a single-frequency-modulated carrier. If a band of frequencies is used to modulate the carrier, then the bandwidth required by the modulated carrier is

$$\mathrm{BW}_{\mathrm{AM}} = 2f_{\max}$$

where $f_{\max}$ is the highest frequency in the modulating signal's range. In practice,

*With the exception of those that are sent with light waves (visible, infrared, etc.), which are not regulated.

†The signal is restricted to two sidebands because modulation greater than 100 percent is not allowed. Figure 2.2 shows an AM signal (time domain) modulated at 100 Hz with 95 percent modulation. Note that the carrier phase does not change.

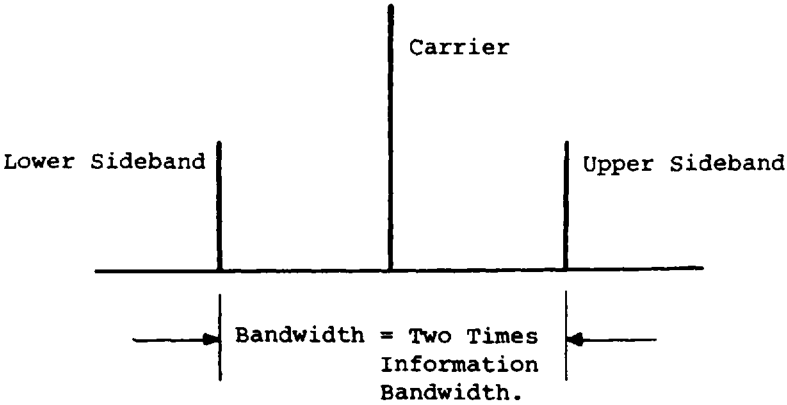


Figure 2.1 AM spectrum showing sideband structure. Sideband amplitude varies with percentage of modulation (50 percent modulation shown).

the 3-dB bandwidth of the modulated carrier is considered to be twice the 3-dB bandwidth of the modulating signal. Thus, a voice-modulated AM signal (assuming a voice bandwidth of 300 to 3000 Hz) would require a 6000-Hz modulated signal bandwidth.

Demodulators for AM signals are usually envelope detectors, product detectors, or phase lock detectors, all of which are examined in detail in Chapter 8. The most common demodulator for AM is an envelope detector, which exists in many forms. AM itself is employed primarily in applications in which a single transmitter is used to send a signal to many receivers (for example, in broadcast-band AM). This reduces the overall cost of the system. Even though an AM modulator and power amplifier is the most expensive type to build and use, the low cost of having multiple AM receivers keeps the overall system cost comparatively low.

The level of the sidebands varies as in the expression

$$\frac{Am}{2}\cos(\omega_c \pm \omega_m)t \qquad 0 \le m \le 1$$

which changes as the percentage of modulation m changes. Note that only the amplitude changes—the bandwidth does not do so—and m is allowed to vary only between a value of zero and one. At $m = 1$, the amplitude of the sidebands is one half the amplitude of the carrier. Therefore at 100 percent modulation, the power in each sideband is one fourth the carrier power. Of course, when the modulation percentage is zero, there are no sidebands.

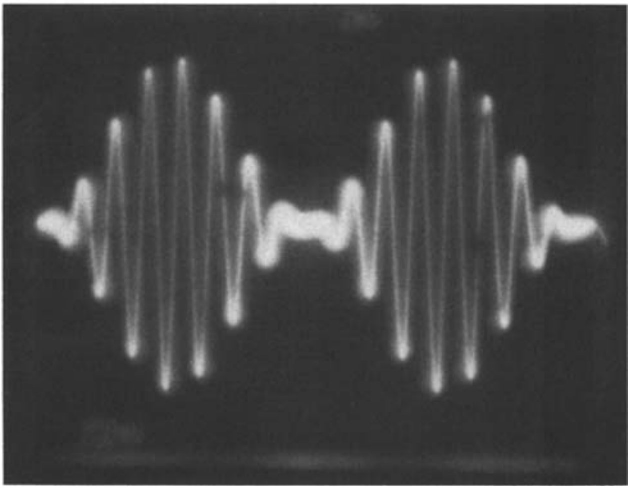


Figure 2.2 Ninety-five percent AM-modulated, 850-Hz carrier with 100-Hz modulation. Note that carrier phase is constant.

With voice systems, the average modulation percentage* is considered to be 30 percent, because the peak-to-average ratio is such that a voice signal that drives a modulator to 100 percent output on voice peaks modulates it only 30 percent on the average. The percentage of modulation, or modulation index, is simply the ratio of the actual modulation to the maximum that the modulator can support, where the modulation index is the fraction

$$\frac{\text{Actual output level}}{\text{Maximum output level}} = \text{modulation index}$$

and the modulation percentage is

$$\text{Modulation index} \times 100 = \text{percent modulation}$$

The AM modulation may be measured by observing the AM waveform as in Figure 2.2. If the peak of the modulated carrier is A volts and the minimum is B volts, then the modulation index is

$$\frac{A - B}{A + B} = m$$

and the modulation percentage is

*Also called "modulation index."

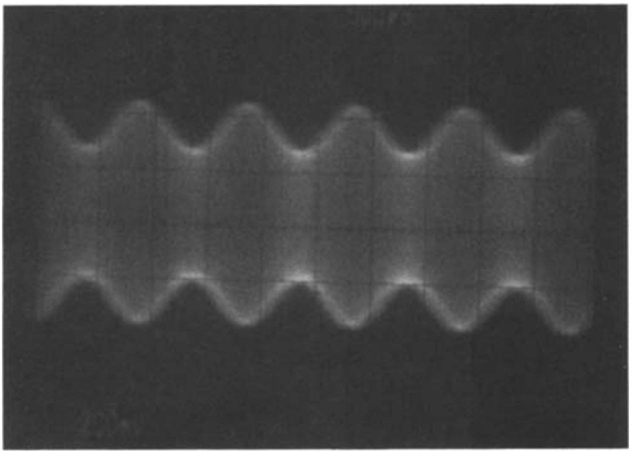


Figure 2.3 Thirty percent AM-modulated carrier, 1-kHz modulation. Time waveform.

$$100 \times \frac{A - B}{A + B} = m \text{ (percent)}.$$

We emphasize that the only difference in the receiver between one modulation index and another is that the higher the index (or percentage), the greater will be the peak signal level. Because the bandwidth does not change, however, a re-

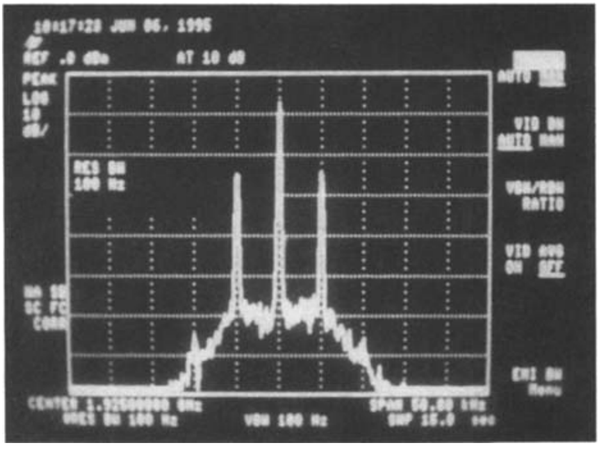


Figure 2.4 Thirty percent AM-modulated carrier spectrum, 5-kHz modulation (spectrum analyzer presentation).

ceiver's sensitivity will appear to improve whenever a higher modulation index is used. This is the reason that receivers are tested with a fixed, or "standard" modulation index (usually 0.3, or 30 percent for voice receivers).

Figures 2.3 and 2.4 illustrate a 30 percent AM-modulation carrier in time domain and frequency domain representations.*

2.2 AM DEMODULATION

Envelope detectors are used almost exclusively for AM demodulation, because of their simplicity and low cost. Product detection and phase lock demodulation may be employed, but performance may not equal that of the simple envelope detector. (See Table 2.3.)

Single-sideband (SSB) modulation is a special form of amplitude modulation that is sometimes employed to enhance the transmission range of radio systems. This method is one in which the carrier and one of the two sidebands are suppressed and the transmitter's output is concentrated in the sideband that remains. The modulated signal bandwidth required is then the same as the information bandwidth (or slightly wider if a vestigial carrier is transmitted).

In AM systems, the maximum modulation index permitted is 1.0 (100 percent), simply because the distortion produced by overmodulating causes sideband "splatter" (i.e., increases the bandwidth modulation, to say nothing of distorting the received signal and interfering with adjacent channel users). This means that the maximum total power in the sidebands (at 100 percent modulation) is one half the carrier power.

That is, if the carrier is $A \cos \omega_c t$ and each sideband is $Am/2 \cos (\omega_c \pm \omega_m)t$, then $m = 1$ produces two sidebands with an amplitude one half that of the carrier. But because the power ratio is the square of the amplitude ratio, we see that the total sideband power is one half the carrier power.

Now, if the carrier and one sideband are suppressed and all of the power of which the transmitter is capable is concentrated in one sideband (it does not matter which one), the transmitter is effectively sending six times as much power.

The primary difficulty with SSB transmissions is in their reception, in which the suppressed carrier must be replaced in the demodulation process. This is the reason for transmission of a vestigial (only partially suppressed) carrier in some systems.

Double-sideband (DSB) modulation is an alternative solution to the receiver demodulation problem of not having a carrier present. In this case the carrier is suppressed but both upper and lower sidebands are transmitted. This technique is

---

*Oscilloscope and spectrum analyzer presentations, respectively.

Table 2.3 AM Demodulators

Demodulator	Characteristics
Envelope detector	Simple, low cost, works well.
Product detector	Similar performance, more complex, requires local oscillator at carrier frequency, a must for SSB.
Phase lock loop	Even more complex. Does not work well with high modulation index.
Costas/squaring loop	Most complex. Used for DSBSC applications.

often called double-sideband transmission, although it is correctly termed double-sideband suppressed carrier (DSBSC) modulation. DSBSC transmission has three times the effective transmitted power of AM but does have the disadvantage of a missing carrier. Because both upper and lower sidebands are present in the modulated signal, the bandwidth is the same as AM. The missing carrier can be regenerated from this or any signal that has symmetrical upper and lower sidebands, however.

Figures 2.5 and 2.6 illustrate the waveforms and spectra of SSB and DSBSC signals.

2.3 FREQUENCY MODULATION

Frequency-modulated carriers usually require much more bandwidth than amplitude carriers do. This is because frequency modulation produces multiple subcarriers unless the frequency deviation of the carrier is highly restricted. In a frequency modulator, instead of varying the carrier amplitude as a function of the information, the carrier frequency is varied. This produces a modulated carrier waveform

$$g(t) = A \cos(\omega_c + \cos \omega_m)t$$

whose bandwidth is a function of the modulation percentage (or index) as well as the modulation rate.* This leads to the introduction of a new term, "deviation ratio," which is defined as

$$\beta = \frac{\Delta f}{f_m}$$

where Δf is the peak frequency shift from the center frequency, due to the modulating signal, and f_m is the modulation rate.

*Remember that the AM signal's bandwidth is not a function of its modulation index.

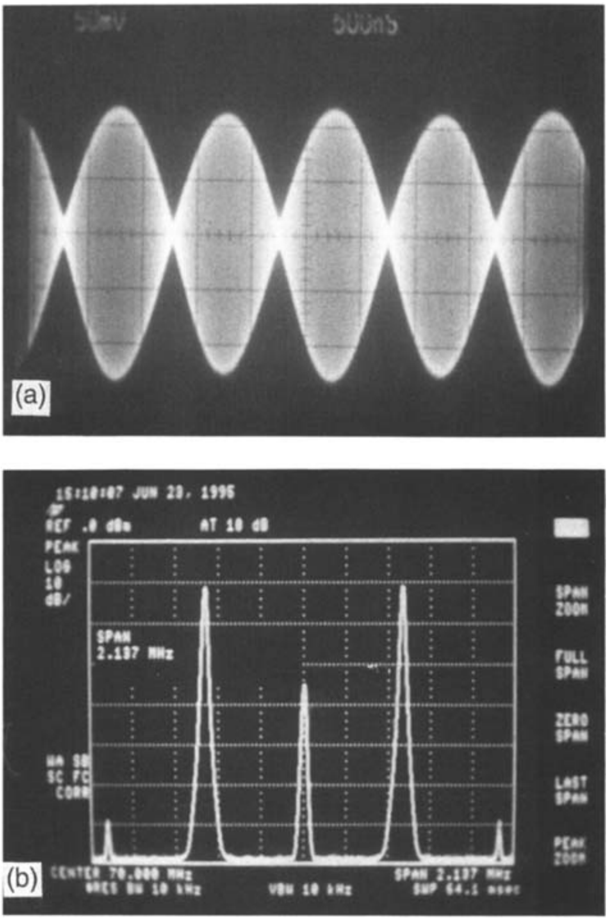


Figure 2.5 Double-sideband, suppressed carrier time and frequency characteristics, (a) DSBC signal (time waveform). Envelope is at modulation rate. (b) DSBSC spectrum. Carrier suppression approximately 25 dB.

For example, a carrier that is shifted by 1 MHz from its unmodulated frequency by a 10-kHz modulating signal has a deviation ratio β that is

$$\beta = \frac{1 \text{ Mhz}}{10 \text{ kHz}} = 100$$

(This, incidentally, is an unusually high deviation ratio. Broadcast FM stations employ $\beta = 5$.)

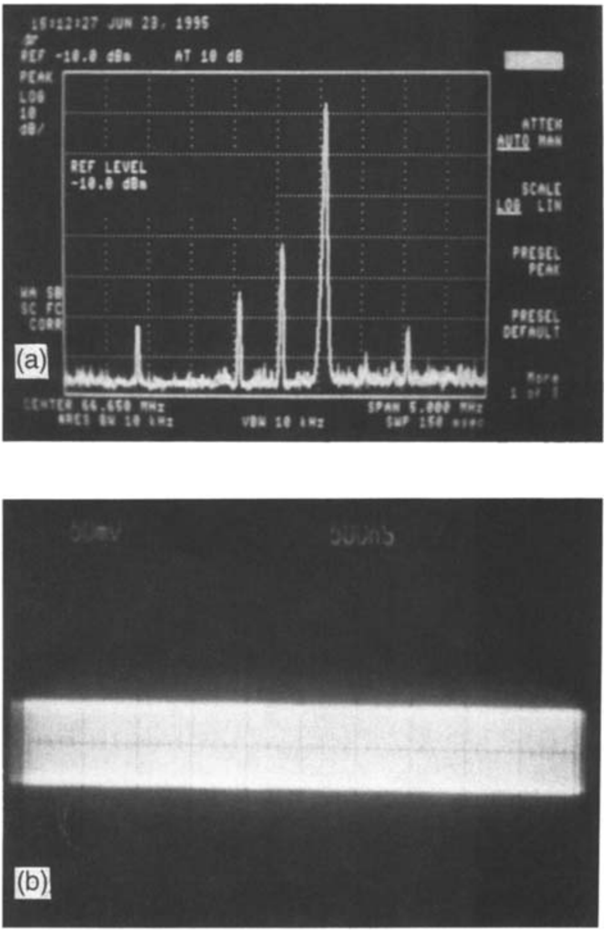


Figure 2.6 Single-sideband time and frequency domain characteristics (single tone modulation). (a) Single-sideband spectrum generated by high-pass filtering a DSBSC signal. (b) Single tone SSB signal at $\omega_c + \omega_m$.

Knowing the deviation ratio of an FM signal allows us to determine the modulated signal bandwidth necessary for a receiver to accommodate the signal with low distortion When the deviation ratio is known, a table or chart of Bessel functions of the first kind can provide the sideband levels to be expected.

Figure 2.7 is a graph of Bessel functions showing the amplitude and phase of the carrier and sidebands, relative to one another. If we consider a modulated

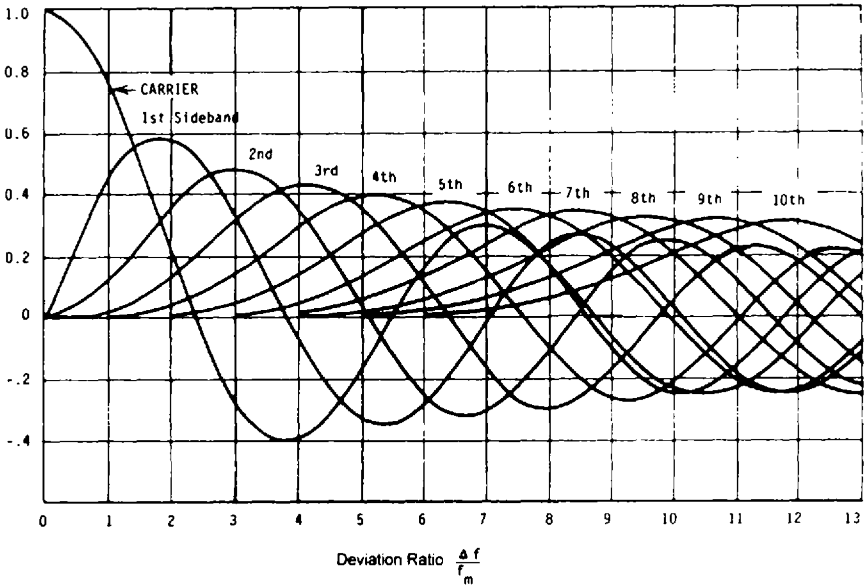


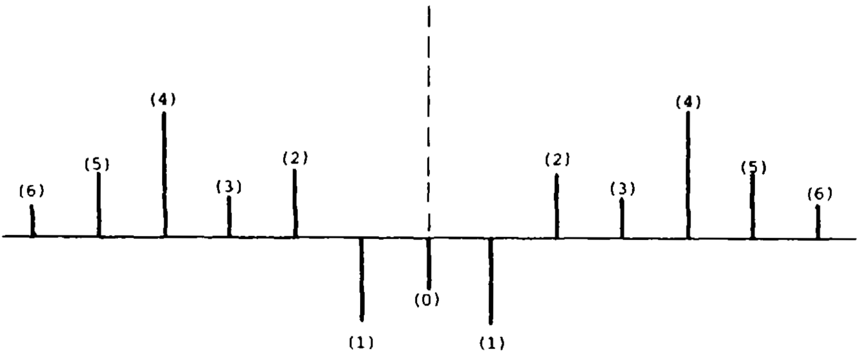
Figure 2.7 Chart of Bessel functions for FM signal analysis.

signal with $\beta = 5$ and move to 5 on the horizontal scale, we can read from the chart the following values of the carrier and sidebands along the vertical scale.

Signal	Level
Carrier	−0.18
1st sideband pair	−0.32
2nd sideband pair	0.3
3rd sideband pair	0.17
4th sideband pair	0.39
5th sideband pair	0.27
6th sideband pair	0.13

The spectrum of this signal would be as shown in Figure 2.8, where negative-going lines represent phase opposed to positive-going lines, and all lines are relative to the *unmodulated* carrier.

Observing Figure 2.7, we see that all sidebands less than pair six are at a higher level and all sidebands greater than pair six are smaller than pair six. Therefore, because pair six is at a low level and all higher order sidebands are



(a)

Figure 2.8 (a) Frequency modulation spectrum, single tone, deviation ratio = 5. (b) Spectra for FM with $\beta = 1$ through 4.

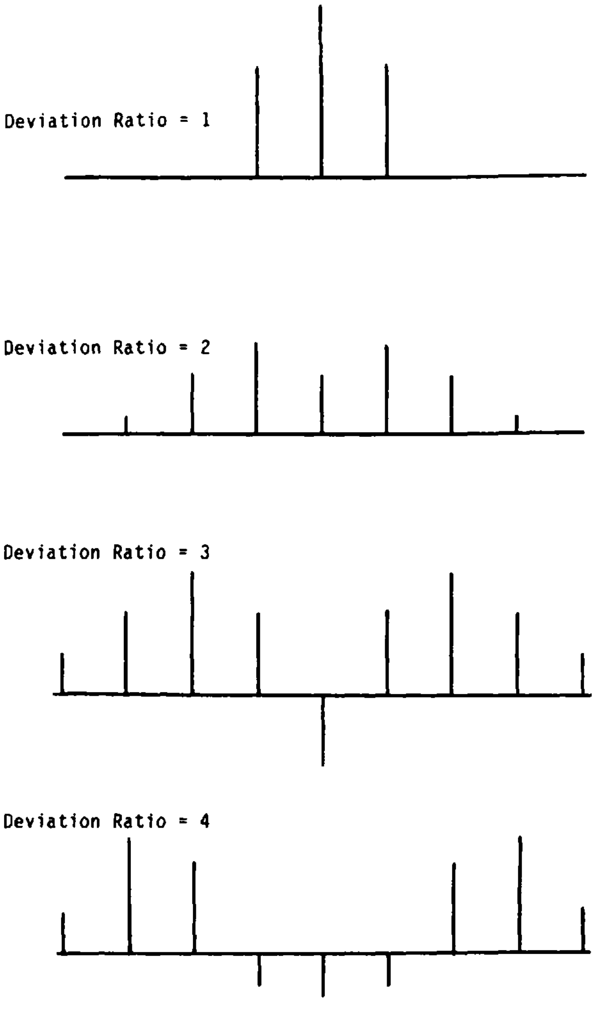
smaller, the bandwidth used would usually be truncated (by filtering) outside the ± 5 sideband range. This is exactly the case in broadcast FM, where the deviation ratio is 5, the modulation bandwidth is 15 kHz, and 150 kHz is the modulated signal bandwidth allowed. It is also of interest to note that "preemphasis" is employed in FM broadcast systems, which is a technique intended to ensure that the deviation ratio does not vary with modulating frequency. This in turn forces the use of a "deemphasis" network in an FM broadcast receiver.

Figures 2.8a and b illustrate the sideband characteristics of FM for deviation ratio 5.0 and 1.0 through 4.0, respectively; Figure 2.9 shows the cumulative power in FM sidebands versus deviation ratio; and Table 2.4 tabulates the results shown in Figure 2.9.

Frequency modulation has the advantage that it has a signal-to-noise ratio improvement capability that is a function of the square of the deviation ratio. When compared with AM, this improvement is $3\beta^2$, because of the characteristic that FM has a fixed* level or "constant envelope," while the average carrier power in an AM signal is approximately 30 percent. This improvement† in signal-to-noise ratio is one of the reasons why FM broadcast reception is typically much better than AM. We will discuss FM improvement and other effects in Chapter 8.

*No FM signal is completely free of unintentional amplitude modulation, just as no AM signal is completely free of unintentional frequency modulation, and these can affect the receiver. Such unintended modulation is termed "incidental FM" or "incidental AM," respectively.

†Called "process gain."


Deviation Ratio = 1
Deviation Ratio = 2
Deviation Ratio = 3
Deviation Ratio = 4

(b)

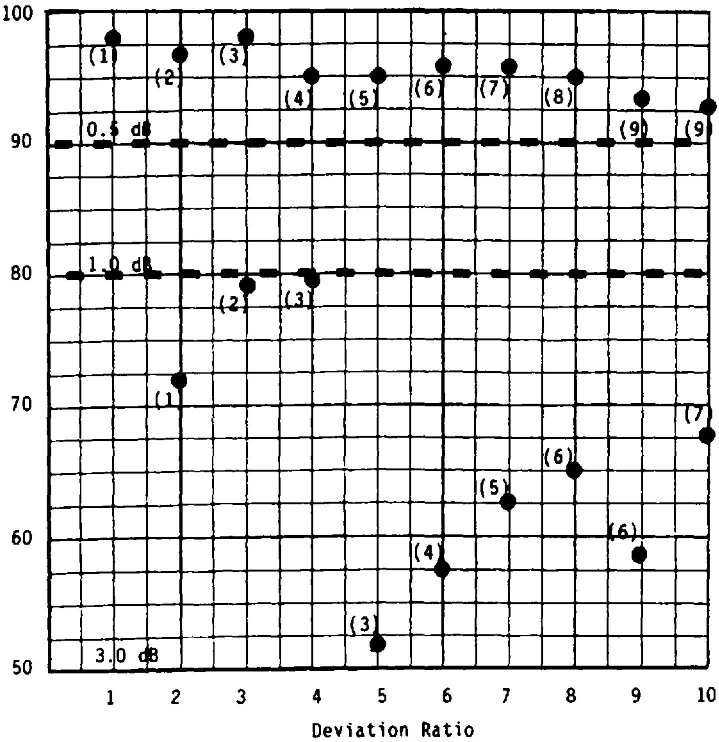


Figure 2.9 Cumulative power in carrier and sidebands versus deviation ratio. (N) is sideband pair by number.

Table 2.4 Number of Sidebands for Power (Two Sided, Including Carrier)

B	$P > 50\%$	$P > 90\%$	3dBBW	1dBBW
1		1 pair (98%)		2 info
2	1 pair (72%)	2 pair (96.5%)	2 info	4 info
3	2 pair (79%)	3 pair (98%)	4 info	6 info
4	3 pair (79.6%)	4 pair (95%)	6 info	8 info
5	3 pair (51.5%)	5 pair (95%)	6 info	10 info
6	4 pair (57.4%)	6 pair (96%)	8 info	12 info
7	5 pair (62.5%)	7 pair (96%)	10 info	14 info
8	6 pair (64.9%)	8 pair (95%)	12 info	16 info
9	6 pair (58.6%)	9 pair (93.6%)	12 info	18 info
10	7 pair (67.9%)	9 pair (92.7%)	14 info	18 info

Table 2.5 Contemporary FM Demodulators

Demodulator	Characteristics
Foster-Seeley discriminator	Sensitive to AM Alignment critical
Ratio detector	Improved AM rejection Alignment critical
Slope detector	Alignment not critical Poor AM rejection Moderate linearity
Pulse count discriminator	Excellent linearity Good AM rejection Alignment not critical
Quadrature detector	IC versions available Alignment not critical (Most common today)
Phase lock loops	Low threshold VCO temperature/aging Can be problem

2.4 DEMODULATORS FOR FM

No other modulation form has so many demodulators available. In addition to the classic Foster-Seeley discriminators, there are ratio detectors, slope detectors, gated-beam* discriminators, quadrature detectors, and others such as phase lock loops. Table 2.5 lists FM demodulators and their most important characteristics. All of these demodulators provide both FM information recovery (from a frequency-modulated carrier) and β^2 signal-to-noise ratio improvement. Except for the phase lock (and Costas) loops, all of these are noncoherent demodulators.

2.5 AMPLITUDE SHIFT KEYING (ASK)

The simplest form of amplitude shift keying is known by at least three different names: on-off keying (OOK), two-ary ASK (2ASK), and pulse AM (PAM). In any case, a binary one is usually signified by the presence (or absence) of a carrier, and a zero is signified by the opposite carrier state. For convenience, the presence of a carrier is usually assumed to convey a one, and any transmitter capable of being turned on and off at the digital data rate may be employed to send such data.

*An FM demodulator requiring a specially designed vacuum tube.

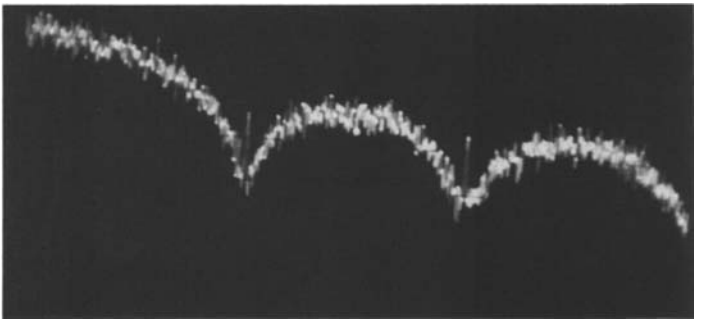


Figure 2.10 Baseband data signal. (Signals at nulls are harmonics of the data clock.)

There is, of course, no reason why the carrier must be turned off completely. If one chose to transmit the carrier at full power to signify a one and 10 percent of full power to signify a zero, then the vestigial power remaining in the zero state might be used to advantage in the receiver. The point is that any two carrier states may be employed to convey binary data. Here we will consider only the more conventional method in which the carrier is turned off or on completely.

The bandwidth of an on-off keyed signal is a function of the modulating signal, which in the case of a data signal is a series of trapezoids* whose period T is a multiple of the reciprocal of the data rate. That is,

$$T_{(N)} = \frac{N}{R \text{ data}}$$

where N is the number of ones or zeros in a row.

Because the frequency content of binary data is a time-varying series of sets of harmonics, the on-off keyed carrier has a spectrum that consists of upper and lower sidebands that have the same content as the baseband data. Figure 2.10 shows the baseband frequency spectrum of a binary data signal, and Figure 2.11 is the on-off keyed spectrum produced by modulating a carrier with the data signal.

If we compare the spectrum of on-off keying, or 2ASK, with that of biphase phase shift keying (BPSK or 2PSK; see Figures 2.11 and 2.12), we see that the spectra are identical except that the 2ASK spectrum has a strong carrier component and 2PSK lacks a carrier, as it is suppressed. The main lobe of the spectrum, which is usually transmitted, has a bandwidth at the first nulls that is two times the data rate.

---

*Assuming finite bandwidth, where the rise and fall times of the ones and zeros are small but still not negligible.

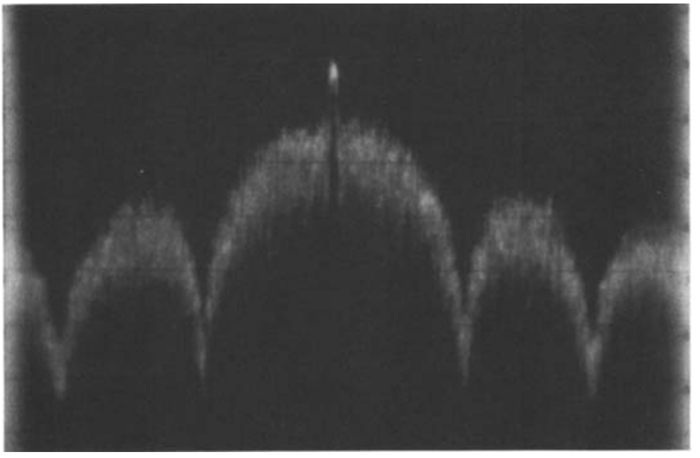


Figure 2.11 Baseband data signal on-off keying a carrier.

A receiver designed to receive such a signal typically has bandwidth between twice the data rate and 0.88* times the data rate. On a noise bandwidth basis, this is a range of 2/0.88 = 2.27.

Amplitude shift keying is not restricted to two levels. Four, eight, or more levels can be employed, where the number of bits of data conveyed by a given level is

$$\text{Information bits} = \log_2 M$$

and M is the number of levels employed. M-ary ASK employs the same amount of bandwidth, whether M is 2 or anything else. However, the signal-to-noise ratio required for demodulation increases as M goes up.

2.6 FREQUENCY SHIFT KEYING

Frequency shift keying will be considered in two categories.

1. Noncoherent FSK, in which frequency is shifted without regard to the phase relationship between frequency one and frequency two. A phase discontinuity is usually produced by such modulation.

*The spectrum shape is $(\sin x/x)^2$, which has a 3-dB bandwidth of 0.88 times the data rate. At twice the data rate (the BW between the first nulls) the power contained is approximately 95 percent, because of the unsuppressed carrier.

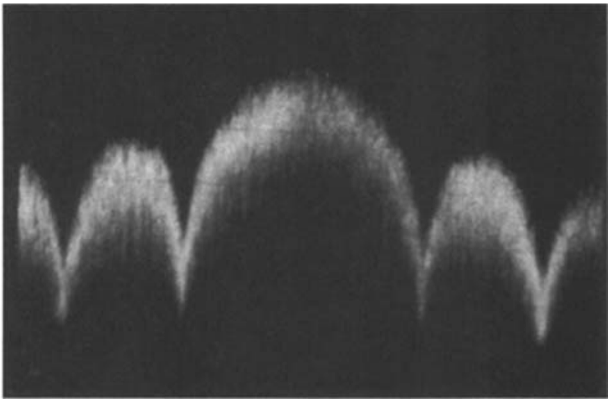


Figure 2.12 Baseband data signal BPSK modulating a carrier.

2. Coherent FSK, where the frequency is shifted without a phase discontinuity.

The reason for this differentiation is that the frequency spectra and demodulators used for coherent and noncoherent FSK are almost different enough to declare them separate forms of modulation.

Noncoherent FSK. A pair of oscillators, one at the frequency for a One, with the other at the frequency for a Zero, and with the two being on-off keyed by a data signal and its complement, produce noncoherent FSK when they are added together (see Figure 2.13). This, of course, is not the only way to generate noncoherent FSK, but it illustrates very well the point that for all practical

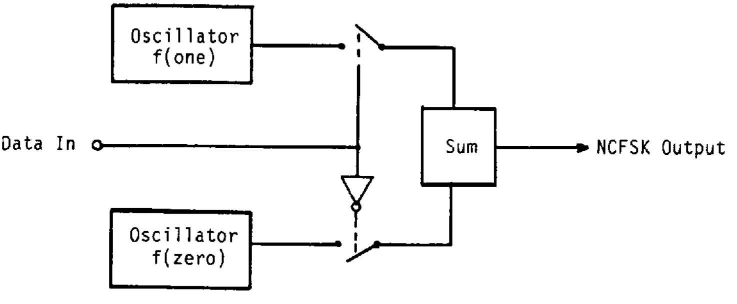


Figure 2.13 Noncoherent FSK generator.

purposes NCFSK is two on-off keyed carriers at different frequencies, that just happen to take turns transmitting depending on whether a One or a Zero is being sent.

2.7 ASK DEMODULATORS

As in AM demodulation, envelope detectors dominate over all others. Any other demodulator is employed only in very unusual applications.

Therefore, the spectrum produced is the same as that of two on-off keyed modulators that are offset from one another by Δf, the amount of the frequency shift between them. This spectrum is shown in Figures 2.14 through 2.17. Remembering that the bandwidth of an on-off keyed signal is two times the data rate (at the first nulls), the bandwidth of an NCFSK signal at the same points must be

$$BW_{NCFSK} = 2R_{data} + \Delta f \qquad \text{(null-null)}$$

which would be the received modulation bandwidth. Alternately, the 3-dB point of the $[(\sin x)/x]^2$ spectra might be used, which is

$$BW_{NCFSK} = 0.88R_{data} + \Delta f \qquad \text{(3dB, neglecting unsuppressed carriers)}$$

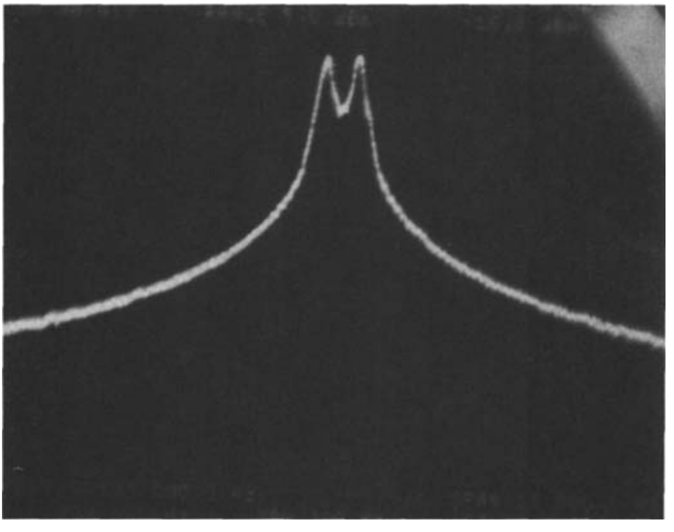


Figure 2.14 NCFSK, 50 bps ± 250 Hz square wave modulation.

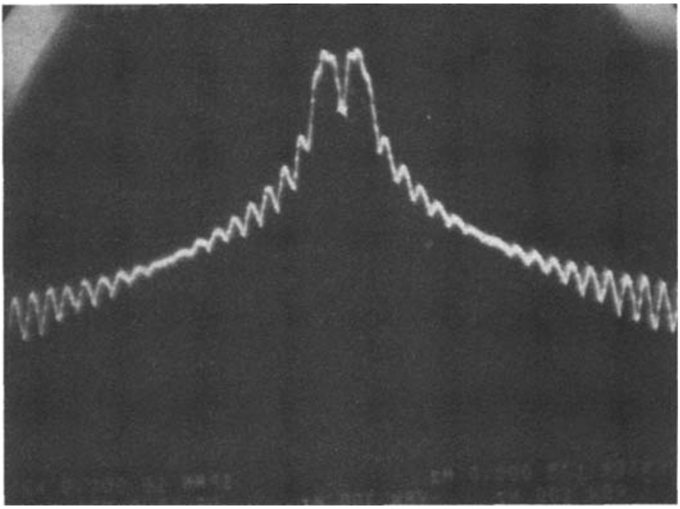


Figure 2.15 NCFSK, 100 bps ± 250 Hz square wave modulation.

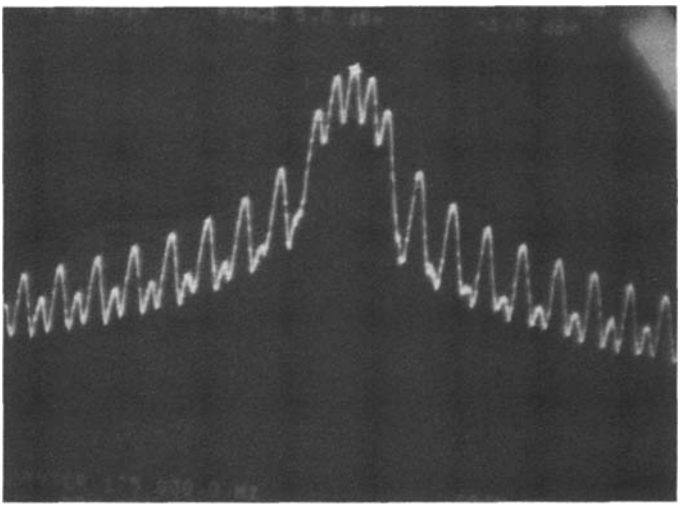


Figure 2.16 NCFSK, 250 bps ± 250 Hz square wave modulation.

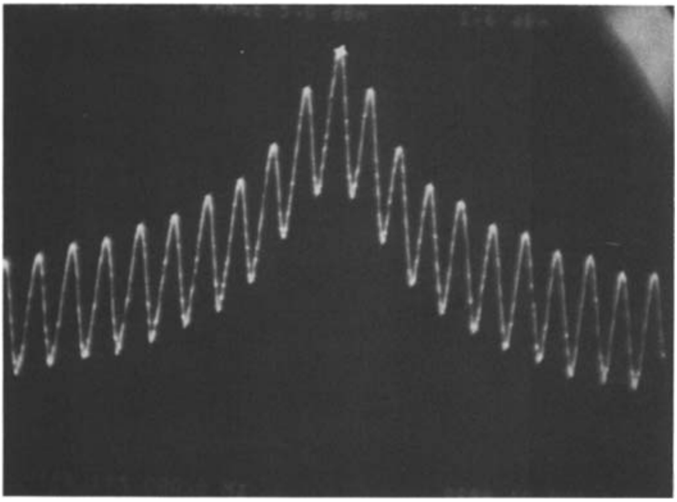


Figure 2.17 NCFSK, 500 bps ± 250 Hz square wave modulation.

In typical FSK systems, where Δf is chosen as 0.7 times the data rate, the bandwidth allowed for NCFSK modulation in a receiver would be between the 3-dB ($1.58R_{\text{data}}$) and null-null ($2.7R_{\text{data}}$) BWs. Another consideration with respect to NCFSK bandwidth is that beyond this bandwidth ($2.7R_{\text{data}}$) the signal rolls off at 6 dB per octave. Where there is no other designation, FSK typically implies noncoherent frequency shift keying.

2.7.1 Coherent FSK

Coherent frequency shift keying is also called digital FM, continuous phase FM (CPFM), and continuous phase shift modulation (CPSM) and in fact includes a whole class of modulation that we will discuss in the category of minimum shift keying (MSK).

Coherent FSK (which we will call CPFSK) differs from NCFSK only in that there is no phase discontinuity as the signal changes from one frequency to the other. This characteristic has two important implications.

Sidelobe energy is reduced (compared with NCFSK) and sidelobe rolloff is faster.

Coherent demodulation is practical. Phase lock loop demodulation can be employed, for example.

Figure 2.18 Simple coherent FSK (or CPFSK) signal generator using voltage-controlled oscillator.

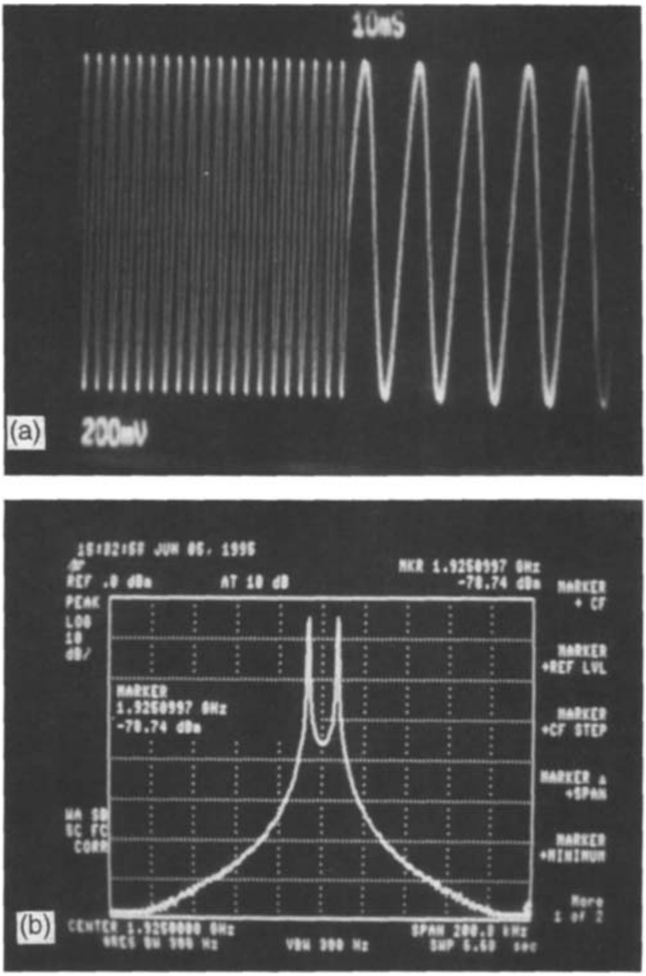


Figure 2.19 Continuous-phase FSK signal, with 14-kHz frequency shift at 200-Hz rate, using square wave modulating signal. (a) CPFSK signal with instantaneous frequency shift at zero crossing. (b) CPFSK signal spectrum.

Figure 2.18 illustrates the difference in the modulator for CPFSK. In this approach, data is low-pass filtered to remove the abrupt level shifts normally found in input data, which allows the voltage-controlled oscillator to change frequency smoothly from f (zero) to f (one) or vice versa. This frequency shift process is coherent, in that there are no phase discontinuities. That is, the VCO "slides" from frequency to frequency. In the linear portion of the shift, the frequency is

$$f = f\,(\text{initial}) + t\frac{df}{dt}$$

where df/dt is a function of the response of the low-pass filter employed.

Figure 2.19 illustrates a form of CPFSK in which any discontinuity is very slight and the sign of the slope of the curve when the frequency changes is always the same, with regard to being positive or negative, before and after the change. This produces a frequency spectrum indistinguishable from that of CPFSK in which the frequency shift is distributed or continuous.

Figures 2.20, 2.21, and 2.22 show the percentage of power in a bandwidth (95 to 100 percent) plotted against the ratio of 3-dB bandwidth to data rate, with deviation ratio (0.5 to 0.7) as a parameter.

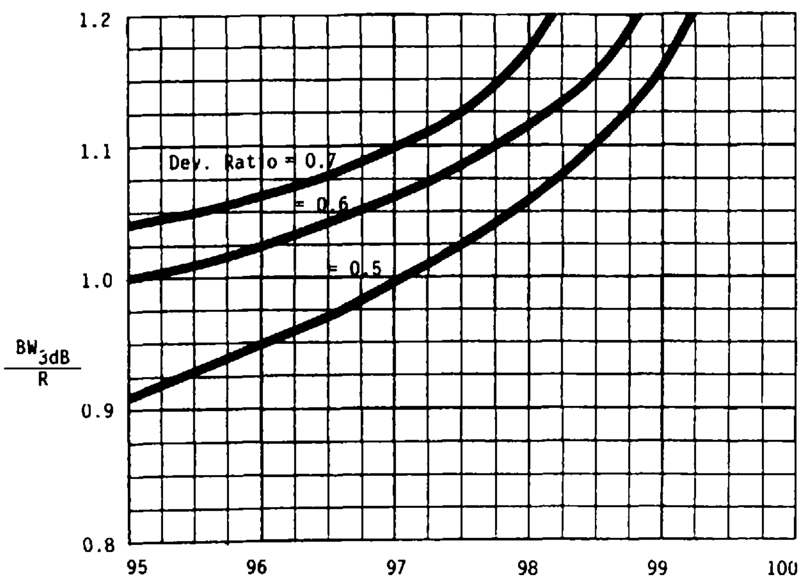


Figure 2.20 Percent of power in bandwidth (CPFSK, no filter).

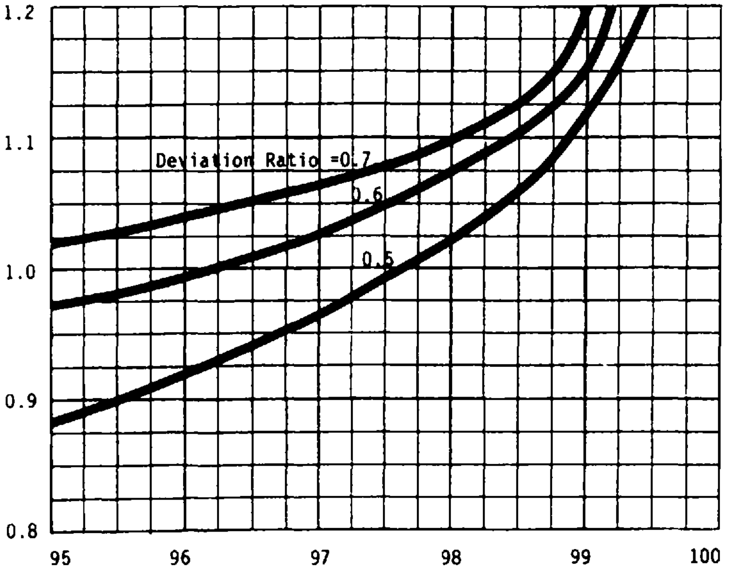


Figure 2.21 Percent of power in bandwidth (CPFSK, RF filter $BW_{3dB} = 2/T$).

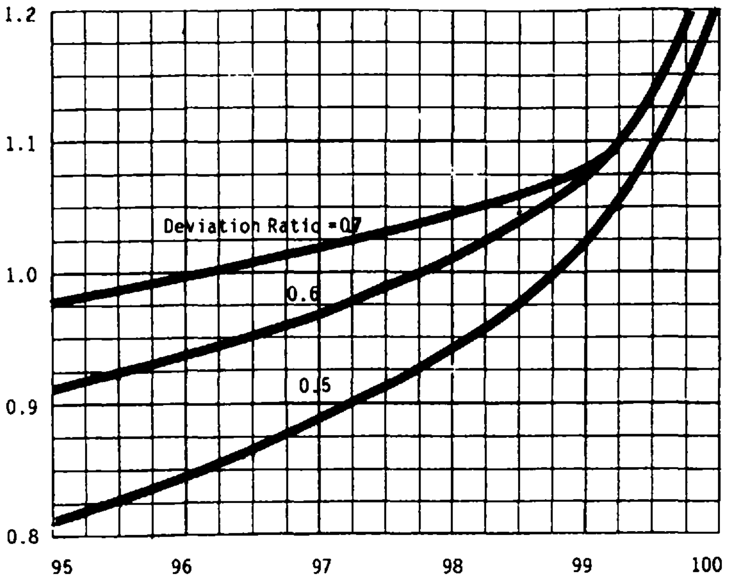


Figure 2.22 Percent of power in bandwidth (CPFSK, RF filter $BW_{3dB} = 1/T$).

2.8 PHASE SHIFT KEYING (PSK)

Phase shift keying and related techniques are the modulation methods most used today in the transmission of digitized information. Three of the PSK methods are most common; these are biphase phase shift key (BPSK or 2PSK), quadriphase phase shift key (QPSK or 4PSK), and offset quadriphase shift key (OQPSK).* Each technique has its advantages and disadvantages, which are

BPSK	Simplest to modulate or demodulate. Works best in linear channels.	Bandwidth (null-null) = $2R_{data}$ 3-dB BW = $0.88R_{data}$ Squaring demodulator, squaring loop, Costas loop
QPSK	Higher data rate than BPSK. Works best in linear channels.	Bandwidth (null-null) R_{data} 30dB BW = $0.44R_{data}$ Quadrupling demodulator, quadrupling loop, Costas loop
OQPSK	Same data rate as QPSK. Preferable to BPSK or QPSK in nonlinear channels.	Same BW as QPSK Less sideglobe regrowth Same demodulators as QPSK

Figure 2.23 shows the basic differences in these three techniques, which lie in the way in which phase shifts occur. Figure 2.23a shows that BPSK has 100 percent modulation at every phase shift, and Figures 2.23b and c show that QPSK and OQPSK have 100 percent modulation at half of the phase shifts and none of the phase shifts, respectively.

For clarity, all three modulation types in Figure 2.23 are shown with their clocks coincidental and the same data present. The BPSK data rate, however, is actually one half of either QPSK data rate. This is done to make a shift-by-shift comparison easier. The data being sent in all cases are

1001011111111001101110

although from these waveforms it is impossible to tell which is a one and which is a zero. Remember that the basic difference between BPSK and QPSK is that every clock time in a QPSK signal represents 2 bits of information whereas a BPSK signal represents only 1 bit per clock time.

QPSK (either simple QPSK or OQPSK) actually consists of a "symbol" in each clock time, where each symbol conveys 2 bits of information and each symbol is differentiated from other symbols by phase. Figure 2.24 shows phase diagrams of BPSK and QPSK signals.

---

*Also sometimes called modified quadriphase, staggered quadriphase, double binary, or one of several other names.

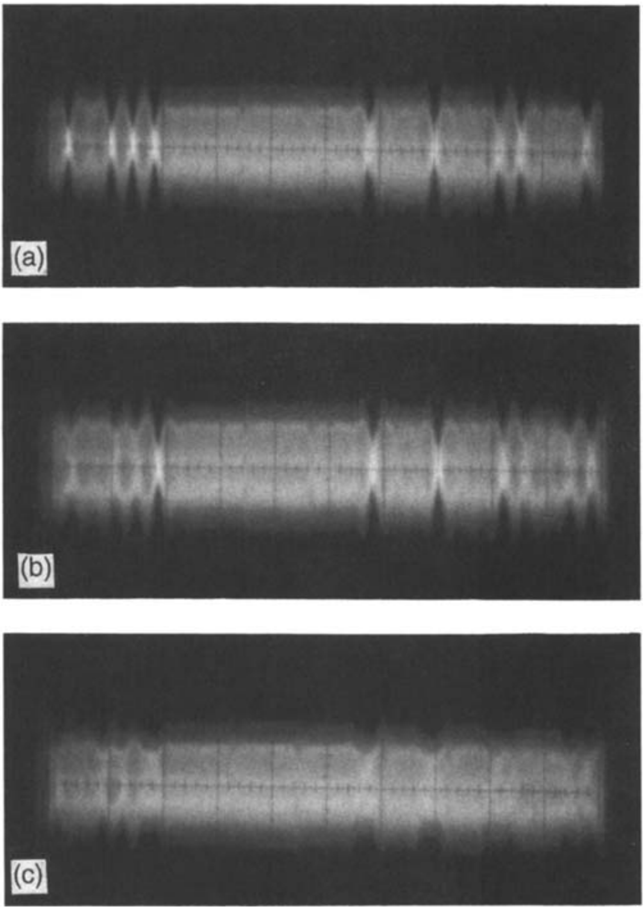


Figure 2.23 Time domain waveforms for biphase, quadriphase, and offset quadriphase modulation. Note difference in one-zero transition. (a) BPSK (2 μsec/division). (b) QPSK (1 μsec/division). (c) OQPSK (1 μsec/division).

The effect of incidental amplitude modulation is to determine the degree of sidelobe regrowth when either the channel or the receiver is nonlinear, with the regrowth being worse as the amplitude modulation is greater. Figure 2.25 illustrates the difference in sidelobe regrowth between OQPSK, in which the degree of incidental AM is 30 percent at every phase shift, and either BPSK or nonoffset QPSK. BPSK and QPSK both contain 100 percent amplitude modulation, which causes a sidelobe structure like that of Figure 2.25b.

Offset quadriphase modulation, unlike BPSK or QPSK, has no 180 degree phase shifts. Instead, all phase shifts are 90 degrees, which results in incidental

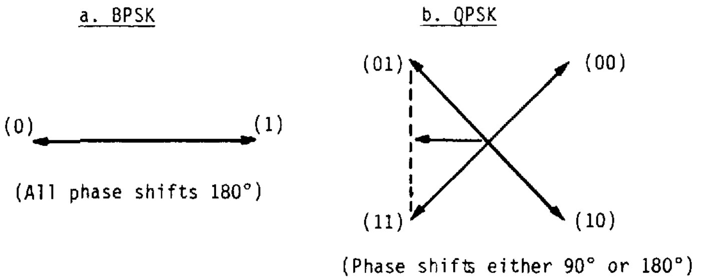


Figure 2.24 BPSK and QPSK phase diagrams. (a) BPSK. (b) QPSK.

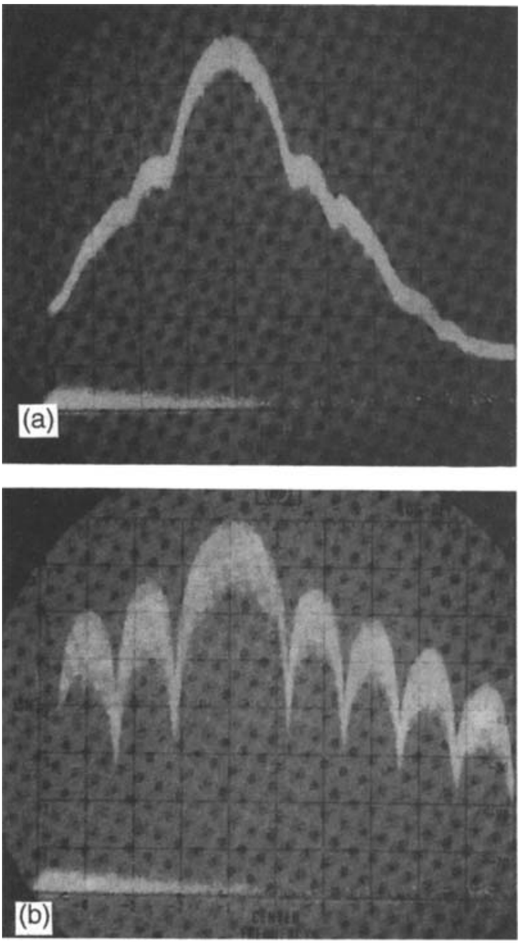


Figure 2.25 Spectra showing the value of offset keying. (a) OQPSK modulation after bandpass filtering ($BW = 2/T$) and hard limiting. (b) QPSK modulation after same filter and

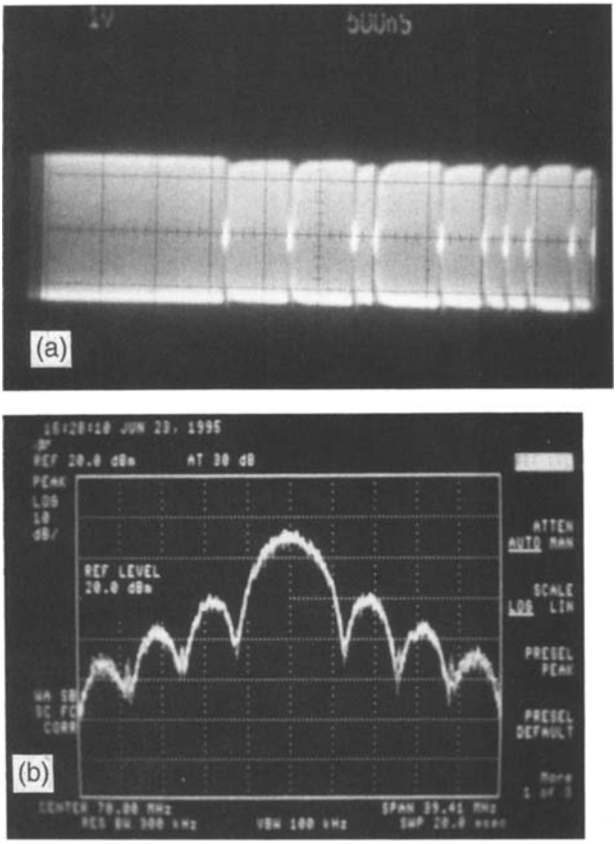


Figure 2.26 Time and frequency domain characteristics of an unfiltered BPSK modulated carrier. (a) Time waveform. (Note that all shifts respond fully in a single bit time.) (b) Frequency spectrum.

amplitude modulation of only 29.3 percent.* The result of this is that an OQPSK signal can be filtered and then passed through a nonlinearity with minimal sidelobe regrowth. Figure 2.25 shows the difference in sidelobe regrowth with OQPSK and QPSK signals that have been filtered and then hard limited, where both had the spectrum of Figure 2.25b before they were filtered.

Bandpass filtering phase shift keyed signals has the result that the rate of change of amplitude, at phase shifts, is slower as bandwidth narrows. As the

*As shown in Fig. 2.24b.

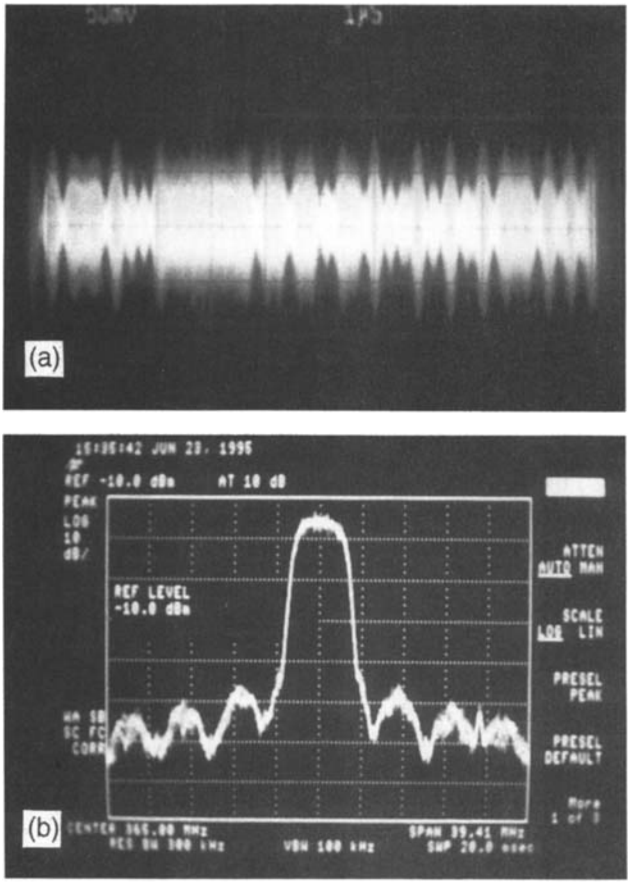


Figure 2.27 Time and frequency domain characteristics of BPSK, bandpass filtered (BW = $1/T$) signal. (a) Time waveform. (Note partial response of single bits due to restricted bandwidth.) (b) Frequency spectrum.

bandwidth is more and more restricted, the amplitude of the narrower pulses does not reach the full level. This is shown in Figures 2.26 and 2.27. In Figure 2.26, which shows an unfiltered BPSK signal, even the single bit times are long enough for full response. The filtered version shown in Figure 2.27 has bits that respond only partially.

As a rule, a BPSK signal that is filtered to a bandwidth $2/T$ (twice the BPSK chip rate) has single bit times whose amplitude is 0.707 times the amplitude of two or more bits in a row.

Table 2.6 PSK Bandwidths and Demodulators

PSK type	3 dBBW	Null-null BW	Demodulator
BPSK	$0.88R_{data}$	$2R_{data}$	Costas loop Squaring loop
QPSK	$0.44R_{data}$	R_{data}	Quadriphase Costas loop Quadrupling loop
OQPSK	$0.44R_{data}$	R_{data}	Quadriphase Costas loop Quadrupling loop
M-PSK	$\frac{0.88R_{data}}{\log_2 M}$	R_{data}	I-Q_M loop $(\ \ )^M$ loop

Modulated signal bandwidths for PSK signals and the demodulators used are listed in Table 2.6. In many cases, the modulated signal is filtered before transmission, which minimizes the bandwidth but in some cases demands a linear channel to maintain this narrow bandwidth. This also means that the receiver must be linear to avoid sidelobe regeneration.

Figure 2.28 shows the percent of power in a filtered BPSK signal's bandwidth, where the bandwidth is compared to the modulating bit rate.

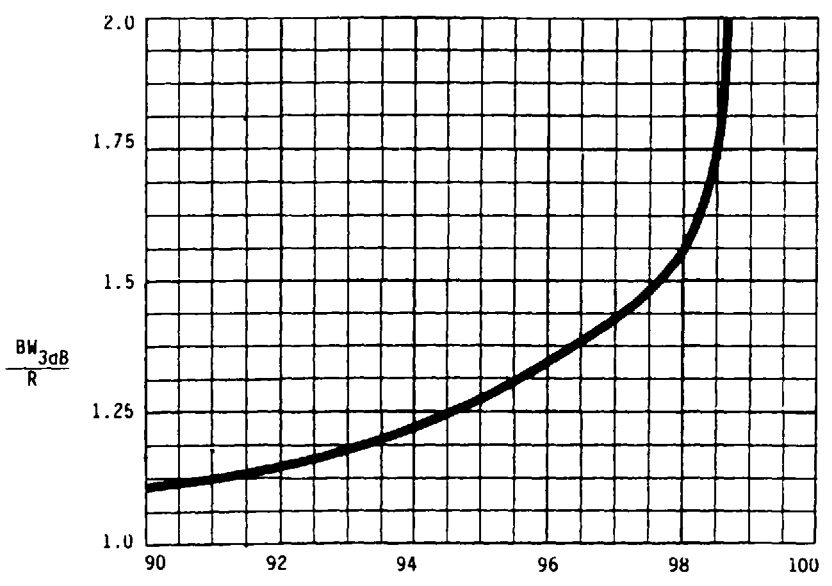


Figure 2.28 Percent of power in bandwidth (BPSK, RF filter $BW_{3dB} = 2/T$).

2.9 MSK AND OTHER EXOTIC WAVEFORMS

Minimum shift keying, or MSK, is the forerunner of a whole class of modulation forms that offer both low sidelobe energy and reduced sidelobe regrowth. These can be correctly considered to be forms of either frequency shift keyed or offset QPSK moddulation. MSK was originally patented in 1963 for the purpose of modulating high-power very low frequency (VLF) transmitters, where the need is for a modulation with minimal discontinuities. This property is valuable for reducing the rate of change of current in the large inductors used for tuning these transmitters but also has a bonus characteristic; the modulated signal bandwidth is also small compared with that for most other modulation techniques.

The MSK bandwidth at the 3-dB points is 0.6 times the data rate, and the bandwidth at the first nulls is 1.5 times the data rate. This is illustrated in Figure 2.29. CPFSK, with a deviation ratio of 0.5, and MSK are identical.

MSK modulation may also be generated as an offset QPSK signal in which cosine-shaped waves are substituted for the rectangular (or trapezoidal) waves that usually modulate the carrier in quadrature with one another. Figure 2.30 shows the difference between OQPSK and MSK modulators. It is of interest to note that the spectrum of an OQPSK signal at the first nulls is narrower than that of an MSK signal at corresponding points. Past the first nulls, however, the MSK signal's spectrum is always lower. A comparison of the two spectra is:

Modulation	Null-to-null BW	3-dB BW	First sidelobe	Rolloff
OQPSK	R_{data}	$0.44R_{\text{data}}$	-13 dB	6 dB/octave
MSK	$1.5R_{\text{data}}$	$0.6R_{\text{data}}$	-23 dB	12 dB/octave

Other exotic waveforms are also practical; their primary difference is in the structure of their sidelobes and their sidelobe levels. Because sidelobes are much

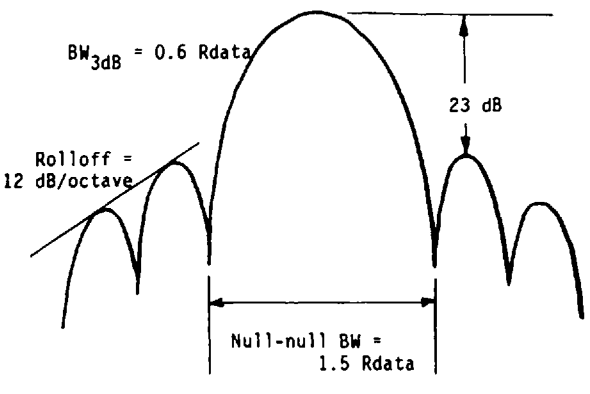


Figure 2.29 Characteristics of MSK spectrum. Random data modulated.

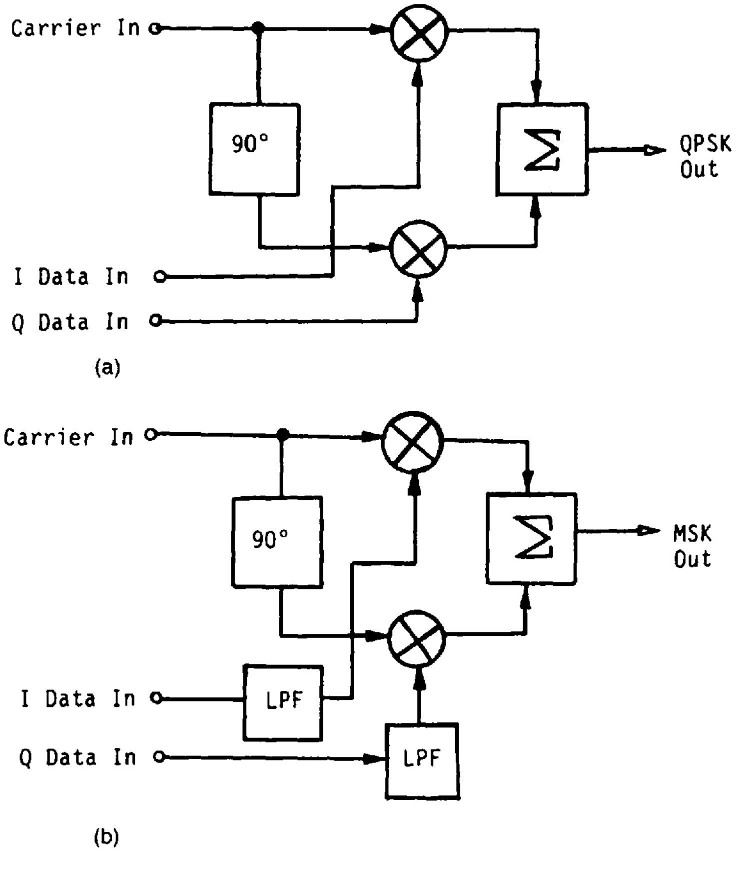
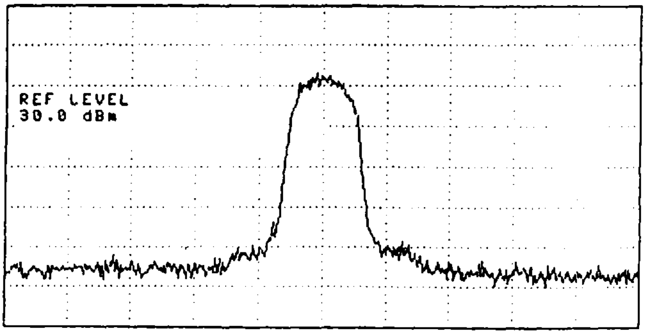


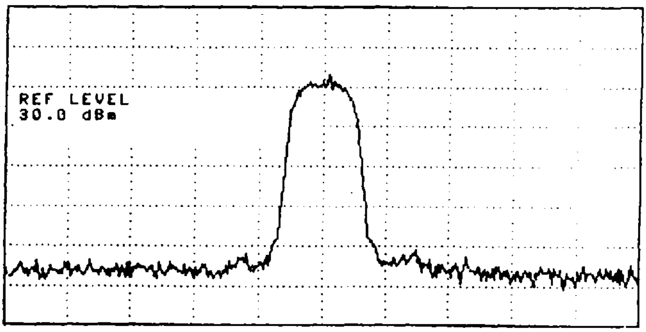
Figure 2.30 Illustration of OQPSK and MSK modulators. (a) OQPSK modulator (data rectangular). (b) MSK modulator (cosine-shaped data).

less critical to a receiver than to a transmitter, we will concern ourselves here with main lobe bandwidth. Receivers should have a bandwidth wide enough to pass the main lobe plus an allowance for frequency uncertainty and Doppler shift, which often cause the receiver to require a bandwidth 20 to 100 percent more than the main lobe bandwidth by itself.

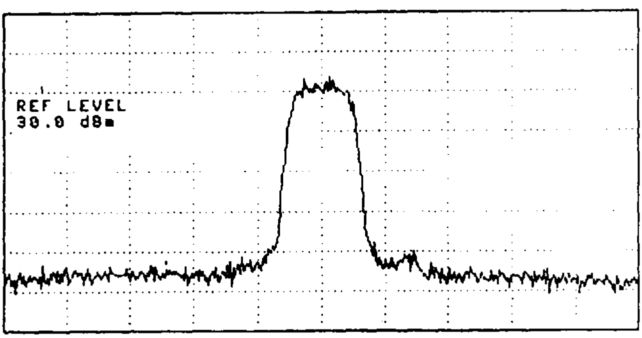
Figures 2.31 and 2.32 illustrate other waveforms that are used to minimize bandwidth and also to reduce the degree of sidelobe regrowth when the modulated signal is passed through a nonlinear amplifier (class C, for example). Figure 2.31 shows various PSK spectra, after filtering, and Figure 2.32 shows GMSK spectra as output in the 1900 MHz band by a PCS1900 GSM (General Systeme Mobile) transmitter.



(a)

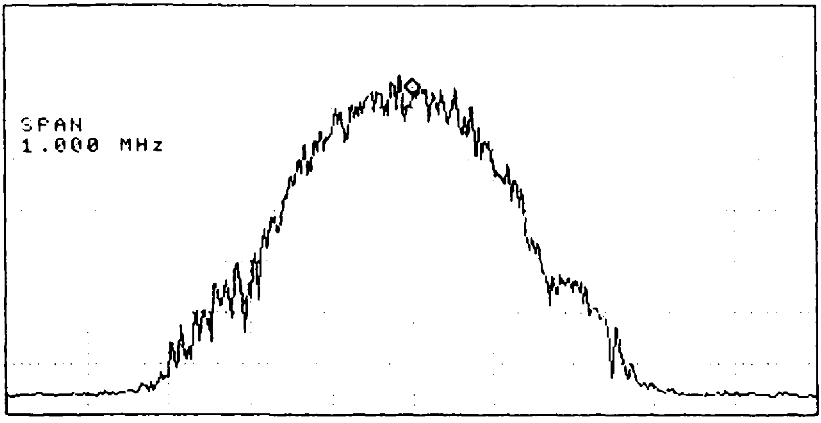


(b)

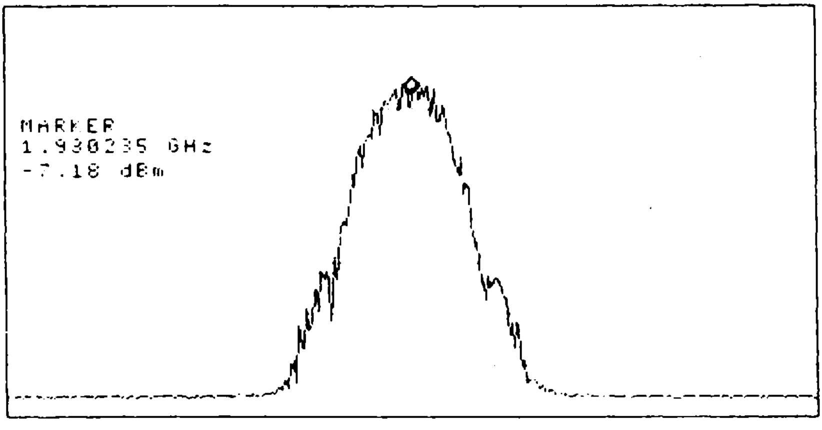


(c)

Figure 2.31 Bandpass filtered phase shift keyed signals at 1920 MHz. 5.0 MHz per division horizontal, 10 dB per division vertical. (a) Filtered BPSK modulation. (b) Filtered QPSK modulation. (c) Filtered OQPSK modulation.



(a)



(b)

Figure 2.32 GSM modulation as seen from a GSM transmitter. Data rate = 200 kbps. Vertical is 10 dB per division. (a) GMSK modulation. 100 kHz per division. (b) GMSK modulation. 500 kHz per division.

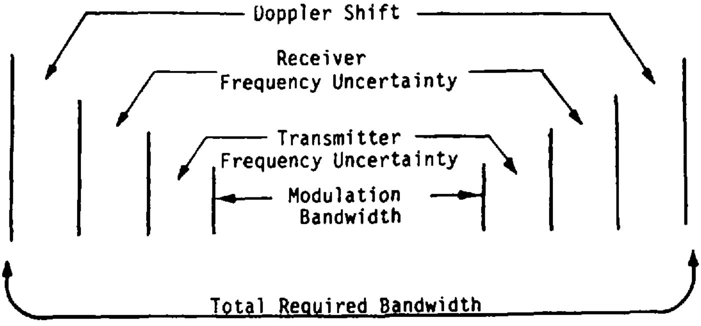


Figure 2.33 Illustration of components making up receiver bandwidth.

Demodulators for MSK and other related waveforms range from those that convert the signal to PSK and demodulate the signal as PSK to discriminators and *I*-*Q* loops. We will discuss these in detail in Chapter 8. Bandwidths for all are within the range of one times R_{data} to twice R_{data}, which corresponds to the main lobe (null-to-null) bandwidth.

2.10 RECEIVER TOTAL BANDWIDTH

The overall bandwidth required of a receiver's IF stages is made up of several components, That is, even though the modulation bandwidth to be received is the major determining factor, there are other significant factors that must be considered. Overall, the bandwidth is made up of

Modulated signal bandwidth
Transmitter frequency uncertainty
Receiver frequency uncertainty
Two times the worst-case Doppler shift

where the relationship between the components is illustrated in Figure 2.33.

This total bandwidth, expressed as an equation, is

$$B_{total} = B_{mod} + \Delta f_{TX} + \Delta f_{RX} + 2D$$

and is in many cases the same as the noise bandwidth of the receiver. That is,

$$B_n = B_{total}*$$

*Assuming that the filter used has at least three *LC* tuned sections or their equivalents.

Table 2.7 Doppler Shift Versus Velocity and Frequency

Velocity		Shift in Hz vs. frequency in MHz					
(mph)	(kph)	100	500	1000	2000	5000	10000
10	16.1	1.5	7.5	15	30	75	150
60	96.6	9.0	45	90	180	450	900
100	160.9	14.9	74.5	149	298	745	1,490
200	321.9	29.8	149	298	596	1,490	2,980
300	482.8	44.7	223.5	447	894	2,235	4,470
500	804.7	74.5	372.5	745	1,490	3,275	7,451
700	1,126.5	104.3	521.5	1,043	2,086	5,215	10,430
1000	1,609.3	149	745	1,490	2,980	7,450	14,900
2000	3,218.7	298	1,490	2,980	5,960	14,900	29,800
5000	8,046.7	745	3,725	7,450	14,900	37,250	74,500
10000	16,093.4	1490	7,450	14,900	29,800	74,500	149,000
20000	32,186.9	2980	14,900	29,800	59,600	149,000	298,000

and KTB_n is the equivalent receiver input noise power. Modulated signal bandwidth is as discussed in the preceding sections of this chapter.

Transmitter frequency uncertainty is a function of the transmitter's carrier source stability. The same is true of the receiver except that its local oscillator stability is the quantity in question. For typical transmitters and receivers, frequency accuracy is in the range of ± 1 to 10 parts per million, which is common in crystal stabilized systems.

Doppler frequency offset occurs any time there is relative velocity between a transmitter and a receiver, with frequency shifting upward (higher) if the distance between the two is decreasing with time and shifting downward (lower) if the distance is increasing with time. The frequency shift is a function of the number of wavelengths per second at a given velocity and frequency. That is,

$$F(d) = \frac{\text{velocity}}{\text{wavelength}} = \frac{v}{\lambda}$$

which is tabulated in Table 2.7.

QUESTIONS

1. What is the difference between analog and digital modulation?
2. What is the difference between NCFSK and CPFSK?
3. What deviation ratio is used in the classic form of MSK?

4. What is the bandwidth required by AM, DSBSC AM, and SSB AM? (Neglect Doppler shift and frequency uncertainty.)
5. What is the level of the second sideband pair in an FM signal with $\beta = 7$?
6. What signal-to-noise ratio gain should one expect when β is increased from 1.0 to 3.0 in an FM system?
7. What is the null-to-null bandwidth of an on-off keyed signal? BPSK?
8. What incidental AM percentage should be expected from BPSK? QPSK? OQPSK?

REFERENCES

1. Tjhung, T. T., Band occupancy of digital FM signals, *IEEE Trans. Comm., Dec. 1964.*
2. Oetting, J. D., A comparison of modulation techniques for digital radio, *IEEE Trans. Comm., Dec. 1979.*
3. Schwartz, M., *Information Transmission, Modulation, and Noise*, John Wiley & Sons, New York, 1959.
4. DeLozier, J. F., Digital Transmission of Analog Signals, Watkins-Johnson Tech-Notes, Sept./Oct. 1981.
5. Watson, B. (R.), FSK: Signals and Demodulation, Watkins-Johnson Tech-Notes, Sept./Oct. 1980.
6. Steber, J. M., PSK Demodulation, Part 1, Watkins-Johnson Tech-Notes, May/June 1984.
7. Dixon, R. C., *Spread Spectrum Systems*, John Wiley & Sons, New York, 1994.
8. Proakis, J. G., *Digital Communications*, McGraw-Hill, New York, 1983.

This page intentionally left blank

3
Receiver Configurations

Many configurations have been developed for use in receivers, although the great preponderance of receivers today are of one general type. The remainder are divided between two other types. The three receiver types are, respectively, superheterodyne, tuned radio frequency, and regenerative. There are many variations on these basic types, which will be discussed in the following pages, but we will concentrate primarily on superheterodyne receivers and their variations.

3.1 TUNED RADIO FREQUENCY (TRF) RECEIVERS

As the first receivers were TRF receivers, they will be the first ones discussed here. A TRF receiver is one in which a received signal is demodulated directly at the frequency at which it is received. In the early days of radio broadcasting, such receivers were the only ones available, and it is fortunate that there were not very many broadcast stations available, because of the difficulty of tuning these early receivers. The only way to improve a receiver's selectivity at that time was to employ several (usually up to five) cascaded stages individually mechanically tuned. (See Figures 3.1 and 3.2.)

Attempts were made to couple the multiple stages together mechanically so that all tuned together, but they were generally unsuccessful. The primary reasons for this lack of success were 1) difficulty in getting the multiple tuned stages to track accurately and 2) mechanical play in the tuning mechanism.

It is just as well that a tunable TRF receiver's selectivity was so difficult to improve upon, as this difficulty was what led to the invention of superheterodyne receivers. TRF receivers could not meet the need for the higher and higher frequencies that became necessary as more and more wireless systems with narrower and narrower bandwidths came about. In the United States, the AM broadcast band covers 550 kHz to 1.6 MHz—a range of approximately 1.5 octaves. Trans-

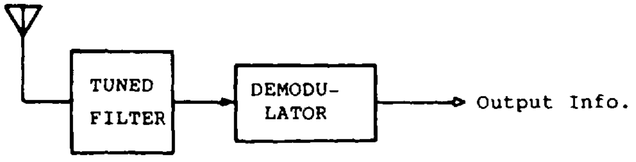


Figure 3.1 Simple TRF receiver.

mitters are spaced at 10.0 kHz within this band, which requires a receiver to be able to select from signals whose channel centers may be only 10 kHz apart and whose bandwidth is such that the outer channel edges abut one another. With a Q of 100, the tunable stages of a TRF receiver at midband would have a 3-dB bandwidth of

$$B_{3\mathrm{dB}} = \frac{1.0 \times 10^6}{100} = \frac{f}{Q} = 10 \text{ kHz}$$

which is enough to pass the modulated signal bandwidth (10 kHz) but lacking enough selectivity to reject a signal in the next channel without using many tuned stages. (For this reason, AM radio stations are not usually assigned to adjacent frequencies within the same geographic area.)

FM broadcast in the 100-MHz band as we know it today would be impractical using TRF receivers, as FM signals are separated by 200 kHz and their modulated signal bandwidth is 150 kHz. A TRF receiver for this band would require multiple tunable stages with a Q of at least

$$Q = \frac{9.8 \times 10^7}{1.5 \times 10^5} = 653$$

when tuned to midband (about 98 MHz). A more practical Q of 100 to 200 would produce a 3-dB bandwidth of 490 to 980 kHz, which means that two to four FM stations would be present in the bandwidth of the TRF receiver. Broadcast FM

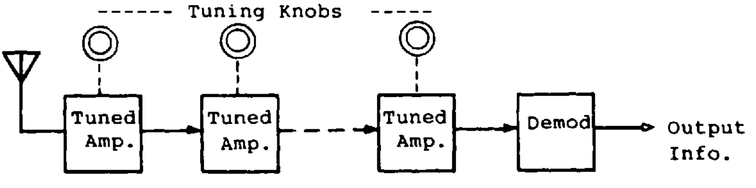


Figure 3.2 Multituned TRF receiver using multiple tuned amplifiers.

radio therefore could not exist as we know it today if it were necessary to depend on TRF receiver techniques.*

This does not mean that TRF receivers are not used today. It simply means that their use is generally restricted to applications in which selectivity is not critical. In fact, as we shall see, every superheterodyne receiver carries a TRF receiver in its heart.

3.2 THE SUPERHETERODYNE TECHNIQUE

Rarely has such an elegant solution to a difficult problem been possible as in the case of the superheterodyne technique for building a receiver. Today, the answer to the question "How can I build filters to provide good selectivity over a widely varying frequency range when Q is limited and also varies inversely with frequency?" may seem obvious. In the late 19th and early 20th centuries, this was not the case. This author is grateful to Edwin Armstrong for conceiving and developing the superheterodyne technique.

Let us suppose that we are in possession of a TRF receiver with all of the desirable properties of the best of receivers, except for one—our TRF receiver cannot operate with equal performance over a wide range of frequencies. Both sensitivity and selectivity vary as we change the frequency.

The solution to our problem is to continue to use our TRF receiver, but to add a frequency converter that shifts the signals we wish to receive to the correct frequency for the TRF receiver. This is the simple principle upon which a superheterodyne receiver works: desired signals are heterodyned to a fixed frequency for amplification, selection, and demodulation. Figure 3.3 shows a block diagram of a simple superheterodyne receiver, which operates as follows:

1. An incoming signal is amplified in a tuned amplifier at the input frequency. (In some superheterodyne receivers this amplifier and/or tuning is deleted.)
2. The amplified, preselected signal is multiplied with a tuned "local oscillator" and the difference between the input frequency and the local oscillator is selected. This difference frequency, which now contains the modulation from the input signal, is called the intermediate frequency.
3. The intermediate frequency amplifier, which usually has very selective filters and high gain, provides the signal received, at the proper level and bandwidth, to the demodulator.

---

*In fact, television, cellular telephones, PCS systems, and all other radio devices that use frequencies greater than a few MHz could not exist as we know them if the superheterodyne receiver had never been invented.

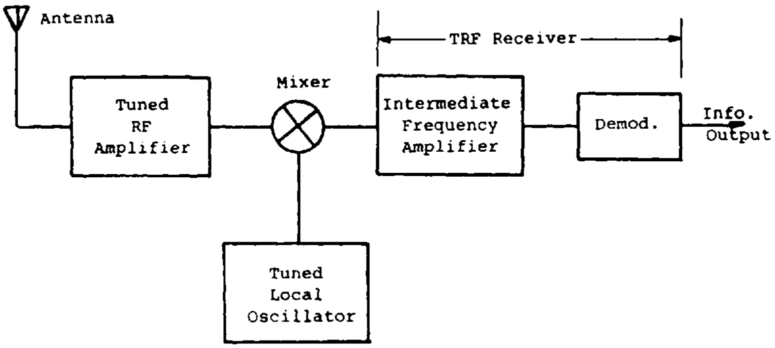


Figure 3.3 Basic superheterodyne receiver.

The intermediate frequency* amplifier and demodulator make up a complete TRF receiver. In this case, however, they operate at a fixed frequency, which can be chosen for best performance and optimal filter design. This in itself is a significant advantage over a TRF receiver that must operate over a wide tuning range.

Superheterodyne receivers outperform any other type of receiver except in a very few specialized applications. We shall return to discuss superheterodyne receivers in much more detail in Chapter 4.

3.3 REGENERATIVE AND SUPERREGENERATIVE TECHNIQUES

In the early days of radio, superheterodyne receivers proved to be so superior to TRF receivers for broadcast applications that those who were not privy to the superheterodyne patents were forced to try other approaches. One of the most successful of these alternative approaches was the regenerative receiver. Figure 3.4 shows a block diagram and a schematic diagram for the part of a regenerative receiver that is of most interest; the regenerative detector.

The idea behind the regenerative detector is to provide an amplifier (see Figure 3.4a) that is allowed to oscillate at the frequency of the desired signal. This is done by either tuning the amplifier as in Figure 3.4b or tuning the feedback as in Figure 3.4a. However, when an input signal at the desired frequency is not present, the feedback level is designed to be inadequate to cause regeneration.

*Conventionally abbreviated IF, which is the designation that will be employed in this book from this point forward.

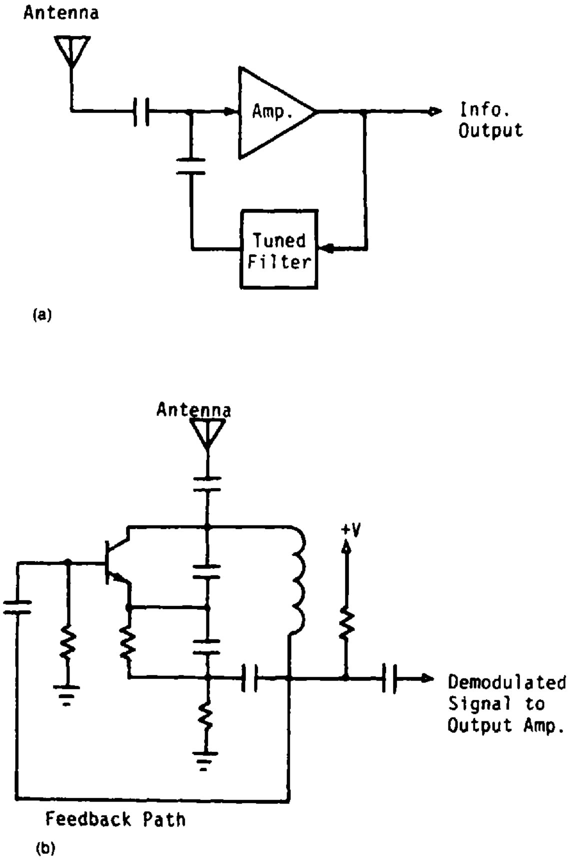


Figure 3.4 Schematic diagram of typical regenerative receiver (output amplifier not shown.) (a) Block diagram of a regenerative receiver. (b) Schematic diagram of regenerative detector as used in Radio Shack Space Patrol Walkie-Talkie. (Radio Shack and Space Patrol are registered trademarks of Tandy Corporation.)

The problem, of course, is in the stability of the circuit. If the feedback level is too high, the circuit will oscillate without an input signal. If, on the other hand, the feedback level is too low, the circuit is ineffective. Because it is very difficult to provide a regenerative detector that will work well under conditions of component aging, temperature extremes, power supply change, etc., regenerative and superregenerative detectors are not usually used in critical applications.

An attempt to control the stability of regenerative detectors resulted in the superregenerative detector, which employs a "quenching" circuit to periodically stop any oscillation. This is done by simply grounding or opening the feedback path at intervals. It allows the feedback level to be set higher than is possible without quenching.

3.4 COMPRESSIVE (MICROSCAN) RECEIVERS

A special form of superheterodyne receiver, used primarily for signal detection at or below the thermal noise level, is the compressive type. (Compressive receivers are also called microscan receivers.) These receivers make use of a spectrum-spreading method called "chirp" or "pulse-FM," as illustrated in Figure 3.5. Instead of using a fixed-frequency local oscillator as in most superheterodyne receivers, compressive receivers use a swept local oscillator. This local oscillator has a frequency range that starts at f_1 and sweeps linearly to

$$F_2 = T\frac{df}{dt}$$

where T is the duration of the sweep and df/dt is the rate of change of frequency with time (i.e., the sweep rate).

When a swept local oscillator signal such as this is multiplied with a received signal, the result is a swept signal at the IF frequency. (Here we are considering only that the received signal is a carrier. If the received signal were modulated, then the result would be a modulated signal whose carrier frequency is swept at the local oscillator sweep rate.)

After amplification,* the swept signal is applied to a compressive (or dispersive) filter whose characteristic is such that its time delay is a function of frequency. That is, its dt/df (time delay per hertz of frequency shift) is the reciprocal of df/dt when the filter is "matched" to the signal. This means that a signal that is caused to sweep, by mixing with the swept local oscillator, and therefore is matched to the internal IF compressive filter, is output from the filter as a pulse. Its power, which was spread over a wide bandwidth by the sweeping process and also over a time T, is compressed into a time τ. The period of τ is the reciprocal of the frequency range of the sweep, Δf.

Thus, because the signal power is compressed in time, its amplitude increases by an amount approximately equal to T/τ, and the pulse signal-to-noise ratio is greater than the input signal-to-noise ratio by an amount

$$\sqrt{T/\tau}$$

which is called the compression ratio or dispersion factor D.

*See Figure 3.5.

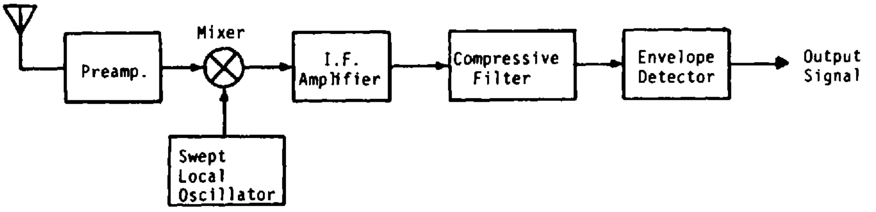


Figure 3.5 Compressive or microscan receiver block diagram.

This is exactly the same as the principle by which "chirp" Radars work, except that in the case of the Radar the transmitter sends the swept signal and the receiver contains the compressive filter. In our case, a compressive receiver, the receiver itself both generates and detects the swept signal.

3.5 RADIOMETRIC RECEIVERS

Radiometric receivers are extremely sensitive. They are employed for detection of very weak signals, such as those emanating from radio stars light years away. Their principle of operation is to compare the energy level that is incident on an antenna with a reference that is calibrated to be equivalent to no signal at all. If the signal from the antenna is greater than the reference level, then a signal source must be present.

Figure 3.6 illustrates the evolution of such a receiver. Although a radiometric receiver cannot distinguish between signals using different kinds of modulation (it depends only on the received signal power level), it can detect signals at lower levels than any other receiver.

3.6 SPECTRUM ANALYZERS

Although spectrum analyzers are very expensive laboratory instruments, and invaluable for their abilities, they are just another superheterodyne receiver. Their main distinguishing feature is that they are much more flexible than any other receiver. Some of their characteristics are

Very wide frequency range
Adjustable observation bandwidth
Adjustable IF bandwidth
Adjustable sweep rate
Adjustable postdetection bandwidth

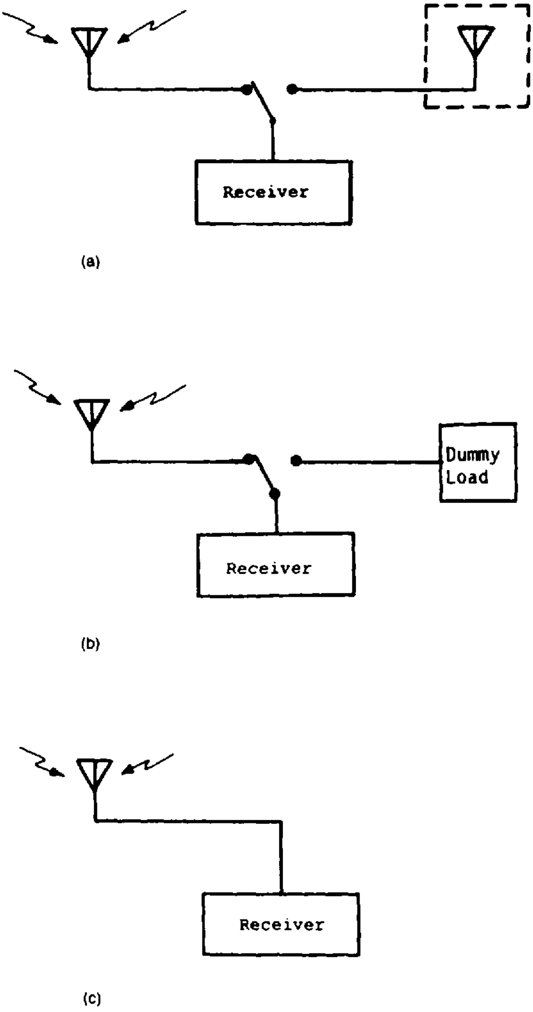


Figure 3.6 Radiometric receiver configurations. (a) Receiver compares signal from two antennas, one shielded. (b) Shielded antenna replaced by dummy load. (c) Receiver compares signal to internal reference.

Variable gain
Availability of a coupled preselector
Programmability
Built-in visual spectrum presentation

With all of these variable parameters, spectrum analyzers tend to be very expensive, although they are not always the best of receivers—especially when they have no preselector.

Except for a few spectrum analyzers at lower frequencies (usually less than 10 MHz), spectrum analyzers are always superheterodyne. At the lower frequencies, there are some spectrum analyzers that employ a direct Fast Fourier Transform (FFT) processor.

Figure 3.7 is a block diagram of a superheterodyne spectrum analyzer, in this case a "double conversion" receiver, because it converts the input signal twice and has two intermediate frequencies.

In many cases there may also be more than one first IF, because it is desirable that this IF is outside the swept frequency range, and it is not unusual for a spectrum analyzer to cover a range of 10 MHz to 40 GHz or even more. A pair of first IF amplifiers that are switched automatically, depending on the band being viewed, is often used. Let us say that the lowest band is 0.1 MHz to 2 GHz and the first IF is at 2.5 GHz (the swept local oscillator could sweep from 2500 to 4500 MHz). Then, when the band is changed to above 2 GHz, the first IF would change to 500 MHz, so that the same swept local oscillator would now allow an input frequency of 2000 to 4000 MHz.*

Spectrum analyzers employ swept local oscillators, like compressive receivers, but do not use a compressive IF filter. Instead, spectrum analyzers use bandpass filters that determine the received bandwidth and the resolution. Resolution is the ability of a receiver (in this case a spectrum analyzer) to detect and separate signals that are close together in frequency, and is a function of both the IF bandwidth and the local oscillator sweep rate.

Spectrum analyzers usually employ a noncoherent amplitude detector, and the display is simply a plot of signal amplitude versus frequency, without phase information. This is the reason that FM spectrum plots seen on a spectrum analyzer do not look the same as the ones derived from a Bessel function chart (see Chapter 2).

Recently developed "frequency domain" or "modulation domain" analyzers substitute a discriminator for the spectrum analyzer's envelope detector.

---

*Note that in neither case does either the input signal range or the swept local oscillator fall at the first IF frequency.

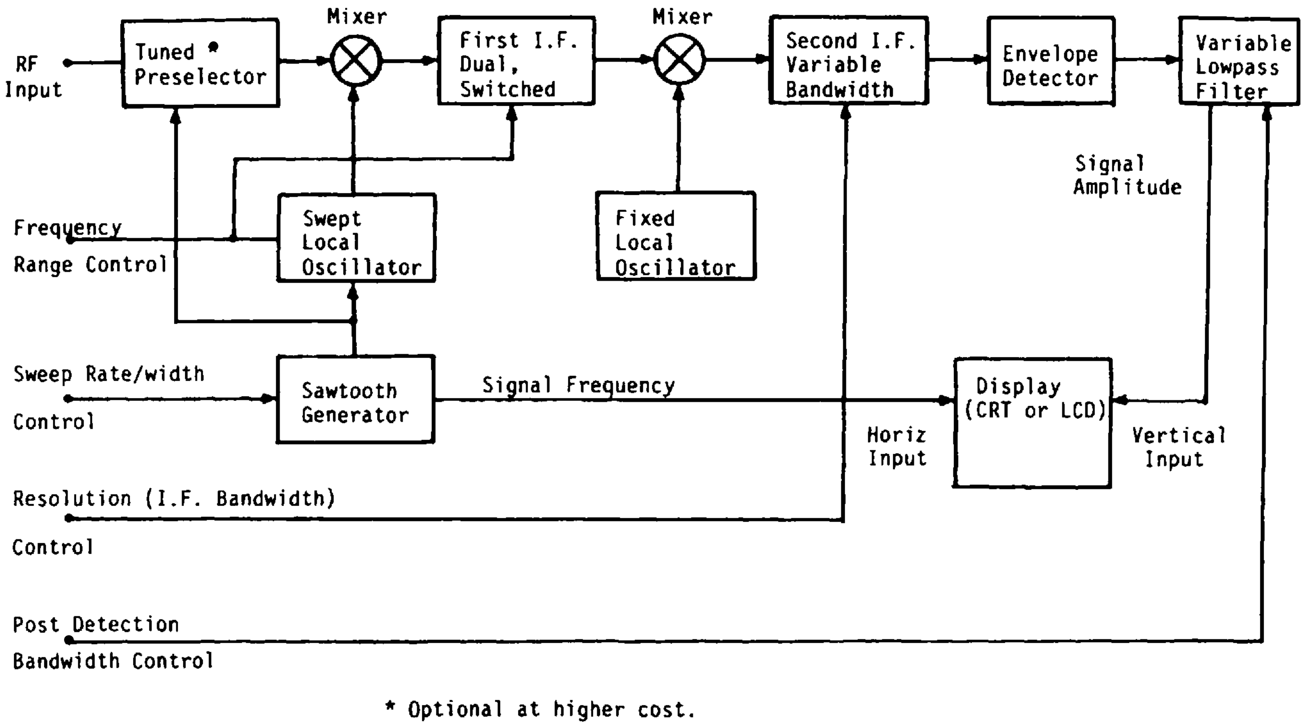


Figure 3.7 Simplified superheterodyne spectrum analyzer block diagram.

Figure 3.8 Crystal video receiver.

3.7 CRYSTAL VIDEO RECEIVERS (see Figure 3.8)

"Crystal video" receivers are really a modernized version of a TRF receiver. They are employed today primarily in extremely wideband signal monitoring applications, where the probability of detection must be high. Suppose that we wish to cover a range of 100 kHz to 20 GHz, that we must detect a signal that consists of a 1-μsec pulse that appears randomly at long intervals, and that the probability of detection* must be 99 percent or better.

A single spectrum analyzer is not a good choice for such an application, simply because it must cover the entire sweep range in 99 out of every 100 μsec, and multiple spectrum analyzers would be far too expensive and would probably interfere with one another. (The real problem with spectrum analyzers is that their instantaneous observation range is the bandwidth of their IF filter. They cannot observe a wide bandwidth all at the same instant of time.)

Crystal video to the rescue! It is not unusual in such a system to employ multiple crystal video (TRF) receivers, each covering perhaps one octave of the band to be observed (as shown in Figure 3.9). In this way, a group of filters, typically each with about one octave in bandwidth, can be used to cover a very wide bandwidth.

The individual crystal video receivers (Figure 3.8) consist of a fixed-frequency filter, an amplifier, and a demodulator. The demodulator is chosen to provide the information being sought and is usually some combination of AM and/or FM demodulators.

With the bandwidth of these receivers, the tendency of the overall system is for the system sensitivity to decrease by 3 dB per octave, because the receiver bandwidth doubles with each octave in frequency. This does not necessarily mean,

---

*Probability of signal detection is the joint probability

$$R(\text{det}) \equiv \text{DFT} \times \text{PA} \times \text{PTH} \times \text{PF}$$

where

DFT ≡ the transmitter's duty factor, $0 < \text{DF} < 1$

PA ≡ probability that the antenna is pointing in the right direction; PL = beamwidth/360

PTH ≡ probability that the signal is large enough to be above the receiver threshold

PF ≡ probability that the receiver is looking at the right frequency

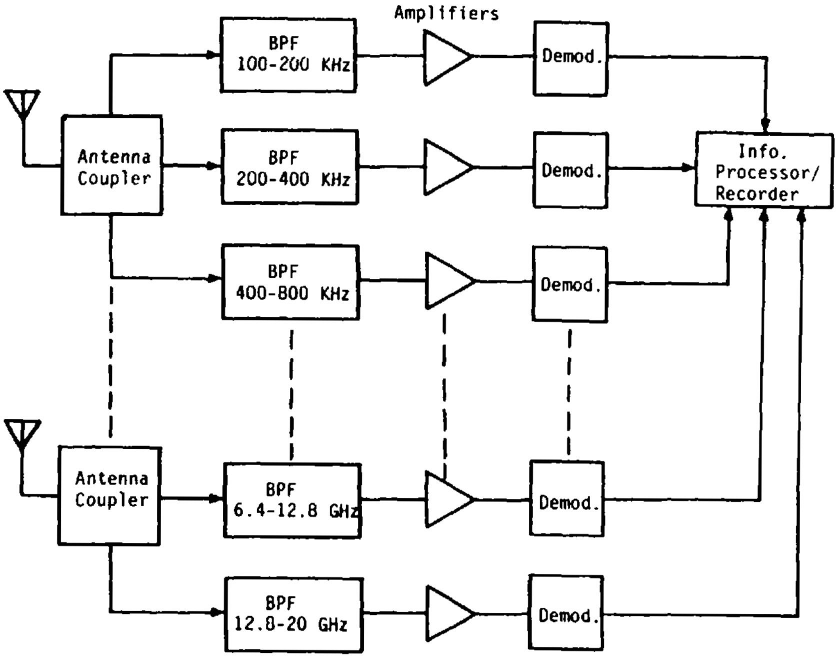


Figure 3.9 Extremely wideband receiver configuration employing multiple (eight) crystal video receivers. (Each receiver covers one octave.)

however, that the higher frequency receivers have a significantly poorer reception range. As the frequency increases, antenna size decreases, and their gain and efficiency tend to increase. This makes up for the reduced bandwidth of the lower frequency receivers in the system.

3.8 INSTANTANEOUS FREQUENCY MEASUREMENT (IFM) RECEIVERS

Another special-purpose receiver, the IFM receiver, deserves mention, although there are currently newer receivers (the frequency domain or modulation domain analyzers) that can perform their job as readily. The task of an IFM receiver is to measure* or monitor the frequency of a signal even as it changes. Examples of

*IFM is also used to designate instantaneous frequency monitoring receivers. (Note the "monitoring" instead of "measuring.")

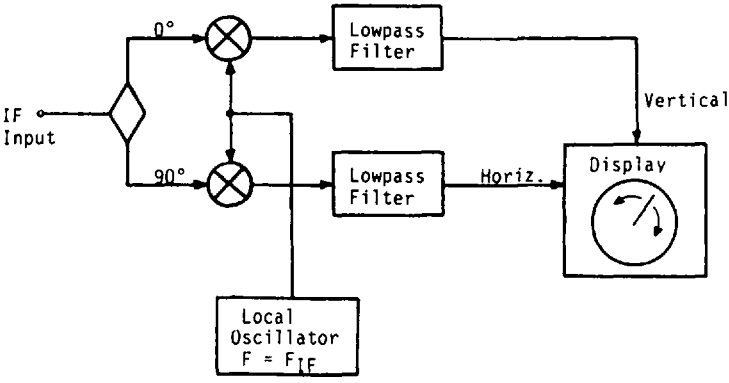


Figure 3.10 IFM receiver (signal processing portion).

such signals are FM signals, frequency agile signals, chirp signals, and frequency hopping signals.

Although IFM receivers use, from an overall standpoint, superheterodyne architecture, the IF signal is processed as shown in Figure 3.10. At the IF frequency, the signal is split, shifted by 90 degrees, and both zero-shift and quadrature signals are multiplied with a local oscillator at the IF frequency.

This produces in phase and quadrature signals that, after lowpass filtering, are fed to the vertical and horizontal plates of a cathode ray tube. This causes the CRT to display a vector whose length is determined by the signal's amplitude and whose angle is a function of the signal's frequency.

Such receivers typically have very fast response and can be used to monitor the most agile signals. Their chief drawback is that it is difficult to monitor more than one signal at a time.

3.9 CORRELATIVE RECEIVERS (DETECTORS)

A technique sometimes employed to detect the presence of a coherent signal in noise is a correlation technique in which a signal and a delayed replica of that signal are multiplied together. Then the product of the two (delayed and undelayed) signals is integrated. This process is a classic correlation.

In this case $g(t)$ is an arbitrary function of time that is repetitive at interval τ. When such a signal is present, the integrated output of the multiplication process can readily be distinguished from that which occurs when random noise or signals without coherence across an interval shorter than τ are present. A simple example

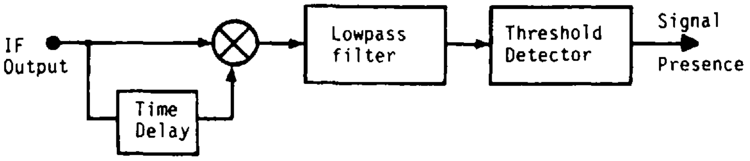


Figure 3.11 Correlative detector used to detect the presence of signals that are coherent across a period of time.

of a signal that can be detected is a single tone that does not change for a period τ or longer.

Spread spectrum receivers often employ the same principle, except that the receiver employs a stored replica of the desired signal, rather than multiplying the received signal with a delayed replica of itself. (See Section 3.10, which follows.)

Figure 3.11 illustrates a correlative detector such as that used in a correlative receiver. The remainder of the receiver would usually be in a superheterodyne configuration, with the correlative detector operating at an IF frequency.

3.10 SPREAD SPECTRUM RECEIVERS

Three basic types of spread spectrum signals are generally considered. Each can also be combined with another kind of modulation, and this is the usual way in which they are used. Although it is possible that a spread spectrum receiver may have to deal with more than one kind of modulation at the same time, one type of modulation is usually associated with the process of spectrum spreading, while another type of modulation conveys the information to be received.

A spread spectrum system is one in which significantly more bandwidth is employed to send the signal than is implied by the information rate that is to be received. This usually means that a spread spectrum receiver has radio frequency (RF) stages that are much wider in bandwidth than its IF amplifier stages and demodulator. This is in concert with the second characteristic of spread spectrum systems; namely, the RF bandwidth is determined by some function that is independent of the data or other information that is being sent.*

Types of spread spectrum signals that we will consider are listed in Table 3.1. Spread spectrum receivers are, like most other receivers, seldom of any type other than superheterodyne. They do, however, add a process that is not normally needed in other more coventional receivers—namely, a despreading process,

*The spreading function is in the usual case a code sequence, designed for noiselike characteristics and separation from other sequences.

Table 3.1 Spread Spectrum Signaling Techniques

Signal type	Spectrum-spreading characteristics	Data modulation characteristics
Direct sequence	Modulation is identical to data transmission systems (BPSK, QPSK, OQPSK, MSK, etc.). Code replaces the usual data to the RF modulator.	Data are FSK or PSK.
Frequency hop	Frequency of operation is chosen by an internal code reference generator.	Data are FSK.
Chirp	Frequency is swept (linearly in most cases) over a range Δf in period T.	Data one = upsweep Data zero = downsweep
Hybrid	A combination of two or more of the above.	Data modulation depends on the spreading type.

wherein the spectrum spreading modulation is stripped away. This leaves the information modulated carrier to be demodulated with the spectrum spreading fuction no longer present.

Spread spectrum techniques, and the receivers necessary for their implementation, are used for a number of reasons. Almost all military communications systems use some spread spectrum method. (Most military systems employ a technique called "frequency hopping.") Many consumer and commercial applications for spread spectrum methods are also rapidly developing. Some of the reasons for using spread spectrum techniques are the following:

- Code division multiple access (CDMA) is made possible.
- Selective addressing is practical, using code selection and/or rejection.
- Spread spectrum ranging (as in the Global Positioning System) is both accurate and robust.
- The power density of spread spectrum signals is lower than that of conventional narrowband signals, which can reduce interference to other systems.
- Privacy is inherent in spread spectrum signals, although encryption is required if true security is desired.
- Spread spectrum receivers can be designed to operate in the presence of either deliberate or inadvertent interference or both.

Today, spread spectrum signals are being used under the pretense that more spread spectrum signals than narrowband signals can be put into a given band. Unfortunately, the physical laws prevail, and this particular fantasy is not true.

Another unfortunate misunderstanding with respect to spread spectrum systems is that designers are sometimes led to believe that spread spectrum

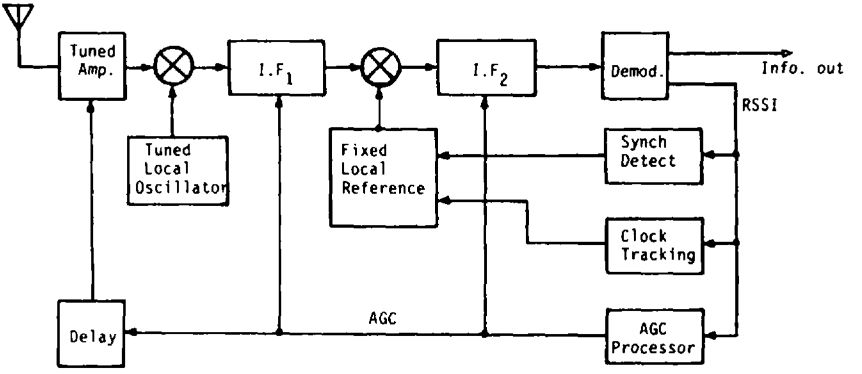


Figure 3.12 Typical spread spectrum receiver block diagram.

receivers are inherently more sensitive than conventional receivers. This misunderstanding is due to the fact that receivers such as those employed in the Global Positioning System (GPS) can and do work at signal levels within their spread spectrum signal bandwidth that are far below the noise level. For example, the noise bandwidth of a GPS receiver is either 1.023 or 10.23 MHz, depending on whether it is designed for military or civilian applications. Therefore, KTB for GPS receivers is either −114 dBm or −104 dBm. The signal level at the GPS receiver input is only −135 dBm, however, which means that the signal-to-noise ratio at the input to a GPS receiver is in the range of −21 to −31 dB! This is far beyond what a conventional receiver can do (at least on the surface).

The truth is that a spread spectrum receiver must work under the same link budget constraints as any other receiver. Although it is true that the spread spectrum receiver can work at a negative signal-to-noise ratio within its RF bandwidth, it cannot do so within its information bandwidth. The GPS receiver's process gain* trades bandwidth at the RF frequency for signal-to-noise ratio in the information bandwidth. The process gain in a GPS receiver is approximately 40 to 50 dB, and when the RF input signal-to noise ratio is −20 dB in a 10.23-MHz bandwidth, it is 50 − 20 = 30 dB in the information bandwidth (which is only 50 Hz).

In a typical spread spectrum receiver (see Figure 3.12) the desired signal is

*Process gain in a spread spectrum system is G_p = 10 log(RF bandwidth/data rate) and is usually expressed in dB. Typical spread spectrum receivers have process gain in the range 10 to 30 dB, because of their information rate and the limitations imposed by frequency allocation. GPS has a process gain of 50 dB because of its low data rate.

amplified and converted to a first IF frequency just as in any other receiver. Any tuning necessary is done by the preamplifier and local oscillator. In the case of spread spectrum receivers, it is very important that the frequency of the first IF be higher than the bandwidth of the preamplifier. Linearity is also very important. The reason for both is to prevent unwanted vulnerability to interference, which we will discuss in the following pages. This concern is especially important in spread spectrum receivers that are employed to resist either deliberate interference or inadvertent noise.

The despreading operation is usually performed at a second mixer, primarily because it is more practical to design the local refernce at a fixed frequency than to make it variable. The local reference itself consists of a code generator and a modulator that generates a replica of the desired, received signal. (If the desired signal is a direct sequence signal, then the local reference is also direct sequence. If the desired signal is frequency hopped, then so is the local reference. The code employed in both cases is the same as that used by the transmitter to spread the signal.)

Following the mixer in which the local reference signal and the received signal are multiplied together, the receiver is a TRF receiver, just like any other superheterodyne receiver, consisting of the second IF and a demodulator. The spread spectrum receiver includes a signal presence detector that is similar to the squelch circuit in other receivers, clock tracking circuits similar to those used in bit synchronizers, and automatic gain control (AGC) processing similar to that of other receivers.

We will discuss the details of both spread spectrum and more conventional receivers further in the succeeding chapters and in Appendix II.

3.11 TRADEOFFS IN RECEIVER DESIGN

At every step in a receiver design, tradeoffs are necessary, and each choice affects the overall performance of the receiver. The following list of tradeoffs and their effects outline the overall problem. There are many more tradeoffs to be made in superheterodyne receivers than in TRF or regenerative receivers and many more possible effects. Nevertheless, the superheterodyne technique produces consistently better overall performance than any of the alternatives. In the chapters that follow, these tradeoffs and their effects will be covered in more detail.

Table 3.2 summarizes receiver design tradeoffs and their effects on a receiver's characteristics, by receiver stage. Other receiver stages are auxiliaries to those listed in Table 3.2. Although they contribute to overall receiver performance, we have listed only those that directly process the desired signal in some way. More information on the auxiliaries (AGC, squelch, etc.) will follow in the succeeding chapters.

Table 3.2 Receiver Design Tradeoffs

Receiver stage	Tradeoff	Effect
TRF preamp	Narrow bandwidth	Improved selectivity
		Difficult tuning and tracking
	Wide bandwidth	Lack of selectivity
		Poor sensitivity
Regenerative preamp	Narrow bandwidth	Tendency to instability
	BPF only	Improved selectivity
Superheterodyne preamp	Narrow bandwidth	Improved spurious rejection
		Reduced interference vulnerability
		More difficult design
	Wide bandwidth	Simpler design
		More vulnerable
		Requires better linearity
	Gain	Gain too high increases spurious responses
		Gain too low increases noise figure
	Noise figure	Low noise figure improves sensitivity
		Low noise figure stages tend to have poor dynamic range
Superheterodyne mixers	Passive	No development cost
		Good isolation
		High LO drive
		6 to 10 dB loss
	Active	Discrete or IC
		Low production cost
		Conversion gain
		Low LO drive
		Lower isolation
IF amplifiers	Frequency	Strong impact on spurious
		Frequency affects BW and gain
	Number of IFs	Determines gain distribution and total gain available
	Bandwidth	Determines receiver predetection bandwidth
	Noise figure	May affect sensitivity
	Gain	Tendency to be unstable if gain too high
	Gain control/limiting	Depends on modulation
Demodulator	Coherent	Lowest threshold
		May allow reduced noise bandwidth
		Favors low IF frequency
	Noncoherent	Simplest design
		Higher threshold

3.12 SELECTING A RECEIVER'S CONFIGURATION

The choice of a receiver configuration is not as simple as it might appear at first glance, although the use of a superheterodyne receiver is often the best choice. Table 3.3 compares the basic receiver architectures (TRF, superheterodyne, regenerative, etc.) on the basis of the advantages and disadvantages of each. We will concentrate on superheterodyne receivers from this point. There are so many choices in specific superheterodyne architectures, however, that the design is not necessarily trivial. These are some of the choices that must be made:

1. Operating frequency and local oscillator frequency range. These are determined by system requirements, the first IF frequency, and
2. The choice of high-side or low-side injection.
3. The number of frequency conversions and IF amplifiers. Single conversion is very common, but dual, triple, or quadruple conversion receivers are not unknown. Where high gain and/or operation at frequencies greater than a few hundred MHz is necessary, at least two conversions are almost always used.
4. Where several conversions are used, the choice of high- or low-side injection must be made several times. With low-side injection, modulation sidebands are preserved, but with high-side injection they are inverted. (That is, upper sidebands become lower sidebands and vice versa.)
5. Bandwidth and gain of each amplifier.
6. Linear or nonlinear processing selected. (With linear processing, automatic gain control may be required.)

Table 3.3 Comparison of Receiver Architectures

Configuration	Advantages	Disadvantages
TRF	Simplicity Few spurious responses	Limited selectivity (Q) Limited sensitivity (gain)
Superheterodyne	Wide frequency range Easily tuned Good sensitivity Good selectivity Uniform performance	Prone to spurious responses
Regenerative	Simplicity Good selectivity	Lack of stability
Superregenerative	Simpler than superheterodyne Good selectivity	Poor long-term stability (better than regenerative, poorer than superheterodyne)

Table 3.4 Superheterodyne Configurations

Configuration	Advantages	Disadvantages
Single conversion	Simplicity	Sensitivity and/or stability poor
Multiple conversion	Good sensitivity Good stability Good spurious rejection	More possible spurious responses
IFM	Rapid reponse	Poor multisignal response Limited to signal identification
Compressive/ microscan	Very low coherent signal detection threshold	Resolution limited to $1/T$ Limited to signal presence
Radiometric	Very low signal detection threshold (coherent or noncoherent)	Limited to signal presence

7. If automatic gain control is needed, then AGC loop bandwidth and response must be specified.
8. Choice of a demodulator.

These and many other design choices must be made. In Chapter 4 we will examine superheterodyne receivers stage by stage.

Superheterodyne configuration tradeoffs are listed in Table 3.4. It is worth noting that multiple conversion superheterodyne receivers will outperform all other configurations, except in a few special applications. Table 3.4 serves as an introduction to the reasons for choosing a multiple conversion superheterodyne configuration.

QUESTIONS

1. What is the main distinguishing feature of a TRF receiver?
2. What are the main problems with TRF receivers?
3. What are the main problems with regenerative receivers?
4. How does a superheterodyne receiver deal with the problems of TRF and regenerative receivers?
5. Can a compressive receiver outperform a TRF receiver? Why?
6. How is a receiver's noise bandwidth defined, and how does it affect a receiver's performance?
7. What is meant by the sensitivity of a receiver?
8. What is meant by the selectivity of a receiver?
9. What is the most used receiver configuration today? Give three reasons why.

REFERENCES

1. Spectrum Analysis, Application Note 63. Hewlett-Packard Company, July 1964.
2. Pulse expansion/compression IF subsystems for radar. *Handbook of Acoustic Signal Processing*, Vol. III, Anderson Laboratories, 1995.
3. BAW (bulk acoustic wave) delay lines, *Handbook of Acoustic Signal Processing*, Vol. II, Anderson Laboratories, 1995.
4. Digital Instantaneous Frequency Measurement Receivers, TRW Microwave, 1984.
5. Terman, F. E., *Electronic and Radio Engineering*, McGraw-Hill, New York, 1955.
6. Langford-Smith, F., *Radiotron Designer's Handbook*. RCA, 1953.
7. Miller, G. M. *Handbook of Electronic Communication*, Prentice-Hall, Englewood Cliffs, NJ, 1979.
8. Krauss, H. L., Bostian, C. W., and Raab, F. H., *Solid State Radio Engineering*, John Wiley & Sons, New York, 1980.
9. Dixon, R. C., *Spread Spectrum Systems, with Commercial Applications*, John Wiley & Sons, New York, 1994.
10. Donaldson, G. J., Competitive Receiver Technologies, Watkins-Johnson Tech-Notes, July/Aug. 1984.

This page intentionally left blank

4
The Superheterodyne Receiver

In the previous chapter, we discussed a number of receiver configurations and their variations. In this chapter and throughout the remainder of this book we will concentrate on the superheterodyne technique.

Although the superheterodyne receiver is the undisputed performance champion and the world as we know it could not exist without superheterodyne receivers, they are far from perfect. Therefore, we will first discuss the primary shortcoming of the superheterodyne reception technique: unwanted spurious responses. These unwanted responses fall into at least three groups:

1. Spurious responses caused by the frequency conversion process and/or the local oscillators(s) used in that process.
2. Spurious responses due to internal amplifier tuning.
3. Spurious responses due to nonlinearity ahead of the frequency converter. (This also applies to TRF receivers.)

In section I we will continue with spurious response description and analysis. Lest we become discouraged by the problem of unwanted responses, let us first look at both the advantages and the disadvantages of superheterodyne receivers side by side.

As a standard, we will consider the dual-conversion superheterodyne receiver that is shown in block diagram form in Figure 4.1. This receiver operates as follows:

- A signal from the antenna (or other transducer) is input to a tuned preamplifier, which has a 3-dB bandwidth wide enough to pass the desired signal but has insufficient rolloff to define the receiver's noise bandwidth (B_N) or selectivity. (In some cases a wideband amplifier with a whole-band filter may be employed here, in the real world.)
- After amplification and preselection, the signal is passed to a mixer

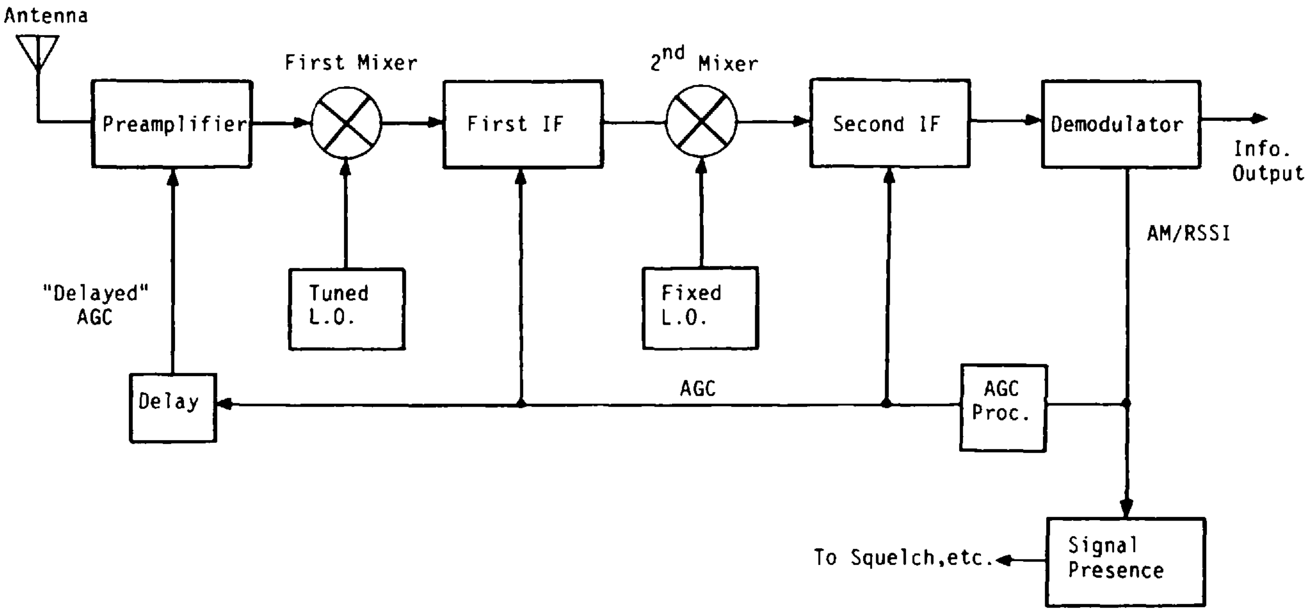


Figure 4.1 "Standard" superheterodyne receiver block diagram.

where it is multiplied with a local oscillator signal whose frequency is a function of the same tuning control that sets the preamplifier frequency. The frequency of the local oscillator is not the same as the frequency to which the preamplifier is tuned, however. The local oscillator is tuned to a frequency that is offset from the input frequency by an amount equal to the first IF frequency. It may be either higher than the input frequency or lower, depending on which is the most advantageous in a particular application.*

- The first IF amplifier is tuned to the difference between the input signal frequency and the local oscillator frequency.† That is, the first IF will select from the mixer output any signal at a frequency

$$f_{IF} = f_{RFIN} - f_{LO}$$

and pass only the signal in a narrow frequency range around f_{IF}. The filter used here is usually much narrower and has much steeper skirts than the filter in the preamplifier. This is possible because the frequency is usually much lower than the RF input signal and much better filters can be constructed.‡

- The filtered and amplified signal at the first IF is now multiplied in the second mixer with a fixed-frequency local oscillator (again either high-side or low-side injection). This process results in converting the signal to a new intermediate frequency where the final filtering and amplification are done before the signal is demodulated. The second IF, being at the lowest signal frequency in the receiver, usually defines the receiver's selectivity, adjacent channel rejection, and noise bandwidth and is a strong contributor to the receiver's signal sensitivity. (Recall that a receiver's noise bandwidth determines the effective noise input power, i.e., $P_N = \mathrm{KTB}_N$.) As the second IF is at the lowest signal frequency in the receiver, the best bandpass filters can usually be obtained here.
- The receiver's signal demodulator follows the final IF amplifier. (Sometimes there are more than two IFs.) It recovers the information from the modulated, down-converted carrier. Remember, the final IF amplifier and demodulator are a TRF receiver, when considered by themselves, and the job of the rest of the superheterodyne receiver is to present the signal to this TRF receiver in a form that it can use.

The demodulator, in addition to recovering the information, often performs other important tasks, one of which is to generate an estimate of the input signal

---

*This is termed "high-side injection" or "low-side injection," respectively.

†Sometimes it may be tuned to the sum of the two in a given application.

‡This is precisely the problem mentioned earlier with TRF receivers. That is, from this point on, the frequency is lower so that limited Q can produce narrower bandpass filters. Remember, $Q = f_c/\mathrm{BW}_{3dB}$.

level. This estimate is then employed to control the receiver's gain (through an automatic gain control or AGC loop), operate a "squelch" control, adjust the local oscillator frequency, or other tasks.

Within the gain control loop shown in Figure 4.1 is a "delay" block. This is not a time delay and should be called an "offset." Because it was named long before most of us were born, however, we will honor the name and continue to call the resulting signal delayed AGC. Its purpose is to prevent the receiver from being desensitized by the AGC loop when signals only slightly larger than the minimum to be received are present.

Now that we have considered our benchmark "standard" receiver as an assemblage of components and know something about the way in which they are related to one another, let us proceed to a summary comparison of the advantages and limitations of superheterodyne receivers. This is given in Table 4.1. From there we will proceed to examine the superheterodyne technique in all its glory (and shortcomings).

Table 4.1 Superheterodyne Receiver Advantages and Limitations

Characteristic	Description and advantages	Limitations
Sensitivity	Determined by noise figure and noise bandwidth. Routinely to −135 dBm in GPS.	Noise figure reduction reaching minimum achievable.
Selectivity	Determined by IF filter(s) and its shape factor.	More selective RF filters needed.
Adjacent channel rejection	To 80 dB in some receivers. A function of IF filter rolloff and ultimate rejection.	More than 60 dB difficult to achieve.
Image rejection	Defined by first IF frequency and RF filter. 80 dB practical.	Tunable preamplifier needed to achieve good image rejection.
IF rejection	Depends on input-to-IF isolation.	Applies only to superheterodyne receivers.
Premixing	Caused by nonlinear preamp.	Applies only to superheterodyne receivers.
Double beats, etc.	Caused by nonlinear preamp.	Applies only to superheterodyne receivers.
Subharmonic rejection	Caused by nonlinear preamp.	Applies to TRF and superheterodyne.

4.1 PREAMPLIFIER CONSIDERATIONS

4.1.1 Noise Figure

A receiver's preamplifier is one of its most important components. The preamplifier can be used to set the receiver's noise figure and to determine much of the spurious response suppression capability of the receiver.

The overall noise figure of a receiver, which is the amount of noise added to the input signal and noise by the receiver itself, is

$$F = 10\log\left(\mathrm{NF}_1 + \frac{\mathrm{NF}_2 - 1}{G_1} + \frac{\mathrm{NF}_3 - 1}{G_1G_2} + \frac{\mathrm{NF}_4 - 1}{G_1G_2G_3}\cdots\right)$$

where F = noise figure in dB

NF_K = noise factor† of stage K

G_K = gain of stage K

It is apparent from this equation that when the gain of the first stage (G_1) is large compared with the noise factor of the second stage (NF_2), the overall noise figure is dominated by the noise factor of the first stage (NF_1). The preamplifier is the first stage in a receiver, and one of the best ways to improve a receiver's sensitivity is to reduce the preamplifier's noise figure and/or increase its gain.

There are some circumstances in which reducing the second-stage noise figure makes a significant difference in a receiver's overall noise figure, but in most applications the second-stage noise figure can be as high as 10 dB without any significant effect. What is a "good" noise figure in a receiver? At this time, amplifiers with a noise figure of less than 0.5 dB at 4 GHz may be purchased for less than $200. The same level of performance would have cost several million dollars in the 1960s.

What has made the difference? The development of low-noise GaAs FET (gallium arsenide field effect transistor) amplifiers has greatly reduced the achievable receiver noise figure in recent years, especially at frequencies over a few hundred megahertz. Today, noise figure of 3 dB or less can be achieved in receivers operating at frequencies in excess of 10 GHz. A "good" noise figure then is usually considered to be near 3 dB, or below.

4.1.2 Signal Handling

The ability of a receiver preamplifier to handle large signals without distortion is critical in many applications. In mobile applications, for example, the attenuation

*Noise factor (NF) is the factor by which an amplifier or other circuit increases its input noise. The relationship between noise factor and noise figure (F) is $F = 10\log\mathrm{NF}$.

that can be expected between a transmitter and a receiver is often as little as 50 dB, and this means that the signal level at the preamplifier input might easily overdrive it.

Consider a cellular telephone system. In such systems, to help prevent mobile-to-mobile interference, a technique called frequency division duplexing or FDD is used. To implement this, the overall band assigned is split into two subbands (824–849 MHz and 864–889 MHz). In these bands, base stations transmit in the high band and mobiles transmit in the low band. This, of course, means that base stations receive in the low band with mobiles receiving in the high band.

This simple approach prevents a pair of mobile units from interfering with one another, because they cannot transmit and receive at the same frequency. It does not prevent base stations and mobile units from interfering with one another, however, which is a very good reason for the tendency to place cellular base stations in locations that are not easily accessible to mobile users.

Where transmitters and receivers can be placed physically close together, at the same frequency or at frequencies that are close enough that they cannot be rejected by filtering at the frequency of operation, it is easy to cause the receiver's preamplifier to become nonlinear. That is, if we suppose that a 100-MHz transmitter whose power output is +40 dBm (10 W) is 50 m away from a receiver, then the signal at the receiver (assuming no antenna gain at either transmitter or receiver) would be

$$+40 \text{ dBm} - 20\log\left(\frac{4\pi \times 50}{3.0}\right) = -6.4 \text{ dBm}$$

Considering that the preamplifier has a nominal 10 to 20 dB of gain, the −5 dBm input signal would cause its output level to be +5 to +15 dBm. This is a very large signal for a preamplifier, which must also handle very small signals.

In fact, one of the most difficult tasks in designing a receiver is to provide a preamplifier that can handle, *at the same time*, large and small signals linearly. The main purpose of tuning the receiver's preamplifier (or employing a bandpass filter with it) is to restrict the range of frequencies over which the preamplifier must accept and amplify wildly disparate signals without distorting them. It is such distortion of one or more signals that leads to a significant part of the superheterodyne receiver's spurious responses.

We will discuss these characteristics in detail in Chapter 5. For the sake of some degree of brevity at this point, we will list a few general rules for preamplifiers:

1. Make the noise figure as low as is practical in the light of cost, signal handling ability (dynamic range), and current drain requirements.
2. Provide sufficient gain to set the receiver's overall noise figure. Gain

should be typically between 10 and 30 dB. Below 10 dB, gain is usually too low. Above 30 dB, the output requirement is too high.
3. Provide an amplifier that is linear over the entire signal range, to prevent spurious signal responses.
4. Filter to the extent possible at the frequency of operation, to suppress image and other spurious responses. As a rule, it is advisable to filter to a bandwidth that is less than or equal to one half of the first IF frequency.

4.2 MIXER/CONVERTER CONSIDERATIONS

Any nonlinear device can be a mixer (or frequency converter). Mixers are used to change a signal's frequency while preserving the modulation being conveyed. This is exactly the process required in a superheterodyne receiver. (This frequency conversion process was originally called heterodyning, hence the name "superheterodyne receiver." Also, at one time, the mixer in a superheterodyne receiver was called the "first detector.")

Mixers can be divided into two general categories: active and passive. Both can perform well over wide signal and frequency ranges. Tradeoffs in active versus passive mixer application are listed in Table 4.2. Active mixers are used in most consumer and commercial applications for one lone reason: in large quantities they are low in cost, even though it may be necessary to amortize the engineering development cost into the overall price.

Passive mixers are often employed in military receivers but seldom in consumer or cost-sensitive applications. Even though passive mixers are inexpen-

Table 4.2 Active and Passive Mixer Considerations

Active	Characteristic	Passive	Characteristic
Transistor	Requires power supply	Ring diode	No power supply needed
	Poor isolation		Good isolation
	May have conversion gain		Moderate conversion loss
	Low cost		High production cost
	Low LO power required		High LO power required
Gilbert cell	Requires power supply	Diode	Very simple
	May have good isolation		Poor isolation
	May have conversion gain		Very low cost
	Readily integrated		Moderate conversion loss
	Low cost		High LO power required
	Low LO power required		

sive and reliable and have very good port-to-port isolation, their cost in large quantities is many times the cost of an active mixer in comparable quantities. The only passive mixers that will be considered here are of the ring diode type, because they are readily available, inexpensive, and require no development time for the user.

Like the preamplifier, the first mixer must be able to operate over a wide range of frequencies and signal levels. The specific frequency range is as shown in Table 4.3.

The choice of high-side or low-side injection of the local oscillator into the mixer is usually made to aid in the implementation of the local oscillator itself. Often, the choice of high- or low-side injection determines the practicality of building the frequency source used for a local oscillator. Another practical consideration is that high-side injection causes the modulation on a carrier to be inverted, whereas low-side injection does not do so. That is, when high-side injection is employed, a signal's upper sideband(s) becomes a lower sideband(s) and vice versa. With some forms of modulation (FSK, for example) such inversion is very important. With others (such as AM), inversion means little or nothing at all.

Together, the first mixer and the preamplifier account for many of the spurious responses of a receiver. The image response, for example, is

$$\mathrm{RF} + 2\mathrm{IF} = \mathrm{RF} + 2(\mathrm{LO} - \mathrm{RF})$$

and

$$\mathrm{RF} - 2\mathrm{IF} = \mathrm{RF} - 2(\mathrm{RF} - \mathrm{LO})$$

for high-side and low-side local oscillator injection, respectively.

Image signals may be rejected by a filter at the preamp frequency or by an "image rejection mixer." The image rejection mixer usually has only 30 dB (approximately) image rejection, however, which can easily be exceeded by a bandpass filter at the preamplifier frequency. For this reason, image rejection mixers are usually employed only to supplement the rejection provided by bandpass filtering.

Table 4.3 Mixer Port Frequency Range Requirements

Port	Frequency range
Input (RF port)	Same range as previous stage
Local oscillator (LO port)	Either RF + IF (high-side injection) Or RF − IF (low-side injection)
IF output	Difference between RF and local oscillator frequencies

As a rule, the higher the frequency of the first IF, the easier it is to reject image frequencies with a bandpass filter in the receiver preamplifier. We will discuss this characteristic in more detail in Chapter 5. Here it is sufficient to remember that image frequencies are separated from the desired signal by twice as much as the frequency of the IF that follows. Therefore the higher the IF frequency, the greater the image frequency separation and the easier it is to suppress by filtering.

4.3 FREQUENCY SYNTHESIZERS

Currently, many receivers (if not the majority) employ frequency synthesis in one form or another in their local oscillators. The primary distinguishing characteristic of these receivers is that they tune in increments. This is as opposed to the continuous tuning ability of older receivers, designed before the availability of low-cost, easy-to-employ "digital" synthesizer integrated circuits.

The primary advantages of employing a phase lock synthesizer, or any other digitally controlled synthesizer are its accuracy and ease of tuning. The drawback is that an ability to fine tune the receiver is seldom provided.

4.4 IF AMPLIFIERS

The choice of an intermediate frequency or frequencies in a superheterodyne receiver and the importance of the choice are difficult to overemphasize. The choice should be made only after a careful analysis of potential spurious responses in the receiver as they are affected by the IF frequency.

Tables 4.4 and 4.5 are intended to show some of the possible spurious responses in a receiver and the actions necessary to suppress them. Choosing an IF frequency or IF frequencies (in a multiple conversion receiver) can be a complicated process. Fortunately, computer programs exist that predict spurious responses up to at least 20th order and help in the choice of the "best" IF range.

One point that should always be considered is that one should avoid using IF frequencies that are close to the frequency of large broadcast transmitters. Be aware that AM broadcast transmitters in the United States are allowed to employ 50,000 W. FM transmitters and television transmitters are also very strong sources of image or IF spurious responses. In most cases it is wise to avoid an IF in any of the broadcast bands.

A second choice of great importance is that of how many conversions (or IFs) to use in a receiver. It is not unusual to employ three, four, or even more separate IF frequencies. For example, the ICOM 7000-series wide range receivers employ quadruple conversion.

Table 4.4 Spurious Responses Affected by IF

Spurious response	Effect	Action required
Image	RF + 2IF − LO = IF	Filter at RF, increase IF
	RF − 2IF − LO = IF	Filter at RF, increase IF
	IF + 2IF − LO = IF	Filter at IF, increase IF
	IF − 2IF − LO = IF	Filter at IF, increase IF
Overwide RF	RF − RF = IF	Increase IF: $F >$ BW
Premixing	2RF ± RF = IF	Increase IF
Double beats, etc.	2RF ± RF in RFBW	Reduce RF bandwidth Increase RF linearity
Higher order beats	MRF ± NRF in RF or IF bandwidth	Increase linearity Choose new IF frequency(ies)
IF response	RF = IF	Isolate input filter
Subharmonic response	$\mathrm{RF} = \dfrac{\mathrm{RF}}{N}$ or $\dfrac{\mathrm{IF}}{N}$	Isolate input filter

Why would one choose to design with more than one IF frequency? Some reasons are:

- Need to operate at very high frequencies*
- Need to operate over a very wide frequency range
- Need for very good sensitivity
- Need for very good selectivity

4.4.1 High-Frequency Operation

In receivers designed for very high frequency* operation (several hundred MHz and higher), it is typical to amplify the signal by a small amount (10 to 20 dB) and convert to a lower frequency. The need for image rejection† dictates that the first IF be at a frequency that allows rejection of the image frequency through the selectivity of the previous RF preamplifier. This forces the first IF frequency to be relatively high, because the image frequency is offset from the desired frequency by an amount equal to twice the IF frequency. It is also desirable to set the first IF frequency high enough that it exceeds the bandwidth of the preamplifier by a factor of 2 or more.

Consider a receiver that has an operating frequency of 5.8 GHz and that requires 60-dB image rejection. With a typical filter Q of 100, the RF filter bandwidth would be 5800 MHz/100 = 58 MHz, and this forces the IF frequency to

*This does not mean to imply the 3 to 30 MHz high-frequency band.

†Many receivers require 80 dB or greater image rejection.

Table 4.5 List of Possible Spurious Responses for Two Frequencies Through 10th Order

Order	Combinations
1	f_1
2	$2f_1, f_1 \pm f_2, 2f_2$
3	$3f_1, 2f_1 \pm f_2, f_1 \pm 2f_2, 3f_2$
4	$4f_1, 3f_1 \pm f_2, 2f_1 \pm 2f_2, f_1 \pm 3f_2, 4f_2$
5	$5f_1, 4f_1 \pm f_2, 3f_1 \pm 2f_2, 2f_1 \pm 3f_2, f_1 \pm 4f_2, 5f_2$
6	$6f_1, 5f_1 \pm f_2, 4f_1 \pm 2f_2, 3f_1 \pm 3f_2, 2f_1 \pm 4f_2, f_1 \pm 5f_2, 6f_2$
7	$7f_1, 6f_1 \pm f_2, 5f_1, \pm 2f_2, 4f_1 \pm 3f_2, 3f_1 \pm 4f_2, 2f_1 \pm 5f_2, f_1 \pm 6f_2, 7f_2$
8	$8f_1, 7f_1 \pm f_2, 6f_1 \pm 2f_2, 5f_1 \pm 3f_2, 4f_1 \pm 4f_2, 3f_1 \pm 5f_2, 2f_1 \pm 6f_2, f_1 \pm 7f_2, 8f_2$
9	$9f_1, 8f_1 \pm f_2, 7f_1 \pm 2f_2, 6f_1 \pm 3f_2, 5f_1 \pm 4f_2, 4f_1 \pm 5f_2, 3f_1 \pm 6f_2, 2f_1 \pm 7f_2,$ $f_1 \pm 8f_2, 9f_2$
10	$10f_1, 9f_1 \pm f_2, 8f_1 \pm 2f_2, 7f_1 \pm 3f_2, 6f_1 \pm 4f_2, 5f_1 \pm 5f_2, 4f_1 \pm 6f_2, 3f_1 \pm 7f_2,$ $2f_1 \pm 8f_2, f_1 \pm 9f_2, 10f_2$

be higher than 116 MHz to prevent two signals that fall within the RF filter bandwidth from being mixed together in the first mixer to produce a spurious signal that falls at the first IF frequency. This IF frequency (116 MHz or greater) would also place the image frequency at an offset of 232 MHz or more from the desired input frequency. With a three-section filter, which is typical of many receivers, the rolloff rate is such that several octaves of separation are required between the filter edge (the 3-dB point) and the image frequency.* Two to four octaves would be required, which places the image at a minimum of 174 to 870 MHz from the filter's band edge and puts the IF at a frequency of 174/2 = 87 MHz to 870/2 = 435 MHz, depending on the type of RF filter. On the other hand, it is usually desirable to place the receiver's demodulator at a much lower frequency, which forces the use of a second IF.

4.4.2 Wide Frequency Range

Dual IF frequencies are often employed in wideband receivers to restrict the range required of the local oscillator. If the local oscillator covers an IF frequency, then the receiver may not be able to operate at some frequency or frequencies because the local oscillator acts as an interferor. An example of the use of a dual IF

*Typical response rolloff is 6 to 12 dB per octave, per section. Rolloff is measured from the 3-dB point and is measured in octaves, with the filter's 3-dB bandwidth as the criterion. That is, if the filter's 3-dB bandwidth is W, then one octave is W Hz from the 3-dB point, two octaves is $2W$ Hz away, three octaves is $4W$ Hz away, etc.

frequency to solve such a problem is the spectrum analyzer described in Chapter 3, where a 2.5 GHz IF frequency is employed for input signals from 0.1 MHz to 2000 MHz and a 500 IF MHz frequency is used for input signals above 2 GHz. The swept local oscillator is the same for all ranges, however, and sweeps from 2500 MHz to 4500 MHz, so that neither the input frequency nor the local oscillator frequency ever fall within the IF, when the desired input is in the correct operating range.

It is also common in widerange receivers to employ dual, triple, or even more IF frequencies, so that the gain required may be distributed over several frequencies with less likelihood of instability at each of the different IFs. In the ICOM R-7000 receivers, for example, that cover a range of 25 MHz to 2.0 GHz, quadruple conversion is used. In these receivers, a standard 455 KHz IF is employed for AM reception, and a standard 10.7 MHz IF is employed for FM reception. (This allows off-the-shelf ceramic filters to be employed for good selectivity at minimum cost.)

4.4.3 Improved Sensitivity

A receiver with a very low noise figure has the potential to provide good sensitivity, but not without high gain in addition to the low noise figure. Let us suppose that a receiver's sensitivity is specified at 1.0 μV (-107 dBm) and that its demodulator requires a 0-dBm input signal. The net gain required between this receiver's RF input and its demodulator is 107 dB.

Because only 20 dB is typically available from the RF preamplifier and some loss can be expected in various places,* at least $107 - 20 = 87$ dB gain must be had from the IF amplifier(s). Although 87 dB gain is practical, it is much easier to achieve if it is distributed among two or more amplifiers at different frequencies. This is especially true as receivers become smaller and smaller and isolation between stages is harder to achieve. As a general rule, with more than 40 dB gain at one frequency, it is difficult to prevent oscillation without very good shielding.

4.4.4. Selectivity

The primary source of a receiver's selectivity is a good IF bandpass filter. In the 1940s, this meant several (five or six) individually tuned, cascaded vacuum tube amplifier stages. Today, the same result or better is achieved with a bandpass filter and an amplifier.

The best IF bandpass filters, however, are available in the lower frequency ranges (i.e., below 100 MHz). Therefore, it is often advantageous to employ dual conversion, with a first IF chosen at a higher frequency to provide good image

---

*Mixer, filters, etc.

rejection and a second chosen where a bandpass filter with the best possible characteristics is available.

4.5 DEMODULATOR CONSIDERATIONS

Demodulators tend to be designed at lower frequencies, almost always at less than 100 MHz and often at 10 MHz or less. There are two basic reasons why this is so.

1. The best bandpass filters are concentrated in the lower ranges. Some of these are crystal filters, ceramic filters, mechanical filters, and surface acoustic wave filters.*
2. It is generally easier to construct a demodulator at lower frequencies, especially when a coherent demodulator is used.

The specific type of demodulator used is a function of the signal being demodulated, as discussed in Chapter 2. Receiver noise bandwidth is determined by the demodulator and/or the preceding IF filter(s). Demodulators and their design are discussed in detail in Chapter 8.

4.6 RECEIVER DESIGN LAYOUT

Returning to the dual-conversion receiver block diagram of Figure 4.1, we must mention one of the most important steps in a receiver design, the process of laying out the receiver. In this case, we have arbitrarily laid out the block diagram for a dual-conversion receiver. In practice, we would start with a specification and then proceed with the design layout.

The next step in the process is to determine the desired signal level in each stage of the receiver, the gain required in each stage, any gain control requirement, and any other characteristics. This process allows detailed specifications to be generated for each block in the receiver.

Chapter 12 discusses the use of these techniques and describes the generation of subsystem specifications.

4.7 AUTOMATIC GAIN CONTROL

Automatic gain control is employed in receivers to make the performance as nearly level as possible as the input signal changes. AGC's main requirement is to

---

*Surface acoustic wave filters can be built at frequencies as high as 1.5 to 2.0 GHz. High insertion loss tends to keep them in the lower frequencies, however.

control the input signal level to the demodulator, which it accomplishes by sensing the signal and adjusting it to a controlled level, by adjusting the gain in the receiver's RF and IF amplifiers. (See Figure 4.2.)

In a superheterodyne receiver it is common to get most of the gain control in the IF amplifiers, as well as most of the gain. Where possible, it is advantageous to provide some gain control in the RF preamplifier. The reason is that such gain control allows the preamplifier output signal to remain at a lower level when large input signals are present. This helps to prevent overdriving the preamplifier and first mixer and reduces receiver spurious response.

The choices for providing AGC are amplifiers whose gain is a function of their current or attenuators that vary the signal by changing their insertion loss. Such attenuators also vary as a function of their current level.

Figure 4.2 does not include the mixers that would be employed for frequency conversion. If the mixers are passive, then loss of 6 to 10 dB should be included per mixer. For active mixers, their gain should be included.

The AGC subsystem is a classic servo loop whose overall gain should be at least (referring to signal amplitudes)

$$\Delta RF_{in} - RF_{out}$$

$$\Delta RF_{in} = RF_{in}(\max) - RF_{in}(\min)$$

$$\Delta RF_{out} = \text{range of signal into demodulator allowable}$$

It is not unusual for RF input signals to a receiver to vary by as much as 120 dB. On the other hand, demodulators require close control of their input signal level, usually to within ± 1 dB. Thus, the *minimum* AGC loop gain required would be 119 dB, a voltage gain of 1 million. (In practice, a loop gain of 10 million or more is common.) This loop gain consists of

$$A_{loop} = \Delta G \times D$$

where ΔG = total gain change per volt

D = detector output in V_{DC} per $V_{(signal)}$

Where AGC delay is employed for the RF preamplifier, it should be remembered that the AGC loop gain is lower* until preamplifier gain control starts.

The AGC voltage detector can be as simple as an envelope detector or as complex as an integrated circuit that includes an oscillator, mixer, IF amplifier, and demodulator. Such integrated circuits usually include an RSSI detector whose output is suitable for AGC. (RSSI is short for received signal strength indication.)

For many receivers, AGC is very important and may be vital. Any receiver that must preserve the signal's amplitude information until it reaches the demod-

*By an amount equal to the preamplifier's gain versus AGC voltage sensitivity.

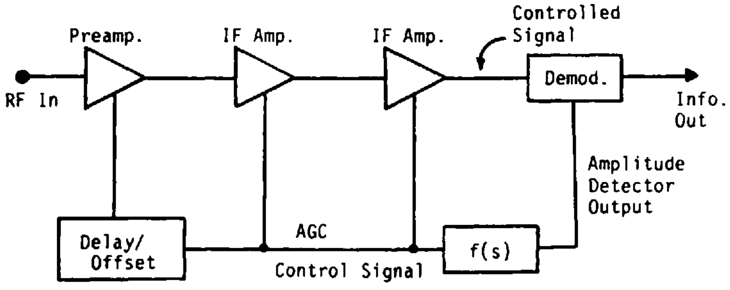


Figure 4.2 Automatic gain control loop. (Does not include frequency conversions.)

ulator can certainly benefit from AGC if it must operate over a wide signal amplitude range. Even FM receivers that employ limiting of the received signal applied to their demodulator benefit from AGC in their RF preamplifier.

The same type of filter used in phase lock loops, a lag-*RC* filter, is used in an AGC loop (see Section 4.8.2) and the loop time constants and bandwidth are determined by the filter. These time constants are

$$T_1 = \frac{K}{\omega_n^2} = (R_1 + R_2)C$$

and

$$T_2 = \frac{2\xi}{\omega_n} = \frac{1.414}{\omega_n} = R_2C$$

which are identical to those of the phase lock loop, where K is the open loop gain, ω_n is the natural resonance of the loop in radians, and ξ is the loop damping factor, 0.707.

The primary difference is that the loop gain is made up of the gain of the controlled stages, the gain of the AGC voltage detector, and any gain in an AGC amplifier, as was previously mentioned. The minimum open-loop gain is equal to or greater than the amount of the signal amplitude change. That is, if the minimum operating signal is 1.0 μV and the maximum operating signal is 1.0 V, then the open loop gain should be $1.0/(1 \times 10^{-6})$ or 1 million. (This assumes that the input to the demodulator is to be held to an exact amount.) The gain required is slightly reduced if the signal input to the demodulator is allowed vary somewhat. For example, if the input signal varies by 120 dB, as above, but the demodulator input is allowed a 1-dB variation, then the minimum open-loop gain is 119 dB, or $\log^{-1}(119/20) = 891{,}251$.

AGC time constants, which are determined by the AGC loop filter, are chosen on the basis of the dynamic conditions under which a receiver is to be employed. Where the signal level is expected to change rapidly, for example, more AGC loop bandwidth is needed, and in many cases proper damping of the loop must be provided to control its transient response.

Television sets and table radios are examples of receivers whose AGC response can be very slow, because the transmitter from which they are receiving a signal certainly does not move, and the receivers also are fixed to within the radius of their power cord. Therefore, the AGC in such units needs very little bandwidth, transient response is not a concern, and many television receivers have been built with a potentiometer for setting the signal level. (Sometimes a near/far switch is used.) Table radios have often used AVC* instead of AGC in the past, because of their greatly reduced dynamic range compared with a mobile receiver.

The rate of change of the signal arriving at a mobile receiver is due to the effect of the change in propagation loss with distance and to the effects of multipath fading. In an aircraft link, where the relative range is changing at 1000 mph and the range is 20 miles, the signal level changes at only approximately 0.12 dB per second for a frequency of operation of 400 MHz. This is a very slow rate for the AGC loop to track and is typical of such links, and only when the aircraft are very close to one another (which is not advisable when the rate of change of distance is so rapid) does the AGC loop have to track at a higher rate. A rate of higher than 100 dB per second is considered to be very high for almost all applications, at least for monotonic changes.

Much more rapid changes occur in multipath situations, as the signal may fluctuate much more often due to alternate multipath cancellation and reinforcement as moving vehicles produce changes in the signal path(s) as seen by a receiver. If the operation is at 400 MHz and two vehicles are moving at a relative speed of 1000 mph, the rate at which fading can be expected to occur is V/λ = (1,608,000 m/hr)/(0.75 × 3600 m/sec) = 595.6 times per second, and the depth of fading can be expected to be approximately 20 to 30 dB. Therefore the rate of change for multipath fading would be 30 dB in 1/(2 × 595.6) = 0.84 msec. Of course, the rate of change for slower vehicles would be much less. For example, with two vehicles moving at relative velocity of 100 mph but operating at 900 MHz, the fading rate would be V/λ = 160,800/(0.33 × 3600) = 134.7 fades per second, at a rate of 30 db in 3.7 msec.

Can AGC loops be designed to track signals at such rates? Of course they can, and it is no more difficult to do so than to design a phase lock loop to track an input signal whose frequency is being modulated at a similar rate. By designing the AGC loop to track the fading rate of a received signal, the receiver can accept the signal without degradation, as long as the signal does not fade below the receiver threshold.

---

*AVC is automatic volume control, which controls only the output audio level in a receiver.

Both phase lock loops and AGC loops are typically type two, second-order servo loops, and the design process is the same. It is especially important in the AGC loop, however, that the loop be properly damped, as shown in Appendix 2. When testing and adjusting AGC loops for proper damping and bandwidth, an amplitude-modulated test signal is employed, with the signal generator modulated at a variable audio rate. The AGC loop filter, which is a lag-*RC* network of the same type used in a phase lock loop, is adjusted in the same way as the phase lock loop: R1 for bandwidth and R2 for damping factor. (See section VIII.B.)

Some receivers (notably in push-to-talk applications) use fast attack, slow decay time constants. This is accomplished by switching the AGC loop filter bandwidth with a carrier detector output.

4.8 BIT AND FRAME SYNCHRONIZATION

Although it is not shown in Figure 4.2, it is necessary in receivers that are employed for data reception to provide both bit synchronization and frame synchronization. Spread spectrum receivers also often employ a bit synchronization process that is usually called clock tracking. In the data receivers, the primary reason for bit synchronization is to be able to supply both output data and the clock that goes with the data to the data users who employ a receiver. In spread spectrum receivers, a bit synchronizer is employed to allow the code generator in the receiver to track the received signal.

Some bit synchronizers operate by transition detection on the recovered data (see Figure 4.3a) whereas others employ a time differential technique (see Figure 4.3b). The second technique is also discussed in Appendix 2.

Frame synchronization is usually accomplished by periodic transmission of a specific symbol that can be recognized by a receiver. Barker codes are often employed for this purpose. Suppose, for example, a Barker code is transmitted at the beginning of each data frame. The receiver, on recognizing the known code, can tell precisely when a frame begins. The code also serves to identify ones and zeros, because its pattern is precisely known. (This is exactly what is done in the GPS system to provide frame synchronization for the 50 bps data stream. The data clock, on the other hand, is derived from a 1/20 countdown of the repetition rate of the spectrum-spreading code, which repeats exactly 1000 times per second.)

4.8.1 Transition Detection

Transition detection is a technique for bit synchronization that takes the recovered data at the receiver and derives the bit clock from the data. The data are in one/zero form, consisting of near-ground and near-power supply voltages with rapid transitions between the two levels. This allows pulses to be generated from the data

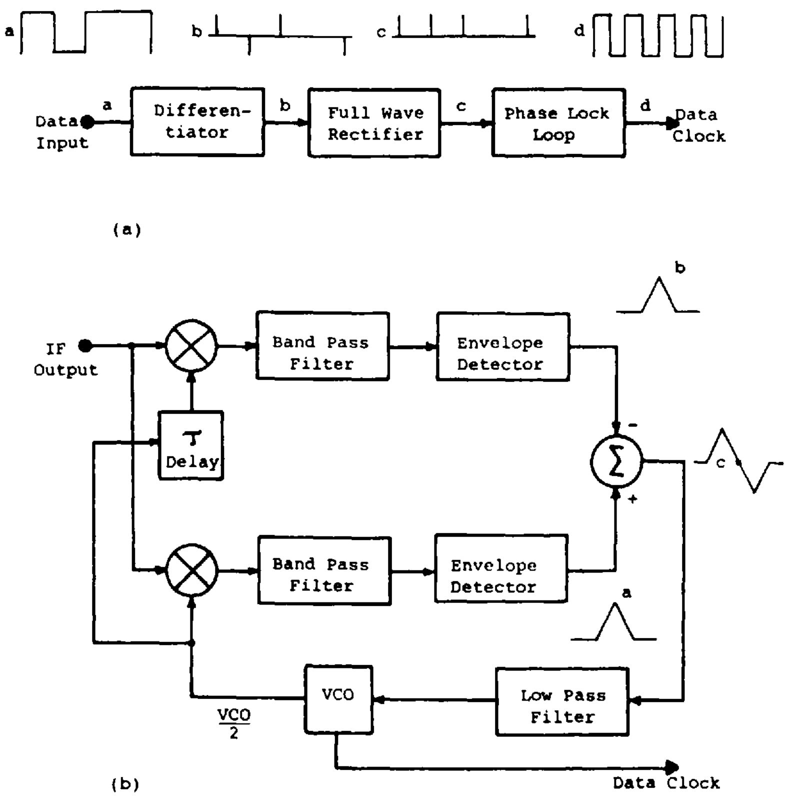


Figure 4.3 Block diagrams of transition detector and delay lock bit synchronizers. (a) Transition detector bit synchronizer. (b) Delay lock bit synchronizer.

stream by coupling the data through a capacitor, where the capacitor's value, combined with the input impedance of the following stage, produces a time constant *RC* that is short compared with a bit time. The signal produced is a positive-going pulse for a zero-to-one transition and a negative-going pulse for a one-to-zero transition in the data. Full wave rectification of the pulses then sets all of the pulses to the same polarization, usually positive going, which produces a pulse train in which the pulse periods coincide with the data (except that pulses are missing when there are two or more bits in a row that are the same). See Figure 4.3a.

Notwithstanding the missing pulses, the pulse stream can be applied to a phase lock loop whose phase detector is triggered by the pulses (such as an *R-S* flip-flop) and whose VCO is at the data rate, and the loop will lock to the data rate.

In such bit synchronizers, the loop filter is chosen to have very narrow bandwidth, so that the VCO is not affected by occasional missing pulses. The Q of the VCO's resonator and the narrow bandwidth of the loop produce a strong flywheel effect to minimize clock jitter, which clock is derived from the VCO.

4.8.2 Delay Lock Loop Bit Synchronizers

A more complex bit synchronizer that is used in both spread spectrum and more conventional receivers is the delay lock loop, which is shown in Figure 4.3b. In this type of bit synchronizer, correlation between the received IF signal and the receiver's data clock is used to synchronize the receiver clock. Let us suppose that the input IF signal is a BPSK modulated signal at rate R. The receiver's clock, divided by two, is multiplied in the bottom multipler with the input IF signal, which is inverted by the clock/2 when it is a one and not inverted when it is a zero. When the transitions in the input signal coincide with the receiver clock/2, the signal level passing through the bandpass filter (which is designed to match the bandwidth of the data-modulated signal) causes the envelope detector to output its highest level. When the transitions do not coincide, the effect is to modulate the received signal with the local signal (the receiver's clock/2), which causes the envelope detector's output to be reduced. The envelope detector's output is the triangular function at "a" in Figure 4.3b, which is maximum when the transitions coincide exactly and minimum when they are offset by half the clock/2 period (which is the same as the input data period).

At the same time as the lower arm of the delay lock loop is operating, the upper arm is multiplying the received signal with a receiver clock/2 that is offset by one half of the clock period. This means that when one of the envelope detectors is outputting a maximum signal, the other is outputting a minimum signal and vice versa. The two signals are shown as "a" and "b" in Figure 4.3b. When these two are combined differentially, they produce a discriminator function (because of their time offset) as shown in c of Figure 4.3b, and this can be used to cause the VCO that is also supplying the data clock to synchronize and track the received data clock.

In practice, because the VCO tracks a point that is halfway between the two clocks multiplied with the received signal, a delay of the receiver's VCO that is one half of the delay between the two reference clocks is used by the receiver and supplied to the data user.

The delay lock loop provides the best performance of all of the bit synchronizers available for both data and spread spectrum code clock tracking. The bandwidth of a typical delay lock loop is from 100 to 1000 times that of the data demodulator that normally operates to demodulate the data; therefore the data clock loop usually continues to function even when a receiver is below its threshold and cannot supply acceptable bit error performance. It is also not unusual to have a loop filter in a delay lock loop that is third or even fifth order, which significantly improves its dynamic performance.

4.9 LIMITATIONS IN SUPERHETERODYNE RECEIVER DESIGN

Superheterodyne receivers are better today than they have ever been. They do have limitations, however. Let us examine these limitations and the prospects for improvement in the light of technology that exists today and may exist in the near future.

4.9.1 Preamplifiers

A significant problem exists today in the preamplifier area with respect to bandwidth restriction and/or tuning. Namely, as operating frequencies become higher and higher, the ability to restrict the bandwidth of preamplifiers becomes more limited.

Several companies manufacture integrated circuit amplifiers (MMICS) today that are capable of providing significant gain over several octaves with a very low noise figure. A typical such amplifier is the AVANTEK (now Hewlett-Packard) INA-03184. This device has 3-dB bandwidth from the very low frequency range to approximately 2.5 GHz and a noise figure less than 3 dB over the same range.

How can such a low-cost, low-noise, easy-to-use amplifier* be ignored? It cannot, but the second part of the problem is how one keeps signals hundreds of MHz away from the desired frequency from swamping the amplifier. Indeed, how can a receiver reject image frequencies when it employs a wideband amplifier? One answer is to employ either a fixed or tunable filter† in series with the wideband amplifier. Usually, this means that the receiver's sensitivity is worsened by 1 to 3 dB because of the insertion loss of a filter. (This, of course, presumes that the filter is ahead of the amplifier.) Remember that the filter is the primary defense against spurious responses, especially image responses.

Another consideration is gain control in the preamplifier. If the preamplifier is a fixed-gain integrated circuit, then an attenuator must be inserted ahead of the preamplifier. This again causes a loss in receiver sensitivity equal to the insertion loss of the attenuator.

---

*Even better IC amplifiers are now available for specific applications.

†More than one filter may also be used.

Table 4.6 Some Mixer Limitations

Limitation	Active mixers	Passive mixers
Local oscillator power	−10 dBm (min.)	+27 dBm (min.)
P_{1dB}	+15 dBm (out)	+24 dBm (in)
F (noise figure)	6 to 15 dB	Less than 1 dB
Conversion loss	Up to 15 dB gain	6 to 10 dB loss

Finally, we must also consider the signal-handling capability of the preamplifier. The device previously mentioned, the INA-03184, has a 1-dB output compression point (P_{1dB}) of −2 dBm, and with gain of 23 dB this means that it can accept only signals that are well below −25 dBm if spurious responses are to be suppressed. In Chapter 5 we will consider these problems in more detail.

4.9.2 Mixers

We have already considered some of the differences between active and passive mixers, especially the cost factors. It is also necessary to consider P_{1dB}, noise figure, and local oscillator power required. Typical limitations for mixers* are listed in Table 4.6. Mixers will be discussed in detail in Chapter 6.

4.9.3 IF Amplifier Limitations

A great deal of progress has been made in the area of integrated circuit IF amplifiers within the recent past. In the period before 1980–1985, integrated IF amplifiers could do no more than a single transistor can. Then IF amplifiers for cellular telephones were developed, but these were intended for 455-kHz IF frequency, and most are not particularly useful at higher IFs (not even 10.7 MHz).

Today's integrated circuit IFs are capable of operating as high as 500 MHz with significant gain (30 dB or more) and gain control for AGC that is greater than the gain available. Primary limitations are that the noise figure is typically in the 6- to 15-dB range and the P_{1dB} (output) is typically about 10 dBm.

Some IC amplifiers are intended for limiting IFs, but many have features such as RSSI output, separate AGC level output, input mixers, and quadrature detectors.

Further information will be given in Chapter 5. These limitations and others will be considered in the remaining chapters of this book.

*These limitations may not apply to the same mixers.

QUESTIONS

1. What is the primary drawback to the use of a superheterodyne receiver?
2. Does a TRF receiver have spurious responses? If so, name at least two.
3. Why are third-order responses more serious than other spurious responses?
4. Name two disadvantages of passive mixers. Name two advantages.
5. What is the primary purpose of automatic gain control?
6. What advantages does a superheterodyne receiver have over a TRF receiver?
7. What problems might a superheterodyne receiver exhibit if it had no preamplifier? Where might such a receiver do well?
8. Name the advantages and disadvantages of multiple conversion in a superheterodyne receiver.
9. What is the purpose of delayed AGC?
10. Name three contributors to the sensitivity of a superheterodyne receiver. Why are these important?

REFERENCES

1. Langford-Smith, F., *Radiotron Designer's Handbook*, RCA, 1953.
2. Terman, F. E., *Electronic and Radio Engineering*, McGraw-Hill, New York, 1955.
3. Spilker, J. J., Jr., and Magill, D. T., The delay lock discriminator—an optimum tracking device, *Proc. IRE*, Sept. 1961.
4. Dexter, C. E., and Glaz, R. D., HF Receiver Design, Watkins-Johnson Tech-Notes, Mar./Apr. 1978.
5. Harper, T., Hybridization of Competitive Receivers, Watkins-Johnson Tech-Notes, Jan./Feb. 1980.
6. Donaldson, G. J., Competitive Receiver Technologies, Watkins-Johnson Tech-Notes, July/Aug. 1984.

5
Amplifier Design and Application

Receivers use amplifiers in many ways: as amplifiers (of course), mixers, oscillators, and even demodulators. Here we will consider the uses and the characteristics required of amplifiers in various selected applications in receivers. In the previous chapter, we discussed some specific aspects of amplifier requirements. This chapter and several of the succeeding chapters will be concerned with the amplifiers themselves and methods used to optimize them for application in receivers.

5.1 PREAMPLIFIERS AND PRESELECTORS

A preamplifier is usually an amplifier whose noise contribution (usually expressed as its noise figure) is low enough that it does not add significantly to the receiver's input noise and whose gain is great enough to boost the desired input signal to a level that is significantly greater than the input noise combined with the noise contributed by the amplifier itself.

That is, the preamplifier's output signal-to-noise ratio is

$$\frac{S_{\text{out}}}{N_{\text{out}}} = G\left(\frac{S_{\text{in}}}{\text{KTB} + N_{\text{preamp}}}\right)$$

and

$$G_{\text{preamp}} = \frac{S_{\text{out}}(\text{KTB} + N_{\text{preamp}})}{N_{\text{out}}S_{\text{in}}}$$

where G is the amplifier gain and KTB is the total input noise in bandwidth B.

If we consider that the input noise and the amplifier-contributed noise are amplified along with the signal, the process illustrated by Figure 5.1 occurs, wherein the input signal-to-noise ratio to the amplifier is 15 dB and the output signal-to-noise ratio is 35 db. In this illustration, the signal level is increased by

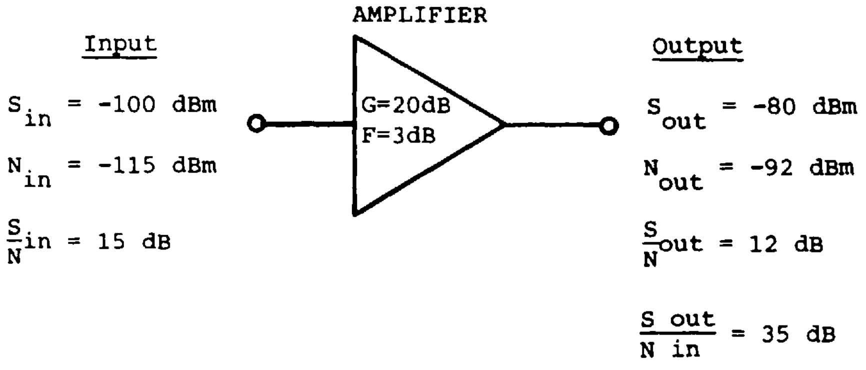


Figure 5.1 Illustration of preamplifier showing relation of signal and noise.

20 dB and the noise level is increased by 23 dB, but the signal is increased to −80 dBm, so the succeeding stages' noise contribution is not significant.

Thus, the action of the amplifier is to amplify both signal and noise and to add to the input noise (KTB). The end result, however, is that even though the output noise level is higher than the input noise level and the output signal-to-noise ratio is worse than the input signal-to-noise ratio, the overall result is a usable signal-to-noise ratio at a level that is high enough to prevent additive noise from further degrading the received signal. For example, if we add another −115 dBm (the level of KTB) to the amplifier output, the signal-to-noise ratio is still 35 dB; the preamplifier has raised the signal level to a point at which KTB is no longer a significant factor. The noise figure of the following stage(s) can also be relatively high—say 6 to 10 dB—without seriously affecting the receiver's sensitivity.*

5.1.1 Preamplifier Bandwidth

If a preamplifier is tuned, for frequency selectivity, or at least is combined with a bandpass filter or filters to restrict its bandwidth, it may be called a preselector. The purpose of reducing the preamplifier's bandwidth is to reject as much of the unwanted spectrum as possible. This has the advantage that a preamplifier's signal handling capability (signified by its P_{1dB}† or IP_3) needs to be designed only

---

*Recall that the overall receiver noise figure is

$$F = 10\log\left(NF_1 + \frac{NF_2 - 1}{G_1} + \frac{NF_3 - 1}{G_1G_2} + \frac{NF_4 - 1}{G_1G_2G_3} + \cdots\right)$$

which is affirmed by our example.

†P_{1dB} is the 1-dB compression point. IP_3 is the third-order intercept.

Table 5.1 Typical Filter Characteristics for Preamplifiers

Filter type	Frequency range	Bandwidth range (%)	Loss[a]
Lumped constant	To 4 GHz	3–10	Low
SAW	To 2 GHz	0.1–50	High
Helical resonator	To 1.5 GHz	1–2	Low
Cavity	To 20 GHz	0.2–30	Low
Combline	1–18 GHz	2–70	Low
Coaxial ceramic	0.4–6 GHz	0.5–10	Low
Waveguide	To 100 GHz	1–10	Low

[a]In this context, low loss means less than 3 dB insertion loss, although multisection filters may have higher loss. Some surface acoustic wave (SAW) filters may have much higher loss.

for the desired signal without considering such things as extremely powerful broadcast signals that fall outside the receiver's band of interest but could cause spurious responses or desensitization of the receiver if the preamp were driven into a nonlinear mode.

Bandpass filtering at the preamplifier frequency does not usually affect a receiver's sensitivity or selectivity,* because the preamplifier bandwidth is much greater than the bandwidth set by the IF filters, which determine the bandwidth presented to the demodulator. This, as discussed previously, is because higher Q and much better filters are available at the IF frequencies.†

Some of the filter types available at preamplifier frequencies are listed in Table 5.1. Table 5.2 lists IC amplifiers.

5.1.2 Preamplifier Gain Control

It is often necessary to provide some form of gain control in a receiver preamplifier to avoid overdriving both the first mixer and the preamplifier itself when large signals are present. For example, if a receiver must operate with RF input signals to its preamplifier that are in the range of 1.0 μV (−107 dBm) to 1.0 V (+13 dBm), which is not an untypical requirement, a considerable amount of gain control capability is necessary.

At −107 dBm signal input, the preamplifier should employ its highest gain,

*Except that if the bandpass filter is used ahead of the preamplifier, its insertion loss may degrade the effective receiver sensitivity by reducing the signal level. It is sometimes said erroneously that a filter's insertion loss ahead of a preamplifier increases a receiver's noise figure. This is not true, but the effect is the same on sensitivity.

†There is promise that superconduction or very high dielectrics may offer very high Q and better high-frequency filters in the future.

Table 5.2 Summary Chart of Integrated Circuit Amplifiers for Preamplifier and IF Application, at 1900 and 350 MHz[a]

	Gain (dB)		Noise figure (dB)		P1dB out (dBM)		IP3 out (dBm)		P1dB in (dBm)		IP3 in (dBm)		
Device number	1900	350	1900	350	1900	350	1900	350	1900	350	1900	350	Manufacturer
1. MAAM 12031	20	—	1.65	—	+7	—	+19	—	−13	—	−1	—	MA/COM
2. MAAM 12032	13	—	1.8	—	+2	—	+13	—	−11	—	+0	—	MA/COM
3. MAAM 02350A2	19	18	3.8	3.9	+14	+14	+25	+24	—	—	+6	+6	MA/COM
4. TQ9132	16	—	4.1	—	+17	—	+27	—	—	—	+11	—	TriQuint
5. RF2304	9.5	13	2.1	1.8	+6	—	+17.5	+21	—	—	+8	+8	RF Monolithics
6. RF2113	—	30	—	18	—	+30	—	—	—	+0	—	—	RF Monolithics
7. MSA0520	—	8.5	—	6.5	—	+23	—	+33	—	+14.5	—	+24.5	Hewlett-Packard
8. MSA0504	—	8	—	6.5	—	+19	—	+29	—	+11	—	+21	Hewlett-Packard
9. MSA1104	—	12	—	3.6	—	+17.5	—	+30	—	+5.5	—	+18	Hewlett-Packard

[a]A portion of the data is extrapolated from data sheet information. Where blanks are shown, either the device is not intended for operation at the frequency or sufficient information is not given on the data sheet.

because of the need to set the receiver's noise figure. When the input signal goes to +13 dBm, however, the same amount of gain in the preamplifier would cause it to saturate and thereby badly distort the signal being amplified. (If the preamplifier has a typical 20 dB of gain, then a +13-dBm input signal would appear at a linear amplifier output as +33 dBm, which is 2 W. Because very few low-noise preamplifiers do not compress at such high output signal levels, a typical preamplifier would badly distort the signal being amplified and give poor spurious performance at such high levels.)

Even if the preamplifier could output a +33 dBm signal without distortion, the mixer that follows would be forced to handle too large a signal. (A passive mixer, for example, would be forced to employ a local oscillator as large as +40 dBm, 10 W.)

All of this reinforces the notion that some form of gain control is often required in a preamplifier if the input signal to a receiver is to vary over a very wide range. Two general approaches are used for preamplifier gain control. The first is to control the gain of the preamplifier itself. The second is to place a variable attenuator in series with the preamplifier. Figure 5.2 shows a typical integrated circuit attenuator. Such attenuators can be employed either ahead of or behind a preamplifier and serve the gain control function very well. The tradeoff is that ahead of the amplifier, the attenuator protects it from being overdriven, and behind it, there is no insertion loss.

Control of the amplifier's gain is desirable if:

- The amplifier's noise figure is not affected. The primary method of adjusting gain in an amplifier is to increase or decrease the bias current in the amplifier. This, however, may also change the amplifier's noise figure. (Noise figure tends to increase as collector current increases.) Figure 5.3 shows a typical curve of gain and noise figure for a bipolar transistor as a function of collector current.
- The amplifier's tuning is not affected. It is not unusual for an amplifier to be detuned as the voltage applied to its electrodes changes. This is due to a varactor-like change in interelectrode capacitance with voltage, and this capacitance change can change amplifier tuning. Figure 5.4 illustrates this detuning mechanism in a typical tuned amplifier.

Although some excellent integrated circuit preamplifiers are available, very few of them incorporate gain control, which may force a designer to incorporate a separate attenuator(s), as discussed earlier in this section. Another type of variable attenuator that may be employed is a diode attenuator, a simple form of which is shown in Figure 5.5.

Silicon diodes have intrinsic resistance that is a function of the current through the diode, and this resistance is approximately

$$R_{\text{diode}} = \frac{26}{i_{(\text{mA})}}$$

M/A-COM
an AMP company

Voltage Variable Absorptive Attenuator, 35 dB 0.5 - 2 GHz

AT-109

V 2.00

Features

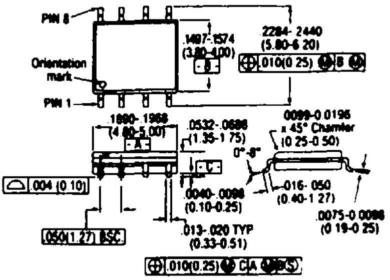
- Single Positive Voltage Control 0 to +5 Volts
- 35 dB Attenuation Range at 0.9 GHz
- ±2 dB Linearity from BSL
- Low DC Power Consumption
- Temperature Range -40°C to +85°C
- Low-Cost SOIC 8 Plastic Package
- Tape and Reel Packaging Available

Description

M/A-COM's AT-109 is a GaAs MMIC voltage variable absorptive attenuator in a low-cost SOIC 8-lead surface mount plastic package. The AT-109 is more linear than the higher attenuation range AT-108. The AT-109 is ideally suited for use where linear attenuation fine tuning and very low power comsumption are required. Typical applications include radio, cellular, GPS equipment and automatic gain/level control circuits.

The AT-109 is fabricated with a monolithic GaAs MMIC using a mature 1-micron process. The process features full chip passivation for increased performance and reliability.

SO-8



8-Lead SOP outline dimensions
Narrow body .150
(All dimensions per JEDEC No. MS-012-AA, Issue C)
Dimensions in ( ) are in mm.

Unless Otherwise Noted: .xxx = ± 0.010 (.xx = ± 0.25)
.xx = ± 0.02 (.x = ±0.5)

Ordering Information

Part No.	Package
AT-109	SOIC 8-Lead Plastic Package
AT-109TR	Forward Tape & Reel*
AT-109RTR	Reverse Tape & Reel*

* If specific reel size is required, consult factory for part number assignment.

Electrical Specifications[1], T_A = +25°C

Parameter	Test Conditions[1]	Unit	Min.	Typ.	Max
Insertion Loss	0.5 - 1.0 GHz 1.0 - 2.0 GHz	dB dB		2.5 3.2	2.7 3.5
Attenuation	0.5 - 1.0 GHz 1.0 - 2.0 GHz	dB dB	35 30		
Flatness (Peak-to-Peak)	0.5 - 1.0 GHz 1.0 - 2.0 GHz	dB dB		±0.5 ±1.2	±0.8 ±1.5
VSWR				2:1	
Trise, Tfall Ton, Toff Transients	10% to 90% RF, 90% to 10% RF 50% Control to 90% RF, Control to 10% RF In-band	µS µS mV		25 35 12	

1. All measurements at 1 GHz in a 50-Ω system, unless otherwise specified. The RF ports must be blocked outside of the package from ground or any other voltage.

Specifications Subject to Change Without Notice.

M/A-COM, Inc. 12-21

North America: Tel. (800) 366-2266 Fax (800) 618-8883 ■ Asia/Pacific: Tel. +81 (03) 3226-1671 Fax +81 (03) 3226-1451 ■ Europe: Tel. +44 (1344) 869 595 Fax +44 (1344) 300 020

Figure 5.2 Integrated circuit attenuator for 500 MHz to 2 GHz. (Courtesy of M/A COM.)

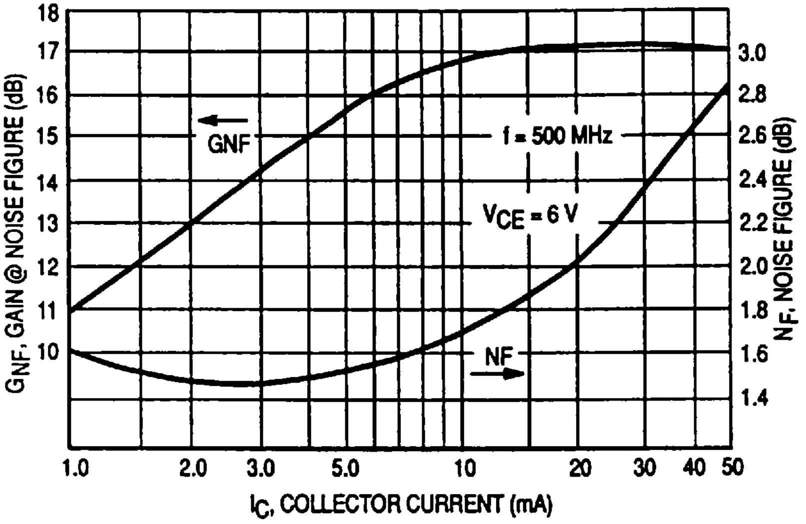


Figure 5.3 Typical gain and noise figure characteristics of bipolar transistor versus collector current. (Courtesy of Motorola.)

That is, a diode with 0.1 mA current has resistance of about 260 ohms, whereas the same diode with 10 mA current has a resistance of about 2.6 ohms. This allows the diode to be used in configurations such as that of Figure 5.5 to adjust signal level. The circuit shown in Figure 5.5 employs the diode as a shunt resistor that forms a voltage divider with the source resistance, R_s.

When the AGC voltage is zero, the diode resistance is high, and the input signal is attenuated by an amount that is determined primarily by the diode's parasitic capacitance. (The diode capacity is in the range of a few hundredths of a picofarad to approximately 2.0 picofarads, where common computer diodes (1N914 etc.) are at the 2.0 pF end of the scale. PIN* diodes, on the other hand, have a capacitance that ranges from about 0.02 to 0.1 pF. Therefore, the insertion loss that can be expected in a 50-ohm system is from less than 1 dB to more than 3 dB, as shown in Table 5.3.

At lower RF and IF frequencies, computer diodes work quite well and their insertion loss is acceptable. At frequencies above a few hundred MHz, however, it is advisable to employ diodes with low capacitance, such as PIN diodes.

*PIN refers to a diode in which silicon is sandwiched between *p*-doped and *n*-doped layers, which leads to the P-Intrinsic-N nomenclature.

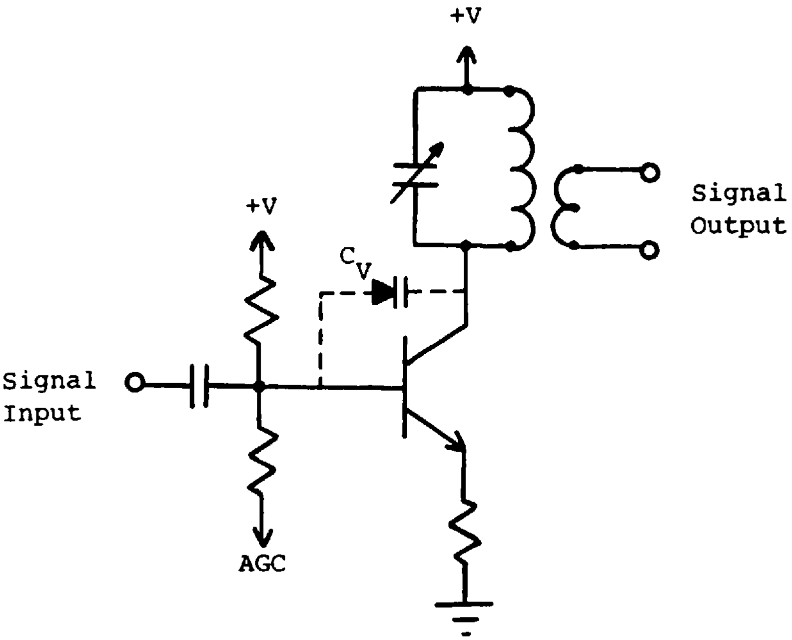


Figure 5.4 Tuned amplifier stage with AGC, showing detuning mechanism due to base-to-collector voltage-variable capacitance.

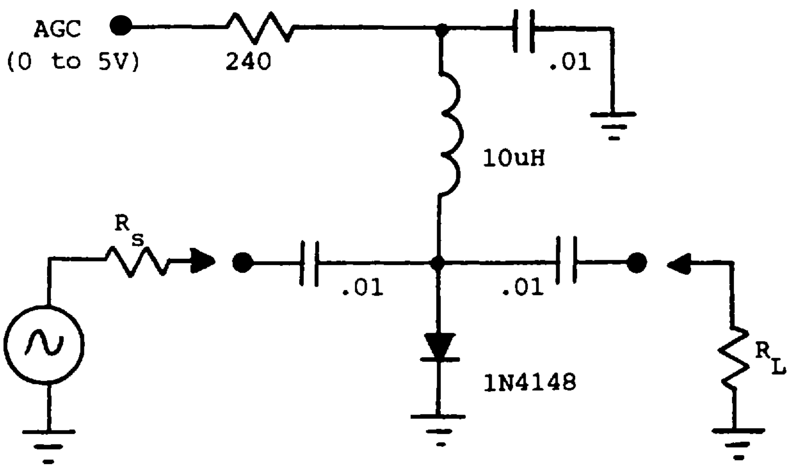


Figure 5.5 Simple diode attenuator using computer diode at 100 MHz.

Table 5.3 Diode Insertion Loss, for Use in Variable Attenuators (50 Ohms)

	Frequency of operation	
Type of diode	100 MHz	2 GHz
Computer diode (2 pF)	0.26 dB	3.5 dB
PIN diode (0.03 pF)	0.004 dB	0.08 dB

5.2 DYNAMIC RANGE

Dynamic range of an amplifier (or any other device) is usually defined in terms of the signal range over which it can operate while meeting all of its specifications for gain, signal-to-noise ratio, suppression of distortion, and any other requirements. Overall, the amplifier's dynamic range is determined by its compression point (or alternatively by its intercept point) and its noise performance. For example, if an amplifier's noise bandwidth is 100 kHz and its noise figure is 3 dB, then its noise floor is

$$-174 \text{ dBm} + 10 \log(1 \times 10^5) \text{ dB} + 3 \text{ dB} = -121 \text{ dBm}$$

Then, if the amplifier's input 1 dB compression point is -10 dBm, its dynamic range is the difference between its noise floor and its input compression point, which is -121 to -10 dBm, or 111 dB. (In many cases, an amplifier's 1 dB compression is specified at its output. The input compression point is different from the output compression point by the amount of the amplifier gain.)

The 1-dB compression point and third-order intercept point are both commonly used to define amplifiers and other devices from the standpoint of their ability to handle large signals and to estimate their spurious signal performance. The fact that both are used leads to the question of why and of how they are related.

First, it must be pointed out that a compression point can be directly measured, whereas third-order intercepts are artificial constructions based on a measurement. Also, the third-order intercept is a more recently developed technique.

If the input and output of a linear device are plotted as in Figure 5.6, the resulting plot is a straight line over the linear range of the device, with slope 1.0. Figure 5.6 shows plots for three devices: an amplifier, a wire or cable, and a lossy device such as a passive mixer. (A mixer, of course, is not a linear device, but its transfer function from one frequency to another is linear. Here, "linear" simply means that a 1-dB change in an input signal produces a 1-dB change in the output.)

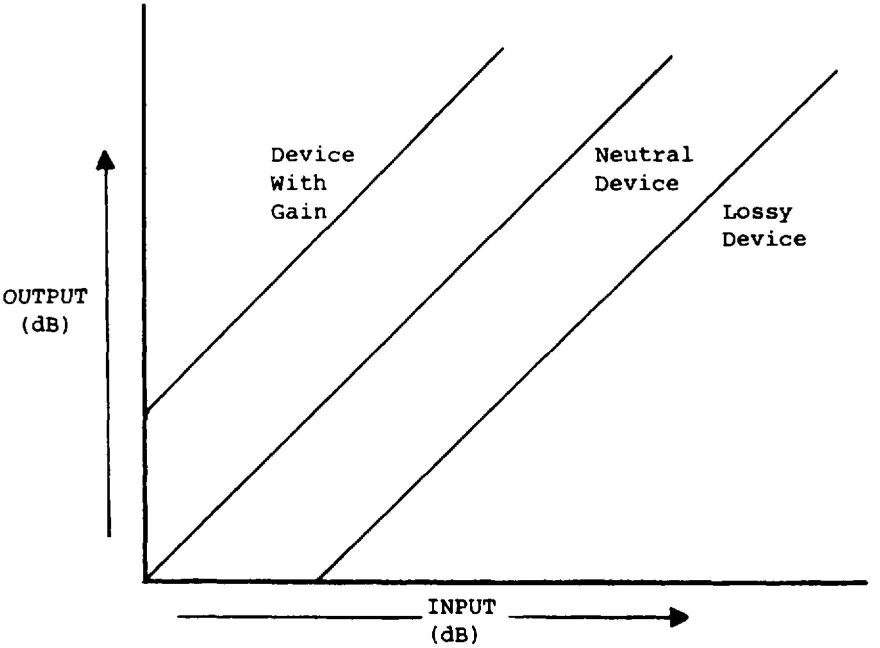


Figure 5.6 Comparison of transfer function of three linear devices: an amplifier, a neutral gain device (a wire), and a lossy device (a passive mixer).

No real device, unfortunately, has a linear transfer function that remains linear no matter how large its input signal. Even Hoover Dam reaches a point at which its electrical output no longer increases with an increase in water flow through its turbines. Here we are interested in the degree to which a device's output is limited when its saturation point is reached, how it behaves in the transition region from linearity to saturation, and where the saturation region begins.

Figure 5.7 shows the transfer function of a more typical device than that in Figure 5.6. The transfer function is initially linear but continues linearly only up to the point at which it can meet the output requirement no longer. (The output requirement, to remain linear, is simply to output a signal that is the input signal level times the gain of the device.) In Figure 5.7, consider that the amplifier output curve is extended as a straight line beyond the point at which saturation begins. Noting the point at which the actual output is less than the extended linear curve by 1 dB, we have the "output 1-dB compression point" or P_{1dB} for the device. (From this plot, we can determine both the input and output compression points.)

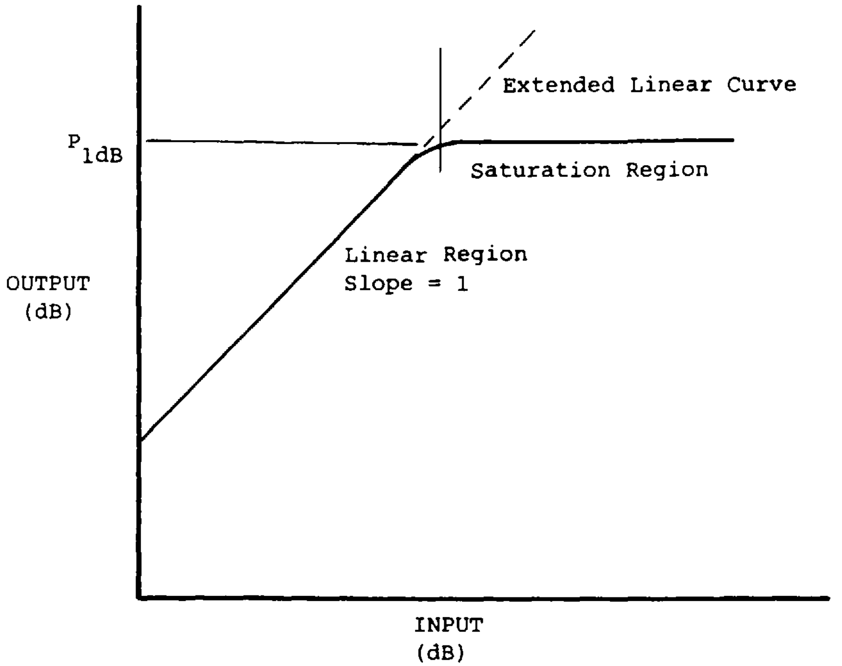


Figure 5.7 Typical amplifier transfer function showing limited output power.

Once we know the 1-dB compression point, the spurious signal output level to be expected can be estimated. The following rule is employed:

If the input signal(s) is P dB below the 1-dB compression point, then the third-order output level will be $2P$ dB below the signal level at the output.

Example. Given an amplifier with two input signals whose power levels are A dBm and B dBm, at frequencies f_1 and f_2, where both of the signals are K dB below an amplifier's input 1-dB compression point, the third-order outputs ($2f_1 \pm f_2$, $2f_2 \pm f_1$, $3f_1$ and $3f_2$) will be $2K$ dB below the desired signal at the amplifier's output. An illustration of this process is shown in Figure 5.8.

The second measure used to estimate third-order (and other) products in linear devices is the "third-order intercept" method. This technique is based on the locus of the third-order products having a slope of 3. (Similarly, the second-order products fall on a curve with a slope of 2.) Therefore, if one plots a single point for the third-order output of an amplifier ($2f_1 + f_2$ for example), a line with slope 3 drawn through that point to cross the extended linear output curve of the

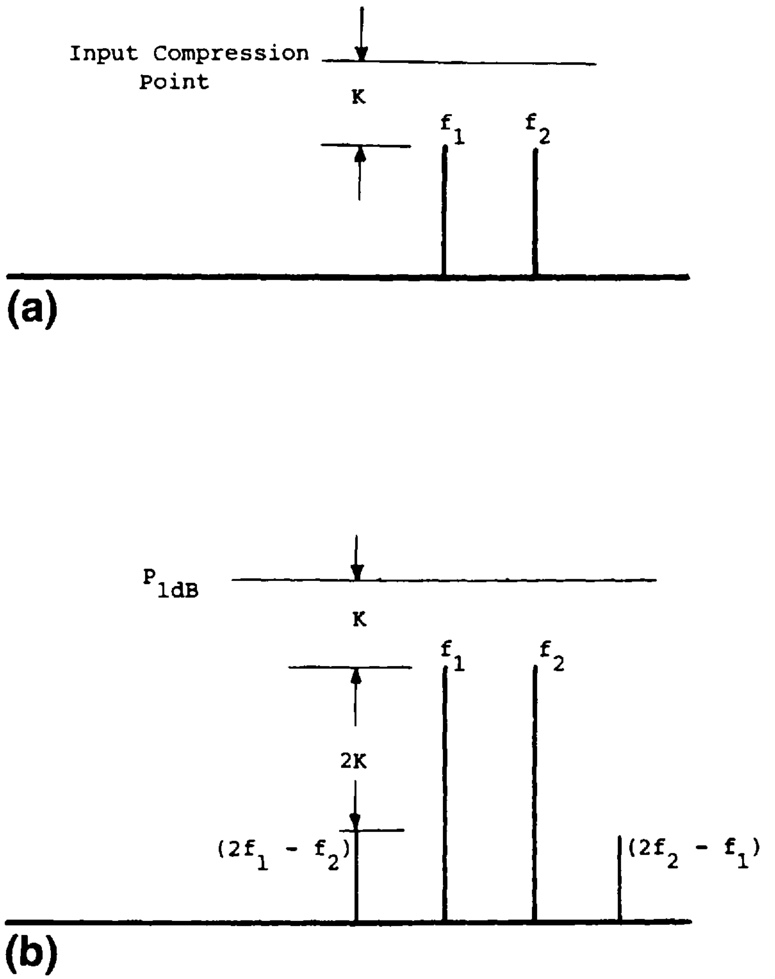


Figure 5.8 Illustration of amplifier spurious output estimation using P_{1dB} as rule of thumb. (a) Input spectrum to amplifier. (b) Partial output spectrum from amplifier.

amplifier's output, a third-order intercept is produced (Figure 5.9). (We have made use of the point-slope theorem.) This third-order intercept is typically from 10 to 15 dB above the 1-dB compression point for the amplifier.

A second-order intercept can also be produced by the same process: after measuring a second-order output ($f_1 + f_2$ for example), drawing a line to intercept the extended linear transfer function, where the line drawn has a slope of 2 and the

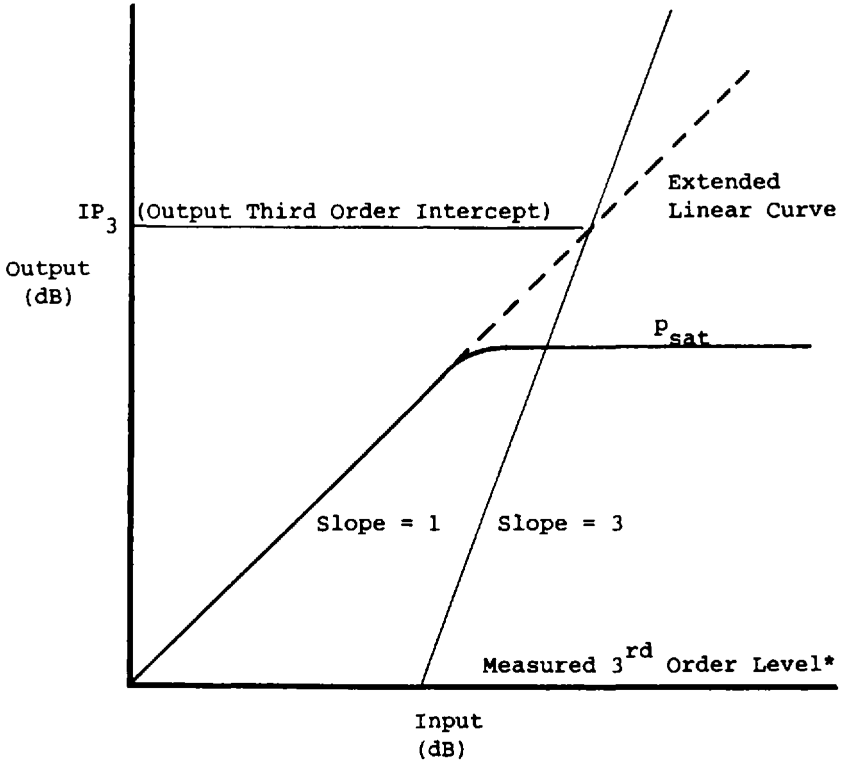


Figure 5.9 Illustration of amplifier output characteristic showing third-order intercept. *Referred to input.

line is drawn from the point of measurement of a second-order output signal.* This produces the second-order intercept.

Why should we be willing to go to such lengths to generate an artificially constructed graph that gives us a quantity that we cannot readily measure directly? The reason is that nomographs such as those in Figures 5.10 and 5.11 allow us to make more accurate estimates of spurious signal levels. Once we know a device's third-order intercept point and the signal level(s), spurious levels can be read directly. Also, manufacturers provide either input or output third-order intercept

*When making use of either the second- or third-order output level, it is most convenient to measure the output level and refer it to the input. That is, simply subtract the amplifier gain from the measurement and use that point on the input axis as the reference point.

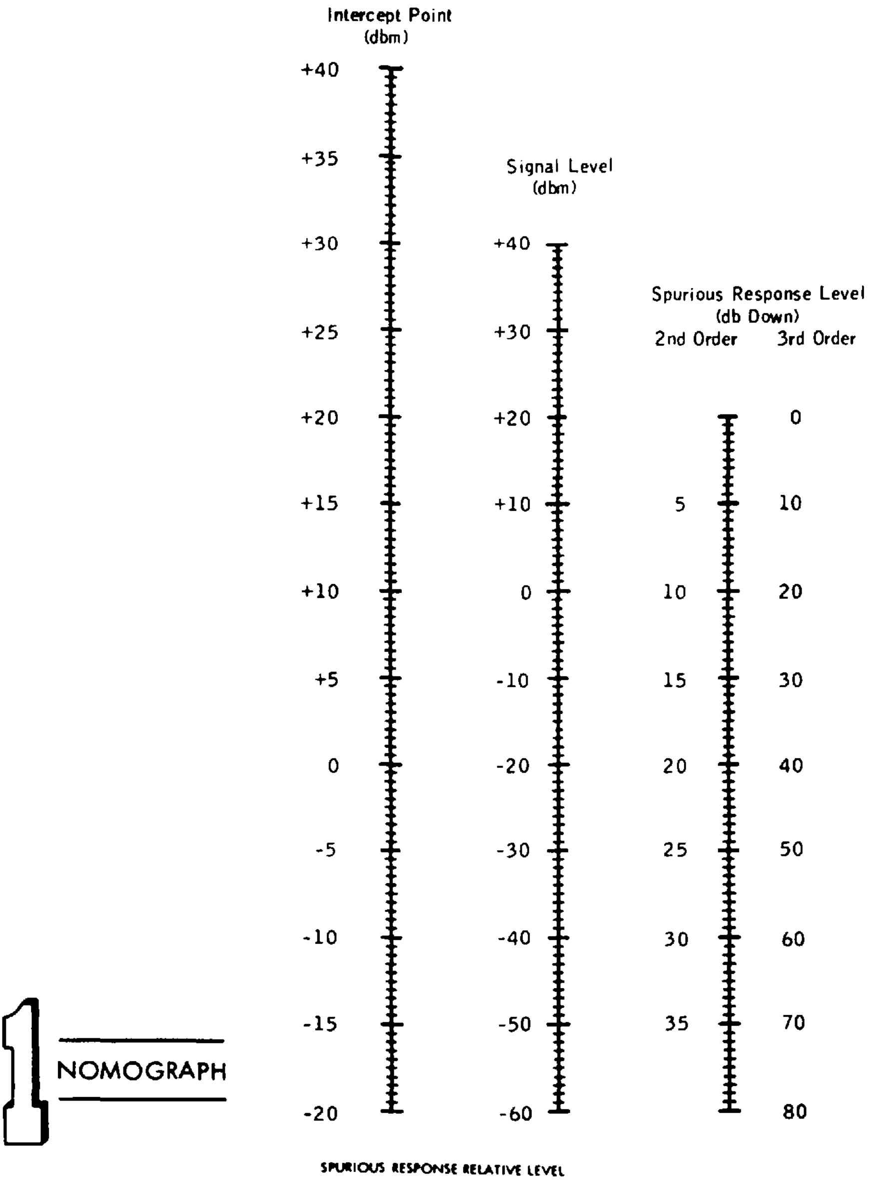


Figure 5.10 Nomograph number 1, for spurious response level from intercept point and signal level. (Courtesy of AVANTEK/Hewlett-Packard.)

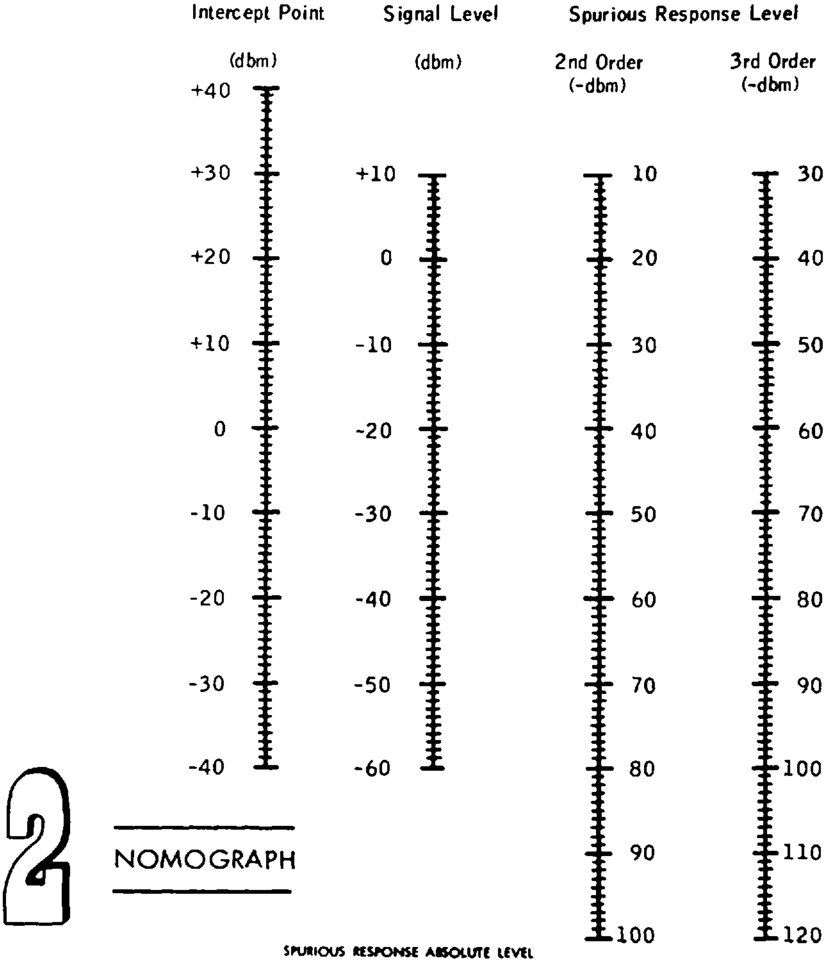


Figure 5.11 Nomograph number 2, for spurious response level from intercept point and signal level. (Courtesy of AVANTEK/Hewlett-Packard.)

information for many of the devices that are available today for use in receiver design and development.

Using the nomographs of Figures 5.10 and 5.11 and placing a straight edge between the signal level and the intercept point, one can read the second-order and third-order spurious response levels directly. Three examples, using the nomographs, are given in the following. (Courtesy of AVANTEK/Hewlett-Packard.)

5.2.1 Spurious-Free Dynamic Range

If we consider the range over which a receiver does not compress an input signal and no spurious signal is above the receiver's noise floor, we have a range that is called the receiver's "spurious-free dynamic range" and that is the best operating region (at least theoretically speaking). The greater this range, the better the receiver's performance in an environment in which the signal input to the receiver varies.

A receiver's noise floor is (in dBm)

$$-174 + 10\log B_N + F$$

where B_N is the receiver's noise bandwidth in hertz and F is the receiver' noise figure in dB. This noise floor is the lower limit of the spurious-free dynamic range (SFDR).

A receiver's input intercept point is

$$\mathrm{IIP}_3 = \text{input (max.) signal level} + \frac{\text{intermod power}}{2}$$

and

$$\text{Intermod power} = 2(\mathrm{IIP}_3 - \text{input level})$$

$$\begin{aligned}\text{Noise floor} &= \text{input level} - \text{intermod power}\\ &= \text{input level} - 2(\mathrm{IIP}_3 - \text{input level})\\ &= 3(\text{input level}) - 2\mathrm{IIP}_3\end{aligned}$$

$$\text{Input level} = \frac{\text{noise floor} + 2\mathrm{IIP}_3}{3}$$

$$\text{Spurious-free dynamic range (SFDR)} = \text{input level} - \text{noise floor}$$

$$= \frac{\text{noise floor} + 2\mathrm{IIP}_3}{3} - \text{noise floor}$$

and finally

$$\text{Spurious-free dynamic range} = \tfrac{2}{3}(\mathrm{IIP}_3 - \text{noise floor})$$

This means that a receiver that has an input intercept point of -10 dBm and a noise floor of -110 dBm has an SFDR of

$$\frac{200}{3} = 66.6 \text{ dB}$$

which is effective between -20 dBm and 66.6 dB below -10, or -76.6 dBm. In other words, the receiver will work best at input signal levels that are between the noise floor and 66 dB above the noise floor.

5.2.2 Avantek Dynamic Range Nomograph

Example 1. Given intercept point +30 dBm and output signal level −10 dBm, find the third-order IM level.

Line up a ruler on +30 on the intercept point scale and −10 dBm on the signal level scale. Read −90 dBm on nomograph 2 or 80 dB down on nomograph 1. The spurious responses are at −90 dBm, or 80 dB down from the signals at −10 dBm.

The intercept point is given for the output level. When input levels are being considered, the amplifier gain must be taken into consideration.

Example 2. Given intercept point +30 dBm, input signal −30 dBm, and amplifier gain 30 dB, find the third-order spurious level at the output. At the output, the two signals will be 0 dBm (−30 dBm +30 dB).

Line up a ruler on +30 dBm on the intercept point scale and 0 dBm on the signal level scale. Read −60 dBm on nomograph 2 or 60 dB down on nomograph 1.

Unequal signals must be converted to equivalent equal signals by subtracting from the stronger signal one third of the difference between the two signals measured in dB.

Example 3. Given intercept point +30 dBm and output signals at −3 dBm and at −18 dBm, find the third-order spurious response.

Step 1 is to find the equivalent equal signal level. The difference between two signals at −3 dBm and −18 dBm is 15 dB. One third of 15 dB is 5 dB. Subtract 5 dB from −3 dBm. The resultant signal level, − 8dBm, is the equivalent equal signal level.

Line up a ruler with +30 dBm on the intercept point scale, −8 dBm on the signal level scale, and read 76 dB down on nomograph 1 or −84 dBm on nomograph 2. This is step 2.

5.3 INTERMEDIATE FREQUENCY AMPLIFIERS

IF amplifiers provide the greatest part of the gain in a receiver and most of the gain control. This is one of the reasons why many high-performance receivers have at least two conversions and therefore two or more IFs: to distribute the gain required over more than one frequency and thereby reduce any possible feedback problems. The first IF is often a major contributor to the receiver's noise figure, and the final IF often defines a receiver's noise bandwidth. These considerations put IF amplifiers near the top of the list of importance in superhetrodyne receivers. (Of course, it should be remembered that an IF amplifier is the vestigial remains of the tuned amplifier in a fixed-frequency TRF receiver.)

Almost all high-quality IF amplifiers today are constructed using packaged, integrated circuit amplifiers, coupled with separate bandpass filters. In contrast, it is still practical to employ discrete, tuned amplifier circuits, and this is common practice in many low-cost consumer products.*

Bandpass filters available today are capable of providing good to excellent selectivity to an IF amplifier over a wide range of frequencies and bandwidths. Although it is practical to choose almost any IF frequency one desires, up to a few GHz, it is still advisable to choose one of the "standard" frequencies wherever possible. The reason for this is that a tremendous amount of time, effort, and expense has gone into development of these standard IF filters, and their high performance and low cost (without the need for new development) make them very attractive. This is the reason that IF amplifiers tend to be concentrated at relatively few frequencies.

It is especially true of the lower frequency IF amplifiers in receivers that they are concentrated at very few frequencies, and extra effort in design is often expended to be sure that one of the standard frequencies is used. This is to ensure that the design can make use of the excellent filters that are readily available. Table 5.4 lists some of the standard IF frequencies for filters that are manufactured by a number of companies and the filter types that are common at those frequencies.

Examples of standard surface acoustic wave filters are given in Tables 5.5 and 5.6. This is a family of filters at two different IF frequencies, with a wide range of bandwidths, that is usable for many receiver IF applications.

Gain control methods employed in IF amplifiers are generally the same as those described for preamplifiers, although a greater selection of integrated circuits is available at IF frequencies that include gain control capability. Some integrated circuits that are intended for IF application also include demodulators as well as RSSI (received signal strength indication) circuits, which can be employed as AGC detectors. Figures 5.12 and 5.13 are examples of such integrated circuits.

In Figure 5.12, a basic description of a gallium arsenide IF amplifier is given (only part of the data sheet is included). This IF amplifier provides approximately 30 dB gain, with 60 dB AGC range (it can attenuate the signal by up to 30 dB) from 30 to 500 MHz, and it does so using a single power supply and less than 5 mA of current. The TQ9114 has a typical noise figure of 6 dB and a typical P_{1dB} of -15 dBm.

The IF/demodulator IC shown in Figure 5.13 (again, only a partial data sheet is shown) includes an IF amplifier, RSSI detection and output circuitry, and a

*As late as the 1960s, it was common practice to employ a series of four to six tuned vacuum tube amplifiers in the IFs of high-performance receivers. This practice ended with the advent of transistors and crystal filters.

Table 5.4 Standard IF Filter Frequencies and Types

Frequency	Type of filter	Remarks
455 kHz	Ceramic	Very low cost, no tuning
	LC	Moderate cost, tunable
	Crystal	High cost, no tuning, high performance
	Mechanical	High cost, no tuning, very high performance
10.7 MHz	Ceramic	Very low cost, no tuning
	LC	Moderate cost, tunable
	Crystal	High cost, no tuning, high performance
21.4 MHz	Same as 10.7 MHz	Similar to 10.7 MHz
60.0 MHz	Crystal	High cost, no tuning, high performance
	Surface acoustic wave (SAW)	Very high cost, no tuning, high performance, high insertion loss
70.0 MHz	Same as 60.0 MHz	Similar to 60.0 MHz
140.0 MHz	Same as 60/70 MHz	Similar to 60 and 70 MHz

wideband FM demodulator. The IF amplifier portion of the device does not allow gain control, but it does provide RSSI output for external AGC (off the chip). Such devices are often employed in FM receivers, where limiting IF amplifiers are usually present, but AGC is valuable to prevent the receiver's preamplifier and downconverter(s) from limiting. The MC13155 is designed to operate to 300 MHz with a single 3- to 5-V power supply and to consume only 7 mA of supply current.

5.4 DISCRETE DEVICE AMPLIFIERS

Even though many very good integrated circuit amplifiers are available today that cover a very wide range of frequencies, it is sometimes necessary to employ a transistor amplifier, an emitter follower, or some other circuit that is made up of discrete devices. Therefore, we include the following (old-fashioned) paragraphs that describe a simple procedure for design with discrete devices, lest such designs become a lost art.

Suppose we wish to design an amplifier to operate at a 10.7 MHz IF. The first step is to select a transistor whose gain-bandwidth product, f_t, is adequate. For example, an MPSH34 (see Figure 5.14) has f_t of over 500 MHz and should produce gain of up to 47 at 10.7 MHz. Second, the transistor's typical current gain, HFE, should be considered. Then, assuming that the supply voltage V_{cc} is known, we can proceed to design an amplifier stage in a few steps.

Table 5.5 (Courtesy of Sawtek.)

Standard 70 MHz SAW Filters

Typical Performance

Part Number	BW1 (MHz)	BW3 (MHz)	BW40 (MHz)	Insertion Loss (dB)	Passband Variation (dB)	Group Delay Variation (nsec p-p)	Phase Linearity (deg p-p)	Ultimate Rejection (dB)	Delay (µsec)	Matching S11-S22
851539	0.09	0.15	0.59*	22.0	0.05	600	1.5	55	8.00	LC-LC
851541	0.18	0.30	0.89	17.5	0.02	200	1.4	60	3.50	L-L
851542	0.33	0.56	1.72	21.6	0.02	100	1.0	60	2.30	L-L
851543	0.61	0.82	1.71	18.2	0.20	150	2.0	60	3.60	L-L
851544	1.06	1.26	2.11	21.5	0.29	80	2.0	60	3.60	L-L
851545	1.47	1.68	2.51	23.2	0.24	120	3.1	60	3.70	L-L
851546	1.91	2.22	3.53	24.3	0.29	45	2.0	60	2.80	L-L
851547	2.25	2.60	4.12	25.2	0.28	45	2.0	60	2.30	L-L
851548	2.73	3.14	4.60	24.5	0.55	45	2.0	60	2.40	L-LL
851549	3.18	3.62	5.31	25.7	0.50	50	3.0	60	2.30	LL-LL
851550	3.87	4.25	5.88	22.5	0.50	85	3.0	60	2.50	LC-LC
851551	4.23	4.73	6.68	24.3	0.32	60	2.0	60	2.40	LC-L
851552	4.78	5.27	7.29	24.1	0.40	75	3.0	60	2.60	LC-L
851553	5.16	5.68	7.51	23.0	0.52	120	6.0	60	2.60	LC-LC
851554	5.73	6.20	8.12	23.0	0.50	100	4.0	60	2.60	LC-LC
851555	6.23	6.72	8.80	23.0	0.45	75	3.3	60	2.60	LC-LC
851556	6.60	7.20	9.56	20.5	0.35	70	2.0	60	1.80	None
851505	7.58	8.10	9.94	22.3	0.32	55	2.6	60	2.50	None
851557	7.63	8.23	10.52	22.0	0.22	40	2.1	60	1.90	None
851558	8.22	8.81	10.92	22.7	0.21	40	2.4	60	2.00	None
851559	8.69	9.26	11.45	22.9	0.30	40	2.0	60	2.00	None
851560	9.24	9.58	12.11	24.0	0.29	30	2.0	60	1.90	None
851475	9.84	10.27	11.92	24.5	0.70	45	2.8	60	2.40	None
851841	10.84	11.47	13.87	22.4	0.38	35	2.7	60	1.95	None
851842	11.83	12.51	15.14	23.5	0.31	45	2.4	60	1.88	None
851843	13.20	13.79	16.66	25.7	0.30	35	2.5	60	1.51	None
851844	13.75	14.63	17.69	24.6	0.26	35	1.9	60	1.50	None
851845	14.79	15.63	18.92	22.7	0.44	35	3.0	60	1.41	L
851846	15.28	16.51	20.58	24.2	0.34	35	2.8	60	1.34	L
851847	17.77	18.84	23.10	24.7	0.33	40	1.6	60	1.42	L
851848	19.43	20.70	25.15	25.2	0.39	20	2.2	55	1.39	L
851849	20.90	22.60	28.00	22.0	0.50	30	3.2	60	1.31	L-L
851850	23.10	24.80	30.80	22.8	0.55	25	4.1	55	1.27	L-L
851851	24.87	26.63	31.00	24.3	0.62	20	3.7	55	1.24	L-L
851852	26.24	28.74	37.00	26.4	0.67	20	3.9	55	1.00	L-L
851853	27.00	30.30	48.40	22.0	0.79	30	3.9	55	2.08	None
851854	29.30	32.50	49.60	22.8	0.63	25	5.3	55	1.94	None
851855	31.35	34.77	52.06	23.3	0.66	20	4.9	50	2.07	None
851856	33.52	36.89	54.65	24.0	0.68	20	5.0	45	1.94	None
851857	35.04	38.65	56.85	24.8	0.50	15	6.0	55	2.07	None
851858	37.10	40.40	58.60	25.3	0.36	10	5.0	45	2.07	None

Sawtek's standard 70 MHz filter family is now a part of the HP-EEsof System Component Library.

*Measured at 35 dB bandwidth.

The maximum recommended continuous RF input power to these devices is +20 dBM

For pricing information and immediate delivery on quantities of 100 or fewer pieces of any of the standard 70 MHz or 140 MHz Sawtek bandpass filters, please contact Penstock at 1-800-736-7862.

To the best of our knowledge, this information was correct at the time of printing. Sawtek reserves the right to modify these specifications when necessary to provide optimum performance and cost.

Table 5.5 Continued

Standard 70 MHz SAW Filters

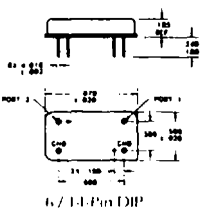
Guaranteed Specifications

Part Number	BW1 (MHz min.)	BW3 (MHz min.)	BW40 (MHz max.)	Center Frequency (MHz)	Insertion Loss (dB max.)	Passband Variation (dB max.)	Group Delay Variation (nsec p-p max.)	Phase Linearity (deg p-p max.)	Ultimate Rejection (dB min.)	Package Style (Pin/DIP)
851539	0.07	0.125	0.7*	70 ± .025	23.5	0.5	800	3	40	6/18 **
851541	0.16	0.25	1.0	70 ± .04	20.0	0.1	300	3	50	6/18
851542	0.31	0.5	1.9	70 ± .05	24.0	0.1	300	3	50	6/14
851543	0.58	0.7	2.0	70 ± .07	23.0	0.4	300	4	50	6/18
851544	0.90	1.0	2.3	70 ± .07	24.0	0.5	150	4	50	6/18
851545	1.10	1.5	2.8	70 ± .08	25.0	0.5	200	6	50	6/18
851546	1.80	2.0	3.8	70 ± .08	26.0	0.7	90	4	50	6/14
851547	2.10	2.5	4.4	70 ± .08	27.0	0.7	100	4	50	6/14
851548	2.50	3.0	4.9	70 ± .09	26.0	0.8	90	8	50	6/14
851549	3.00	3.5	5.6	70 ± .09	27.0	0.8	100	8	50	6/14
851550	3.70	4.0	6.1	70 ± .09	25.0	0.8	100	6	50	6/14
851551	4.00	4.5	6.9	70 ± .09	26.0	0.8	90	4	50	6/14
851552	4.50	5.0	7.5	70 ± .09	26.0	0.8	100	6	50	6/14
851553	5.00	5.5	7.7	70 ± .09	25.0	0.8	140	9	50	6/14
851554	5.60	6.0	8.4	70 ± .10	25.0	0.8	130	7	50	6/14
851555	6.00	6.5	8.9	70 ± .10	25.0	0.8	100	6	50	6/14
851556	6.50	7.0	9.7	70 ± .10	22.0	0.8	100	4	50	6/14
851505	7.30	7.5	10.2	70 ± .10	24.0	0.8	90	5	50	6/14
851557	7.50	8.0	10.6	70 ± .11	24.0	0.8	60	4	50	6/14
851558	8.00	8.5	11.1	70 ± .11	25.0	0.8	60	5	50	6/14
851559	8.60	9.0	11.7	70 ± .12	25.0	0.8	60	4	50	6/14
851560	9.10	9.5	12.3	70 ± .12	26.0	0.8	60	4	50	6/14
851475	9.00	10.0	12.0	70 ± .15	26.0	1.0	60	4	60	6/14
851841	10.70	11.0	14.2	70 ± .20	24.0	0.6	70	4	55	6/14
851842	11.60	12.0	15.4	70 ± .20	25.0	0.6	60	4	55	6/14
851843	12.75	13.0	17.1	70 ± .20	28.0	0.6	50	4	55	6/14
851844	13.50	14.0	17.9	70 ± .20	27.0	0.6	60	4	55	6/14
851845	14.60	15.0	19.2	70 ± .20	25.0	0.7	60	5	50	6/14
851846	15.00	16.0	20.9	70 ± .15	26.0	0.7	50	4	50	6/14
851847	17.60	18.0	23.5	70 ± .20	26.0	0.6	60	3	55	6/14
851848	19.30	20.0	25.4	70 ± .20	27.0	0.7	35	4	50	6/14
851849	20.60	22.0	28.2	70 ± .20	23.0	0.8	45	5	55	6/14
851850	22.70	24.0	31.7	70 ± .20	24.0	1.0	40	6	50	6/14
851851	24.50	26.0	34.0	70 ± .20	25.5	1.0	35	6	50	6/14
851852	25.60	28.0	37.7	70 ± .20	27.5	1.0	40	6	50	6/14
851853	26.30	30.0	49.5	70 ± .20	22.5	1.0	40	5	45	6/14
851854	29.00	32.0	50.0	70 ± .20	23.5	1.0	30	7	50	6/14
851855	31.00	34.0	52.8	70 ± .20	25.0	1.0	30	7	45	6/14
851856	32.70	36.0	55.5	70 ± .20	25.0	1.3	30	8	40	6/14
851857	34.80	38.0	57.5	70 ± .20	26.0	0.8	30	8	50	6/14
851858	36.60	40.0	59.6	70 ± .50	26.0	0.7	30	8	40	6/14

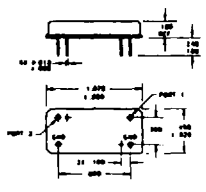
*Measured at 35 dB bandwidth.
All parts are guaranteed at 25° C.

Sawtek reserves the right to modify the above-listed specifications without notice when necessary to provide optimum performance and cost.

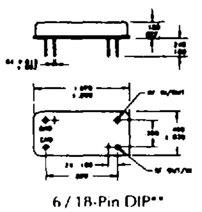
Package Drawings



6 / 14-Pin DIP



6 / 18-Pin DIP



6 / 18-Pin DIP**
(for PN 851539 only)

Table 5.6 (Courtesy of Sawtek.)

Standard 140 MHz SAW Filters

Typical Performance

Part Number	BW1 (MHz)	BW3 (MHz)	BW40 (MHz)	Insertion Loss (dB)	Passband Variation (dB)	Group Delay Variation (nsec p-p)	Phase Linearity (deg p-p)	Ultimate Rejection (dB)	Delay (μsec)	Matching S11-S22
851900	0.16	0.27	0.84	23.2	0.01	130	0.5	60	3.62	None
851901	0.32	0.56	1.73	19.5	0.03	100	1.5	68	2.78	None
851902	0.57	0.77	1.58	20.7	0.25	110	2.0	65	3.93	None
851903	0.81	1.04	2.04	21.9	0.28	100	1.6	60	3.20	None
851904	1.27	1.53	2.56	22.8	0.16	70	1.5	60	3.31	L
851905	1.75	2.07	3.60	23.8	0.28	45	1.6	60	2.70	L-L
851906	2.13	2.56	4.20	25.6	0.21	35	1.2	60	2.39	L-L
851907	2.62	3.06	4.65	25.5	0.22	35	0.9	55	2.46	L-L
851909	3.85	4.22	5.54	23.3	0.40	45	1.7	60	2.82	L-LC
851911	4.88	5.30	6.77	22.5	0.43	70	2.0	60	2.59	LC-LC
851913	5.75	6.30	8.22	26.5	0.35	44	1.8	60	2.35	LC-LC
851915	6.70	7.30	8.80	25.0	0.45	54	3.2	65	3.00	LC-L
851917	7.60	8.18	10.10	23.6	0.45	72	3.5	65	2.31	None
851919	8.75	9.35	11.30	21.4	0.35	70	3.5	65	2.10	None
851921	9.66	10.35	12.44	22.0	0.45	60	3.5	65	1.92	None
851923	11.42	12.20	14.50	22.3	0.45	35	2.0	60	1.83	None
851925	13.40	14.16	17.00	22.0	0.36	30	3.1	60	1.52	None
851927	15.18	16.19	19.50	22.8	0.40	20	2.8	63	1.59	None
851929	17.12	18.30	22.05	22.0	0.37	24	3.5	60	1.36	None
851931	19.20	20.40	24.50	22.0	0.35	19	2.3	60	1.31	None
851933	22.76	24.27	29.88	24.7	0.30	12	3.3	60	1.28	None
851935	26.75	28.35	34.40	24.5	0.45	17	3.0	65	1.17	LC
851937	30.05	32.20	39.20	24.6	0.40	11	3.0	60	1.17	L-L
851939	34.50	36.60	44.00	23.5	0.45	13	3.4	60	1.07	L-L
851941	37.30	40.10	49.10	23.6	0.45	11	3.6	60	1.10	L-L
851943	41.80	44.50	53.70	23.0	0.45	11	3.8	65	1.07	L-L
851945	45.40	48.40	58.80	26.5	0.50	12	5.5	60	1.04	L-L
851947	51.00	56.50	71.00	27.0	0.52	12	8.0	60	1.00	LC-L
851948	60.70	64.70	80.00	29.5	0.55	8	6.0	60	0.93	L-L
851949	66.50	74.50	97.00*	25.0	0.80	12	10.0	40	1.42	None
854101	72.30	81.70	111.10*	26.7	0.90	10	6.0	40	1.43	None

Sawtek's standard 140 MHz filter family is now a part of the HP-EEsof System Component Library.

*Measured at 30 dB bandwidth

The maximum recommended continuous RF input power to these devices is +20 dBM.

For pricing information and immediate delivery on quantities of 100 or fewer pieces of any of the standard 70 MHz or 140 MHz Sawtek bandpass filters, please contact Penstock at 1-800-736-7862.

To the best of our knowledge, this information was correct at the time of printing. Sawtek reserves the right to modify these specifications when necessary to provide optimum performance and cost.

Table 5.6 Continued

Standard 140 MHz SAW Filters

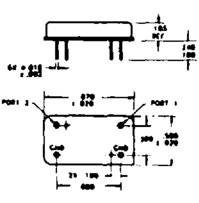
Guaranteed Specifications

Part Number	BW1 (MHz min.)	BW3 (MHz min.)	BW40 (MHz max.)	Center Frequency (MHz)	Insertion Loss (dB max.)	Passband Variation (dB max.)	Group Delay Variation (nsec p-p max.)	Phase Linearity (deg p-p max.)	Ultimate Rejection (dB min.)	Package Style (Pin/DIP)
851900	0.14	0.25	0.95	140 ± .05	25.0	0.10	400	1	50	6/18
851901	0.30	0.50	1.9	140 ± .06	21.5	0.20	300	3	50	6/18
851902	0.47	0.75	1.8	140 ± .10	23.0	0.45	240	4	50	6/18
851903	0.77	1.0	2.2	140 ± .10	24.0	0.50	270	3	50	6/18
851904	1.20	1.5	2.7	140 ± .10	25.0	0.40	150	3	50	6/18
851905	1.65	2.0	3.9	140 ± .12	26.0	0.60	120	3	50	6/14
851906	2.00	2.5	4.5	140 ± .15	27.5	0.50	100	3	50	6/14
851907	2.40	3.0	5.0	140 ± .15	27.5	0.50	90	3	50	6/14
851909	3.70	4.0	6.0	140 ± .15	25.0	0.70	120	4	50	6/14
851911	4.70	5.0	7.0	140 ± .20	25.0	0.70	180	4	50	6/14
851913	5.60	6.0	8.4	140 ± .20	28.0	0.70	110	4	50	6/14
851915	6.50	7.0	9.2	140 ± .20	27.0	0.70	130	6	50	6/14
851917	7.30	8.0	10.6	140 ± .20	25.5	0.70	210	7	50	6/14
851919	8.50	9.0	11.7	140 ± .20	23.5	0.70	210	7	50	6/14
851921	9.30	10.0	13.0	140 ± .20	24.0	0.70	150	7	50	6/14
851923	11.10	12.0	15.0	140 ± .20	25.0	0.75	90	4	50	6/14
851925	13.20	14.0	17.6	140 ± .20	24.0	0.60	90	6	50	6/14
851927	15.00	16.0	20.2	140 ± .20	25.0	0.70	60	6	50	6/14
851929	16.80	18.0	22.6	140 ± .20	25.0	0.70	60	7	50	6/14
851931	18.80	20.0	25.0	140 ± .25	24.0	0.70	60	5	50	6/14
851933	22.50	24.0	30.3	140 ± .25	26.5	0.60	35	6	50	6/14
851935	26.40	28.0	36.0	140 ± .25	26.5	0.70	50	6	50	6/14
851937	29.70	32.0	40.0	140 ± .25	27.0	0.70	40	6	50	6/14
851939	33.00	36.0	45.0	140 ± .25	26.0	0.70	50	7	50	6/14
851941	37.00	40.0	50.0	140 ± .25	25.0	0.75	25	6	50	6/14
851943	41.40	44.0	55.0	140 ± .25	25.0	0.70	30	6	50	6/14
851945	45.00	48.0	60.0	140 ± .30	28.0	0.80	30	9	50	6/14
851947	47.50	56.0	73.0	140 ± .35	28.5	0.80	40	10	50	6/14
851948	60.00	64.0	81.0	140 ± .35	32.0	0.80	25	9	50	6/14
851949	62.00	72.0	102.0*	140 ± 1.2	27.0	1.50	35	15	30	6/14
854101	64.00	80.0	113.0*	140 ± 1.2	28.0	1.50	35	12	30	6/14

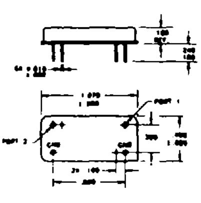
*Measured at 30 dB bandwidth.
All parts are guaranteed at 25°C.

Sawtek reserves the right to modify the above-listed specifications without notice when necessary to provide optimum performance and cost.

Package Drawings



6 / 14-Pin DIP



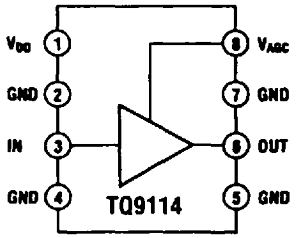
6 / 18-Pin DIP

TQS
WIRELESS COMMUNICATIONS

S E M I C O N D U C T O R , I N C .

TQ9114

IF/AGC Amplifier



The TQ9114 IF/AGC (Intermediate Frequency/Automatic Gain Control) Amplifier is part of TriQuint's RFIC Downconverter Building Block family. Intended for use as an Automatic Gain Control Amplifier in an IF stage, a wide range of gain control is available. The TQ9114 provides wide-bandwidth operation from a standard +5 V power supply. Its low current consumption and small, plastic surface-mount package are ideally suited for low-cost hand-held and battery-powered applications.

Features

- 30 – 500 MHz operation
- 30 dB gain @ 100 MHz
- 60 dB AGC range
- Single + 5 V supply
- 3.1 mA supply current
- SO-8 plastic package

Applications

- GPS (Global Positioning Systems)
- Cellular Communications
- Spread-Spectrum Receivers

Electrical Specifications

Test Conditions: V_{DD} = +5 V, T_A = 25 °C, Frequency = 100 MHz

Parameter (1)	Min.	Typ.	Max.	Units
Frequency of operation	30		500	MHz
Gain MHz	27	30		dB
AGC Range (2)	50	60		dB
DC Supply Current		3.1	4.5	mA
Gain Control Voltage	0		5	V

Note: 1. Min/Max values listed are production tested.
2. V_{AGC} = 0 V, Max. Gain, V_{AGC} = +5 V, Min. Gain
3. Voltages which produce Min and Max Gain

Figure 5.12 TQ9114 IF integrated circuit with 60 dB AGC range. (Courtesy of TriQuint Semiconductor.)

MOTOROLA

MC13155

Advance Information

Wideband FM IF

The MC13155 is a complete wideband FM detector designed for satellite TV and other wideband data and analog FM applications. This device may be cascaded for higher IF gain and extended Receive Signal Strength Indicator (RSSI) range.

- 12 MHz Video/Baseband Demodulator
- Ideal for Wideband Data and Analog FM Systems
- Limiter Output for Cascade Operation
- Low Drain Current: 7.0 mA
- Low Supply Voltage: 3.0 to 6.0 V
- Operates to 300 MHz

WIDEBAND FM IF

SEMICONDUCTOR TECHNICAL DATA

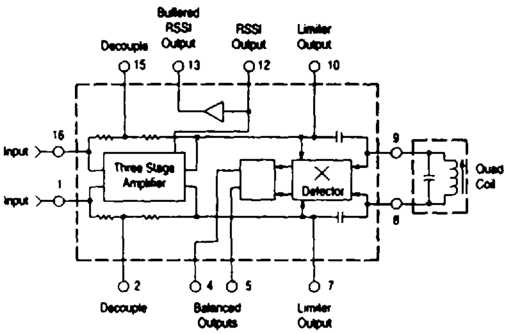
D SUFFIX
PLASTIC PACKAGE
CASE 751B
(SO–16)

MAXIMUM RATINGS

Rating	Pin	Symbol	Value	Unit
Power Supply Voltage	11, 14	V_{EE} (max)	6.5	Vdc
Input Voltage	1, 16	V_{in}	1.0	Vrms
Junction Temperature	–	T_J	+150	°C
Storage Temperature Range	–	T_{stg}	– 65 to +150	°C

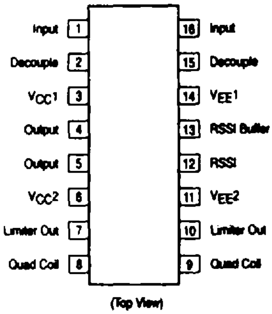
NOTE: Devices should not be operated at or outside these values. The "Recommended Operating Conditions" provide for actual device operation.

Figure 1. Representative Block Diagram



NOTE: This device requires careful layout and decoupling to ensure stable operation.

PIN CONNECTIONS



(Top View)

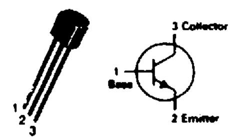
ORDERING INFORMATION

Device	Operating Temperature Range	Package
MC13155D	T_A = – 40 to +85°C	SO–16

Figure 5.13 MC13155 IF integrated circuit, with demodulator. (Courtesy of Motorola.)

MPSH34

CASE 29-04, STYLE 2
TO-92 (TO-226AA)



IF TRANSISTOR

NPN SILICON

MAXIMUM RATINGS

Rating	Symbol	Value	Unit
Collector-Emitter Voltage	V_{CEO}	40	Vdc
Collector-Base Voltage	V_{CBO}	40	Vdc
Emitter-Base Voltage	V_{EBO}	4.0	Vdc
Collector Current — Continuous	I_C	50	mAdc
Total Device Dissipation @ T_A = 25°C Derate above 25°C	P_D	350 2.8	mW mW/°C
Operating and Storage Junction Temperature Range	T_J, T_{stg}	−55 to +135	°C

THERMAL CHARACTERISTICS

Characteristic	Symbol	Max	Unit
Thermal Resistance, Junction to Ambient	$R_{\theta JA}$	357	°C/W

Refer to MPSH24 for graphs.

ELECTRICAL CHARACTERISTICS (T_A = 25°C unless otherwise noted.)

Characteristic	Symbol	Min	Typ	Max	Unit
OFF CHARACTERISTICS					
Collector-Emitter Breakdown Voltage (I_C = 1.0 mAdc, I_B = 0)	$V_{(BR)CEO}$	40	—	—	Vdc
Collector-Base Breakdown Voltage (I_C = 100 μAdc, I_E = 0)	$V_{(BR)CBO}$	40	—	—	Vdc
Emitter-Base Breakdown Voltage (I_E = 10 μAdc, I_C = 0)	$V_{(BR)EBO}$	4.0	—	—	Vdc
Collector Cutoff Current (V_{CB} = 30 Vdc, I_E = 0)	I_{CBO}	—	—	50	nAdc
ON CHARACTERISTICS					
DC Current Gain (I_C = 7.0 mAdc, V_{CE} = 15 Vdc) (I_C = 20 mAdc, V_{CE} = 2.0 Vdc)	h_{FE}	 40 15	 — —	 — —	—
Collector-Emitter Saturation Voltage (I_C = 7.0 mAdc, I_B = 2.0 mAdc)	$V_{CE(sat)}$	—	—	0.5	Vdc
Base-Emitter On Voltage (I_C = 7.0 mAdc, V_{CE} = 15 Vdc)	$V_{BE(on)}$	—	—	0.95	Vdc
SMALL-SIGNAL CHARACTERISTICS					
Current-Gain — Bandwidth Product (I_C = 15 mAdc, V_{CE} = 15 Vdc, f = 100 MHz)	f_T	500	720	—	MHz
Collector-Base Capacitance (V_{CB} = 10 Vdc, I_E = 0, f = 1.0 MHz)	C_{cb}	—	0.25	0.32	pF

Figure 5.14 Data sheet for MPSH34 transistor (typical low-cost device, characterized as an IF amplifier). (Courtesy of Motorola.)

5.4.1 Untuned Amplifier Design Procedure (see Figures 5.15 and 5.16)

1. Choose a collector operating point at $V_{cc} - V_{sat}/2$ if no resistor in the emitter is to be used. V_{sat} is the transistor saturation voltage at a given operating current (usually about 0.2 to 0.5 V).
2. Choose the transistor current to be used, based on its f_t specification, which is given at a particular collector current level.

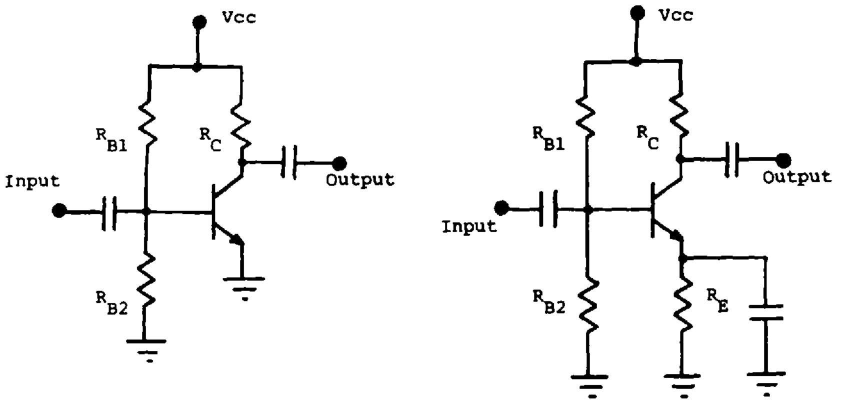


Figure 5.15 Untuned transistor amplifiers, with and without emitter resistor, employing voltage divider bias.

3. The collector resistor R_c is then

$$\frac{(V_{cc} - V_{sat})/2}{i_c} = \frac{V_{cc} - V_{sat}}{2i_c} = R_c \text{ (ohms)}$$

4. Using the transistor's specified current gain (beta or HFE), the base bias current is calculated as

$$i_b = \frac{i_c}{\beta} = \frac{i_c}{\text{HFE}}$$

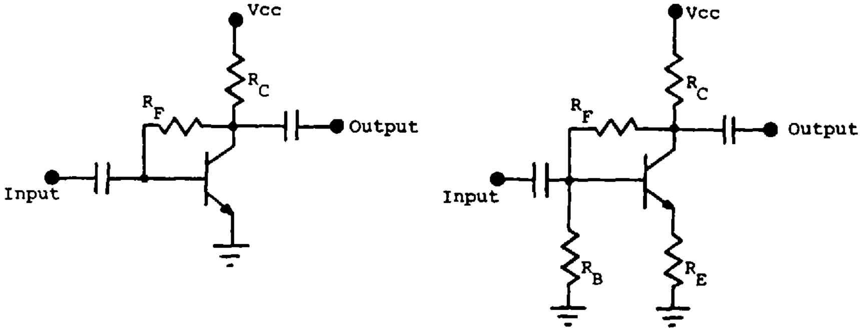


Figure 5.16 Transistor amplifiers employing feedback bias. Emitter and base resistors to ground are optional.

5a. If a voltage divider is to be used for biasing the base, then the values of R_{B1} and R_{B2} can be calculated as

$$\frac{V_{cc} - V_e - 0.7}{i_b} = \frac{R_{B1}R_{B2}}{R_{B1} + B_{B2}} = R_{source}$$

$$R_{B1} = \frac{R_{B2}R_{source}}{R_{B2} - R_{source}}$$

$$R_{B2} = \frac{R_{B1}R_{source}}{R_{B1} - R_{source}}$$

or with a typical value of R_{B2} at 100 K ohms,

$$R_{B1}\ (\text{ohms}) = \frac{[(V_{cc} - V_e - 0.7)/i_b]1 \times 10^5}{1 \times 10^5 - [(V_{cc} - V_e - 0.7)/i_b]}$$

In the case of no emitter resistor, V_e is zero. If an emitter resistor is used, then V_e is $(i_c + i_b)R_E$. (Here, 0.7 is the approximate forward voltage drop for the base to emitter junction of a silicon transistor.)

5b. Using feedback bias, as in Figure 5.16, a single resistor may be used (although a high-value 100K resistor from transistor base to ground may be added for stability). The value of the feedback resistor may be calculated as

$$R_{FB}\ (\text{ohms}) = \frac{V_c - V_b}{i_b}$$

Often, a resistor is added in the emitter circuit to ground, which increases the input impedance looking into the amplifier from the base by approximately

$$R_E \times H_{FE}$$

When such a resistor is present, the gain of the amplifier is reduced, because its gain is approximately

$$\frac{R_c}{R_E} \approx \text{amplifier gain}$$

Therefore, in high-frequency amplifiers, a capacitor (C_E in Figures 5.15 and 5.17) is connected across the emitter resistor to bypass the resistor at the frequency of interest. Typically, the capacitor's value is chosen to have reactance that is no more than about one tenth of R_E at the lowest frequency to be amplified.

It should be realized that when employing the feedback biasing approach, the amplifier's gain is reduced, because the feedback resistor not only sets the amplifier's DC operating point but also acts to provide negative feedback with respect to the signal being amplified. This provides higher fidelity and good stability but does reduce overall gain.

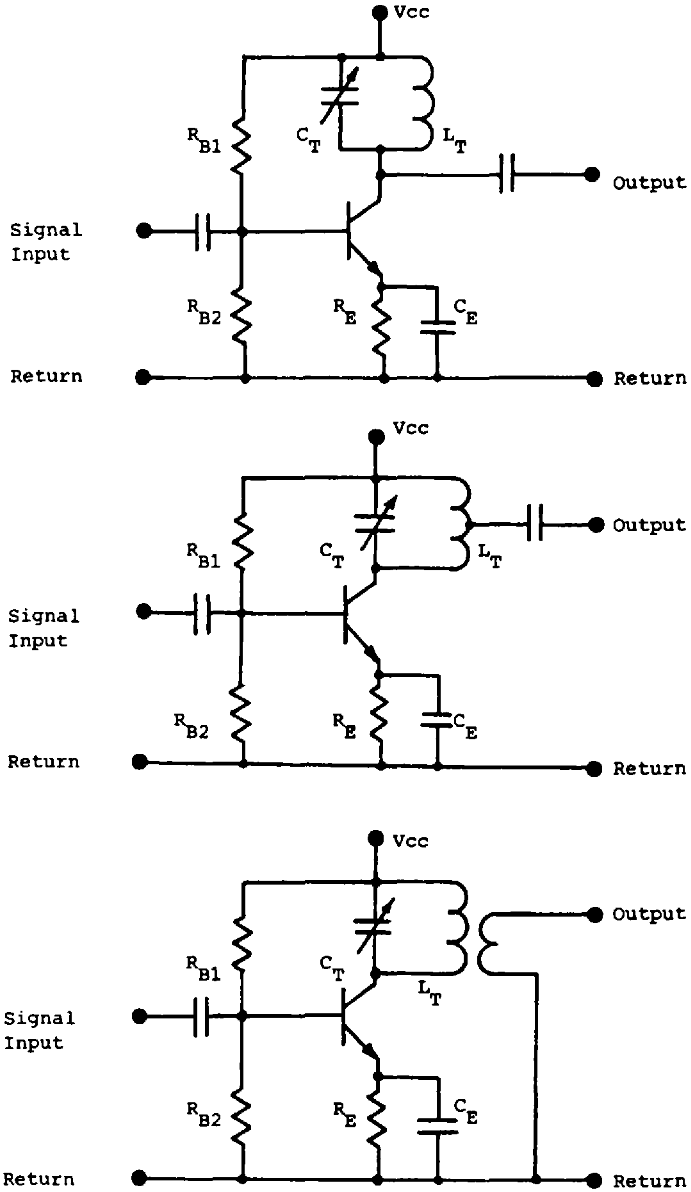


Figure 5.17 Tuned amplifier circuits with three output methods.

5.4.2 Tuned Amplifiers

Tuning an amplifier, by substituting either parallel or series tuned circuits for other components in the amplifier, is often done to make the amplifier more frequency selective. Figure 5.17 illustrates the most common tuned circuit configuration, where a parallel tuned circuit is used in the collector circuit. The primary difference between these alternatives is the method of taking an output signal from the amplifier, and the second and third approaches usually offer more flexible matching to other circuits.

Tuned amplifiers such as these (usually double tuned as in Figure 5.18) were the mainstay of older receivers developed before the advent of crystal, ceramic, and surface acoustic wave filters. They are still sometimes employed in the IF stages of low-cost mass-produced consumer receivers. In double-tuned amplifier circuits such as that of Figure 5.18, some bandwidth flexibility as well as frequency tuning capability is provided. It is possible to vary both the degree of transformer coupling and the tuned frequency of both the primary and secondary tank circuits. The primary drawback is that such amplifiers require periodic alignment, but they provide a certain degree of flexibility not available from fixed tuned filters.

In a given parallel tuned circuit, where the circuit is connected as shown in Figure 5.17, the bandwidth is a function of the Q of the tuning capacitor and its parallel capacitor. Normally, the Q of the capacitor is so much higher than the Q of the inductor that it is reasonable to assume that the inductor determines the circuit bandwidth. That is, for all practical purposes, the circuit Q is presumed to be

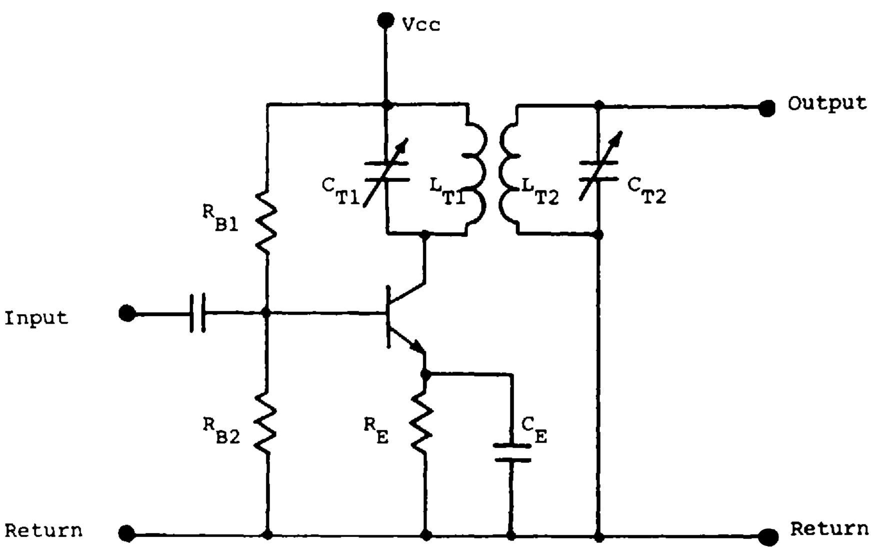


Figure 5.18 Double-tuned amplifier circuit.

$$Q = \frac{X_L}{r_L}$$

where the frequency to which the circuit is tuned is simply

$$F_c = \frac{1}{2\pi\sqrt{LC}}$$

Q is also defined as

$$Q = \frac{F_c}{BW_{3dB}}$$

A technique often employed to adjust the bandwidth of a tuned circuit is to add a resistor to the circuit in parallel with the inductor and capacitor. (The resistor could also be inserted in series with the inductor, but this is physically more difficult in most cases.) It should also be realized that the resistance that limits the Q of a tuned circuit is the resistance of the wire r_L used to construct the inductor.

Assuming that the circuit Q is X_L/r_L when the resistor R_P is not present, the question to be answered is, what is the value of resistance that should be added to adjust the tuned circuit for wider bandwidth? The circuit's impedance when resonant is resistive, $X_L = X_C$, and assuming $r_L \ll X_L$ as is the case for high Q, the impedance is $X_L/2 = X_C/2 = (X_L + X_C)/2$.

I. C. Hansen, in the *Radiotron Designer's Handbook*, shows that a parallel tuned circuit has Q that is

$$Q = \frac{1}{(1/R_p)\sqrt{L/C} + r_L\sqrt{C/L}}$$

Solving for R, we have

$$R_p = \frac{1}{Q(\sqrt{L/C} + R_p r_L\sqrt{C/L})}$$

and

$$1 = R_p(Q\sqrt{L/C} + QR_p r_L\sqrt{C/L})$$

which is of the quadratic form

$$0 = R_p^2 Q r_L\sqrt{C/L} + R_p Q\sqrt{L/C} - 1$$

Using the quadratic formula, therefore,

$$R_p = -\frac{Q\sqrt{L/C} \pm \sqrt{(Q\sqrt{L/C})^2 + 4Qr_L\sqrt{C/L}}}{2Qr_L\sqrt{C/L}}$$

(where only positive resistance values are valid).

5.5 NOISE BANDWIDTH OF FILTERS

The close relationship that exists between the amplifiers and filters used in modern receivers is such that together they determine the entire receiver's performance. The noise bandwidth of a receiver determines the noise power (KTB) to which the receiver is subjected, and the receiver's filters determine the noise bandwidth. In TRF receivers, the RF filter determines the noise bandwidth, whereas in superheterodyne receivers it is the IF filters that usually do the job. Receivers that employ coherent demodulators sometimes have a noise bandwidth that is determined by the bandwidth of the demodulator, but this occurs only when the tracking range of the (coherent) demodulator is much wider than the information bandwidth of the signal to be received. Such cases are found in satellite systems, for example, where a large amount of Doppler shift is expected and the filter ahead of the demodulator must be wide enough to accommodate the Doppler, but the demodulator is designed to track the signal within the tracking range, with a narrower loop bandwidth.

Further discussion of the noise bandwidth of coherent demodulators may be found in Chapter 8. Here we will discuss the noise bandwidth of filters such as those found in RF and IF stages of receivers. These filters are designed to have just enough bandwidth to pass the desired signal and are usually (for IF filters at least) the primary determiners of the noise bandwidth of receivers with noncoherent demodulators. Table 5.7 lists demodulators from the standpoint of coherence or lack thereof.

Where a noncoherent demodulator is preceded by an IF filter, a receiver's bandwidth is determined by the IF filter. The reason for this is that the noncoherent demodulator is not frequency selective and has greater bandwidth than the filter. The IF filter then must supply both frequency differentiation and noise rejection.

Coherent demodulators have the ability to act as tracking filters and center themselves around the incoming signal. If the tracking filter has bandwidth less

Table 5.7 Demodulator Listing According to Coherence or Noncoherence

Noncoherent demodulators	Coherent demodulators
Envelope detectors	Phase lock loops
Foster-Seeley discriminators	Squaring loops
Ratio detectors	Feedback-FM demodulators
Pulse-count discriminators	Costas loops
Quadrature detectors	*I-Q* loops
Dual-filter demodulators	Decision-directed loops
Squaring demodulators	

Table 5.8 B_N/B_{3dB} as a Function of Number of Tuned Circuits

Tuned, maximally flat series-connected circuits	Noise bandwidth to 3 dB bandwidth ratio
1	1.571
2	1.111
3	1.047
5	1.017
20	1.001

than that of the IF filter that precedes it, the overall receiver noise bandwidth is determined by the coherent demodulator.

The noise bandwidth of a coherent demodulator such as those listed in Table 5.7, compared with the demodulator's closed-loop bandwidth, is

$$B_N = 3.2B_L*$$

where B_N is the closed loop noise bandwidth and B_L is the closed loop 3-dB bandwidth. This holds for any loop that is type 2, second order, with damping factor 0.707, all of which are standard parameters for closed-loop, coherent demodulators.

Bandpass filters themselves have been characterized with respect to their noise bandwidth compared with their 3-dB bandwidth. In one case, the noise bandwidth of 1, 2, 3, 5, and 20 cascaded tuned circuits is given as a multiple of the 3-dB bandwidth of a single circuit. The comparison is shown in Table 5.8. Table 5.9 lists filter order (number of tuned circuits) versus B_N/B_{3dB} for three types of filters: Butterworth, Bessel, and Chebyshev with 1-dB ripple (chosen as a nominal value). (It should be noted that, regardless of filter type or bandpass ripple, the B_N/B_{3dB} ratio for a single filter section is 1.57. Also note the agreement between Table 5.8 and the Butterworth section of Table 5.9.)

When one considers the noise bandwidth of a receiver with respect to its IF filters, the most typical (and reasonable) approach is to consider that if there are at least three tuned circuits (or their equivalent) cascaded to determine IF bandwidth, the 3-dB bandwidth of the IF is for all practical purposes the noise bandwidth of the receiver. That is,

$$B_N \text{ (rcvr)} \approx B_{3dB} \text{ (narrowest IF)}$$

The exception that has already been mentioned is the case in which a coherent demodulator is employed, and

*See section 8.2 and the references in Chapter 8.

Table 5.9 B_N/B_{3dB} for Three Filter Types as a Function of the Number of Filter Sections

Type of filter	No. of sections (order of filter)	Noise bandwidth / 3 dB bandwidth
Butterworth	1	1.57
	2	1.11
	3	1.045
	4	1.025
	5	1.015
	6	1.01
Bessel	1	1.57
	2	1.155
	3	1.075
	4	1.04
	5	1.04
	6	1.04
Chebyshev (1 dB ripple)	1	1.57
	2	1.205
	3	0.995
	4	1.145
	5	0.915
	6	1.13

$$B_{3dB} \text{ (narrowest IF)} > 3.2B_L$$

so that

$$B_N \text{ (rcvr)} \approx 3.2B_L$$

where B_L is the closed-loop 3-dB bandwidth of the coherent demodulator.

5.6 OUTPUT AMPLIFIERS

Receiver output amplifiers are employed to interface with and drive devices that make use of demodulated information. These can include audio amplifiers as used in stereo systems and telephones, relay drivers as used in garage door openers and automobile door locks, line drivers for computer interfaces, and many other applications. In these varied applications, the impedances driven are in the range of 4 ohms to a few thousand ohms, and it is seldom necessary to employ discrete amplifiers. Output amplifiers for receivers are discussed in more detail in Chapter 8.

5.7 LIMITING AMPLIFIERS

A limiting amplifier is an amplifier that is designed to have a fixed output signal amplitude, no matter what its input signal amplitude. Such amplifiers are usually designed to incorporate a large amount of gain (often 80 dB or more) so that a small input signal will drive the amplifier's output to the maximum peak-to-peak level it can achieve, which is set by the saturation and cutoff points for the amplifier. (Saturation is the output voltage of an amplifier when it is being driven to its maximum current level, and cutoff is the output voltage when the amplifier current is zero.) A typical limiting amplifier might have output voltage when limiting of 2 V peak to peak, which with a gain of 80 dB would give it a "limiting sensitivity" of $2.0/10{,}000 = 2.0 \times 10^{-4}$ V peak to peak, or 141.4 μV RMS. This means that any time the input signal to the limiting amplifier is greater than 141.4 μV RMS, its output would be fixed at 0.707 V RMS. Such amplifiers may be employed instead of automatic gain control in some receivers.

There is, of course, a limitation to the ability of an amplifier to maintain exactly the same output level for all input levels. The degree to which the limiting amplifier (usually termed a "limiter") does maintain the same level determines whether it is described as a "hard" or a "soft" limiter. A hard limiter maintains its output level over a very small range, and a soft limiter's output may vary significantly more.

Although no specific rules about what a hard limiter is are known to this author, a good rule to consider is that used for AGC loops; that is, the output of a hard limiter should vary by less than 1 dB while its input varies from the minimum to the maximum level.

Historically, limiting amplifiers have been used in the intermediate frequency amplifier stages of FM receivers to reduce their sensitivity to incidental AM from all causes—whether from problems in the transmitter or the receiver or from propagation effects. It is important to realize that such receivers do not employ limiting before they pass the signal through a bandwidth-restricting filter. Such a filter is necessary to prevent signals from outside the band of interest from limiting the receiver. That is, it is desirable for a signal to which the receiver is tuned to limit, but it is undesirable for more than one signal to be present in the limiter, especially when the desired signal is not the largest signal present. Cahn [1], in a classic paper written in 1961, showed that whenever two signals are present in a limiter at the same time, the larger of the two signals suppresses the smaller signal (at the limiter output) by up to 6 dB. Experimental evidence shows that this is true. (See Figures 5.19 through 5.24.) Cahn's paper shows that, given two inputs to a limiter, $A_1 \cos \omega_1 t$ and $A_{1a} \cos \omega_2 t$, the limiter's output is

$$A_L = \frac{A_1(\cos \omega_1 t + a \cos \omega_2 t)}{A_1[1 + a^2 + 2a \cos(\omega_1 + \omega_2)]^{1/2}}$$

where a is the amplitude ratio of the two signals and is less than 1.0.

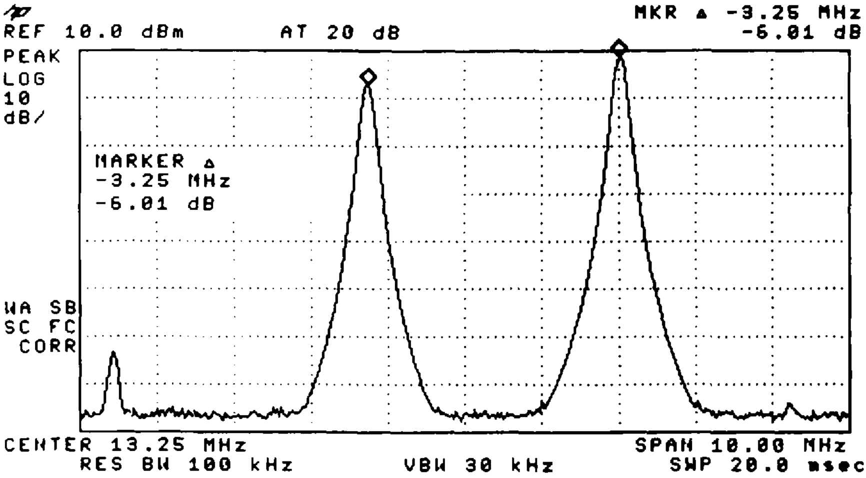


Figure 5.19 Limiter input, one signal 6 dB greater than the other.

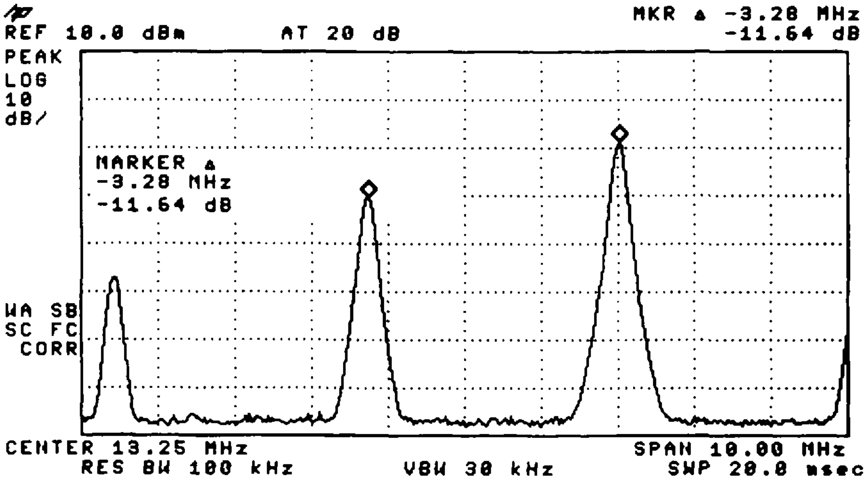


Figure 5.20 Limiter output, with input signals 6 dB different in power level. (Note that the output levels are 12 dB different—6dB more than at the input.)

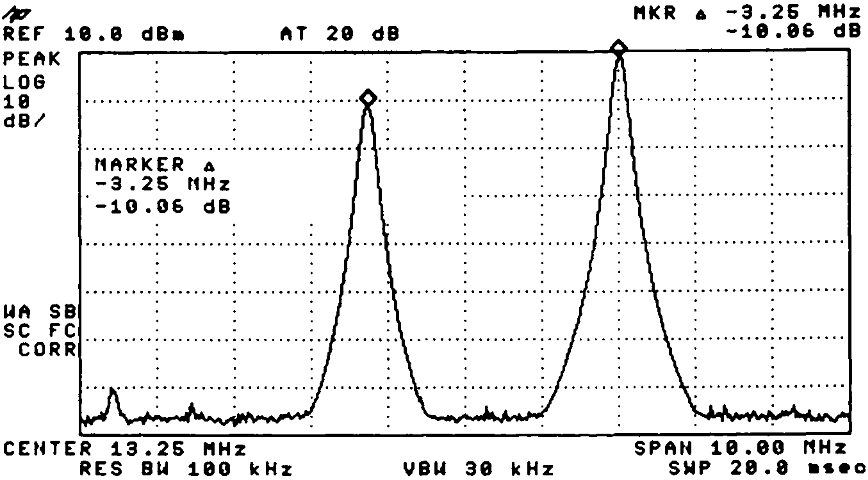


Figure 5.21 Limiter input, one signal 10 dB greater than the other.

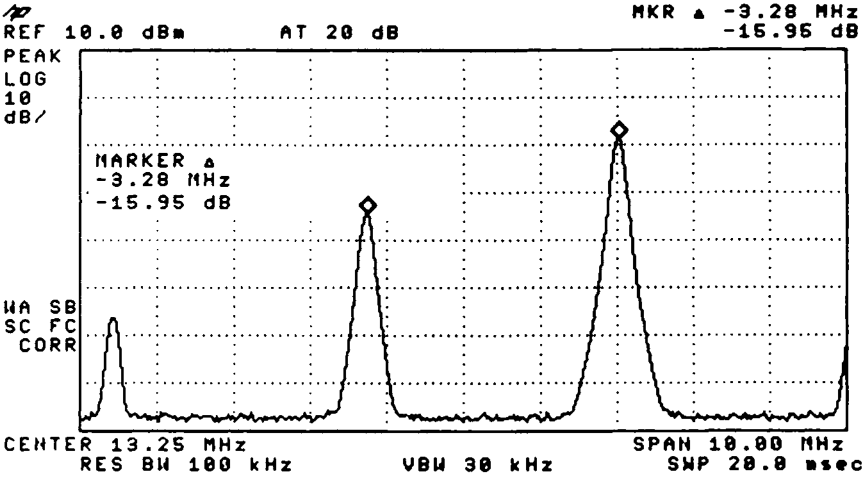


Figure 5.22 Limiter output, with input signals 10 dB different in power level. (Note that the output levels are 16 dB different—6 dB more than at the input.)

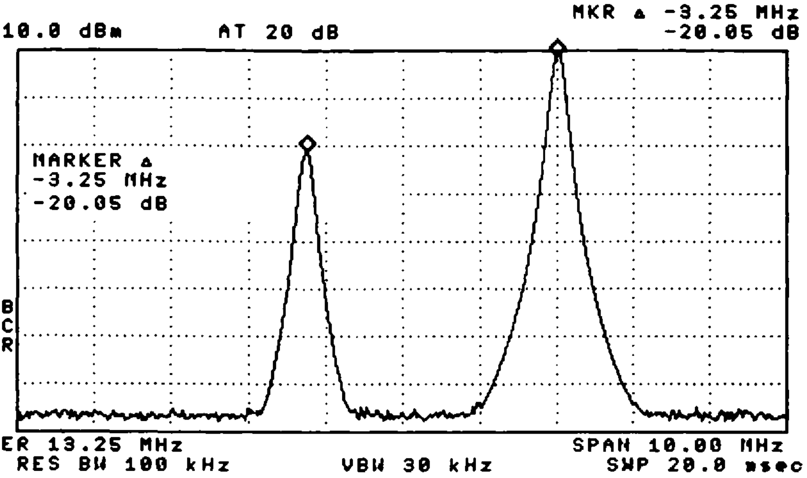


Figure 5.23 Limiter input, one signal 20 dB greater than the other.

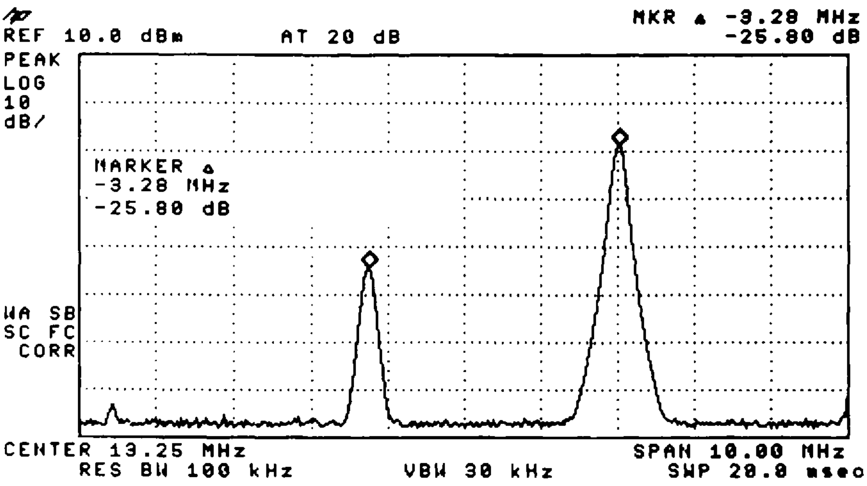


Figure 5.24 Limiter output, with input signals 20 dB different in power level. (Note that the output levels are 26 dB different—6 dB more than at the input.)

$$A_L = \cos\omega_1 t + \frac{a}{2}\cos\omega_2 t - \frac{a}{2}\cos(2\omega_1 - \omega_2)t \cdots$$

which shows that the larger signal at the input is still the largest at the limiter output, and the smaller signal at the input is smaller still at the output, when compared with the larger signal. Effectively, the limiter steals power from the smaller of the two input signals to generate the higher order products in the limiting process.

Any amplifier will limit if driven hard enough. It may not limit symmetrically, however. That is, the signal may not be limited at an equivalent level at both the top and the bottom of the output wave simultaneously. Such unsymmetrical limiting can readily occur if an amplifier is not biased so that its output is at the center of its output voltage range.

Sometimes, passive diode limiters may be employed instead of limiting amplifiers. Of course, amplifiers must be used to increase the signal level to a point at which a diode limiter can work. Figure 5.25 illustrates two different types of

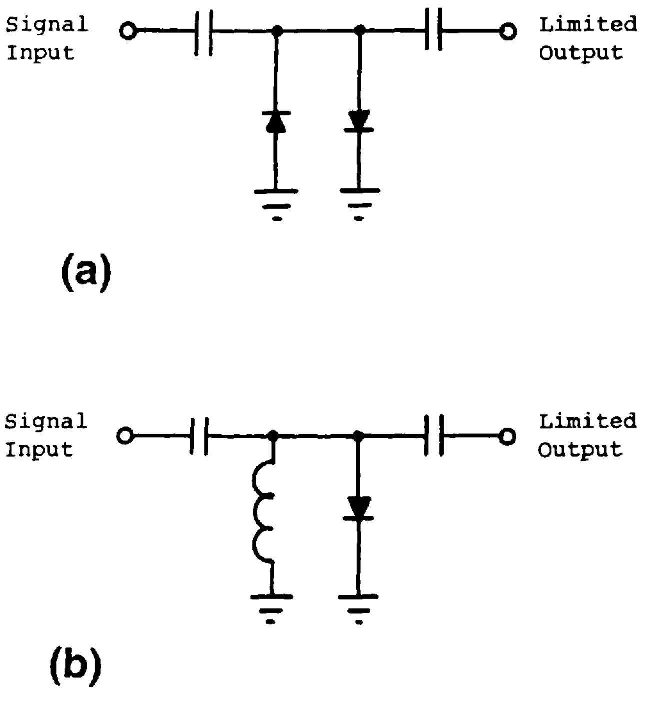


Figure 5.25 Diode limiter configurations employing back-to-back computer and PIN diodes. (a) Back-to-back diode limiter. (b) PIN diode limiter.

diode limiters. The back-to-back type can employ almost any type of diode, with silicon diodes limiting at approximately 1 V peak to peak and germanium diodes limiting at approximately 0.4 V peak to peak.

It is also worthwhile to note that limiters are sometimes employed at the input of a receiver for protection from large signals that otherwise might damage the preamplifier. Such limiters are usually designed to limit at a level of approximately 0 to +10 dBm, because integrated circuits are typically rated for a maximum of +10 to +15 dBm at their input, with respect to not being damaged by input signals. When a limiter is placed at the input of a receiver for protection, it should not be permitted to affect the desired signal, other than through whatever incidental insertion loss it may have. (This can be as low as a few tenths of a dB.)

5.8 LOG AMPLIFIERS

In some applications, amplifiers whose output level is a logarithmic function of their input level are employed to advantage. The reason for their use is usually that the signals being amplified are short-duration pulses, such as those that are found in a radar or IFF (identification friend or foe) receiver. In radar and IFF applications, received signals consist of short-duration pulses (usually 1 to 5 μsec in length) from multiple sources whose distances vary over a wide range. The resulting wide variation in signal amplitude between signals and the short signal duration make it impractical to employ automatic gain control in a receiver, as a receiver's AGC loop cannot respond quickly enough to the multiple independent signals that are present. This often forces the use of either a fixed-gain amplifier or a log amplifier. One common use of log amplifiers is in spectrum analyzers.

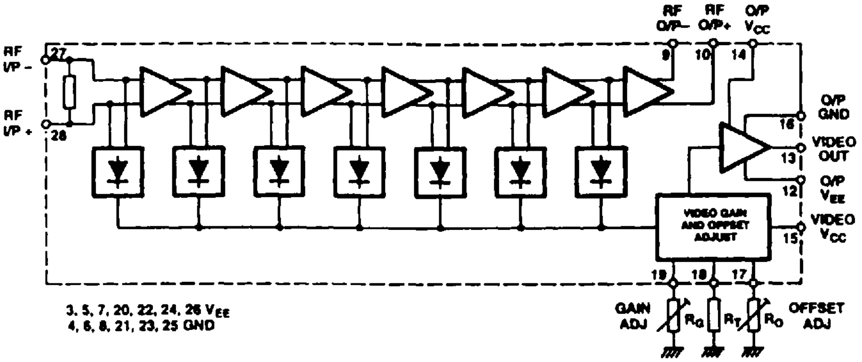


Figure 5.26 Block diagram of GEC Plessey type SL 3522 logarithmic/limiting amplifier. (Courtesy of GEC Plessey Semiconductors.)

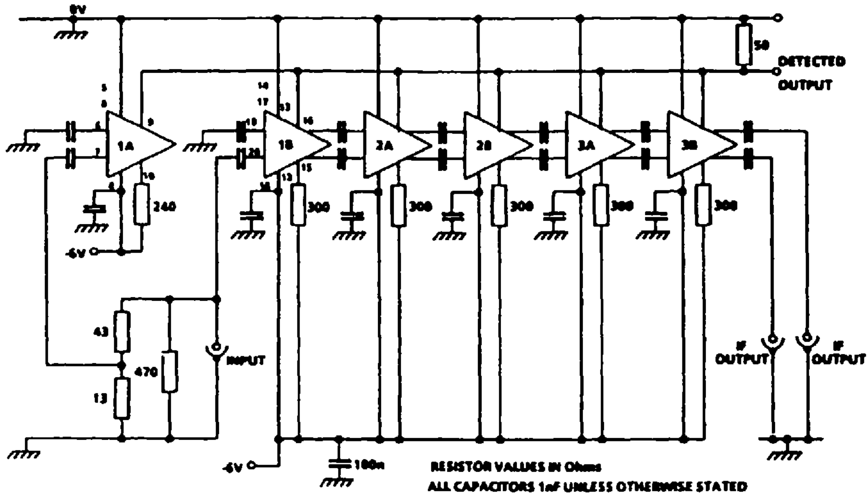


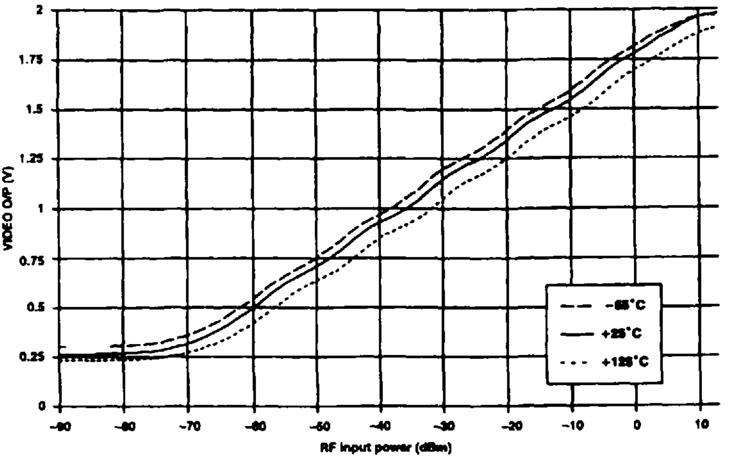
Figure 5.27 Block diagram of GEC Plessey type SL 2524, wideband logarithmic amplifier. (Courtesy of GEC Plessey Semiconductors.)

Log amplifiers often consist of several individual amplifiers in a series configuration, as shown in Figure 5.26. In the particular case shown in Figure 5.26, six log amplifiers are used, with an output limiting amplifier added. Figure 5.27 shows a somewhat different configuration that is found in another integrated circuit. Performance of these log amplifiers is shown in Figures 5.28 and 5.29. It should be noted that the logarithmic function of these amplifiers is facilitated through the use of the interstage diode detectors, which adjust the bias of the amplifiers, as a function of the signal level, thereby adjusting the signal output to provide a logarithmic transfer function. The diodes also serve to provide a detected signal level output (ideal for application in a spectrum analyzer).

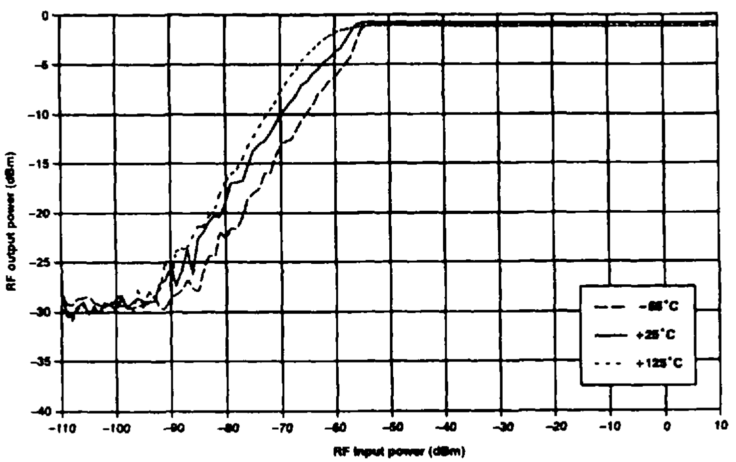
5.9 GENERAL-PURPOSE RF AMPLIFIER ICS (ALSO CALLED MMICS)

A very useful type of amplifier that fits well into many applications because of its size and low cost is the Hewlett-Packard (previously Avantek) MSA and INA series amplifiers.* These amplifiers are as small as many single transistors but require only a pair of coupling capacitors and a resistor to operate over a wide

*Other manufacturers also make similar devices.

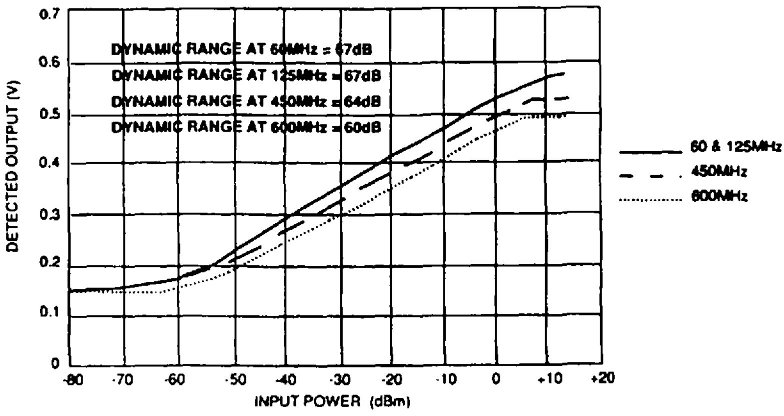


(a)

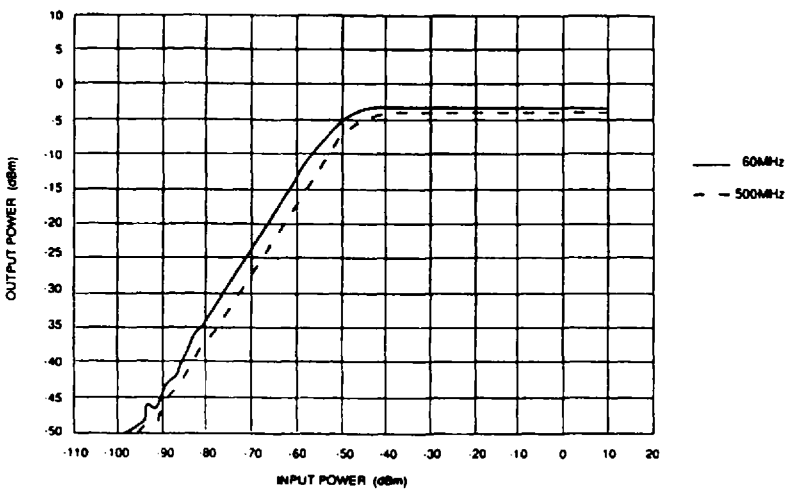


(b)

Figure 5.28 Transfer function for video and limiting outputs of SL 3522. (a) Video output versus RF input at 350 MHz. (b) Limiting output versus RF input at 350 MHz. (Courtesy of GEC Plessey Semiconductors.)



(a)



(b)

Figure 5.29 Detected output and limiting outputs of SL 2524. (a) Detected output versus RF input at several frequencies. (b) Limiting output versus RF input at two frequencies. (Courtesy of GEC Plessey Semiconductors.)

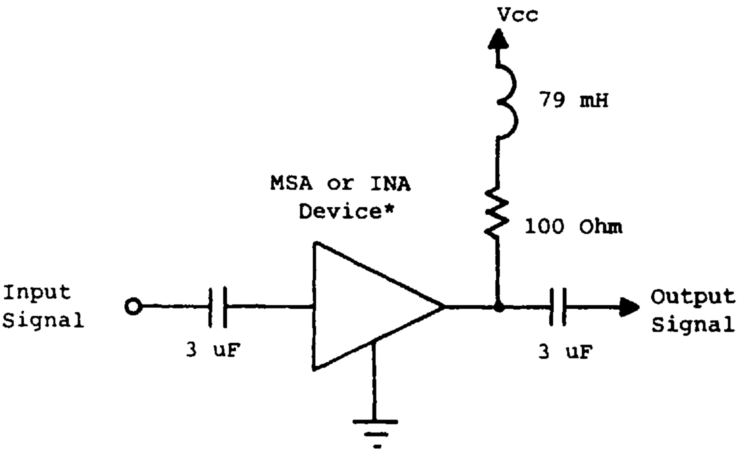


Figure 5.30 Schematic diagram showing amplifier using general-purpose IC wideband amplifier. Such devices are often termed MMICs, or monolithic microwave integrated circuits. *Values shown are for MSA-03184 with 5-V power supply, V_{cc}.

range of frequencies (up to about 2 GHz). They are also designed to provide 50-ohm input and output impedances, which makes them adaptable to many uses in receivers. Figure 5.30 illustrates the very simple construction of an amplifier circuit using an MSA-03184.

Coupling capacitors need only be large enough in value to have low impedance at the lowest frequency at which the amplifier is expected to operate. Typically, impedance of less than 1/10 the 50-ohm input and output impedance is acceptable. If the amplifier is expected to operate at 1000 Hz (the devices are rated to DC), then the coupling capacitor should be large enough to have 5 ohms impedance or less at that frequency. This would be

$$C = \frac{1}{5\omega^*} = 3.18 \text{ mF}$$

Of course, much smaller capacitors would be acceptable in higher frequency applications.

The biasing resistor required may be determined from the supply voltage minus the device voltage (4.0 V for an MSA-03184) with the difference divided by the device current (10 mA). Thus, for operation from a 5-V supply, the biasing resistor would be 100 ohms.

$$R = \frac{5 - 4}{0.01} = 100 \text{ ohms}$$

*ω is the usual $2\pi f$.

Because the 100-ohm resistor appears to be in parallel with the output, an inductor should be added in series with the resistor. For 500 ohms inductor impedance, its value should be $L = 500/\omega = 79$ mHy, for operation at 1000 Hz. Again, this component value would be much smaller for operation at higher frequencies. Dissipation in the 100-ohm resistor would be only $i^2R = 0.01^2 \times 100 = 10$ mW. With careful layout, this amplifier can provide more than 20 dB of gain up to 2 GHz.

QUESTIONS

1. If an amplifier's input signal is 0.1 V and its output is 15 V, what are its voltage and power gain? (Assume that input and output impedances are matched and are 50 ohms.)
2. If the preceding amplifier's input impedance is 50 ohms and its output impedance is 300 ohms (both matched), what are the voltage and power gain?
3. What is the output power in questions 1 and 2?
4. What is the difference between P_{1dB} and IP_3 for an amplifier?
5. Given that IP_3 is specified as +15 dBm and gain is 21 dB, what is IIP_3 for an amplifier?
6. If an amplifier's IIP_3 is given, how can one estimate P_{1dB} ?
7. Given an amplifier with IP_3 at +10 dBm and input signals at −20 dBm, what level of third-order output should be expected?
8. In a receiver whose preamplifier has a noise figure of 2.5 dB, an attenuator whose insertion loss is 1.2 dB is to be added. What is the effect on the amplifier? What is the effect on the receiver?
9. An amplifier with 2 dB noise figure and 18 dB gain is to be used to drive a mixer with 10 dB noise figure and 6 dB gain. Can this combination work for a receiver with 100 kHz noise bandwidth and 1.0 μV specified sensitivity? Assume a 10 dB IF noise figure.
10. What Q would be required for an amplifier to operate with a bandwidth of 50 kHz at 10.7 MHz?
11. What is the effect of adding a variable attenuator ahead of a preamplifier? After the preamplifier?
12. What are the advantages and disadvantages of employing a limiting IF amplifier in a receiver?

REFERENCES*

1. Cahn, C. R., A note on signal-to-noise ratio in bandpass limiters, *IEEE Trans. Inform. Theory*, Jan. 1961.

*Also see references for Chapter 6.

2. *Professional Products IC Handbook*, GEC Plessey Semiconductors, Jan. 1996, pp. 13–43.
3. *Wireless Communications Products Databook*, TriQuint Semiconductor, Spring 1994, pp. 2–3, 2–9, 2–21.
4. Ayrandjian, N., Simple computation of spurious-free dynamic range, *RF DESIGN*, Jan. 1987.
5. Emery, F. E., Solid-State Limiting Amplifiers, Watkins-Johnson Tech-Notes, Sept./Oct. 1978.
6. Henderson, B. C., and Zazkowski, J. M., Integrated Mixer-Preamplifiers, Watkins-Johnson Tech-Notes, May/June 1988.
7. Galla, T. J., Cascaded Amplifiers, Watkins-Johnson Tech-Notes, May/June 1989.
8. Galla, T. J., A handy "how-to-use" guide for MAR monolithic drop-in amplifiers, *Mini-Circuits*.
9. Galla, T. J., Plain Talk on Log Amps, App. Note M529, M/A-COM, 1994.
10. Spectrum Analysis ... Distortion Measurement, Spectrum Analyzer Series App. Note 150-11, Hewlett-Packard, Oct. 1976.
11. McVay, F. C., Don't guess the spurious level of an amplifier, *Electron. Design*, Feb. 1967.
12. Shelton, R. D., and Adkins, A. F., Noise bandwidth of common filters, *IEEE Trans. Comm.*, Dec. 1970.
13. Useful Formulas, Tables, and Curves for Random Noise, IN-103, General Radio Co., June 1963.
14. Froelich, R., Measurement of GaAs FET Noise Parameters, Watkins-Johnson Tech-Notes, Nov./Dec. 1986.
15. Spectrum Analysis ... Noise Figure Measurement, Spectrum Analyzer Series App. Note 150-9, Hewlett-Packard, Apr. 1976.

6
Mixer/Converter Design

Mixers, or frequency converters, are also sometimes called multipliers, and in the early days of superheterodyne receivers they were often called "detectors." (In old, 1920–1950 vintage, receiver diagrams, the first mixer was often called the first detector.) Whatever name it is known by, the requirement is that the device that we will call a mixer has the ability to:

- Convert the desired signal from the received frequency to the receiver's first IF frequency (or other IF frequency)
- Convert the signal with minimum distortion and/or additive noise
- Convert the signal with minimum loss
- Convert the signal with good frequency accuracy

6.1 MIXER ACTION

The process by which a mixer operates is the nonlinear processing of a pair of signals in such a way that the sum and/or the difference of the two signals' frequencies is produced. This can be accomplished with either diodes or amplifiers, as any device with a nonlinear transfer function will work. In some cases, a "harmonic mixer" may be employed. Such mixers are designed to emphasize the products of harmonics of either the input signal, the local oscillator, or both. They are usually used when a higher local oscillator frequency is required than is actually available.

Ideally, the nonlinearity is a perfect square law function, although in practice even near perfection is difficult to achieve. Given a pair of signals $A \cos \omega_1 t$ and $B \cos \omega_2 t$ being processed in a nonlinear device with square law properties, the result is

$$
\begin{aligned}
&(A\cos\omega_1 t + B\cos\omega_2 t)^2 \\
&= A^2\cos^2\omega_1 t + AB\cos(\omega_1 + \omega_2)t + AB\cos(\omega_1 - \omega_2)t + B^2\cos^2\omega_2 t \\
&= \frac{A^2}{2}(1 + \cos 2\omega_1 t) + \frac{B^2}{2}(1 + \cos 2\omega_2 t) \\
&\quad + AB\cos(\omega_1 + \omega_2)t + AB\cos(\omega_1 - \omega_2)t
\end{aligned}
$$

which consists of two second-harmonic terms and the sum and difference of the original two frequencies. A bandpass filter is normally employed to select the desired frequency, which is usually the difference of the local oscillator and input frequencies. The bandpass filter also rejects the second-harmonic signals from the mixing process.

Note that the square law device as a mixer suppresses the local oscillator and the signal frequencies. Such a mixer is called a "balanced" mixer and will be discussed further in the following pages. Where mixers that are not balanced are employed, the sum and difference frequencies, harmonics, and the original input frequencies are present in the mixer output. Figures 6.1 and 6.2 show output spectra of a balanced mixer and another mixer that is not balanced.

Table 6.1 lists the basic characteristics of typical mixers of three types, diode (passive), transistor, and Gilbert cell. Although transistor mixers may be balanced,

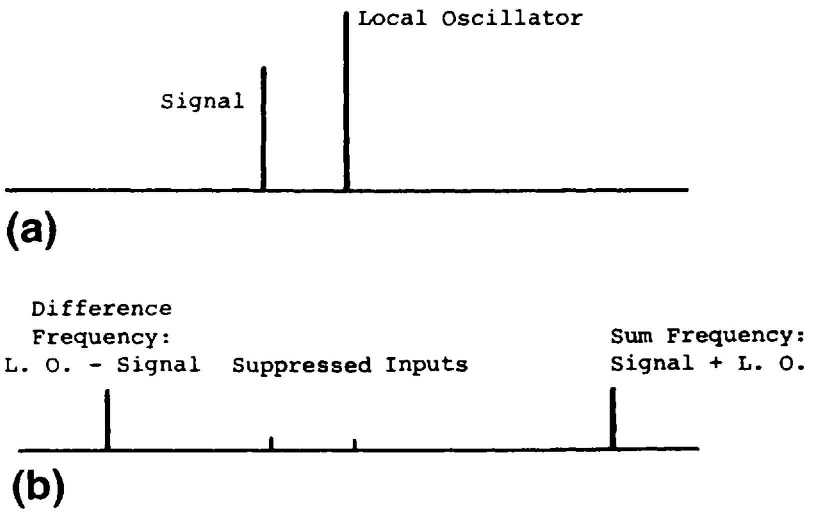


Figure 6.1 Typical input and output signal spectra for a balanced mixer (in this case, double balanced). Signals shown are not to scale. (a) Mixer input, signal and local oscillator. (b) Balanced mixer output.

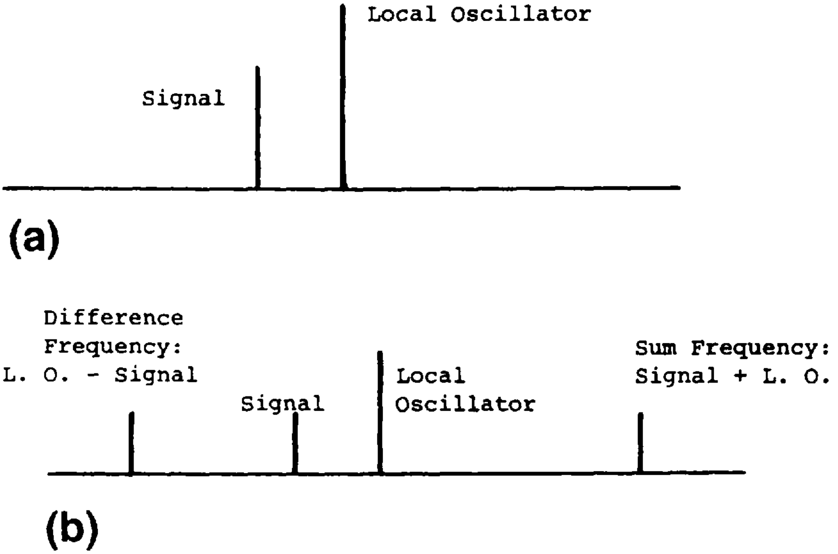


Figure 6.2 Typical input and output spectra for standard mixer (not balanced). Signals shown are not to scale. (a) Mixer input; signal and local oscillator. (b) Mixer output; mixer not balanced.

such mixers are usually difficult to design and produce. Transistor mixers of both bipolar and field effect types are also possible.

With respect to other mixer actions, mixers may add noise to the signals that they convert due to either noise on their local oscillator input or noise generated within themselves. In general, passive mixers have very low internal noise (less than 1 dB), and active mixers have noise figures that are in the range 3 to 20 dB. Remember that whatever the noise figure of the first mixer in a receiver, the preamplifier must have lower noise figure and higher gain to prevent the mixer from dominating the receiver sensitivity.

Table 6.1 Some Typical Mixer Characteristics

Mixer type	Active/ passive	Gain	LO level	Balance	Noise figure
Diode	Passive	−6 to −10 dB	0 to +27 dBm	Yes/no	Very low
Transistor	Active	0 to +20 dBm	−10 to 0 dBm	Not usual	To 10 dB
Gilbert cell	Active	5 to +20 dBm	−10 to +10 dBm	Yes/no	To 20 dB

Mixers that have no active amplifier within themselves must have loss and this loss is a minimum of 6 dB, with typical loss in a passive mixer at 7 to 10 dB, depending on the amplitude of the local oscillator used. A general rule for the local oscillator drive level in a passive mixer is that it should be at least 7 dB greater than the largest signal to be converted. It is often a practice to "starve" a mixer with respect to local oscillator drive level, but this is done at the expense of increased "conversion loss."

Active mixers amplify their inputs and/or their output, and this can provide net conversion gain. The advantages of active mixers are that they are low in cost and they usually require a lower local oscillator drive level.

Frequency accuracy is not a function of a mixer. The mixer must rely on the accuracy of its input signals, as its output is a product of its input signals.

6.2 PASSIVE MIXERS

Passive mixers, which typically use one or more diodes as their nonlinear element, are used primarily in high-quality applications such as military receivers. The reasons for this are that

1. High-performance passive mixers are readily available at low development cost,
2. These mixers cover wide frequency ranges and signal levels, and
3. They offer double-balanced operation.

A typical double-balanced, diode ring mixer of the type currently used in many applications is shown in schematic form in Figure 6.3. This type of mixer is called double balanced because both of the input signals are suppressed at the output. The suppression capability may be understood by observing Figure 6.4,

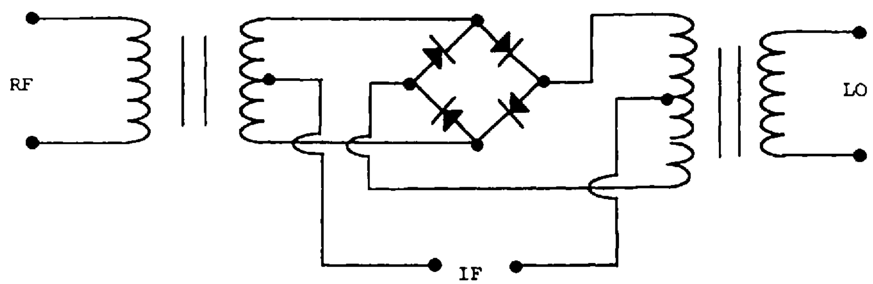


Figure 6.3 Schematic diagram of double-balanced, diode ring mixer. Signals may be interchanged on RF and LO ports.

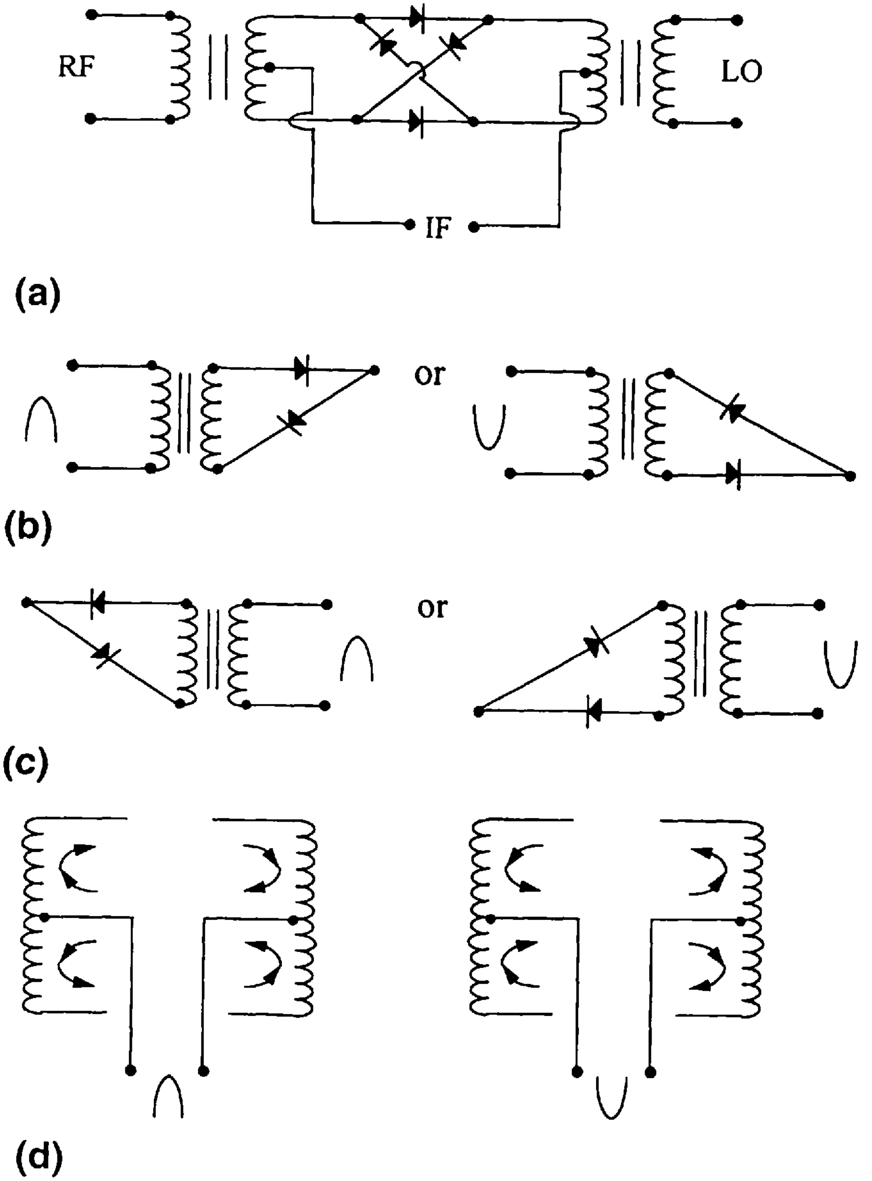


Figure 6.4 Illustration showing input signal suppression of diode ring mixer. (a) Alternate drawing of Figure 6.3. (b) View of diode ring mixer from RF port. (c) View of diode ring mixer from LO port. (d) View of diode ring mixer from IF port.

which considers the diode ring mixer at its input ports. Figure 6.4a is the same as Figure 6.3, except that the diodes are drawn as a lattice instead of a ring, and Figures 6.4b through d are breakdowns that show portions of the overall mixer.

In Figure 6.4b, the mixer is shown from the viewpoint of the RF port. When the input signal at the RF port swings positive, two diodes in the ring conduct, and when the input signal swings negative the other two diodes conduct. Therefore the input signal at the RF port sees a transformer whose secondary is shorted through a pair of diodes. So does a signal input at the LO port.

Signals input at the IF port are suppressed by a different mechanism. Balanced transformer windings at both input and return connections of the IF port provide the signal cancellation desired.

At this point, if all of the input signals are either shorted through diodes or canceled by a transformer, you may be wondering, "How can the diode ring mixer work?" It does so because of the mutual intermodulation of the current in the diodes, which generates products of the input signals but does not pass the input signals themselves to the mixer output. (In practice, input signals are not completely suppressed at the output because of imperfections in the components [transformers and diodes] used. Suppression from input port to output port that is around 30 dB is practical, however.)

Note: The port-to-port suppression of signals is commonly referred to as port-to-port isolation. That is, when concerned with the amount of local oscillator signal appearing at a mixer's signal input port, one would specify the local oscillator to input port-to-port isolation.

Passive mixers are manufactured by a number of companies. They are small, inexpensive, and offer excellent performance. For these reasons, passive mixers are employed in many applications, but not usually employed in consumer products. The reason is that the best passive mixers (the diode ring type) do not lend themselves to large-scale mass production as well as active mixers do. The amount of labor necessary to construct and test a passive mixer limits cost reduction, so that even in very large quantities, the price of a passive mixer is still in the $2 range. An active mixer designed for a specific application costs more in development time (no development is necessary for a passive mixer), but once it is designed, it costs a few cents and can be constructed on a printed circuit board by a machine. Of course, simple diode mixers can be just as inexpensive as an active mixer, but their performance is much poorer than that of either a diode ring passive mixer or a good active mixer.

Figure 6.5 shows three other passive, diode mixers, one of which is "single balanced." All three are easily constructed and relatively inexpensive but not readily implemented in integrated circuit form. The single diode mixers work best with forward bias, but bias is not required for the balanced mixer of Figure 6.5b.

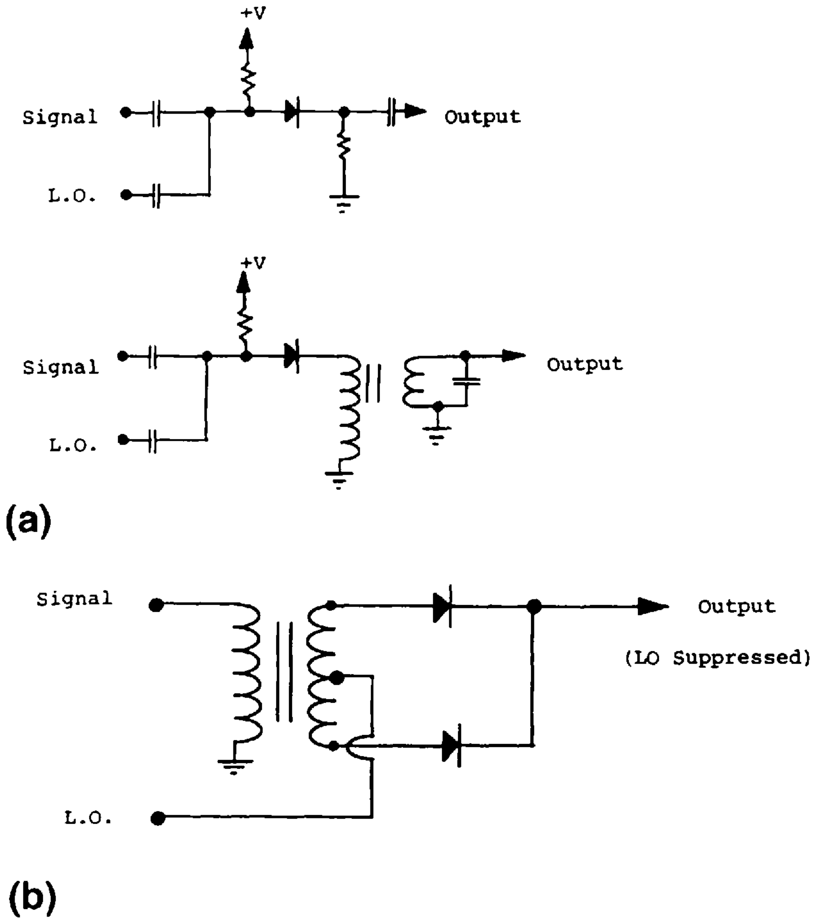


Figure 6.5 Diode mixers, unbalanced and single balanced. (a) Single diode mixers, untuned and tuned output. (b) Single-balanced diode mixer. (For suppression of the signal instead of the local oscillator, reverse the inputs.)

6.3 ACTIVE MIXERS

Active mixers are basically amplifiers that have nonlinear transfer functions. That is, an amplifier (with or without significant gain) that distorts its input signal(s) can be used as a mixer. All that are required are some method of input for more than one signal and some method of selecting the desired product of the input signals.

This process can be simply implemented as shown in the circuits of Figures 6.6 and 6.7.

In Figure 6.6, bipolar transistor mixers are shown in two configurations. In the first (Figure 6.6a) an input signal and a local oscillator signal are both connected to the transistor's base. This results in both signals being amplified simultaneously, which causes nonlinear interaction between the two signals and produces the desired mixing products.

Figure 6.7 shows mixers similar to those in Figure 6.6 but employing field effect transistors (FETs). One significant difference between the FET mixers and bipolar mixers is that dual-gate FETs are readily available wherein improved isolation is possible.

Figure 6.8 illustrates the difference between two signals that are linearly processed and the same two signals when processed nonlinearly (mixed). The difference between the linear and nonlinear processing cases may be as simple as a different biasing point on the transistor stage.

A dual-gate FET mixer as employed in the ARC-164 receiver is shown in Figure 6.9. The same basic mixer is employed in the ARC-164 for both the first and second mixers; the first mixer operates over an input RF range of 225 to 400 MHz, and the second mixer's input is 70 MHz. The first IF frequency in the ARC-164 receiver is 70 MHz, and the second IF frequency is 30.1 MHz. FET mixers are employed to provide wide operational dynamic range, with good local oscillator to input isolation.

An active mixer using three transistors in a differential amplifier configuration is shown in Figure 6.10. A derivative of this type of mixer, known as a Gilbert cell, is common in integrated circuits, where they are employed to frequencies as high as 12 GHz. It is interesting to note that the input and the output from the differential pair of transistors may be either single ended or differential (push-pull). This leads to the possibility of a balanced output for suppression of some fundamental and odd harmonic components.

Selecting the desired frequency component is accomplished by either tuning the differential transistor collector circuits or placing a filter at the mixer output.

Again, we emphasize that any amplifier will work as a mixer* if it is either overdriven or biased at a point at which the signals input to it are distorted. At this point its transfer function is nonlinear, and although it may not be an ideal mixer because of producing unwanted products, it will still work as a mixer.

6.4 IMAGE-REJECTION MIXERS

A special mixer architecture that is sometimes employed in receivers is an "image-rejection mixer." Such mixers are actually a pair of mixers, with the

*Unfortunately, any mixer will not work as an amplifier.

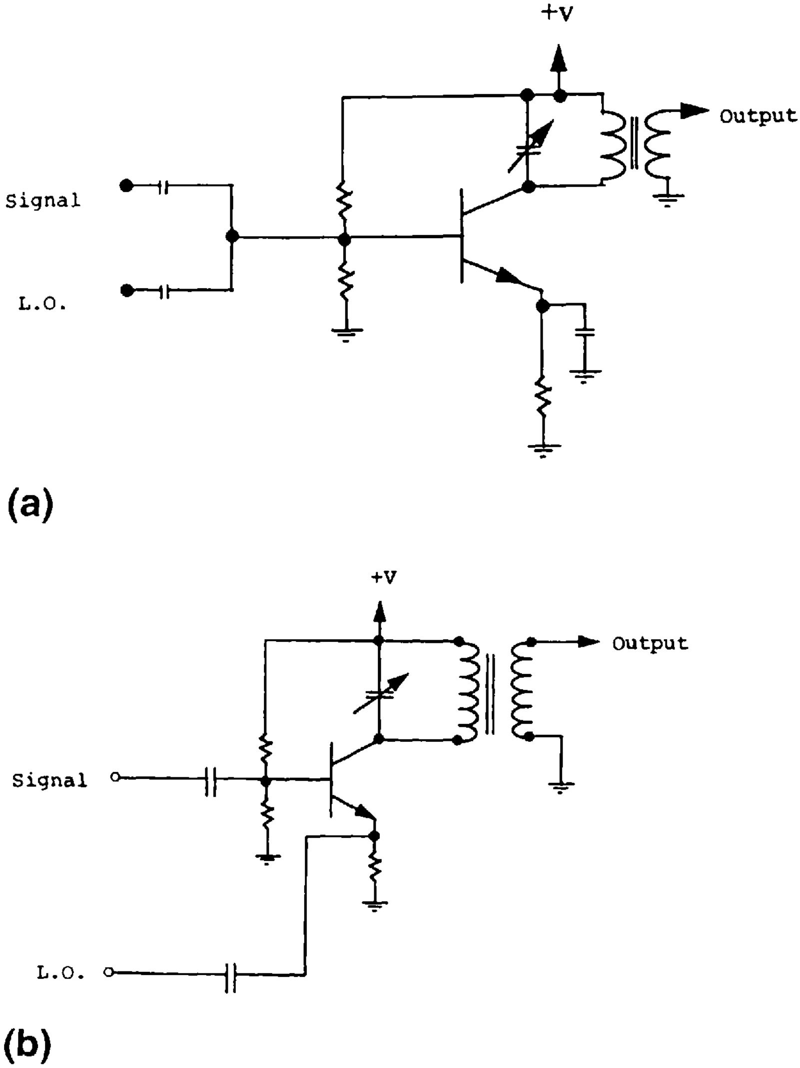


Figure 6.6 Typical, active bipolar transistor mixers. (a) Base input, bipolar mixer. (b) Base and emitter input, bipolar mixer.

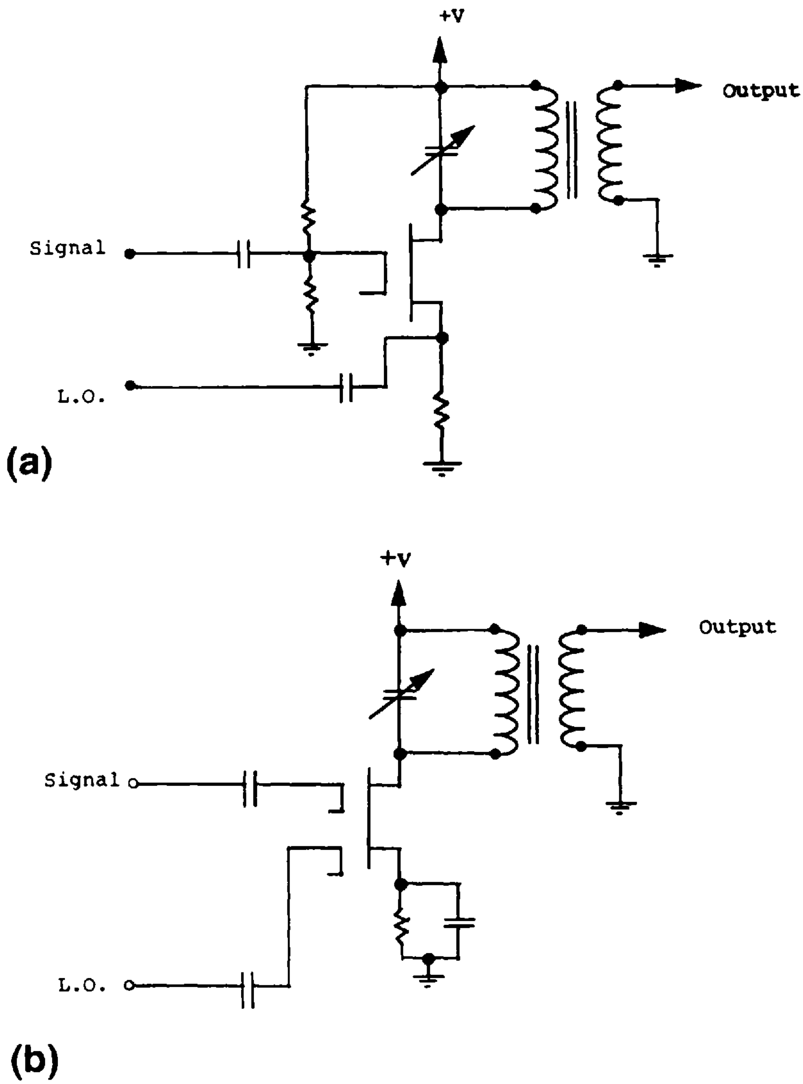
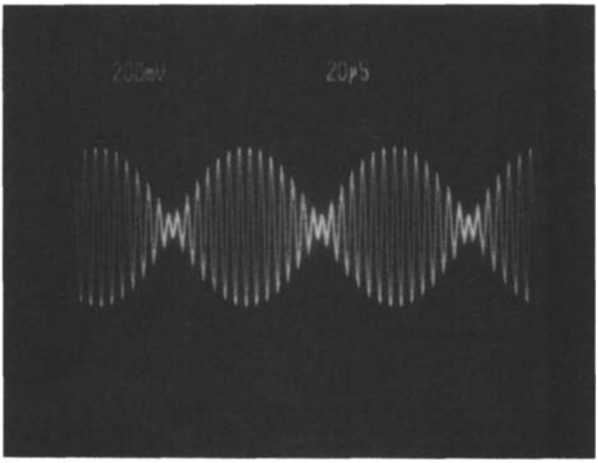
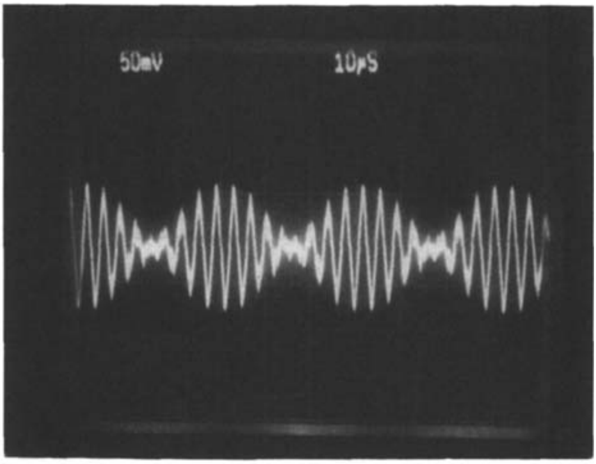


Figure 6.7 Dual-gate field effect mixers. (a) Gate and source input. (b) Dual gate input (gate bias not shown).



(a)



(b)

Figure 6.8 Illustration of difference between linear and nonlinear processing of two signals. (a) Frequency A and frequency B summed together linearly. Frequency difference = 15 kHz. (Note that there is no phase shift in the apparent carrier at any point.) (b) Frequency A and frequency B multiplied together. Frequency A = 250 kHz, frequency B = 15 kHz. DSBSC modulation. (Note phase shifts between amplitude packets.)

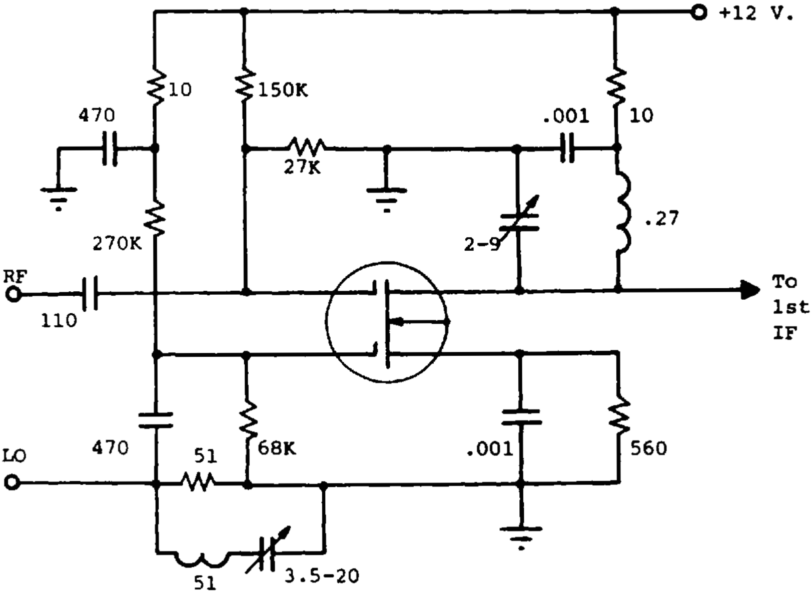


Figure 6.9 First mixer as used in ARC-164 receiver.

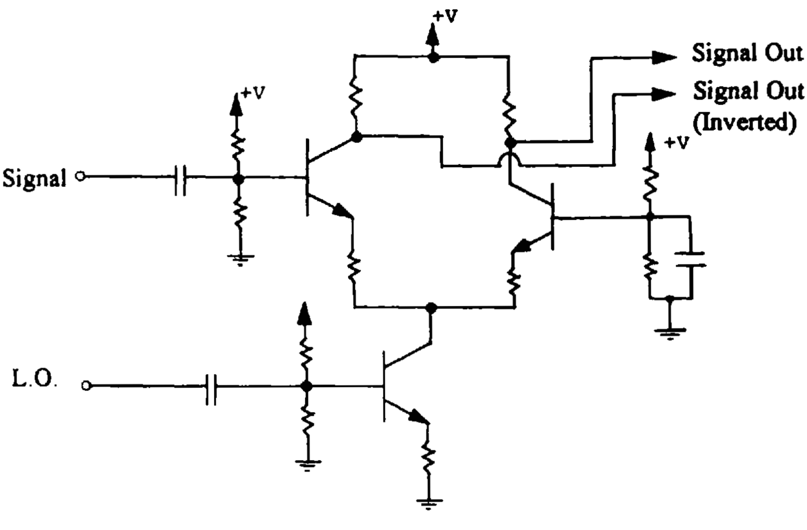


Figure 6.10 Differential amplifier as a mixer. Both input and output can be either differential or single ended.

components arranged to help suppress a receiver's image response. It should be realized, however, that an image-rejection mixer cannot provide sufficient rejection by itself to meet typical receiver requirements. (Typical receiver image rejection specifications are 80 dB or more, but an image-rejection mixer usually does well to provide 30 dB or image suppression.)

Image-rejection mixers are the same as the device also called a "quadrature single-sideband generator." Figure 6.11 is a block diagram of such a device, which consists of two 90 degree power splitters, two double balanced mixers, and either a zero degree or 180 degree power adder.

Two input signals, $A \cos \omega_1 t$ and $B \cos \omega_2 t$, input to the signal and local oscillator ports, are each split into quadrature components. These are $A \cos \omega_1 t$ with $A \sin \omega_1 t$, and $B \cos \omega_2 t$ with $B \sin \omega_2 t$, respectively. Of these, $A \cos \omega_1 t$ is used to modulate $B \cos \omega_2 t$, and $A \sin \omega_1 t$ is used to modulate $B \sin \omega_2 t$.

The modulated signals produced are

$$A \cos \omega_1 t \, B \cos \omega_2 t = \frac{AB}{2} \cos(\omega_1 + \omega_2)t + \frac{AB}{2} \cos(\omega_1 - \omega_2)t$$

and

$$A \sin \omega_1 t \, B \sin \omega_2 t = \frac{AB}{2} \cos(\omega_1 - \omega_2)t + \frac{AB}{2} \cos(\omega_1 + \omega_2)t$$

When these modulated signals are added together in phase, the signal produced is $AB \cos(\omega_1 - \omega_2)t$ (the desired result when down-converting), but when they are added together out of phase the signal produced is

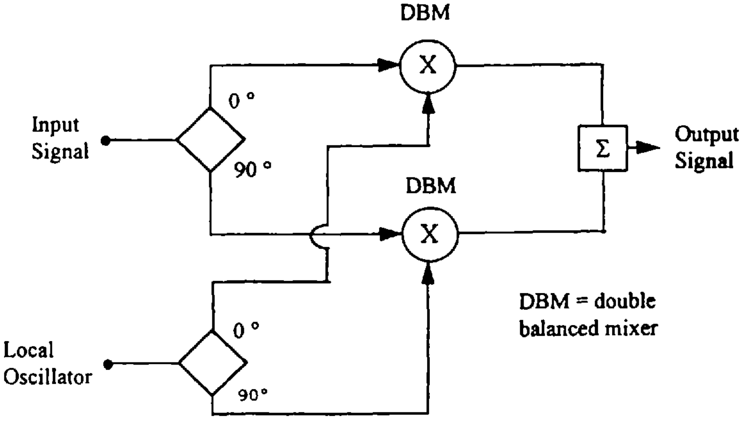


Figure 6.11 Block diagram of image-rejection mixer.

$$AB\cos(\omega_1 + \omega_2)t \qquad \text{(the desired result when up-converting)}$$

and these two signals are the sum and difference of the two inputs, where the undesired product is suppressed. Now, if an image signal is present, which is separated from the desired signal by an amount $2\omega_{IF}$, then one of the signals is

$$A\cos\omega_1 t$$

and the other is

$$B\cos(\omega_2 + 2\omega_{IF})t \qquad \text{(high-side injection)}$$

or

$$B\cos(\omega_2 - 2\omega_{IF})t \qquad \text{(low-side injection)}$$

which are paired with $A\sin\omega_1 t$ and $B\sin(\omega_2 + 2\omega_{IF})t$ or $B\sin(\omega_2 - 2\omega_{IF})t$, respectively.

When down-converting, the local oscillator $B\cos\omega_2 t$ frequency is higher than $A\cos\omega_1 t_1$ if high-side injection is used, and lower if low-side injection is used. That is, ω_2 may be less than or greater than ω_1.

We will let $\omega_3 = \omega_2 \pm 2\omega_{IF}$. Now, the pair of modulators produce

$$A\cos\omega_1 t\, B\cos\omega_3 t = \frac{AB}{2}\cos(\omega_1 + \omega_3)t + \frac{AB}{2}\cos(\omega_1 + \omega_2)t$$

and

$$A\sin\omega_1 t\, B\sin\omega_3 t = \frac{AB}{2}\cos(\omega_1 + \omega_3)t - \frac{AB}{2}\cos(\omega_1 - \omega_3)t$$

If these are summed in phase, the result is $AB\cos(\omega_1 + \omega_3)t$, and if they are summed out of phase, the result is $AB\cos(\omega_1 - \omega_3)t$. We are interested here in the first result, $AB\cos(\omega_1 + \omega_3)t$.

Replacing ω_3 with $\omega_2 \pm 2\omega_{IF}$, we have

$$AB\cos(\omega_1 + \omega_2 \pm 2\omega_{IF})$$

which falls outside the IF filter range.

The image-rejection mixer does accomplish the purpose of suppressing the image frequency, but unfortunately it typically provides only 30 dB image rejection, and this is not enough for most applications. Remember that many receivers require 80 dB or more image rejection, so an image-rejection mixer can help, but it cannot do the job alone.

6.5 MIXER DYNAMIC RANGE

Mixer dynamic range, like amplifier dynamic range, is the range of signal level over which the mixer can operate while maintaining a linear transfer function. Although this may seem to be an anomaly, because the mixer itself is a nonlinear device, a mixer's output can be linear if the output changes dB for dB with its input.

In this linear range, the mixer can operate from its noise floor up to the point at which the signal being converted begins to be compressed. The bounds of this range are determined by the mixer's noise figure at one end and by both the mixer's local oscillator input level and its intrinsic signal handling ability at the other.

As a general rule, passive mixers exhibit a wider dynamic range than active mixers, because the passive mixer noise figure is typically well below 1 dB and passive mixers are readily available that are designed to handle input signals up to about +20 dBm. (Unfortunately, such high-level mixers also typically require local oscillator drive levels of +27 dBm or more.)

Very few active mixers are capable of handling input signals that are larger than about 0 dBm. Also, active mixers have noise figures that are typically in the 5 to 10 dB area. On the good side, active mixers do not usually require as great a local oscillator drive level as passive mixers.

Note: The local oscillator drive level for a passive mixer should be approximately 7 dB greater than the largest signal to be converted.

6.6 MIXER EXAMPLES

Examples of various mixers available today are shown in the following pages. Figure 6.12 lists a group of passive mixers designed to employ a local oscillator whose power level is 200 mW and whose input signal is intended to be as high as +15 dBm. Similar mixers are available that are designed to handle input signals as high as +24 dBm, with a +27 dBm (500 mW) local oscillator level.

A pair of integrated circuit mixers is shown in Figures 6.13 and 6.14. The first of these is a silicon device (IAM-82008) and the second is gallium arsenide (TQ9172). Their noise figures are 19 and 13 dB, respectively (typical). These two devices have a wide range of practical applications but are only representative of the devices that are available, many of which include other functions in addition to frequency conversion. Like most active mixers, the IAM-82008 and the TQ9172 also boast conversion gain, with 15 dB for the IAM and 6 dB for the TQ device.

Many other tradeoffs must be considered for any particular application, so it is not possible to choose on the basis of the few parameters we have mentioned. (Also, the complete data sheets for these mixers have not been included here, where the intent is only to show that such devices currently exist in many forms.)

A more complete form of integrated circuit, from a receiver's overall viewpoint, is shown in Figure 6.15. This integrated circuit includes not only a mixer but also a low-noise amplifier (LNA) intended to be employed as a preamplifier in receivers operating in the range 1.5 to 2.5 GHz. The overall noise figure provided for the combined LNA/mixer combination is 3.5 dB, with input-to-output gain of 23 dB. (This specification is usually termed "conversion gain" for an integrated active mixer or mixer-amplifier.)

ultra-low distortion

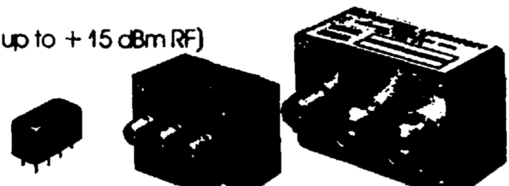
Frequency Mixers

LEVEL 23 (+23 dBm LO, up to +15 dBm RF)

10 KHz to 2.5 GHz

case style selection

outline drawings see Table of Contents



RAY ZMY ZAY

	MODEL NO.	FREQUENCY MHz LO/RF f_L–f_U	IF	CONVERSION LOSS dB Mid-Band m $\bar{x}$	σ	Max.	Total Range Max.	LO-RF ISOLATION, dB L Typ.	Min.	M Typ.	Min.	U Typ.	Min.	LO-IF ISOLATION, dB L Typ.	Min.	M Typ.	Min.	U Typ.	Min.	PRICE $ Qty. (1-9)	DISTRIBUTOR FACTORY	LOCAL
RAY case A01	• RAY-1	5-500	DC-500	6.67	.09	7.5	8.5	55	46	40	30	30	25	56	46	40	30	30	20	40.95	•	•
	• RAY-2	10-1000	DC-1000	6.89	.22	8.5	10	50	35	40	30	35	25	50	35	35	26	25	20	64.45	•	•
	• RAY-3	07-200	DC-200	5.53	.08	7.5	8.0	55	45	40	30	30	25	56	45	40	30	30	20	40.95	•	•
	RAY-6	0.03-50	DC-50	5.36	.15	7.5	8.5	60	50	45	30	35	25	60	45	40	25	30	20	52.45	•	•
	RAY-6U	0.01-100	DC-100	5.09	.15	7.0	6.0	60	50	50	40	40	30	50	45	40	30	35	25	46.95	•	•
	RAY-11	100-2500	DC-500	6.23	.21	9.0	9.0	35	25	32	25	32	25	14	10	20	10	20	10	74.95	•	•
ZMY case M21	• ZMY-1	5-500	DC-500	6.62	.10	7.5	8.5	56	45	40	30	30	25	56	45	40	30	30	20	67.95	•	
	• ZMY-2	10-1000	DC-1000	6.89	.22	8.5	10	50	35	40	30	35	25	50	35	35	26	25	20	84.95	•	
	• ZMY-3	07-200	DC-200	5.53	.08	7.5	8.0	55	45	40	30	30	25	55	45	40	30	30	20	69.95	•	
ZAY case M22	• ZAY-1	5-500	DC-500	6.57	.09	7.5	8.5	56	45	40	30	30	25	56	46	40	30	30	20	64.95	•	
	• ZAY-2	10-1000	DC-1000	6.89	.22	8.5	10	50	35	40	30	35	25	50	35	35	25	25	20	84.95	•	
	• ZAY-3	07-200	DC-200	5.53	.08	7.5	8.0	55	45	40	30	30	25	55	45	40	30	30	20	64.95	•	

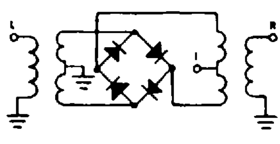
L = low range (f_L to 10 f_L) M = mid range (10 f_L to $f_U/2$) U = upper range ($f_U/2$ to f_U)

m = mid band (2f_L to $f_U/2$)

NOTES:

- • HTRB tested
- 1 For quality control procedures, environmental specifications, and Hi-Rel, ML and TX description see Table of Contents.
- 2 Absolute Maximum Ratings: RF power 350 mW for level 23, 23S and 500 mW for level 27. Peak IF current 40 mA.
- 3 For connector types and case mounting options, see case style outline drawing, see Table of Contents.
- 4 Prices and specifications subject to change without notice
- $\bar{x}$ – Average of conversion loss at center of mid-band frequency (f_L + f_U/4)
- σ – Standard deviation

schematic



pin and coaxial connectors

see case style outline drawing

Series	RAY				ZMY	ZAY	SAY	ZFY	VAY
Models	-1 -3	-2	-6 -6U**	-11	all models	all models	all models	all models	all models
LO	8	8	8	1	1	1	8	1	8
RF	1	1	1	8	3	3	1	2	1
IF	3.4*	3.4*	3.4*	3	2	2	3	3	3
GND♦	2.5.6.7	2.5.6.7	2.5.6.7	2,5.6.7	—	—	2.5.6.7	—	2,5,6.7
CASE GND	2	2.5.6.7	—	2,5.6.7	—	—	2.5.6.7	—	2,5.6.7
NOT USED	—	—	—	—	—	—	4	—	4

*pins must be connected externally

**Pin 2 case gnd Pin 5 6 7 gnd

♦Ground externally. All measurements made with GND pin(s) grounded externally

MIL-M-28837/1*, NSN GUIDE

MCL NO.	NSN
RAY-1	5895-01-106-6188
RAY-2	5895-01-111-7368
RAY-3	5895-01-064-5082
RAY-6	5896-01-317-5882
SAY-11	5895-01-199-3893
VAY-1	5895-01-232-5890
ZMY-1B	5895-01-213-3888
ZMY-2	4935-01-080-7636

*units are not QPL listed

Figure 6.12 Typical passive mixer specifications. (Courtesy of Mini-Circuits.)

HEWLETT PACKARD

Silicon Bipolar MMIC 5 GHz Active Double Balanced Mixer/IF Amp

Technical Data

IAM-82008

Features

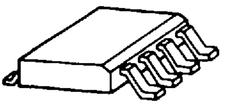
- RF-IF Conversion Gain: 15 dB from 0.05-5 GHz
- IF Conversion Gain from DC to 2 GHz
- IF Output P_{1dB}: +8 dBm Typical
- Single Polarity Bias Supply: V_{CC} = 7 to 13 V
- Load Insensitive Performance
- Conversion Gain Flat over Temperature

Description

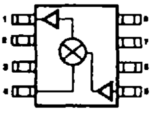
Hewlett-Packard's IAM-82008 is a complete moderate-power double-balanced active mixer housed in a miniature low cost surface mount package. It is designed for narrow or wide bandwidth commercial and industrial applications having RF inputs up to 5 GHz. Operation of RF and LO frequencies below 50 MHz can be achieved using optional external capacitors to ground. The IAM-82008 is particularly well suited for applications that require load-insensitive conversion gain and good spurious signal suppression and moderate dynamic range with low LO power. Typical applications include frequency down-conversion, up-conversion, modulation, demodulation, and phase detection. Markets include fiber-optics, GPS satellite navigation, mobile radio, and communications transmitters and receivers.

The IAM series of Gilbert multiplier-based frequency converters is fabricated using Hewlett Packard's 10 GHz f_T 25 GHz f_{MAX} ISOSAT™-1 silicon bipolar process. This process uses nitride self-alignment, submicrometer lithography, trench isolation, ion implantation, gold metallization, and polyimide inter-metal dielectric and scratch protection to achieve excellent performance, uniformity and reliability.

Plastic SO-8 Package



Functional Block Diagram and Pin Configuration



Pin Description	
1 IF Output	8 RF Ground (optional)
2 V_{ee} AC Ground	7 V_{CC}
3 V_{ee} AC Ground Thermal Contact	6 LO Ground (optional)
4 RF Input	5 LO Input

Figure 6.13 IAM-82008 integrated circuit mixer/IF amplifier. (Courtesy of Hewlett-Packard.)

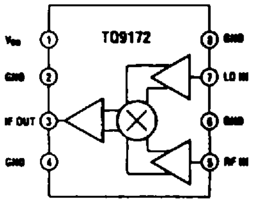
S E M I C O N D U C T O R, I N C

WIRELESS COMMUNICATIONS

TQ9172

Active Frequency Mixer

ICs



The TQ9172 Active Frequency Mixer is part of TriQuint's RFIC Downconverter Building Block family. It is intended for use in frequency downconversion and demodulation of RF and high-IF signals. The TQ9172 provides wide-bandwidth operation from a single +5 V power supply. Its modest current consumption and small, plastic surface-mount package are well suited for low-cost hand-held and battery-powered applications. The TQ9172 features simplified interfaces which provide matched inputs in a 50-ohm environment.

Electrical Specifications

Test Conditions: V_{DD} = +5 V, T_A = 25 °C, LO = 1490 MHz, RF = 1575 MHz, P_{LO} = -5 dBm

Parameter[1]	Min	Typ	Max	Units
LO/RF Frequency Range	150		2500	MHz
IF Frequency Range	30		500	MHz
Conversion Gain	1	6		dB
Noise Figure		13	16	dB
Input 3rd Order Intercept[2]		-6		dBm
Supply Current	11	17	24	mA

Notes: 1 Min/Max values listed are production tested.
2 Frequency separation of the two signals is 500 KHz.

Features

- 150 – 2500 MHz LO/RF frequency operation
- 30 – 500 MHz IF operation
- Single + 5 V supply
- SO-8 plastic package
- All ports matched to 50 Ω
- 6 dB conversion gain
- 17 mA operating current

Applications

- GPS (Global Positioning Systems)
- Cellular Communications
- Spread-Spectrum Receivers
- ISM-Band Systems

Figure 6.14 TQ9172 integrated circuit mixer, intended for down-conversion. (Courtesy of TriQuint Semiconductor.)

RF MICRO·DEVICES

RF2431

HIGH FREQUENCY LNA/MIXER

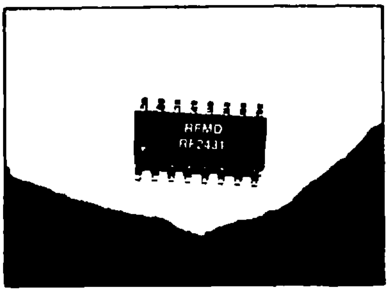
Typical Applications

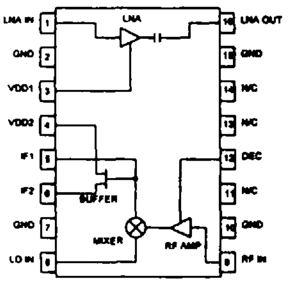
- UHF Digital and Analog Receivers
- Digital Communications Systems
- Spread Spectrum Communications Systems
- Commercial and Consumer Systems
- Portable Battery Powered Equipment
- General Purpose Frequency Conversion

Product Description

The RF2431 is a monolithic integrated UHF receiver front-end. The IC contains all of the required components to implement the RF functions of the receiver except for the passive filtering and LO generation. It contains an LNA (low-noise amplifier), a second RF amplifier, a dual-gate GaAsFET mixer, and an IF output buffer amplifier which will drive a 50 Ω load. Alternatively, the IF output may be matched to a higher impedance at significantly reduced current. The output of the LNA is made available as an output to permit the insertion of a bandpass filter between the LNA and the RF/Mixer section. The LNA output is buffered to permit a wide range of choices for the interstage filter without altering the VSWR or noise figure at the LNA input and to provide high isolation from the LO to the input port. The LNA section may be disabled to conserve power.

Handle with care: ESD sensitive.





Functional Block Diagram

Features

- *Single 3 V to 6 V Power Supply*
- *1.5 GHz to 2.5 GHz Operation*
- *23 dB Small Signal Gain*
- *3.5 dB Noise Figure*
- *13 mA DC Current Consumption*
- *-14 dBm Input IP3*

RF Micro Devices, Inc. • 7341-D West Friendly Ave. • Greensboro, NC 27410 • Tel (910) 855-8085 • Fax (910) [illegible]

RF2431-1 [illegible]

Figure 6.15 Combined low-noise preamplifier and mixer integrated circuit RF2431. (Courtesy of RF Microdevices.)

6.7 HIGH-SIDE VERSUS LOW-SIDE INJECTION

The choice of high-side or low-side injection of the local oscillator into a mixer is based on

1. The effect on the signal being converted,
2. The frequency required of the mixer, and
3. The percentage bandwidth* required of the local oscillator.

When high-side injection is employed (the local oscillator frequency is higher than the input signal), the signal's upper and lower sidebands, if any, are inverted. For FSK signals, the sense of the modulation is also inverted. In many cases, this may not be a problem, but in some cases such an inversion could cause ones to become zeros and vice versa. Where such inversion is a problem, one should employ low-side injection (use a local oscillator frequency that is below the desired signal frequency).

If the mixer being employed has a limited local oscillator frequency range, it may be necessary to use low-side injection.

Percentage bandwidth of the local oscillator (often a frequency synthesizer) is reduced whenever high-side injection is employed, and this is the primary reason for using it.

6.8 MIXER ISOLATION

Port-to-port isolation is a very important parameter for a mixer and may determine the choice made for a particular application. It is obvious that the local oscillator level required and the level of leakage through the mixer to an antenna are important in a military system, because they may well determine whether an antiradiation missile can home in on a given receiver.

Such leakage is also important to a television viewer who does not want a neighbor's local oscillator to interfere with his enjoyment of Monday night football.

Another illustration of the need for port-to-port isolation is the case in which a pulsed signal might load the local oscillator by changing the loading when the signal is present, as opposed to when it is not present. This, in turn, can cause the local oscillator frequency to "pull,"† and this in turn degrades the signal in the

*Percentage bandwidth is the ratio of the bandwidth of a device to the center frequency at which it is operating, expressed in percent.

†Oscillator pulling (or pull) is the effect on the frequency of an oscillator that is caused by changes in the loading, the power supply voltage, or other peripheral factors that may cause the frequency of the oscillator to vary.

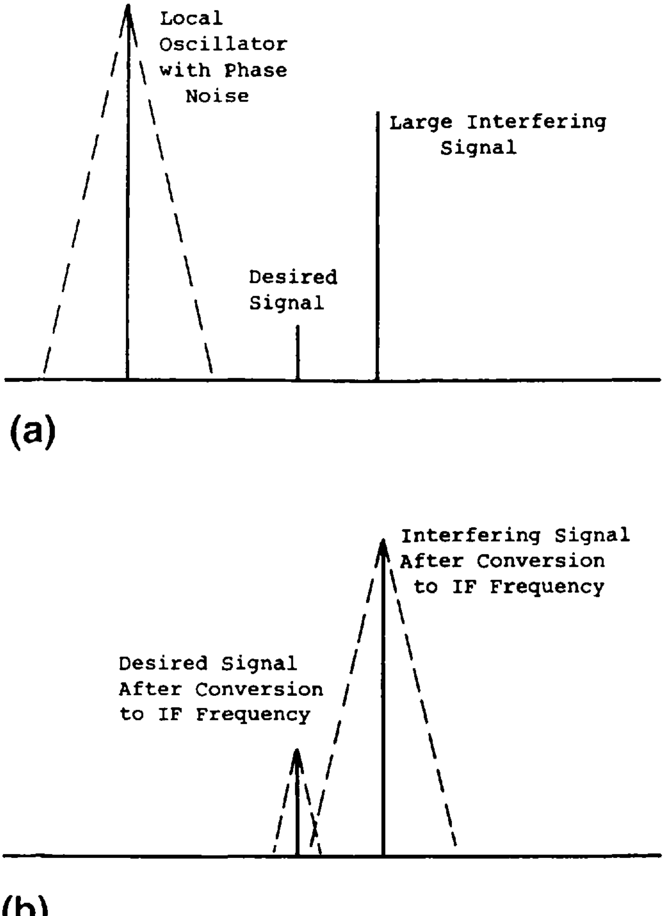


Figure 6.16 Illustration of "reciprocal mixing" process. Note that interfering signal, with added phase noise due to local oscillator, overlaps onto desired signal. (a) Spectrum of input signals and local oscillator having phase noise (spectrum centered at interference frequency). (b) Desired signal and interfering signal after conversion to IF frequency (spectrum centered at IF).

conversion process. One solution to this problem is to provide better RF and/or IF to local oscillator port-to-port isolation.

In practice, mixers can provide about 30 dB port-to-port isolation over a wide range of environmental and aging conditions.

6.9 RECIPROCAL MIXING

"Reciprocal" mixing is a process by which a small desired signal can be degraded due to a combination of phase noise on the local oscillator in a receiver and a strong signal at a nearby frequency. (Reciprocals have nothing to do with the process.)

Whenever a local oscillator is employed to convert a received signal to an intermediate frequency, whether up or down, the spectrum of any incidental modulation on the local oscillator is impressed on the signal whose frequency is being converted. Thus, the phase noise spectrum of a local oscillator is impressed on both a desired signal and any unwanted signal that may be converted to the intermediate frequency region along with the desired signal. (Remember that the bandwidth of a receiver before the IF filter is usually wide enough that many undesired signals must be processed and converted to the IF range that will be rejected by the IF filter.)

If the phase noise* added to a strong undesired signal overlaps the desired signal, then the added noise within the signal bandwidth degrades the signal-to-noise ratio of the receiver. This is the reciprocal mixing process. Figure 6.16 illustrates this process by which a receiver may be desensitized.

Given that a strong CW signal is offset from the desired signal by an amount K Hz and has amplitude A, when that signal is modulated by the phase noise of the local oscillator, which is at a level $-P$ dBc at offset K, the level of the phase noise on the translated undesired signal is

$$A - P + 10 \log W$$

where W is the receiver's noise bandwidth, and the receiver's sensitivity will be degraded by the amount of the added noise within the receiver bandwidth.

QUESTIONS

1. How does a mixer affect a receiver's sensitivity? What if the first mixer has a 10-dB noise figure? What if the second mixer has a 10-dB noise figure?
2. What are the disadvantages of using a passive mixer? An active mixer?

*See Section 7.7.

3. How does the local oscillator level applied to a mixer affect a receiver's dynamic range?
4. What is meant by a "high-level" mixer and what makes it high level?
5. What are at least two advantages of employing an active mixer?
6. What are the disadvantages of using low-side injection?
7. What is the primary reason for using high-side injection?
8. How does a mixer affect local oscillator radiation from a receiver?
9. What is the meaning of a mixer's conversion gain or loss?
10. What is the effect of mismatch on a passive, double-balanced mixer?

REFERENCES*

1. Reynolds, L. D., LNA/mixer ICs ease wireless receiver design, *Microwaves & RF*, Nov. 1993.
2. *Designer's Handbook*, RF Micro Devices, 1995, pp. 65–106.
3. Solutions for Wireless Communications, Pacific Monolithics.
4. The application of GaAs MMICS to DBS receivers, *Microwave Engineering Europe*, Dec./Jan. 1992.
5. Your GaAs Source, Anadigics.
6. M/A-COM RF and Microwave Semiconductors; Discrete, Monolithic and Multi-Function, pp. 13-1 to 13-12.
7. Communications Components, GaAs & Silicon Products, Hewlett-Packard, pp. 8-33 to 8-44.
8. Reynolds, L. D., *RF/IF Designer's Handbook*, Mini-Circuits, pp. 1-1 to 1-340.
9. Grebenkemper, C. J., Local Oscillator Phase Noise and Its Effect on Receiver Performance, Watkins-Johnson Company Tech-Notes, Nov./Dec. 1981.
10. Watson, R. E., Receiver Dynamic Range: Part 1, Watkins-Johnson Company Tech-Notes, Jan./Feb. 1987.
11. Henderson, B. C., and Zazkowski, J. M., Integrated Mixer-Preamplifiers, Watkins-Johnson Tech-Notes, May/June 1988.
12. Henderson, B. C., Predicting Intermodulation Suppression in Double-Balanced Mixers, Watkins-Johnson Tech-Notes, July/Aug. 1983.
13. Henderson, B. C., and Cook, J. A., Image-Reject and Single-Sideband Mixers, Watkins-Johnson Tech-Notes, May/June 1985.
14. Henderson, B. C., Mixers in Microwave Systems, Parts 1 and 2, Watkins-Johnson Tech-Notes, Jan.-Apr. 1990.
15. Termination Insensitive Mixers (TIM), App. Note M532, M/A-COM, 1994.
16. *RF/IF Designer's Handbook*, Mini-Circuits, 1992.

*Also see references for Chapter 5.

This page intentionally left blank

7
Local Oscillator Design

Local oscillators in receivers provide a signal that is used to convert the input signal to a frequency at which it may demodulated with as little distortion as possible and with the effects of noise and interference reduced as much as possible. The local oscillator (or oscillators in a multiple conversion receiver) must be accurate in frequency, be adequate in drive level, and contribute as little unwanted noise or other spurious content as possible in the conversion process. A usable local oscillator can be as simple as a manually tuned single oscillator or as complex as a high-speed multiple loop frequency synthesizer; the choice is most often governed by switching speed from one frequency to another and/or by frequency accuracy.

The choice may also be made on the basis of operating frequency and whether high-side or low-side injection is used. In any case, there are as many ways to implement a local oscillator as there are answers. Here we will concentrate on the use of frequency synthesizers, as their cost is low and their use today is nearly universal.

Local oscillator noise affects the demodulation process in the same way as noise from the preamplifier or noise from a jammer; it causes the demodulator to make mistakes in bit decisions or causes the demodulator's output signal-to-noise ratio to be degraded. In data systems, such as PSK systems, where the phase margin for decision making is one half of the phase shift between phase positions, added phase noise can significantly increase the bit error rate, and this is the reason for placing such emphasis on local oscillator noise reduction.

7.1 TUNED OSCILLATORS

For many years, the most common way to tune a receiver was to change manually a tuning capacitor* (or capacitors) whose capacity varied as a function of the rotation of a shaft with a tuning knob attached. This sufficed quite well but suffered from the fact that either a person or a tuning motor had to turn the capacitor from one value to another to change frequencies. This method even had the advantage that more than one capacitor could be tuned at once, so that both preamplifier and local oscillator could be tuned with the same tuning knob. But, horror of horrors, turning a tuning knob does not adapt readily to digital (incremental) control, and in the early 1940s military radios began to use synthesized local oscillators.

Today, only the very simplest of consumer receivers employ manually tuned local oscillators with mechanically variable tuning capacitors. But they work, and they allow listeners to fine-tune the receiver to meet whatever listening need they may have or to overcome whatever the vagaries of the broadcast bands may produce. The tuning capacitors used in these consumer receivers look much different from those used in old receivers, but they work the same way. Figure 7.1 shows tuning capacitors as employed over the past 60 to 70 years.

Of course, tuned oscillators used in frequency synthesizers today are no longer tuned mechanically (except for "trimmer" capacitors that are still used). Instead, varactor diodes are employed, whose junction capacity varies as a function of the voltage applied (see Section 7.4 on voltage-controlled oscillators). Fixed-tuned oscillators are often employed in receivers as local oscillators for

1. Single-frequency receivers
2. Multiple conversion receivers, after the first conversion

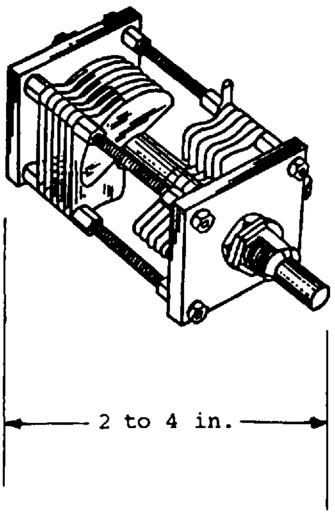
and the resonator used in the oscillator determines the oscillator's performance.

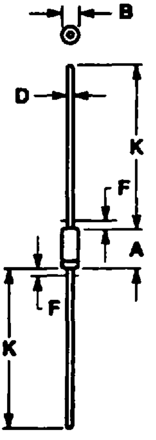
The following resonators are common today:

7.1.1 LC

A resonant circuit that consists of an inductor and a capacitor and that resonates at a frequency $1/(2\pi\sqrt{LC})$. (Varactor-tuned oscillators are included.) The inductor and capacitor used determine the frequency of the oscillator, but every component of the oscillator circuit contributes to its stability (or lack thereof). The circuit Q in lumped LC resonators is usually limited to about 200.

*Some receivers used variable inductors, in which the inductance was changed by moving a core within a coil. The technique was called "slug tuning," which brings up all kinds of interesting mental pictures.





DIM	MILLIMETERS		INCHES	
	MIN	MAX	MIN	MAX
A	5.84	7.62	0.230	0.300
B	2.16	2.72	0.085	0.107
D	0.46	0.56	0.018	0.022
F	—	1.27	—	0.050
K	25.40	38.10	1.000	1.500

Figure 7.1 Comparison of tuning capacitors. Dual, air variable (left). Solid state, single (right).

7.1.2 Helical Resonator

Helical resonators are actually a transitional type of resonator that falls between lumped constant *LC* and cavity resonators. They consist of an inductor within a cavity, where the inductor serves to excite the cavity. Such resonators may have a *Q* of 1000 to 2000. Their frequency ranges from less than 100 MHz to over 1 GHz. Figure 7.2 shows typical helical resonators in a filter configuration.

7.1.3 SAW

SAW or surface acoustic wave oscillators are typically employed in the same range as helical resonator oscillators but are more stable because of their higher *Q*. (Helical resonators are better for wide-range voltage-controlled applications; SAW oscillators are difficult to tune over more than a few percent bandwidth.) A typical SAW resonator-controlled oscillator is shown in Figure 7.3. The SAW resonator *Q* is about 10,000 or more.

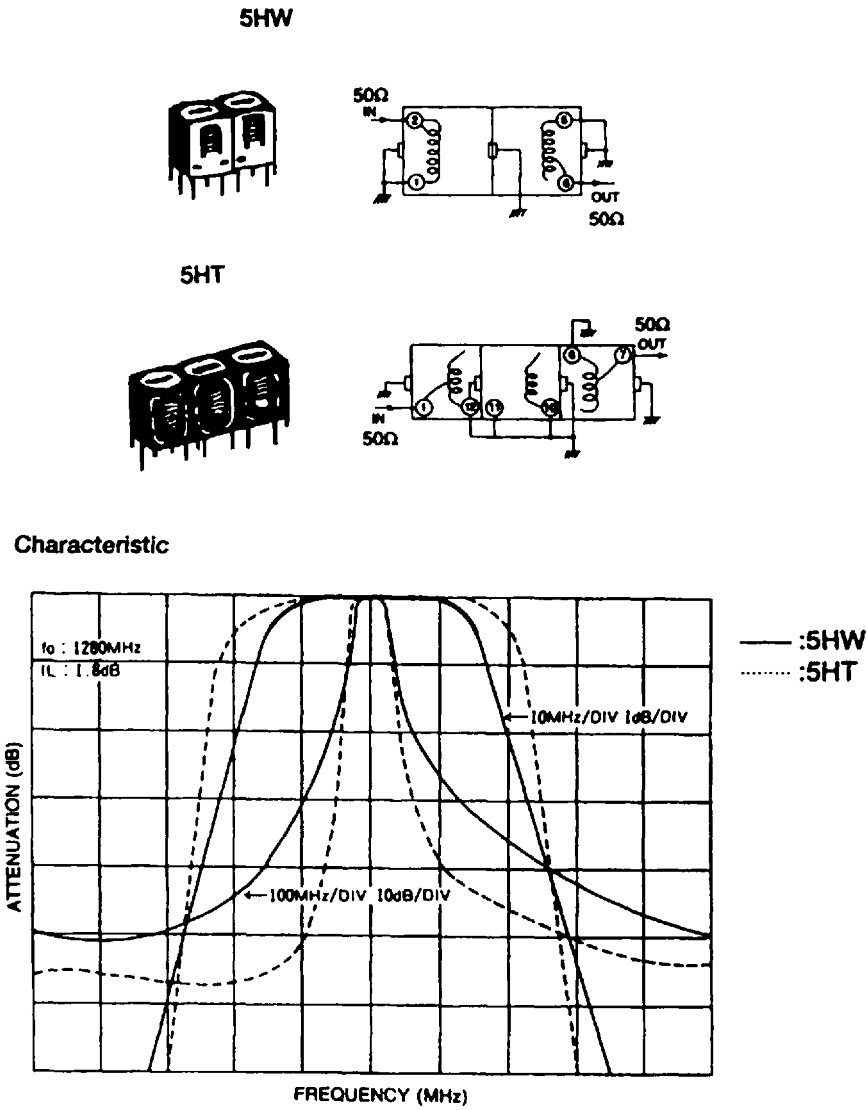


Figure 7.2 Helical resonators (in filter configuration). Oscillator typically uses only one resonator. (Courtesy of Toko America Inc.)

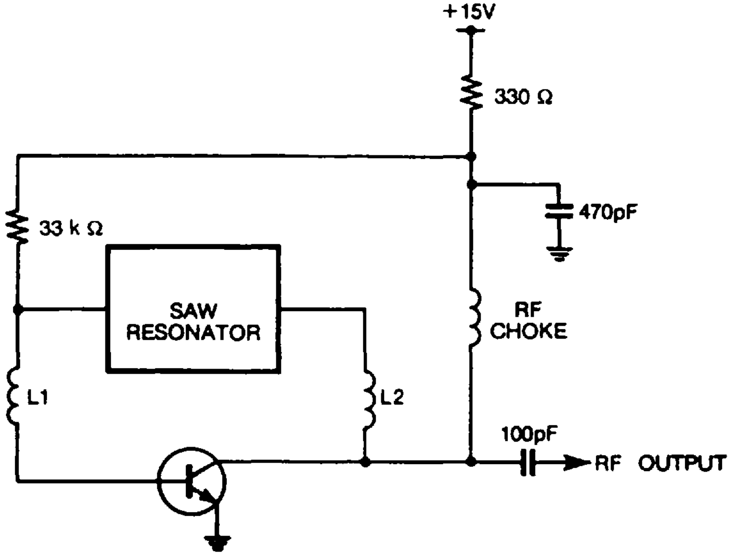


Figure 7.3 Typical SAW oscillator. L1 and L2 values are chosen as a function of the resonator parameters.

7.1.4 Crystal

Quartz crystal–controlled oscillators have been the mainstay of accurate receiver designs for many years. The crystals at the heart of the oscillators are available from low frequencies (in the kilohertz range) to around 200 MHz. At the higher frequencies, crystal-stabilized oscillators actually work at a harmonic of the fundamental frequency, and harmonics as high as the 9th or 11th are employed. Figure 7.4 shows a high-frequency crystal oscillator that is useful over a wide range of operating frequencies.

Crystal oscillators can be expected to be accurate to a few parts in 10^5 over a wide temperature range, but only if the crystals are specified properly. When specifying a crystal to operate at a particular frequency, one technique is to provide the oscillator circuit to the crystal manufacturer, who can make the crystals work in the circuit in which they will be used. Short of that, however, it is possible to buy crystals that are adequate for prototyping by specifying the following:

- Crystal type and holder: Crystal type has to do with frequency and temperature characteristics. For most receiver applications, an AT cut or ST cut crystal is used. The crystal holder is chosen from one of several

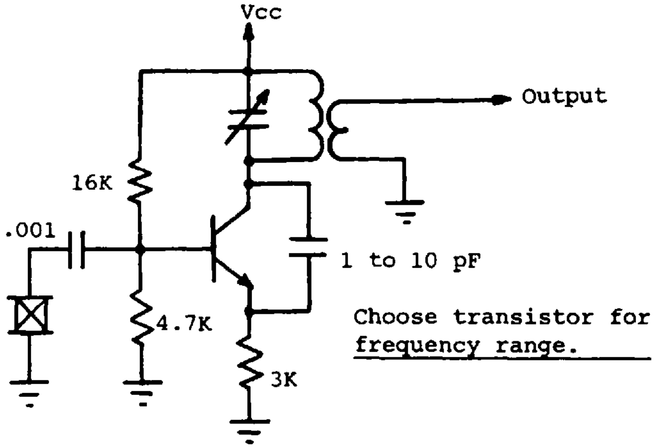


Figure 7.4 Crystal oscillator using series mode crystal, for application up to 200 MHz.

standard sizes, depending on the crystal frequency. Figures 7.5 and 7.6 show temperature characteristics of typical quartz crystals and drawings of standard crystal packages.

- Crystal calibration circuit: The type of circuit configuration in which crystals provide the proper frequency must be specified. This has to do with whether the crystal is employed in a series or parallel circuit mode. In the series mode, the crystal frequency must be specified along with the series resistance at resonance. In the parallel mode, crystal frequency and parallel capacitance must be specified.

A typical circuit for use of a series mode crystal is shown in Figure 7.4. This circuit oscillates at the crystal frequency because the impedance of the base circuit is lowest in the series mode at the resonant frequency. Away from the resonant frequency, the crystal impedance increases and the oscillator ceases to operate.

Figure 7.7 is a typical oscillator circuit used in computer and frequency synthesizer applications. The active devices are CMOS inverters as employed in integrated circuits. The crystal is connected between the input and output of the inverter, with a high resistance in parallel. A resistor may also be placed in series with the crystal to limit crystal current. Two capacitors are connected to ground from the inverter's input and output, and their capacitance is such that each capacitor's value is twice the parallel capacitance specified for the crystal.

A crystal's Q is typically in the range of 100,000 or higher, which is the reason for a crystal-stabilized oscillator's accuracy and low noise. The high Q is also the reason that crystal-stabilized oscillators generally do not make good voltage-controlled oscillators (VCOs), as their range of control is not usually more than about 50 to 100 Hz per million Hertz at the operating frequency.

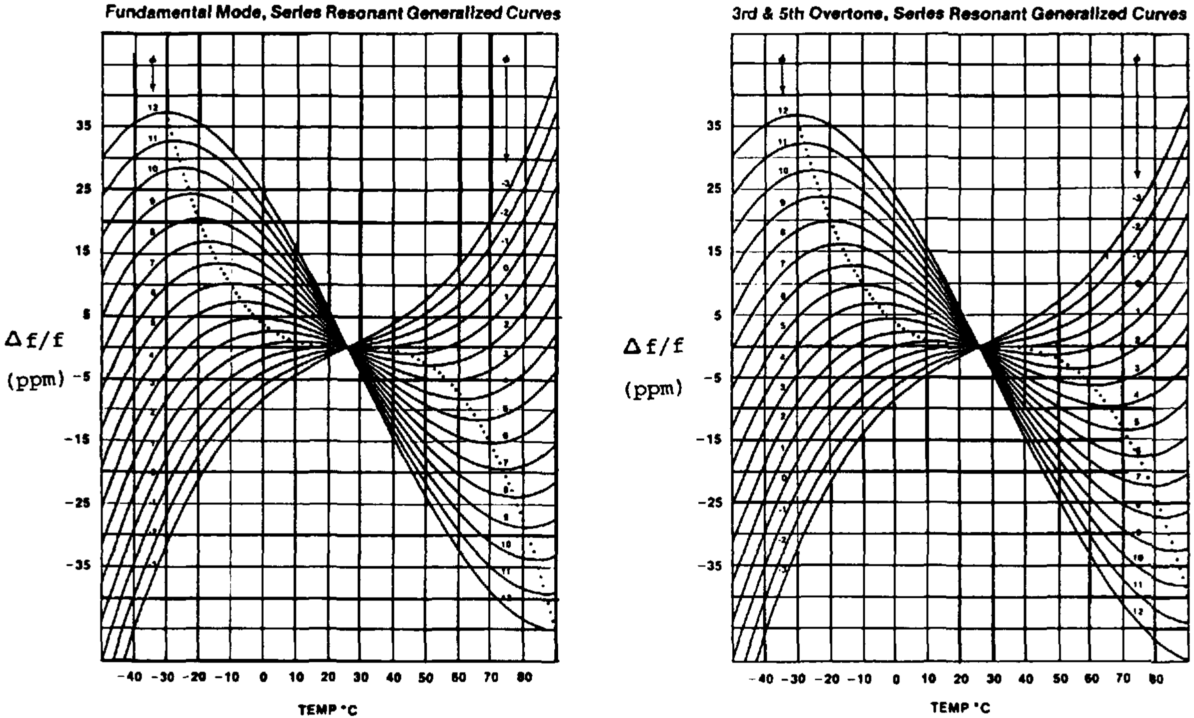
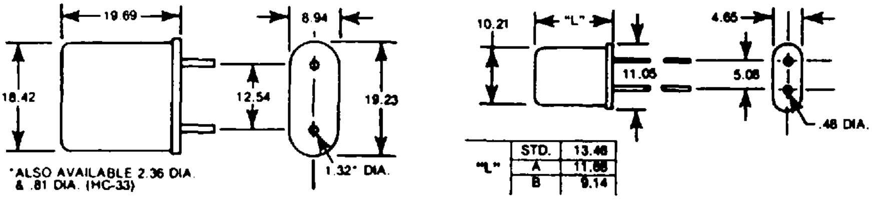


Figure 7.5 Temperature characteristics for fundamental and overtone crystals as a function of the angle of the cut (quartz). (Courtesy of Piezo Crystal Company.)



HC-6 Package   HC-18 Package

Figure 7.6 Common, standard crystal packages.

Among the most important characteristics of a crystal are the type of package that encloses it and the method used to seal the package. The sealing method is, in fact, the most important determiner for long-term crystal frequency stability. The three packages and sealing methods are shown in Table 7.1. The stability listed is for a well-aged crystal after at least 24 hours of operation.*

Table 7.2 shows typical accuracies of 30 low-cost commercial crystals. In a typical application, oscillators using such crystals would be trimmed to the nominal frequency if better accuracy is required. The best approach is to provide the manufacturer with the exact circuit that will be used with the crystal, but this is not practical for crystals that are made in large quantities for unspecified use.

7.1.5 Cavity Resonators

At sufficiently high frequencies, at which the physical size of the cavity required is not intimidating, a cavity whose dimensions determine its frequency of oscillation may be employed. The Q exhibited by such resonators is in the range of 1000 or more, but is highly affected by the way the excitation and output coupling is accomplished. The most common means of coupling to a cavity resonator is a small coupling loop within the cavity, which is as simple as a short piece of wire that is connected to a coaxial connector's center pin on one end, and to the cavity wall (the outer conductor for the coax, and usually the ground return for everything including the cavity) at the other end. For a cylindrical cavity, the wavelength of resonance is approximately 2.61 times the radius of the cylinder.

*Crystals do not age at the same rate when vibrating as they do when not vibrating, and when vibrating they age in such a way that they asymptotically approach a final frequency. For this reason, new crystals for critical applications should be aged for 30 days or more under the conditions in which they will be used. Also, even after being well aged, a crystal should be turned on in its oscillator for up to 24 hours before it is expected to meet its accuracy specifications. For best performance, crystal-stabilized oscillators should never be turned off.

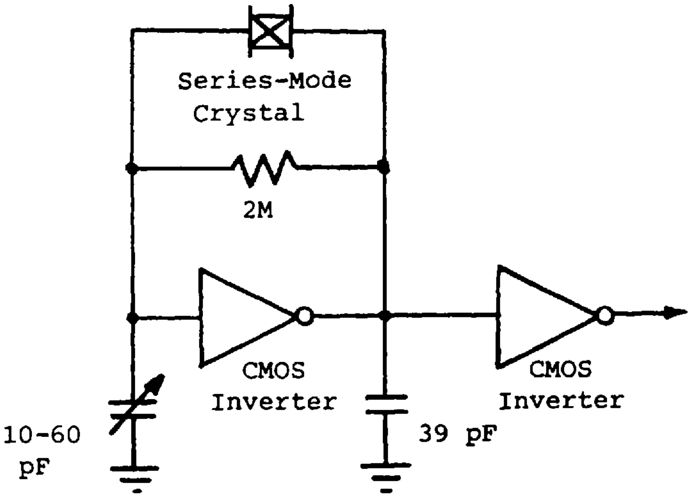


Figure 7.7 Typical crystal-stabilized inverter oscillator as employed in computers and frequency synthesizers. The crystal should be specified as parallel, with 20 to 40 pF parallel capacity (typical). CMOS inverters or CMOS NAND gates work as well as oscillators.

7.1.6 Dielectric Resonators

Dielectric resonator–stabilized oscillators can be designed to operate at frequencies up to about 20 GHz. The dielectric resonator Q, however, is not as high as that of crystals or SAW resonators. Figure 7.8 shows a typical dielectric resonator oscillator.

7.2 SYNTHESIZERS

Today, frequency synthesizers, as opposed to simple tuned oscillators, are employed in even the lowest cost receivers. Automobile receivers, television receivers, and many other receivers that must operate at multiple frequencies benefit from synthesizers that are often as accurate as the broadcast transmitters from which they are designed to receive signals.* Broadcast band AM and FM transmitters, for example, are required to maintain frequency to within $\pm 1 \times 10^{-5}$ times the assigned frequency, and this is within the realm of commercial crystal stability.

Synthesizers are divided into two general groups, direct and indirect. Indi-

*In many cases, such receivers lack a fine tuning control, which could improve performance. This is the greatest drawback in the use of frequency synthesizers.

Table 7.1 Crystal Packaging Characteristics

Crystal package material	Sealing method	Characteristics
Metal	Solder	Lowest cost Poorest stability
	Cold weld	Medium to high cost Stability to about 1×10^{-9}/day
Glass	Fusion	Highest cost Largest size Best stability Stability to about 5×10^{-11}/day

rect synthesizers are defined as those that employ one or more phase lock loops to generate the desired signal, and direct synthesizers do not employ a phase lock loop.

7.2.1 Indirect Synthesizers

Figure 7.9 illustrates both a simple phase lock loop and a single-loop phase lock frequency synthesizer. The difference between the two is the divide-by-N counter added to the frequency synthesizer.

In a simple phase lock loop, the voltage-controlled oscillator is caused to synchronize its frequency to an input frequency. This operation is a function of the action of the phase detector, whose output is filtered and used to correct the frequency of the VCO. (The key to both simple phase lock loops and phase lock frequency synthesizers is to remember that the feedback loop in both serves to force the frequency of both inputs to the phase detector to be identical.)

In a phase lock synthesizer (indirect) the $\div N$ counter in the feedback path causes the VCO to operate at a frequency that is N times the reference frequency, so that the two inputs to the phase detector are the same. That is,

Table 7.2 Typical Commercial Crystal Accuracy

	Crystal accuracy (sample of 10 crystals)		
Nominal frequency	Average frequency	Average error	Worst-case error
1.000000 MHz	1.00004084	4.084×10^{-5}	4.09×10^{-5}
3.000000	2.99999067	3.11×10^{-6}	3.2×10^{-6}
6.553690	6.5536955	8.4×10^{-7}	1.07×10^{-6}

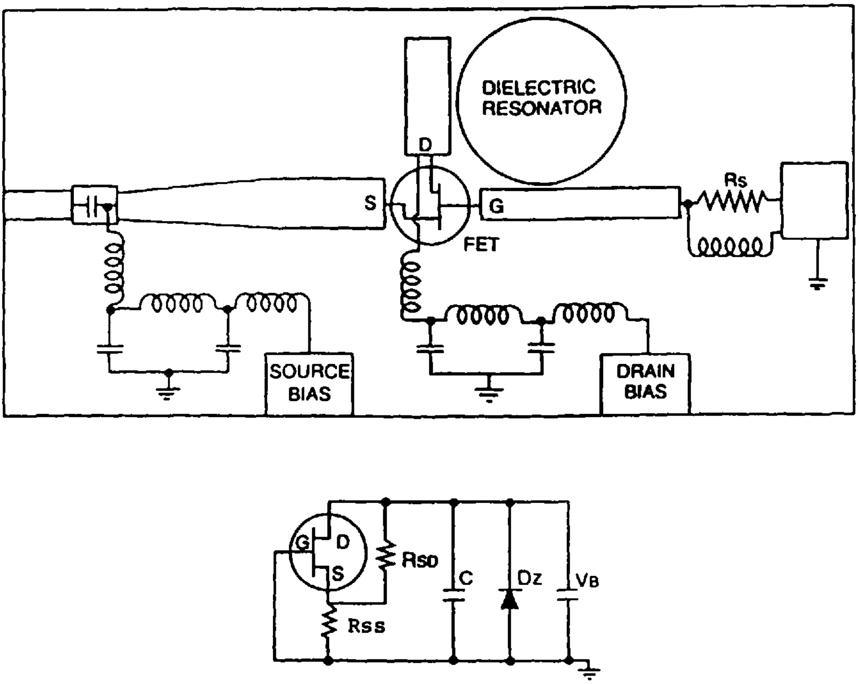


Figure 7.8 Dielectric resonator oscillator. Circuit layout and equivalent circuit. (Courtesy of Murata Erie North America.)

$$F_{\text{reference}} = \frac{F_{\text{vco}}}{N}$$

and therefore

$$F_{\text{vco}} = N \times F_{\text{reference}}$$

This means that by controlling the counter's division ratio, the frequency synthesizer's output frequency can be selected.

Example. Let us suppose that a frequency synthesizer for a broadcast band FM receiver is required. Such a receiver must have 200 kHz channel spacing, but the channels begin at 88.1 MHz and end at 107.9 MHz. (For this reason, a 100 kHz frequency reference would be employed, with the counter programmed to employ only the even increments, starting at 988 and ending at 1886. This assumes the receiver is to employ high-side injection.) Figure 7.10 is a block diagram of one approach to this frequency synthesizer.

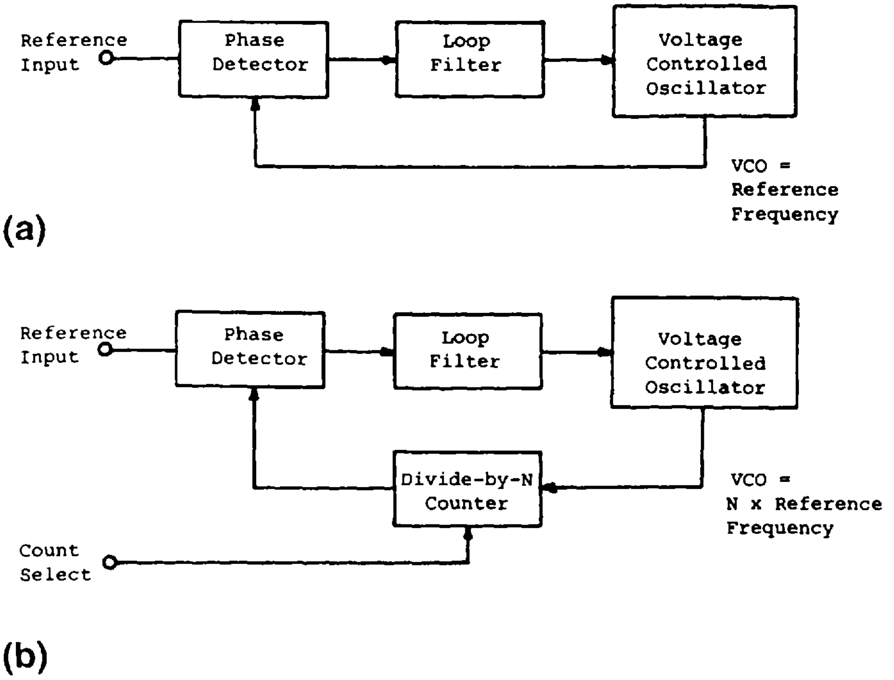


Figure 7.9 Comparison of simple phase lock loop with phase lock synthesizer. (a) Simple phase lock loop. (b) Phase lock frequency synthesizer.

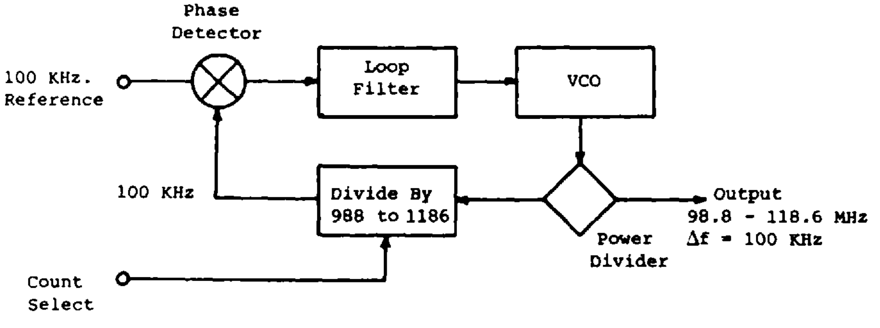


Figure 7.10 Block diagram of phase lock synthesizer for FM broadcast band.

It should be noted that there are some very important differences in the phase lock loops employed in receivers for frequency synthesis and demodulation. The differences are (at the least) the frequency counter within the loop and the phase detector. In frequency synthesizers, the phase detector is usually a digital phase detector, whereas demodulators employ analog phase detectors in most applications (but not always).

Table 7.3 describes the characteristics of analog and digital phase detectors. Phase detectors are described in more detail in section 7.5.

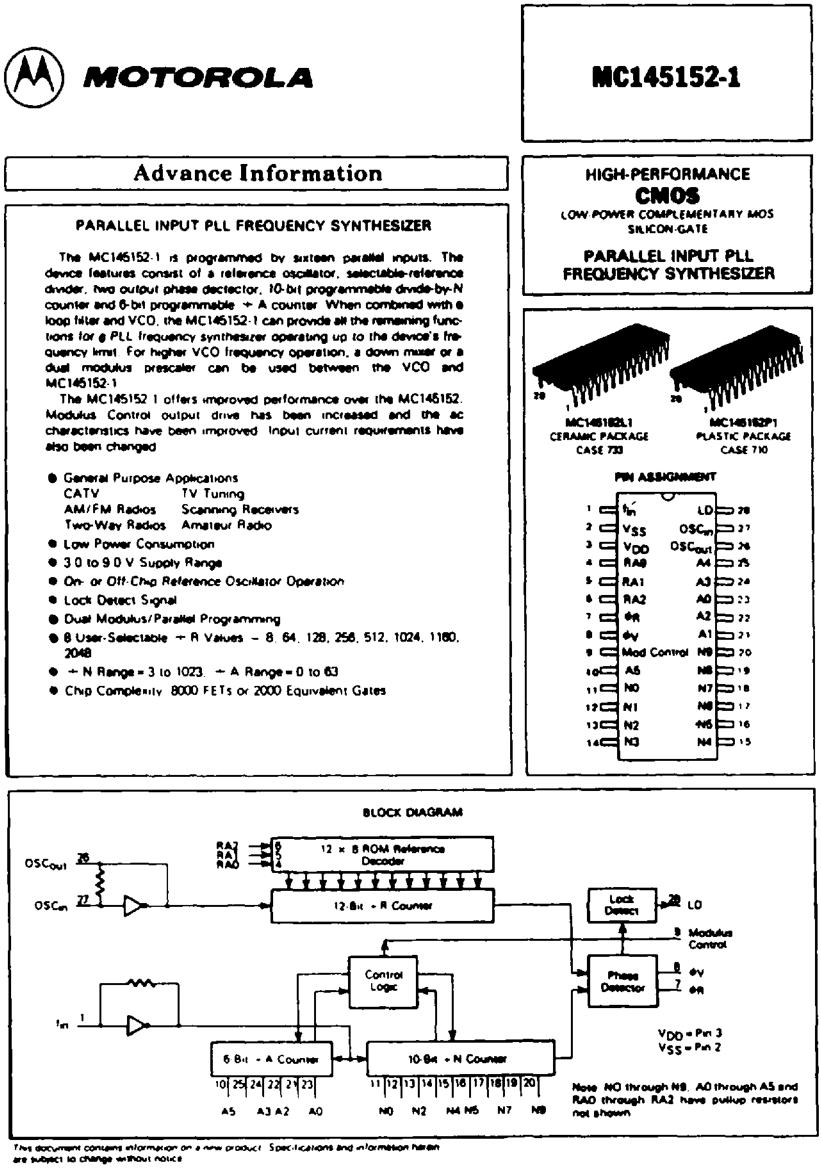
Conveniently, there exist today a number of integrated circuit frequency synthesizers that often include phase detectors, and these phase detectors are usually digital phase detectors. In most cases, it is practical to construct a frequency synthesizer with a synthesizer chip, a separate VCO, and a few discrete components used for bypassing and frequency determination.

7.2.2 Typical Integrated Circuit Synthesizers

Two frequency synthesizer integrated circuits are shown in the following pages. Both are indirect synthesizers, but they differ in their method of frequency selection control. The first, a Motorola MC145152, employs a control counter that is controlled in parallel and is more readily employed in applications such as frequency hopping. The second, a Fujitsu MB1502, requires bus control from a microprocessor for frequency selection. Figures 7.11 through 7.13 show these synthesizers and basic applications.

Table 7.3 Characteristics of Phase Detectors

Phase detector	Characteristics
Analog	Lowest sensitivity (in volts per radian). Good linearity over operating range. Locks at 90 degrees.
Digital	Higher sensitivity than analog detectors. All require logic-level inputs.
Exclusive-OR	Requires signal symmetry. Locks at 90 degrees (like analog).
R-S flip-flop	Works with pulses or square waves. Locks at 90 degrees (like analog).
Edge triggered	Works with pulses or square waves. Locks at zero or 180 degrees. Has discriminator/phase detector action.



MOTOROLA

MC145152-1

Advance Information

HIGH-PERFORMANCE
CMOS
LOW-POWER COMPLEMENTARY MOS
SILICON-GATE

PARALLEL INPUT PLL
FREQUENCY SYNTHESIZER

PARALLEL INPUT PLL FREQUENCY SYNTHESIZER

The MC145152-1 is programmed by sixteen parallel inputs. The device features consist of a reference oscillator, selectable-reference divider, two output phase dectector, 10-bit programmable divide-by-N counter and 6-bit programmable ÷ A counter. When combined with a loop filter and VCO, the MC145152-1 can provide all the remaining functions for a PLL frequency synthesizer operating up to the device's frequency limit. For higher VCO frequency operation, a down mixer or a dual modulus prescaler can be used between the VCO and MC145152-1.

The MC145152-1 offers improved performance over the MC145152. Modulus Control output drive has been increased and the ac characteristics have been improved. Input current requirements have also been changed.

- General Purpose Applications
  - CATV — TV Tuning
  - AM/FM Radios — Scanning Receivers
  - Two-Way Radios — Amateur Radio
- Low Power Consumption
- 3.0 to 9.0 V Supply Range
- On- or Off-Chip Reference Oscillator Operation
- Lock Detect Signal
- Dual Modulus/Parallel Programming
- 8 User-Selectable ÷ R Values – 8, 64, 128, 256, 512, 1024, 1160, 2048
- ÷ N Range = 3 to 1023, ÷ A Range = 0 to 63
- Chip Complexity: 8000 FETs or 2000 Equivalent Gates

MC145152L1
CERAMIC PACKAGE
CASE 733

MC145152P1
PLASTIC PACKAGE
CASE 710

PIN ASSIGNMENT

1	f_{in}	LD	28
2	V_{SS}	OSC_{in}	27
3	V_{DD}	OSC_{out}	26
4	RA0	A4	25
5	RA1	A3	24
6	RA2	A0	23
7	ϕR	A2	22
8	ϕV	A1	21
9	Mod Control	N9	20
10	A5	N8	19
11	N0	N7	18
12	N1	N6	17
13	N2	N5	16
14	N3	N4	15

BLOCK DIAGRAM

Note: N0 through N9, A0 through A5 and RA0 through RA2 have pullup resistors not shown.

This document contains information on a new product. Specifications and information herein are subject to change without notice.

Figure 7.11 Motorola MC145152-1 integrated circuit frequency synthesizer, with parallel frequency control input. (Courtesy of Motorola.)

FUJITSU

June 1991

DATA SHEET

MB1502

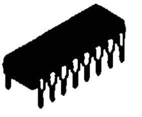
Serial Input PLL Frequency Synthesizer

The Fujitsu MB1502 fabricated in Bi-CMOS technology, is a single chip serial input PLL synthesizer with pulse-swallow function.

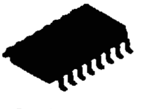
The MB1502 contains the following: analog switch to speed up lock up time, control signal generator, 16-bit shift register, 15-bit latch, programmable reference divider (binary 14-bit programmable reference counter), 1-bit switch counter, phase comparator with phase conversion function, charge pump, crystal oscillator, 19-bit shift register, 18-bit latch, programmable divider (binary 7-bit swallow counter and binary 11-bit programmable counter) and a 1.1 GHz two modulus prescaler that can select either a 64/65 or 128/129 divide ratio.

It operates supply voltage of 5 V typ. and achieves very low power supply current of 8 mA typ. realized through the use of Fujitsu Advanced Process Technology.

- High operating frequency: $f_{IN\,MAX} = 1.1$ GHz ($V_{IN\,MAX} = -10$ dBm)
- Pulse swallow function: 64/65 or 128/129
- Low supply current: $I_{CC} = 8$ mA typ.
- Serial input 18-bit programmable divider consisting of:
  - Binary 7-bit swallow counter: 0 to 127
  - Binary 11-bit programmable counter: 16 to 2047
- Serial input 15-bit programmable reference divider consisting of:
  - Binary 14-bit programmable reference counter: 8 to 16383
  - 1-bit switch counter (SW) sets divide ratio of prescaler
- On-chip analog switch achieves fast lock up time
- Two types of phase detector output:
  - On-chip charge pump (Bipolar type)
  - Output for external charge pump
- Wide operating temperature: −40 °C to +85 °C
- 16-pin Plastic DIP Package (Suffix: −P)
  16-pin Plastic Flat Package (Suffix: −PF)

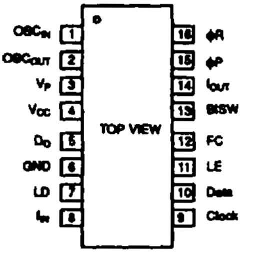


Plastic Package
DIP-16P-M04



Plastic Package
FPT-16P-M06

Pin Assignment



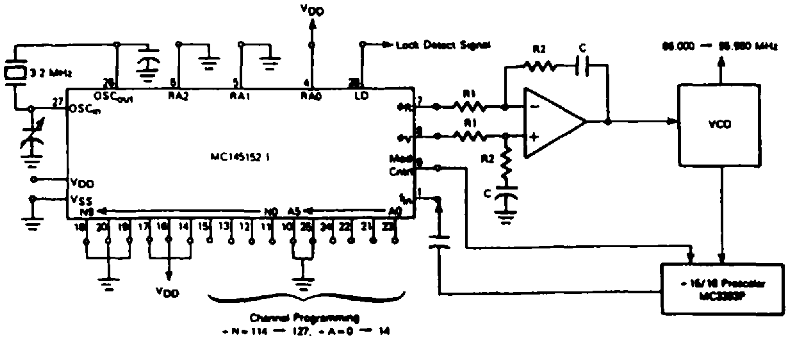
[illegible]

ABSOLUTE MAXIMUM RATINGS

Parameter	Symbol	Ratings	Unit
Power Supply Voltage	V_{CC}	−0.5 to +7.0	V
	V_P	V_{CC} to 10.0	V
Output Voltage	V_{OUT}	−0.5 to V_{CC} +0.5	V
Open-drain Voltage	V_{OOP}	−0.5 to 0.8	V
Output Current	I_{OUT}	±10	mA
Storage Temperature	T_{STG}	−55 to +125	°C

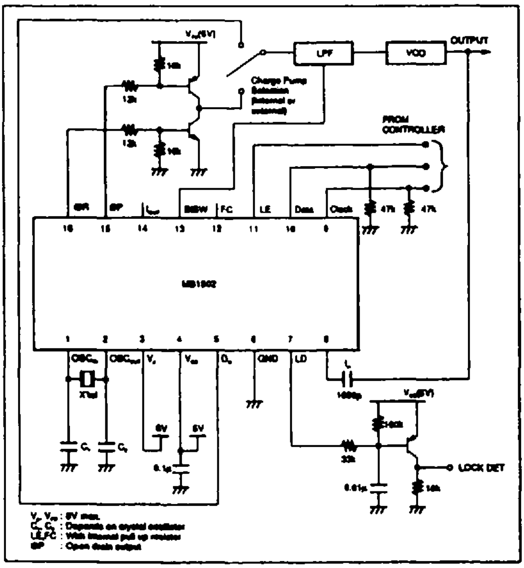
Note: Permanent device damage may occur if absolute maximum ratings are exceeded. Functional operation should be restricted to the conditions as detailed in the operation sections of this data sheet. Exposure to absolute maximum rating conditions for extended periods may affect device reliability.

Figure 7.12 Fujitsu MB1502 integrated circuit frequency synthesizer, with serial frequency control input. (Courtesy of Fujitsu.)



(f_R = 50 KHz, ÷ R = 64, N = 1720-1919, IF = 22 MHz)

(a)



(b)

Figure 7.13 Application diagrams for two integrated circuit synthesizers. (a) Aircraft navigation receiver synthesizer employing the Motorola MC145152-1. (b) General application diagram for Fujitsu MB1502. (Courtesy of Motorola and Fujitsu.)

7.3 LOCK-IN TIME REDUCTION IN INDIRECT SYNTHESIZERS

Several techniques can be used to speed up a phase lock synthesizer. For any such synthesizer, it can be shown that the lock-in time is approximately

$$t_{\text{lock}} \approx \frac{35\Delta f^2}{(f_{3\text{dB}})^3}$$

This means that if Δf, the frequency shift, is reduced by half, the lock-in time is reduced by four, and if the loop bandwidth is doubled, the lock-in time is reduced by eight.

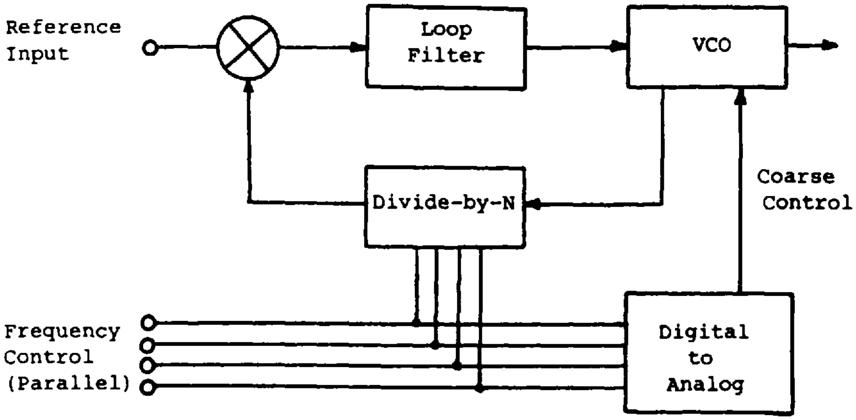
Two methods for reducing Δf are illustrated in Figure 7.14. The first (Figure 7.14a) employs a digital-to-analog converter to translate the control signals used to select the operating frequency to a DC level that sets the VCO to approximately the correct frequency, reducing Δf. The second method also sets the VCO to approximately the correct frequency using a voltage derived from a discriminator tuned to the reference frequency.

The edge-triggered phase detectors discussed in Section 7.5 may also be employed to reduce synthesizer lock-in time. They provide discriminator action that is two valued: when the frequency reference is lower in frequency than the counted-down VCO frequency, the edge-triggered phase detector's output is a logic zero, but when the frequency reference is higher than the counted-down VCO frequency, the edge-triggered phase detector puts out a logic one. This causes the VCO to be driven in the direction needed to synchronize to the frequency reference. Then, when the counted-down VCO and the frequency reference are within 360 degrees of synchronism, the edge triggered phase detector becomes a conventional phase detector and causes the synthesizer loop to phase lock and maintain synchronization.

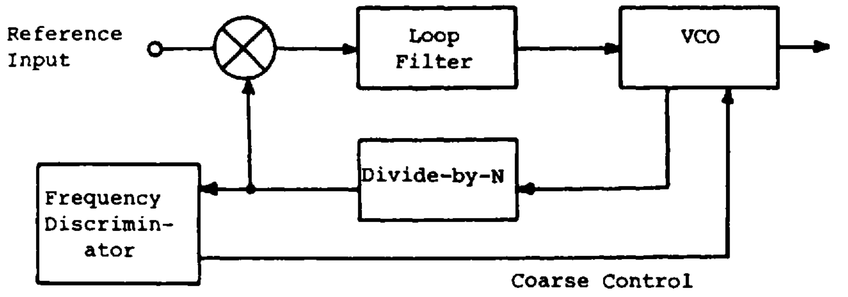
7.3.1 Multiple Indirect Synthesizers

A technique sometimes used, when switching speed from one frequency to another is of paramount importance, employs more than one phase lock synthesizer. This technique multiplexes several synthesizers together and is sometimes called a "ping-pong" synthesis technique because the output is alternately taken from two or more synthesizers.

The problem with this technique is that the frequency control input to the two (or more) synthesizers must be supplied in such a way that the synthesizer that is settling to the next frequency must have information that is different, and ahead in time, when compared with that which is current. Figure 7.15 illustrates this timing requirement, which is simply that the receiver must be able to calculate



(a)



(b)

Figure 7.14 Methods of reducing lock-in time for phase lock synthesizers using digital-to-analog conversion or frequency discrimination. (a) Digital-to-analog converter used to provide coarse steering to VCO. (b) Frequency discriminator used to provide coarse steering.

	Period 1	Period 2	Period 3	Period N→
Synthesizer One	Output	Settling	Output	Settling to N+1
Synthesizer Two	Settling	Output	Settling	Output N
Control Data	Synthesizer One Current Synthesizer one set to Period 2 Frequency	Synthesizer Two Current Synthesizer two set to Period 3 Frequency	Synthesizer One Current Synthesizer two set to Period 4 Frequency	Synthesizer Two Current Synthesizer one set to period N+1 Frequency

Figure 7.15 Illustration of timing in multiple or "ping-pong" synthesis technique.

the frequency that is required for up to N periods before it is actually required, where N is the number of synthesizers multiplexed together.

The advantage of this technique is that one synthesizer's output is already available while the other is being switched and settling to the next frequency. This means that two synthesizers can be used to provide a switching rate (usually called a hopping rate) that is M times faster than is practical with a single synthesizer, where M is typically a factor of 10 for two synthesizers.

If a single synthesizer can settle in time t_{lock}, and this is allowed to be 10 percent of the period at a given frequency, then the hopping rate must be

$$R_{\text{hop}} = \frac{1}{10t_{\text{lock}}}$$

When two such synthesizers are used, with one settling while the other is putting out the desired frequency, the hop rate can be increased to

$$R_{\text{hop2}} = \frac{1}{t_{\text{lock}}}$$

Does this mean that three synthesizers could be used to increase the hop rate by a factor of 100? No, unfortunately, the hopping rate for N synthesizers, under the same conditions, is

$$R_{\text{hop}N} = \frac{1}{10(N-1)t_{\text{lock}}}$$

Furthermore, when more than a few synthesizers are combined, the problem of providing multiplier frequency control signals with the required timing becomes more complex than using a direct synthesizer, which in itself is more complex than a single phase lock synthesizer but is less complex than multiple synthesizers with their control mechanism.

7.3.2 Settling Time

Settling time in a receiver's local oscillator is determined by the requirements of the particular receiver. A noncoherent receiver (such as an AM receiver), for example, requires only that the frequency of the received signal be within the bandwidth of the demodulator. However, a coherent receiver that employs a phase-locking demodulator requires that the signal be within the lock range of the demodulator, and this is a much more stringent requirement.

Figure 7.16 shows typical lock-in performance of a phase lock (indirect) synthesizer. This is a typical plot of the control voltage of the VCO in an indirect synthesizer on being given a step function change in its frequency control input.

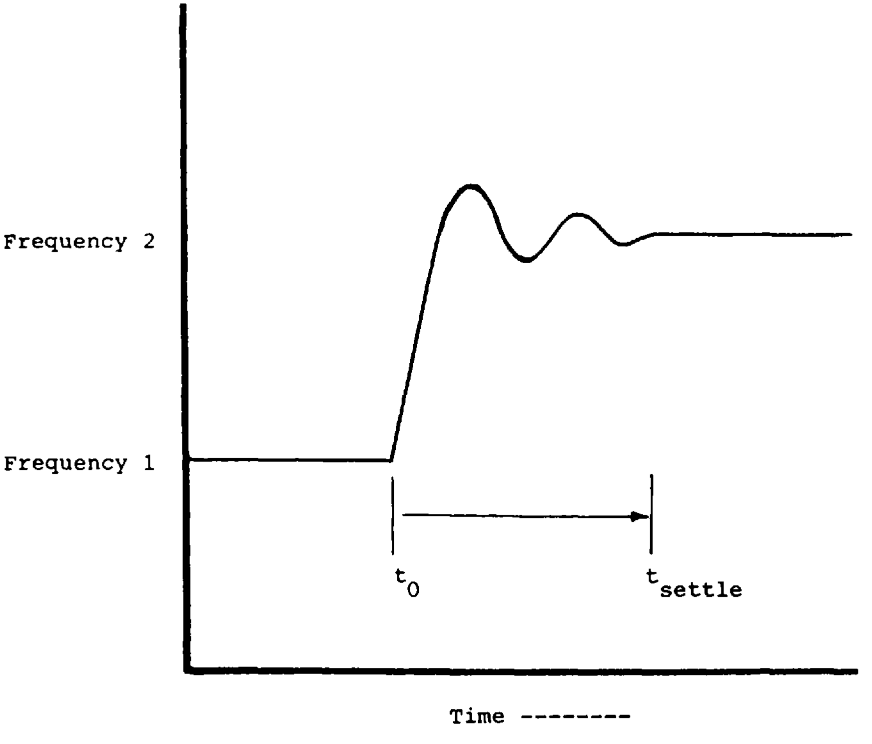


Figure 7.16 Typical settling action of a phase lock synthesizer. (Settling function is due to damping factor.)

Because the VCO frequency is a function of the control voltage, it is apparent that settling time is a function of the frequency tolerance of the receiver.

7.4 VCOs

Voltage-controlled oscillators are key to the design of a high-performance indirect frequency synthesizer. The VCO must cover the frequency range required, with sufficient sensitivity to the control voltage available to provide adequate loop gain, and yet must be insensitive to noise on its power supply or any other possible noise source. It must also produce as little internal noise as is practical. In many cases, the best course is to employ a VCO from one of the excellent suppliers that exist today. Characteristics of some typical VCOs are given in the following pages.

Some of the important characteristics of VCOs are:

Frequency range: The range of frequencies covered by the VCO, within the specified characteristics.

VCO sensitivity: Sometimes called VCO gain, sensitivity is the change in frequency per volt of change in the VCO's control voltage. It is expressed in either hertz per volt or radians per second per volt. VCO sensitivity is one of the most important components of loop gain in both phase lock frequency synthesizers and demodulators.

Linearity: Usually expressed in terms of deviation from a linear frequency versus control voltage curve. (One percent maximum deviation from linear is considered to be good linearity.) Linearity in frequency synthesizers is usually not as critical as in a phase lock demodulator. It is extremely important that the VCO frequency versus voltage curve be monotonic and have no discontinuities.

Modulation bandwidth: Modulation bandwidth for a VCO has to do with the rate of change of the control voltage. For use in frequency synthesizers, the primary concern is that the VCO does not limit the switching time. In demodulators, the VCO must track the maximum modulation rate. Although it is usually common and necessary to employ some bypassing on the VCO control input, it is advisable to make it as small as possible.

Operating voltage: The VCO must operate on the system's available supply voltage, and its control voltage must not be greater than that which can be readily derived from the supply voltage. (Having a perfect VCO does not include a 20-V control voltage range in a hand-held receiver that operates on 3 V.)

Noise: A VCO can generate noise as well as pick it up from its control and power inputs. Therefore shielding, bypassing, and high Q are all important considerations.

7.4.1 Integrated Circuit VCOs

VCOs that are incorporated into IC phase locks, loops, or synthesizers are seldom good enough to be employed in a high-quality receiver as the local oscillator, because of their noise characteristics. However, a few are included here for reference. (The same comment holds for integrated circuit VCOs that are stand-alone, in a separate package. They are typically too noisy.)

7.4.2 Custom VCOs

Anyone, of course, is welcome to design his own VCO. It is, in fact, often necessary to do so if only to fill a gap while the purchasing process takes place. A few basic designs follow.

7.4.2.1 PNP Transistor VCO

Figure 7.17 is a schematic diagram of a simple VCO that can typically be tuned over an octave or more of bandwidth without the tuning voltage exceeding its power supply voltage. It can also be designed to operate from low frequencies to frequencies in the 1- to 2-GHz range.

A low-cost PNP transistor that operates well in this circuit is the 2N5202.

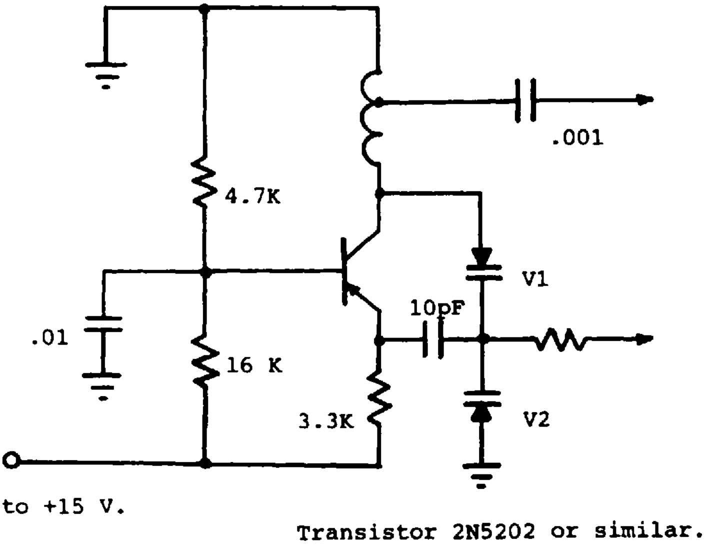


Figure 7.17 Schematic diagram of a versatile VCO for general applications.

Any PNP with good F_t will generally suffice. Replacing varactor V_2 with a fixed capacitor will result in a reduced tuning range.

The critical components for determination of operating frequency are the inductor in the collector circuit (note that the inductor is at ground potential, which is possible because a PNP transistor is used), the varactors, and the small feedback capacitor in the emitter circuit.

Use of a PNP transistor also makes it more practical to employ a positive-going control voltage and helps to provide a low-impedance output point that is taken from a tap on the tuning inductor. The tuning inductor itself is usually a short strip of copper or brass (one-half turn) at high frequencies, with the output tap somewhere near the center. This is illustrated in Figure 7.18. Many variations are practical, of course, including printed circuit loops and coils of wire. (The particular configuration depends on the VCO frequency of operation.)

7.4.2.2 Tunnel Diode VCO

Another easily constructed VCO is one that makes use of a neglected component, a tunnel diode. Tunnel diodes make excellent oscillators that take advantage of their negative resistance property. (RF engineers will appreciate that if a tunnel diode is biased in its negative resistance region, it is difficult to prevent it from oscillating.) Figure 7.19 shows a typical current-versus-voltage curve for a tunnel

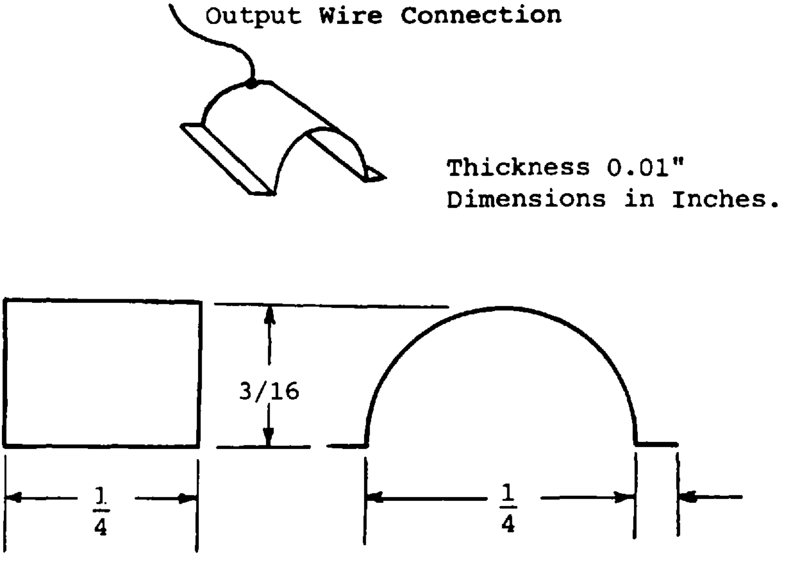


Figure 7.18 Typical inductor for high frequencies, for use with VCO of Fig. 7.17.

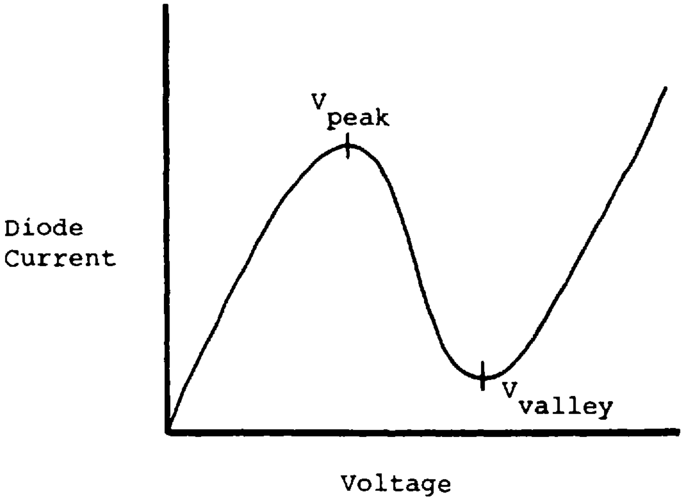


Figure 7.19 Tunnel diode current versus voltage curve, showing negative resistance region.

diode, which exhibits a strong negative resistance region in which the diode current decreases while the voltage increases.

Depending on the material from which the diode is made (germanium, silicon, gallium arsenide, etc.) V_{peak} and V_{valley} vary, but both are normally less than 1 V. Two conditions are required for oscillation:

1. The diode must be biased in the negative resistance region.
2. The resistance of the biasing circuit must be lower than the negative resistance of the diode.

The center of the negative resistance region is

$$\frac{V_{peak} + V_{valley}}{2}$$

and the current at this point is

$$\frac{I_{peak} + I_{valley}}{2}$$

which means that the negative resistance at the center of the region is

$$\frac{V_{center}}{I_{center}} = \frac{V_{peak} + V_{valley}}{I_{peak} + I_{valley}}$$

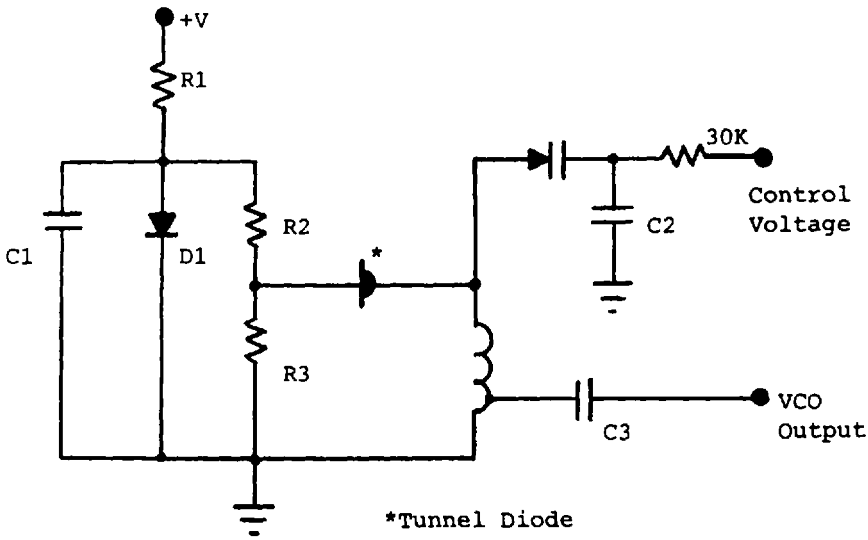


Figure 7.20 Tunnel diode VCO circuit diagram.

For typical diodes, the negative resistance is 25 to 75 ohms, and the negative resistance region is centered at about 0.1 to 0.5 V.*

Figure 7.20 illustrates a simple circuit for a tunnel diode oscillator that meets the criteria listed. In this circuit, D_1 is used to regulate the voltage applied to R_2 and R_3, and R_1 is used to drop the supply voltage. C_1 is a bypass to provide low impedance at the operating frequency. R_2 and R_3 divide the voltage across D_1† to produce the tunnel diode (TD) bias and must provide a source resistance less than the TD's negative resistance. Therefore, the resistance of R_2 and R_3 must be such that

$$\frac{R_2 R_3}{R_2 + R_3} < R_{\text{negative}}$$

and

$$\frac{R_2 R_3}{R_2 + R_3} \times V_{\text{D1}} = V_{\text{TD}}$$

*The negative resistance curve of a tunnel diode can be measured on a standard diode/transistor curve tracer, but it is usually necessary to place a resistor (perhaps 100 to 200 ohms) in series with the tunnel diode. Otherwise, it will oscillate.

†D_1 may actually consist of one or two diodes. Almost any silicon diode is acceptable, including computer diodes and rectifiers.

C_1 and C_2 are bypass capacitors, and C_3 provides DC blocking for the output signal, which is taken from a tap on the tuning coil.

Tunnel diode oscillators tend to be stable with temperature, and the tunnel diode itself is a very low-noise device. The main drawback to their use is that only a few manufacturers make tunnel diodes today, and they are made primarily for extremely high-speed digital applications.

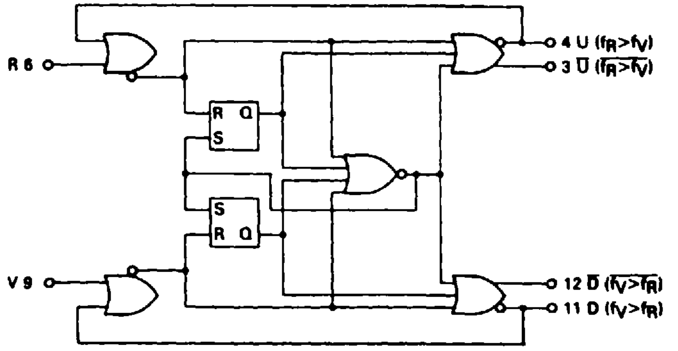
7.5 PHASE DETECTORS FOR FREQUENCY SYNTHESIS

In frequency synthesizers, unlike phase lock demodulators, the phase detector employed is digital. The simplest type is an exclusive-OR gate, but the best and most prevalent today is the edge-triggered, multi-flip-flop type such as the Motorola MC 4043. This type is found within most of the integrated circuit synthesizers available today because of two significant advantages:

1. Edge triggering allows operation without the need for symmetry of either the reference signal or the divide-by-N frequency counter output.
2. Discriminator and phase detector action are combined, which reduces synthesizer settling time by a factor of 5 or more in typical applications.

Figure 7.21 shows a block diagram of a typical edge-triggered discriminator/phase detector.

Other phase detectors that are useful in frequency synthesizers are exclusive-OR gates and RS flip-flops. However, none of the digital phase detectors is recommended for use in a phase lock demodulator, for reasons that will be discussed in Chapter 8, although all of the digital phase detectors offer higher gain



INPUT		OUTPUT			
R	V	U	D	$\overline{U}$	$\overline{D}$
0	0	X	X	X	X
0	1	X	X	X	X
1	1	X	X	X	X
0	1	X	X	X	X
1	1	1	0	0	1
0	1	1	0	0	1
1	1	1	0	0	1
1	0	1	0	0	1
1	1	0	0	1	1
1	0	0	0	1	1
1	1	0	1	1	0
1	0	0	1	1	0
1	1	0	1	1	0
0	1	0	1	1	0
1	1	0	0	1	1

X = Don't Care

Figure 7.21 Block diagram and truth table for Motorola MC 12040 phase-frequency, edge-triggered phase detector. (Courtesy of Motorola, Inc.)

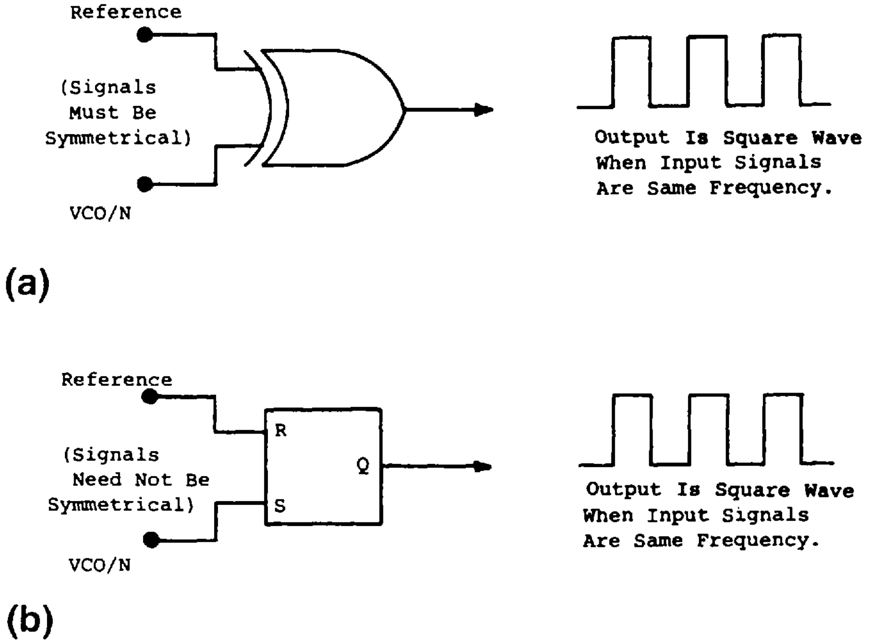


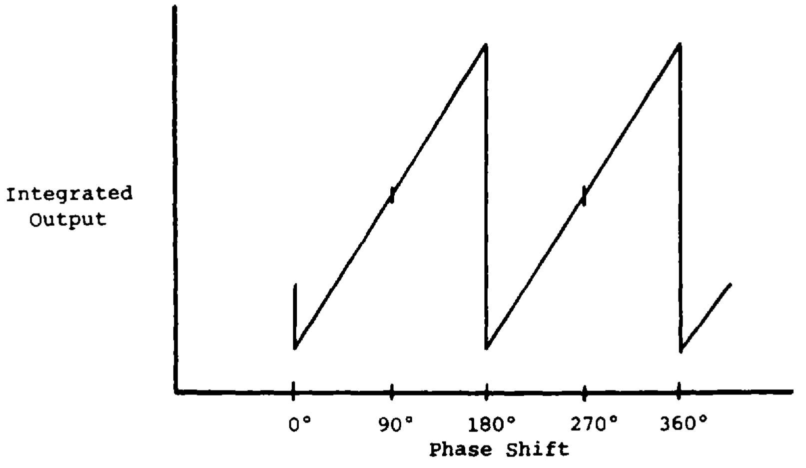
Figure 7.22 Exclusive-OR and R-S flip-flop phase detectors. (a) Exclusive-OR phase detector. Symmetry of output varies as input square wave frequencies differ. (b) R-S Flip-flop phase detector. Output is similar to that of exclusive-OR phase detector, but input signals can be unsymmetrical.

than the analog phase detectors that should be employed in demodulation. Phase detector gain as referred to here is the phase detector output voltage as compared with the amount of phase shift between its input signals.

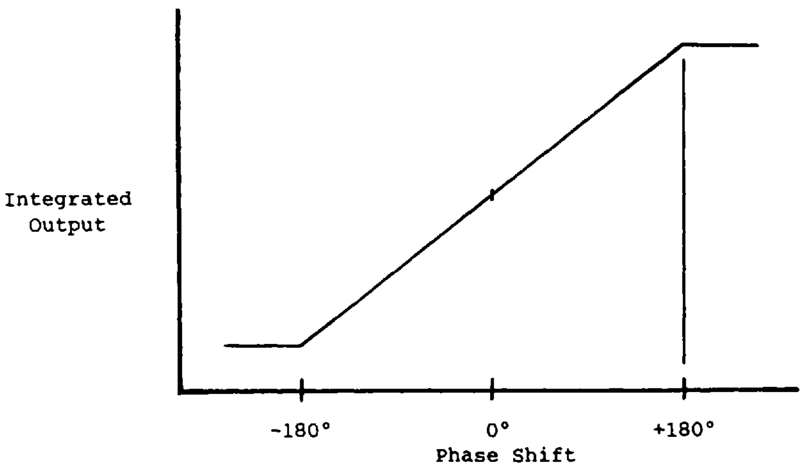
Exclusive-OR gate and RS flip-flop phase detectors are shown in Figure 7.22. Both phase detectors, as well as the edge-triggered phase detector, offer more linear operation than an analog phase detector. They also offer higher gain, but this is all at the expense of their requirement for digital input signals. Figure 7.23 shows typical digital phase detector output versus input signal phase shift.

7.6 DIRECT SYNTHESIZERS

When it is required to have a local oscillator that can shift rapidly one frequency to another, it is usually necessary to employ a "direct" synthesizer (i.e., a synthesizer that does not employ a phase lock loop). Such a synthesizer may be as simple as a

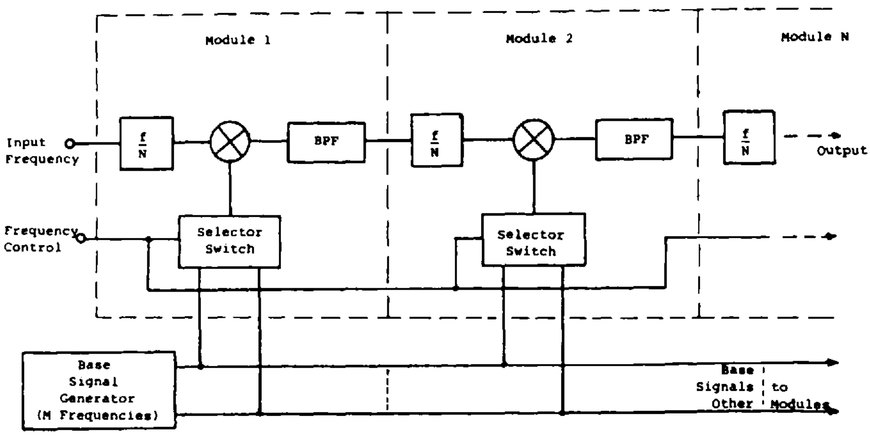


(a)



(b)

Figure 7.23 Integrated voltage output for digital phase detectors as used in typical frequency synthesizers. Output is a function of phase shift between the two input signals. (a) Exclusive-OR and R-S flip-flop phase detector, integrated output. (b) Edge-triggered phase detector, integrated output.



Number of frequencies available = M^k (M is the number of base frequencies, and k is the number of modules.)

Frequency Separation = $\frac{\Delta M}{M^{(k-1)}}$ (ΔM is the separation between the frequencies in the base set.)

Figure 7.24 Add-and-divide frequency synthesizer for high-speed frequency-to-frequency switching.

pair of oscillators with a switch to select between them. Modern receivers usually require far more than two frequencies, however. Even broadcast band receivers require 98 frequencies for the FM band, and 103 frequencies for the AM band, but it is not practical to provide 100 oscillators and a 100 position switch to select a desired frequency.*

A much more practical synthesis technique is illustrated in Figure 7.24. This direct synthesizer configuration (called "add and divide" or "mix and divide") was described in 1963 by Hastings and Stone and is employed today in high-speed multiple frequency synthesizers. The advantage of this particular direct synthesis technique is that it can generate M^K frequencies with spacing $\Delta M/(K - 1)$ and switch between any of them in less than a microsecond. (The typical switching time is 100 to 200 nsec.) Here M is defined as the number of base frequencies, K is the number of modules, and ΔM is the frequency separation between the base frequencies. With the exception of very high-speed frequency hopping receivers[3], the capabilities of add-and-divide synthesizers are seldom fully exercised. Their greatest drawback, however, is that there are no commercially available integrated circuit modules for implementing add-and-divide synthesizers.

*What was done before the advent of low-cost integrated circuit synthesizers? A simple manually tuned oscillator served as the local oscillator in all but the most complex and expensive receivers.

7.7 LOCAL OSCILLATOR/SYNTHESIZER PHASE NOISE

Noise in a receiver's local oscillator(s) is impressed upon any signal that is mixed with the oscillator. That is, an input signal that has no modulation, being mixed with a local oscillator that is noisy, will have the local oscillator's noise spectrum when translated to the IF. Likewise, if the input signal has modulation, then its modulation will be convolved with the noise from the local oscillator when translated to the IF. This process then degrades the performance of the receiver if the local oscillator's noise level is significant compared with the noise accompanying the input signal when it reaches the mixer.

Local oscillator and synthesizer* noise is characterized in one of two ways. One of these is termed "single-sideband phase noise" and is a frequency domain measure. The second is "Allan variance," which is a time domain measure.

For all practical purposes, single-sideband phase noise ($\mathscr{L}$) is a spectrum analysis tool, and Allan variance $\sigma^2_{(T)}$, is a form of time jitter analysis. Historically, oscillator and frequency synthesizer manufacturers have specified performance in terms of single-sideband phase noise, but some have specified Allan variance as well. Figure 7.25 is a typical graph showing SSB phase noise for an HP8642 signal generator operating at 1.0 GHz.

Single sideband phase noise is the noise (in dBc) in a 1-Hz bandwidth, measured at a frequency that is offset from the carrier by a specific amount. For the Hewlett-Packard 8642 signal generator, Figure 7.25 shows that SSB phase noise is approximately −70 dBc at 100–Hz offset, −95 dBc at 1–kHz offset, and −145 dBc at 100-kHz offset. (This, incidentally, is excellent performance.) Receivers can operate with SSB phase noise that is far worse (as much as 60 to 70 dB worse) in many instances than is shown in Figure 7.25.

It is practical to employ a laboratory spectrum analyzer to measure SSB phase noise. Remember, however, that very few spectrum analyzers have 1-Hz bandwidth measurement capability. Therefore, a conversion in measurement must be made, which is to add an amount 10 log(actual bandwidth/1 Hz) to the measurement made. If the actual (noise) bandwidth of the spectrum analyzer being used is unknown, the manufacturer should be contacted.

Allan variance is measured with a frequency count technique, so that, conceptually, a laboratory frequency counter could be used. The problem, however, is that laboratory frequency counters do not usually have the capability to measure over the shorter time intervals needed for Allan variance measurement [i.e., Allan variance, $\sigma^2_{(T)}$, is often measured over millisecond intervals or less, and standard frequency counters are not designed to do this]. These measurements in specifications are usually presented in a tabular form as shown in Table 7.4.

*Here we will not differentiate between fixed local oscillators and frequency synthesizers.

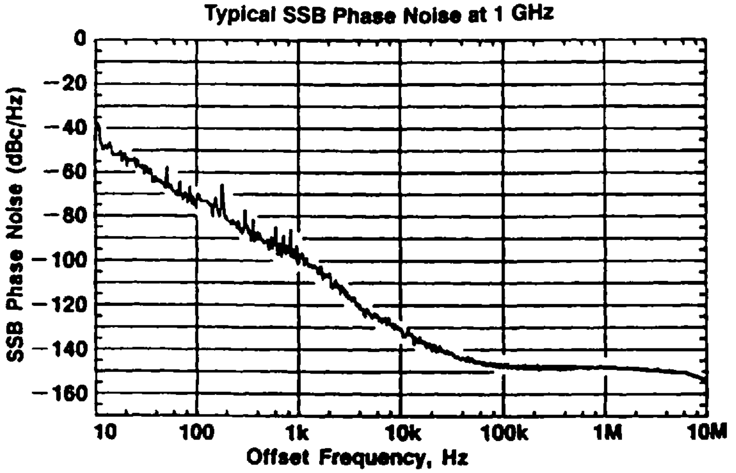


Figure 7.25 Hewlett-Packard HP8642 phase noise plot. (Courtesy of Hewlett-Packard.)

Observing Table 7.4 one can see that Allan variance is a measure of the stability of a frequency source over various time intervals. It is measured by counting the frequency* over a reasonable number (10, for example) of contiguous equal intervals. The method of measuring Allan variance used by Tracor and described by them is as follows:

1. Make N frequency measurements in N consecutive, contiguous, equal time intervals.
2. Find the $N - 1$ frequency differences in the N time periods, where

$$\Delta f_{(N-1)} = f_{(n)} - f_{(n-1)}$$
$$\Delta f_{(N-2)} = f_{(n-1)} - f_{(n-2)}$$
$$\vdots$$
$$\vdots$$
$$\Delta f_{N-(N-1)} = f_{n-(n-1)} - f_{n-(n-2)}$$

3. Square the differences (there are $N - 1$ differences).
4. Sum the $N - 1$ squares.
5. Find the square root of the sum.
6. Divide by $2N$.

*Remember that a frequency counter gives the average frequency per unit time.

Table 7.4 Typical Allan Variance Presentation

T	$\sigma^2(T)$	T	$\sigma^2(T)$
1 msec	5×10^{-6}	10 sec	5×10^{-9}
10 msec	2×10^{-7}	100 sec	3×10^{-9}
100 msec	3×10^{-8}	1000 sec	1×10^{-9}
1.0 sec	1×10^{-8}		

The answer is the Allan variance for that particular time interval and is

$$\sigma^2_{(T)} = \frac{\sqrt{\Delta f_1^2 + \Delta f_2^2 + \Delta f_3^2 \cdots + \Delta f^2{}_{N-1}}}{2N}$$

A separate measurement must be made in N time intervals for each time interval of interest and the calculation done to determine the Allan variance. Fortunately, automatic instruments are available to measure and calculate the results. Otherwise, few would ever consider anything other than SSB phase noise measurements.

In practice, the choice between SSB phase noise and Allan variance may already be moot, as the 1995 Hewlett-Packard catalog no longer gave Allan variance measurements of its instruments. It did, however, describe software that allows a standard laboratory spectrum analyzer to make SSB phase noise measurements directly, without considering correction factors for bandwidth. One can convert from SSB phase noise to Allan variance, or vice versa, although it may be easier to make the measurement itself than to convert from one to the other. Both Rohde [15] and Spilker [16] address the conversion problem and provide conversion tables. The reason for this problematic conversion is that the noise spectrum of a frequency source is made up of (at least) five components:

1. Random frequency walk
2. Flicker frequency noise
3. Random phase walk (white FM)
4. Flicker phase noise
5. White phase noise

These components each dominate within a particular region of the phase noise spectrum, and the conversions are a piecewise linear approximation to the spectrum. Each component is converted separately. Therefore, it may well be quicker and more accurate to make the measurement in question than to convert from one to the other.

Figure 7.26 is a spectrum analyzer presentation of the output of a typical phase lock frequency synthesizer. The spurs, which are down approximately 45 to 50 dB, are at the reference frequency and its harmonics, and are the most difficult-

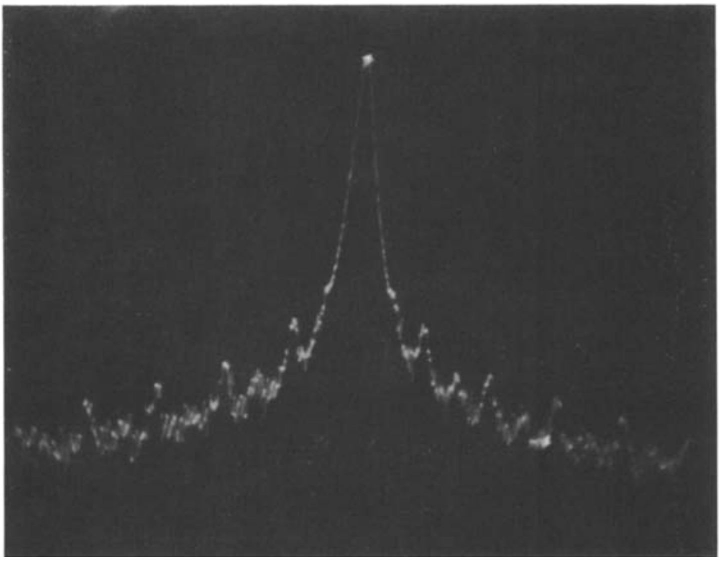


Figure 7.26 Typical phase lock synthesizer output spectrum showing reference-related sidebands.

to-suppress unwanted output from a phase lock synthesizer. They are produced by undesired frequency modulation of the VCO by the reference frequency output of the phase detector, which should ideally be rejected by the loop filter.

7.8 REDUCTION OF LOCAL OSCILLATOR PHASE NOISE

A natural question that must arise when considering oscillator phase noise is, "Why do we consider only phase noise?" If we assume a signal such as that from a local oscillator, we can consider that the noise is made up of two components in quadrature, one of which is an amplitude noise component and the second a phase noise component. Figure 7.27 illustrates this.

If the noise is random, then the noise power in $B \sin \omega_n t$ is equal to $B \cos \omega_n t$. However, the action of an oscillator is such that its output amplitude continues to build up until it cannot increase any more. Why? Because it is regenerative.

This means that an oscillator's output normally suppresses amplitude noise, although it can appear as AM-to-PM* noise but at a greatly reduced level. Of

*Amplitude modulation to phase modulation.

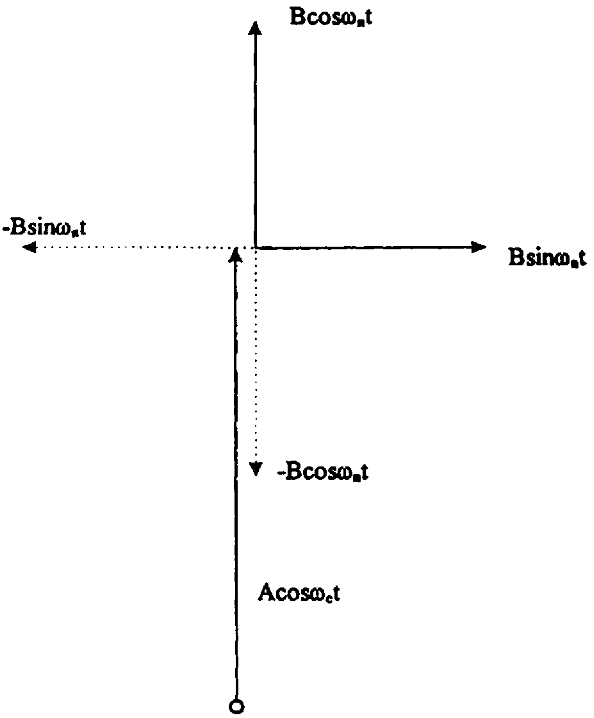


Figure 7.27 Local oscillator signal with amplitude and phase noise components.

course, the higher the Q of the resonator, the lower is the effect of AM-to-PM noise conversion, as well as the phase noise itself. The overall result is that there is very little AM noise to start with, and what little there is, is converted to phase noise.

The best approach to phase noise reduction is to increase the Q of the resonator in the signal source. Proceeding from the lowest Q to the highest Q, we can employ the resonators listed in Table 7.5.

7.9 FREQUENCY OFFSET DUE TO *G* LOADING

Crystals exhibit frequency offset due to the effects of gravity or G loading such as that experienced in maneuvering high-speed aircraft. The effect of simply turning a crystal from side to side is shown in Table 7.6.

Table 7.5 Resonators Listed in Order of Increasing Q

Resonator	Value of Q
RC	Up to 100
LC	Up to 500
Dielectric	Up to 1000
Cavity	Up to 1000
SAW	Up to 10000
Crystal	Up to 10^6 (and higher)

Specially designed crystals, intended to compensate for the effects of G loading, have been developed. These are termed stress-compensated ("SC cut") crystals or "doubly rotated" crystals and are very effective in reducing frequency shift due to G loading. A report from Hewlett-Packard has shown that in a production run of 1000 oscillators using SC cut crystals, all oscillators exhibited a change of 2×10^{-9} or less for a $2G$ change. (This $2G$ change can be accomplished by simply turning an oscillator over.)

7.10 COMPARISON OF TECHNIQUES FOR GENERATION OF HIGH FREQUENCIES

Figure 7.28 shows the result of tests carried out to compare the single-sideband phase noise produced by three different means of generating high-frequency signals:

1. Direct multiplication of a quartz crystal–stabilized oscillator, generating a low-frequency signal, to a high frequency.
2. Use of a phase lock loop synthesizer to multiply a low crystal frequency reference signal to a high frequency.
3. Use of a high-frequency surface acoustic wave resonator in an oscillator to generate a high-frequency signal directly.

At offsets of more than approximately 8 kHz, Figure 7.28 shows that the direct, surface acoustic wave oscillator approach has the lowest phase noise. Below 1 kHz offset, however, the phase lock technique is lowest in phase noise.

It should be pointed out that all of the test results shown here would be acceptable for many receiver applications. Only the most critical receiver designs require single-sideband phase noise less than -100 dBc/Hz at 1 kHz offset.

Table 7.6 Typical Frequency Shift for Commercial Crystals as a Function of Orientation

Nominal frequency (MHz)	10-crystal average side 1 frequency	Side 2 (avg.)	Side 3 (avg.)	Side 4 (avg.)	Side 5 (avg.)	Side 6 (avg.)	Maximum/ (average offset)
1.000000	1.0004084	1.0004110	1.0004088	1.0004102	1.0004104	1.0004062	2.6×10^{-6}/ (2.8×10^{-7})
3.000000	2.99999067	2.9999047	2.99999014	2.99999021	2.99999025	2.99999325	3.2×10^{-7}/ (2.2×10^{-7})
5.000000	5.0006085	5.0006078	5.0006071	5.0006078	5.0006069	5.0006073	3.2×10^{-7}/ (1.8×10^{-6})
6.553690	6.5536955	6.5536979	6.5536903	6.5536888	6.553926	6.5537072	1.78×10^{-6}/ (8.8×10^{-7})

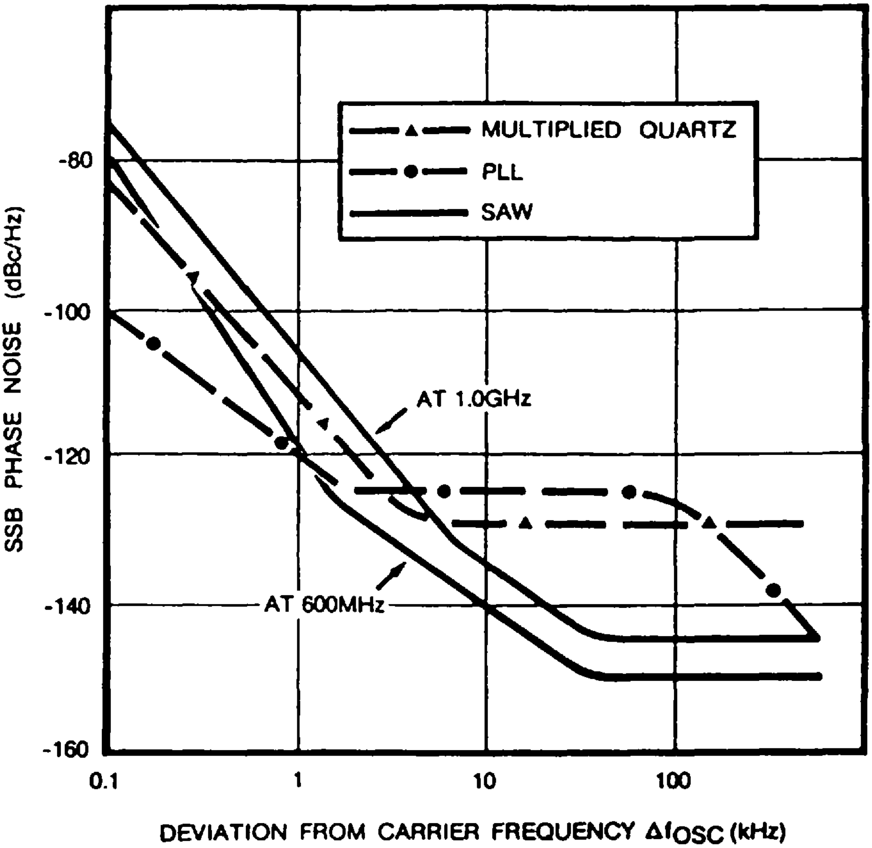


Figure 7.28 Comparison of single-sideband phase noise characteristics of three techniques for high-frequency local oscillator signal generation.

QUESTIONS

1. What is the effect on a receiver of noise and/or spurious content from a local oscillator?
2. What are the advantages of employing a phase lock synthesizer as a local oscillator? Disadvantages?
3. What is the difference between "accuracy" and "stability" in a local oscillator? Does a specification mean anything when it says "Frequency stability—1 × 10^{-6}"?
4. Compare the advantages and disadvantages of direct and indirect frequency synthesizers as local oscillators.

5. What is the meaning of "single-sideband phase noise"? What is meant by "Allan variance"?
6. Why do phase lock frequency synthesizers tend to employ digital phase detectors?
7. What is the effect of choosing high-side injection in a receiver? Can high-side and low-side injection be used in the same receiver?
8. How is the reference frequency chosen in a phase lock synthesizer? What is the source of spurious outputs from a phase lock synthesizer?
9. How can the reference-related spurious content in a phase lock synthesizer be reduced? How much reduction can be expected?
10. What can be done to reduce the switching time needed by a phase lock frequency synthesizer?

REFERENCES

1. Telecommunications Products, 1992 Data Book, Fujitsu Microelectronics, pp. 3-3 through 3-176.
2. Product Catalog, A Guide to SAW Products and Applications, SAWTEK Incorporated, pp. 26–33.
3. CMOS/NMOS Special Functions Data, Motorola, pp. 6-1 through 6-128.
4. Solutions for Wireless Communications, Pacific Monolithics, Inc.
5. Kiser, J. H., Miniature VCOs cut current in portable designs, *Microwaves & RF*, Jan. 1997.
6. Zavrel, R. J., Jr., and Edwards, G., *The DDS handbook*, Stanford Telecommunications, 1990.
7. Direct Digital Synthesis, 21 Questions & Answers for RF Engineers, Qualcomm.
8. Q 3036 1.6 GHz PLL Frequency Synthesizer, Qualcomm, 1990.
9. User's Guide for Q 0710-1 Direct Digital Synthesizer (DDS) Driven Phase-Locked Loop (PLL) Frequency Synthesizer, Qualcomm, Sept. 1992.
10. Q 3036 Phase Locked Loop Frequency Synthesizer Phase Noise Measurements, App. Note AN3036-1, Qualcomm, May 1992.
11. VCO Product Catalog, Z-Communications, Inc., 1992.
12. Beam, W. C., and Rezin, J. R., Design Considerations for Fast Switching PLL Synthesizers, Watkins-Johnson Tech-Notes, Mar./Apr. 1991.
13. Buswell, R. N., How to Specify and Use a DTO/VCO, Watkins-Johnson Tech-Notes, July/Aug. 1987.
14. Anderson, H. W., Low-Power VHF/UHF Synthesizer Design, Watkins-Johnson Tech-Notes, Sept./Oct. 1983.
15. Rohde, U. L., Digital PLL Frequency Synthesizers, Prentice-Hall, Englewood Cliffs, NJ, 1983.
16. Spilker, J. J., Digital Communications by Satellite, Prentice-Hall, Englewood Cliffs, NJ, 1977.

8
Demodulator Design and Application

No part of a receiver is more critical to its performance than its demodulator. The demodulator is often the most difficult single part of a receiver to design and to optimize, and as modulation has become more and more complex, demodulators have been forced to keep pace. On the modulation side, more and more data are being sent in less and less bandwidth, while demodulators are being called on to provide lower error rates under ever-deteriorating conditions. Here, we will categorize demodulators on the basis of the type of modulated signal they are intended to work with.

It should also be realized that a receiver's demodulator input is the point that must be controlled with respect to the receiver's noise bandwidth and signal amplitude. The receiver will not work unless the signal level and its signal-to-noise ratio are adequate at the demodulator input. With this in mind, we will proceed to discuss demodulation and demodulators and their role in the reception of both analog and digital information. (Please note that no mention has been made of "digital" modulation, as there is no such thing as far as a receiver is concerned.) We will discuss the use of digital techniques in receivers, but the signals being received are *not* digital. Tables 2.1 and 2.2 (from Chapter 2) are reproduced here, and they list types of demodulators and the types of modulation considered. We will begin with demodulators intended for analog information.

8.1 AMPLITUDE MODULATION

In this section we will discuss envelope detectors, product detectors, and phase lock loops, which are the most common types of demodulators for use with amplitude-modulated signals.

Table 2.1 Analog Modulation Techniques, Bandwidths, and Demodulators

Analog modulation method	Modulated RF bandwidth	Modulated signal representation	Demodulator(s) typically employed
AM (DSB)	$2B_{\text{info}}$	$A\cos\omega_c t + \frac{Am}{2}[\cos(\omega_c + \omega_m)t + \cos(\omega_c - \omega_m)t]$	Envelope detector Product detector Phase lock loop
AM (DSBSC)	$2B_{\text{info}}$	$\frac{Bm}{2}[\cos(\omega_c + \omega_m)t + \cos(\omega_c - \omega_m)t]$	Product detector Costas loop Squaring loop
AM (SSB)	B_{info}	$Bm\cos(\omega_c + \omega_m)\,t$ or $Bm\cos(\omega_c - \omega_m)\,t$	Product detector (except vestigial carrier SSB)
FM	$\geqslant 2(f_m + f_d)$	$B\cos(\omega_c t + \cos\omega_m t)$	Discriminator Ratio detector Quadrature detector Phase lock loop

Table 2.2 "Digital" Modulation Techniques and Bandwidths

Digital modulation method	Modulated RF (null-null) bandwidth	Modulated signal representation	Demodulator(s) typically employed
M-ASK	$\frac{4R_{\text{data}}}{M}$	$\frac{DA}{M-1}\cos\omega_c t$	Envelope detector
NCFSK	$2R_{\text{data}} + \Delta f$	$f(0) = A\cos\omega_0 t$ $f(1) = A\cos\omega_1 t$	Any FM demodulator except phase lock loop Dual filter demodulator
CPFSK[a]	1.3 to $1.5R_{\text{data}} + \Delta f$	$A\cos\left(\frac{\pi t}{2T} - U\sin\frac{2\pi t}{T}\right)$ $0 < U < 1,\ -T < t < T$	Dual filter demodulator
M-FSK	$2R_{\text{data}} + (m-1)\Delta f$	$f(M) = A\cos\omega_m t$	Dual filter demodulator
BPSK	$2R_{\text{data}}$	$A\cos\omega_c t \pm 90°$	Squaring loop Costas loop Squaring demodulator
QPSK/OQPSK	R_{data}	$A\cos\omega_c t \pm 90° \pm 45°$	Multiplier loop Costas loop Multiplier demodulator
M-PSK	$\frac{4R_{\text{data}}}{M}$	$A\cos\omega_c t \pm 90° \pm 45° \pm 22.5° \cdots \pm 180°$	Multiplier demodulator

[a]Including MSK and other exotic waveforms.

8.1.1 Envelope Detector

Envelope detectors are the most used demodulators for amplitude-modulated signals. This is because of their simplicity, ease of use, and very low cost. An envelope detector may consist of as few parts as one diode, one resistor, and one capacitor, as shown in Figure 8.1.

The design of an envelope detector is dependent on its RC time constant being long compared with the period of the RF carrier and short compared with the period of the modulating signal. This is usually accomplished by employing an IF frequency in the range of 100 times higher than the modulating frequency, which in turn permits an RC time constant 10 times shorter than the shortest modulating period but 10 times longer than the RF carrier period.

Standard broadcast band receivers are excellent examples of the choice of an IF frequency, where the AM modulation is voice or music with a bandwidth of 5 kHz. In such AM receivers, the IF frequency is usually 455 kHz. This puts the carrier frequency into the envelope detector at 91.6 times the modulation bandwidth, which makes the choice of a time constant in the detector relatively simple. Choosing the time constant at 20 μsec allows the RC filter to be able to follow the modulation envelope, because the period of the highest modulating frequency is 5000 Hz (period 200 μsec) but the carrier frequency is 455 kHz, so the filter cannot follow the carrier. (The carrier period is 2.2 μsec.)

This demodulator operates ideally in a square law mode, although a combination of square law and higher order modes is actually more common. At any rate, the envelope detector does not rectify the received signal, as is usually supposed, because this requires too large a signal level. Instead, it employs its nonlinear transfer function (see Figure 8.2) to modify the signal in such a way that it produces a modulation component that can be selected by proper filtering.

If we consider only the square law action of the diode, the demodulator works as follows:

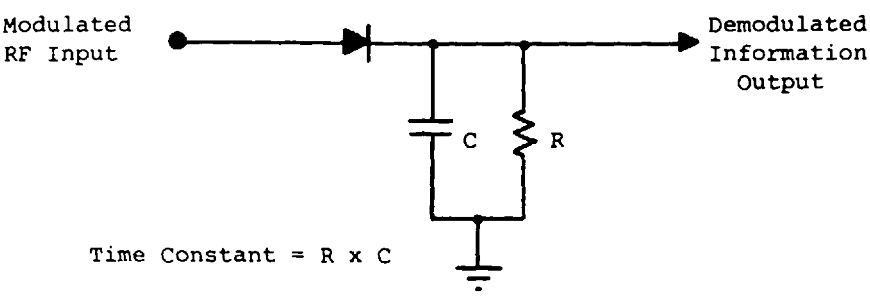


Figure 8.1 Typical envelope detector.

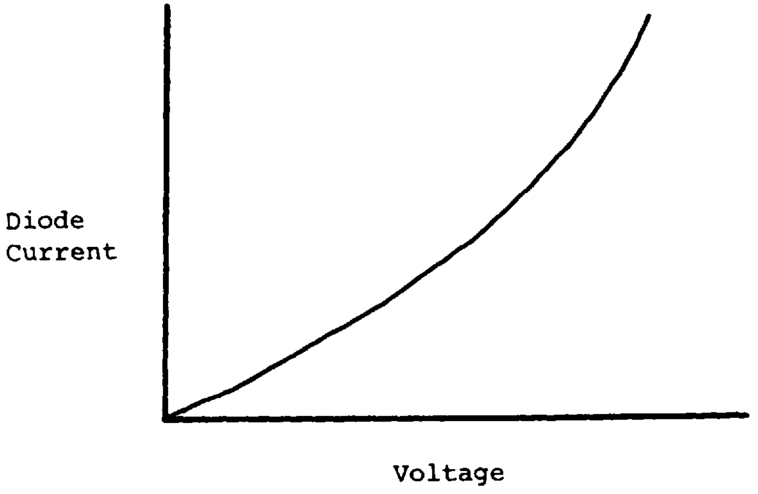


Figure 8.2 Diode voltage versus current curve. (Specific curve is dependent on diode type and material.)

Input $\quad A\cos\omega_c t + \frac{Am}{2}\cos(\omega_c + \omega_m)t + \frac{Am}{2}\cos(\omega_c - \omega_m)t$

Squared $\quad \left[A\cos\omega_c t + \frac{Am}{2}\cos(\omega_c + \omega_m)t + \frac{Am}{2}\cos(\omega_c - \omega_m)t\right]^2$

$$
\begin{aligned}
&= \frac{A^2}{2}[\cos(\omega_c + \omega_c)t) + \cos(\omega_c - \omega_c)t] \\
&+ \frac{A^2m}{4}[\cos(\omega_c + \omega_c + \omega_m)t + \cos(\omega_c - \omega_c - \omega_m)t] \\
&+ \frac{A^2m}{4}[\cos(\omega_c + \omega_c - \omega_m)t + \cos(\omega_c - \omega_c + \omega_m)t] \\
&+ \frac{1}{2}\frac{Am^2}{2}[\cos(\omega_c + \omega_m + \omega_c + \omega_m)t \\
&\quad + \cos(\omega_c + \omega_m - \omega_c - \omega_m)t] \\
&+ \frac{A^2m}{4}[\cos(\omega_c + \omega_m + \omega_c)t + \cos(\omega_c + \omega_m - \omega_c)t] \\
&+ \frac{1}{2}\frac{Am^2}{2}[\cos(\omega_c + \omega_m + \omega_c - \omega_m)t \\
&\quad + \cos(\omega_c + \omega_m - \omega_c + \omega_m)t]
\end{aligned}
$$

$$+\frac{1}{2}\frac{Am^2}{2}[\cos(\omega_c - \omega_m + \omega_c - \omega_m)t$$

$$+\cos(\omega_c - \omega_m - \omega_c - \omega_m)t]$$

$$+\frac{A^2m}{4}[\cos(\omega_c - \omega_m + \omega_c)t + \cos(\omega_c - \omega_m - \omega_c)t]$$

$$+\frac{1}{2}\frac{Am^2}{2}[\cos(\omega_c - \omega_m + \omega_c + \omega_m)t$$

$$+\cos(\omega_c - \omega_m - \omega_c - \omega_m)t]$$

$$=\frac{A^2m}{4}\cos-\omega_m t + \frac{A^2m}{4}\cos\omega_m t + \frac{A^2m}{4}\cos\omega_m t + \frac{A^2m}{4}\cos-\omega_m t$$

$$= A^2m(\cos\omega_m t) \qquad \text{after filtering}$$

where the output remaining after filtering to remove second-harmonic carrier and DC terms is the modulation term that is needed by the receiver.

8.1.2 Product Detectors

In single-sideband as well as double-sideband suppressed carrier systems, a carrier must be reinserted to demodulate the received signal. This is accomplished in a "product detector" (Figure 8.3), which multiplies a local oscillator at the IF center frequency with the signal to be demodulated. This produces a baseband signal that is separated from the carrier and its harmonics by a low-pass filter.

The major problem encountered in the use of a product detector is in demodulation of single-sideband signals, where any error in the frequency of the local carrier emerges as a shift in the frequency of the demodulated audio signal. This often results in a "Donald Duck" effect due to shifting the audio pitch as well as changing the phase and harmonic relationships of signals in the audio band. Single-sideband receivers often include a beat frequency or variable frequency oscillator, which is in effect a vernier control for the reinserted local oscillator in the receiver. It allows an operator to adjust the sound until it sounds acceptable to his or her ear.

The product detector's input signals are the desired SSB signal

$$A\cos(\omega_c + \omega_m)t^*$$

and the local oscillator

$$B\cos\omega_c t$$

and the output of the product detector is the product of the two signals

*This is the upper sideband.

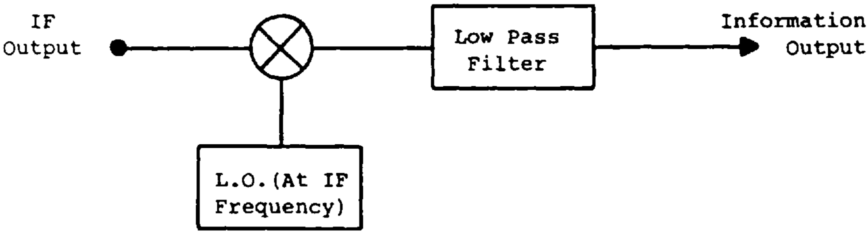


Figure 8.3 Product detector block diagram.

$$A\cos(\omega_c + \omega_m)t \times B\cos\omega_c t$$
$$= \frac{AB}{2}[\cos(\omega_c + \omega_c + \omega_m)t + \cos(\omega_c - \omega_c + \omega_m)t] = \frac{AB}{2}\cos\omega_m t$$

when the second-harmonic term $(AB/2)\cos(2\omega_c + \omega_m)t$ is filtered out by the low-pass filter following the multiplier in the detector.

The problem with the product detector occurs when there is a Doppler shift of more than about 30 Hz or there is a frequency shift in either the transmitter or the receiver local oscillator. Let us assume that a frequency shift in the input signal occurs, so that the input signal becomes $A\cos(\omega_c + \omega_m + \omega_D)t$. When multiplied with the local oscillator signal $B\cos\omega_c t$, the result is

$$A\cos(\omega_c + \omega_m + \omega_D)t \times B\cos\omega_c t$$
$$= \frac{AB}{2}[\cos(\omega_c + \omega_c + \omega_m + \omega_D)t + \cos(\omega_c - \omega_c + \omega_m + \omega_D)t]$$
$$= \frac{AB}{2}\cos(\omega_m + \omega_D)t \qquad \text{after the same low-pass filter}$$

Thus we see that the received, demodulated signal ω_m is shifted by the amount of the Doppler or other frequency shift in the input signal.

The same effect occurs if the local oscillator is shifted instead of the input signal, and worst of all, in either case the phase relationships in the received signal are not maintained. This is the effect that causes garbling of the received signal. For example, let us suppose that the frequency shift on the input signal is 100 Hz. The lowest frequency in the baseband would be shifted by 100 Hz, and 30 Hz would become 130 Hz. At the top of the band, 3000 Hz would become 3100 Hz, and the signals are no longer harmonically related (i.e., 3000/30 = 100, but 3100/130 = 23.846). Therefore, voice signals are not only shifted in pitch but also distorted by the shifting of the phase relationships in the voice components.

This is one of the primary reasons that single-sideband techniques are not usually employed in high-speed aircraft. At frequencies above about 30 MHz, Doppler shift produces intolerable distortion of received voice signals.

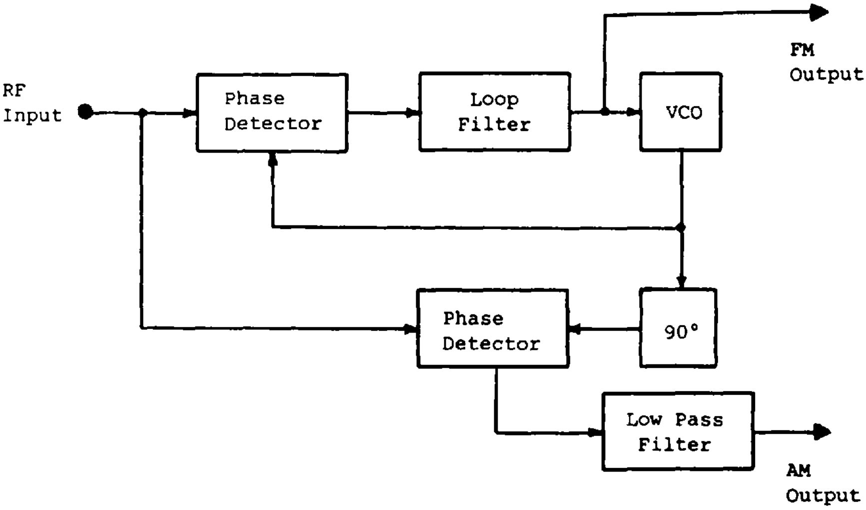


Figure 8.4 Block diagram of phase lock demodulator for either AM or FM signals. (AM output may also be employed for automatic gain control or RSSI.)

8.1.3 Phase Lock Loop AM Demodulator

The phase lock loop as an AM demodulator is quite practical, although in most applications such a demodulator is not required. That is, the usual 10- to 12-dB output AM signal-to-noise ratio from a receiver typically requires approximately the same signal-to-noise ratio into the demodulator, whether it is an envelope detector or a phase lock loop detector. There is no signal-to-noise improvement on using a phase lock loop AM demodulator, and the phase lock loop's lower threshold is not required. Furthermore, if a phase lock AM demodulator is employed, the percentage of AM modulation must be restricted. That is, if a 100 percent modulated signal is input to a phase lock loop, the loop has no choice but to lose synch when the carrier drops to zero level in the modulation troughs.

Nevertheless, in an AM/FM receiver, it could prove advantageous to employ a single phase lock loop for both AM and FM (or FSK) demodulation. Figure 8.4 is a block diagram of a phase lock loop configured for both AM and FM reception.

AM demodulation in a phase lock loop is facilitated by the fact that the voltage-controlled oscillator (VCO) in the loop synchronizes 90 degrees* out of phase with the input signal carrier. Therefore, if the VCO is phase shifted by 90 degrees, the phase-shifted VCO will be either in phase or 180 degrees out of phase

*Assuming an analog phase detector.

with the received carrier, and it can be used as a coherent reference for AM demodulation.

8.2 FM DEMODULATORS

No other type of modulation has so diverse a group of demodulators. Since the advent of FM radio in the 1920s and 1930s, many different approaches to FM demodulation have been taken. Today, the most common form of demodulator is the quadrature detector, implemented as part of an integrated circuit IF amplifier and demodulator stage. Quadrature detectors and other types of FM demodulators are described in the following pages.

8.2.1 Foster-Seeley Discriminator

The Foster-Seeley discriminator was one of the very first FM demodulators and was incorporated in many FM radios and TV receivers through the 1940s and 1950s. Its primary disadvantage is its sensitivity to amplitude variations and its need for periodic realignment. Few are used today. Figure 8.5a shows the basic diagram of a Foster-Seeley discriminator.

8.2.2 Ratio Detector

Ratio detectors replaced Foster-Seeley discriminators because of their reduced sensitivity to amplitude variations. In schematic diagram form the ratio detector and Foster-Seeley discriminator are very similar, using almost the same number and types of components. Ratio detectors also require periodic realignment. Few ratio detectors are used today. Compare the ratio detector diagram of Figure 8.5b with the diagram of the Foster-Seeley discriminator.

Both the Foster-Seeley discriminator and the ratio detector consist of a pair of envelope detectors whose inputs are tuned above and below the center frequency of their input signal. When the input signal is above the center frequency, the detector on that side outputs a larger DC signal than the detector on the other side, and vice versa. Therefore, the output that is taken across the two detector outputs (one output is grounded) varies as a function of the input signal frequency.

The ratio detector's primary difference is that one of the detector diodes is reversed, and this causes the sum of the detector output voltages to be constant although their ratio changes. For this reason, the detected output is taken between the two tuning capacitors.

Neither the Foster-Seeley discriminator nor the ratio detector is well adapted to receivers in which integrated circuitry is used, for either size or cost reasons. Both of them also convert their input FM signal to an AM signal before

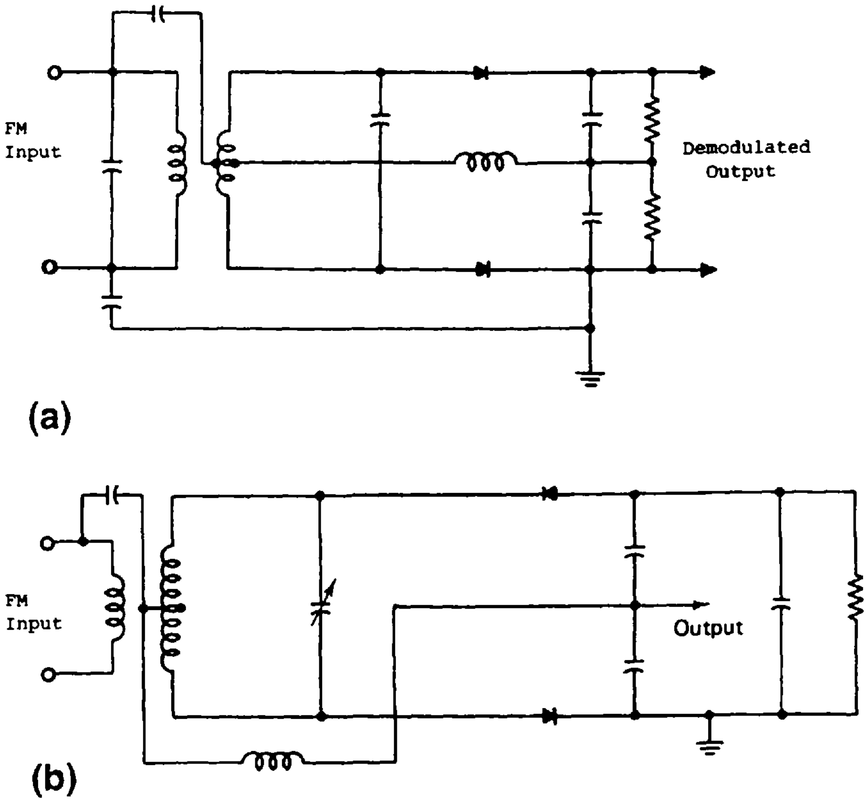


Figure 8.5 Schematic diagrams of Foster-Seeley and ratio detector FM demodulators. (a) Foster-Seeley discriminator. (b) Ratio detector.

envelope detection, which is the technique employed in the slope detector that follows.

8.2.3 Slope Detectors

Slope detection followed Foster-Seeley and ratio demodulators and solved some of their problems with its sheer simplicity. A slope detector actually works by converting the frequency modulation to amplitude modulation and then AM demodulating the signal. It does this by passing the FM signal through an amplifier or bandpass filter that is deliberately detuned with respect to the input FM carrier. For example, let us suppose that the input is a typical 10.7-MHz frequency-modulated signal. This signal is taken to an RF stage that is tuned at least 100 kHz above or below the 10.7-MHz signal. Thus, the signal falls on the slope of the amplifier's response curve, causing signals at one frequency to be amplified a different amount than signals at another frequency and converting the frequency modulation from FM to AM. The signal can then be demodulated with an envelope detector or, with careful design, in the same amplifier used to convert from FM to AM.*

Slope detectors tend to be more stable (they seldom require realignment) than Foster-Seeley and ratio detectors, as they will continue to work (within reason) as long as the signal falls in the amplifier's tuning slope. They are, in general however, less accurate than many other FM demodulators because of the imperfect linearity in a typical tuning slope. (See Figure 8.6.)

8.2.4 Pulse Count Discriminators

One of the most accurate, and linear, FM demodulators is the pulse count discriminator. It is often used in critical applications such as missile telemetry for this reason, but it is not usually used in consumer or industrial receivers. A pulse count discriminator consists of a limiter, a pulse generator, and an integrator (see Figure 8.7).

The frequency-modulated input signal is limited and the limiter output is fed to a pulse generator that triggers on the rise of the limited signal. As the frequency increases, more pulses are generated. As the frequency decreases, fewer pulses are generated. Each pulse has a fixed period that is short compared with the period of the FM carrier, and each charges the integrator by a fixed amount. Therefore, the higher the frequency, the higher the integrator's output, and the lower the frequency, the lower its output. The integrator's output then is a voltage that corresponds to the modulating function very accurately, and this is why pulse count discriminators are used in critical applications.

The degree of accuracy and linearity provided by pulse count discrimination is not usually required in other than military or space applications.

8.2.5 Quadrature Detectors

Currently, quadrature detection is the most common form of FM demodulation, because of the ease with which it may be implemented. That is, a number of integrated circuits exist that include both IF amplifiers and quadrature detectors in the same package. Two examples are the CA 3089, made by several companies, and the NE 604, made by Philips. A block diagram of a quadrature detector is shown in Figure 8.8.

---

*Actually, the slope detector does not convert from FM to AM, but it does add amplitude modulation as a function of the signal frequency, and the amplitude modulation can be recovered by an envelope detector that effectively ignores the presence of frequency modulation.

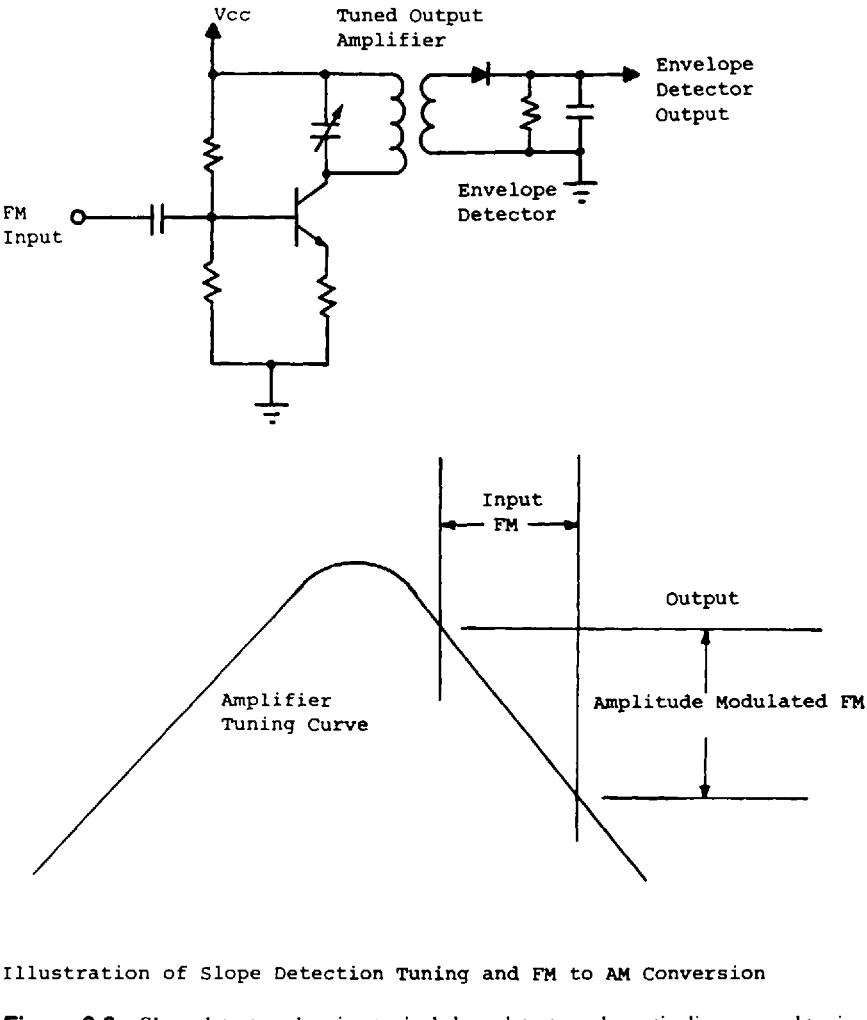


Figure 8.6 Slope detector, showing typical slope detector schematic diagram and tuning curve. Amplifier is deliberately tuned to an offset frequency so that the FM signal is on one side of the tuning curve, which causes the output signal to be both amplitude and frequency modulated, which can be demodulated by an envelope detector as an AM signal.

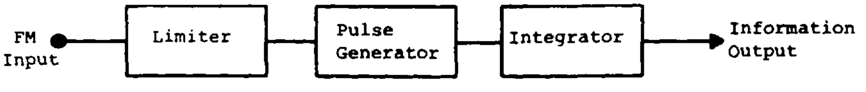


Figure 8.7 Pulse count discriminator block diagram.

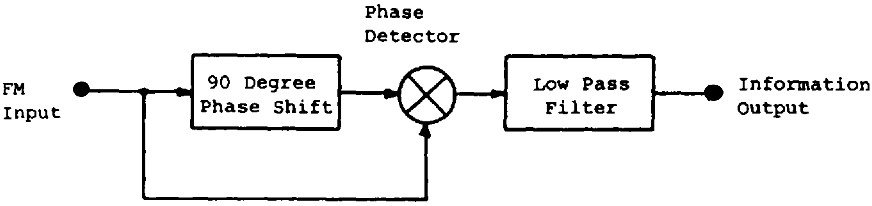


Figure 8.8 Quadrature detector block diagram.

The 90 degree phase shift used in a typical quadrature detector is provided by a bandpass filter, which provides the 90 degree shift when it is tuned to the center of the input frequency range plus a phase shift that varies as the signal deviates from the center frequency. This phase deviation allows the multiplier (phase detector) to output a varying signal as a function of the modulation present.

For all practical purposes, the quadrature detector demodulates a signal in exactly the same way as a phase lock loop demodulates the same signal. That is, the quadrature detector shifts the input signal by 90 degrees and multiplies the phase-shifted and unshifted signals together. The 90 degree phase shifter used is usually a bandpass filter centered at the carrier frequency. Thus the demodulation process is one in which an input signal $A\ \cos(\omega_c + \phi_m)t$ is multiplied with a filtered and shifted version of itself, $A \sin \omega_c t$:

$$A \cos(\omega_c + \phi_m)t A \sin \omega_c t$$
$$= \frac{A^2}{2}[\cos(\omega_c + \omega_c + \phi_m)t + \cos(\omega_c - \omega_c + \phi_m)t]$$
$$= \frac{A^2}{2}\cos \phi_m t \qquad \text{(after low-pass filtering)}$$

This remaining term is the desired modulation function.

Comparing the phase lock loop with the quadrature detector, the phase detector in the phase lock loop multiplies the input signal with the loop VCO, which maintains a 90 degree phase shift with respect to the input signal carrier. Thus the signals multiplied together in the phase lock loop are essentially the same as those multiplied together in the quadrature detector. The primary difference is that in one case (the phase lock loop) the VCO is multiplied with the input signal, and in the other case the signal itself is bandpass filtered and multiplied with the input. With respect to performance, the phase lock loop can be expected to have a lower threshold, as its carrier reference signal (the VCO) is cleaner. For this reason, a phase lock demodulator is often considered to be a "threshold extension" demodulator.

8.2.6 Phase Lock FM Demodulators

Phase lock loops make excellent FM or FSK demodulators, as long as their limitations are considered. Unlike other FM demodulators, phase lock loops must synchronize before they can work, and the lock-in time required must be taken into account. The lock-in time for a phase lock loop is approximately (from Gruen [7])

$$t_{\text{lock}} \approx \frac{35\Delta f2}{f_{3\text{dB}}^{3}} \qquad \text{(see also Table 8.3)}$$

where t_{lock} = lock time in seconds. This is the time required for the VCO to synchronize its frequency with the input signal.

Δf = difference in the VCO's open-loop frequency and the input signal frequency.

$f_{3\text{dB}}$ = closed-loop 3-dB bandwidth.

Hoffman [8] gives a relation between loop noise bandwidth, B_n, and natural resonance, ω_n:

$$B_n = \frac{4\zeta^2 + 1}{4\zeta}\omega_n \text{ Hz}, \qquad \text{where } K \gg \omega_n/\zeta$$

If $\zeta = 0.7$, $B_n = 1.06\omega_n$ Hz.

Hoffman [8] also shows that

$$f(3\text{ dB}) = 0.3088 B_n \text{ Hz} \qquad (\zeta = 0.7)$$

Therefore,

$$B_n = \frac{f(3\text{ dB})}{0.3088} = 3.238 f(3\text{ dB})$$

Expressing B_n in Hz, we have

$$B_n = 2\pi \times 1.06 = 6.658 f_n \text{ Hz}$$

Since $f(3\text{ dB}) = 0.3088B_n$ and $B_n = f(3\text{ dB})/0.3088 = 6.658f_n$,

$$f(3\text{ dB}) = 0.3088 \times 6.658 f_n = 2.056 f_n$$

This shows not only that the noise bandwidth of a phase lock is 3.238 times the loop 3-dB bandwidth but also that the relationship between the loop bandwidth and the natural resonant frequency is 2.056. This analysis also holds for Costas and squaring loops.

An FM-oriented phase lock loop, like the AM demodulating loop previously discussed, consists of three main parts: 1) a phase detector, 2) a loop filter, and 3) a VCO. (Missing from the FM demodulator are the second phase detector and low-pass filter needed for AM demodulation. [See Figure 8.9.])

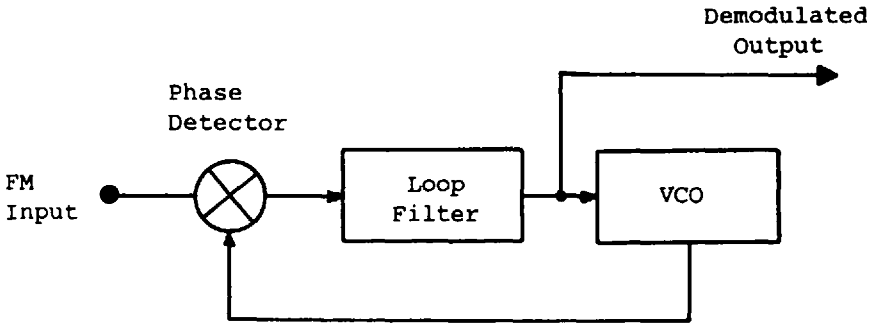


Figure 8.9 Phase lock loop FM demodulator block diagram.

The phase detector is usually an analog* detector such as a double-balanced mixer.

When an analog phase detector is in a loop under locked conditions, its inputs are the desired signal, usually from an IF amplifier output, and the VCO signal, fed back as a phase reference. These two signals, in the locked condition, are maintained by the loop at a 90 degree phase relationship with one another (plus or minus the phase error). Therefore we have $A \sin \omega_c t$ as one input and $B \cos(\omega_v \pm \phi_e)t$ as the other, where ϕ_e is zero when there is no phase error.

With no phase error, $A \sin \omega_c t$ and $B \cos \omega_v t$ are multiplied together in the phase detector to produce $(AB/2)[\cos(\omega_c + \omega_v)t - \cos(\omega_c - \omega_v)t]$, which is a sine wave at the input frequency, switched by 180 degrees every time the reference (VCO) signal changes polarity. Because the VCO (when locked) is at the same frequency and 90 degrees out of phase with the desired signal, it switches the desired signal phase twice in every cycle, at the 90 and 270 degree points, and this produces a signal that is symmetrical around zero, which integrates to zero volts in the loop filter. As phase error occurs, the phase detector output becomes unsymmetrical above or below zero, and this causes a DC shift in the output to the VCO, correcting its frequency. At one extreme then, where there is no frequency error, the phase detector (ideally) causes the output voltage to the VCO to be zero.

At the other extreme, when the input signal is either in phase or 180 degrees out of phase with the VCO, the VCO reference signal causes the input signal to be switched by 180 degrees at its zero and 180 degree points. This causes the phase detector to behave as a coherent full-wave rectifier, with the rectified peaks going either positive or negative, depending on the phase of the VCO. Because a full-

* Analog phase detectors, such as double-balanced mixers or Gilbert cells, are almost always employed as phase detectors in phase lock loops used as demodulators. (Digital phase detectors are often used in phase lock frequency synthesizers.) The reason for this is that it is necessary to have a digital input to a digital phase detector, but it is not always desirable to digitize the input to a phase lock demodulator.

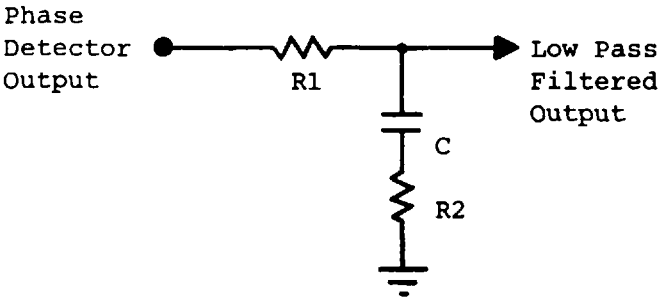


Figure 8.10 Typical second-order phase lock loop filter.

wave-rectified sine wave, when filtered, has a level of $0.63V_{peak}$, the maximum output range of the analog phase detector is $0.63V_{peak}$ minus zero, or simply $0.63V_{peak}$. This points out the reason why even phase lock loops are very sensitive to their input signal level. *The loop gain is a function of the input signal level, and the loop bandwidth is a function of the loop gain.*

Furthermore, it has just been shown that a phase lock loop with an analog phase detector and 0 dBm input will have detector sensitivity of only $(0.63 \times 0.314\text{V})/\pi/2$ radian = 0.126 V per radian of phase shift. This means that the loop's VCO will have to provide most of the required loop gain.

The loop filter is a lag-RC circuit made up of an RC lowpass filter with an added resistor (as shown in Figure 8.10) in series with the capacitor. This resistor is the "damping" resistor. This filter configuration causes the phase lock demodulator to be termed a "second-order" loop.*

The loop filter determines the loop bandwidth and response characteristics. Its time constants (with the loop closed) are

$$T_1 = (R_1 + R_2)C = \frac{K}{\omega_n^2}\dagger$$

$$T_2 = R_2C = \frac{2\zeta}{\omega_n}\dagger$$

From these equations, one can solve for the loop filter values. A straightforward phase lock loop design procedure may be outlined as follows, based on the two equations for the loop time constants:

---

*Phase lock loops used as demodulators are usually type 2, second-order loops. This means that the loop contains two integrators (type 2) and has two poles in its transfer function (second order). VCOs have a limited rate of change of frequency, and this is effectively a low-pass response, equivalent to integration.

†See Gruen [7].

1. Determine the open loop gain K

   $$K \geqslant 20\pi\,\Delta f$$

   assuming phase error is $\leqslant 0.1$ radian, where

   $$K = \underset{\text{(V/radian)}}{K_{\text{phase detector}}} \times \underset{\text{(radian/volt)}}{K_{\text{VCO}}}$$

   (Therefore K is dimensionless.)

   The phase error in a phase lock loop is such that

   $$\sin \Delta\phi = \frac{\Delta\omega}{K} + \frac{f_m}{(f_n)^2}$$

   The maximum modulating frequency f_m, however, for a type 2, second-order phase lock loop with damping factor 0.707 is larger than f_n by a factor of 2.06. This means that f_n^2 is always much larger than f_m, so that the $f_m/f_n{}^2$ term contributes very little to the overall phase error.

   Therefore $\sin\Delta\phi \approx \Delta\omega/K$ and $K_{min} \approx \Delta\omega/\sin\Delta\phi_{max}$. If $\Delta\phi_{max}$ is 0.1 radian, then

   $$K_{min} \approx \frac{\Delta\omega}{0.09} = \frac{2\pi\,\Delta f}{0.09} = 20\pi\,\Delta f$$

2. Define closed-loop bandwidth f_{3dB}, based on modulation bandwidth and/or lock-in time. In the case of an FM-demodulating loop, the loop bandwidth is determined by the modulation bandwidth, but for non-FM-demodulating loops, lock-in time is usually the criterion. Loop bandwidth as a function of lock-in time is

   $$f3\text{db} = \sqrt[3]{\frac{35\,\Delta f^2}{t_{lock}}}$$

3. Find $\omega_n = 2\pi f_n = 2\pi\, f_{3dB}/2.06 \approx \pi f_{3dB}$.
4. Determine ζ, the damping factor. For communications systems, ζ is usually chosen to be 0.707.
5. Choose a reasonable value for C. (A good starting point is 0.1 μF.)
6. Solve for $R_2 = 2\zeta/\omega_n C = 1.414/\omega_n C$.
7. Solve for $R_1 = K/\varphi_n{}^2 C - R_2$

Once the loop phase detector and VCO are chosen, which provides the values for K, and the loop filter is designed on the basis of the loop time constants, the only thing left is to assemble the components and test the loop's bandwidth.

It is often desirable to insert an amplifier into a phase lock loop between the loop filter and the VCO, where the VCO's sensitivity is reduced by the amount of the gain possessed by the amplifier. In this way, the VCO's sensitivity to noise on

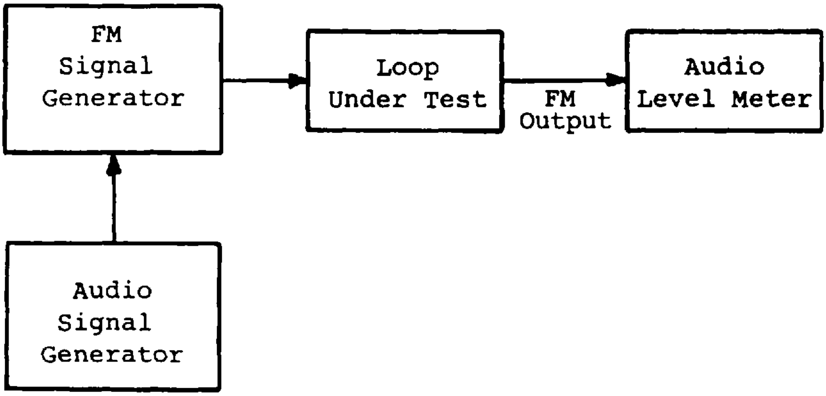


Figure 8.11 Test setup for phase lock loops, Costas loops, and other loops.

its power supply or other inputs is reduced, while an amplifier with good common mode rejection of noise on its power supply provides the needed loop gain and reduces overall loop sensitivity to noise from outside the loop.

A test setup and procedure for ensuring that a phase lock loop is designed properly are mandatory. As a phase lock loop is a feedback device, a small error in design can cause a large deviation in performance from that which is desired. Testing a phase lock loop must be accomplished by measuring its response to frequency-modulated input signals. Figure 8.11 shows a phase lock loop test setup.

The loop is tested by inputting a frequency-modulated carrier signal and measuring the demodulated output level as the modulating frequency is varied. This will result in output values that may be plotted as shown in Figure 8.12. (This plot of response versus frequency is a Bode plot.)

When evaluating the Bode plot, one makes use of the design parameters. That is, the 3-dB bandwidth (f_{3dB}) should be at the right frequency. If not, R_1* should be adjusted (higher to reduce f_{3dB} or lower to increase it). The natural resonant frequency (f_n) should be equal to $f_{3dB}/2.06$ and the resonant peak should be in the range of 0.5 to 1.0 dB above the low-frequency response. If this peak is too high, R_2 should be increased, and if it is too low then R_2 should be decreased. This procedure will result in a phase lock loop that has the proper bandwidth and transient response.

If transient response is a primary criterion, then it is advisable to input a signal (FM modulation is not required) that can be shifted abruptly from one

*In the lag-RC loop filter.

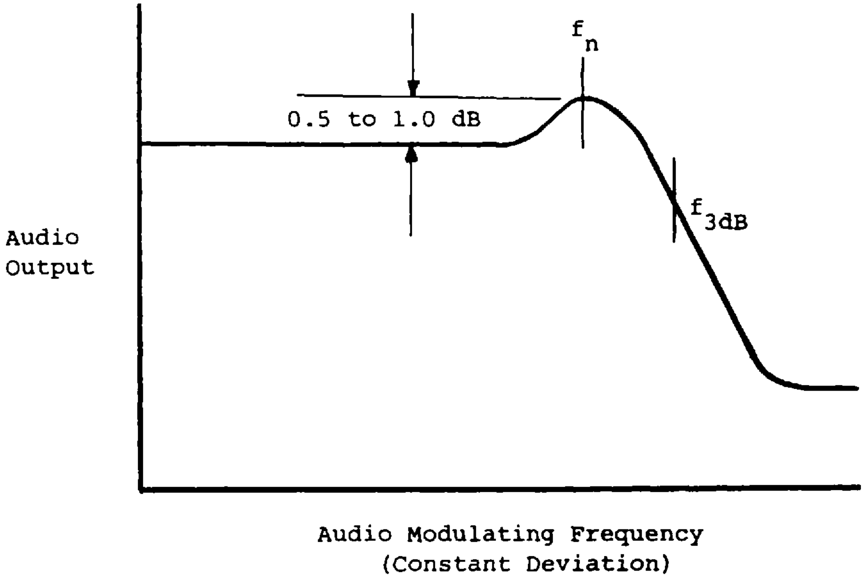


Figure 8.12 Plot of closed-loop frequency response of phase lock loop to input FM modulated signal.

frequency to another within the loop's lock range. Now, with the loop locked, shift the frequency and observe the signal input to the VCO. The amount of ringing on the VCO control input before settling to a steady state, after a step function change in input frequency, is a measure of the loop's damping factor. Varying R_2 allows adjustment of this damping factor. (Increasing R_2 reduces ringing; decreasing R_2 increases it but also reduces rise time.)

Table 8.3 Phase Lock Loop Characteristics (for Type 2, Second-Order Loop with Damping Factor 0.707)

Loop parameter	Characteristic
Lock-in time (sec)	$\dfrac{35\ \Delta f^2}{f_{3dB}^{\ 3}}$
Lock-in range (Hz)	$(2\zeta\omega_n K)^{1/2}$
Tracking range (Hz)	$2K$
ω_n (radians/sec)	πf_{3dB}
B_n (Hz)	$3.2B_L$

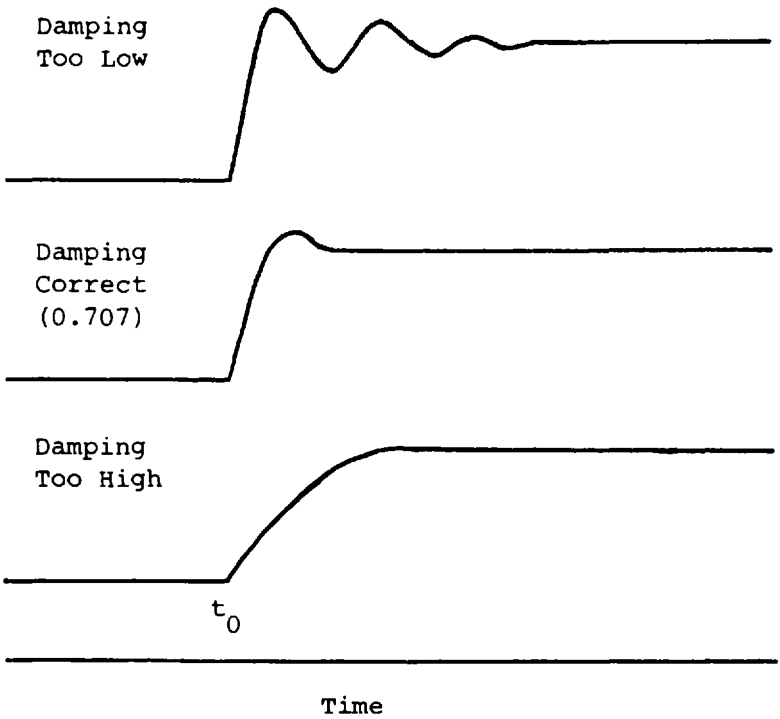


Figure 8.13 Typical phase lock loop response as a function of loop damping factor. (The same response can be expected from Costas and multiplier loops.)

Typical damping factor versus response is shown in Figure 8.13. Overall characteristics are taken from, or derived from, Gruen [7] and/or Hoffman [8].

8.3 "DIGITAL" DEMODULATORS

ASK: see AM demodulators; FSK: see FM demodulators and dual-filter demodulators.

8.3.1 Dual-Filter Demodulator

A dual-filter demodulator consists of two envelope detectors, each with an input filter tuned to one of two FSK input frequencies and with their outputs compared to determine which frequency has been used. This demodulator configuration is illustrated in Figure 8.14. A significant problem that exists in implementation of dual-filter demodulators is that of procuring filters that allow sufficient selectivity

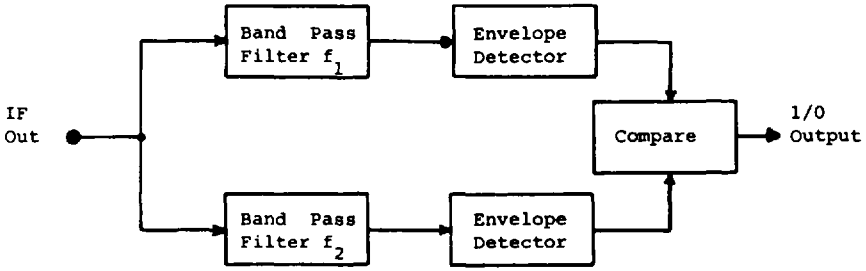


Figure 8.14 Dual-filter demodulator block diagram.

to discriminate between the two channels used and still provide enough bandwidth to allow a signal that falls in the filter bandwidth to rise quickly.

One solution to this problem is to employ a pair of local oscillators and frequency converters, as shown in Figure 8.15. In this case, where FSK is received at $f_{IF} + \Delta f$ to signify a one and $f_{IF} - \Delta f$ to signify a zero, let us suppose that one local oscillator using high-side injection and a second local oscillator using low-side injection are employed. The high-side LO drives one mixer, and the low-side

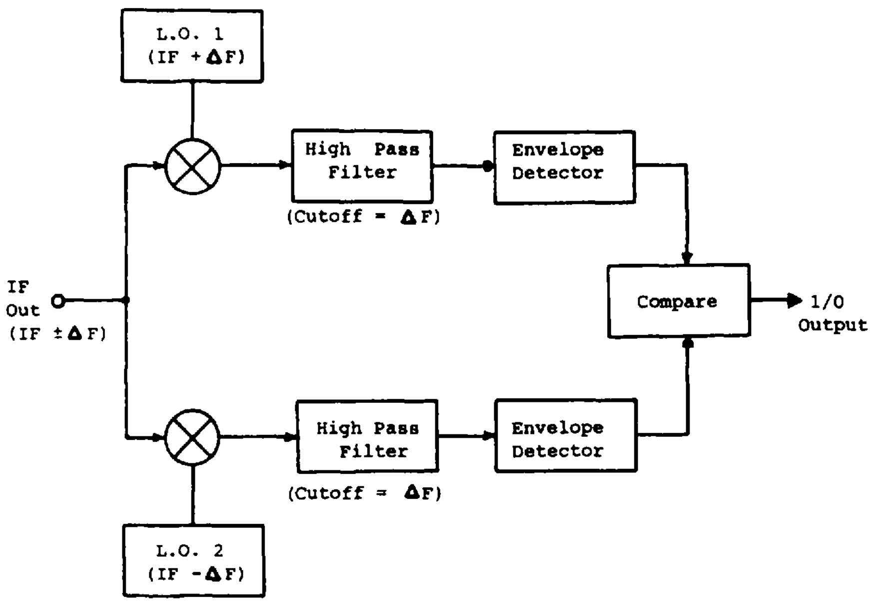


Figure 8.15 Dual-filter FSK demodulator using high-pass filters.

LO drives the other. The two mixer outputs separately connect to a pair of high-pass filters, set to roll off at a frequency near Δf.

When a one is sent and the input signal is at $f_{IF} + \Delta f$, the mixer employing high-side injection has an output that is near $\Delta f - \Delta f$ and the mixer employing low-side injection has an output that is near $2\Delta f$. (The mixer outputs are reversed if a zero is sent.) Therefore, if the high-pass filters have a cutoff frequency near Δf, one will have an output when a one is sent, and the other will have an output when a zero is sent, and neither will have a significant output when the opposite signal is intended. The advantage of this approach is that it employs high-pass filters that can have wide bandwidth, rather than depending on narrow-bandpass filters.

8.3.2 PSK

A Costas loop is a phase lock loop configuration that is intended for double-sideband, suppressed carrier demodulation (see Figure 8.16). It is most commonly used for PSK data signal recovery but may also be employed for AM, double-sideband AM, FSK, or frequency modulation. For this reason, Costas loops are sometimes called "universal" demodulators, as they can simultaneously recover more than one information stream as long as only one carrier is used.

The function of a Costas loop, like other coherent demodulators, is to regenerate a carrier and then employ that carrier to recover the information conveyed by a received signal. The Costas loop does this for a wide range of carrier modulation forms.

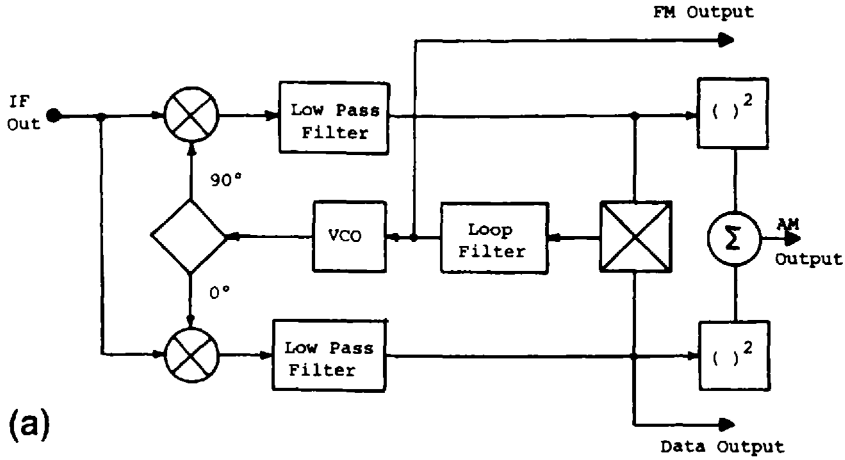
Costas loop design is similar to simple phase lock loop design. The main difference is that the open-loop gain (K) for the Costas loop is made up of three components instead of the two components of the phase lock loop. These components are

1. Phase detector sensitivity in volts per radian
2. VCO sensitivity in radians per volt
3. Third multiplier gain

Otherwise, Costas loop design and test procedures are the same as those for a simple phase lock loop. It should be noted that two different types of multipliers are shown in the Costas loop block diagrams. (See Figure 8.16.) Those that are represented by a circle with an X can be implemented with a common phase detector, such as a double-balanced mixer. Those that are represented by an X enclosed within a square require a DC* multiplier. The low-pass filters in the loop are intended to reject the carrier frequency but should have enough bandwidth to pass the data components without significant distortion.

---

*A multiplier whose response extends from zero Hertz (DC) through a frequency range greater than the information bandwidth.



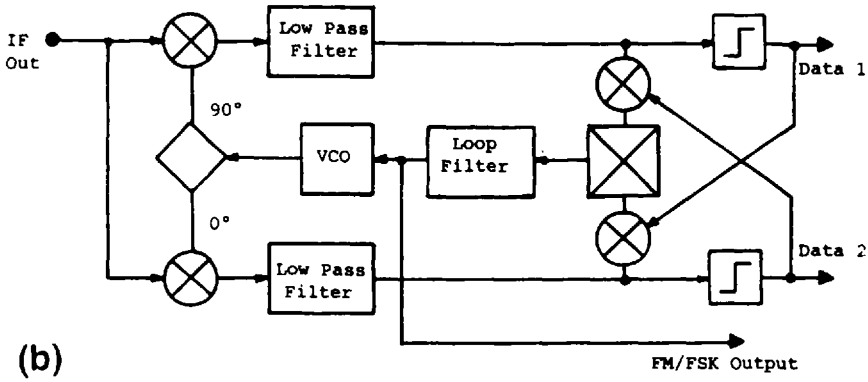


Figure 8.16 Costas demodulator block diagrams. (a) BPSK Costas loop "universal" demodulator. (b) QPSK Costas loop "universal" demodulator (AM output not shown).

When the Costas loop is synchronized (locked), its input signal is a PSK-modulated carrier $A \cos \omega_c t \pm 90°$. Its VCO, however, is locked at a 90 degree phase shift, so that the upper multiplier has inputs $A \cos \omega_c t \pm 90°$ and $B \sin \omega_c t$. When these are multiplied together, the result is

$$A \cos \omega_c t \pm 90° \times B \sin \omega_c t = \pm AB/2 \sin^2 \omega_c t \qquad \text{(quadrature terms)}$$

The lower multiplier's input signals are similar, but are in-phase terms, because the VCO is in phase with the input carrier. That is, the result from the lower multiplier is

$$A \cos \omega_c t \pm 90° \times B \cos \omega_c t = \pm AB/2 \cos^2 \omega_c t \qquad \text{(in-phase terms)}$$

These in-phase terms, after low-pass filtering, are the desired PSK data, recovered.

Figure 8.17 illustrates the waveforms seen in a Costas loop demodulator.

The QPSK Costas loop shown in Figure 8.16b is modified to provide independent data streams from the two independently modulated quadrature carriers of a QPSK data signal.

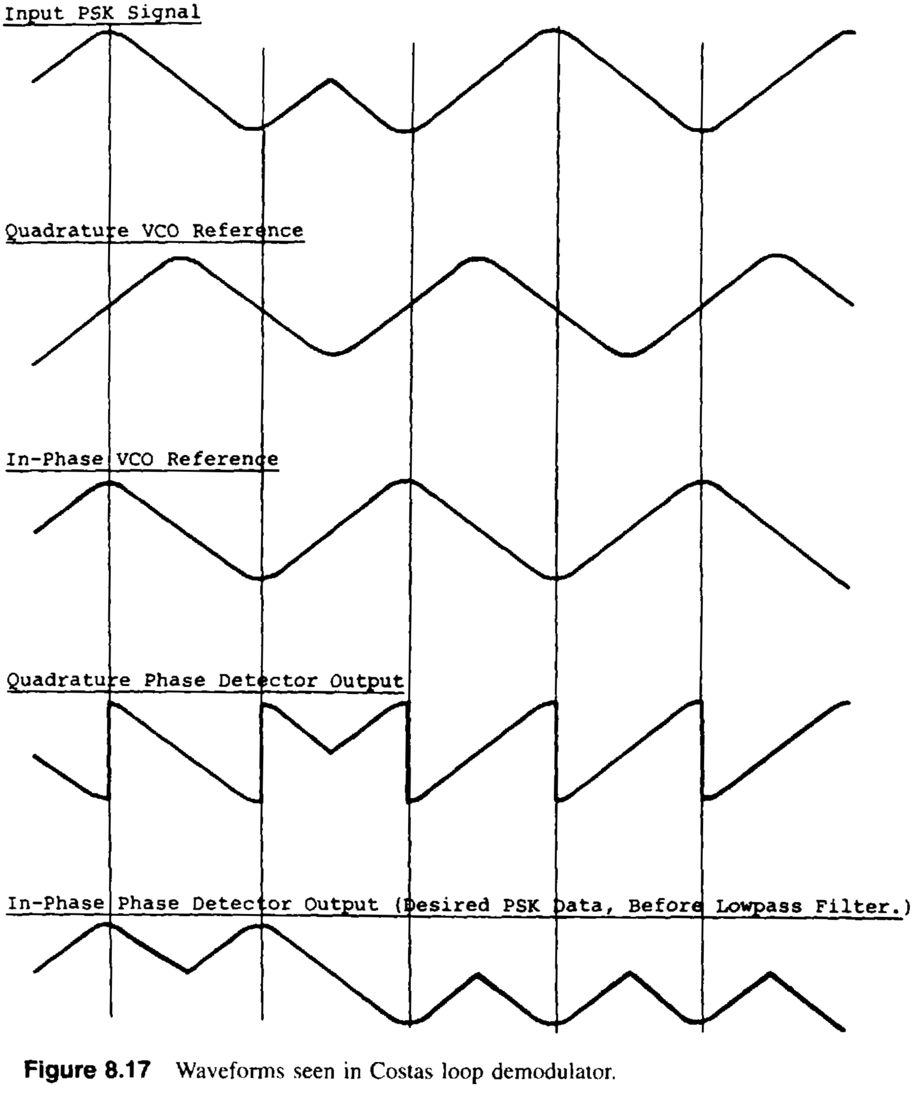


Figure 8.17 Waveforms seen in Costas loop demodulator.

When testing a Costas loop to determine its bandwidth and damping factor, which is very important to help in the reduction and/or prevention of false lock, a Costas loop is tested in the FM mode. Costas loops are sometimes called universal demodulators because of their ability to demodulate FM, AM, or PSK signals without the need for mode switching. Costas loops should be tested in the FM mode, just as a simple phase lock loop is tested. That is, a frequency-modulated input signal is used, with the modulating frequency being swept from a low frequency to a frequency well above the target loop bandwidth. The FM output is then plotted as a function of modulating frequency and examined as in the case of the simple phase lock loop, with adjustments being made to the loop filter components as necessary. (The filter for a Costas loop* is a lag-RC filter, just as in a phase lock loop.)

8.3.3 Squaring Loop

An alternative type of phase lock loop called a squaring or quadrupling loop is also designed to demodulate phase shift keyed signals. Squaring loops depend on the fact that squaring a double-sideband symmetrical signal produces a twice-frequency carrier without the data modulation. This carrier is then divided by two and used as a carrier reference for demodulating the PSK-modulated carrier. Squaring and quadrupling loops are illustrated in Figures 8.18 and 8.19. Both are forms of the more general "multiplier loop."

It should be noted that, although the squaring and quadrupling loop block diagrams clearly include squaring and quadrupling devices, they are not strictly required. In fact, a common frequency doubler such as the Minicircuits RK-3 works very well to produce a second-harmonic carrier from a BPSK-modulated signal. This type of frequency doubler is actually a transformer coupled full-wave rectifier identical in architecture to the full-wave rectifiers commonly used in power supplies. Such a device ignores 180 degree phase shifts such as those found in BPSK signals and simply puts out a full-wave rectified signal as if the phase shifts were not present. This is illustrated in Figure 8.20. For quadrupling loops, doubling the signal twice, to the fourth harmonic, is practical. In many cases, however, there is sufficient fourth-harmonic output from a frequency doubler to allow a phase lock loop to synchronize. In such a case, a frequency doubler could be used alone to meet the requirements of a quadrupling loop.

Multiplier loops may also make use of the approach used for AM demodulation with a phase lock loop, to develop a signal for use in AGC, signal presence, or bit synch/clock tracking applications. (See Section 8.1.)

When comparing Costas and squaring loops, it should be realized that the

---

*The loop filter is labeled $f_{(s)}$.

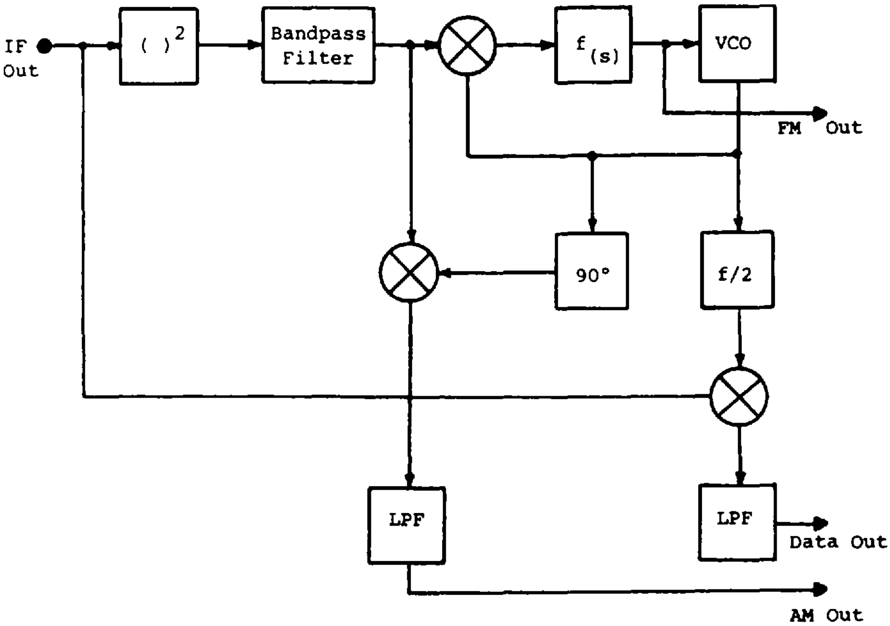


Figure 8.18 Squaring loop "universal" demodulator for BPSK data.

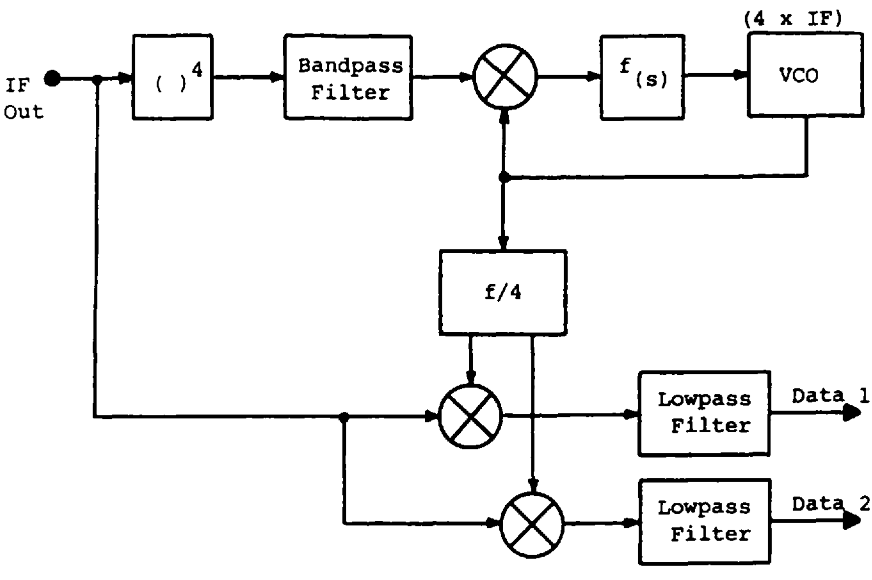


Figure 8.19 QPSK squaring loop or quadrupling loop (AM and FM outputs not shown).

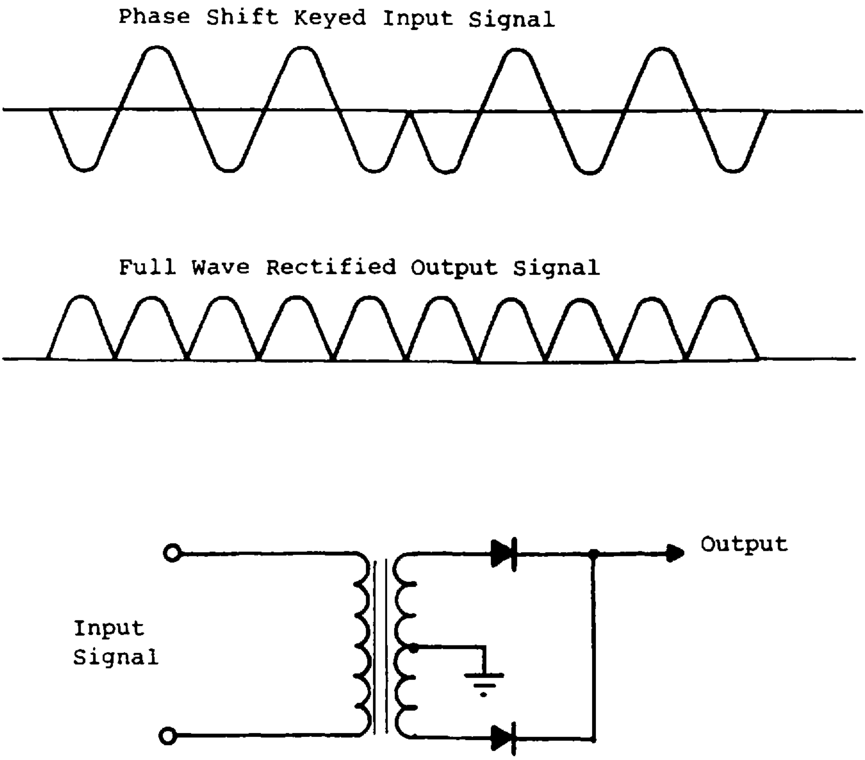


Figure 8.20 Full-wave frequency doubler for use in squaring loop or other applications.

two have identical threshold performance [1]. Most often, the choice between them is made on the basis of which is most available and with which of the two the user has had the most experience. It should be realized, however, that Costas loops have been employed more often in the past than squaring loops. In general, experience has led this author to believe that it is more difficult to develop a Costas loop than a squaring loop and that the primary drawback to squaring loops is that they work at twice their input frequency, which means a 100-MHz IF would force a squaring loop to work at 200 MHz. Another drawback with multiplier loops is that they do not readily adapt to demodulation of signals in which the increments of phase shift are not integer submultiples of 360 degrees. (The difficulty is that fractional dividers are not readily available for deriving phase references from a VCO.)

Squaring/multiplier loops are designed in the same way as simple phase lock demodulators. It is important to note, however, that the phase lock loop part of the demodulator must be designed for ΔF twice as great as that which exists at the demodulator input.

8.4 CARRIER RECOVERY

Recovery of a carrier from a received signal varies from unimportant to vital, depending on the specific type of modulation being received and the level of performance required. Certainly, AM and FM signals may be demodulated without "carrier recovery"; however, having a coherent carrier reference available can improve the demodulation performance of almost any receiver.

What do we mean by carrier recovery? In the context of this chapter, carrier recovery is intended to mean the process by which a replica of the original carrier signal, sent by the transmitter whose signal is of interest, may be derived in a receiver. The purpose of being able to derive, or replicate, the original carrier is to make it available for use in the demodulation process.

Examples of carrier recovery techniques are common. All phase lock demodulators employ carrier recovery. So do squaring demodulators and Costas demodulators. The fact that explicit carrier recovery is not vital, however, is borne out and confirmed by envelope detectors and product detectors, which do not require that a received carrier be separated out from the composite modulated signal. Even phase shift keyed signals may be demodulated without a carrier reference signal by employing a differential detection method, as shown in Figure 8.21. In this method, a delay line whose period is the same as the data period is used to provide both delayed and undelayed versions of a PSK-modulated signal.

When the undelayed and delayed versions of the signal are multiplied together, the output signal from the multiplier (after low-pass filtering) is a positive- or negative-going signal, depending on whether the delayed and undelayed multiplier inputs are in phase or out of phase with one another. Figure 8.21 shows the waveforms that are typical for such a demodulator. Note that no carrier recovery process is necessary for this device to work.

Why place emphasis on recovery of a carrier when there are obviously many examples of demodulators in which carrier recovery is not needed? The reason is that if the best possible performance is required of a receiver, then a coherent demodulator should always be used. Does this mean that better broadcast band AM and FM receivers could be built with coherent demodulators? Yes, it does. Why don't we do it then? Because the performance is good enough as it is. Broadcast stations are not meant to transmit over long distances,* and the amount of transmitter power employed is more than sufficient to cover a local area. This is

*High-power, "clear channel" stations do exist whose power is up to 50,000 W and that employ directional antenna arrays, but these are not the norm.

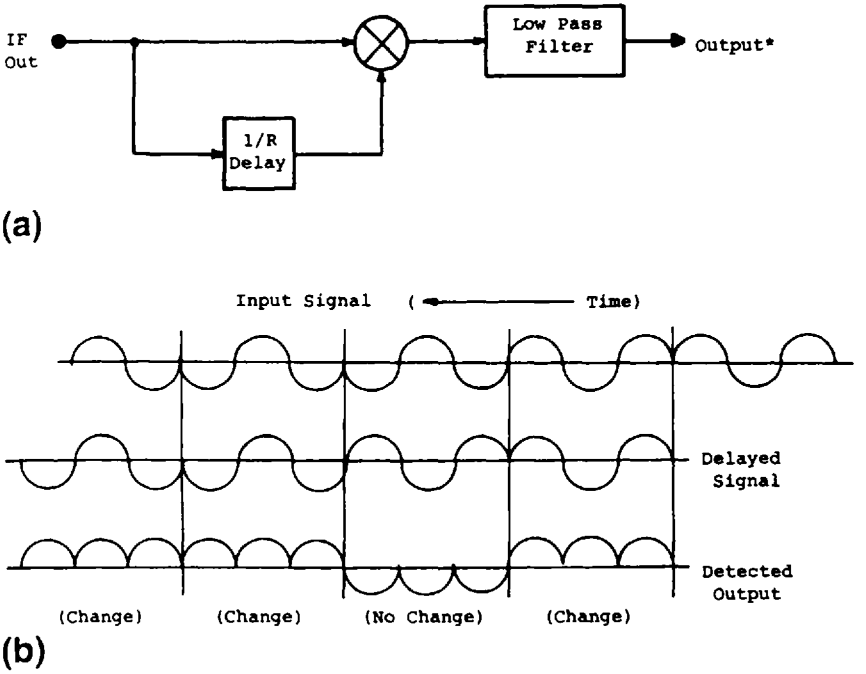


Figure 8.21 Differential demodulator for BPSK signals. (a) Differential demodulator block diagram. *Output corresponds to transitions (high = change). (b) Waveform in differential demodulator.

why many broadcast stations must reduce the power that they employ at sundown each night—so that they do not interfere with other stations at the same frequency in other areas.

In satellite and space applications, however, everything possible is done to improve receiver performance, including the use of coherent demodulators. Such receivers usually employ low-noise preamplifiers, high-gain antennas, and coherent demodulators to get the best possible performance from a power-limited satellite downlink.

In recent years, a move toward digital implementation in receivers has resulted in another class of noncoherent demodulation that makes up for some of the drawbacks of not having a coherent carrier to use in recovery of information. After all, what is it about a coherent carrier reference that is so advantageous? Figure 8.22 shows the reason why a coherent carrier reference is so desirable in a demodulator: a coherent carrier allows direct demodulation of the composite modulated signal using a carrier identical to that with which it was generated.

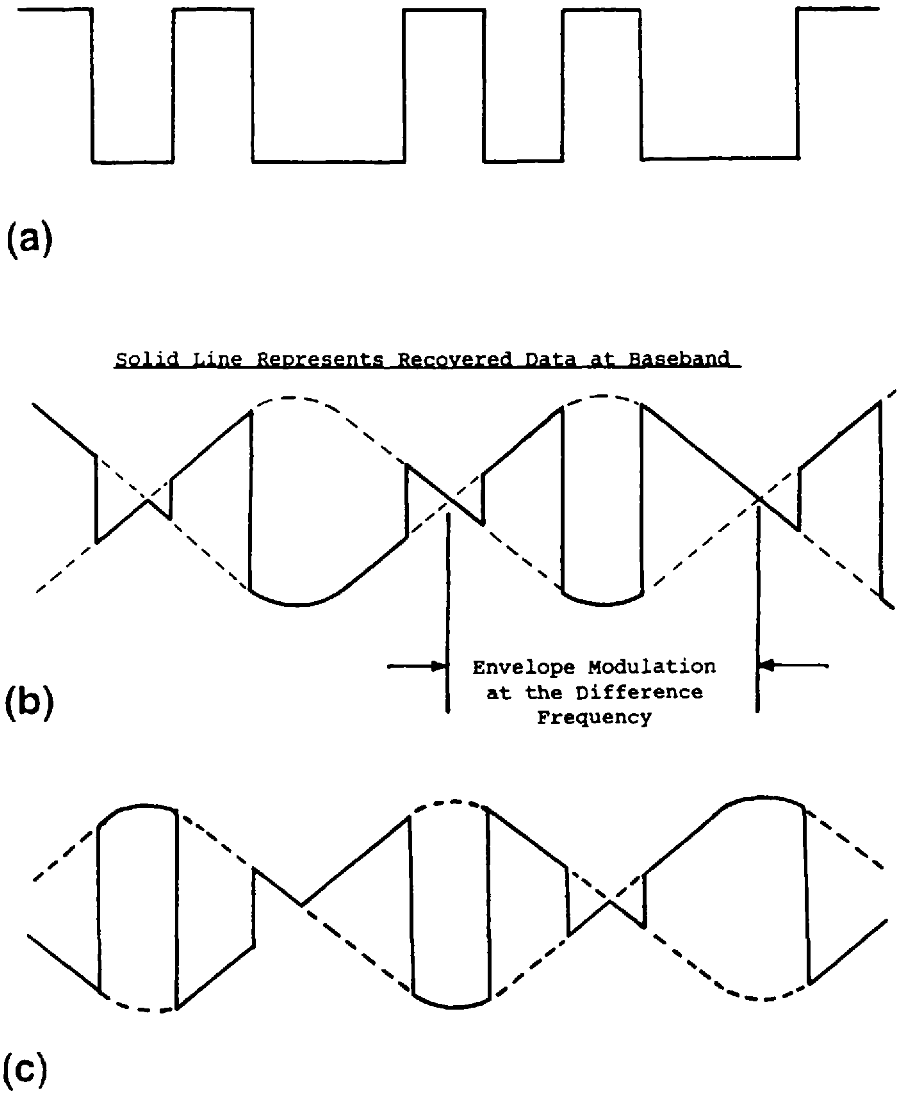


Figure 8.22 Direct-conversion signals as seen at baseband, due to frequency error between input carrier and local oscillator. (Difference frequency is the frequency error.) (a) Data as transmitted. (b) Q channel at baseband. (c) In-phase channel at baseband.

Coherence means that the carrier is identical in frequency and phase, but not necessarily in amplitude, with the original carrier.

The approach used in digital receivers is to employ either two carrier references that are in phase quadrature with one another or a single carrier reference and two quadrature versions of the modulated, received signal. This solves the basic problem that is created by down-converting directly to baseband from a higher frequency: any frequency error that exists between the incoming signal's carrier and the translating carrier (local oscillator) causes amplitude and phase modulation of the signal at baseband at the rate of the difference frequency (*i.e., there is modulation of the baseband sig- at the rate of the "zero beat" between the incoming and local carriers, and this may be very destructive to the demodulation process*).

The solution is provided by employing two signals generated by multiplying

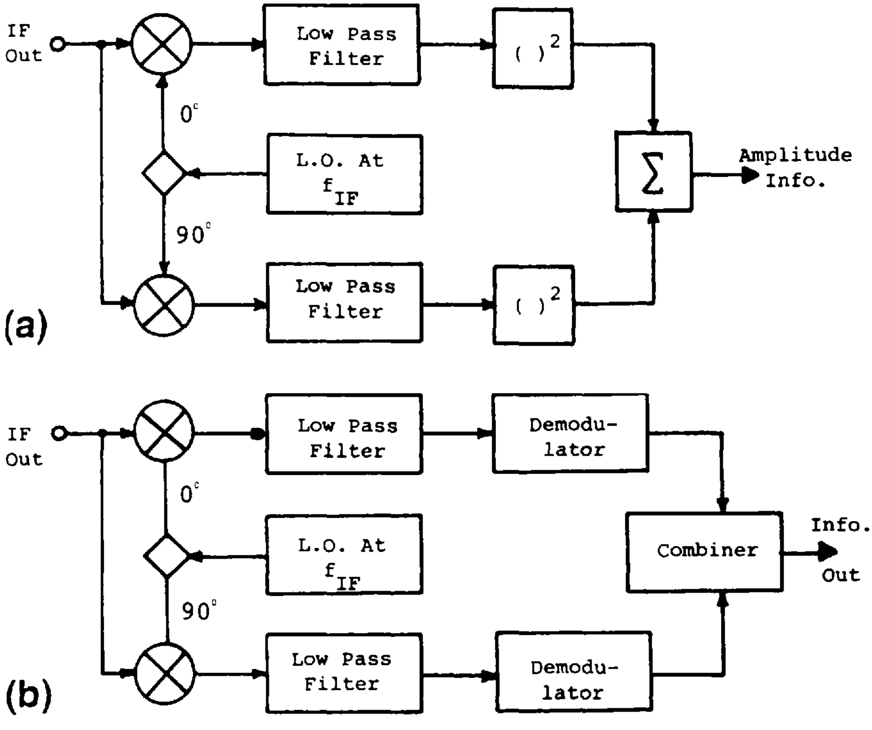
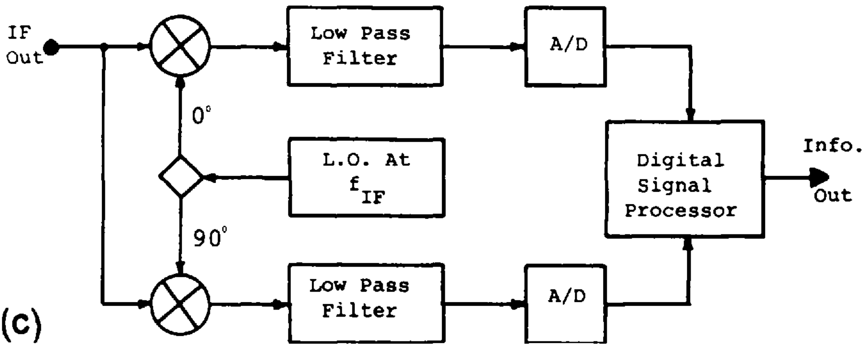


Figure 8.23 Demodulation methods for baseband I–Q signals. (a) I and Q baseband demodulator for amplitude information. (b) I and Q baseband demodulator using separate demodulators for the quadrature channels. (c) Digital baseband demodulation approach for I and Q signals. (d) Baseband correlation for I and Q applications, using digital correlator.

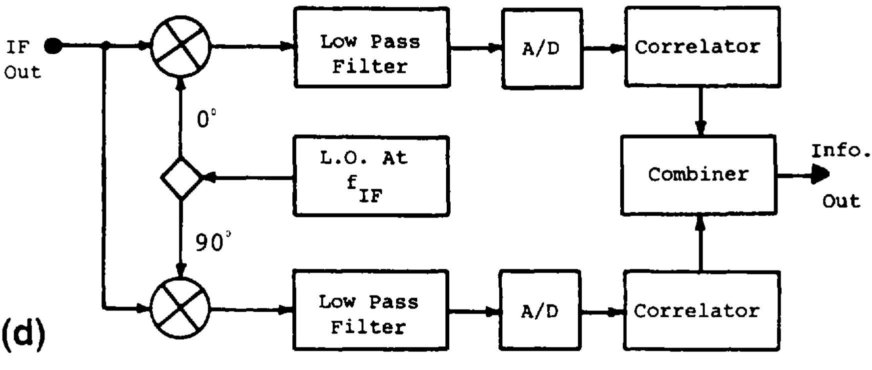
the received signal with separate, quadrature local oscillators in that the incidental modulation produced by a frequency difference in the local oscillators and the received signal is also in quadrature, which means that when one signal decreases, the other increases. By demodulating both, or by combining the two signals in some way, the signal can be recovered.

Figure 8.22 illustrates this problem. Depending on the type of modulation, various approaches are possible for employing the resulting signal, some of which are illustrated in Figure 8.23. Demodulators that make use of the the quadrature approach are termed quadrature or I-Q demodulators and may in fact be some form of Costas or multiplier loop.

It is emphasized that any type of signal modulation may be processed in this way, whether analog or "digital." Also, any analog process used may be approximated by digital processes, which we will examine further.



(c)



(d)

Figure 8.23 Continued

8.5 TRADEOFFS IN DEMODULATOR DESIGN

It is almost always possible to make tradeoffs between such things as threshold performance, bit error rate, output S/N ratio, cost, power dissipation, and information rate when designing a demodulator. For example, if the bandwidth of the demodulator is increased because of a change in required information, it may be necessary to raise the demodulator's input signal-to-noise ratio if the bit error rate is to be maintained.

A general rule that should always be considered is that the information rate in a system is a function of both the bandwidth used and the signal-to-noise ratio in that bandwidth. This is, of course a loose interpretation of Shannon's theorem, which states that the capacity of a system to send *error-free* information is

$$C = W \log_2 (1 + S/N)$$

expressed in bits per second. If we divide both sides of this equation by W (W is bandwidth in hertz) we have

$$C/W = \log_2 (1 + S/N)$$

Therefore,

$$2^{C/W} = 1 + S/N$$

and

$$S/N = 2^{C/W} - 1$$

But

$$S/N = \text{signal power/noise power}$$

$$= \frac{\text{energy per bit} \times \text{bit rate}}{\text{noise density} \times \text{bandwidth}} = \frac{E_b}{N_0} \times \frac{R}{W}$$

and we see that the relationship between signal-to-noise ratio and E_b/N_o is

$$\frac{E_b}{N_o} = \frac{S}{N} \times \frac{W}{R}, \qquad \text{where } R \text{ is considered to be at capacity, } C$$

Therefore, the relationship between E_b/N_o and S/N is the bandwidth per Hertz used by the modulation scheme.

Furthermore, if we consider that $E_b/N_o = W(2^{R/W} - 1)/R$ and we let the ratio W/R approach infinity, then E_b/N_o approaches a value of 0.69, which is the minimum value that is possible for error-free information, no matter how much bandwidth is used.

Expressed in dB, the smallest possible E_b/N_o with zero error rate is

$$10 \log 0.69 = -1.59 \text{ dB}$$

which is known as *Shannon's limit.*

Confusion often exists because it is not unusual to see curves of bit error probability as a function of signal-to-noise ratio, as well as identical curves of bit error probability as a function of E_b/N_o. How is this possible?

If it is assumed that the number of bits per second per hertz of bandwidth is 1.0, then the sets of curves are identical. Remember that

$$S/N = E_b/N_o \times R/W$$

Therefore, if $R = W$, then $S/N = E_b/N_o$.

When we compare different forms of modulation with respect to their output bit error probability, it is easy to see that a receiver's performance can readily be improved by changing the demodulator (if the transmitter will cooperate by also changing its modulation format). Figure 8.24 shows probability of bit error versus

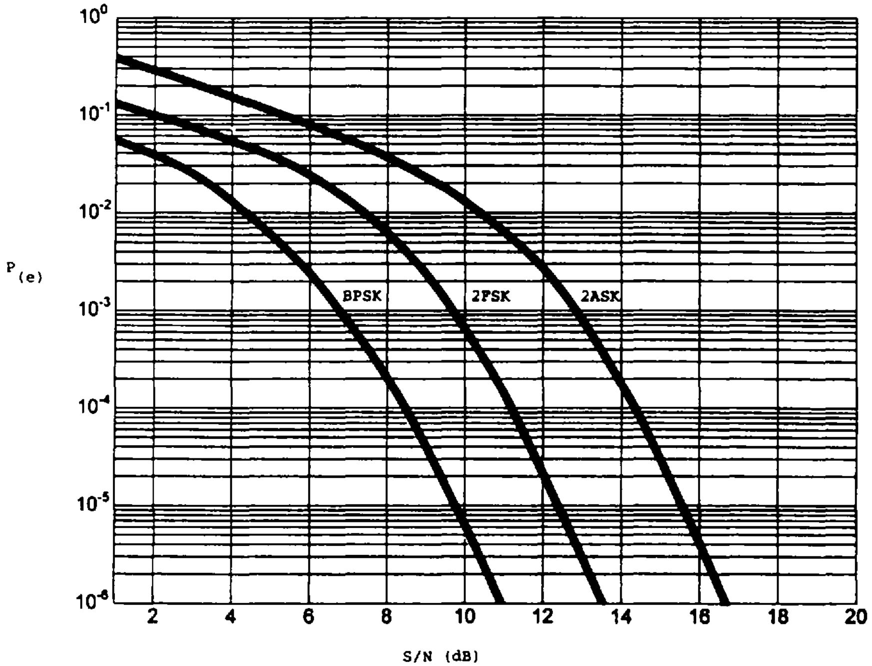


Figure 8.24 Probability of bit error versus signal-to-noise ratio for ASK, FSK, and PSK.

signal-to-noise ratio curves for phase shift key, frequency shift key, and amplitude shift key modulation with theoretically perfect demodulation. These curves also do not allow anything other than additive white Gaussian noise. Thus, it can be expected that the actual performance in a typical real receiver will be in the range of 1 to 3 dB worse than what is shown. (Often, losses are attributable to frequency error.)

Note that once the bit error rate reaches a probability of 1×10^{-2} or less, frequency shift keying is about 3 dB better than amplitude shift keying, and phase shift keying is about 3 dB better than frequency shift keying. Furthermore, they remain so over a wide range of bit error rates (beyond 1×10^{-8}). It should also be noted here that bit error rates lower than 1×10^{-3} are noteworthy in wireless systems, and bit error rates less than 1×10^{-6} are downright remarkable. At least this is true of mobile wireless systems. In fixed systems, with error correction coding, error rates in the vicinity of 1×10^{-5} or better are more common.

Table 8.4 lists the number of bits per second per hertz of bandwidth for various modulation schemes. The number of bits per second per hertz of bandwidth is also defined as the modulation efficiency of a system. Of those listed, we see that the range is 0.8 to 3.1 bits per second per hertz, a ratio of approximately 3.9 or 5.9 dB. Table 8.5 includes the E_b/N_0 requirement for the different modulation types, with a calculation of signal-to-noise ratio for comparison.

It is interesting to note that for the binary modulation schemes, where 1 bit per symbol is the general rule, S/N ratio is less than or equal to E_b/N_o (except for minimum shift key). Once more than 1 bit per symbol is sent, then S/N ratio becomes greater than E_b/N_o.

Table 8.4 Modulation Format and Number of Bits per Second per Hertz of Bandwidth[a]

Modulation	Bits per second per Hertz
On-off key (coherent detection)	0.8
FSK (noncoherent detection)	0.8
CPFSK $\beta = 0.7$ (noncoherent detection)	1.0
CPFSK $\beta = 0.5$	1.9
BPSK (coherent detection)	0.8
QPSK (coherent detection)	1.9
9PSK (coherent detection)	2.6
16PSK (coherent detection)	2.9
QAM (coherent detection)	1.7
QPR (coherent detection)	2.25
16APK (coherent detection)	3.1

[a]See Oetting [2].

Table 8.5 Comparison of E_b/N_o and S/N Ratio for Various Modulation Techniques (BER = 1×10^{-4})

Modulation	E_b/N_o (dB)	S/N (dB)	Difference (dB)
On-off key[a]	12.5	11.5	−1.0
FSK[b]	11.8	10.3	−1.5
CPFSK[b] $\beta = 0.7$	10.7	10.7	0.0
CPFSK[c] $\beta = 0.5$	9.4	12.2	2.8
BPSK[a]	9.4	8.4	−1.0
DPSK	10.6	9.6	−1.0
QPSK	9.9	12.7	2.8
DQPSK	11.8	14.4	2.6
8PSK[a]	12.8	16.9	3.1
16PSK[a]	17.2	21.8	4.6
QAM	9.5	11.8	2.3
QPR	11.7	15.2	3.5
16APK	13.4	18.3	4.9

[a]Coherent demodulation.
[b]Noncoherent demodulation.
[c]MSK.

How does this affect our choice of a demodulator, and how does the choice affect a receiver's performance? Unfortunately, we are seldom free to change the modulation in a specific application, but we may be able to change demodulators. The demodulator's required input S/N ratio (or E_b/N_o) is the minimum acceptable in the receiver's noise bandwidth, and the receiver's noise figure adds to the input noise. Thus, the minimum receiver input S/N ratio is

$$S/N = S/N \text{ (demod input)} + \text{receiver noise figure}$$

and the receiver's minimum input signal is

$$-174 \text{ dBm} + 10 \log B_n + \text{noise figure} + S/N \text{ (demod input)}$$

neglecting any losses or process gain.

To improve a receiver, the choices that we have are

1. Reduce its noise bandwidth, which usually entails a change in the modulation format.
2. Reduce the receiver noise figure.
3. Reduce the demodulator's required input S/N ratio.
4. Reduce the demodulator's losses.

Note that only a reduction in the receiver's noise figure does not involve the demodulator. This emphasizes the fact that the rest of a receiver exists for the purpose of presenting its demodulator with a usable signal from which it can recover the intended information.

8.6 DEMODULATOR PERFORMANCE CRITERIA

Criteria for judging the performance of a demodulator are dependent on its application. Demodulators for data reception are expected to provide a maximum bit error rate under specified input signal and noise conditions. The maximum bit error rate usually allowable is in the 1×10^{-3} to 1×10^{-6} range, where voice applications may find error rates as high as 1×10^{-2} acceptable, but computer-to-computer applications may struggle with any error rate greater than 1×10^{-6}.

The actual demodulator error rate may in many cases be improved by employing an error detection and/or correction scheme that operates in series with the demodulator. That is, where bit error rate is critical (and it always seems to be critical) error correction coding can be employed, which entails coding at the transmitter and decoding after the demodulator. It must be understood, however, that the use of error detection coding implies that the demodulator must operate at a higher data rate than if such coding is not used. In many cases, the demodulator's rate must be higher by two to three times the actual information rate. This in turn implies that the bandwidth in which the demodulator must operate is increased by a like amount, and the overall receiver sensitivity must be degraded because of the need to have a bigger received signal to overcome the increased noise in the demodulator's widened bandwidth.

The message here is that addition of an error correction code to a system does not always automatically improve performance. Only if the coding gain is greater than the demodulator degradation, due to its wider noise bandwidth, is there a system improvement when adding error correction coding.

Demodulators for voice applications do not employ bit error rate as a criterion for performance (unless the voice is digitized). Instead, they employ signal-to-noise ratio at the demodulator's output. Tests made over the course of many years have led to the use of a 10-dB signal-plus-noise-to-noise criterion for many receivers. The reason for this is that the average listener has been found to be able to understand approximately 90 percent of what is said when the receiver to which he (or she) is listening has an output $(S + N)/N$ ratio of 10 dB. Tests to confirm such intelligibility are done with standardized tapes under controlled conditions, to ensure that a receiver meets the minimum output specification.

An alternative requirement sometimes used is a SINAD measurement, in which the output is required to meet a 12-dB signal-plus-noise-plus-distortion-to-noise criterion.

In many cases, the demodulator's output is not measured directly when a receiver's performance is tested, because there is often an output driver or other interface device between the demodulator and the actual output. In such cases, the output device's contribution is considered to be negligible, except for SINAD measurements, where distortion contributed by an output amplifier would be considered a part of the output ratio.

Some demodulators also have a form of "process gain" wherein there is an improvement in the signal-to-noise ratio between the demodulator input and its output. One example of such a demodulator is any form of FM demodulator, including discriminators, ratio detectors, quadrature detectors, and phase lock loops. In FM systems, there is a gain in S/N ratio that is a function of the FM deviation ratio (see Chapter 2) that occurs in the demodulator. Another type of demodulator that exhibits process gain is a pulse-position-modulation demodulator, whose process gain is a function of the ratio of the time difference in pulse positions and the rise time of the pulses themselves.

Process gain in a PPM receiver is the ratio of the pulse rise time to the pulse time position difference. That is,

$$G_{\mathrm{p/ppm}} = \frac{\Delta t}{t_{\mathrm{r}}}$$

(see Black [13]). Spread spectrum receivers also exhibit process gain, although the process gain is not a function of the demodulator (as it is in an FM receiver, where the process gain is β^2). In spread spectrum receivers, process gain is (for direct sequence or frequency hopping)

$$G_{\mathrm{p/ss}} = \frac{\mathrm{BW}_{\mathrm{RF}}}{R_{\mathrm{data}}}$$

(see appendix II and Dixon [14]) and

$$G_{\mathrm{p/chirp}} = t_{\mathrm{pulse}}/2\Delta_f \qquad \text{for chirp systems.*}$$

In these
Δt = time shift in a PPM signal pulse
t_{r} = rise time of the PPM pulse
$\mathrm{BW}_{\mathrm{RF}}$ = spread spectrum bandwidth
R_{data} = data rate sent in the spread bandwidth
t_{pulse} = recovered pulse width
Δ_f = swept bandwidth
β = FM deviation ratio

In effect, the despreading process used in a spread spectrum receiver is a demodulation process, but it is not the data demodulation process.

*Chirp is also a form of spread spectrum modulation.

8.7 DEMODULATOR COMPARISONS

It is difficult to draw direct comparisons between dissimilar types of demodulators for all applications, and some comparisons may be either irrelevent or unfair, but it is certainly practical to compare demodulators on a general basis. That is, comparison of an envelope detector with a quadrature detector for use in FM demodulation would be a nearly useless exercise, but comparing their general abilities and requirement is useful. Table 8.6 is a summary of various demodulators and their typical performance capabilities, for comparison.

Envelope detectors require an input signal-to-noise ratio of approximately 10 dB to provide the 10-dB $(S + N)/N$ output usually specified. At reduced input signal-to-noise ratios, an envelope detector typically has lower output signal-to-noise ratio than that which exists at its input, while the detector signal-to-noise ratio input and output approach a unity relationship at about a 10-dB input ratio. Product detectors, which are employed to demodulate single-sideband as well as conventional AM signals, must also output signal-to-noise ratios of 10 dB or more (with single-sideband signals, 12 dB is often required), and this brings about the need for a similar input signal-to-noise ratio. Because no process gain is provided by AM or SSB AM signaling, there can only be a loss, and the loss approaches zero when the detector's input signal-to-noise ratio approaches 10 dB or higher.

Phase lock loops can also be employed as AM detectors, but their performance is no different from that of a product detector; even though the phase lock loop itself can acquire the carrier at input signal-to-noise ratios as low as 3 dB, it still does no more than act as a carrier source for multiplication with the input signal, and that is exactly the process by which the product detector works.

FM demodulators, such as Foster-Seeley discriminators, ratio detectors, quadrature detectors, and phase lock loops, all have the advantage that they can

Table 8.6 Comparison of Demodulators and Their Capabilities

Demodulator	Intended modulation	S/N input	S/N out
Envelope detector	AM	10 dB	10 dB
Product detector	AM	10–12 dB	10–12 dB
Phase lock loop	AM	10 dB	10 dB
Foster-Seeley discriminator	FM	6–8 dB	20–22 dB
Ratio detector	FM	6–8 dB	20–22 dB
Quadrature detector	FM	6–8 dB	20–22 dB
Phase lock loop	FM	6–8 dB	20–22 dB
Costas loop	PSK	6.8 dB	BER 10^{-3}
Squaring loop	PSK	6.8 dB	BER 10^{-3}

provide process gain that is a function of the deviation ratio being employed. For broadcast band signals, for example, the deviation ratio employed is 5.0 (the modulation bandwidth is 15 kHz, and the peak deviation is 75 kHz) and this produces a process gain of (75 kHz/15 kHz)^2 or 25. At lower modulating frequencies, preemphasis is used that maintains the deviation ratio, and in the receiver deemphasis is employed after demodulation.

A threshold input signal-to-noise ratio of 6 to 8 dB must be reached before the process gain is effective, so this means that an FM demodulator can be expected to output a signal-to-noise ratio (when the deviation ratio is 5.0) that is equal to its input signal-to-noise ratio plus the deviation ratio (β) squared. Thus, in dB, an FM demodulator's output signal-to-noise ratio would be

$$S/N_{\text{output}} = S/N_{\text{input}} + 10 \log \beta^2$$

For broadcast applications, this is approximately

$$S/N_{\text{output}} = 6 \text{ to } 8 \text{ dB} + 10 \log 25 = 20 \text{ to } 22 \text{ dB}$$

Although the phase lock loop can acquire and track signals below a 6- to 8-dB signal-to-noise ratio, the threshold for FM gain prevents practical employment at such levels. For this reason, phase lock demodulators are not often used in consumer FM receivers.

Costas and squaring loops have limitations that are similar to those of phase lock loops, but their threshold is 3 dB higher, and they are seldom employed for FM detection, so FM gain usually does not apply.

8.8 DEMODULATORS FOR QAM AND OTHER EXOTIC MODULATION

In the previous section it was shown that when m-ary phase shift key modulation is employed and the number of phase shift positions is large, other forms of modulation may actually outperform PSK, which usually outperforms both ASK and FSK when m is small. At $m = 16$ or greater, the phase shifts in PSK modulation become small enough that it becomes advantageous to combine amplitude modulation with phase shift keying, which actually improves performance. Three of the techniques that outperform 16PSK are (from the standpoint of E_b/N_0 required) QAM (quadrature amplitude modulation), QPR (quadrature partial response), and 16APK (16-ary amplitude-phase keying). The typical performance for 16PSK compared with these other forms of combined amplitude and phase modulation is shown in Table 8.7, when operating at a bit error rate of 1×10^{-4}. Figures 8.25 through 8.28 are diagrams of the phase/amplitude positions for 16PSK, 16QAM, 16QPR, and 16APK signals for comparison purposes.

Because all except 16PSK are combinations of AM and PSK, it is necessary

Table 8.7 Comparison of E_b/N_o Required to Produce a Bit Error Rate of 1×10^{-4} for Several 16-ary Modulation Formats

Modulation	E_b/N_o (dB)	Modulation	E_b/N_o (dB)
16PSK	17.2	16QPR	11.7
16QAM	9.5	16APK	13.4

for demodulators for these signals to be capable of operating with both kinds of modulation simultaneously. Figures 8.29 and 8.30 illustrate forms of I–Q demodulators that may be employed.

8.8.1 16PSK Demodulator

Figure 8.25 illustrates a 16PSK demodulator based on a multiplier loop. As in many PSK demodulators, a phase detector or multiplier is employed for each pair

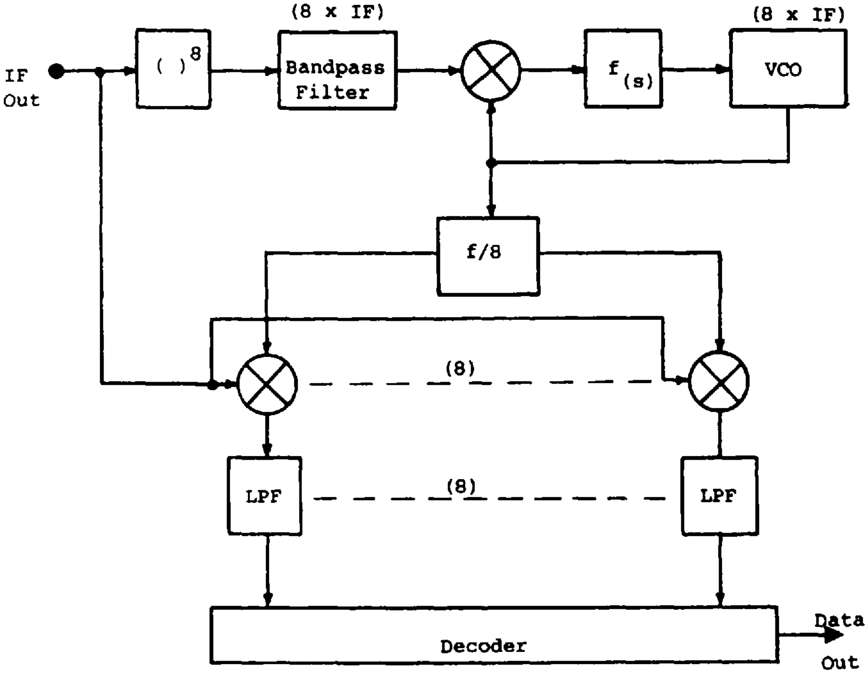


Figure 8.25 16PSK demodulator employing multiplier loop for carrier recovery.

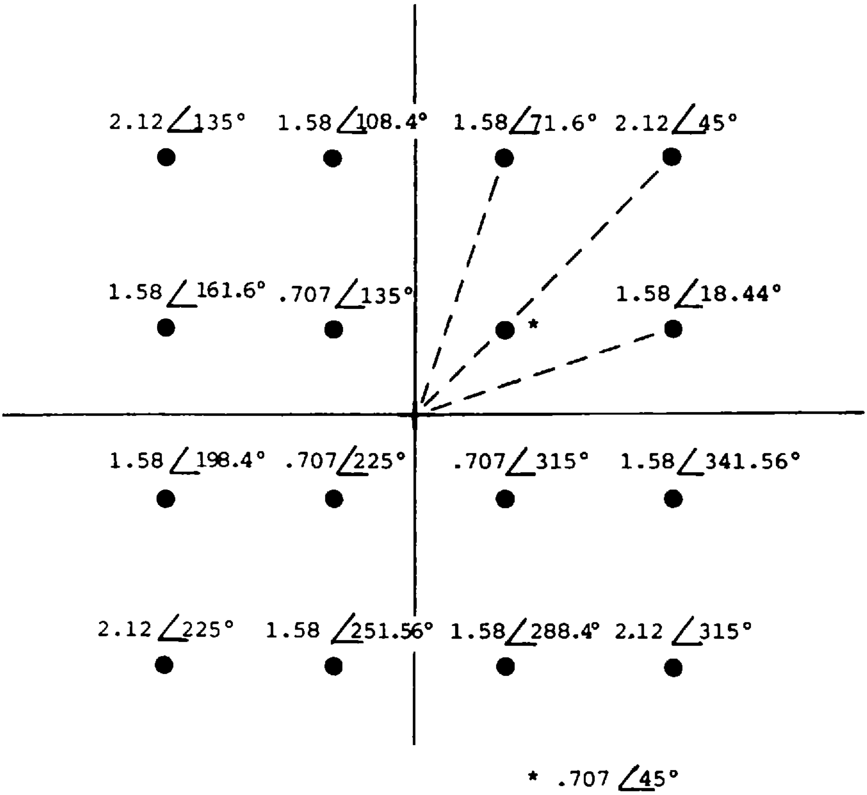


Figure 8.26 Phase/amplitude diagram for 16QAM signal, showing four positions per quadrant and all phase/amplitude positions. (Amplitudes are normalized to distance from position to position in vertical and horizontal planes.)

of phases. The entire demodulator consists of only a multiplier, a carrier recovery loop, a divider, and the multipliers used for signal recovery (the decoder is not considered to be a part of the demodulator itself). The times-eight multiplier produces a carrier to which the phase lock loop synchronizes. The divide-by-eight counter then provides eight outputs, in 45 degree shifted increments, at the input frequency. These eight carrier references then act as the phase references for demodulating the 16-phase input signal.

The decoder's task is to select the phase detector output that is the largest and translate that information as one of 16 four-bit data words. It must also, as part of the process, determine which is the largest output from the eight phase detectors

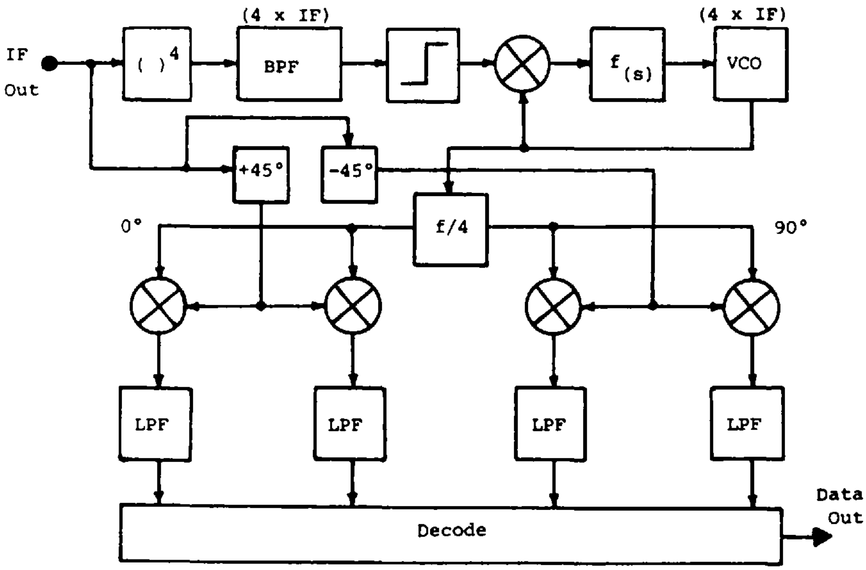


Figure 8.27 16QAM demodulator with multiplier (quadrupling) loop for carrier recovery.

and then whether that output is plus or minus. Three bits are determined by the phase detector and the fourth bit by whether its output is plus or minus.

8.8.2 Quadrature Amplitude Modulation

QAM signals have the characteristic phase/amplitude diagram shown in Figure 8.26 and the distribution of phases and amplitudes listed there. Please note that both the amplitudes and phases are irregular; that is, they are not evenly spaced and are not integer submultiples in either amplitude or phase (although some are).

From the listings in Figure 8.26, we see that the minimum phase shift is 26.565 degrees. This is greater than the smallest phase shift used in 16PSK by 4.065 degrees, which reduces the receiver's sensitivity to phase noise. Figure 8.26 also shows that the smallest amplitude shift is 25.5 percent of the peak amplitude. This, of course, is worse than 16PSK, for which all phase shifts have the same amplitude. Overall, however, QAM outperforms 16PSK.

The QAM demodulator shown in Figure 8.27 is based on the multiplier loop shown in Section 8.3. Note that only times four multiplication is employed, even though a 12-phase modulation scheme is being used. This is made possible by

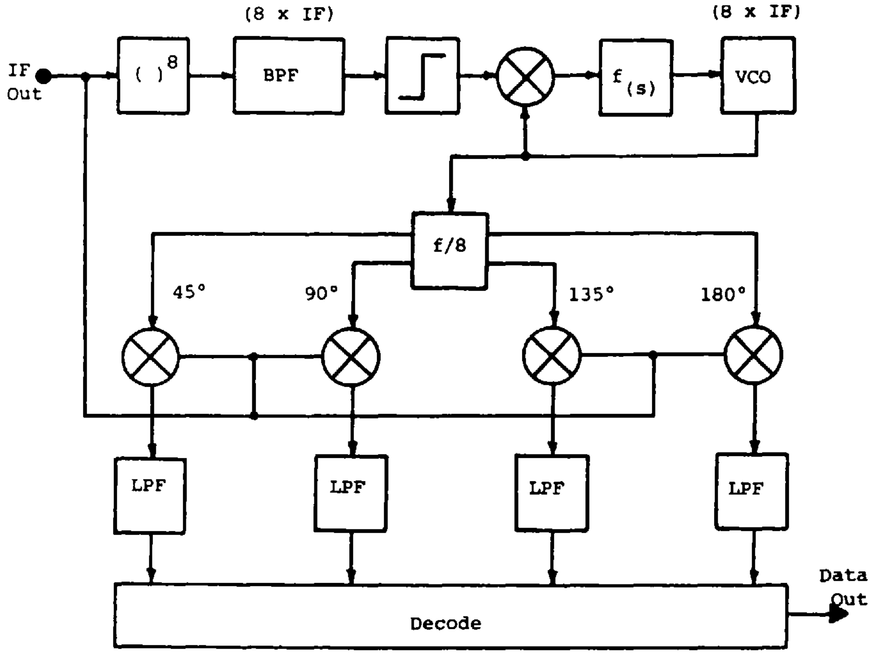


Figure 8.28 9-ary QPR demodulator using multiplier loop. (Filtered 16-ary QAM.)

phase shifting the input signal as shown. The decoder in this demodulator must decode on the basis of both amplitude and phase.

8.8.3 Quadrature Partial Response Demodulator

QPR signaling is employed for two purposes. One is to make it possible to provide a fixed equalizer whose response is the complex conjugate of a transmit filter that does far more to the signal than is possible by the propagation link. (Therefore the link contributes very little unknown distortion compared with what has already been done.) The second reason is to reduce the bandwidth of a QAM signal.

When a QAM signal is filtered to produce QPR, the number of phase/amplitude positions usable decreases. That is, 16QAM becomes 9QPR, and 64QAM becomes 49QPR, where 9QPR employs 8 phases and 3 amplitudes and 49QPR employs 32 phases and 37 amplitudes. Neither of the two demodulators can operate at the same level after filtering as before, however, because having 9 states allows only 3 bits per symbol, and 49 states allows only 5 bits per symbol.

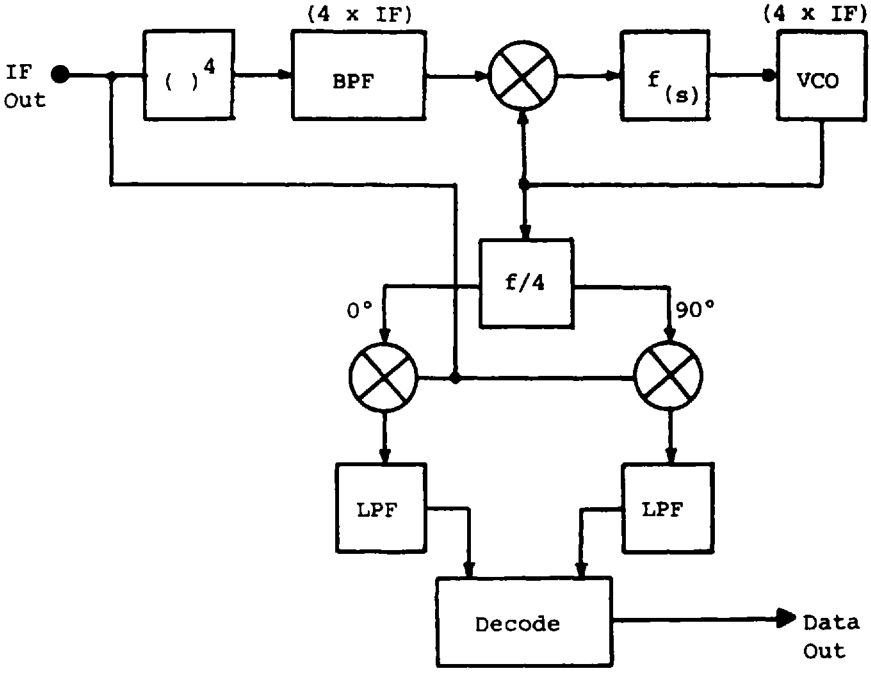


Figure 8.29 16-ary APK demodulator using multiplier loop for carrier recovery.

Figure 8.28 shows an implementation of a 9QPR demodulator, which employs a times-eight multiplier loop for carrier recovery but four phase detectors to determine which phase (or its complement) is being used at any given symbol time. As in the 16PSK demodulator, part of the recovered word is determined by which phase detector's output is highest, and an additional bit is determined by whether the output is plus or minus.

8.8.4 16-ary Amplitude/Phase Demodulator

16-ary APK signals employ only four phases, but they employ four amplitudes for each phase. A quadriphase demodulator based on either a Costas loop or a multiplier loop can be used to recover a carrier from such signals. Two bits of each 4-bit symbol are derived from the phase, and the other 2 bits are derived from the level of the signal at the specific phase.

Figure 8.29 is a block diagram of a 16-ary APK demodulator of the multi-

National Semiconductor

LM380 Audio Power Amplifier

General Description

The LM380 is a power audio amplifier for consumer application. In order to hold system cost to a minimum, gain is internally fixed at 34 dB. A unique input stage allows inputs to be ground referenced. The output is automatically self centering to one half the supply voltage.

The output is short circuit proof with internal thermal limiting. The package outline is standard dual-in-line. A copper lead frame is used with the center three pins on either side comprising a heat sink. This makes the device easy to use in standard p-c layout.

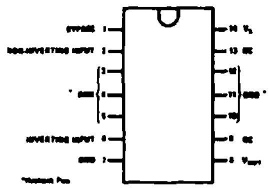
Uses include simple phonograph amplifiers, intercoms, line drivers, teaching machine outputs, alarms, ultrasonic drivers, TV sound systems, AM-FM radio, small servo drivers, power converters, etc.

A selected part for more power on higher supply voltages is available as the LM384. For more information see AN-69.

Features

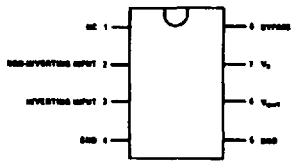
- Wide supply voltage range
- Low quiescent power drain
- Voltage gain fixed at 50
- High peak current capability
- Input referenced to GND
- High input impedance
- Low distortion
- Quiescent output voltage is at one-half of the supply voltage
- Standard dual-in-line package

Connection Diagrams (Dual-In-Line Packages, Top View)



TL/H/6977-1

Order Number LM380N
See NS Package Number N14A

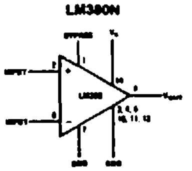


TL/H/6977-2

Order Number LM380N-8
See NS Package Number N08E

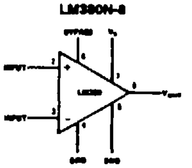
Block and Schematic Diagrams

LM380N

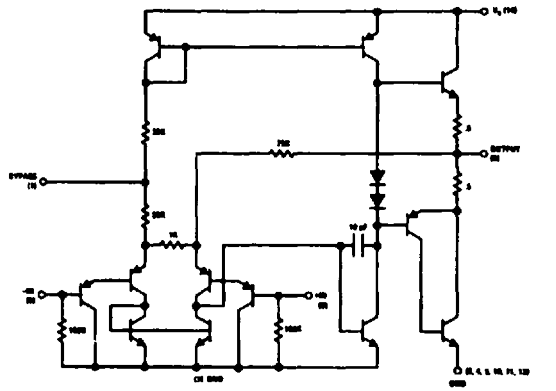


TL/H/6977-3

LM380N-8



TL/H/6977-4



TL/H/6977-5

Figure 8.30 Low-power drain integrated circuit power amplifier useful for portable receiver applications. (Courtesy of National Semiconductor.)

plier loop variety. It should be noted that it is identical* to any other QPSK multiplier loop (usually called a quadrupling loop) with the exception of the decode block, which is not necessary on a simple quadriphase demodulator. Please note that adding the decoder to decipher the amplitude information doubles the data rate when compared with simple QPSK, but the phase shift rate has not changed. In other words, the data rate has doubled without significantly increasing the bandwidth required.

8.9 DEMODULATOR AND RECEIVER OUTPUT

If receivers had no output, there would be little reason for their existence. Receivers exist because their owners want to make use of information that is available (free of cost) within the range of frequencies for which it is practical to build transmission and reception equipment and systems. The information, of course, may be anything from rock and roll music to messages of vital national importance. It can also be data, voice, or video. As the type of information and its destination must vary greatly, so must the criteria used to judge the quality of the output vary greatly. Here, we will consider the various forms of output and the measures used to judge them.

8.9.1 Criteria for Receiver Output

Receivers, from the standpoint of their output signals, may be sorted according to their use. Some of the common uses and outputs are found in Table 8.8.

The criteria used for judging and comparing the various outputs are functions of the fidelity with which information can be reproduced by a receiver and the conditions under which a receiver can reproduce its output information.

For receivers used in voice and music (or similar applications) the criteria usually applied are signal-to-noise ratio and distortion, together with flatness of response. In stereo receivers, for example, specifications would typically be similar to those in Table 8.9. These, of course, are the specifications that have to do primarily with the output signal from a stereo receiver. Each specification would apply to the two outputs separately.

Receivers intended for digital data do not usually specify output quality in terms of signal-to-noise ratio, although it is practical to do so. Instead, data output may be given in terms of E_b/N_0 or bit error rate. Although E_b/N_0 is not directly measured and is calculated from other parameters such as signal-to-noise ratio or bit error rate, demodulator performance is often expressed in terms of E_b/N_0,

*Except for the limiter that precedes the phase lock loop.

Table 8.8 Receiver Classification According to Application

Receiver use	Output signal
AM reception	Voice, music, telegraphy
FM reception	Voice, music, telemetry
ASK reception	Data, including digitized analog information
FSK reception	Data, including digitized analog information
PSK reception	Data, including digitized analog information

where bit error rate may then be estimated using curves such as those shown in Figure 8.24.

Output devices used by receivers are vastly different for analog and data receivers. Analog receivers (i.e., receivers designed for receiving analog information) often employ very high power output amplifiers that are intended to drive very low impedance loads. It is not unusual to see an amplifier that can drive 100 or more watts into a 2-ohm load, for example. This requires the ability to output a bit over 7 amps of current from the output amplifier. Some typical integrated circuit output amplifiers are shown in the following pages. (See Figures 8.30 and 8.31.)

For data receivers, it is often necessary to drive a line whose impedance may not be defined to less than a factor of 10 (say 100 to 1000 ohms) and which may require differential output instead of single-ended output. Line driver integrated circuits are used in such applications, of which a typical example also follows. Figures 8.32 and 8.33 show a typical line driver integrated circuit and two illustrations of its application. The particular integrated circuit shown is capable of either single-ended or differential operation.

Table 8.9 Typical Output Specifications for a Stereo Receiver

Parameter	Specification
Sensitivity	1.0 μV for 10 dB output *S/N* ratio, with 30% modulation at 1.0 kHz
Deemphasis	Six dB per octave
Signal-to-noise ratio	Not less than 10 dB
Distortion	Less than 0.1% at maximum output
Output impedance	8 ohms
Output power	50 W per channel
Channel separation	Not less than 40 dB at 1.0 kHz

National Semiconductor

LM3886 *Overture*™ Audio Power Amplifier Series
High-Performance 68W Audio Power Amplifier w/Mute

General Description

The LM3886 is a high-performance audio power amplifier capable of delivering 68W of continuous average power to a 4Ω load and 38W into 8Ω with 0.1% (THD + N) from 20 Hz–20 kHz.

The performance of the LM3886, utilizing its Self Peak Instantaneous Temperature (°Ke) (SPiKe) Protection Circuitry, puts it in a class above discrete and hybrid amplifiers by providing an inherently, dynamically protected Safe Operating Area (SOA). SPiKe Protection means that these parts are completely safeguarded at the output against overvoltage, undervoltage, overloads, including shorts to the supplies, thermal runaway, and instantaneous temperature peaks.

The LM3886 maintains an excellent Signal-to-Noise Ratio of greater than 92 dB with a typical low noise floor of 2.0 μV. It exhibits extremely low (THD + N) values of 0.03% at the rated output into the rated load over the audio spectrum, and provides excellent linearity with an IMD (SMPTE) typical rating of 0.004%.

Features

- 68W cont. avg. output power into 4Ω at V_{CC} = ±28V
- 38W cont. avg. output power into 8Ω at V_{CC} = ±28V
- 50W cont. avg. output power into 8Ω at V_{CC} = ±35V
- 135W instantaneous peak output power capability
- Signal-to-Noise Ratio ≥ 92 dB
- An input mute function
- Output protection from a short to ground or to the supplies via internal current limiting circuitry
- Output over-voltage protection against transients from inductive loads
- Supply under-voltage protection, not allowing internal biasing to occur when $|V_{EE}| + |V_{CC}| \le 12V$, thus eliminating turn-on and turn-off transients
- 11-lead TO-220 package

Applications

- Component stereo
- Compact stereo
- Self-powered speakers
- Surround-sound amplifiers
- High-end stereo TVs

Typical Application

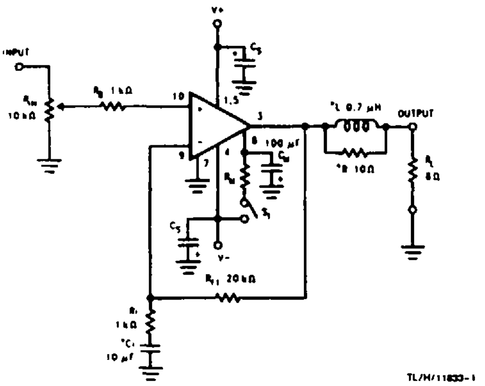
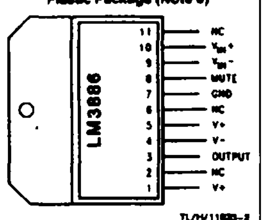


FIGURE 1. Typical Audio Amplifier Application Circuit

*Optional components dependent upon specific design requirements. Refer to the External Components Description section for a component functional description.

Connection Diagram

Plastic Package (Note 8)



Top View

**Order Number LM3886T or LM3886TF
See NS Package Number TA11B for Staggered Lead Non-Isolated Package or TF11B* for Staggered Lead Isolated Package**

*Preliminary: call you local National Sales Rep. or distributor for availability

Figure 8.31 High-power integrated circuit output amplifier for use in stereo or similar receiver applications. (Courtesy of National Semiconductor.)

MOTOROLA
SEMICONDUCTOR
TECHNICAL DATA

AM26LS30

Advance Information

Dual Differential (EIA-422-A)/ Quad Single-Ended (EIA-423-A) Line Drivers

DUAL DIFFERENTIAL/ QUAD SINGLE-ENDED LINE DRIVERS

SILICON MONOLITHIC INTEGRATED CIRCUIT

The AM26LS30 is a low power Schottky set of line drivers which can be configured as two differential drivers which comply with EIA-422-A standards, or as four single-ended drivers which comply with EIA-423-A standards. A mode select pin and appropriate choice of power supplies determine the mode. Each driver can source and sink currents in excess of 50 mA.

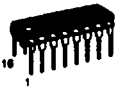
In the differential mode (EIA-422-A), the drivers can be used up to 10 Mbaud. A disable pin for each driver permits setting the outputs into a high impedance mode within a ±10 V common mode range.

In the single-ended mode (EIA-423-A) each driver has a slew rate control pin which permits setting the slew rate of the output signal so as to comply with EIA-423-A and FCC requirements and to reduce crosstalk. When operated from symmetrical supplies (±5.0 V), the outputs exhibit zero imbalance.

The AM26LS30 is available in a 16-pin plastic DIP and surface mount package. Ambient operating temperature range is –40° to +85°C.

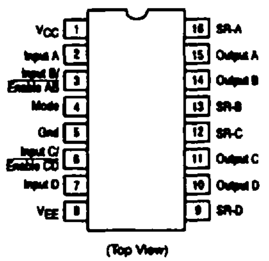
- Operates as Two Differential EIA-422-A Drivers, or Four Single-Ended EIA-423-A Drivers
- High Impedance Outputs in Differential Mode
- Short Circuit Limit In Both Source and Sink Modes
- ± 10 V Common Mode Range on High Impedance Outputs
- ± 15 V Range on Inputs
- Low Current PNP Inputs Compatible with TTL, CMOS, and MOS Outputs
- Individual Output Slew Rate Control in Single-Ended Mode
- Replacement for the AMD AM25LS30 and National Semiconductor DS3691

PC SUFFIX
PLASTIC PACKAGE
CASE 648



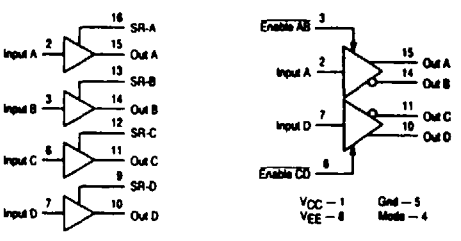
D SUFFIX
PLASTIC PACKAGE
CASE 751B
(SO-16)

PIN CONNECTIONS



(Top View)

BLOCK DIAGRAMS



ORDERING INFORMATION

Part No.	Ambient Temperature Range	Package Type
AM26LS30PC	–40° to +85°C	Plastic DIP
MC26LS30D		SO-16

Figure 8.32 Typical line driver integrated circuit suitable for application in the output of data receivers. (Courtesy of Motorola.)

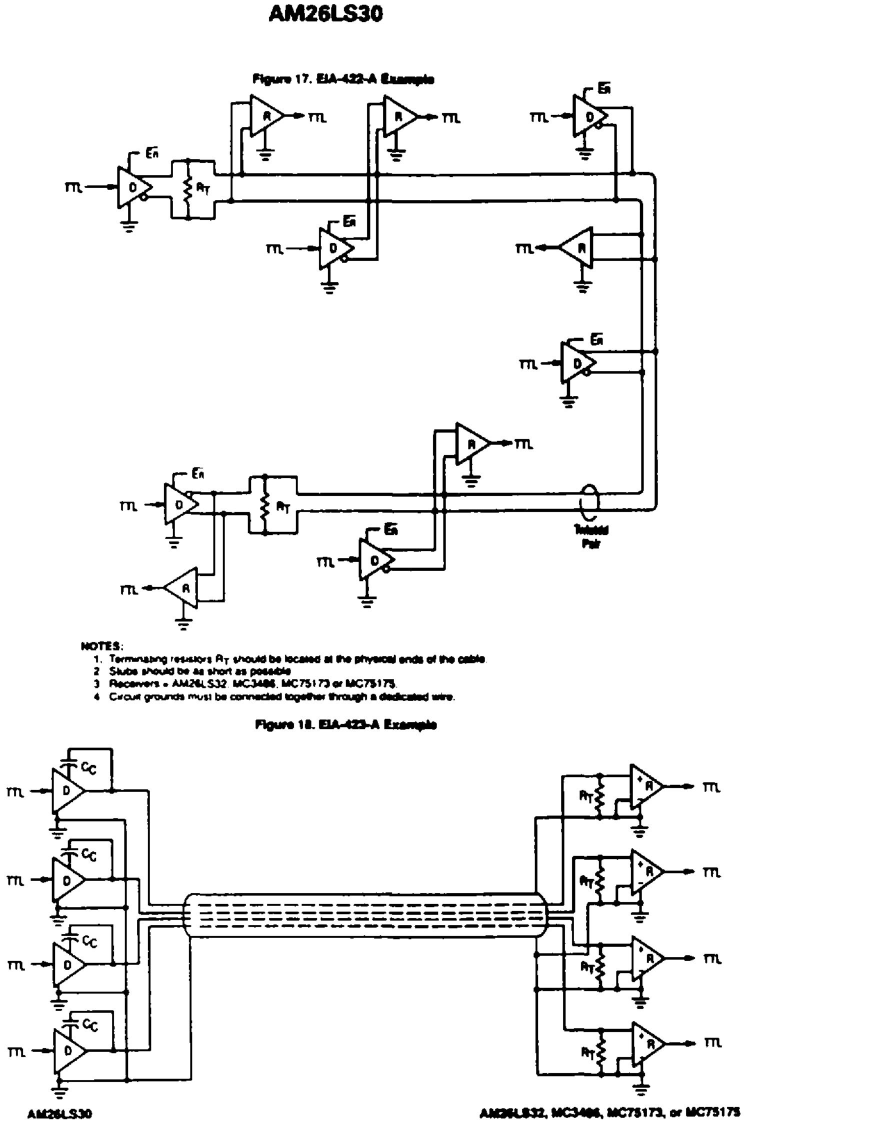


Figure 8.33 Illustrations showing the application of an AM 26LS30 integrated circuit line driver. (Courtesy of Texas Instruments.)

QUESTIONS

1. Why should a coherent demodulator's threshold be better than a noncoherent demodulator's threshold?
2. Define B_N, B_L, and K for a phase lock loop.
3. What are the differences between a phase lock demodulator and a phase lock synthesizer?
4. Design an envelope detector for a 5.0 MHz IF with an AM bandwidth of 50 kHz.
5. What is the effect of overdamping on a phase lock demodulator?
6. What is the approximate bandwidth required for a Costas loop that must lock within 10 msec when the maximum frequency offset is 100 kHz?
7. Can a product detector be used for AM demodulation?
8. What kind of demodulator would you choose for use with a noncontinuous phase frequency shift keyed signal? Why?
9. In a receiver for PSK signals that has a noise figure of 2.5 dB and a Costas loop demodulator with 1.3 dB loss, what sensitivity should be expected?
10. What does E_b/N_0 mean to a demodulator?

REFERENCES

1. Didday, R. L., and Lindsey, W. C., Subcarrier tracking methods and communication system design, *IEEE Trans. Comm.*, Aug. 1968.
2. Oetting, J. D., A comparison of modulation techniques for digital radio, *IEEE Trans. Comm.*, Dec. 1979.
3. Garner, W. J., Bit error probabilities relate to data-link S/N, *Microwaves*, Nov. 1978.
4. Lindsey, W. C., and Chie, C. M., A survey of digital phase-locked loops, *IEEE Proc.*, Apr. 1981.
5. Goldman, S., Understanding the limits of quadrature detection, *Microwaves & RF*, Dec. 1986.
6. Steber, J. M., Understanding PSK demodulation techniques, *Microwaves & RF*, Mar. 1984.
7. Gruen, W. J., Theory of AFC synchronization, *Proc. IRE*, Aug. 1953.
8. Hoffman, L. A., Receiver design and the phase-lock loop, *IEEE Electron Space Exploration Ser.*, Aug. 1963.
9. Cahn, C. R., Performance of digital phase-modulation communication systems, *IRE Trans. Comm.*, May 1959.
10. Amoroso, F., and Kivett, J. A., Simplified MSK signaling technique, *IEEE Trans. Comm.*, Apr. 1977.
11. Keiser, B. E., Digital modulation techniques compared, *Microwaves & RF*, Apr. 1983.
12. Gardner, F. M., *Phaselock Techniques*, John Wiley & Sons, New York, 1979.
13. Black, H. S., *Modulation Theory*, Van Nostrand, New York, 1953.

14. Dixon, R. C., *Spread Spectrum Systems*, John Wiley & Sons, New York, 1994.
15. Proakis, J. G., *Digital Communications*, McGraw-Hill, New York, 1983.
16. Watson, B. (R.), FSK: Signals and Demodulation, Watkins-Johnson Tech-Notes, Sept./Oct. 1980.
17. Steber, J. M., PSK Demodulation, Parts 1 and 2, Mar.–June 1984.
18. *VCO Designer's Handbook*, Mini-Circuits, 1996.

9
Defining and Measuring Performance

Receiver performance definition and measurement are a function of the application of the specific receiver. For general-purpose receivers, such as those that are usually called "communications receivers," which operate in the range 2.0 to 30.0 MHz, a superheterodyne configuration is common. On the other hand, banks of TRF receivers are sometimes used to monitor wide bands of frequencies with high probability of detection. These two very different kinds of receivers have very different requirements, even though functionally they perform almost the same kinds of tasks. Here, we will consider overall receiver specifications. Later, in Chapter 12, we will consider specifications for the various subsystems of receivers.

9.1 RECEIVER SPECIFICATIONS

A typical receiver specification can be summarized in 1 or 2 pages, but a complete specification for a military receiver may take up to 50 pages or more. An example is the AN/ARC-182, which is a VHF-UHF transceiver whose summary specification sheet is only 2 pages but whose military specification fills 78 pages. (To be more precise, the AN/ARC-182 includes both a receiver and a transmitter and is not a transceiver but is actually a receiver-transmitter.) The second page of the summary specification is given here as an example of the parameters that may often be seen in such summaries. It should be noted that this summary specification does include such things as sensitivity but leaves out other important receiver characteristics such as adjacent channel rejection. (See Table 9.1.)

It is also interesting to note the specification for "frequency stability" at ± 1 PPM. When specifying frequency stability, a period for which the stability is valid must be given, or else nothing has been specified. That is, the specification should say

Table 9.1 AN/ARC-182(V) VHF-UHF AM/FM Transceiver

SYSTEM SPECIFICATION	
Frequency range	30.00 to 87.975 MHz FM. 108.00 to 117.975 MHz, receive only, AM. 118,00 to 155.975, AM. 156.000 to 173.975, FM. 225.000 to 399.975, AM/FM
Channel spacing	25 kHz in all bands.
Guard receiver	40.50 MHz FM. 121.50 MHz AM. 156.80 MHz FM. 243.00 MHz AM. The frequency of operation of the guard receiver is determined by the band of operation selected for the main receiver.
Sensitivity	3 μV for 10 dB (s + n)/n ratio of 30-percent modulation AM. 1 μV (156-174 MHz) for 10 dB (s + n)/n ratio for 8-kHz deviation FM. 0.6 μV (30-88 MHz)
Transmitter power output	10 watts CW minimum—AM. 15 watts CW minimum—FM.
Duty cycle	5 minutes receive. 1 minute transmit.
Transmitter modulation capability	
AM	NLT 80 percent upward for 90 percent downward.
FM	Deviation capability of NLT 25 kHz, (5.6 kHz nominal—normal voice, secure voice).
Transmitter modulation response	
AM	Narrow band, 10 Hz to 12,000 Hz. Wide band, 10 Hz to 25,000 Hz.
FM	10 Hz to 12,000 Hz.
Receiver audio output	250 milliwatts into 150/600 ohms.
Receiver audio response	400 to 3000 Hz AM/FM
Normal voice	
Wide band audio	10 to 25,000 Hz
Image	80 dB minimum
Spurious rejection	70 dB minimum.
Selectivity	6 dB, 37 kHz minimum.
Narrow band	60 dB, 50 kHz maximum.
Wide band	6 dB, 69 kHz minimum. 60 dB, 170 kHz maximum.
Environmental	MIL-E-5400 class II with NMT 3-dB power reduction at 71° C.
Input power	+28 V dc, in accordance with MIL-STD-704B.
Frequency stability	± 1 PPM.
PHYSICAL CHARACTERISTICS	
Weight	4.54 kg (10 lb).
Size	14.6 cm wide (5-3/4 in), 12.4 cm high (4-7/8 in), 16.5 cm deep (6-1/2 in), behind the mounting surface.
EQUIPMENT LIST	
Options	Panel or remote mount transceivers, remote control, indicator.
ADF/homing	ADF in vhf-AM and uhf-AM, homing in vhf-FM (30-88 MHz).

Frequency stability ... ± 1 ppm per year*

to define the period for which the receiver's (and in this case also the transmitter's) frequency will be within ± 1 Hz per MHz of the frequency to which it is supposed to be tuned. This also defines, by default, the time period that can be allowed between recalibrations of the radio. The complete specification has not been included here because of its size.

A listing of electrical parameters that should be specified for a receiver is given in the following pages, with explanations of the reasons for that particular specification. This listing is not necessarily exhaustive but is typical of what is found in specifications for general receiver applications.

Frequency Range. This is the range of frequencies over which a receiver must operate and meet all of the other requirements given in the list of specifications.

Channel Spacing. Channel spacing defines the resolution that must be provided by the receiver's tuning mechanism within the receiver's operating frequency range.

Frequency Accuracy. This is the accuracy of the receiver's tuning for all causes, including temperature, altitude, humidity, and long-term stability. Frequency accuracy is usually specified in terms of parts per million as a fraction of the intended operating frequency.

Sensitivity. Sensitivity is the minimum signal level required to cause the receiver to output a specified signal-to-noise ratio or bit error rate, when the input signal to the receiver is modulated at a "standard" level by a specific modulating signal. In voice receivers, the standard level is 30 percent of the peak modulation permitted, and the modulation rate used is 1 kHz.

In data receivers, the output of the receiver is judged on the basis of bit error rate when the input signal to the receiver is modulated by data. Of course, data-modulated signals are always modulated at a level that is effectively 100 percent.

With both data and voice receivers, sensitivity is determined on the basis of the minimum input signal (modulated with the proper information for testing) that produces a given minimum output signal-to-noise ratio or maximum bit error rate.

Modulation. Modulation, with respect to a receiver, refers to the type or types of modulation that a receiver must demodulate. It is not unusual for a receiver to be able to demodulate several types of analog and digital (discrete) modulation such as AM, FM, and FSK, or even PSK.

Modulation Bandwidth. This refers to the baseband bandwidth of the information signal. For voice, the modulation bandwidth specified is

*Or other time period of interest.

usually 300 to 3000 Hz, although there is some variation according to application. For music (in FM radio, for example) the modulation bandwidth is typically 20 to 15,000 Hz. In a receiver, the idea is to demodulate and pass the information with a flat frequency response and minimum distortion. Modulation bandwidth is not usually defined in receivers that are intended for data reception, unless they are also expected to handle analog signals.

Data Rate. Received data rate is specified for a receiver, as the receiver is expected not only to demodulate the received signal (recover the data conveyed) but also to recover and provide the data clock at the output interface. Often, a range of data rates is required.

Distortion. In receivers intended for reception of analog information, especially FM broadcast signals, distortion is an important consideration. The receiver, of course, cannot output a signal that is better than what it receives, but it should not materially degrade that which it receives. This measurement is made on the basis of having a low-distortion input signal, with the effect of the receiver being measured at the receiver output, after demodulation and amplification.

Output Impedance. Output impedance is specified on the basis of application. Aircraft intercommunication systems typically operate at 150- or 300-ohm impedances, for example, while stereo receivers are often designed to drive 8-ohm loads. Data receivers may have line driver ICs at their output, driving either single-ended or balanced lines. In any case, the receiver must be specified to meet the particular application.

Output Power. Output power also varies as a function of application. Stereo receivers may have output power as high as several hundred watts. Communication receivers seldom require more than a few hundred milliwatts to perhaps 1 W. (The average power for a comfortable listening level in a moderately sized room is around 250 mW, with reasonably efficient loudspeakers). Military intercommunication systems require output levels of 150 mW to 1 W (they also often allow distortion as high as 10 percent).

Input Signal Range. This refers to the RF input signal range that a receiver can operate with, while affecting the receiver's output signal-to-noise ratio as little as possible. For example, a receiver might have its input signal change by as much as 100 dB, while its output bit error rate is maintained at a maximum of 1×10^{-3}. Or, in a broadcast receiver whose input is rapidly changing over 60 to 80 dB, present an output that changes by no more than 1 dB. It is not unusual for a receiver in mobile or aircraft applications to have an input that varies over a 140-dB range.

RF Input Impedance. Input impedance of receivers is usually either 50, 75, or 300 ohms, where television receivers employ 75-ohm input for interface with low-cost (RG-59) coax cable and 300-ohm input for use with low-cost twin lead. Many television receivers provide for both 75- and 300-ohm input connections. Outside the television world, 50-ohm inputs are most common, with an occasional 90-ohm input to match 90-ohm cable.

Input VSWR. VSWR (voltage standing wave ratio) for receivers is specified to limit the mismatch at the input to a receiver between the antenna, or the cable from the antenna, and the receiver's input circuits. An alternative, and equivalent, specification is that of input return loss. VSWR is usually specified as less than 2:1 and return loss as 10 dB or greater. In narrowband receivers or critical applications, VSWR as low as 1.1:1 or 1.2:1 might be specified, or return loss as high 20 dB.

Image Rejection. Military receivers often call for image rejection of at least 80 dB. Commercial and consumer receivers may not require as much image rejection, and it may not be practical to do so well in broadcast band receivers. In fact, with a 455-kHz IF frequency, the image frequency of an AM broadcast band receiver falls right in the broadcast band itself. In such applications, the best defense is a tunable filter in the receiver front end, but this often leads to other problems.

The designer's best judgment must be exercised in specifying image rejection, based on the specific application and the particular part of the frequency spectrum in which the receiver must operate.

Adjacent Channel Rejection. Military specifications require 80 dB or more adjacent channel rejection. They are not usually successful at reaching this level of performance in today's receivers. It is easy to see why when the receivers are examined. Adjacent channel rejection is provided by the IF filters in a receiver, and in military receivers these filters are either surface acoustic wave or crystal filters (often both). Although such filters are capable of providing 80 dB isolation under ideal conditions, they cannot do so when simply mounted on an open printed circuit board, as is often necessary. Consumer and commercial receivers have even more of a problem with isolation because of cost considerations. Observance of Table 9.1 will show that adjacent channel rejection is not even mentioned, and this is common in today's military radios. Their adjacent channel rejection is typically in the range of 40 to 60 dB, and this is also true of nonmilitary receivers.

Selectivity. Adjacent channel rejection is actually a special case of selectivity, which in turn is simply a convenient way to express the response of a receiver to signals that are at a frequency to which the receiver is not tuned and for which it is hoped that there will be no response at all. Selectivity is also a function of the receiver's IF filters. It is specified as rejection (compared with the receiver's response to a signal at the frequency to which it is tuned) at frequencies offset from the desired center frequency.

For example, a receiver selectivity specification may call for response down 60 dB when the input signal is offset by 100 kHz from the frequency to which the receiver is tuned.

Spurious Rejection. All spurious responses other than image response are often lumped together with a common minimum level of rejection (again compared with the receiver's response to a signal at the receiver's desired frequency). As in the previous cases, military receivers strive to reach 80 rejection levels, whereas other receivers may be content with 40 to 60 dB rejection. As most of the spurious responses in receivers apply primarily to superheterodyne receivers, because of their mixing process, spurious rejection is not as much of a concern in other architectures.

Frequency Accuracy. Frequency accuracy is very important to a receiver, because any offset in its tuning can degrade the demodulation process. Overall frequency accuracy is determined by the type of frequency-determining device employed in the receiver. Unfortunately, cost is also determined by the type of frequency-determining element. In most applications, the only practical, reasonably priced solution for nominal frequency accuracy of 1×10^{-5} to 1×10^{-6} is to employ quartz crystals.

Other specifications, such as operating temperature range, size, and weight, must also be carefully considered.

9.2 RECEIVER PERFORMANCE MEASUREMENTS

Almost all receiver measurements are based on the same test setup that is employed for measurement of receiver sensitivity, and many of the criteria for receiver performance are based on the sensitivity measurement. For this reason, we will proceed by discussing receiver sensitivity and then go on to other receiver measurements and the techniques used to carry them out.

9.2.1 Sensitivity

Sensitivity is one of the most important characteristics of a receiver, inasmuch as it defines the receiver's ability to accept a minimum-level signal and derive from that signal useful output information. Sensitivity, in effect, is the essence of the receiver's entire reason for existing, and in measuring sensitivity we attempt to examine the receiver's performance in a controlled situation.

A typical test setup for receiver sensitivity measurement is shown in Figure 9.1. (Today, specialized test generators exist that incorporate all of the functions shown here into a single instrument.) The signal input to the receiver is generated

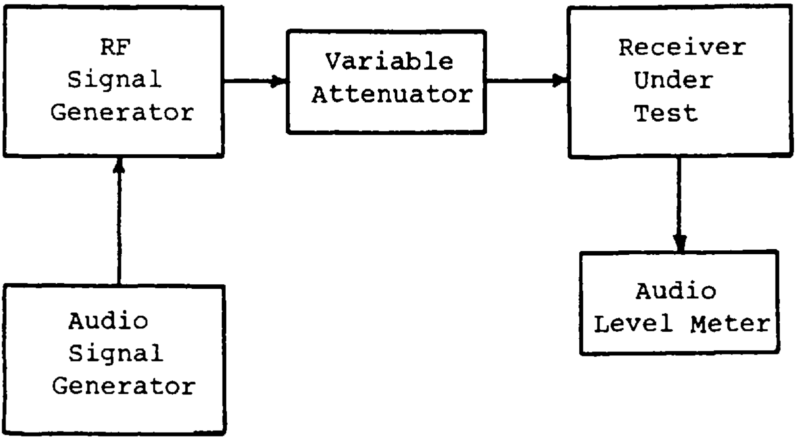


Figure 9.1 Test setup for receiver sensitivity using $(S + N)/N$ ratio. (This is the basic setup used for all receiver measurements.)

by a modulated RF generator, whose modulating signal comes from another signal generator that serves as an information source for testing purposes. In the case of voice (or music) receivers, for example, the information signal generator would be a simple audio tone generator. For data receivers, however, the information source would be a data simulator. At the receiver's output, to measure the demodulated signal level and the signal-plus-noise-to-noise ratio, a measurement device such as a voltmeter, oscilloscope, or distortion analyzer may be used.

In the case of data systems, the receiver's sensitivity is tested by measuring bit error rate versus input RF signal level. Sensitivity for both data and analog receivers is, of course, just a measure of the performance of a receiver in the presence of the receiver's own noise plus KTB in the receiver's noise bandwidth, with the smallest signal that will produce a usable output. Figure 9.2 illustrates a test setup using bit error rates for receiver testing.

For some types of receivers it is necessary to employ a distortion analyzer in measuring sensitivity. The reason is that (obviously) if the measurement is a SINAD sensitivity measurement, which involves signal-plus-noise-plus-distortion-to-noise as a criterion for acceptability, then a distortion analyzer is needed. Not so obvious is the need for a distortion analyzer in single-sideband or double-sideband suppressed carrier receivers. In such receivers, if the modulating signal is removed from the RF input, there is no carrier remaining and the standard technique for measuring sensitivity is not practical.

The standard technique for measuring sensitivity by measuring receiver output with and without input signal modulation is as follows:

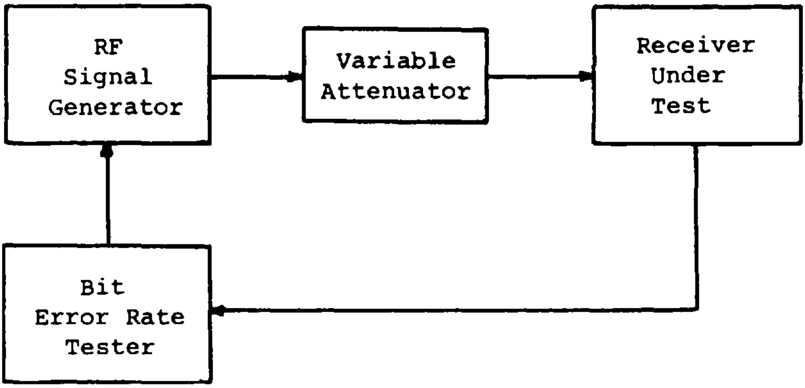


Figure 9.2 Variation of receiver sensitivity test setup used for data receivers.

1. Set up the receiver to be tested and the test equipment to be employed, as shown in Figure 9.1.
2. (Assuming a voice receiver.) Set the signal generator being used for modulation to 1 kHz and its level to that required to drive the RF signal generator to a 30 percent modulation level.
3. Set the RF generator to the frequency to which the receiver being tested is tuned, with 30 percent modulation at 1 kHz.
4. Measure the receiver's output (signal-plus-noise)* level, using the voltmeter or other device.
5. Set the modulation percentage to zero and repeat step 4. The ratio of the measurements in steps 4 and 5 is the receiver's signal-plus-noise-to-noise ratio $(S + N)/N$. At a 10 dB $(S + N)/N$ ratio or higher, the noise is small compared with the signal level, and this often leads to the mistake of specifying output in terms of S/N ratio instead of $(S + N)/N$ ratio. Table 9.2 shows the relationship between the two quantities, which is small but worthy of noting.
6. Adjust the RF signal level either up or down as required until the difference between the measurements made in steps 4 and 5 is 10 dB (or 12 dB if the specification requires it). The RF level that produces the proper minimum $(S + N)/N$ ratio is the receiver sensitivity.

If the receiver being tested is a single-sideband or a double-sideband, suppressed carrier receiver, then a distortion analyzer should be substituted as the output measuring device and step 5 of the sensitivity measurement process de-

*In some cases, a signal-plus-noise-plus-distortion level will be measured here.

Table 9.2 Comparison of Signal-to-Noise Ratio and Signal-plus-Noise-to-Noise Ratio (Normalized to Signal Power = 1)

Noise power	S/N	S/N (dB)	$(S + N)/N$	$(S + N)/N$ (dB)	Difference (dB)
0.2	5.0	6.98	6.0	7.78	0.8
0.111	9.0	9.54	10.0	10.0	0.46
0.10	10.0	10.0	11.0	10.41	0.41
0.05	20.0	13.0	21.0	13.22	0.22

leted. The distortion analyzer allows any single tone modulation output from the receiver to be suppressed, and this leaves only the noise and distortion components at the receiver's output, which can be measured independently. (This technique can also be applied to AM or FM receivers, so that the modulation does not have to be removed from the test RF signal generator, but a distortion analyzer is needed.)

Data receivers are tested for sensitivity using the same type of test as analog receivers, but with two basic differences:

1. Data-modulated RF signals are used to test the receiver.
2. Bit error rate is employed as the criterion for output performance.

Because of these differences, a bit error rate analyzer may be substituted for both the input information source (random or pseudorandom data) and the output measurement device. Such a bit error rate tester is shown in Figure 9.2. It consists of a word generator, whose output is used as the information to modulate the RF signal generator, a counter that counts the number of bits transmitted, a comparator that compares the receiver output data with what was input to the RF test generator and counts the number of errors, and finally a numerical unit that divides the number of errors detected by the number of bits sent and produces a "bit error rate."

The sensitivity test itself is run by modulating the RF signal generator with data, measuring the receiver output error rate, and adjusting the RF signal level until the bit error rate is equal to the maximum allowed. This RF level is the data receiver's sensitivity.

Other types of sensitivity measurements are made with the same test setup but with somewhat different criteria or slight modifications. Some examples are:

"Hard" sensitivity is measured by adding a 6-dB attenuator at the output of the RF signal generator and proceeding with the test in exactly the same way as previously described. The sensitivity measured will be worse than when the attenuation was not present (an obvious conclusion).

Surveillance receivers, for which the primary purpose is to recognize that a signal is present but not necessarily to demodulate any information that might be conveyed by it. (Such receivers are also often found in radar applications, where unmodulated pulses might be employed.)

Sensitivity for non-information-demodulating receivers is often defined in terms of either "minimum discernible signal" (MDS) sensitivity or "tangential" sensitivity. MDS sensitivity is measured by observing the signal detector (usually an envelope detector) output and adjusting the RF input level to the receiver until turning the RF input on and off produces a 3-dB difference in the output level. This RF level is the MDS sensitivity.

Tangential sensitivity does not depend as heavily on subjective judgment. (Three dB is theoretically the minimum difference that can be detected by a human observer or listener.) Instead, tangential sensitivity is taken at the point in signal level at which having the signal present produces a signal in the receiver whose envelope is tangent to the envelope of the noise alone. Figure 9.3 illustrates the concept of tangential sensitivity.

9.2.2 Selectivity

Selectivity tests employ the same test setup used for sensitivity testing. The difference is that instead of employing an input signal that is tuned to the frequency to which the receiver is tuned, the input frequency is varied over a wide range. The overall process is as follows:

1. Measure the receiver's sensitivity.
2. Change the frequency of the RF generator by an increment (determined by the receiver's bandwidth and the specified requirements).
3. Measure sensitivity again.
4. Repeat steps 2 and 3 over the complete frequency range to be covered.
5. Compare the sensitivity measured at each of the points across the range

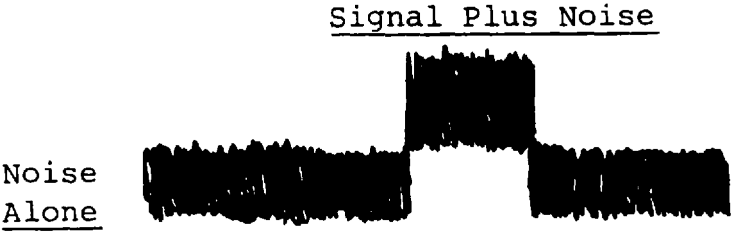


Figure 9.3 Waveforms as seen in tangential sensitivity measurement. The noise is tangent to the signal plus noise. (Note that the lower edge of the signal plus noise is at the same level as the upper edge of the noise alone.)

of interest to the sensitivity of the receiver at the frequency to which the receiver is tuned. A plot of the difference in sensitivity as a function of offset from the receiver's center frequency will yield a "selectivity curve" for the receiver, and selectivity at specific points can be read directly from the curve. (In many specifications, only a few specific frequency offsets are called out, so it may not be necessary to make more than a few selected measurements.)

9.2.3 Adjacent Channel Rejection

Adjacent channel rejection is actually a special case of selectivity, wherein a selectivity test is made with the RF signal generator set to the increment of frequency next to that at which the receiver is tuned. Measurements are often made of the adjacent channel performance of the receiver when the signal generator is set both above and below the receiver's tuned frequency.

9.2.4 Dynamic Range

Receiver dynamic range measurement is done in the same test setup used for sensitivity and other measurements. It can well be started with a sensitivity measurement, in fact, because the dynamic range of a receiver is truly the signal range over which the receiver can operate, while meeting the specified output information specifications, and one end of the signal range is set by the receiver's sensitivity. The other end of the range is determined by the largest signal that the receiver can handle without evidence of output signal degradation.

A direct measurement can be made of single-signal dynamic range, as such a measurement is intended to show the performance of the receiver with a single signal set to the frequency to which the receiver is tuned. When more than one signal is present, however, there are so many different possible conditions that testing for every contingency is more than is practical. Therefore, the usual procedure is to simulate the performance based on predictions derived from 1-dB compression and third-order intercept information and to spot test at various points selected by observation of these analyses and simulations.

A "spurious-free dynamic range" can be predicted on the basis of the intercept point for the receiver and its components. (See Chapter 5.)

9.2.5 Spurious Responses

Spurious responses and their rejection are also tested in the sensitivity test setup. In military applications, almost all spurious responses are specified at −80 dB with respect to the receiver's sensitivity. That is, any signal at a frequency specified as a spurious frequency (such as an image frequency) should require 80 dB more receiver input to give the same level of receiver output as a signal tuned to the same frequency as the receiver. Commercial and consumer receivers do not usually do as well, because their operating conditions *may not* be as stringent as those of military receivers, and it is often difficult and expensive to reduce spurious signal responses to a −80 dB level.

Some spurious responses for which specific tests are usually conducted and their frequencies are listed in Table 9.3.

Tests for spurious responses are done in the same test setup used for sensitivity testing and are often done in exactly the same way but with the test signal set to the spurious response test frequency. Where two frequencies are employed, a second RF signal generator is added, but the second signal generator is not usually modulated, the idea being to see if the receiver will employ the two signals in a way that produces a component that the receiver recognizes as valid.

As in other cases, spurious response is judged on the basis of a comparison of the receiver's sensitivity with the receiver's response to a signal at the spurious frequency. That is, if the receiver sensitivity is −95 dBm but it takes −25 dBm to get the same output from the receiver at the image frequency, then the receiver's image response is the difference between the two, or −70 dB. That is, it takes 70 dB more input to the receiver at the image frequency to get the same output as the signal at the frequency to which the receiver is tuned.

Table 9.3 Some Spurious Responses and the Frequencies at Which They Should Be Expected

Spurious response	Expected frequency or frequencies
Image response	RF ± 2 × IF (dependent on whether high- or low-side injection is employed)
IF response	IF frequency or frequencies
$\frac{\text{RF}}{\text{N}}$ response	$\frac{\text{RF}}{\text{N}}$ frequencies
$\frac{\text{IF}}{\text{N}}$ response	$\frac{\text{IF}}{\text{N}}$ frequencies
Double and triple beats	Any two frequencies, such that $f_1 \pm f_2$, or $2f_1 \pm f_2$, or $f_1 \pm 2f_2$ fall in the receiver signal range
Premixing	Any two frequencies, such that $f_1 \pm f_2$, or $2f_1 \pm f_2$, or $f_1 \pm 2f_2$ fall at the first IF frequency
Birdies	Spurious responses for which there is no reasonable prediction method

9.2.6 Receiver Desensitization

Receivers may be desensitized by signals that are internal to themselves (usually "birdies" generated by the receiver's own frequency synthesizer[s]) or by external signals input to the receiver along with a desired signal.

When the desensitizing signal is internal, it may fall at or near the RF input frequency, or it may fall within the bandwidth of one of the IFs. In either case, it can be just as effective at desensitizing a receiver. If, for example, the internal undesired signal is larger than the desired signal at the second IF frequency, the input RF signal must be increased, and the receiver's effective sensitivity has been degraded. It is not unusual for a new receiver design to be tested at every frequency in its operating range, so that any desensitized frequencies may be found. In some cases, automated testing is performed on high-performance receivers at every frequency as well, where a critical application is involved, but general practice does not involve such testing for multichannel receivers once they reach high-volume production.

A receiver may also be desensitized by the presence of a large signal at its RF input frequency or close to it. In fact, there is always some signal level at which a receiver becomes nonlinear, and its performance begins to degrade, if a large unwanted signal is present along with the desired signal at the receiver input. This is a case for which "jamming resistance" tests are called out in military specifications and for which *C/I* (carrier-to-interference ratio) tests are specified in other applications.

Again, the receiver sensitivity test setup is employed, with the addition of a second signal generator that acts as the interferor. Tests are performed by measuring the receiver performance as a function of signal-to-interference ratio.

9.2.7 Distortion

Distortion measurements in a receiver presuppose that the received signal itself has no distortion or that its distortion is known. Therefore, the test equipment employed for making the distortion test should be of good quality and well calibrated. The receiver output should be measured with a distortion analyzer and with various levels up to the maximum required. The test setup itself is the same as that used for sensitivity measurements.

In military receivers, distortion as high as 10 percent is allowed. For consumer systems, especially stereo FM receivers, the receiver may be required to contribute less than 1 percent distortion.

Distortion in a receiver is measured by modulating the RF generator with a test tone at the standard level (usually 30 percent), setting the receiver output to a specified level, and finding the output distortion level with a distortion analyzer.

9.2.8 Frequency Accuracy

The accuracy of a receiver is a function of its local oscillator(s), which is often synthesized from a single crystal-stabilized source, because having multiple frequency sources in a receiver can cause spurious responses that would not exist if all the frequencies in the receiver had an integer relationship with one another.

One technique for measuring the frequency accuracy of a receiver is to input an accurate RF signal and measure the frequency of the IF signal input to the receiver's demodulator. This IF signal will be offset by any inaccuracy in the local oscillators used to convert it to the final IF frequency. A spectrum analyzer can be employed to perform this measurement.

A second method is to measure the local oscillator frequencies directly with a frequency counter. No input signal is necessary for this test. The first local oscillator's frequency is the RF input frequency, offset by an amount equal to the first IF frequency (higher than the IF with high-side injection and lower than the IF with low-side injection). A second local oscillator would be at a frequency offset from the first IF frequency by an amount equal to the second IF frequency. That is,

$$f_{\mathrm{LO1}} = f_{\mathrm{RFin}} \pm f_{\mathrm{IF1}}$$

and

$$f_{\mathrm{LO2}} = f_{\mathrm{IF1}} \pm f_{\mathrm{IF2}}$$

Of course, if the receiver is a triple conversion, or more, receiver, then each local oscillator frequency would be measured. Where all local oscillators are generated from a single source, then only that source would need to be measured.

One other special case, that of a receiver with a coherent demodulator (such as a phase lock or Costas loop), is worthy of mention. In such a receiver, the accuracy can be measured by putting in an accurate RF signal and measuring the frequency of the VCO in the demodulator when the receiver is locked.

As usual, the test setup for receiver sensitivity would be employed for frequency accuracy measurements, except that direct measurement of the local oscillators with a frequency counter does not require any test setup at all; it is necessary only to supply power to the receiver so that the oscillators are operating and stabilized.

A test of frequency stability differs from one of frequency accuracy only in that the same tests should be performed at regular intervals to determine the change in frequency as a function of time.

9.2.9 Cross-Modulation

Modulation transferred from one signal to another because of nonlinearity in a receiver sometimes occurs when two signals are close enough in frequency to fit within the receiver's input bandwidth. This transfer can be quantified by adding a signal generator to the sensitivity test setup, where that signal generator is modulated but is offset in frequency by a specified amount (usually one or more channels) while the generator set to the desired frequency is unmodulated.

Then, after the receiver's output is set for a standard level with a standard modulated RF signal, the cross-modulation level from having a modulated signal in a nearby channel is compared with that due to a desired signal in the correct channel (the channel to which the receiver is tuned). The difference is the cross-modulation performance. It is normal for the small-signal performance and the large-signal performance to differ, so tests should be performed at various signal levels. It may also be necessary to input the offset, modulated signal at a level significantly higher than the centered, unmodulated signal.

9.2.10 Interference Rejection

All multiple access receivers should work in the presence of interference, but the signal-to-interference ratio for most receivers, such as personal communications system (PCS) receivers, is almost always positive. Some receivers, such as those used in military antijamming systems, are expected to operate when their input signal is considerably smaller than any interfering signal that may be present with the desired signal. Spread spectrum receivers often offer good examples of such operation. Figure 9.4 shows a test configuration that is adapted for antijamming measurements.

The jamming or interference* signal used in this testing is often very specific, because the receiver used in a particular system is often more vulnerable to one type of signal than another. For example, direct sequence receivers are typically most vulnerable to unmodulated carrier interference (CW) near their center frequency and frequency hopping receivers are typically most vulnerable to broadband interference. Of course, jamming is often much more sophisticated and system specific, because its purpose is to cause as much havoc as possible.

Jamming tests are usually done in two parts. One set of tests is intended to measure the ability of a receiver to synchronize or acquire the signal in the presence of jamming. These tests are run at various jamming-to-signal ratios, and a number of trials are done at each ratio. Then the number of successes versus trials is plotted as a function of the J/S ratio to show the probability of acquisition versus interference. The second set of tests is done to determine a receiver's ability to track and hold the signal in the presence of interference, after it has acquired the desired signal.

The ability of a receiver to operate in the presence of signals that happen to be there but are not intended for the receiver, while one signal is present that is

*Receivers cannot tell the difference between a jammer and an interferer, nor do they care. An interferer is usually considered to be another user who happens to interfere; a jammer may be an active, deliberate, malignant interferer who does all that he can to be a problem.

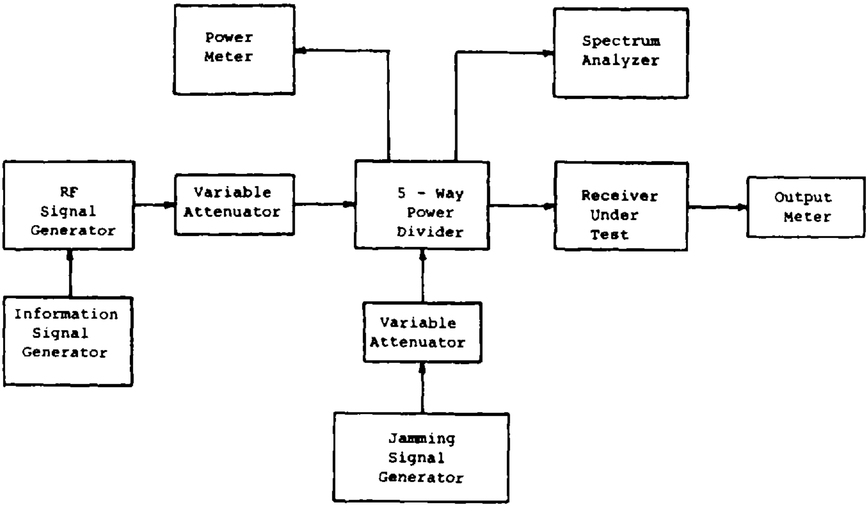


Figure 9.4 Receiver jamming resistance test setup.

intended for it, is also of interest and is tested in the same way. The measurement is usually called a *C/I* or carrier-to-interference test, however, and is done using a transmitter similar to the one that is used to generate the desired signal itself.

9.2.11 Group Delay and Phase Shift

At this time, it is practical to measure phase shift directly, using either a vector voltmeter or a network analyzer. In some cases it is practical to make such a measurement with an oscilloscope, but the frequency range of modern receivers limits one's ability to make a straightforward oscilloscope measurement.

This limitation leads directly to the original reason for the first development of a delay/phase shift measurement method that was employed long before the availability of network analyzers or vector voltmeters. The technique makes use of relatively low-frequency instruments and is called a "group delay" measurement.

To make a group delay measurement, an amplitude-modulated signal generator is used as a signal source and input to the device being tested. At the device output, a diode AM detector is used to recover the signal envelope. Then the modulating signal and the recovered envelope signal are compared on a dual-trace oscilloscope, and the group delay can be measured directly from the difference in the two traces. Further, both diffential delay and delay slope can be measured by simply sweeping the input signal frequency or making measurements in discrete steps across the band of interest.

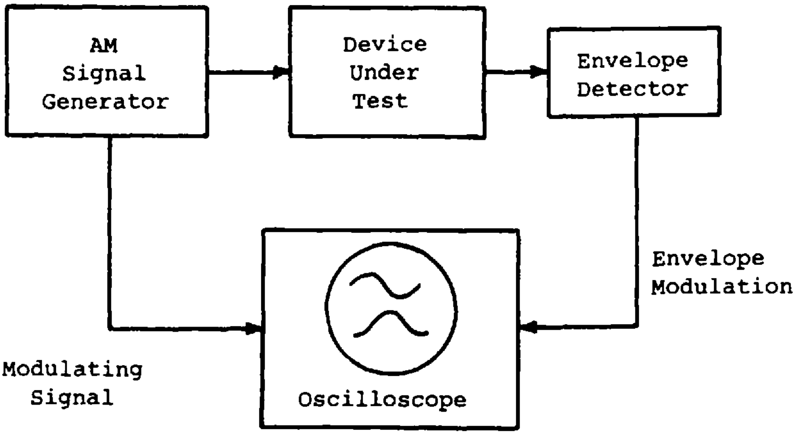


Figure 9.5 Illustration of test setup for group delay measurement.

Figure 9.5 illustrates a test setup for group delay measurement. Delay through a receiver or through a single device such as an amplifier can be measured with this setup. Delay through the demodulator may be accomplished if the receiver demodulator can pass the envelope modulation frequency. (The higher the modulating frequency, the better the delay resolution.) Otherwise, group delay would be done only between a receiver's RF input and its final IF output.

Neither a network analyzer nor a vector voltmeter can make this delay measurement directly, as they are not designed to employ a different input and output frequency and must therefore convert a receiver's IF output signal to the input frequency. This process itself can introduce significant potential sources of error. Group delay, on the other hand, does not depend on having the same frequency at both the input and the output of the measuring device and can operate with an input at RF and an output at IF.

9.2.12 Intermodulation, 1-dB Compression, and Third-Order Intercept

Although intermodulation, P1dB, and IP3 are important contributors to a receiver's performance, they are not usually measured with the receiver on a complete system basis. Instead, they are measured on a subsystem level as necessary and the effect on a complete receiver is extrapolated from the individual specifications. It is practical to monitor the output of a receiver to determine the maximum signal level (input) at which the receiver puts out an acceptable signal-to-noise ratio or bit error rate, but it may be difficult to say precisely which stage in the receiver has caused the problem at that input level. At least this is true without looking at the signal in the individual stages of the receiver.

Chapters 4 and 5 discuss signal handling and spurious response of individual amplifers and mixers.

QUESTIONS

1. What is the difference between a receiver sensitivity measurement and an adjacent channel rejection measurement?
2. Name three spurious response tests that should be performed with a superheterodyne receiver.
3. Given a receiver whose sensitivity is 5 μV for a 10-dB output S/N ratio and whose IF bandwidth is 25 kHz, calculate the receiver's noise figure.
4. How can a bit error rate test be performed without a bit error rate tester?
5. Calculate the image frequencies for an AM receiver with a 455-kHz IF.
6. How is image response measured?
7. Should AGC be disabled when measuring receiver sensitivity?
8. What is the difference between a "hard" and a "soft" sensitivity measurement?
9. What image response frequencies should be tested for a 1-GHz receiver that has a 350-MHz IF, a 40 MHz IF, and high-side injection? What about low-side injection?
10. How is signal-to-noise ratio measured, and how is E_b/N_0 measured?

REFERENCES

1. Kinley, R. H., *Standard Radio Communications Manual*, Prentice-Hall, Englewood Cliffs, NJ, 1985.
2. Dixon, R. C., *Spread Spectrum Systems—With Commercial Applications*, John Wiley & Sons, New York, 1994.
3. Scrupski, S. E., *Electronic Engineering Measurement Filebook*, TAB Books, 1965.
4. Wheeler, G. J., *Introduction to Microwaves*, Prentice-Hall, Englewood Cliffs, NJ, 1963.
5. Kemp, B., *RF Interference Control Handbook*, H. W. Sams, 1962.
6. Watson, R., Guidelines for receiver analysis, Parts 1 and 2, *Microwaves & RF*, Dec. 1986.
7. Cokenias, T., Spread-spectrum operators face compliance issues, *Wireless Systems Design*, Apr. 1997.
8. Federal Communications Commission, Part 24—Personal Communications Services, 47 CFR Chapter 1.
9. Federal Communications Commission, Part 15.247 through 15.307, 47 CFR Chapter 1, ISM bands and unlicensed PCS.

10
Receiving Antennas

Antennas used by receivers have the ability to influence the performance of a communications system just as much as the antennas used by transmitters. Somehow, they never seem to attract as much attention, however. Perhaps this is because losing a few dB of output power at a transmitter seems so much worse than losing the same amount at a receiver, especially when the signal has just been attenuated by the medium through which it has traveled to the receiver by as much as 100 dB. After all, losing 3 dB of transmitter power means dropping the power by half, but what is 3 dB more loss when you have just lost 100 dB in the link?

As a result of this kind of thinking, plus the fact that in broadcast applications the power transmitted is up to a megaWatt, many receivers may have no more than a small piece of wire for an antenna, and this may be poorly matched in both wavelength and impedance to the receiver it is supposed to serve. Much of the time, the primary reason for any effort spent in designing an antenna for a receiver is that there happens to be a transmitter also attached to it. Fortunately, a sensitive receiver can operate well enough to satisfy all but the most critical applications, when its antenna is a window screen, a set of bedsprings, or a coat hanger. (Consider how much performance is lost.)

Here we will consider a world in which receiving antennas are well matched to their applications. What a concept!

10.1 POLARIZATION

Antenna polarity is an important consideration in a receiver, although in some cases it may not matter. The polarization used in many applications is vertical, not necessarily because it is better but because it fits in. For example, it is hard to imagine an AM broadcast band antenna lying flat on the ground (it would have to be twice as long) or elevated at the top of a mast, still horizontal and twice as long as what is now used. Remember that the low end of the AM band is 560 kHz, for which the wavelength is 536 m. Thus a half-wavelength horizontal antenna would be 268 m long.

Vertical antennas are no shorter than horizontal antennas, but a quarter-wavelength vertical antenna standing in a marsh (broadcast band antennas are not placed in marshes to save the environment) can be made to believe that it is a half-wavelength long. Not only that, but it is easy to connect the transmitter to the antenna, because it is standing on its low-impedance point, at ground level, and the coaxial cable from the transmitter does not have to be attached halfway up the tower. Figure 10.1 illustrates this condition, which is also used to advantage at much higher frequencies in high-speed aircraft. That is, a high-speed aircraft maker does not want to employ a half-wave antenna any more than a broadcaster does, and fortunately the aircraft skin is made to order as the high-conductivity ground plane. This brings up the question, what does the broadcaster do who does not have a bog in which to locate his antenna? He buries either a wire mesh or radial cables in the ground around his antenna for at least a quarter-wavelength radius, connected to the antenna feed line return.

What does this have to do with the receiver antenna? The receiver's antenna should be of the same polarity as the transmitter for best reception. Therefore, if the transmitter employs vertical polarity for any reason, then the receiver usually employs vertical polarity also. (This, however, is not always the case.) What other choices does the receiver have?

Four possible polarization choices are available (slant polarizations are not counted):

1. Vertical: Used in many applications for convenience.
2. Horizontal: Broadcast FM and TV. Also used for separation from vertically polarized signals.

3a. Right-hand circular: Broadcast FM. Satellite downlinks. Multipath rejection.

3b. Left-hand circular: Satellite downlinks. Multipath rejection.

It is possible to employ one polarization or another for the purpose of signal differentiation. That is, a receiver with a vertically polarized antenna will reject a signal that is horizontally polarized, and vice versa. The amount of rejection, optimally, is in the range of 20 dB, typically, if the two antennas are oriented exactly 90 degrees away from one another (orthogonal) *and* sufficient scattering has not occurred to destroy the orthogonality.

Similarly, right- and left-hand circularly polarized antennas display orthogonality under good conditions, with respect to one another's signals. In fact, under the right conditions, a circularly polarized receiving antenna having one polarity sees multipath signals in the opposite polarity, which allows it to discriminate against them.

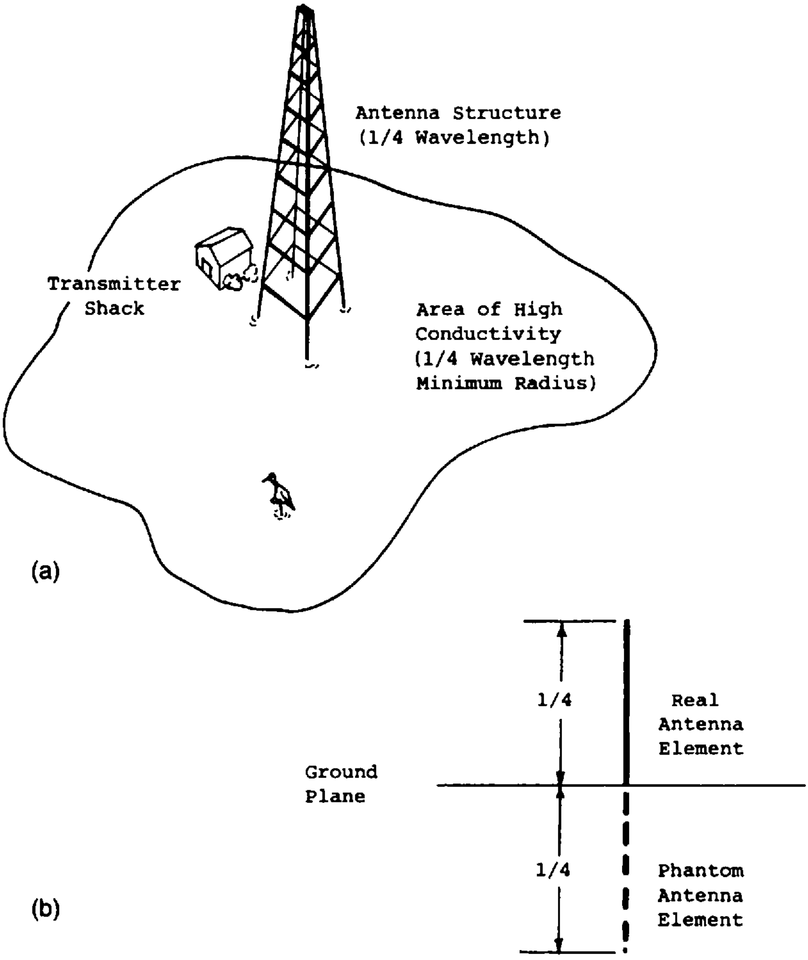


Figure 10.1 Illustration of 1/4 wavelength antenna over high-conductivity ground plane. (a) Typical broadcast antenna structure. (b) Electrical equivalent (behaves as 1/2 wave.)

There is also some advantage to having a transmitter send its signal in circular polarization, with its receivers using either vertical or horizontal (linear) polarization. Under these conditions, the receiver takes a 3-dB loss due to its polarization not being the same as that of the incoming signal, but the receiver cannot be placed in an orientation that is orthogonal to the incoming signal, so the loss is never more than 3 dB. This is a particularly valuable technique where the receiver is a handset with no guarantees about the way it is held by the user and should be employed for applications such as cellular and PCS systems.

Some antennas are capable of more than one polarization. An antenna with two orthogonal, linearly polarized sets of elements is sometimes used for diversity. That is, since multipath fading is not fully correlated in both vertical and horizontal polarities, it is possible to get some improvement in receiver performance by observing and making use of both polarities. Antennas with cross-polarized elements can also be used to either generate or receive circularly polarized signals.

From the standpoint of diversity, it should be realized that tests have shown that polarization diversity in personal communication systems has not proved to be as effective as spatial diversity. Therefore let us not celebrate this form (polarization) of diversity prematurely.

10.2 ANTENNA GAIN

All real antennas have some gain. This simply means that they do not transmit or receive exactly the same in all directions. An antenna that transmits the same amount of power evenly, or receives with exactly the same capability, in all directions is called an "isotropic" antenna. It cannot exist in the real world, if only because there is no way to feed a signal to such an antenna without affecting its distribution of transmitted energy. (The coax necessary to feed or take away the signal would affect the antenna pattern.) An isotropic antenna would, in effect, be a truly wireless antenna. Think about that.

At any rate, one of the two ways of stating antenna gain is in comparison to an isotropic antenna. For example, a half-wave dipole antenna has 2.1-dB gain in a direction perpendicular to its axis, when compared to an isotropic antenna, so the term dBi is used to describe its gain. The most common statement with respect to antenna gain is to state that an antenna has N dB gain, where it is understood that the gain is in comparison with an isotropic antenna, which has 0 dB gain.

The second way of stating antenna gain is to give the gain in comparison with a dipole antenna, in which case the gain is stated in terms of dBd. This is much less common, however, and sometimes leads to confusion, since dBd is a number that is 2.1 dB less than the dBi rating. Which one is used is of little consequence, as long as its use does not lead to confusion. In an R^2 propagation environment, the difference in range (for a 2.1-dB difference in link budget) is approximately a factor of 1.3.

Figures 10.2 to 10.4 show typical plots of antenna patterns for several commercially available "omnidirectional" antennas. The term omnidirectional as commercially used does not imply that the antenna is actually capable of transmitting or receiving in all directions equally. Instead, omnidirectional is conventionally used to mean that the antenna being described is omnidirectional in

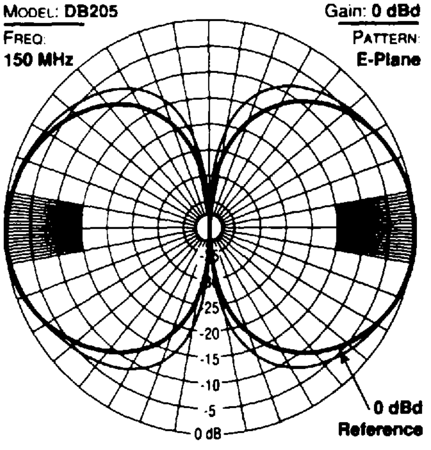


Figure 10.2 Unity gain antenna (0 dBd/2.1 dBi). Note agreement with theoretical pattern. Omnidirectional. (Courtesy of Allen Telecom Group Inc.)

azimuth and usually has a null in its pattern in the vertical direction. (These are called the "H plane" and "E plane," respectively.) Omnidirectional antennas typically cover a volume that is shaped like a doughnut or a bagel (depending on your preference). As the gain of an omnidirectional antenna increases, the vertical coverage decreases, which increases the radius of horizontal coverage.

If we start with a sphere of equal coverage, with the isotropic antenna, and

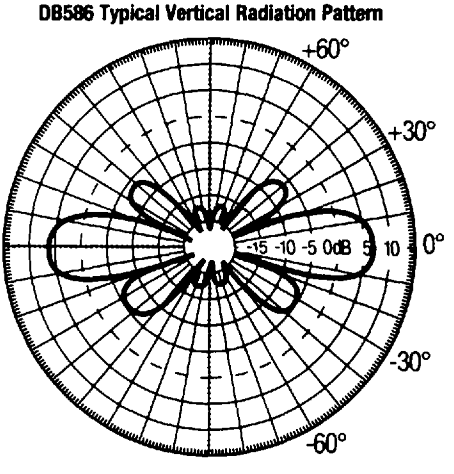


Figure 10.3 Three dB gain dipole antenna. Omnidirectional. (Courtesy of Allen Telecom Group Inc.)

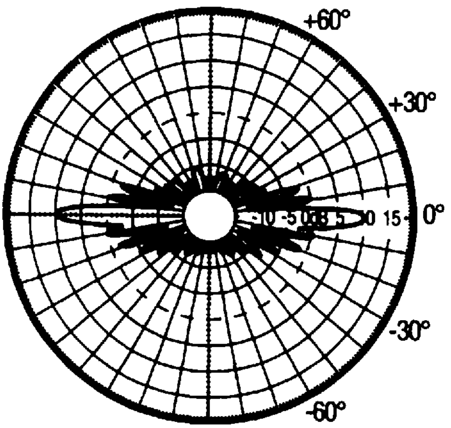


Figure 10.4 Ten dB gain dipole antenna. Omnidirectional. (Courtesy of Allen Telecom Group Inc.)

then reduce the spherical surface by cutting the sphere in half, the surface area is also reduced by half. But if the same amount of power is transmitted, then twice the power density will exist at the surface of the hemisphere as existed at the surface of the sphere. Therefore an antenna with the same surface area would see a 3-dB increase in its received signal power, and every time the covered area is decreased by a factor of 2, the signal power in the same antenna goes up by another 3 dB. Antenna gain, in general, is

$$G = 10\log\left(\text{Eff.} \times \frac{4\pi A}{\lambda^2}\right)$$

where A = antenna area
λ = wavelength, consistent with A
Eff. = antenna efficiency, typically 0.5 to 0.65

and the area of a half-wave dipole is approximately $0.13\lambda^2$. Parabolic antennas, which are common in reception of downlink signals from satellites and point-to-point earthbound systems, have a relationship between their gain and their beam-width that is

$$\theta = (27{,}000/G)^{1/2} \quad \text{or}$$

$$G = 27{,}000/\theta^2$$

Since many receivers have only a length of wire as their antenna, it is usual for broadcast stations to specify their "field strength," in their area of coverage, in terms of microvolts per meter. Thus, in an area of 100 μV/m, a 1-m antenna length should produce 100 μV at its receiver input (ignoring any questions of impedance matching).

Figures 10.5 and 10.6 show antenna patterns for directional antennas, for comparison with omnidirectional antennas. Directional antennas usually have higher gain, although omnidirectional antennas with up to 13 dBi gain are not unusual.

The effect of antenna gain is to increase the signal level seen at the input of a receiver. For all practical purposes, the increase is the same as if the transmitter power had been increased or the distance from the transmitter had been decreased, except that the angle over which the signal can be received may be significantly restricted. Figures 10.5 and 10.6 illustrate antenna patterns in which the antenna beamwidth is as narrow as a few degrees, and effective communications can occur only when the transmitter is within this restricted window (unless it is close enough to make up with power for being outside the beamwidth).

10.3 PORTABLE ANTENNAS

Portable antennas for receivers range from those that are found on the inside of transistor portable, AM/FM, broadcast band receivers to those that are found on portable television receivers. Unfortunately, the definition of a portable television has often meant any television set that has a handle on it (even if it weighs 900 pounds). We will also include as portable the antennas that are mounted on vehicles.

One of the most difficult of the many problems that exist with portable resistance applications is due to the fact that the wavelength of the signals that are most desirable to receive is much longer than the dimensions of the receiver (or even the vehicle in which it is mounted) will allow. Therefore, we are faced with the problem to which we have already alluded; the antenna must be small with respect to wavelength, and therefore signal reception must suffer.

Hand-held, portable, receivers are being used over a wide range of frequencies for telephone conversations, military command and control, entertainment, and many other applications that were not even conceived just a few years ago. The need to operate at higher and higher frequencies has in some ways been a boon to radio designers, because the antennas needed have become smaller. Unfortunately, they have not yet disappeared.

In a hand-held telephone, a two-way link is required. This is unlike most military systems, in which a time-shared one-way link suffices. In the military system, only one person talks at a time, so that only one frequency is needed. This also means that the antenna may be much narrower in bandwidth than in a simultaneous transmit/receive system. Military radio systems often employ tuned

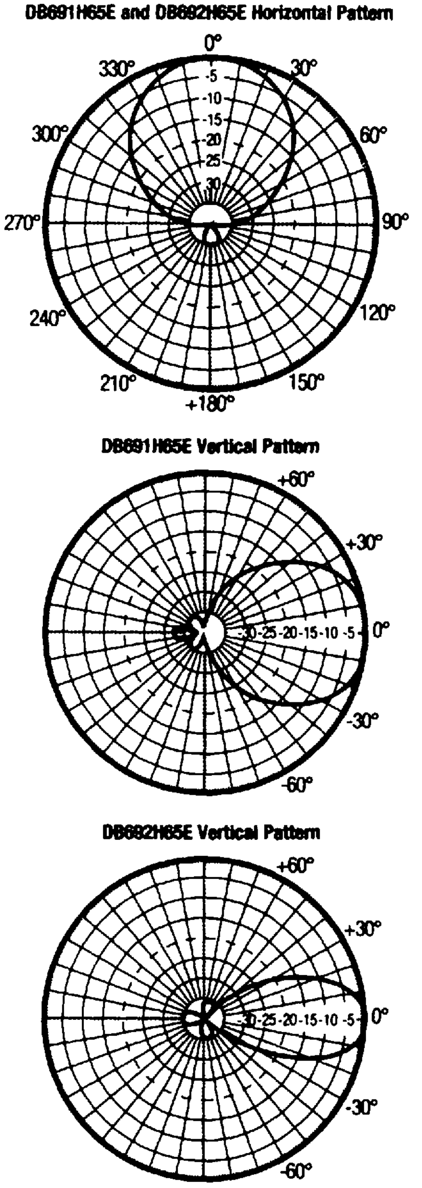


Figure 10.5 Directional antennas with 6 dB gain (middle) and 9 dB gain (bottom). Top pattern is the horizontal pattern for both. (Courtesy of Allen Telecom Group Inc.)

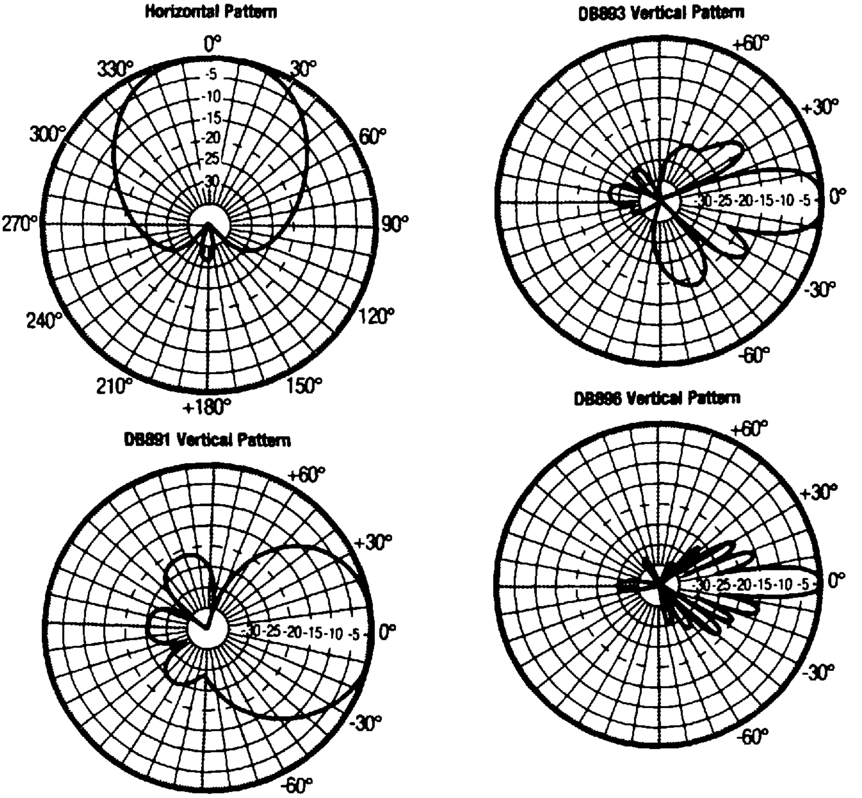


Figure 10.6 Directional antennas with 7.5 dB (bottom left), 12 dB (top right), and 15 dB gain (bottom right). Top left is horizontal pattern for all. (Courtesy of Allen Telecom Group Inc.)

antennas for two reasons: 1) Mobility requirements force the use of small antennas, and 2) the frequencies used are often in the lower bands (2–30 MHz and 30–88 MHz, for example). Hand-held telephones for nonmilitary applications operate in the 27- and 49-MHz bands (cordless telephones); the 824–849 MHz/869–894 MHz (cellular) bands; the 902–928 MHz instrument, scientific, and medical (ISM) band; and the 1850–1910/1930–1990 personal communications system (PCS) bands. (These are the U.S. bands, but similar allocations exist in other countries.)

Antennas for civilian hand-held telephones are not usually tuned (variable) because either the band allocation is so narrow that fixed tuning is acceptable or the frequency is high enough that the antenna's percentage bandwidth is no more than 10 percent. (In the PCS band, only 7 percent bandwidth is required, and only 8 percent is required for the cellular band.) Therefore, antenna tuning is not required.

The antennas employed in hand-held telephones are summarized in Table 10.1. (Remember that these antennas are employed in an application that includes a transmitter, and in some cases the transmitted output at 3 m from the antenna is only 50,000 μV per meter.) The rule for these is *whatever works*. In some cases, the antenna is a quarter-wavelength, and in others it is not. Obviously, at 27 MHz, where 1/4 wavelength is over 2.7 m, such an antenna is not practical on a hand-held device. The "rubber duck" antennas (flexible antennas typically made from speedometer cable) often used in portable/hand-held units certainly are not longer than one-tenth wavelength. The point is that *they do work*, although they might work better with a bigger antenna.

In the higher frequency bands, a quarter-wavelength or longer antenna is not out of the question. At 915 MHz (the center of the 902–928 MHz band) a quarter-wavelength is only 3.2 inches and in the PCS band (centered at 1920 MHz) only 1.5 inches. This makes it much easier to provide an antenna that is well matched to the application. Cellular telephones typically employ an antenna that is up to about 3 inches in length.

A significant problem in hand-held devices, with respect to their antenna, is the ground plane. Antennas are designed to behave as half-wave devices (see Figure 10.7). In a half-wave antenna, the current distribution is such that it is maximum at the center (also the feed point) and minimum at the ends. If the antenna were the same length but fed from the end, it would have maximum current at the input end, but minimum current at the opposite end, and behave like

Table 10.1 Antenna Summary for Hand-Held Telephones

Frequency range (MHz)	Telephone system	Antenna
27	Cordless, home use	Sectorized[a] Rod or flexible Short at wavelength
49	Cordless, home use	Sectorized[a] Rod or flexible Short at wavelength
902–928	Cordless, home use	1/4 to 1/2 wavelength[b]
824–894	Cellular telephone, public use	1/4 to 1/2 wavelength[b]
1850–1990	Personal communications, public use	1/4 to 1/2 wavelength[b]

[a]Partially retractable.
[b]Usually fully retractable.

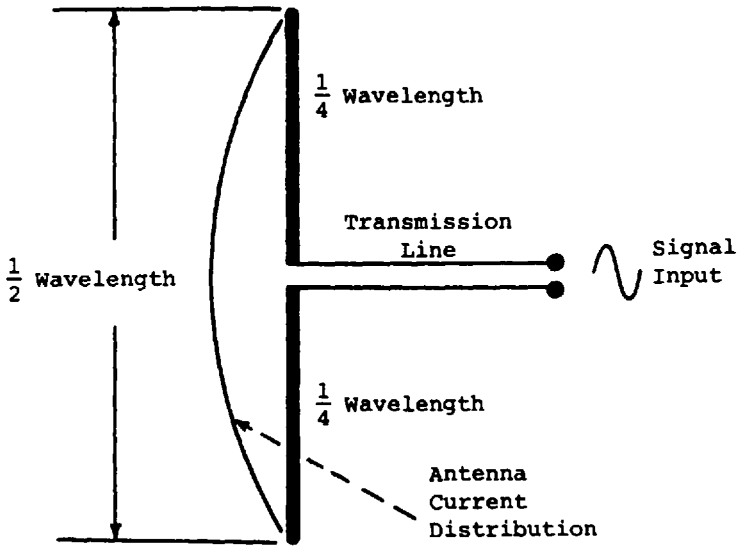


Figure 10.7 Illustration of half-wave antenna showing current distribution.

a quarter-wavelength at half the frequency. (Of course, this is not a strange condition to a receiver. Receivers must regularly contend with such conditions.) If a hand-held device presents an adequate ground plane to a quarter-wave antenna, then the other half appears as a reflection in the ground plane.

Unfortunately, very few hand-held devices can provide a satisfactory ground plane for an antenna, especially when a hand and a head are in close proximity to the hand-held device. Therefore, if possible, a half-wave antenna, center fed, is a much better choice. Figure 10.8 shows the construction of a raised-feed, half-wave antenna that is useful in hand-held devices—especially at the higher frequencies. This antenna is made by folding the shield of the coax back over its outer cover to a length of 1/4 wavelength and exposing the inner conductor for the same length. This provides what is in effect a half-wavelength, center-fed antenna, where the coaxial cable that feeds the antenna can be any length desired. This antenna does not require a ground plane and is minimally affected by proximity of hands and other body parts. (A human head in the vicinity of an antenna can attenuate the signal, received or transmitted, by as much as 20 dB.) The extended feed feature allows the active antenna portion to be raised, which may help to overcome head-blocking attenuation.

Other types of receiving antennas that have served well in portable applications are those used in automobiles, loops, and loopsticks. In automobiles, the

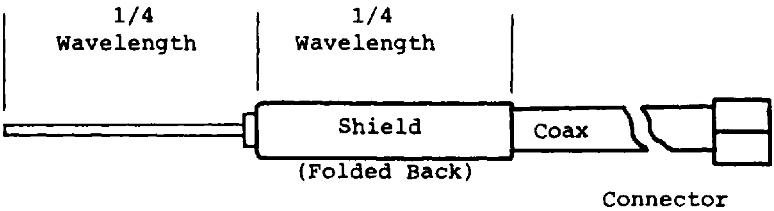


Figure 10.8 Half-wavelength extended-feed antenna structure. (Shown without molded cover.)

standard at the present time is a stainless steel whip that is usually mounted vertically near the windshield.* Its length is about 30 inches, typically, and it serves as both an AM and an FM antenna, even though the two bands are nearly six octaves apart. Experimental antenna systems for automobiles have been tested that provide a separate antenna at each corner of the vehicle, together with a sampling processor that chooses the best antenna at any given time. (This approach is intended primarily to solve multipath problems, particularly for the FM reception mode.)

Loop antennas were used in many AM receivers in the 1930s through 1950s. The loops themselves were no larger than about 8 by 12 inches and were wound on a flat piece of cardboard that also served as the back of the receiver cabinet. Both portable and other small but not so portable receivers used this type of antenna. It usually employed around 30 turns of wire, woven around the edge of the previously mentioned cardboard. This loop antenna also made the receiver direction sensitive. A connection was usually provided for a supplemental external long wire antenna, which solved the directionality problem.

A miniaturized version of the loop called a "loopstick," which is still in use, came into use in the 1950s. The loopstick is a ferrite core, usually much longer than its diameter, which is tuned by a coil of wire that is wrapped around the core and a tuning capacitor that is coupled to the receiver tuning mechanism. The permeability of the ferrite core reduces the size of the loop required and permits smaller receivers to be made, without significant loss in performance. Loopstick antennas are not used above a few MHz, however, because the materials needed to build them are not readily available. An illustration of that fact is seen in the area of AM/FM portable radio receivers: for AM reception, such receivers employ a loopstick, while they employ a multisection, extendable antenna for FM reception.

*Some autos have rear-mounted antennas, and some have them at a rakish angle.

10.4 FIXED ANTENNAS

Receiving antennas for fixed locations are a far cry from those that are needed for portability. (Dish antennas that are portable with a semitruck and trailer and take an hour or more to set up do not count as portable.) High gain and broad bandwidth are the stock in trade of fixed antennas, although there are plenty of fixed antennas that have neither high gain nor broad bandwidth. Some of the limiting characteristics of the antennas that are available are listed in Table 10.2.

The omnidirectional antennas have limited gain, primarily because they are not free to reduce their beamwidth in the H plane. However, they can provide gain of up to 14 to 15 dB while still maintaining full 360 degree coverage. Broad bandwidth and high gain are difficult to achieve at the same time in an omnidirectional antenna. One omnidirectional dipole antenna that is worthy of note is the type employed on high-speed aircraft, in the 225- to 400-MHz band. This antenna is a quarter-wave antenna, mounted on a ground plane (the aircraft's aluminum skin). Its bandwidth [(175/312.5) × 100 = 56 percent] is provided by shaping of the dipole itself, which has a length-to-width ratio of approximately 1:3. In other words, the antenna is a "fat" antenna. (Its thickness is no more than approximately 3/4 inch, which is the side presented as a wind load.) Because the antenna is a fat antenna, with a small length-to-width ratio, its inductance is low, and its bandwidth is much wider than that of a thin antenna of the same length. (If a dipole antenna is lacking in bandwidth, make it thicker.)

An example of a discone antenna is shown in Figure 10.9. This particular antenna is sold by Radio Shack as part number 20-013 and is made up of 16

Table 10.2 Characteristics of Some Common Fixed Antennas

Antenna type	Gain (dBi)	Beamwidth	Bandwidth
Omnidirectional			
Dipole	2.1	Up to 40°	Up to 60%
Stacked dipole	Up to 14	Down to 4°	Up to 23%
Discone	2.1	Up to 40°	Up to six octaves +
Directional			
Corner reflector	Up to 12	30°H/30°E	Up to 23%
Yagi array	Up to 12 (six element)	40°H (vertical polarity)	Up to 10%
Log-periodic array	Up to 12	40°H (vertical polarity)	Up to two octaves
Helix	Up to 15	40°H/40°E	Up to 10%
Parabolic	Up to 60	Down to 0.2°	Up to 10%

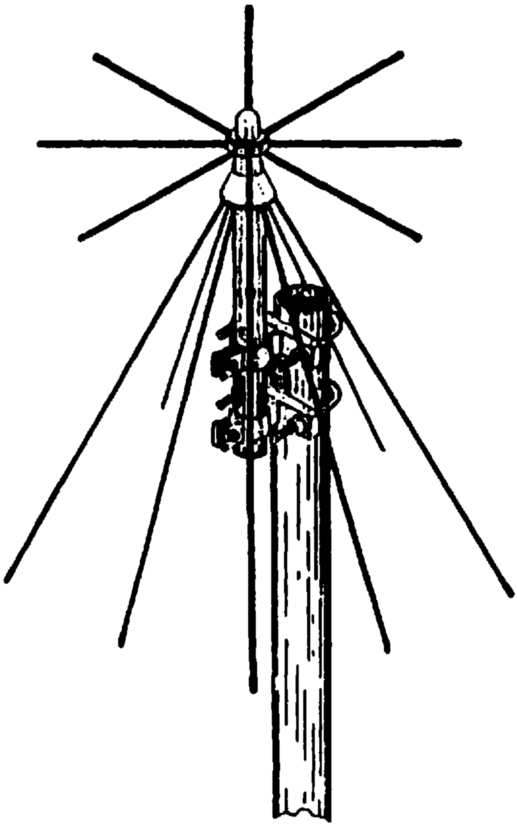


Figure 10.9 Radio Shack 20-013 discone antenna using stainless steel rods to construct the disk and cone. (Intended for use from 25 MHz to 1.3 GHz.) (Courtesy of Tandy Corp.)

elements to form a 22-inch disk at the top and a cone at the bottom with a diameter of 32 inches and a length of 29 inches. At higher frequencies, it is advantageous to construct the disk and the cone from sheet metal, such as brass. Either way, the discone antenna works well over a wide bandwidth. Discone antennas are vertically polarized.

Directional antennas of several types are listed in Table 10.2. One type worthy of mention that is not listed is the simple long wire antenna, which has been employed for many years for long-range reception. Long wire antennas are directional and exhibit gain in the direction perpendicular to the wire's axis and a null along its axis. Other directional antennas offer up to 60 dB gain. There exist antennas with even higher gain and arrays of antennas capable of both gain in the direction of a desired signal and steerable nulls that can be pointed in the direction of interfering signals. Here, only the most fundamental antenna types have been listed, for the purpose of showing the wide range of options open to use with a receiver.

It is common in base stations, especially at the higher frequencies, to employ more than one antenna for the base station's receiver(s) where only one is used for the transmitter. This is done to provide multipath mitigation for the handset-to-base path, as the handset transmitter has limited output power. In the other direction, base to handset, simple brute force power is employed, because it is usually difficult to place more than one antenna in a handset when the handset is only 2 to 4 inches long and the user will not tolerate more than one antenna sticking out of a handset. (For multipath mitigation, it is necessary to separate multiple antennas from one another by at least one-half wavelength.)

Such techniques are known as antenna diversity techniques and are commonly used. The GSM system, for example, typically employs two antennas for diversity in the handset-to-base link, where the handset transmitter ouput is a maximum of 2 W, and 20 W or more in the base-to-handset link, where there is no antenna diversity. This form of antenna diversity (more than two antennas can readily be used) has been shown to be more effective than the use of "rake" receiver techniques.

It is also practical, where more than one antenna is available, to employ the antennas for direction finding. Multiple antenna methods have been employed in military applications for many years and are becoming an interesting subject for use in location of 911 calls today.

Some of the antennas listed may seem very exotic, but almost everyone has experience with array antennas and parabolic dishes, which are used every day for television reception. Helix antennas are also employed for consumer applications in many areas of the world. (Helix antennas have circular polarization; all of the other antennas listed have either vertical or horizontal polarization only.)

10.5 ANTENNA TESTING

Directionality and gain testing of antennas is best performed in an anechoic chamber, where the antenna's pattern is not affected by its surroundings. Some rudimentary testing can be performed by comparing an antenna with another antenna whose characteristics are known, but unless an environment is available wherein the antenna being tested is not affected by objects within its field, significant test data errors are possible. The anechoic chamber serves the purpose of isolating the antenna being tested from its surroundings.

VSWR testing of antennas is important testing that can be conducted to determine not only the impedance match but also the frequency range over which

Table 10.3 Return Loss and Voltage Standing Wave Ratio (VSWR)

Return loss (dB)	VSWR	Loss due to VSWR (dB)
∞	1.0:1	0.0
26.4	1.1:1	0.01
20.8	1.2:1	0.04
17.7	1.3:1	0.07
15.6	1.4:1	0.12
14.0	1.5:1	0.18
12.7	1.6:1	0.24
11.7	1.7:1	0.3
10.9	1.8:1	0.37
10.2	1.9:1	0.44
9.5	2.0:1	0.51
8.3	2.25:1	0.7
7.4	2.5:1	0.9
6.6	2.75:1	1.07
6.02	3.0:1	1.25

the antenna works well. For this test, an anechoic chamber is not required, as long as the antenna does not have significant objects within its near field.* A hand, for example, on or near an antenna can distort its pattern and change its VSWR readings.† VSWR is normally measured as a function of "return loss" for an antenna. This is the measure of the forward power (to the antenna) versus reflected power (from the antenna). Ideally, an antenna should accept all of the signal power input to it and pass it on as electromagnetic energy, except for what is lost as heat. If the antenna is perfect, no power will be returned to the source. Therefore a high return loss figure signifies an antenna that is doing something with the input power, with the assumption that it is being transmitted. VSWR is the ratio of returned power to incident power, so the lower the ratio the better. Table 10.3 is a listing of return loss versus VSWR comparisons. In many applications, a VSWR of 2.0:1 is the maximum that is considered to be acceptable, and a plot of antenna return loss as a function of frequency, where return loss of 10 dB is the threshold value, is used to determine an antenna's frequency range. Figure 10.9 is a typical network analyzer plot of antenna S_{11} or return loss, using a Hewlett-Packard HP8753C network analyzer.

Observing Figure 10.10, in which an antenna for the 1850- to 1990-MHz

*An antenna's near field is the region contained by the radius $0.4D^2/\lambda = A/2\lambda$, where D is the largest dimension of the antenna, λ is the wavelength, and A is the effective area of the antenna. The effective area of a half-wave dipole is $0.130\lambda^2$ and of a short dipole is $0.119\lambda^2$.

†A hand with rings or watches may be significantly worse.

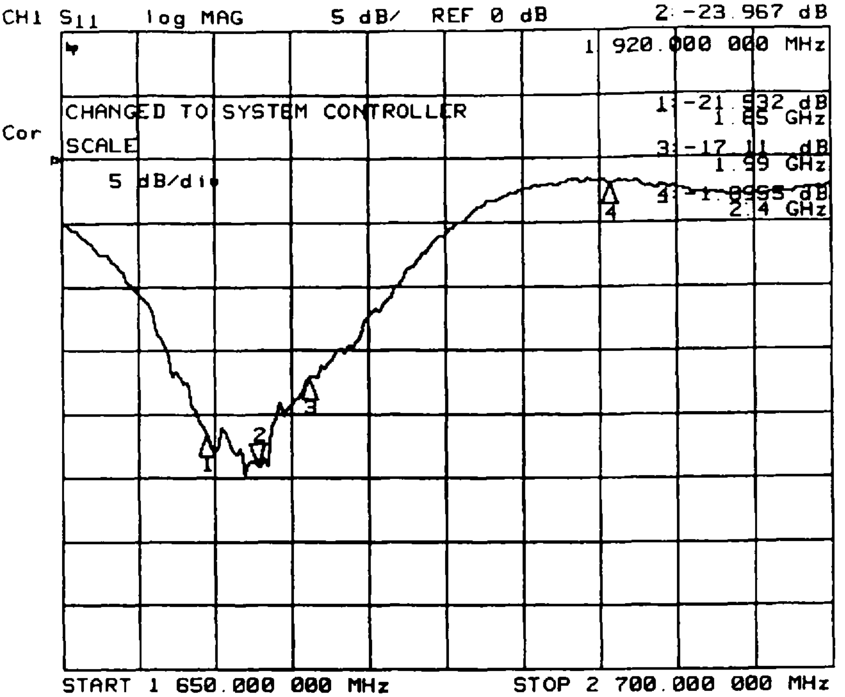


Figure 10.10 Plot of return loss versus frequency for an 1850- to 1990-MHz PCS band antenna.

band was being tested, we see that this particular antenna had a return loss of 21.532 dB at the lower end of the band and 17.11 dB at the upper end. At the 10-dB return loss crossovers, we see that this antenna could be used over a 360-MHz range, with its center at about 1940 MHz.

Return loss is related to VSWR as follows:

$$\text{VSWR} = \frac{1 + \rho}{1 - \rho}$$

and

$$\rho = \log^{1-}\frac{\text{return loss}}{20}$$

where ρ is the reflection coefficient, which is

$$\frac{V\ \text{reflected}}{V\ \text{incident}} = \sqrt{\frac{P\ \text{reflected}}{P\ \text{incident}}}$$

Also, return loss = 20 log ρ, the percentage of power transferred is $100(1 - \rho^2)$, and the loss of transferred power due to VSWR is

$$10 \log (1 - \rho^2) \text{ dB}$$

Antenna tests are usually made with a signal source, looking into the antenna. Here, we are interested in looking from the antenna into the receiver. The same mismatch and VSWR considerations hold. That is, if a given return loss is seen due to a mismatch at the receiver input, then the same amount of loss will occur in the antenna-to-receiver input as would occur in a signal generator-to-antenna output.

Two test setups for measuring return loss are shown in Figure 10.11, in

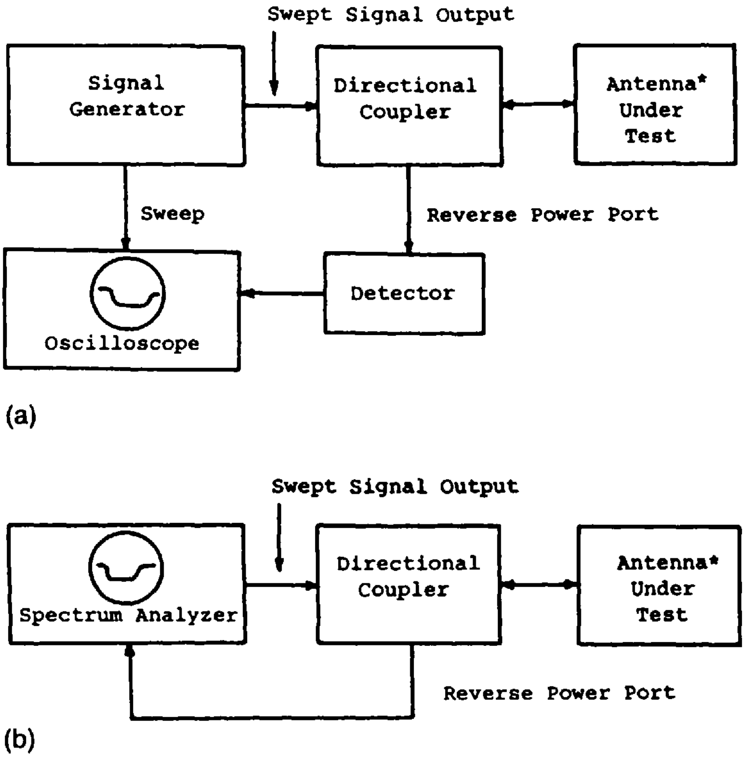


Figure 10.11 Two test setups for measuring antenna VSWR. Also useful for other VSWR testing. (a) Test setup for VSWR test without network analyzer or spectrum analyzer. (b) Test setup for VSWR test with spectrum analyzer that has coupled sweep output. (*Or other device.)

antenna measurement configuration. They are also useful for measuring return loss of other sections of a receiver, such as the preamplifier or IF amplifiers. Another variation of Figure 10.11a is to replace the detector and oscilloscope shown with a spectrum analyzer that does not have a coupled sweep output, but it uses the spectrum analyzer's built-in functions to replace those separate test units.

In calibrating any of these test setups, the following calibration process or its equivalent should be employed:

1. Remove the antenna or other device being tested, and replace it with a load that is of the correct impedance.
2. Calibrate the oscilloscope or spectrum analyzer for zero return loss.
3. Remove the load from the connector that would normally be connected to the antenna or other device under test, and replace it with a calibrated open circuit.
4. Calibrate the oscilloscope or spectrum analyzer for maximum return loss. (This will be the full output of the signal generator, minus losses in the directional coupler.)
5. Replace the open circuit with the device being tested, and measure the return loss as function of frequency.

QUESTIONS

1. Give at least two reasons why antennas tend to be vertically polarized.
2. How much loss can be expected (if any) when receiving a vertically polarized signal with a horizontally polarized antenna?
3. What if the receiving antenna is circularly polarized?
4. What is the mechanism that produces antenna gain?
5. What is meant by the term 10 dBi? 10 dBd?
6. How is wavelength related to the size of an antenna?
7. When testing the reflected power of an antenna, which is better, 3-dB return loss or 15 dB? Why?
8. How does ground conductivity affect an antenna?
9. If an antenna has 10-dB gain and is omnidirectional in the H plane, what is its approximate beamwidth in the E plane?
10. If an antenna does not have enough bandwidth, what would you do to increase it?

REFERENCES

1. Stutzman, W. L., and Thiele, G. A., *Antenna Theory and Design*, John Wiley & Sons, New York, 1981.

2. *Reference Data for Radio Engineers*, H. W. Sams, 1975.
3. Base Station Antennas, Catalogue 220/94, Celwave.
4. Base Station Systems Equipment, Allen Telecom Group, Decibel Products Division (Catalog).
5. Wide Band Antennas and Control Systems, Tecom Industries, Inc. (Catalog).
6. Clingerman, D. A., "Lab Notes" (Periodical), The Olde Antenna Lab of Denver.
7. Orr, W. I., *Radio Handbook*, Editors and Engineers Division of H. W. Sams, 1978.
8. Carr, J. J., *The TAB Handbook of Radio Communications*, TAB Books, 1984.
9. Facts about Proper VHF Vertical Antenna Design, Advanced Electronic Applications, Inc., 1996.
10. Cunningham, J. E., *The Complete Broadcast Antenna Handbook*, TAB Books, 1977.
11. Langford-Smith, F., *Radiotron Designer's Handbook*, RCA, 1960.
12. Hawthorne, D. L., Polarization Measurement, Parts 1 and 2, Watkins-Johnson Tech-Notes, Nov.-Feb. 1987.
13. Lee, W. C. Y., *Mobile Communications Engineering*, McGraw-Hill, New York, 1982.
14. Jakes, W. C., Jr., *Microwave Mobile Communications*, John Wiley & Sons, New York, 1974.

11
Implementation with Digital Methods

For many years, long before the development of microprocessors, designers have dreamed of building receivers with digital techniques that would relieve them from the woes of analog design. In fact, the ideal has been a receiver that consists of an antenna, an analog-to-digital converter, and a signal processing computer. (Sometimes an amplifier is grudgingly added after the antenna in this dream.)

Today, the latest in cellular telephones has made a giant step in the direction of digital implementation. That is, their receivers consist of an antenna, a pre-amplifier, down-converter, IF amplifier, and an analog-to-digital converter, with the rest of the receiver being implemented with a digital processor. Sooner or later, it will be practical to build almost completely digital receivers that do not have to down-convert before the A/D conversion process, at least up to operating frequencies in the GHz range.

After all, it is possible to do almost anything with digital processes that can be done with analog processes. The only question is, how accurately must the result of the digital process represent the analog product of the analog process that it is replacing? The success of digital compact discs (CDs) seems to confirm that the process used in that particular area has done very well, for example. However, there exists a group of people (audiophiles) who claim that, for them, the final analog result of the analog-to-digital recording and digital-to-analog playback is not acceptable.

An answer, then, to the question of whether or not a digitally implemented receiver can replace an analog receiver is another question: Does the digital receiver perform the same functions as well as the analog receiver? If so, will it fit in the same box for the same price? If the answer to even one of these questions is yes, then we will most certainly see fully digitized* receivers in the near future.

11.1 ANALOG-TO-DIGITAL CONVERSION

The analog-to-digital conversion process is dependent on sampling its analog input at a rate that is sufficient to preserve all of the information that is to be recovered when converted back to an analog output signal. To do this, it is necessary to sample the analog input signal at a rate that is at least twice the highest frequency in the bandwidth of the signal to be recovered. (This is a restatement of the Shannon/Nyquist sampling theorem, which states that "If a function of time $f(t)$ contains no frequencies higher than W hertz, it is completely determined by giving the value of the function at a series of points spaced $1/2W$ seconds apart.")

Figure 11.1 illustrates this by showing the spectra generated in the sampling and recovery process. For this purpose, the sampling process is assumed to use pulse-amplitude modulation for simplicity, although any other process would be equivalent. PAM is the case in which the amplitude of a series of pulses at a rate equal to the sample rate chosen is set to the same level as the analog signal being sampled, at the sample time. This produces amplitude modulation at the sample rate, which is a double-sideband signal, and as long as the sample rate is much higher than the baseband (as in Figure 11.1a) there is no interference between the two. It is important to realize that it is only the higher frequency signal that is transmitted to the receiver and that the baseband signal is not sent.

If the sample rate is conveniently at the desired frequency, then f_s modulated by the baseband could be transmitted, but in most instances the sampled signal would be converted to a new frequency. At the receiver, the signal (now an analog carrier pulse-amplitude modulated by the sampled baseband signal) can be converted to baseband (or IF) either by a conventional down-conversion process or by sampling.†

Because of the sampling theorem, it is often considered that a signal at 500 MHz (for example) must be sampled at least 1 billion times per second. This is not the case, however, unless one wishes to reproduce the 500-MHz carrier from the digitized, sampled version of the signal. If one wishes only to reproduce the

*Fully digitized means here that the analog-to-digital conversion process occurs directly at the operating frequency, although there may be other processing such as amplification ahead of the A/D converter.

†In a receiver, an analog-to-digital conversion process is down-conversion. Just as a product detector mixes a signal to baseband, A/D conversion takes an input signal at an RF or IF frequency and produces a baseband data stream.

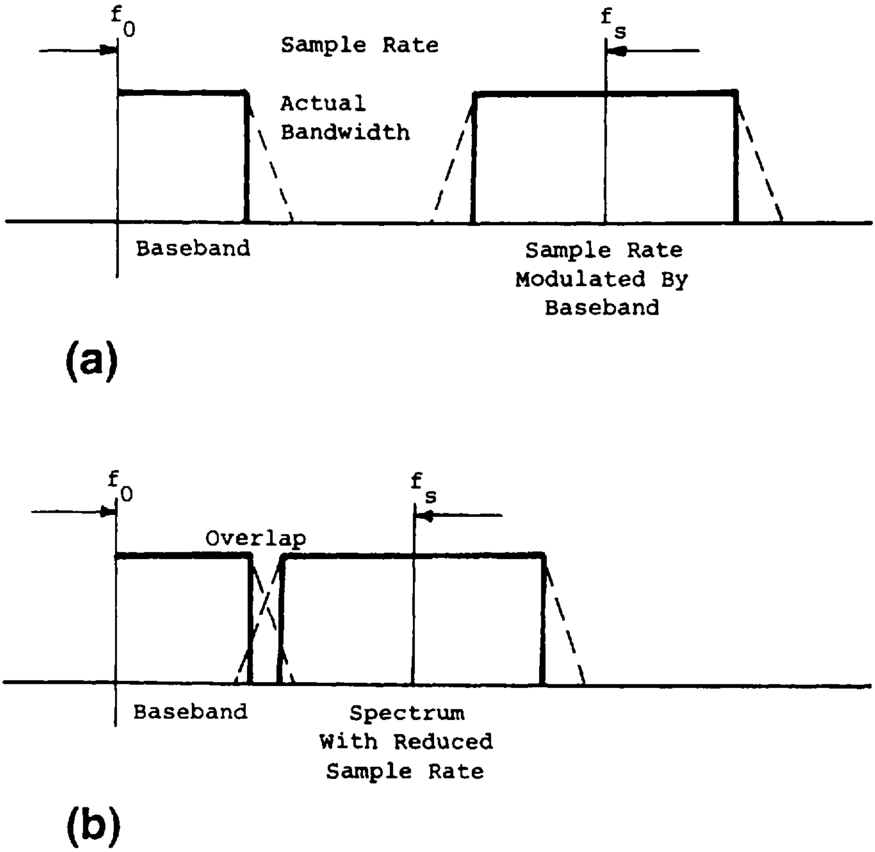


Figure 11.1 Illustration of Shannon/Nyquist theorem showing spectral overlap due to filters and limitation that does not permit sampling at rate less than twice the baseband bandwidth. (a) Spectrum showing baseband information and baseband modulated sample rate. (b) Spectrum showing overlap of spectra as sample rate. Approaches 2 × baseband bandwidth.

modulating information conveyed by the 500-MHz carrier, it is necessary only to sample at twice the rate of the modulating signal, and this itself is an advantage ... the modulating information is produced in its sampled form at baseband. The sampling process then obviates the need for a separate down-converter, but it is advantageous to sample at the highest frequency that is practical.

The next question that presents itself is, why not sample at the operating frequency? Today, the answer is that we can do exactly that, up to at least 900 MHz. Figure 11.2 shows an integrated 1000 megasample per second analog-to-

SPT
SIGNAL PROCESSING TECHNOLOGIES

SPT7760

8-BIT, 1000 MSPS FLASH A/D CONVERTER

3

FEATURES

- 1:2 Demuxed ECL Compatible Outputs
- Wide Input Bandwidth - 900 MHz
- Low Input Capacitance - 15 pF (MQUAD)
- Metastable Errors Reduced to 1 LSB
- Monolithic for Low Cost
- Gray Code Output

APPLICATIONS

- Digital Oscilloscopes
- Transient Capture
- Radar, EW, ECM
- Direct RF Down-Conversion

GENERAL DESCRIPTION

The SPT7760 is a full parallel (flash) analog-to-digital converter capable of digitizing full scale (0 to -2 V) inputs into eight-bit digital words at an update rate of 1000 MSPS. The ECL-compatible outputs are demultiplexed into two separate output banks, each with differential data ready outputs to ease the task of data capture. The SPT7760's wide input bandwidth and low capacitance eliminate the need for external track-and-hold amplifiers for most applications. A proprietary decoding scheme reduces metastable errors to the 1 LSB level. The SPT7760 operates from a single -5.2 V supply, with a nominal power dissipation of 5.5 W.

The SPT7760 is available in a 68L PGA and an 80L surface-mount MQUAD package over the industrial temperature range of -25 to + 85 °C. Contact the factory for availability of die and /883 versions.

BLOCK DIAGRAM

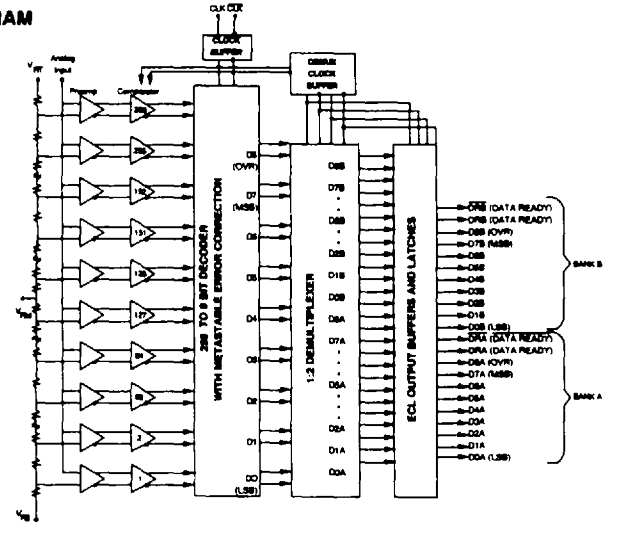


Figure 11.2 Eight-bit, 1 gigasample per second analog-to-digital converter. (Courtesy of Signal Processing Technologies.)

digital converter, whose input bandwidth is rated at 900 MHz. This means that this device can sample a signal at a frequency as high as 900 MHz, up to 1 billion times per second, and produce 8 bits per sample. (The output data word is eight parallel bits per sample, with each bit stream at 1 gigabit per second, maximum.)

This device could operate at an IF or RF frequency as high as 900 MHz or with a modulating frequency as high as 500 MHz. (It cannot do both, as the upper sideband of a 900-MHz carrier with 500-MHz modulation would stretch to 1400 MHz.) If we assume that modulation rate is not greater than 10 MHz, then a carrier as high as $900 - 10 = 890$ MHz is conceivable. (This integrated circuit could work in the cellular telephone range.) For higher frequencies, a down-conversion can be employed to bring the signal within the operating frequency range.

For modulation rates up to 10 MHz, only 20 megasamples per second (MSPS) are required, and even with four-times oversampling, 80 MSPS. Because many applications require much narrower bandwidths (GSM, for example, has a modulation rate of only 270.8 kbps), the sampling rate needed for many receivers is far below the rate that could be achieved. Why would one choose an A/D converter that is so fast, when the speed is not required for the modulation rates of most applications?

With a 1 gigasample per second sampling rate, the SPT7760* must complete all of the process necessary to output an 8-bit representation of the signal being sampled, within $1/(1 \times 10^9) = 1.0$ nsec. Since its input bandwidth is rated at 900 MHz, and a 900-MHz signal has a period of $1/(9 \times 10^8) = 1.11$ nsec, a sample is taken and the digitized output generated in less than a cycle of the input signal. This is more than adequate to prevent an ambiguity in the sampling point, although there is a fixed time delay between the sampling clock and the data (plus/minus jitter, or time uncertainty). The short interval over which a sample is taken serves to prevent the sample from being averaged across a range of values of an RF carrier that is changing continuously at a very high rate.

For the A/D converter that we have discussed, the SPT 7760, the input signal can vary over a range of 0 to -2 V DC, which could reach an RF level as great as $+10$ dBm in a 50-ohm system. This device, however, cannot handle such large signals at the higher frequencies. In fact, the dynamic range listed is approximately 35 to 40 dB. With 8-bit resolution and a 2-V input range, the number of steps that can be resolved is $2^8 = 256$, which gives a resolution of $2/256 = 7.81$ mV. This suggests that a receiver with 1-μV sensitivity would need

$$20 \log \left(\frac{7.81 \times 10^{-3}}{2.83 \times 10^{-6}} \right) = 68.8 \text{ dB}$$

*Figure 11.2.

ANALOG DEVICES

12-Bit, 41 MSPS Monolithic A/D Converter

AD9042

FEATURES
41 MSPS Minimum Sample Rate
80 dB Spurious-Free Dynamic Range
595 mW Power Dissipation
Single +5 V Supply
On-Chip T/H and Reference
Twos Complement Output Format
CMOS-Compatible Output Levels

APPLICATIONS
Cellular/PCS Base Stations
GPS Anti-Jamming Receivers
Communications Receivers
Spectrum Analyzers
Electro-Optics
Medical Imaging
ATE

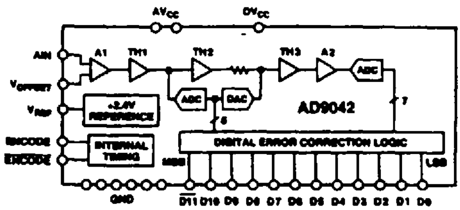
PRODUCT DESCRIPTION

The AD9042 is a high speed, high performance, low power, monolithic 12-bit analog-to-digital converter. All necessary functions, including track-and-hold (T/H) and reference are included on chip to provide a complete conversion solution. The AD9042 runs off of a single +5 V supply and provides CMOS-compatible digital outputs at 41 MSPS.

Designed specifically to address the needs of wideband, multichannel receivers, the AD9042 maintains 80 dB spurious-free dynamic range (SFDR) over a bandwidth of 20 MHz. Noise performance is also exceptional; typical signal-to-noise ratio is 68 dB.

The AD9042 is built on Analog Devices' high speed complementary bipolar process (XFCB) and uses an innovative multipass architecture. Units are packaged in a 28-pin DIP; this custom cofired ceramic package forms a multilayer substrate to which internal bypass capacitors and the 9042 die are attached and a 44-pin TQFP low profile surface mount package. The AD9042 industrial grade is specified from –40°C to +85°C. However, the AD9042 was designed to perform over the full military temperature range (–55°C to +125°C); consult factory for military grade product options.

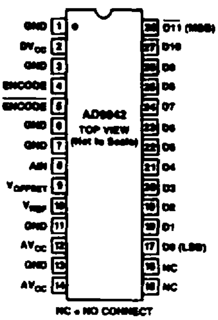
FUNCTIONAL BLOCK DIAGRAM



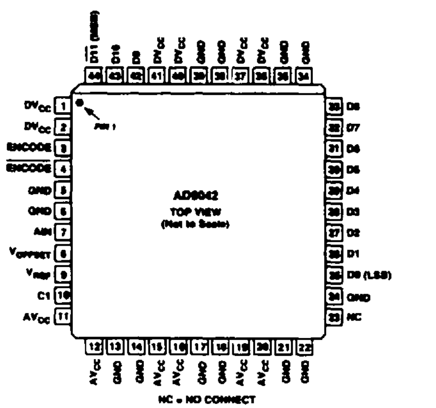
PRODUCT HIGHLIGHTS

1. Guaranteed sample rate is 41 MSPS.
2. Dynamic performance specified over entire Nyquist band; spurious signals typ. 80 dBc for –1 dBFS input signals.
3. Low power dissipation: 595 mW off a single +5 V supply.
4. Reference and track-and-hold included on chip.
5. Packaged in 28-pin ceramic DIP and 44-pin TQFP.

AD9042AD PIN DESIGNATIONS



AD9042AST PIN DESIGNATIONS



REV. A

One Technology Way, P.O. Box 9106, Norwood, MA 02062-9106, U.S.A.
Tel: 617/329-4700 Fax: 617/326-8703

Figure 11.3 Twelve-bit, 41 megasample per second A/D converter. (Courtesy of Analog Devices.)

gain ahead of the A/D converter. This is a very good reason to employ frequency conversion in a digitized receiver, since 69-dB gain is not easy to get at a single high frequency.

In a receiver that must work with signals that have a range of 1 μV to 1 V (120 dB), either the A/D converter must meet the dynamic range requirement or there must be some gain control ahead of the A/D converter to bring the signal within its signal-handling ability. Such an A/D converter would need as much as $\log_2 (1 \times 10^6) = 20$ bit conversion capability. Figure 11.3 shows a 12-bit, 41-MSPS A/D converter with 80-dB dynamic range up to 20 MHz.* This device could work well in a receiver that has down-conversion to a 10.7 MHz IF, 1.0 V peak-to-peak maximum IF output, and 70 μV minimum IF output. (To operate at 1 μV, the receiver would need at least $20 \log 70 = 36.9$ dB gain ahead of the A/D converter and to operate at 1.0 V, the ability to reduce that gain to zero.)

The ideal A/D converter would have the ability to operate at a receiver's input RF frequency, which requires that its input circuits must be designed to match the impedances found at the antenna and/or amplifier interfaces. It should have sufficient dynamic range to work with the signal levels input to the receiver or with the input signals after they have been conditioned by preamplification, by gain control, and by filtering. The A/D converter should sample the analog signal within a period that is very short compared with the period of the input signal. The sampling rate of the ideal A/D converter would be sufficiently higher than the information bandwidth to allow oversampling of the signal by as much as 10 times, if desired. Finally, it should do all of these things on minuscule power and require a single low-voltage supply.

An analog-to-digital converter and the digital-to-analog converter needed at the end of all of the digital processing (in a receiver with digital output, the D/A converter is not needed) are the only differences between an analog receiver and a digital receiver from the standpoint of the tasks they must carry out. All of the other functions must be present in either analog or digital receivers. Table 11.1 lists these functions and the methods used to implement them in digital and analog receivers.

Remember that a digital receiver is one that has been implemented with digital techniques, not a receiver for receiving "digital" modulation. Both analog and digital receivers are intended to receive analog signals, whether the information being sent is analog information or digital data, and the purpose of the A/D converter is simply to change the form of the signal to be processed by the receiver from an analog signal to a digital approximation to it, so that digital techniques can be used. If the digital approximation of the signal is adequate and the digital processing is accurate, then the final output may be acceptable.

*It should be noted that the A/D converters shown in Fig. 11.2 and in Fig. 11.3 require too much operating current to be employed in hand-held, battery-powered receivers.

Table 11.1 Comparison of Analog and Digital Receiver Requirements and Techniques

Requirement	Analog technique	Digital technique
Frequency selection	LO and RF tuning	Sampling rate and filter coefficients
Unwanted signal rejection	RF and IF filtering	Sampling rate and filter coefficients
Wide dynamic range	Low noise figure and high IIP3	Increased bits per sample
Demodulation	Modulation-specific demodulator	I and Q signal processing
Down-conversion	Mixing with LO	Sampling and A/D conversion
Amplification	Analog level increase by amplification	Not always equivalent

11.2 DIGITAL EQUIVALENTS OF ANALOG PROCESSES

There exists for all of the analog processes that go on in a receiver a digital process that can do the same thing if implemented correctly. Table 11.1 lists some of the analog and digital tasks that are required and the equivalent processes by which they are carried out. This list covers the major tasks expected of a receiver.

11.2.1 Frequency Selection

The first task expected of a receiver is to select the signal or signals that fall within a specific frequency range. The ability of a receiver to select a particular signal by its frequency is limited by several factors:

1. The frequency accuracy of the transmitter
2. The frequency accuracy of the receiver
3. The fact that that there may be more than one signal at the selected frequency
4. The inability of the receiver to reject signals in adjacent channels

Frequency accuracy on both analog and digital receivers depends on the accuracy of their reference oscillator, which is typically crystal stabilized. Neither of the two has an inherent advantage with respect to the oscillators themselves. However, there may be an advantage for a digital approach in that the down-conversion process employed in a typical analog receiver involves phase lock multiplication of a receiver's frequency source, whereas the sampling rate needed by a digital receiver usually does not require such multiplication. Although a phase lock multiplier uses the same reference and has the same offset in parts per million as the sampling clock, it causes more frequency error because its signal is used to translate the signal to an IF frequency, and its entire multiplied offset is transferred to the signal at IF.

One problem with frequency selection in digital receivers is that it is very difficult at the higher frequencies to build a digital filter that is useful at the RF input frequency. Lack of a tuned filter at the receiver RF input (or at the output of the preamplifier) is one of the primary reasons for intermodulation and premixing in receivers today. Of course, if there is more than one signal in the input frequency band, there may be no way to separate the desired signal from the undesired one(s), unless the signals are some form of multiple access signal* especially designed for recognition and demodulation in the presence of other signals.

Adjacent channel rejection is a function of the IF filter or filters in an analog receiver. A digital receiver that converts to baseband directly from the RF input frequency without converting to an IF must do all of its filtering at baseband, with a low-pass filter. While excellent low-pass filters can be provided with digital techniques, the problem in digital receivers is to prevent an adjacent channel signal from being mapped into the low-pass bandwidth by the analog-to-digital conversion process. For example, if the adjacent channel signal is presented to the A/D converter along with the desired signal, then it will also be translated to baseband, appearing as shown in Figure 11.4. Figure 11.5 shows the harmonic spectra that result when an adjacent channel interferor is present, compared with that which is present when there is no adjacent channel signal.

Observing Figure 11.4, we see that it does not matter whether an interfering adjacent channel signal is on the high side of the signal being sampled or the low side; at baseband the interferor appears above the baseband, because the process of sampling folds the desired signal at zero frequency, just as it would if the signal were simply being multiplied with a local oscillator at the RF frequency. Figure 11.5 shows the harmonic structure due to sampling and illustrates that when adjacent channel signals are present, one should sample at a minimum rate of six times the bandwidth of the information (assuming the adjacent channels are from identical systems) and preferably higher. In other terms, remembering that the Nyquist criterion demands sampling at a rate at least twice the bandwidth of the information being sampled, to avoid possible intermodulation between the baseband adjacent channel signal and the lower side harmonic of the sampled signal, h_1 should be at least six times m_1. Further, if the sampling rate h_1 is not at least four times m_1, then the lower side harmonic of the sampled interference will fall into the information baseband bandwidth.

---

*Two forms of multiple access signals are particularly well suited to use in digital receivers: 1) time division multiple access, or TDMA, and 2) code division multiple access, or CDMA.

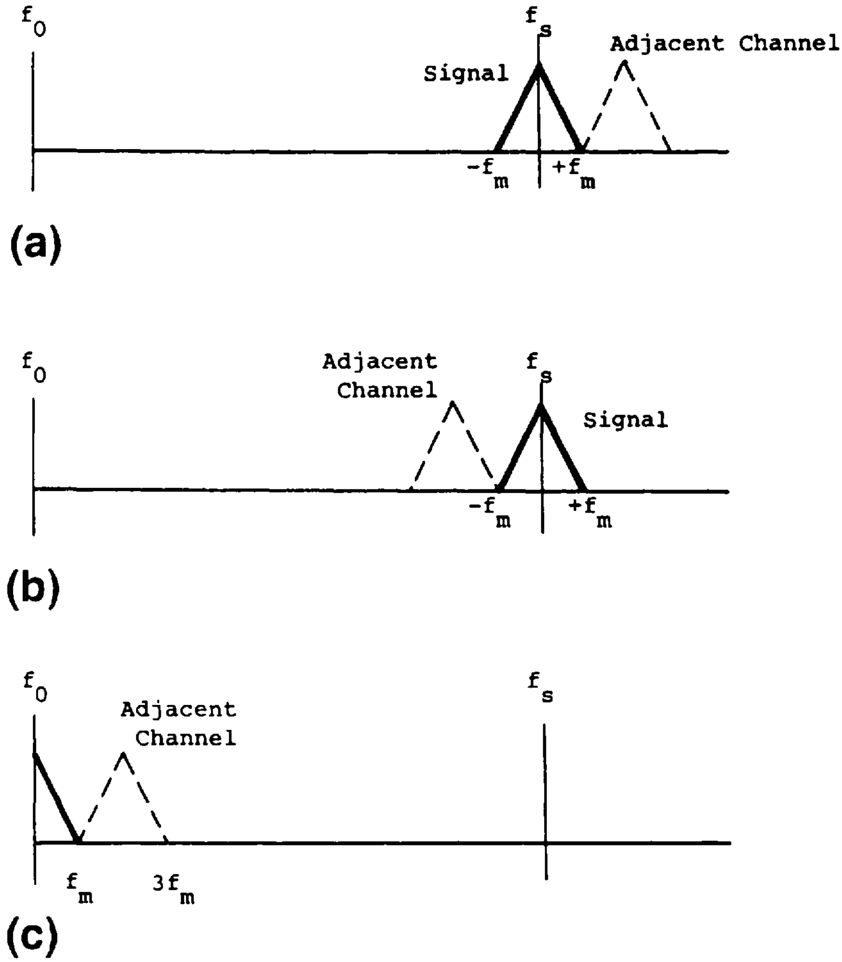


Figure 11.4 Illustration of translation to baseband by sampling at a rate higher than $2 \times 3f_m$ but much less than f_s. (a) Signal with adjacent channel interference on high side. (b) Signal with adjacent channel interference on low side. (c) Information at baseband, with adjacent channel signal, due to sampling.

A safe rule would be to sample at 10 times the information bandwidth, if at all possible. This would reduce the possibility of aliasing (the correct term for interference between signals produced by sampling) and allow more separation between sampling products, which makes the task of filtering easier.

Imperfect sampling (not at the correct rate) produces the same kind of problem that mixing with an incorrect local oscillator produces in a product

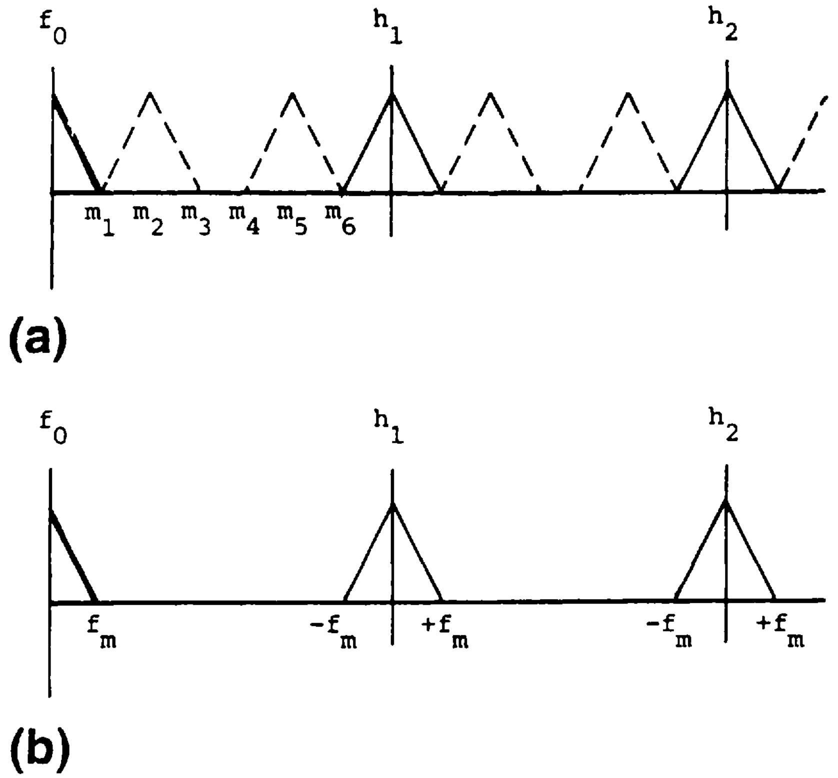


Figure 11.5 Illustration showing spectrum due to sampling, with and without adjacent channel interference. (a) Spectrum due to sampling with adjacent channel interference present. (b) Spectrum due to sampling without adjacent channel interference.

detector or zero IF down-conversion. That is, the spectrum does not fold over around zero, and the digital equivalent of a zero beat occurs and causes the sample points to rotate at the rate of the error in the sampling frequency. This means that in-phase and quadrature processing is required.

Low-pass filtering to remove the effects of adjacent channel interference should be implemented as part of the digital I and Q processing.

11.2.2 Unwanted Signal Rejection

Rejection of adjacent channel signals was discussed in the previous section. Other unwanted signal responses have been discussed in almost all of the preceding chapters. Digitization in a receiver is not a solution to spurious responses that are produced by premixing, intermodulation in the receiver preamplifier, double

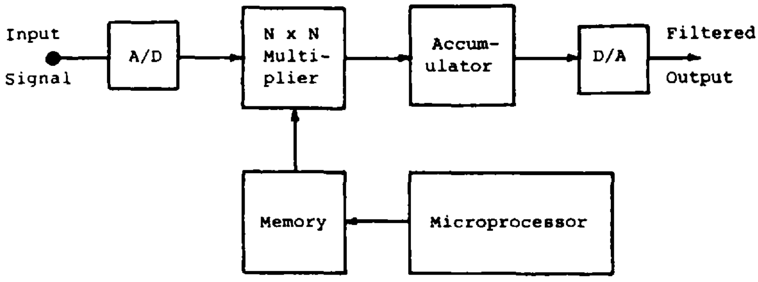


Figure 11.6 Generic digital filter.

beats, or any other such response that is a function of the linearity or band rejection characteristics of the receiver, unless the operating frequency is relatively low (no higher than a few hundred MHz).

Digital filters excel at the lower frequencies but are not capable of performance equaling that of analog filters at frequencies above the VHF band, and they will never match the low power consumption capability of analog filters, because analog filters consume no operating power at all. A generic block diagram of a digital filter is shown in Figure 11.6. Analog signals are converted to digital signals by the A/D converter and multiplied by the filter coefficients stored in memory. (The filter coefficients are a stored, digitized replica of the desired signal.) The order of the coefficients and their rate are determined by the microprocessor. After multiplication, the product is input to an accumulator and converted back to an analog signal. (If further digital processing were to be done, D/A conversion would not be carried out until the digital processing was completed.)

Other forms of digital filters exist, one of which is illustrated in Figure 11.7. (The analog equivalent of this filter is shown in Figure 11.8.) The problem with digital filters at the higher frequencies (a few hundred MHz and higher) is that the sampling rate necessary to implement a bandpass filter at RF is twice the RF frequency. It is one thing to sample a signal at 1 GHz, for example, when you are processing only the information that is being conveyed by the 1-GHz carrier, and it is quite another to sample at the rate necessary to process the carrier itself. The difference in sampling rates is easily an order of magnitude or more, as the carrier must be sampled at least 2 billion times per second, while the rate required to sample the information and recover it from the carrier would seldom be higher than 2 million samples per second. The sampling rate for a standard cellular telephone signal at 825 MHz would be in the range of 60 to 100,000 samples per second, but a filter at 825 MHz would require sampling at more than 1650 million samples per second.

Figure 11.9 illustrates the type of performance that can be achieved in

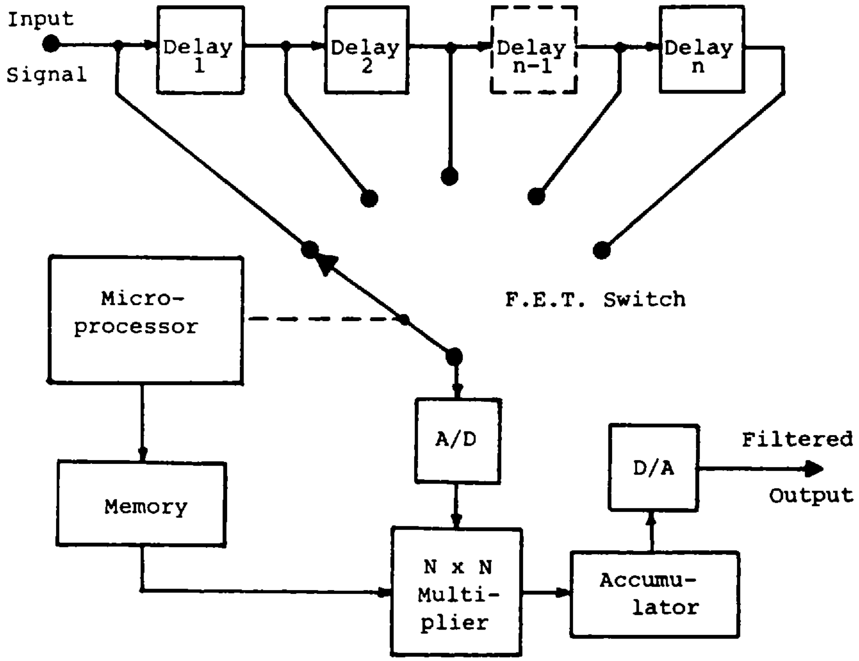


Figure 11.7 Digital finite impulse response filter. Delay elements may be analog or digital (such as a shift register).

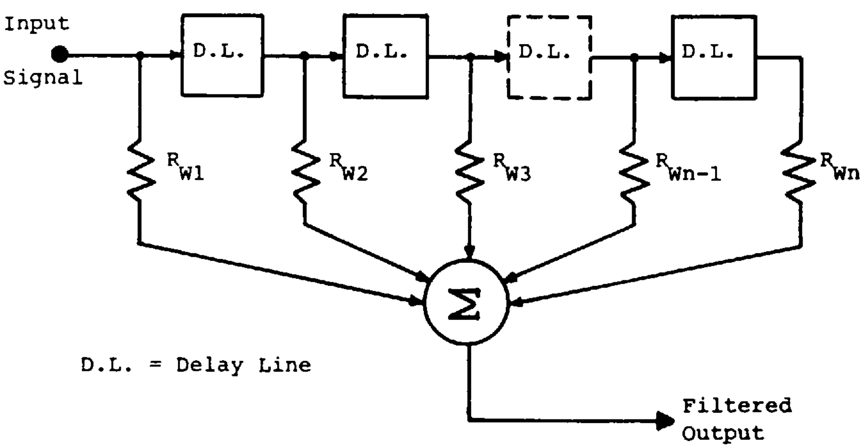


Figure 11.8 Analog equivalent of digital finite impulse response filter (FIR), called a transversal filter.

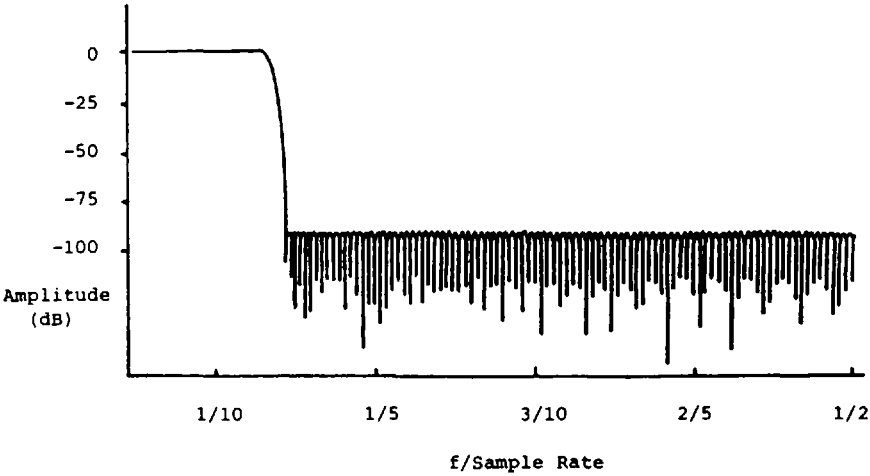


Figure 11.9 Plot of response of digital low-pass filter.

digital filters. The filter response shown is for a low-pass filter, but similar performance is possible in bandpass and high-pass filters as well. Such response is but one of the capabilities that make digital filtering attractive for receiver applications; the other is the possibility of having filters that are flexible enough to have bandwidths that may be adapted for a particular use, where changing the filter bandwidth can be as simple as using a different clock rate in the digital processor. The barrier to using such digital filters in receivers is the sampling rates necessary for their application at higher RF frequencies. Today, analog filters are necessary for unwanted signal rejection at RF, and usually at IF frequencies as well.

11.2.3 Oversampling and Undersampling

The terms used to describe the sampling rate of a digital subsystem with respect to the minimum or Nyquist rate are "oversampling" and "undersampling." It is common to oversample or employ a sampling rate that is higher than twice the highest frequency in the band of interest. A good example is the technique employed in telephone applications, in which a voice signal with 3-kHz bandwidth is sampled at 8000 samples per second (one third faster than the minimum of 6000 samples per second) and 8 bits are sent per sample. The primary reason for oversampling is that practical filters cannot prevent the lower sideband of a sampled signal from interfering with the upper edge of the baseband (as illustrated in Figure 11.1) because their rolloff in response is not fast enough to prevent aliasing unless the sampling rate is increased. A second reason for oversampling is that in some cases oversampling can be employed to improve the resolution of a digitized signal for which it is possible to sample at an adequate rate but it is not practical to provide the number of bits per sample that is implied by the dynamic range of the signal being sampled.

Undersampling is the name given to the technique used in sampling the information directly from a carrier without bothering with the carrier frequency itself. (Actually this technique saves a lot of bother, functionally.) That is, the samples are not taken at the Nyquist rate with respect to the carrier but do meet the requirement for the information bandwidth of the information being conveyed by the carrier. By sampling at a rate lower than that required to digitize the carrier, the carrier is suppressed. This does not mean that the information desired is not being oversampled, however. At the same time, a modulated signal may be both under-sampled and oversampled; it is undersampled with respect to the carrier and oversampled with respect to the information being conveyed.

11.2.4 Multipliers and Mixers

The sampling process in digital receivers may in some part do away with the down-conversion process as we know it today, at least up to the frequency at which it is no longer economical to do the necessary sampling and A/D conversion. After all, even in this day and age when integrated circuitry seems to be so advantageous that nothing cannot be built with it and there seems to be no reason to employ anything but integrated circuits, there are many consumer products (especially receivers) that still employ discrete transistor designs. Therefore it is doubtful that digital techniques will ever completely take over the world of receiver design. (And even if they do, there will still be a group of people who stoutly maintain that analog receivers sound better. Further, they will also maintain that analog receivers that employ vacuum tubes are best of all. In some ways they are correct.)

Digital multipliers can perform many of the functions that mixers and multipliers perform in receivers today, although it is not generally advantageous to use a digital multiplier unless it is part of an overall digital implementation. To replace a mixer with an A/D convertor, a memory with local oscillator coefficients stored in it, a multiplier, an accumulator, a controller for the memory, and a D/A converter would be lunacy unless there were some completely separate, overriding reason to do so. If all of the rest of the receiver is digital, then it is perfectly logical to do the multiplication digitally, rather than convert to analog, mix signals, and then convert back to digital.

Digital multiplication is a simple process. It consists of taking two input words in binary form; let us say word P and word R, where word P is added $R - 1$ times to generate a product

$$P \times R = \sum_{Q=0}^{R-1} P + PQ$$

When in a receiver, the multiplication process is often used in conjunction with an addition process, wherein the product of a given multiplication is added to the previous cumulative result. This is the equivalent of the analog process of mixing and filtering, which we find everywhere in receivers. Digital multipliers can also be valuable components for use in demodulators and can serve as phase detectors or as the DC-coupled third multiplier in Costas or phase lock loops.

11.2.5 Wide Dynamic Range

Signals whose amplitude varies over a 120-dB or greater range are common in mobile receivers, whereas those that reside at a fixed location may never see more than about 20 to 40 dB variation in signal level. This means a great deal of difference in the design of the two types of receivers, where the mobile receiver's signal range implies that a 20- to 21-bit-per-sample A/D converter may be required, but the fixed receiver needs only 7- to 14-bit-per-sample A/D conversion. Of course, if the receiver does not require linearity (FM receivers often have limiting IFs), then A/D conversion requirements may be reduced.

One method for reducing the dynamic range requirement for an A/D converter is to employ AGC in a preamplifier and/or an IF ahead of a receiver's A/D converter. This is one of the reasons why digitized commercial telephones in the cellular band use single conversion; the dynamic range problem is greatly alleviated as far as the A/D converter is concerned. This approach does not help the preamplifier, mixer, or first IF amplifier with respect to their dynamic range, however. That is, they can also employ AGC, but with the same problems that cause us to investigate switching from analog to digital processes in the first place.

11.2.6 Demodulation

Digital implementation of demodulators ranges all the way from replacement of analog phase detectors with exclusive-OR gates to embedding complete phase lock and Costas loops in software. In some cases, it is difficult to see why one would want to replace an analog demodulator with a digital version. (For example, why replace a simple envelope detector with a digital process?) One of the best answers of all is, for flexibility!

Many demodulators consist of multiply-and-add operations (phase detect-and-integrate in the analog world) that lend themselves readily to digital implementation. For example, AM can be demodulated by converting the analog AM signal to digital, squaring the digital signal by multiplying it by itself, converting the squared signal back to analog, and low-pass filtering the result. (There was no

Table 11.2 Digital Demodulation Techniques for Receivers

Signal modulation	Analog demodulator	Digital process
Amplitude modulation		
Standard	Envelope detector	Sample, digitize, square, D/A convert, filter
SSB	Product detector	Sample, digitize, multiply with digitized carrier, D/A convert, filter
DSB	Product detector	Sample, digitize, multiply with digitized carrier, D/A convert, filter
Frequency modulation[a]	Quadrature detector	Sample, digitize, multiply with 90 degree phase shift (delay of digitized signal), D/A convert, filter
	Phase lock loop	Sample, digitize, multiply with digitized carrier, calculate phase error, adjust carrier phase,[b] output phase correction to D/A converter, filter
Amplitude shift key	Same as AM	Same as AM
Frequency shift key	Same as FM	Same as FM
Phase shift key[a]	Squaring loop	Sample, digitize, multiply with digitized 2× carrier, calculate phase error, adjust carrier phase,[b] divide carrier by two, multiply divided carrier with digitized input signal, D/A convert, filter

[a]Not all possible demodulators listed.
[b]Loop filter included as phase and phase rate control.

guarantee that it would be as simple as an envelope detector.) Similar methods can be employed for other types of modulated signals. Table 11.2 lists some of the possible approaches to implementing digital demodulators, for different modulation schemes, in receivers.

11.2.7 Down-Conversion

The specific digital version of the down-conversion process used in analog receivers would include sampling the input RF signal, digitizing the samples, multiplying the digitized input with a digitized local oscillator, digitally filtering the result (another multiply-and-add operation), and outputting the filtered signal as either an analog or a digital signal. All of this is not necessary, because the sampling and digitizing process, using undersampling as previously described, produces the digitized RF modulating signal at baseband as its normal product.

Down-conversion is not required if the sampling and digitization process is done directly at the RF input frequency.

The alternative being employed today in commercially available products is to amplify the input signal, down-convert it with an analog mixer, amplify it again at an IF frequency, and then convert it to a digital signal at baseband by sampling and digitizing. This increases the sampling duration requirement and makes the sampler less difficult but does not change the minimum sampling rate. It may, however, if gain control is included in the RF and IF amplifier chain, reduce the dynamic range required of the A/D conversion process.

11.2.8 Amplification

Digital amplification is a conceptually viable operation that is as simple as multiplying each digitized sample of a signal by a fixed binary number. (An $n \times n$ multiplication gives a product that has up to $2n$ bits. That is, a signal digitized to 16 bits and multiplied with a 16-bit word would produce a multiplier output of up to 32 bits.) However, when the input signal is as small as a microvolt, then the input signal is 2.8 μV peak to peak, and the A/D conversion process should have resolution capable of recognizing a change of perhaps 10 percent of that level. In other words, if the signal increases by 1 dB, its amplitude would increase to 3.14 μV peak to peak, a change of $3.14 - 2.8 = 0.34$ μV, approximately 12 percent. Therefore, resolution to around 0.3 μV would be reasonable to seek as a goal, and 16-bit A/D output would permit a dynamic range of

$$0.3 \times 10^{-6} \quad \text{to} \quad 2^{16} \times 0.3 \times 10^{-6} \text{ Volts}$$

a factor of $2^{16} = 65{,}536$, or $10 \log 65{,}536 = 48.2$ dB.

On the other hand, analog amplifiers are readily available whose dynamic range is much higher and whose noise performance is better.

11.3 COMPARISON OF DIGITAL AND ANALOG METHODS

Receivers have been built for many years using analog components and techniques, and it is difficult to meet the same performance specications with a newer approach. Nevertheless, digital techniques are known to work at the lower frequencies and offer great promise for application at higher frequencies. The primary drawback with digital receiver design today is the unavailability of digital circuits that can operate at the higher frequencies, with power requirements low enough to allow them to operate in hand-held and/or battery-operated receivers. An A/D converter that requires an ampere or more of power supply current is not a viable solution in a system that has batteries whose capacity is 1 ampere-hour or less. Gallium arsenide or perhaps silicon-germanium integrated circuits may offer

Table 11.3 Comparison of Analog and Digital Receiver Subsystems

Function	Analog performance	Digital performance	Winner
Antenna matching	Excellent matching over wide bandwidth.	Many possible undesirable effects.	Analog
Tuning at RF	Practical within limits of Q. Fixed tuning used at higher frequencies, varactors at VHF and low UHF.	Digital filters not practical today at higher frequencies.	Analog
Amplification	Noise figure in 1 to 2 dB range with gain of 20 dB practical; IP3 at +19 dBm.	Does not compute.	Analog
Down-conversion	Conversion gain of −10 to +15 dB common. Active and passive mixers are available.	Except for undersampling, not practical above VHF.	Analog, digital possible
IF filter and amplifier	Up to 80 dB gain, with gain control and excellent passive filters.	If undersampling, IF filtering, amplification, and gain control carried out at baseband	Digital, to low VHF
Frequency synthesis	Phase lock synthesizers useful up to 20 GHz and higher.	Direct digital synthesis impractical above low VHF.	Analog
Demodulation	Different demodulator needed for each class of modulation.	I-Q digital demodulators can provide demodulation of a wide range of signals.	Digital

some relief, but they have not done so at this time. New materials and techniques will undoubtedly be found that will ease the difficulty, but at the same time they will also make analog circuitry better.*

Subsystem comparisons between the analog and digital methods for receiver applications are listed in Table 11.3. Observation of the "win" column might lead one to believe that if a vote were taken and each subsystem listed had only one vote, then analog implementation would surely win. This would be an

*Distinctions between analog and digital circuits and signals blur as digital rates and analog frequencies come closer together. Only a little distortion is needed to make an analog signal at a GHz or higher indistinguishable from a digital square wave at the same rate. Therefore the same engineers may very well do both jobs, and the traditional separation between circuits may well fade away. Furthermore, digital signals at a GHz look suspiciously like sine waves with a few phase inversions and a few missing cycles thrown in.

erroneous conclusion, because the real promise of digital receiver design is in its flexibility, which is something that is very difficult to achieve in analog receivers without physically duplicating filters, demodulators, and other circuits that are necessary to change from one type of modulation and/or frequency to another. On the other hand, digital filters and demodulators may change their operating parameters by reprogramming the algorithms and the clock rates employed in a receiver. Therefore a properly designed digitally implemented receiver should be able to cover a wider range of frequencies and demodulate more different signals than has ever been practical with a completely analog receiver.

In practice, it is hard to imagine that receivers will ever be completely digital, as some devices (such as antennas and their matching, or low-noise amplifiers) simply "do not compute."

Today, and for some time to come, receivers will consist of at least an antenna with its matching circuits, an amplifier, an A/D converter, and a digital processor, in its most futuristic digital guise. Other receivers must pick and choose from their menus to select the best possible performance at the lowest cost, and if flexibility is not a paramount requirement, it is doubtful that digital implementation will be an important consideration. One very practical thing could change this outlook ... the development of complex, low-cost, integrated circuit components that require extremely low power supply current and are very simple to use. Not only that, but they must work just as well as the circuits they replace and be transparent to the user.

That is, the developer of these revolutionary receivers and their circuits must perform a heroic task for which the ultimate compliment will be that the user does not notice it and for which the user will pay less than he ever did before for a comparable device. At best, the user (or consumer if you prefer) may somehow be convinced that the new technology is better than the old, but that does not mean that he or she will pay more. After all, today's cellular telephone, which is a marvel of technological achievement, is something for which the user either has paid nothing or perhaps has received payment to induce him or her to accept it.

QUESTIONS

1. What is the minimum sampling rate for a 3-kHz voice signal, according to the Shannon/Nyquist sampling theorem?
2. What is the minimum sampling rate for the 3-kHz voice signal if it happens to be present as modulation on a 1-MHz carrier, using amplitude modulation?
3. Why is it not practical to employ the minimum sampling rate in real applications?

4. Given a signal with a 43-dB dynamic range of operation, how many bits per sample are required to represent the signal to 1 bit accuracy?
5. Explain the difference between a zero-IF receiver (a receiver that converts its input RF signal directly to baseband) and a receiver that employs an A/D converter at the RF input.
6. Why does a D/A converter generally employ a low-pass filter at its output?
7. Why does an A/D converter require a low-pass filter at its input?
8. What serial data rate would be required for a processor following an A/D converter whose input signal is 1.0 VRMS maximum, whose resolution is 1 mV, and whose input signal bandwidth is 100 kHz? (Assume oversampling of one third.)
9. What is meant by the term oversampling? Undersampling? What advantages and disadvantages are there to each? (Undersampling is sometimes referred to as subsampling.)
10. What parallel data rate would be required by the processor of question 8?

REFERENCES

1. Rabiner, R. R., and Gold, B., *Theory and Application of Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1975.
2. Oppenheim, A. V., and Schafer, R. W., *Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1975.
3. Bozic, S. M., *Digital and Kalman Filtering*, Edward Arnold, London, 1979.
4. Schweber, B., Converters restructure communication architectures, *EDN Magazine*, Aug. 3, 1995.
5. Bucklen, W., Eldon, J., Schirm, L., and Williams, F., Designer's guide to: Digital signal processing—Parts 1 through 5, *EDN Magazine*, March/April/May 1981.
6. Nardi, B. J., Digital Signal Processing for Multichannel Receiving Systems, Watkins-Johnson Tech-Notes, May/June 1982.
7. Moore, S. E., Microwave Sampling for Ultra-Broadband Frequency Conversion, Watkins-Johnson Tech-Notes, Nov./Dec. 1983.
8. Dexter, C. E., and Poffenberger, A. W., Software Design of Microprocessor-Controlled Receivers, Watkins-Johnson Tech-Notes, Parts 1 and 2, Nov.-Jan. 1985.
9. Shapiro, A. L., Rapid Broadband Signal Acquisition, Watkins-Johnson Tech-Notes, Mar./Apr. 1989.
10. Goldman, S., Understanding the limits of quadrature detection, *Microwaves & RF*, Dec. 1986.
11. Won, M. S., and Crome, C., Building FIR Filters in Programmable Logic, *Embedded Systems Programming*, Aug. 1996.
12. Hosking, R. H., *Digital Receiver Handbook*, Pentek Inc., 1996.
13. Brannon, B., Wide-dynamic-range A/D converters pave the way for wideband digital-radio receivers, *EDN*, Nov. 1996.

14. Wepman, J. A., and Hoffman, J. R., RF and IF Digitization in Radio Receivers: Theory, Concepts, and Examples, NTIA Rpt. 96-328, Mar. 1996.
15. Brannon, B., Using wide dynamic range converters for wide band radios, *RF Design*, May 1995.
16. *Digital Signal Processing Solutions Selection Guide*, Texas Instruments, 1996.
17. Aschwanden, F., Direct conversion—how to make it work in TV tuners, *IEEE Trans. Consumer Elec.*, Aug. 1996.

12
Level Diagrams in Receiver Design

This chapter presents a method for developing detailed specifications for a receiver design, along with its subsystems, where only the most basic requirements are given. The process employed is to begin with a block diagram and develop the characteristics required of each block (subsystem) based on its signal processing requirements in conjunction with other blocks. Then, as each block is defined, a detailed specification for its development can be put together.

We will assume that a receiver is to be designed whose sensitivity is 1.0 μV or less, that has not less than 10 dB output $(S + N)/N$ ratio, with AM modulation at 30 percent, and a modulation bandwidth of 5 kHz. An alternative considered is a mode in which the receiver must receive BPSK data at a 38.4 kbps rate, with a maximum bit error rate of 1×10^{-4}. Again, 1 μV sensitivity or better is assumed.

In both the AM and data modes, the maximum input signal is assumed to be at least 1 V with specified operation and 3 V without damage.

Operating frequency for the receiver is assumed to be 800 to 900 MHz.

12.1 RECEIVER LAYOUT FROM A SPECIFICATION

Figure 12.1 is an initial block diagram layout for a receiver to meet the requirements listed above. A dual-conversion receiver architecture has been chosen because of the frequency range and the amount of gain that is needed. (At least 107 dB of RF and IF gain is needed if the receiver's demodulator is to work at a 0 dBm input level.) Automatic gain control is also included, as limiting is not desirable with AM signals, and gain control is advisable in the preamplifier even in receivers that are allowed to limit.

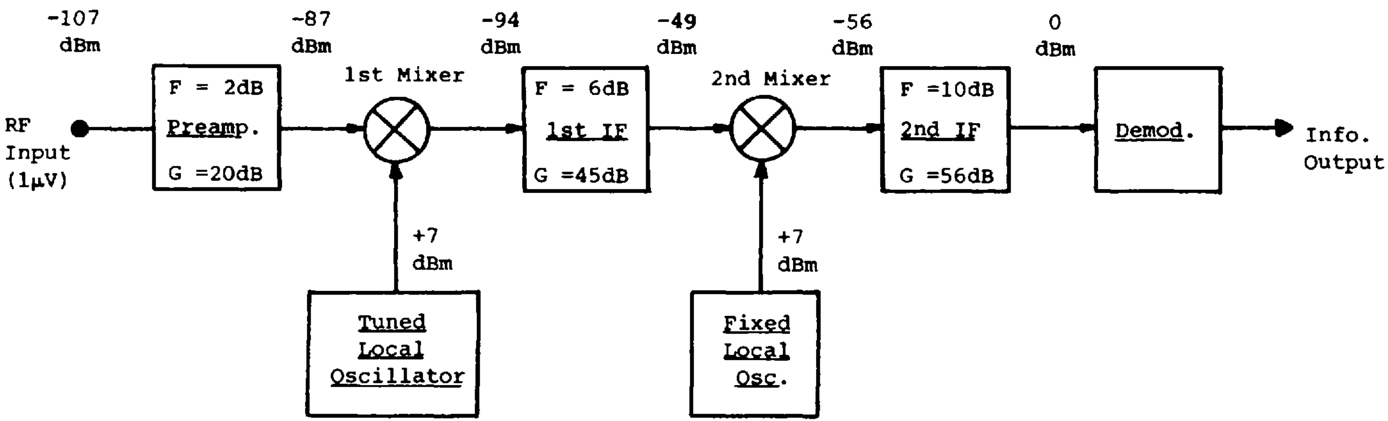


Figure 12.1 Double-conversion receiver layout, with signal levels for minimum RF signal input.

12.2 SIGNAL LEVEL PROJECTION

Using the nominal gain numbers from Figure 12.1, the input and output signal levels become

	Input (dBm)	Output (dBm)
Preamplifier	−107	−87
First mixer	−87	−94
First IF	−94	−49
Second mixer	−49	−56
Second IF	−56	−0
Demodulator	0	

when the RF input to the receiver is the minimum that the receiver is designed to receive (1.0 μV = −106.9897 dBm). Under these conditions, only the output from the second IF must have a compression point that is of any concern, and the local oscillator levels (+7 dBm) are also relatively low. It is practical to operate the first and second mixers with lower power local oscillators, but at the cost of greater conversion loss and possibly higher spurious signal levels. Active mixers with lower local oscillator power requirements could also be employed. In this example, we have chosen to employ passive mixers.

Figure 12.2 shows the signal levels that would exist in the receiver if the maximum specified RF input signal were present and the various stages of the receiver were able to handle the signal levels that would result from the amplification that is shown. With a total net gain of 107 dB, the input to the demodulator would be +120 dBm, which is 1 billion watts and is slightly impractical. This is a very good reason to employ either gain control or limiting in a receiver. Ridiculous as this result may be, it tells us that an AGC loop that has the ability to hold the signal at the demodulator to within ±1 dB of the 0 dBm level shown in Figure 12.1 must have gain of not less than

$$\text{Total RF input signal range} - \text{demodulator input range}$$
$$= 120\ \text{dB} - 1\ \text{dB} = 119\ \text{dB}$$

A block diagram of this type is the first step in the design process. Here, passive mixers have been assumed, where the type that employs a +7 dBm local oscillator is shown. Each block in the signal path, up to the demodulator, lists its gain and noise figure, and the input to the demodulator is considered to be approximately 0 dBm. The demodulator's gain is not specified, because the controlled point for the receiver is at the input to the demodulator, and the demodulator's performance is constant if its input level and minimum input signal-to-noise ratio are constant. Frequency of operation and bandwidth are not listed in Figure 12.1 or in the succeeding figures (12.2 through 12.4) as they are not necessary to this part of the process. They will, however, be considered as we proceed further.

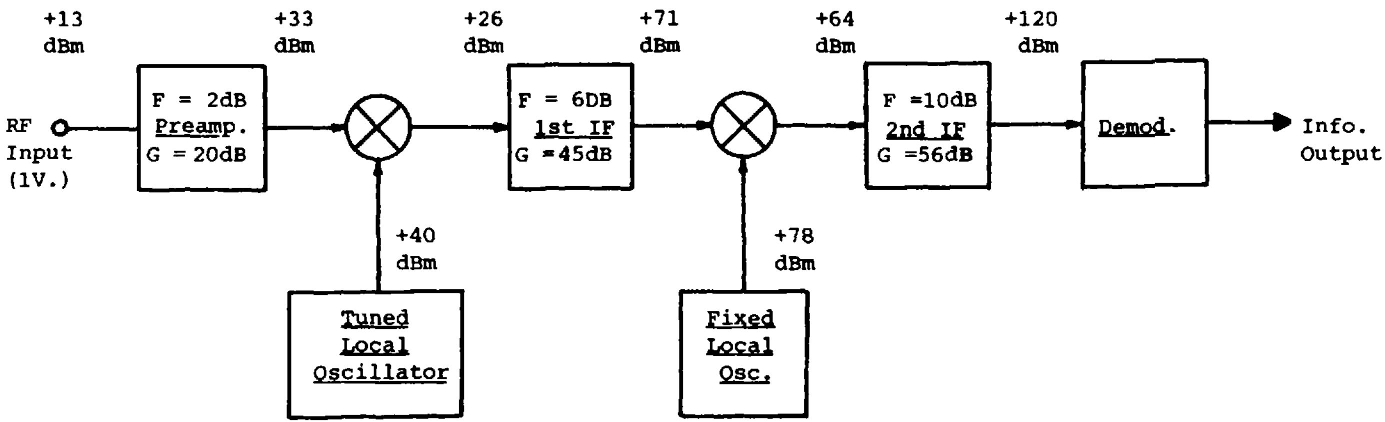


Figure 12.2 Double-conversion receiver layout, with signal levels for maximum RF signal input without AGC or limiting.

Given that the minimum RF signal level is 1 μV (−107 dBm) and the demodulator input is chosen as 0 dBm, one must conclude that the net gain required in the signal path is 107 dB. This is provided in Figure 12.1 as follows:

Gain		
Preamplifier	20 dB	
First IF	45 dB*	
Second IF	45 dB*	
	121 dB	
Loss		
First mixer	7 dB	
Second mixer	7 dB	
	14 dB	
Gain minus loss	121 dB	
	−14 dB	
	107 dB	(net gain)

12.3 GAIN CONTROL DEFINITION

The AGC loop control range must be equal to the input signal range, minus the allowable 1-dB error at the demodulator, which is the same 119 dB, minimum. This means that the preamplifier and both IFs, together, must have at least 119 dB of variability in their overall gain, controllable as a function of the signal level present. As it is necessary to provide for excess gain in the IFs and preamplifier, the total excess gain must also be provided for in the AGC design. That is, the gain variability should be increased to at least

119 dB + total excess gain

In Figure 12.3, the gain for each stage is adjusted to the amount needed to present a 0-dBm signal at each mixer and to the demodulator, which presents the gain figures that allow us to define the gain variation in each stage. This is:

	Max. gain (dB)	Min. gain (dB)	Minimum AGC variability† (dB)
Preamplifier	20	−13	33
First IF	45	7	52
Second IF	56	7	63

* In practice, the first and second IFs would be designed to have up to 10 dB excess gain. This makes up for losses that may not have been accounted for and allows for variations in the production process. Any excess gain must be accompanied by an increase in the AGC control range, however, so that the demodulator input may remain as intended.

† Does not include allowance for excess gain, which must be added. Note that excess gain does not add to the signal variability but must be accounted for because of variations from receiver to receiver.

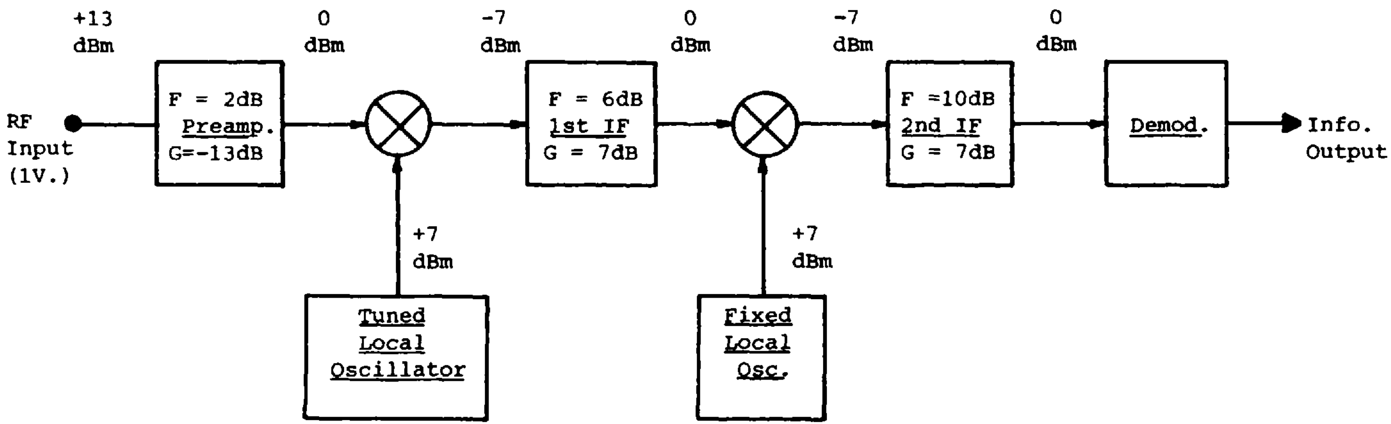


Figure 12.3 Double-conversion receiver layout, with maximum RF input signal level and gain adjusted to maintain signal linearity.

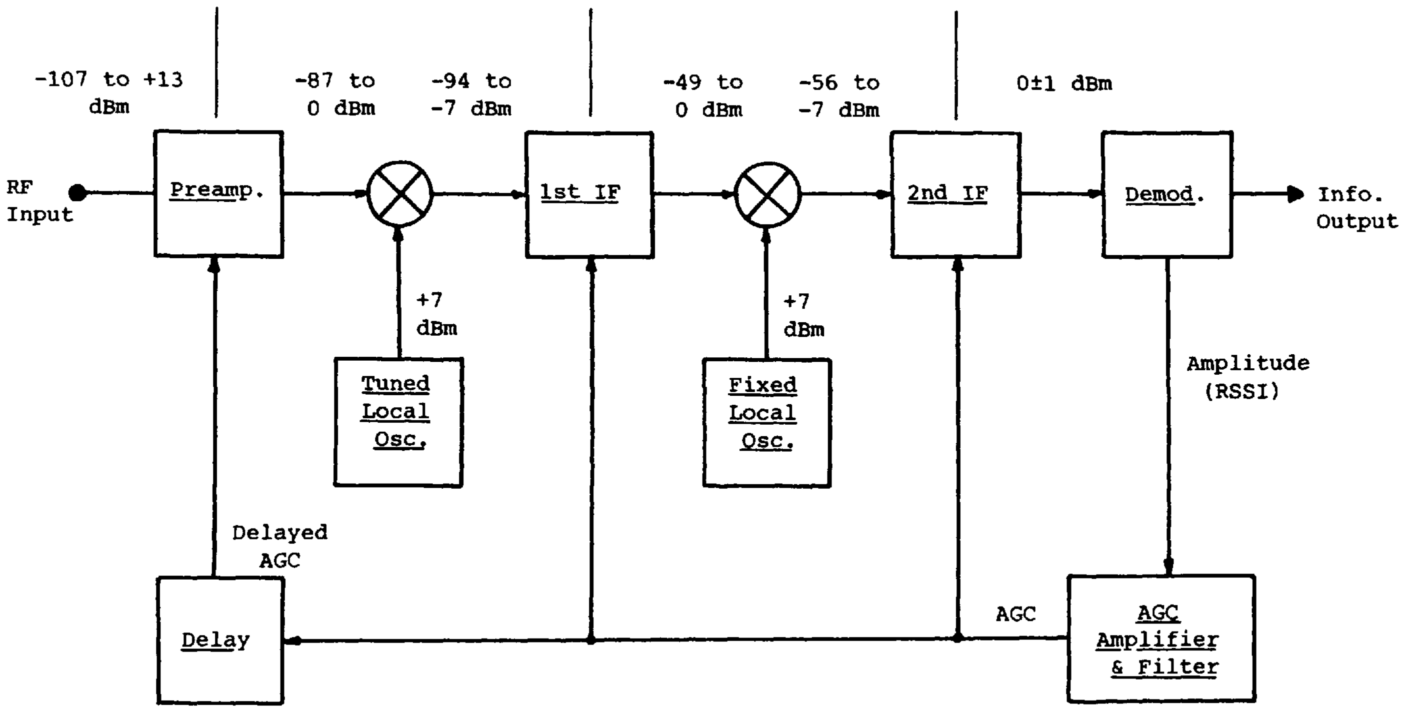


Figure 12.4 Double-conversion receiver layout, with AGC loop and signal levels throughout receiver.

At this point, we are beginning to have sufficient information on our receiver to be able to specify requirements for the various stages. Prior to doing so, however, let us go to Figure 12.4, which shows not only the received signal path but also the AGC loop that controls the gain within the received signal path. Figure 12.4 also gives the composite signal range through the signal path within the receiver, and these, together with gain and noise figure from Figures 12.1 through 12.3, permit us to generate specifications for the subsystems needed to make the overall receiver work.

12.4 SUBSYSTEM SPECIFICATIONS

Gathering all of the specifications that were postulated at the start of this chapter and combining them with those that have been developed in the course of constructing the diagrams shown in Figures 12.1 through 12.4, we have the summary given in Table 12.1. Also included are typical specifications for adjacent channel rejection and image rejection, as well as allowances for typical standard channel spacings.

We will now proceed to set down specifications for all of the major subsystems of the 800- to 900-MHz receiver that we started out to design, beginning at the receiver's RF input.

Table 12.1 Summary of Specifications for Overall Receiver

Sensitivity	AM, 1.0 μV for 10 dB $(S + N)/N$ ratio with 30% modulation at 1.0 kHz BPSK, 1.0 μV for bit error rate of 1×10^{-4} at data rate of 38.4 kbps
Adjacent channel rejection	Not less than 80 dB
Operating range	1.0 μV to 1.0 V (-107 to $+13$ dBm)
Maximum nonoperating input without damage	+22.5 dBm (3.0 V)
Frequency range	800–900 MHz
Image rejection	Not less than 80 dB
Channel spacing	AM, 25 kHz BPSK, 100 kHz
Frequency accuracy	1×10^{-6} or better
Modulation	AM, 5 kHz information bandwidth BPSK, 38.4 kbps data rate
Output $(S + N)/N$ ratio	Not less than 10 dB
Output bit error rate	Not greater than 1×10^{-4}

Preamplifier

Operating frequency	800–900 MHz
Signal input range	Operating, 1.0 μV to 1.0 V
	Nonoperating, up to 3.0 V
Input/output impedance	50 ohms
VSWR	Not greater than 1.5:1
Noise figure	Not greater than 2.0 dB
P1dB (input)	Not less than +10 dBm
AGC range	Not less than 36 dB
Attenuator loss (insertion loss)	Not greater than 2 dB
Bandwidth	100 MHz
Rolloff	Not less than 40 dB/octave

First mixer

Input frequency	800–900 MHz
Local oscillator frequency range	1035–1135 MHz
Local oscillator level	+7 dBm
Conversion loss	Not greater than 7 dB
Port-to-port isolation	Not less than 30 dB
P1dB input	Not less than 0 dBm
VSWR (all ports)	Not greater than 2:1
Impedance (all ports)	50 ohms

First IF

Center frequency	235 MHz*
Bandwidth	100 kHz
Gain	7–45 dB
AGC range	38 dB minimum
Noise figure	Not greater than 6 dB
P1dB	Not less than +10 dBm
Input/output impedance	50 ohms
VSWR	Not greater than 2:1

Second mixer

Input frequency	235 MHz
Local oscillator frequency	10.7 MHz
Local oscillator level	+7 dBm
Conversion loss	Not greater than 7 dBm
Port-to-port isolation	Not less than 30 dB

*Chosen arbitrarily for this example. Should be chosen on the basis of spur analysis.

P1dB input	Not less than 0 dBm
VSWR (all ports)	Not greater than 2:1
Impedance (all ports)	50 ohms

Second IF

Center frequency	10.7 MHz*
Bandwidth	AM, 10 kHz BPSK, 100 kHz
Gain	7–56 dB minimum
AGC range	49 dB minimum
Noise figure	Not greater than 10 dB
P1dB	Not less than +10 dBm
Input/output impedance	50 ohms
VSWR	Not greater than 2:1

Demodulator

Input frequency	10.7 MHz
Modulation	AM, bandwidth 5 KHz BPSK, data rate 38.4 kbps
Input level	0 dBm ± 1 dB
Input impedance	50 ohms
VSWR	Not greater than 2:1
Noise figure	Not greater than 15 dB
Baseband output	AM—not less than 1.0 W, ±1 dB from 20 Hz to 5 kHz BPSK—0/5V ± 0.5V into 2K ohm
Output impedance	AM, 8 ohms BPSK, not greater than 100 ohms; single-ended or differential connection provided
Output $(S + N)/N$ ratio	Not less than 10 dB over receiver operating range
Output BER	Not greater than 1×10^{-4} over receiver operating range
RSSI† output	0 to 5 VDC over receiver operating range

*Arbitrarily chosen to take advantage of the large variety of excellent filters available at 10.7 MHz.

†Received signal strength indication ... DC voltage output as a function of signal level. Used for automatic gain control.

AGC processor

Loop gain*	Not less than 119 dB
Closed-loop bandwidth	10 Hz†
Attack time	1 msec maximum
Decay time	35 msec minimum
AGC delay‡	To 100 μV level
AGC range (VDC)	0 to VCC for full control of gain-controlled stages
AGC adjustment	To allow setting demodulator input level

These subsystem specifications may be used to channel the efforts of a group of designers, so that (if each can meet his particular specification) the receiver can be designed by a group working together to reduce the time of development. The specifications could also be employed as the basis for outside procurement.

This process, or its equivalent, must be employed to ensure that a receiver can operate over its entire signal range without becoming nonlinear at the higher signal levels and that sufficient gain is provided to meet the small-signal requirements. If done carefully, the results can define every receiver stage well enough to allow a working receiver to be built.

Of course, every receiver design is different, with its own special requirements.

12.5 SUMMARY

The overall process of laying out a receiver and calculating its internal signal levels, for the purpose of generating a detailed specification for each of its blocks, can be summarized in a few steps. These are listed in the following pages.

It is presupposed that a general requirement for the receiver is available and that the frequency of operation, type of modulation, and other such defining requirements are known. (Otherwise, why bother designing a receiver?) The procedure begins by generating a receiver block diagram. We will assume that the receiver is to use superheterodyne architecture.

1. Lay out the receiver in block diagram form, showing each subsystem in the signal path, as in Figure 12.1.

---

*Loop gain is the total range of the IF and RF gain control in gain change per volt, times the AGC voltage sensitivity in volts per unit change in signal level. The AGC loop amplifier is included.

†Chosen to prevent loss of modulation at 20 Hz.

‡Chosen to prevent desensitization of the receiver at signal levels less than 100 μV (−67 dBm).

At frequencies below 100 to 200 MHz, single conversion is usually acceptable. Otherwise, two or more conversions are required. Remember that higher gain is more practical when more frequency conversions are employed. Also, the frequency of the demodulator can more readily be reduced, which provides for a wider selection and more freedom in the selection of a final IF filter.

Choice of active or passive mixers need not be made at the time that the layout is done but is necessary before the assignment of gain to the different receiver stages is done. This choice is usually based on development time and/or cost. Development time is typically negligible for a passive (usually double-balanced) mixer, because they are purchased ready made. On the other hand, active mixers may be much less expensive in large quantities, even though they may require development time with its attendant cost.

A decision must also be made as to the type of IF amplifier that is to be employed, and this decision is based on the modulated signal's characteristics. Either gain-controlled, logarithmic, or limiting IF amplifiers may be chosen, where the choice is based on whether or not the signal must be processed linearly and whether the signal is of constant duration or pulsed.

2. Calculate the gain required in the receiver between its RF input and its demodulator input.

$$\text{Gain required} = \frac{\text{demod. input level (V)}}{\text{RF input level (V)}} + \text{losses}$$

Losses include the insertion loss of filters, attenuators, the conversion loss of mixers, and any other loss that causes a reduction of the signal level within the signal path.

3. Distribute the gain required:

Assign 15 to 20 dB gain to the preamplifier. Less gain may make the first mixer design more difficult from the noise figure standpoint; more gain may make it too difficult to handle the amplified signal level in both the preamplifier and the first mixer.

The remainder of the required gain must be distributed to the IF stages. Note that no more than 60 dB gain should be attributed to any one frequency, with no more than 40 dB being preferred. (The higher the gain at any one frequency, the more difficult is the shielding problem.)

In cases in which active mixers are employed, it is possible that some gain may be attributed to the mixers as well.

4. With minimum RF input, calculate the signal level at the input and output of every block in the signal path between the RF input and the demodulator input. The signal levels found will be the levels that exist when the receiver is operating at its maximum sensitivity. Mark the levels on the diagram.

5. With maximum RF input level, recalculate the signal level at the input and output of every block in the signal path between the RF input and the input to the demodulator, using the same gains as in the previous step. (This is the gain determined in step 3, where the gain was distributed among the stages.)

Mark these levels on the diagram.

The difference between the maximum signal that would exist (step 5) if a stage could output it linearly and the signal that it can actually output is the amount of gain control required in that stage. That is, the gain control required by a particular stage is

AGC control (dB) =
(max. output if linear) − (actual max. output), both in dBm

6. Noting the actual signal level output capability of each stage in the diagram:

Typical preamplifier P1dB is 0 to +10 dBm.
Typical passive mixer P1dB (input) is 0 dBm, with an LO at +7 dBm. (Other passive mixers have LO levels of +17 and +27 dBm.)
Typical active mixer P1dB (input) is −15 to 0 dBm.
Typical IF amplifier P1dB is −10 to +10 dBm.

Mark the actual signal level capability on the diagram.

7. Gain for each stage under maximum RF input signal conditions may now be calculated as

Gain (under max. AGC)
$= (\text{output}_{(n)}\ \text{max., dBm}) - (\text{output}_{(n-1)}\ \text{max., dBm})$

We have examined the design layout of a superheterodyne receiver. This process allows the designer to define the operation of each stage of the receiver under maximum and minimum signal conditions, so that specifications for each can be generated, and aids in predicting their expected performance.

Tuned radio frequency receivers may be designed by the same process, by leaving out the mixers and examining the signal path just as we have done. In general, the same approach could also be used for designing a receiver that includes a regenerative detector, although such receivers are more akin to TRF than they are to superheterodyne receivers. The point is that in the design process it is important to view the signal path in the receiver as an integrated whole, with all parts depending on all of the others. The process described in this chapter is a systematic technique for design of a receiver, and that is all. If you have a method that is equivalent or better, please use it.

QUESTIONS

1. In a receiver whose input RF signal covers a range of −90 to −5 dBm, what is the net gain required in the receiver if the demodulator requires an input level of +5 dBm?
2. If this receiver is to operate in the AM broadcast band, should it be a double-conversion receiver? Why?

3. Could a passive, double-balanced mixer be employed? What advantages and disadvantages would it have?
4. Should AGC be employed in the above receiver? If so, what should be the minimum AGC control range?
5. What is the noise figure of the receiver shown in the layout of Figure 12.1?
6. Can the receiver whose subsystems are specified in section 12.4 meet the 1 μV sensitivity requirement in both the AM and BPSK modes? Show your calculations.
7. Why can an FM broadcast band receiver get away with only a single conversion?
8. Generate a block diagram for a receiver to operate in the 225 to 400 MHz band, receiving FSK at 100 kbps.
9. Can your receiver from question 8 operate at 5 μV input? At that input level, what is the expected BER?
10. What is/are the IF frequency(ies) used in your receiver design, and what is/are the local oscillator frequency(ies)?

13
Error Detection and Correction

Error detection and correction, as it pertains to receivers, is the process by which a terminal determines that a mistake has been made and attempts to correct that mistake. A human operator, for example, can make corrections to a message received under poor operating conditions based on a priori knowledge of what is to be received and on the context of the message itself. We are not so concerned here, however, with the process by which a human operator can detect errors and make decisions on what was intended as we are with detection and correction of errors without the intervention of an associative human brain as an interpreter. That is, we are interested in the process of sending data at low bit error rates (one error per million bits or less) between terminals that have little or no idea at all if the data they are receiving are in error. (The average computer is too stupid to recognize anything other than data not being sent in the proper format, such as bytes with the wrong length or that do not agree with their parity.)

13.1 ERROR DETECTION AND CORRECTION BACKGROUND

In general, there must be cooperation between the transmitter and the receiver before error detection and/or correction can occur. This is true for the human error detection and correction system as well. The human must know the general rules of the language and the context before errors can be detected. Certainly, an English-speaking error detector cannot detect errors in Swahili.

In teletype systems, for example, messages are often sent wherein each character is sent as a 7-bit word, where each 7-bit word has exactly three ones in it and four zeros. This means that there are $\binom{7}{3} = 35$ characters in the set of possibilities, even though there are $2^7 = 128$ possible 7-bit words. Since the transmitter sends only the 7-bit words that have three ones in them, however, the receiving terminal can detect any single error (but not two errors) in the 7-bit word simply by counting the number of ones. If the number of ones is not three, then an error has occurred. In many applications, the occurence of an error causes the receiver to ask for any character in which an error has been detected to be sent again. This is called an "ARQ" technique.

Please note that in this approach, the transmitter must condition the signals that it sends, so that the receiver can detect the occurrence of an error. (This is the cooperation that was mentioned earlier, along with an agreement to retransmit any character that the receiver finds to be in error.)

A simple method that allows detection of an error, but does not provide identification of it, is a parity check. This technique is often employed in computer memory subsystems, wherein a byte* that is to be stored is processed by counting the number of ones in the byte. Then, if the number of ones is odd, a one is added at the end of the byte, but if the number of ones is even, a zero is added at the end. (This causes the byte to be 1 bit longer, so that a typical 8-bit byte is then 9 bits long, but we have added the ability to detect a single error.) On receiving a now-9-bit byte, the receiving unit observes the parity bit that has been tacked onto the first 8 bits. If the parity bit is a one, then the first 8 bits should contain an odd number of ones, but if the parity bit is a zero, then there should be an even number of ones in the first 8 bits. The total number of ones in the data and the parity are always even.

Using this simple "even parity" method, any single error in the first 8 bits (the original byte of information) can be detected. An error in the parity bit that is added to the end cannot be detected, however, so the byte to which the defective parity bit is attached will appear to the receiver to be in error, even though it may be flawless.

The problem that exists in error detection alone is that the receiver may not be able to make use of the knowledge that it has, other than to apply an "erasure" to that particular character and allow the end user to do what he or she can with the knowledge that it was wrong. ARQ (automatic repeat request) is a step better, but it has the flaw that as the error rate increases, more and more of the system capacity is employed to repeat characters that were in error.

Another step comes about when one realizes that when only a few of the 2^m symbols in a set are being used and all are being chosen from the same m-bit set, it may be possible to select the "best" symbols from the set and thereby improve the performance of the error correction process. This is, in fact, the whole error correction process in a nutshell. To be more explicit, the whole science of error correction coding (at least for block codes) is that of finding the best set of m-ary symbols from the set of all symbols of a given length, where "best" is

*A typical byte is 8 bits long—but not always.

defined as having the greatest mutual distance between all members of the chosen symbol set.

Suppose that we wish to send information using the English alphabet, with a few punctuation marks, and numerals zero through nine, so that we have 39 symbols plus an upper/lowercase symbol available to double the possibilities. This would give us the possibility of sending 78 different characters but would require only 40 different symbols. A 6-bit symbol would be sufficient to provide the symbol set, as $2^6 = 64$ is the number of (binary) symbols, with 6 bits, available. The problem is that the 40 symbols that we need leave only 24 in the set we have available, and this is not enough to ensure that the 40 we use are sufficiently different from one another to ensure that we can tell them apart once errors are made. It is therefore necessary to increase the symbol length, so that there are more symbols in the set from which we can choose. (We are now slowly closing in on the basis from which error correction coding was developed.)

Given that we employ 7-bit symbols, we now have $2^7 = 128$ symbols to choose from, where we need to choose only the 40 symbols that have the best characteristics. And what are those characteristics? They were formulated by R. W. Hamming, which is the reason why they are usually termed "Hamming weight" and "Hamming distance." Hamming weight is the number of ones in a word or symbol, and Hamming distance is the number of differences between a pair of words or symbols. That is, the binary word 1111111 of length 7 bits has a weight of seven, and if we compare this word with another word of the same length, 0000000, their distance is seven. If we add these two words together modulo two (without a carry) we have

$$\begin{array}{r} 1111111 \\ \oplus\ \underline{0000000} \\ 1111111 \end{array}$$

and their modulo-two sum has a *weight* that is the *distance* of the two from one another.

Suppose now that we transmit the first word, 1111111, to represent a one, and the second word, 0000000, to represent a zero, even though it would be notoriously inefficient to do so. (We call the two words that we are using "symbols.") Inasmuch as the two symbols are different in seven places (their distance is seven), up to three errors in the seven used to transmit the information can be tolerated. Let us see why.

If, for example, a one was transmitted, using the symbol 1111111, and 1101111 was received (the third bit in the received word is in error), then it could be detected by comparison of the received symbol with the expected possible symbols. This is done by adding the received symbol, modulo two, to the expected symbols:

$$\begin{array}{rl} & 1101111 \text{ Received symbol} \\ \oplus & \underline{1111111} \text{ Expected one symbol} \\ & 0010000 \text{ (Weight} = 1) \end{array} \qquad \begin{array}{rl} & 1101111 \text{ Received symbol} \\ \oplus & \underline{0000000} \text{ Expected zero symbol} \\ & 1101111 \text{ (Weight} = 6) \end{array}$$

The modulo-two sum with the lowest Hamming weight denotes the correct symbol, and the incorrect bit is denoted by the one in the modulo-two sum.

To detect and correct more bits in error, we need to have greater distance between the symbols, and the pairs of symbols (in the set of 7-bit symbols) with the greatest distance are 1111111 paired with 0000000,* plus all of the other pairs consisting of XXXXXXX and $\overline{X}\overline{X}\overline{X}\overline{X}\overline{X}\overline{X}\overline{X}$. Since any 7-bit symbol, paired with its bit-by-bit complement, has a distance of seven with that complement, if the pair is used to send information, then up to three errors in the symbol can be detected and corrected. This can be readily demonstrated.

Given a second error in the transmitted word, in the fourth position, we would transmit 1111111 and receive 1100111. Modulo-two adding the received word with the two possible words, we have

$$\begin{array}{rl} & 1100111 \text{ Received symbol} \\ \oplus & \underline{1111111} \text{ Expected one symbol} \\ & 0011000 \text{ (Weight} = 2) \end{array} \qquad \begin{array}{rl} & 1100111 \text{ Received symbol} \\ \oplus & \underline{0000000} \text{ Expected zero symbol} \\ & 1100111 \text{ (Weight} = 5) \end{array}$$

and the correct symbol corresponds to the sum with the lowest Hamming weight, just as when there was only one error.

Adding a third error in position six (it does not matter in which position the error is added), the transmitted signal would still be 1111111, and the received signal would be 1100101. Now when we add the received word to the possible words we have

$$\begin{array}{rl} & 1100101 \text{ Received symbol} \\ \oplus & \underline{1111111} \text{ Expected one symbol} \\ & 0011010 \text{ (Weight} = 3) \end{array} \qquad \begin{array}{rl} & 1100101 \text{ Received symbol} \\ \oplus & \underline{0000000} \text{ Expected zero symbol} \\ & 1100101 \text{ (Weight} = 4) \end{array}$$

Here, the one symbol, modulo-two added to the received symbol, still has a lower Hamming weight and would be chosen correctly.

If we add one more error, in position one for example, the result of modulo-two summing with the possible symbols becomes

$$\begin{array}{rl} & 0100101 \text{ Received symbol} \\ \oplus & \underline{1111111} \text{ Expected one symbol} \\ & 1011010 \text{ (Weight} = 4) \end{array} \qquad \begin{array}{rl} & 0100101 \text{ Received symbol} \\ \oplus & \underline{0000000} \text{ Expected zero symbol} \\ & 0100101 \text{ (Weight} = 3) \end{array}$$

*The special technique in which a bit is sent more than once is called "replication coding," and even though it is inefficient ($1/m$ times the data rate) it does have some application.

Now the modulo-two sum with the lowest Hamming weight corresponds to the zero symbol, and we can see that four errors in the received symbol could cause the wrong symbol to be chosen, whereas three errors would not.

A general rule that can be employed is that the number of bits that can be corrected is the next lowest integer that is less than half the distance between the words in the symbol set.* Of course, if multiple symbols are employed, then the number of errors that can be corrected is determined by the minimum distance between all of the members of the set.

It is very inefficient to employ a technique in which the same information is sent several times, even though the technique is very powerful (replication coding can accept more errors than any other coding method) but has a maximum of 1/3 transmission rate. That is, the rate at which the symbols are sent is at least three times the data rate. For this reason, more efficient encoding and decoding methods have been developed.

All error correction systems rely on redundancy in the symbols used, as well as maximum distance between those symbols.

13.2 TYPES OF ERROR DETECTION CODES AND THEIR PERFORMANCE

Error detection and correction codes are usually divided into two types: "block" codes and "stream" codes. Block codes are codes in which a specific input word to the encoder produces a specific output symbol, and the number of bits in the input and encoded output are fixed. In such cases, the code "rate" is considered to be the input rate divided by the output rate. An encoder whose input is a 5-bit word and whose output is a 10-bit symbol would be termed a "rate 1/2" encoder, and the code generated would be called a rate 1/2 code. Similarly, such a code is often called a (10,5) code.

Stream codes do not operate in such a way that an identifiable specific input word produces a specific output symbol, although they have a "rate" that is simply the input data rate divided by the encoded output rate. This, however, is the point from which the similarities between block and stream coders and their operation begin to diverge, but they can be compared on the basis of their coding gain.

Today, the type of error correction coding employed in a particular application is influenced by the availability of a reasonably simple decoder, more than any

---

*For example, if a set of symbols has a distance of 5, then the error correction capability is 5/2 = 2.5, and two can be corrected. Note that it does not matter what the symbol length is, as long as the distance is 5. It also does not matter how many symbols are in the set, if the minimum distance between symbols is at least 5.

Table 13.1 Common Error Detection and Correction Techniques

Encoding technique	Decoding technique	Typical characteristics
Block codes		
Hamming	Hamming	(7,4) Single error correction. Rate 4/7
Golay	Golay	(24,12) Up to three errors corrected. Rate 1/2.
BCH (Bose-Chaudari-Hocquenghem)	BCH	(23,12) Up to three errors corrected. Rate 12/23.
Reed-Solomon	Reed-Solomon	(21,6) Up to three errors corrected. Rate 2/7.
Stream codes		
Convolutional	Viterbi	Rate 1/2, constraint length seven. Coding gain variable.

other factor. Table 13.1 lists the coding and decoding methods that are most common, where the reason for their use is that they are either the lowest in cost to implement or an integrated circuit is available that makes them easy to use, or both.

Golay and convolutional coders/decoders are available in integrated circuit form.* Diagrams of convolutional encoders and Viterbi decoders from Qualcomm and Stanford Telecom are shown in Figure 13.1 and Figure 13.2. Figures 13.3 through 13.5 illustrate the performance of such encoders and decoders. Table 13.2 is a comparison of three† convolutional encoders with Viterbi decoders, using rate 1/2, $K = 7$, soft decision decoding. (Hard decision decoding provides 2.0 to 2.5 dB less coding gain at the same uncorrected, or raw, input error rate.)

The performance of convolutional encoders with Viterbi decoders from Stanford Telecom and Qualcomm, but with hard decision decoding, is shown in Table 13.3, along with the performance of a Golay (24,12) encoder/decoder. The Golay encoder/decoder is a rate 1/2 block system but can be compared with the convolutional encode/Viterbi decode stream systems on a coding gain basis. Note that the STEL and Qualcomm units' performances agree exactly in their hard decision decoding mode, and that Golay encoder/decoder performance is within 1 dB of both over the 10^{-2} to 10^{-7} operating range.

Another block code of interest is the Hamming single error correction code.

*Golay encoders are available from Space Research Technology, Inc., and convolutional coders with Viterbi decoders are available from Qualcomm Corporation and Stanford Telecom.

†STEL 5269, Qualcomm 1650, and Joint Tactical Information Distribution System.

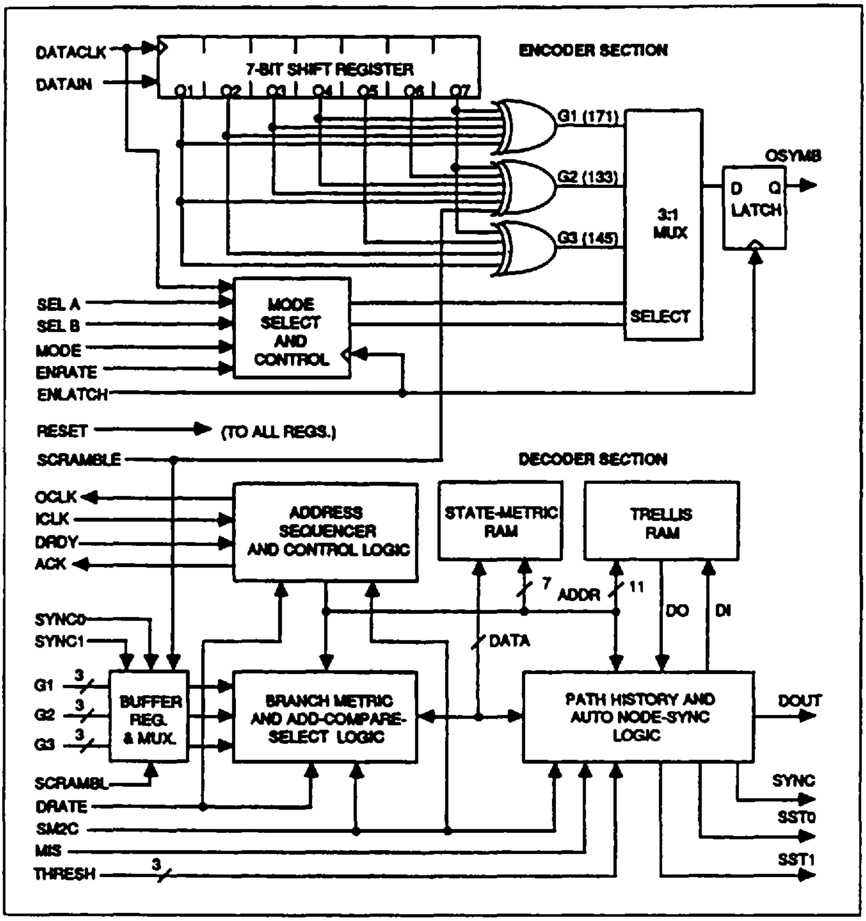


Figure 13.1 Block diagram of Stanford Telecom STEL-5269 convolutional encoder/Viterbi decoder integrated circuit. (Courtesy of Stanford Telecom.)

Although it is relatively simple to implement, this code type has been used in some very demanding applications. The encoding technique is based on addition of parity bits, and decoding is based on parity checks, where $m - 1$ parity checks are required to detect an error in $m + p$ data bits with parity. In the case of the Hamming codes, however, errors in the parity bits can also be detected. Thus, any single error can be detected and therefore corrected. (One of the joys of working with binary signals is that when an error is detected, the correct information must be the complement of the bit that is found to be in error.)

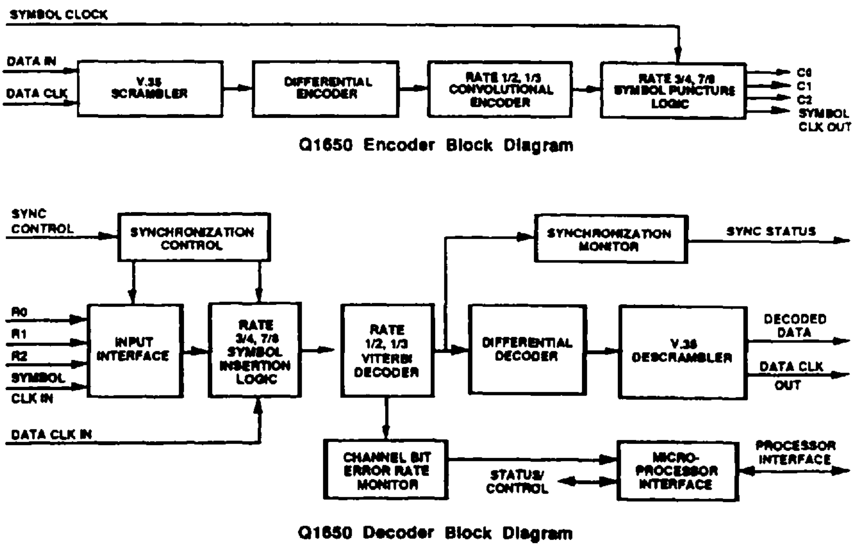


Figure 13.2 Block diagrams of Qualcomm Q1650 convolutional encoder and Viterbi decoder (resident on same chip). (Courtesy of Qualcomm Corp.)

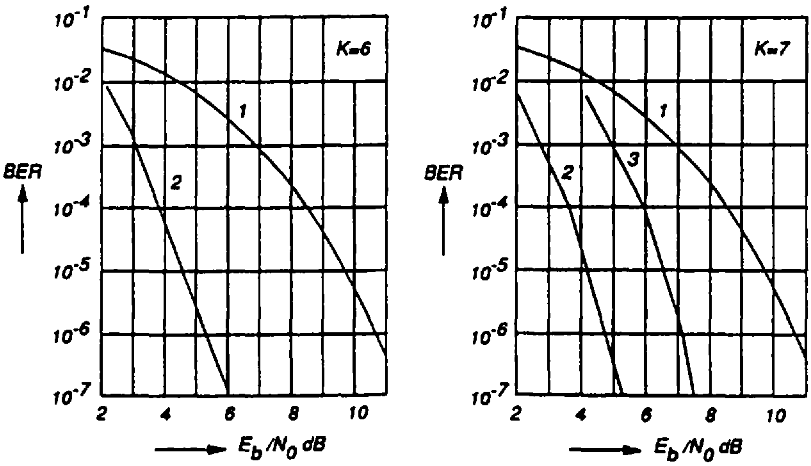


Figure 13.3 Performance of constraint length six ($K = 6$) and constraint length seven ($K = 7$) Stanford Telecom STEL-5269 integrated circuit, with hard and soft decision decoding. Curve 1 shows unencoded PSK performance. Curve 2 shows performance with soft decision decoding, and curve 3 shows performance with hard decision decoding. (Courtesy of Qualcomm Corp.)

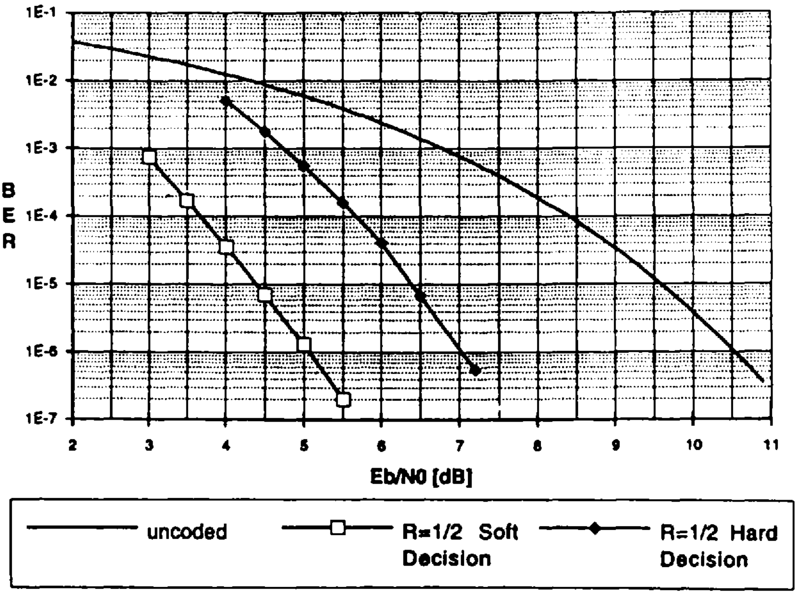


Figure 13.4 Performance of Qualcomm Q1650, constraint length seven, convolutional encoded/Viterbi decoded integrated circuit. Comparison of uncoded PSK and rate 1/2 coded performance using soft and hard decisions. (Courtesy of Qualcomm Corp.)

Example. Let us assume that we have a data word $d_1\ d_2\ d_3\ d_4$ where each of the characters is binary (0 or 1), and we wish to encode it using the Hamming single error correction technique.

The parity word required has a length that is not less than $\log_2(m + p + 1)$. In this case, with four data bits and three parity bits, $4 + 3 + 1 = 8$, and $\log_2 8 = 3$. We find the parity word by generating the parity bits individually in three steps;

1. $p_1 = d_1 \oplus d_2 \oplus d_3$
2. $p_2 = d_1 \oplus d_2 \oplus d_4$
3. $p_3 = d_1 \oplus d_3 \oplus d_4$

The parity word, $p_1p_2p_3$, is simply added at the end of the data word, and this characteristic is typical of a "systematic" code. The symbol resulting from adding the parity to the data is

$$\underbrace{d_1d_2d_3d_4}_{\text{data}}\ \underbrace{p_1p_2p_3}_{\text{parity}}$$

and this is the symbol that would be transmitted.

Detecting any errors that may be made in the transmission process is done through a similar process; three parity checks are performed on the received

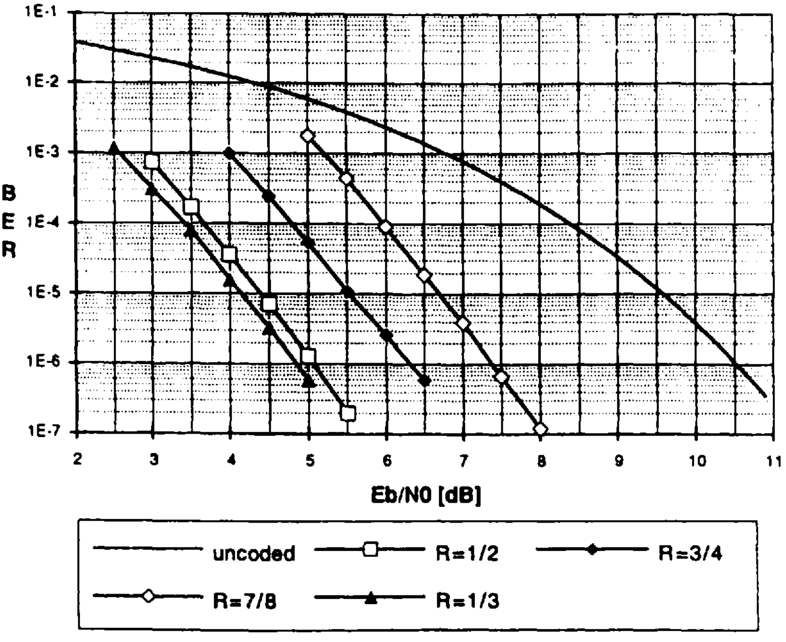


Figure 13.5 Performance of Qualcomm Q1650 encoder/decoder using soft decisions and various rates. (Courtesy of Qualcomm Corp.)

Table 13.2 Comparative Performance of Three Rate 1/2, $K = 7$ Convolutional Encoders with Viterbi Soft Decision Decoding

	Coding gain (dB)		
Raw error rate	STEL	Qualcomm	JTIDS
10^{-2}	2.7	2.1	2.2
10^{-3}	4.2	3.8	3.8
10^{-4}	4.9	4.7	4.6
10^{-5}	5.5	5.1	5.1
10^{-6}	6.0	5.4	5.3
10^{-7}	6.2	5.6	—

Table 13.3 Performance of the STEL and Qualcomm Encoders/Decoders Shown in Table 13.2 Using Hard Decision Decoding Plus Golay

	Coding gain (dB)		
Raw error rate	STEL	Qualcomm	Golay
10^{-2}	0.5	0.5	0.9
10^{-3}	2.0	2.0	1.6
10^{-4}	2.6	2.6	2.0
10^{-5}	3.1	3.1	2.4
10^{-6}	3.4	3.4	2.7
10^{-7}	3.6	3.6	2.8

symbol, and these are shown in Table 13.4. If all three parity checks are zero, then there is no error, or else there is more than one error (remember that this is a single-error-correcting code).

Observing the parity checks made at the receiver, we see that

d_1 occurs in all three checks,
d_2 occurs in checks 1 and 2,
d_3 occurs in checks 1 and 3, and
d_4 occurs in checks 2 and 3.

At this point we have all of the information that we need to determine a data error, but we can also determine whether or not an error has been made in the parity, because

p_1 occurs only in check 1,
p_2 occurs only in check 2, and
p_3 occurs only in check 3.

Table 13.4 Parity Checks Used with Hamming (7,4) Single Error Correction Code to Determine Error Position in Received Symbol

Check no.	Result
1	Does $d_1 \oplus d_2 \oplus d_3 \oplus p_1 = 0$?
2	Does $d_1 \oplus d_2 \oplus d_4 \oplus p_2 = 0$?
3	Does $d_1 \oplus d_3 \oplus d_4 \oplus p_3 = 0$?

Table 13.5 Parity Check Error Listing Versus Error Position

If parity check = 1 in check number	Then bit in error is
1,2,3	d_1
1,2	d_2
1,3	d_3
2,3	d_4
1	p_1
2	p_2
3	p_3

Therefore we can set up a table that identifies any error in either the data or the parity, as illustrated in Table 13.5.

Given data word 1101, using the rules that have just been described, the parity that would be generated is 010, and the symbol transmitted would be 1101010. Since 7 bits are employed to transmit 4 bits of information, this is a (7,4) or rate 4/7 code, and it is capable of detecting and correcting one error in every 7-bit symbol.

Implementation of this encoder would be very simple with logic devices but would be even simpler with a read-only memory. Noting that there are only $2^4 = 16$ possible 4-bit input words and that each one of the input words maps to a specific 7-bit output word, we see that only $7 \times 16 = 112$ bits of memory are required to store the entire library of symbols needed.

Examination of the 16 possible 7-bit words generated by the 4-bit inputs combined with their parity shows that we have the list of data words, parity, and symbols that are shown in Table 13.6 as the only possible symbol set using the rules that have been enumerated. Further examination of these symbols shows that the minimum distance between each of these symbols and any other symbol in the set is three. Therefore, a simple way to decode the received information while detecting and correcting any single error that may be present is to modulo-two add every possible symbol from the list in Table 13.6 to the symbol that has been received. Then, the most likely symbol is the one that has the lowest Hamming weight from the resulting modulo-two sums. Any errors will be identified in the correct symbol by a one in the sum with lowest weight. If no error is present, the lowest weight symbol will be all zeros.

A wired-logic error detection and correction decoder for 4-bit data words using the Hamming single error correction codes is shown in Figure 13.6. This decoder is shown in block diagram form, for implementation using standard logic devices. It requires only 12 exclusive-OR gates, 3 inverters, 7 AND gates, and 7 D

Table 13.6 The Symbol Set Generated by the Hamming Single Error Correction Encoding Technique, with 4-Bit Data Words

	Data words	Parity	Transmitted symbol
1.	0000	000	0000000
2.	0001	011	0001011
3.	0010	101	0010101
4.	0011	110	0011110
5.	0100	110	0100110
6.	0101	101	0101101
7.	0110	011	0110011
8.	0111	000	0111000
9.	1000	111	1000111
10.	1001	100	1001100
11.	1010	010	1010010
12.	1011	001	1011001
13.	1100	001	1100001
14.	1101	010	1101010
15.	1110	100	1110100
16.	1111	111	1111111

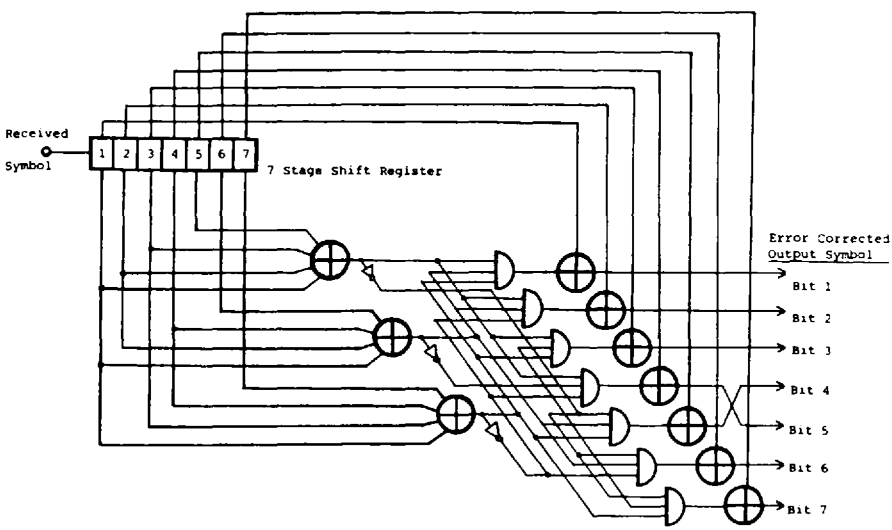


Figure 13.6 Logic diagram for Hamming (7,4) single error correction code with serial input and parallel symbol output (decoder).

flip-flops for a serial input, parallel output decoder function. Symbol synchronization is not shown in Figure 13.6.

If one wishes to detect and correct more than one error, it is possible to carry out the Hamming single error correction process more than once, although it produces very inefficient coding. For example, if one wanted to correct two errors in a 4-bit word, it would be necessary to perform two sets of parity checks, adding first three parity bits and then four more parity bits. This would produce 11 bit symbols, for an (11,4) rate 4/11 code, and the ability to correct three errors would require 15 bit symbols with a rate 4/15 code. The reason for such poor efficiency in the Hamming technique is that as more bits are corrected, parity checks must be added to find errors in the parity that was added in prior iterations. More efficient encoding methods are available, fortunately.

Two of these encoding methods are the Golay and BCH coding techniques, both of which can be implemented with a rate of approximately 1/2 and can correct up to three errors in a 23- or 24-bit encoded symbol. (This is a percentage of errors of 12.5 to 13 percent.) Another coding technique is Reed-Solomon, which is actually a BCH coding technique that is usually employed in m-ary transmission schemes. A (21,6)* Reed-Solomon decoder can detect and correct three errors in 21, for a percentage corrected of 14.28 percent. A significant advantage of the Reed-Solomon technique is that it can correct errors of the same order as rate 1/2 BCH or Golay approaches but does not double the bandwidth required. (Simultaneously going from BPSK to QPSK when adding a rate 1/2 code accomplishes approximately the same thing.)

The key to the BCH codes is their flexibility. For any m, such that $m = \log_2(n + 1)$, a BCH code can be derived that will correct t errors. For Golay codes, this is not the case. While the same is true of Hamming codes with respect to the data block length, they are too inefficient for correction of multiple errors. Important parameters for the BCH codes are

Encoded block length: $n = 2^m - 1$, $m = 3,4,5 \ldots$ (or a factor of n)
Number of bits corrected: $t = (n - k_{\min})/m$
Data bits per block: $k_{\min} = n - mt$

Example. Suppose that a block of 16 bits is to be encoded and that 3 bits must be corrected. That is, $k = 16$, and $t = 3$. The encoded block length must be longer than the data length, so that if we choose 31 for approximately rate 1/2 encoding, m can be chosen to be 5. Therefore,

$$n = 2^m - 1 = 2^5 - 1 = 31$$

*(21,6) Reed-Solomon coding is typically used in applications such as 8-ary FSK or frequency hopping (also 8-ary). The 21-bit symbol is divided into seven 3-bit subsymbols and each subsymbol is transmitted as one of eight frequencies.

and the number of data bits per block is

$$k_{\min} = n - mt = 31 - (5 \times 3) = 16$$

which means that there are 15 parity bits per block, and this provides 3 bits of error correction in a 31-bit symbol. The percent of errors corrected is $3/31 \times 100 = 9.67$ percent.

Table 13.7 is a listing of the characteristics of some BCH codes with respect to their block length, number of bits corrected, and number of information bits per block, as a function of m from 3 through 10. Please note that once m is greater than 4, the number of symbols employed is too large to allow practical decoding of BCH codes. (For $k_{\min} = 4$, the number of symbols is only $2^4 = 16$, and for $k_{\min} = 7$, $2^7 = 128$ symbols are needed, but for $k_{\min} = 16$, $2^{16} = 65{,}536$ symbols would be required. This is too many to expect to compare on a symbol-by-symbol basis to find the one with the lowest Hamming weight with respect to a received symbol.) See the references given for a description of other BCH decoding methods, which are beyond the scope of this chapter.

Reed-Solomon, BCH, Golay, Hamming, and all other block codes can be decoded by employing the lookup table technique previously mentioned, in which all of the symbols that are being used are stored, with any received symbol compared with the stored set of allowed symbols. Then the most likely is chosen by finding the one with the smallest Hamming distance.

Stream codes such as convolutional codes cannot readily be decoded with a lookup table, even though some convolutional codes are systematic (the data themselves appear as part of the encoded symbol). In the convolutional encoder shown in Figure 13.7, the encoded output does not contain the data but is the data modulo-two added sequentially to two parity sums on the previous six input bits.

Table 13.7 Table of Characteristics of BCH Codes of Rate Approximately 1/2, with m from 3 Through 10

m	n	$k_{\min}$	t	Rate
3	7	4	1	4/7
4	15	7	2	8/15
5	31	16	3	16/31
6	63	32	5	32/63
7	127	64	9	64/127
8	255	127	16	127/255
9	511	256	28	256/511
10	1023	512	51	512/1023

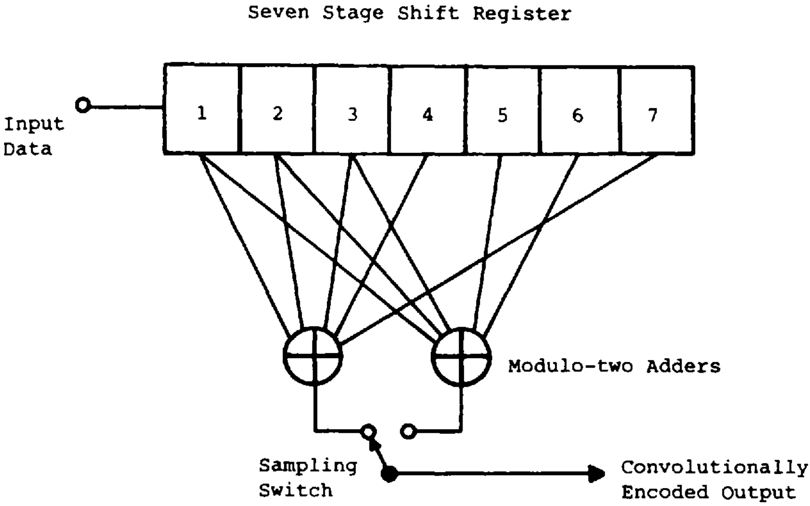


Figure 13.7 Convolutional encoder. Constraint length seven, rate 1/2 shown.

That is, the output of the encoder shown in Figure 13.7 during any input data period is

$$\text{during } t_1 = D_{\text{in}} \oplus D_2 \oplus D_3 \oplus D_4 \oplus D_7$$
$$\text{and during } t_2 = D_{\text{in}} \oplus D_2 \oplus D_3 \oplus D_5 \oplus D_6$$

where t_1 is the first half of the input data bit period and t_2 is the second half. The encoded stream of output information is simply a series of parity checks, modulo-two added to the input data, and there are two parity checks per input data bit. Also notable is the fact that the parity checks are sent directly as generated if the input bit is a zero, but they are inverted if the input bit is a one.

The shift register taps employed in generating the parity checks correspond to those for a linear maximal code of degree six, together with a second linear maximal code of degree five. Specifically, these codes correspond to the primitive polynomials $1 + x + x^4 + x^5 + x^6$ and $1 + x + x^2 + x^4 + x^5$ (referring to Figure 13.7), which are 63 and 31 chip linear maximal codes, respectively. The data, shifting through the register, are effectively convolved with these two codes, and this is why the process is called "convolutional" encoding.

Decoding the signals received is a different matter. The information itself is not received. Instead, each bit is represented by two parity checks that are either inverted or not inverted, and the best method of decoding the data is to employ a sequential decoder that derives the data from that which is received, based on a knowledge of the encoding process. However, the most common decoder in use is the Viterbi decoder, which employs a simplified algorithm and is available in integrated circuit form. Performance of Viterbi decoders is shown in Figures 13.1 through 13.5.

13.3 PERFORMANCE LIMITATIONS IN ERROR DETECTION AND CORRECTION

In case you have have not gotten the idea already, dear reader, error correction coding techniques have not yet reached perfection. That is, when a system no longer works because its link budget is not adequate, or the interference present is at too high a level, or the modulation technique being employed is not robust enough, then adding an error detection and correction coding and decoding scheme to the system *may* or *may not* make a significant difference in the system's performance. Whether error correction can solve a receiver's problem depends on the user's expectations and the specific characteristics of the propagation paths in which the system is being employed. For example, a link in which the errors that are present occur in bursts may have an average error rate of 1×10^{-6} but have 10 errors in a row, with no errors occurring at all for another 10 million bits. None of the error correction methods that we have considered here can accommodate a burst of errors that is 10 bits long.

A block code designed to correct up to 10 bits in a single symbol would have to employ symbols that are approximately 80 bits long (considering that most codes can correct about 1 bit for every 8 bits of symbol length) and the symbols used would need a minimum distance of at least 21. No doubt, such codes could be designed, but the encoding and decoding process would be complex and time consuming. The solution that is usually employed is to combine an error correction scheme with a technique called interleaving, which distributes bursts of errors over a wider span of time so that they can be managed by an error correction subsystem in small increments.

Some of the limitations of error correction techniques that should be considered, with respect to receiver performance and overall system design, are discussed in the following paragraphs.

13.3.1 Error Correction Limit

No error correction system known can correct as much as 50 percent errors. In fact, the majority of error detection and correction techniques can correct no more than approximately 15 percent. Replication coding with majority decoding (this is actually an error-ignoring scheme) can correctly select data when they are sent p times, if $(p + 1)/2$ is an integer, and an error is made no more than $(p - 1)/2$ times out of p receptions. Therefore if the data are sent three times, an error can be made in reception $(3 - 1)/2 = 1$ time out of the three, for a 33.33 percent error rate. This is a rate 1/3 technique.

It is simple to see that as p is increased, we can increase the error tolerance, but we can never correct (ignore) more than $(p - 1)/2$ errors and even if the symbol length approaches infinity, the number of errors tolerated only approaches 50 percent.

Further, if it were practical to correct 50 percent or more errors in a message, it might be possible to do away with communications systems as we know them. If I write down a random message sent to me in binary form, not having a receiver, I will be correct in 50 percent of my trials and the 50 percent error correction scheme should be able to correct the errors that I make (or at least most of them). A simple single error correction code should correct the rest.

Error correction coding schemes require redundancy in the transmitted message: at least three checks per error that is correctable. With only two checks, an error detected by one of the two checks can detect an error but does not identify it.

13.3.2 Bandwidth Expansion

When adding the redundancy needed to detect and correct errors, the effective information rate of a transmitter is increased. Part of the information that is sent is that which is necessary to identify an error or errors, but it still increases the rate of transmission necessary.

This means that a system to which error correction coding is added must employ more bandwidth than it would without the error correction coding, unless the modulation technique being employed is also changed to allow more bits to be sent in the same bandwidth (i.e., binary to m-ary). Higher signal-to-noise ratio or E_b/N_0 may or may not be required, depending on the specific modulation method being used before and after the addition of error correction coding.

13.3.3 Error Correction Coding Does Not Always Add to Performance

Many systems must operate at error rates (before correction) that are not conducive to production of sufficient coding gain to make it worthwhile to incorporate error correction coding. An example of such a system is a frequency hopping system, in which it is not unusual to have 10 percent or more of the available channels that the receiver sees occupied by interferors. This produces an error rate of 10 percent or more at the receiver's decoder input.

Observing Figures 13.3 through 13.5 and Tables 13.2 and 13.3, one can see that the encoding and decoding techniques for which these apply are not intended to operate at 10 percent error rates. Also, the maximum gain that they provide is in the range of 2.5 to 6 dB in the area for which they are intended. Unfortunately, the coding gain is greatest where it is needed least and least where it is needed most.

When we add the bandwidth expansion factor, it is possible that increasing the bandwidth of a receiver to accommodate the higher symbol rate necessitated by error correction encoding may degrade the receiver's performance at low E_b/N_0 levels.

13.3.4 Shannon and Error Correction Coding

Error correction coding is one of the techniques that are suggested by Shannon's capacity theorem. That is, the capacity of a system to transmit error-free information, expressed in bits per second, is

$$C = W \log_2(1 + S/N)$$

and error correction coding with its tendency to increase the bandwidth employed in sending information can reduce the signal-to-noise ratio requirement while maintaining the information rate. The same principle is also employed in spread spectrum systems, but the encoding methods are considerably different. In fact, some spread spectrum systems also employ error correction coding in addition to their spectrum spreading processes.

Figure 13.8 shows a plot of the error-free capacity of a system for various signal-to-noise ratios as bandwidth is increased. Figures 13.9 and 13.10 show the effect on bandwidth when capacity is constant but signal-to-noise ratio is changed and the effect on capacity when bandwidth is constant but signal-to-noise ratio is changed.

13.3.5 Error Burst Limitations

Very few error correction codes can correct bursts of errors, even though the total number of errors may well be within the average rate that can be tolerated by the particular coding scheme. That is, a decoder that can operate at an error rate of 1×10^{-2} may be able to correct 1 error in every hundred, or perhaps up to 3 in every block, but it cannot correct 10 in a thousand if they occur in a burst, even though the average is 1×10^{-2}. As previously mentioned, interleaving is the most common solution to the burst error problem, but an interleaver must work in conjunction with an error correction coder and decoder. It can only redistribute the errors in a burst, to prevent them from exceeding the number of errors correctable in a single block.

The interleaver introduces delay into the encoding and decoding process that is not less than twice the length of the maximum burst that can be corrected with the interleaver operating.

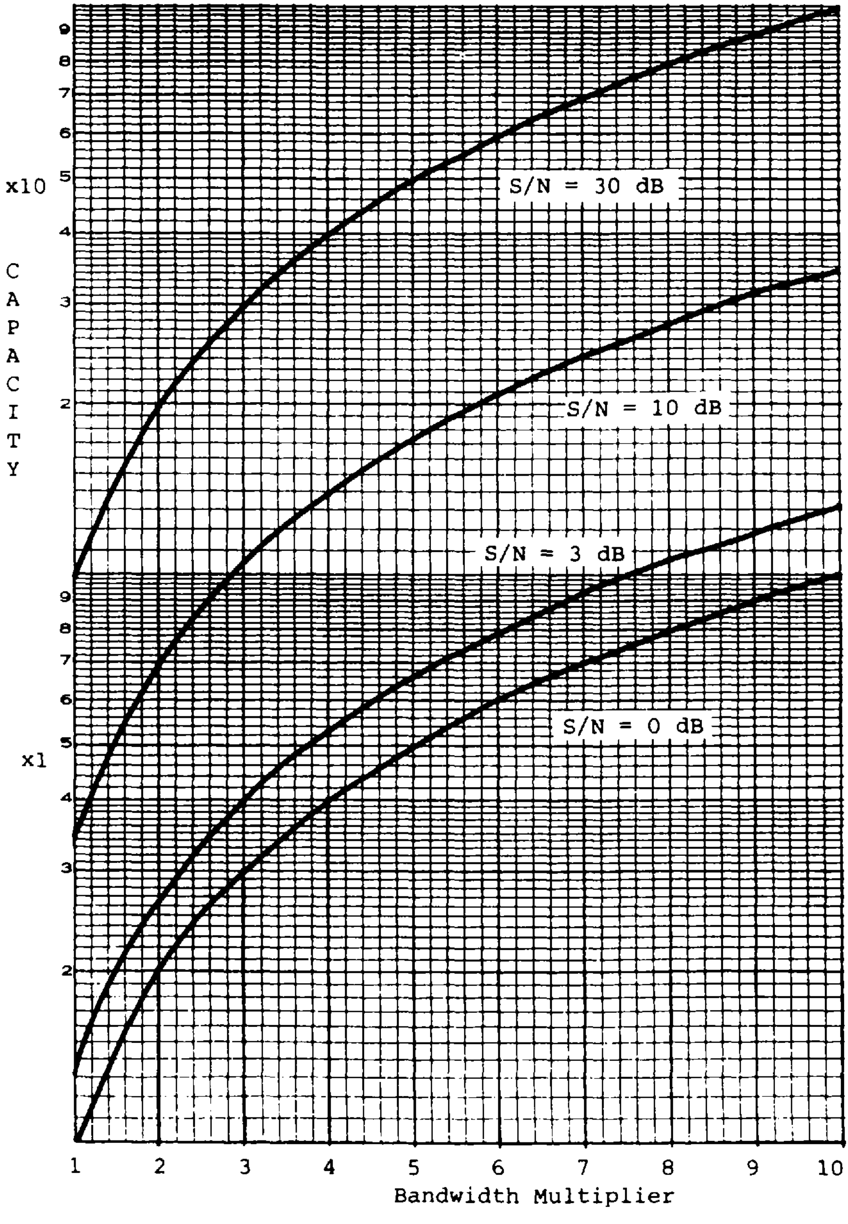


Figure 13.8 Capacity as a function of bandwidth and signal-to-noise ratio.

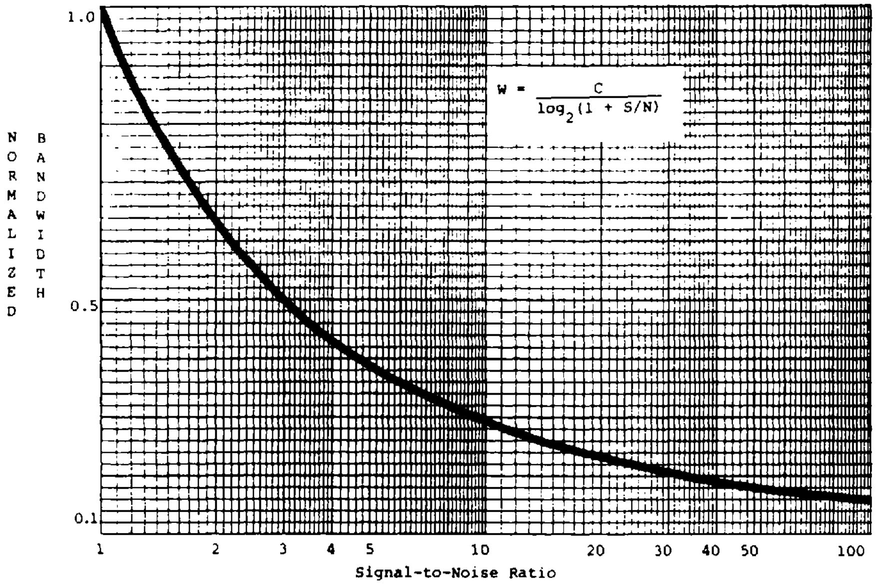


Figure 13.9 Bandwidth required as a function of signal-to-noise ratio. Capacity constant.

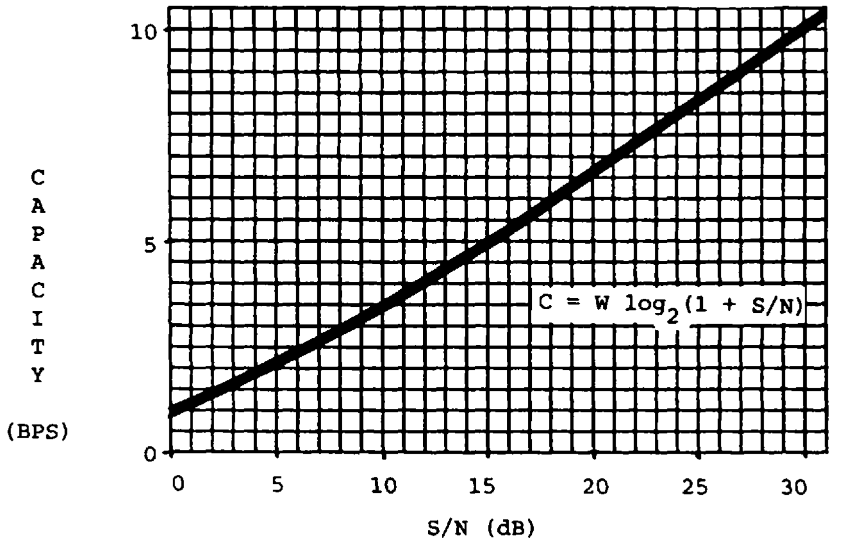


Figure 13.10 System capacity with constant bandwidth as a function of signal-to-noise ratio.

13.4 INTERLEAVER OPERATION

Interleavers and deinterleavers are used in communications systems to enable the use of error correction coding, when the errors expected are clustered in relatively infrequent bursts and the bursts are longer than the maximum span of errors that can be corrected by an error correcting code. Interleaving is employed in conjunction with an error correction scheme.

Suppose that we wish to send a data symbol 1001100, which is the symbol produced by adding parity 100 to data 1001 using Hamming single error correction encoding, and interference exists within the channel that has a duration of up to 7 bit times. An interleaver could be employed that would allow as many as 7 bits in error in a burst to be corrected.

Figure 13.11 shows a 7×7 interleaver consisting of a memory in which data symbols are entered in rows but are output in columns. That is, the data input would be

$$A_1A_2A_3A_4A_5A_6A_7\ B_1B_2B_3B_4B_5B_6B_7\ C_1C_2C_3C_4C_5C_6C_7 \cdots$$

while the output would be

$$A_1B_1C_1D_1E_1F_1G_1\ A_2B_2C_2D_2E_2F_2G_2\ A_3B_3C_3D_3E_3F_3G_3 \cdots$$

where the symbols have been rearranged in such a way that each of the input symbols $(X_1X_2X_3 \cdots X_7)$ is distributed over a seven-symbol distance at the interleaver output. This reassembled symbol stream is transmitted.

The deinterleaver (also shown in Figure 13.11) is a memory that is identical to the interleaver, but the received data symbols are entered in columns and output in rows, which restores each symbol's data to the same order in which it was input to the interleaver before transmission.

Now, if any group of up to 7 bits (one complete symbol) is interfered with by a burst of interference within the link, the deinterleaver distributes the effects of the burst across seven symbols. This means that input symbol $B_1B_2B_3B_4B_5B_6B_7$ is transformed along with six other symbols to

(Symbol 1)	(2)	(3)	(4)
$A_1\underline{B_1}C_1D_1E_1F_1G_1$	$A_2\underline{B_2}C_2D_2E_2F_2G_2$	$A_3\underline{B_3}C_3D_3E_3F_3G_3$	$A_4\underline{B_4}C_4D_4E_4F_4G_4$
(5)	(6)	(7)	
$A_5\underline{B_5}C_5D_5E_5F_5G_5$	$A_6\underline{B_6}C_6D_6E_6F_6G_6$	$A_7\underline{B_7}C_7D_7E_7F_7G_7$	

(Underlines show the distribution of the bits from symbol B to seven interleaved symbols.)

If a burst of interference coinciding with one of the interleaved symbols occurs [say symbol (4)], then received symbol (4) becomes

$$X_4X_4X_4X_4X_4X_4X_4$$

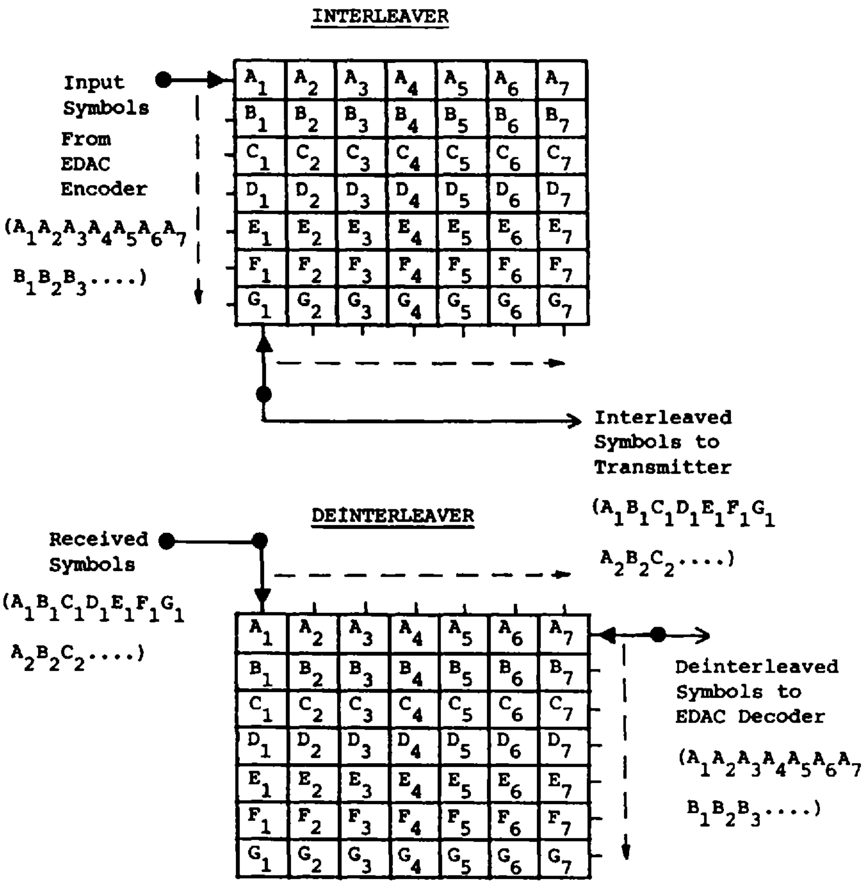


Figure 13.11 Interleaver and deinterleaver operation using 7 × 7 memory matrix. Interleaver stores input data in rows and outputs data in columns. Deinterleaver stores received data in columns and outputs data in rows.

After deinterleaving, however, the entire set of seven interleaved symbols becomes

$$A_1A_2A_3X_4A_5A_6A_7 \quad B_1B_2B_3X_4B_5B_6B_7 \quad C_1C_2C_3X_4C_5C_6C_7 \quad D_1D_2D_3X_4D_5D_6D_7$$
$$E_1E_2E_3X_4E_5E_6E_7 \quad F_1F_2F_3X_4F_5F_6F_7 \quad G_1G_2G_3X_4G_5G_6G_7$$

each of which contains only one bit that may be in error, and this potential error (potential in that the receiver may have guessed right even though the interference was present) is correctable by the embedded Hamming single error correction code.

In this way, an interleaver can distribute a burst of errors that cannot be corrected by the most common error correction codes, to occur in smaller groups that can be corrected by the same codes. The primary drawback with the use of an interleaver is the delay introduced in filling the memories employed by the interleaver and deinterleaver.

QUESTIONS

1. What is the distance between code words 00101100, 10100110, and 11101101?
2. How many errors can be corrected when these code words are used?
3. What is the Hamming weight of the word 110100100010101?
4. Generate the symbol that would be transmitted using a Hamming single error correction code with 0101 as input data.
5. What is meant be coding gain, and how is it measured?
6. What is the rate of a (21, 6; 3) encoder?
7. Given an input rate of 28 kbps and a rate 1/2 encoder, what would be the bandwidth of the output signal from an unfiltered BPSK modulator?
8. After removing the encoder, what would be the modulated signal bandwidth?
9. What would be the bandwidth of the signal in question 7 if the modulator were QPSK? OQPSK?
10. What is the meaning of the constraint length in a convolutional encoder?

REFERENCES

1. Hamming, R. W., Error detecting and error correcting codes, *Bell Syst. Tech. J.*, Vol. 29, 1950.
2. Peterson, W. W., *Error-Correcting Codes*, MIT Press, Cambridge, MA, 1961 (BCH code reference).
   Peterson, W. W., and Weldon, E. J., Jr., *Error Correction Codes* 2nd ed., MIT Press, Cambridge, MA, 1972.
3. Golay, M. J. E., Notes on digital coding, *Proc. IRE*, June 1949.
4. Reed, I. S., and Solomon, G., Polynomial codes over certain finite fields, *SIAM J.*, June 1960.
5. Shannon, C. E., A mathematical theory of communication, *Bell Syst. Tech. J.*, July and Oct. 1948.
6. Bhargava, V. K., et al., *Digital Communications by Satellite*, John Wiley & Sons, New York, 1981 (BCH decoding reference).

7. Edwards, G., *The Forward Error Correction Handbook*, Stanford Telecommunications, Inc., 1990.
8. Q 1650 k = 7 Multi-Code Rate Viterbi Decoder, Qualcomm, 1991/1992.
9. Hyper-FEC III CMOS Hard Decision Forward Error Correction Encoder/Decoder, Space Research Technology, Inc., 1987.
10. 32-Bit Parallel Error Detection and Correction Circuits, SN74AS632 and 634, Texas Instruments, 1986.
11. Rennie, L. J., Technical Notes, Space Research Technology, Inc., 1987/88.

This page intentionally left blank

14
Receiver Examples

Many excellent receivers exist today. They are found in a wide variety of applications, from very low frequencies to extremely high frequencies. In the following pages, a few examples of receivers designed for different purposes (all superheterodyne receivers, as it happens) are described. No effort was made to select the "best" receiver in any case, although those described are certainly not the worst that could be found with a little effort.

The first receiver described is a general-purpose AM/FM/SSB receiver that is useful over a frequency range of more than six octaves and is intended for general consumer (including amateur radio) use. This receiver is a quadruple conversion receiver when being used in the 1000- to 2000-MHz range and is a triple conversion receiver when operated in the 25- to 1000-MHz range. Note that it does not cover most of the 2.0- to 30-MHz band that is covered by what is usually termed a "communications receiver."

Because many of the users of the 2.0- to 30-MHz band are radio amateurs and others who routinely employ a communications band receiver in conjunction with a transmitter, and there are many good high-frequency (another commonly used name for the 2.0 to 30 MHz band) transceivers/receiver-transmitters available, section II of this chapter is a listing of tests done by the American Radio Relay League on currently available "2.0 to 30 MHz transceivers." One of the most interesting characteristics of all receivers in this band, whether part of a transceiver or not, is that they often do not employ a low-noise amplifier at the receiver input.

Military receivers are represented by both the AN/ARC-164 and AN/ARC-182 units, which are receiver-transmitters, and by the GPS (Global Positioning System) receivers, which are remarkable in that they must work with input signals at approximately -135 dBm.

It is noted that none of the receivers shown is a "digital" receiver, although digital processing is certainly vital in the latest GPS receivers, which simultaneously track as many as 10 satellites. Neither is it suggested that one cannot do better with either digital of analog techniques. Both are here to stay.

Indeed, it is certain that future receivers will be better, smaller, and lower in cost than anything other than our dreams can achieve today. Unless great strides are made in extrasensory perception, however, they will still have to conform to the same physical laws that limit us today.

14.1 ICOM R-7000 RECEIVER

An excellent example of a contemporary receiver that covers a very wide frequency range (25 MHz to 2 GHz) with excellent sensitivity is the ICOM R-7000. This receiver operates in AM, FM, narrowband FM, and single-sideband modes over its full operating band. Its architecture employs triple conversion, as illustrated in Figure 14.1. (In the 1000- to 2000-MHz range, quadruple conversion is used.)

A selectable 20-dB attenuator, to allow strong signal reduction, is provided at the receiver input, which is followed by four bandpass filters, selected according to the part of the band being received. The bandpass filters are followed by a preamplifier and a low-pass filter whose cutoff frequency is also selected by the frequency select control.

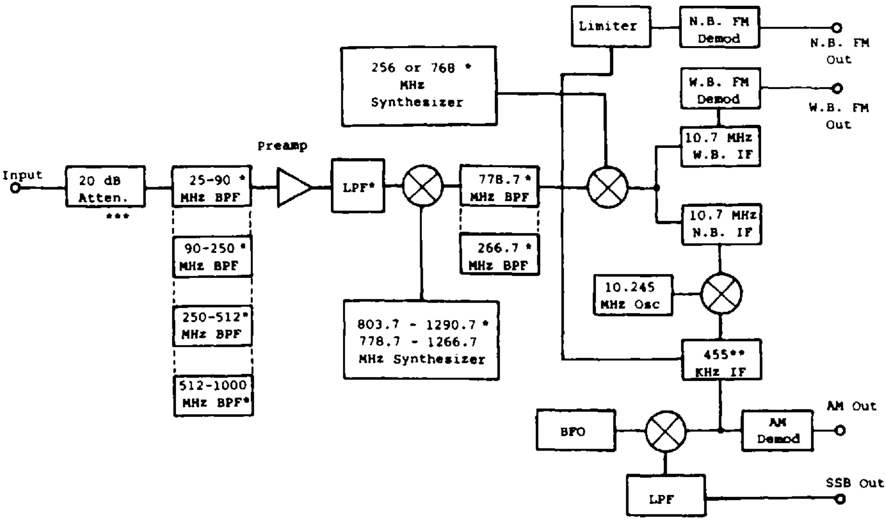


Figure 14.1 Block diagram of ICOM R-7000 receiver. (Simplified, 25 to 1000 MHz mode.) *Automatically selected with frequency. **Selected by bandwidth control. ***Individually selected.

The first mixer is driven by a frequency synthesizer whose output is in two bands. Band one is used in the 25 to 512 MHz received signal range and is 803.7 to 1290.7 MHz. Using high-side injection, and subtracting the input signal, produces a first IF frequency of 778.7 MHz. Band two covers 778.7 to 1266.7 MHz and is used with input signals in the 512- to 1000-MHz range. Again, subtracting the input signal from the local oscillator produces a first IF frequency of 266.7 MHz. High-side injection is also employed in this part of the band.

A pair of IFs, at 778.7 MHz in the lower frequency range and 266.7 MHz in the upper frequency range, are used. It is also of interest to note that the 1000- to 2000-MHz band extension mode is accommodated by adding an input down-converter that employs a 1000-MHz fixed local oscillator, which causes 1000 MHz to be subtracted from the input signal.

The 778.7-MHz IF signal is mixed with a fixed 768-MHz local oscillator signal, and the 266.7-MHz IF signal is mixed with a fixed 256-MHz local oscillator signal, which in both cases produces 10.7 MHz as the receiver's second IF frequency. This frequency is particularly advantageous because of the availability of excellent low-cost filters in both crystal and ceramic construction.

In the normal, wideband FM mode (200 kHz bandwidth) the signal is amplified, limited, and demodulated. Narrowband FM signals are converted to 455 kHz and filtered to either 15, 6, or 3 kHz bandwidth before limiting and demodulation.

AM and single-sideband signals are also converted to 455 kHz, with a bandwidth of 15, 6, or 3 kHz before demodulation. As with 10.7-MHz IFs, the use of 455 kHz as an IF frequency makes available a wide selection of the best bandpass filters at low cost.

Figure 14.2 shows the front panel of an R-7000 receiver, whose size is only

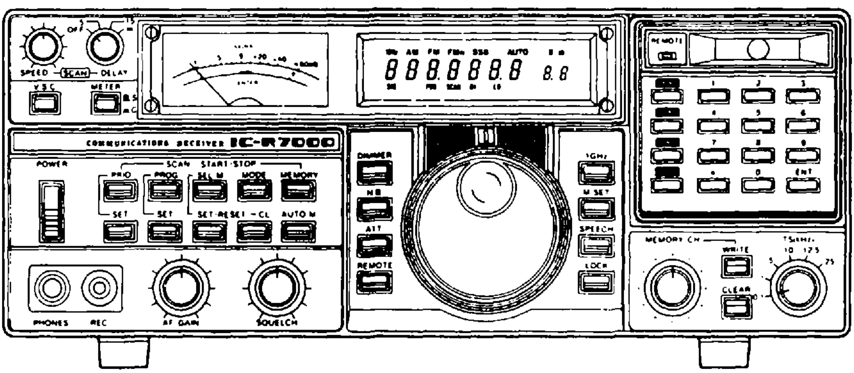


Figure 14.2 Front panel of the ICOM R-7000 receiver. Front panel width is 11.25 inches. (Courtesy of ICOM Corp.)

Table 14.1 ICOM R-7000 Receiver Specifications

• Receive frequency range

VERSION	FREQUENCY COVERAGE (MHz)
USA and EUROPE	25 ~ 999.999 *1025 ~ 1999.999
AUSTRALIA and FRANCE	**25 ~ 999.999 *1025 ~ 1999.999

* Specifications guaranteed from 1240 to 1300MHz.
** Excluding 87.5 to 108MHz.

• Receive modes	: A3E (AM), F3E (FM), J3E (SSB)
• Sensitivity	: 25 ~ 999.999MHz FM : Less than 0.5μV for 12dB SINAD FM (wide) : Less than 1.0μV for 12dB SINAD AM : Less than 1.0μV for 10dB S/N SSB : Less than 0.3μV for 10dB S/N 1240 ~ 1300MHz FM : Less than 0.5μV for 12dB SINAD FM (wide) : Less than 2.0μV for 12dB SINAD AM : Less than 2.0μV for 10dB S/N SSB : Less than 0.3μV for 10dB S/N
• Squelch sensitivity	: FM (Threshold) Less than 0.2μV for noise squelch FM (Tight) More than 32mV for meter squelch at S9+60dB SSB (Threshold) More than 3.0μV for meter squelch
• Selectivity	: FM, AM ±7.5kHz minimum at −6dB FM (narrow), AM (narrow) ±3.0kHz minimum at −6dB FM (wide) ±75kHz minimum at −6dB SSB ±1.4kHz minimum at −6dB
• Spurious and image response rejection	: More than 60dB
• Frequency stability	: 25 ~ 999.999MHz ±5ppm at 0°C ~ +50°C 1240 ~ 1300MHz ±10ppm at 0°C ~ +50°C
• Receive system	: 25 ~ 999.999MHz FM, AM, SSB Triple-conversion superheterodyne FM (wide) Double-conversion superheterodyne 1240 ~ 1300MHz: FM, AM, SSB Quadruple-conversion superheterodyne FM (wide) Triple-conversion superheterodyne
• Intermediate frequencies	: 25 ~ 512MHz: 1st 778.7MHz 2nd 10.7MHz 3rd 455kHz excluding FM (wide) mode 512 ~ 999.999MHz 1st 266.7MHz 2nd 10.7MHz 3rd 455kHz excluding FM (wide) mode
• Frequency control	: CPU based 100Hz step digital PLL synthesizer
• Number of memory channels	: 99 channels
• Supply voltage	: 117, 220 or 234V AC (50/60Hz)
• Current drain	: Receiving 1.7A at maximum audio output Squelched 1.4A
• Antenna impedance	: 50 ohms
• Audio output	: More than 2.5W at 10% distortion with an 8 ohm load
• Audio output impedance	: 4 ~ 8 ohms
• Usable temperature	: −10°C ~ +60°C
• Dimensions	: 286(303)mm(W) x 110(127)mm(H) x 276(319)mm(D) Bracketed values include projections.
• Weight	: Approximately 8.0kg (excluding options)

• All stated specifications are subject to change without notice or obligation.

Courtesy of ICOM Corp.

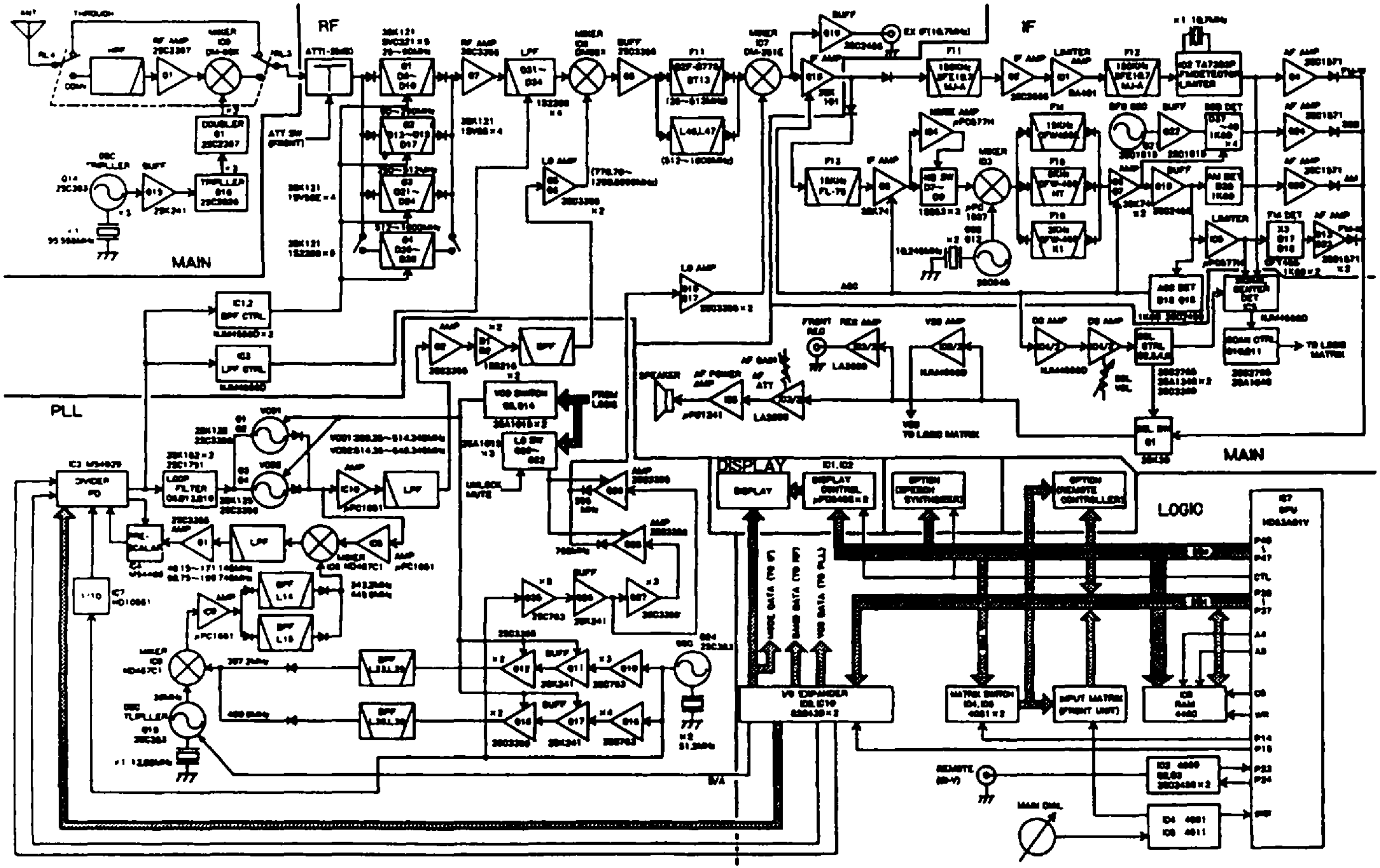


Figure 14.3 Complete block diagram of R-7000 receiver. (Courtesy of ICOM Corp.)

$11.25 \times 4 \times 11.5$ inches (not including handle or mounting feet). Specifications are listed in Table 14.1.

A complete block diagram of the R-7000 receiver is shown in Figure 14.3, including frequency synthesis and the down-conversion method employed to enable the receiver frequency range to be extended from 1000 to 2000 MHz.

14.2 HIGH-FREQUENCY (2 to 30 MHZ) TRANSCEIVER TEST RESULTS*

The results of tests done by the American Radio Relay League on over 90 transceivers from 18 manufacturers are given in Table 14.2. Test conditions are quoted in the following paragraphs.

> This section compares key test results from previously tested equipment. Since most HF equipment covers all HF bands, only exceptions are noted. A minus sign indicates that a band is not present. -WARC indicates that 30M, 17M, and 12M are not present. A plus sign indicated that the radio has an additional band. CW, USB, and LSB are assumed for modes. If any additional modes are available, they are indicated.
>
> * indicates that the measurement was noise limited at the value shown.
>
> nl indicates that the measurement was noise limited. The level at which it was noise limited is not known.
>
> -- indicates that data are not available.
>
> ? indicates that this was just added to the table, based on partial information. We still need to look up the rest of the data in the original review. It should be available on the next go around, but we wanted to offer what information we had readily available.

Key to notes found in the table:

1. The test was noise limited. Test records do not indicate at what level the noise limiting occurred.
2. The test was noise limited at the value shown. This is the level of dynamic range at which the noise increased by 1 dB.
3. This unit was tested on 15M instead of 20M.
4. Test data was taken with the preamplifier on.
5. Test frequencies are unknown.
6. Not tested on 20M.
7. IMD DR measured at 30 kHz spacing.
8. Not tested on 80M.
9. Also has 6 meter coverage.
10. BDR not performed on 20M.

---

*We are interested here in the test results given for the receiver portion of the transceiver.

11. AM receive only.
12. Measured on 6M only.
13. All HF bands available with optional plug-in modules.
14. Transmit IMD data taken on 24.950 MHz.
99. Measured by outside laboratory.

Table 14.2 shows three test results that are of special interest for receivers: MDS, which refers to minimum discernible signal sensitivity; BDR, or blocking dynamic range; and IMD, which is intermodulation distortion. Figures 14.4, 14.5, and 14.6 illustrate the test methods used by ARRL to make the tests that result in the numbers given in Table 14.2. These figures and descriptions are given through the courtesy of ARRL.

Table 14.2 HF Transceiver Comparative Table

Manufacturing Model	Date	Bands	Modes	Typical Output Power	Xmit Spurious dBc	Xmit -IMD 3rd/5th	MDS 80M/20M	BDR 80M/20M	IMD DR 80M/20M	Notes
Alinco										
DX-70T	12/95	+6 M	AM,FM	100		29/51	-138/-136	129*/126*	90/92	4
Atlas										
350XL	–	–	–	--	--	–	-131	117	81	99
210/215X	--	–	--	–	–	–	-120	123	76	99
Collins										
KWM-380	10/82		AM	100	-59	33/40	-131/-131	nl/nl ,	nl/nl ,	
Cubic										
Ast102	12/81	?	?	?	-49	-28/	-131	nl	84	
Drake										
TR-4C	–	–	–	–	–	–	-124	105	74	99
TR-7	5/79	-WARC	AM	100	-46	32/36	-133/-133	120/120	84/90	
Heathkit										
HW-5400	10/84	-160		100	-48	30/31	-135/-133	110/112	82/90	
HW-9	7/85	-160,-WARC	CW only	5	-42	n/a	-130/-128	124/122	98/88	
HW-99	3/86	-160,-20, -WARC	CW only	50		n/a	-124/-116	117/112	91/87	3
HW-104	12/76	?	?	?		40/	-125	94	71	
SB-104	–	–	–	–	–	–	-123	92	79	99
SB-1400	10/89		AM	100	-53	30/44	-136/-136	113/113	nl/nl	
SS-9000	2/84			100	-55	29/44	-138/-140	119/118	90/92	
ICOM										
IC-701	4/79	-WARC		100	-45	45/49	-133/-133	nl/nl ,	89/87	
IC-706	3/96	+6M, +2M	WBFM, AM,FM	100 HF 100 6M, 10 2M		35/32 HF	-140/-129	106*/104*	88*/87*	4
IC-707	4/94		AM,FM	100		28/36	-138/-138	115*/128*	93/87	4
IC-720A	8/82		AM	100	-58	28/52	-132/-132	nl/nl ,	97/92	
IC-725	3/90			100	-56	35/38	-138/-138	nl/nl ,	92/91	4
IC-728	2/93		FM	100	-50	39/40	-137/-137	116*/123*	91/92	4
IC-729	2/93	+6M	FM	100		48/44	-141	111*	85	4,12
IC-730	12/82	-160	AM	100	-50	40/46	-140/-140	nl/nl ,	nl ,/96	
IC-735	1/86		AM,FM	100	-65	33/39	-134/-133	nl/nl ,	92/88	4
IC-736	4/95	+6 M	AM,FM	100		24/37	-139/-139	118*/130*	92/92	4
IC-737	8/93		AM,FM	100		39/48	-139/-137	118/118	94/95	4
IC-738	4/95		AM,FM	100		40/40	-138/-139	116*/119*	93/94	4
IC-740	9/83			100	-63	30/40	-141/	125/	94/	4,5
IC-745	9/85		AM,FM	100	-65	34/40	-140/-144	115/116	92/94	4
IC-751	1/85		AM,FM	100	-60	36/44	-142/-138	nl/nl ,	91/93	4
IC-761	9/88		AM,FM	100	-56	38/45	-140/-139	120/122	95/96	4
IC-765	12/90		AM,FM	100	-64	40/44	-142/-142	148/146	98/96	4
IC-775DSP	1/96		AM,FM RTTY	200		30/45	-143/-143	135/132	104/103	4
IC-781	1/90		AM,FM	150		37/39	-137/-134	133/133	97/100	4

Table 14.2 Continued

Manufacturing Model	Date	Bands	Modes	Typical Output Power	Xmit Spurious dBc	Xmit -IMD 3rd/5th	MDS 80M/20M	BDR 80M/20M	IMD DR 80M/20M	Notes
Index										
QRP-Plus	9/96			5	-35	25/38	-132/	111*/111*	96*/93*	
JRC										
JST-135HP	3/92		AM,FM	150	-63	30/43	-132/-132	121*/117*	95/91*	
JST-245	9/95	+6 M	AM,FM	150		39/56	-138/-138	123*/126*	92/95	4
Kenwood										
TS-120S	2/80	-160,-WARC		100	-49	39/40	-139/	108/	75/	6
TS-130S	7/81	-160		100	-45	38/39	-138/-138	109/110	79/78	
TS-140S	6/88		AM,FM	100	-46	30/42	-137/-137	115/114	92/91	4,7
TS-180S	5/80			100	-50	40/42	-139/-139	112/114	82/83	
TS-430S	3/84		AM	100	-51	31/34	-138/-137	nl/nl ₁	95/89	
TS-440S	12/86		AM,FM	100	-43	28/46	-140/-139	112/111	89/89	
TS-450S	4/92		AM,FM	100	-50	35/49	-140/-141	109/108	70/71	4
TS-50S	9/93		AM,FM	100		30/42	-139/-139	110/109	86/88	4
TS-520S	5/78	-WARC		100		36/47	/-133	/104	/69	8
TS-530S	3/82			100	-42	28/40	-135/-136	112/120	88/90	
TS-570D	1/97		AM,FM,FSK	100	-58	28/42	-140/-139	119*/115*	99/98*	
TS-680S	10/88	+6M	AM,FM	100	-45	32/43	-140/-140	108/107	92/95	4,9
TS-690S	4/92	+6M	AM,FM	100	-50	32/38	-140/-141	109/108	70/71	4,9
TS-820	9/76	-WARC		100	-45	39/45	/-136	/114	/85	4,8
TS-830S	5/81			100	-45	32/50	-136/-136	129/nl ₁	83/82	
TS-850S	7/91		AM,FM	100	-64	28/40	-143/-141	141/148	100/99	
TS-870S	2/96		AM,FM	100		18/30	-141/-139	124/123	95/95	4
TS-930S	1/84		AM	100	-50	35/42	-139/-139	nl/nl ₁	88/87	
TS-940S	2/86		AM,FM	120	-54	37/43	-140/-139	141/138	93/97	
TS-950SD	1/91		AM,FM	150	-55	42/46	-143/-142	139/139	99/101	
TS-950SDX	12/92		AM,FM	150	-40	35/41	-139/-138	132/132	93/94	4
MFJ										
MFJ-9017	7/93	17M only	CW only	4		n/a	-130	101	80	
Oak Hills Research										
QRP-40	5/92	40 M only	CW only	3		n/a	-129	108	83	
Radio Shack										
HTX-100	2/92	10 M only	CW, USB only	25		30/44	-136	98	75	
Ranger										
RCI-2950	2/92	10 M only	AM,FM	25		21/37	-130	80	62	
S&S Engineering										
ARK-40	5/94	40 M only	CW	5		n/a	-127	95*	94	
Swan										
Astro-150	7/80	?	?	?	-44	29/	-127	114	84	
Ten Tec										
Argonaut II	1/92		FM, AM	5	-53	30/36	-139/-137	109/120	82/84	
Argosy	10/82	-160,-17,-12		50	-48	31/46	-133/-133	99/98	64/64	
Century 22	5/85	-160,-17,-12	CW only	20	-46	n/a	-131/-128	112/109	82/81	
Corsair	--	--	--	--	--	--	-131	130	93	99
Corsair II	8/87			100	-45	29/48	-127/-124	117/	84/80	10
Delta II	1/92		FM,AM	100	-47	33/44	-134/-130	109*/104	98/88	
OMNI B	--	--	--	--	--	--	-136	129	87	99
OMNI D	1/80	-17/-12		100	-48	30/36	-128/-139	115/125	94/90	
OMNI V	11/90			95	-48	30/45	-135/-136	135/135	95/97	
OMNI VI	1/92		FM,FSK	100	-41	39/39	-134/-136	124*/128*	95/100	
Paragon	5/88		AM	100	-56	33/49	-140/-137	138/136	102/101	
Scout	12/93	Note 13		50		25/35	-129/-125	119/119	86/87	

Table 14.2 Continued

Manufacturing Model	Date	Bands	Modes	Typical Output Power	Xmit Spurious dBc	Xmit IMD 3rd/5th	MDS 80M/20M	BDR 80M/20M	IMD DR 80M/20M	Notes
Yaesu										
FT-1000D	3/91		AM,FM	200	-45	36/42	-136/-136	137/154	94/98	4
FT-1000MP	4/96		AM,FM	100	-55	27/45	-130/-135	139/137	91/94	4,14
FT-101ZD	12/79	-WARC		100	-45	38/47	/-139	/112	/78	8
FT-101E	9/76	-WARC		100		34	108	81		
FT-102	10/83			100	-44	40/40	-127/-127	nl/nl₁	97/98	
FT-107M	4/81		AM	100	-47	32/41	-133/-133	nl/nl₁	82/90	
FT-301S/D	10/77	-WARC	AM	100	-55	40/46	-133	100	75	5
FT-650	10/91	6M,10M, 12M only	AM,FM	100		31/42	-139	114	82	
FT-707	7/81	-160	AM	120	-49	34/44	-126/-127	nl/nl₁	77/83	
FT-747GX	8/89		AM	100	-54	32/45	-136/-136	110/120	90/92	
FT-757GX	12/84		AM,FM	100	-58	34/45	-140/-137	nl/nl₁	90/89	4
FT-767GX	9/87		AM,FM	100	-56	40/45	-131/-136	117/115	92/85	4
FT-77	11/83	-160	AM,FM	100	-54	35/44	-140/-140	99/99	92/94	
FT-840	5/94		AM,FM	100		28/49	-137/-138	108*/113*	90/90	
FT-890	9/92		AM,FM	100	-50	30/46	-138-137	127*/127*	93/96	4
FT-900AT	2/95		AM,FM	100		35/43	-138/-137	120*/124*	91/99	4
FT-901DM	11/78	-WARC	AM,FM	100	-46	38/43	-137/-137	114/118	85/90	
FT-980	11/84		AM,FM	100	-56	37/45	-137/-138	nl/nl₁	nl/nl₁	
FT-990	11/91		AM,FM	100	-49	38/47	-133/-129	130*/131*	94/92	
FT-ONE	8/83		AM	100	-53	38/40	-133/-138	nl/nl₁	nl/nl₁	

This and related test information is available from the ARRL. Used by permission.

Receiver Noise Floor (Minimum Discernible Signal) Test:

Test Description: The noise floor of a receiver is the level of input signal that gives a desired audio output level that is equal to the noise output level. This is sometimes called "minimum discernible signal" (MDS), although a skilled operator can detect a signal up to 10 dB or so below the noise floor. Most modern receivers have a noise floor within a few dB of "perfect." A perfect receiver would hear only the noise of a resistor at room temperature. However, especially for HF receiving systems, the system noise is rarely determined by the receiver. In most cases, external noise is many dB higher than the receiver's internal noise. In this case, it is the external factors that determine the system noise performance. Making the receiver more sensitive will only allow it to hear more noise. It will also be more prone to overload. In many cases, especially in the lower HF bands, receiver performance can be improved by sacrificing unneeded sensitivity by placing an attenuator in front of the receiver. The more negative the sensitivity number expressed in dBm, or the smaller the number expressed in voltage, the better the receiver.

Key Test Conditions:
50-ohm source impedance for generators.; Receiver audio output to be terminated with specified impedance.
Receiver is tested using 500 Hz bandwidth, or closest available bandwidth to 500 Hz.

Block Diagram:

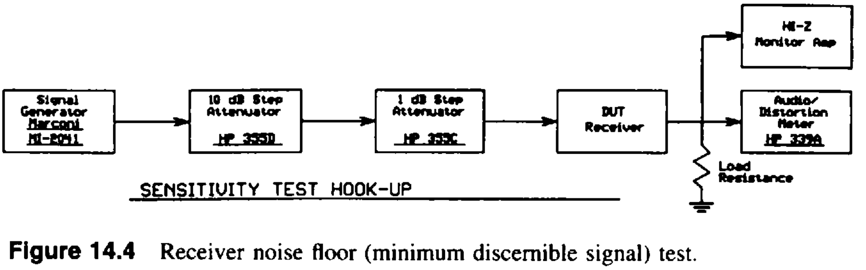


Figure 14.4 Receiver noise floor (minimum discernible signal) test.

Blocking Dynamic Range Test:

Test Description: Dynamic range is a measurement of a receiver's ability to function well on one frequency in the presence of one or more unwanted signals on other frequency. It is essentially a measurement of the difference between a receiver's noise floor and the loudest off-channel signal that can be accommodated without measurable degradation of the receiver's response to a relatively weak signal to which it is tuned. This difference is usually expressed in dB. Thus, a receiver with a dynamic range of 100 dB would be able to tolerate an off-channel signal 100 dB stronger than the receiver's noise floor.

In the case of blocking dynamic range, the degradation criterion is receiver desense. Blocking dynamic range (BDR) is the difference, in dB, between the noise floor and a off-channel signal that causes 1 dB of gain compression in the receiver. It indicates the signal level, above the noise floor, that begins to cause desensitization. BDR is calculated by subtracting the noise floor from the level of undesired signal that produces a 1-dB decrease in a weak desired signal. It is expressed in dB. The greater the dynamic range, expressed in dB, the better the receiver performance. It is usual for the dynamic range to vary with frequency spacing.

Key Test Conditions:
AGC is normally turned off; the receiver is operated in its linear region. Desired signal set to 10 dB below the 1-dB compression point, or 20 dB above the noise floor in receivers whose AGC cannot be disabled. The receiver bandwidth is set as close as possible to 500 Hz.

Block Diagram:

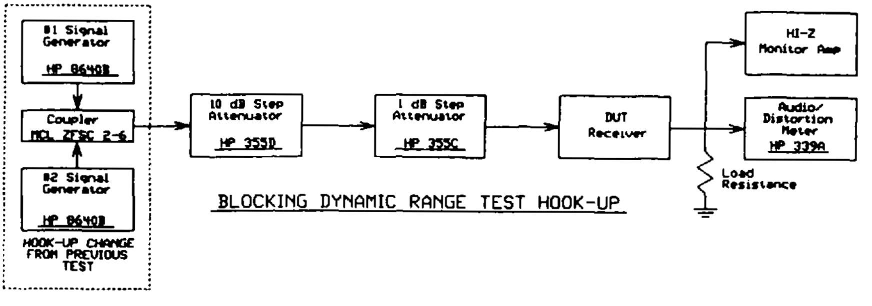


Figure 14.5 Blocking dynamic range test.

14.3 AN/ARC-164 RECEIVER (PART OF THE AN/ARC-164 RECEIVER-TRANSMITTER)*

A good example of a high-performance military receiver is the ARC-164 receiver, which is shown in block diagram form in Figure 14.7. This receiver is part of a receiver-transmitter, designed and built by the Magnavox Company, that operates in the 225- to 400-MHz band. It is currently used in many aircraft, primarily by the U.S. Air Force, for voice-mode applications. The ARC-164 is also capable of frequency hopping and is the base around which the "Have Quick" systems were developed.

*The difference between a "receiver-transmitter" and a "transceiver" is that a transceiver employs the same amplifiers and other subsystems in both the transmit and receive modes, while a receiver-transmitter is functionally much more autonomous, usually sharing only a frequency synthesizer, an antenna, and some switching/filter components between the receive and transmit modes.

In-Band Receiver IMD Test:

Test Description: This test measures the intermodulation that occurs between two signals that are simultaneously present in the passband of a receiver. Two signals, at levels of 50 uV (nominally S9), spaced 100 Hz are used. The receiver AGC is set to FAST. The receiver is tuned so the two signals appear at 900 Hz and 1100 Hz in the receiver audio. The output of the receiver is viewed on a spectrum analyzer and the 3rd- and 5th order products are measured directly from the screen. The smaller the products as seen on the graph, the better the receiver. Generally, products that are less than 30 dB below the desired tones will not cause objectionable receiver intermodulation distortion.

Key Test Conditions:

S9 or S9 + 40 dB signals
Receiver set to SSB normal mode, nominal 2 - 3 kHz bandwidth

Block Diagram:

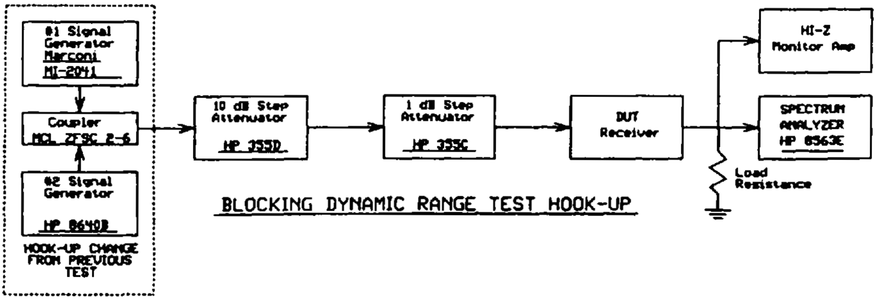


Figure 14.6 In-band receiver intermodulation distortion test.

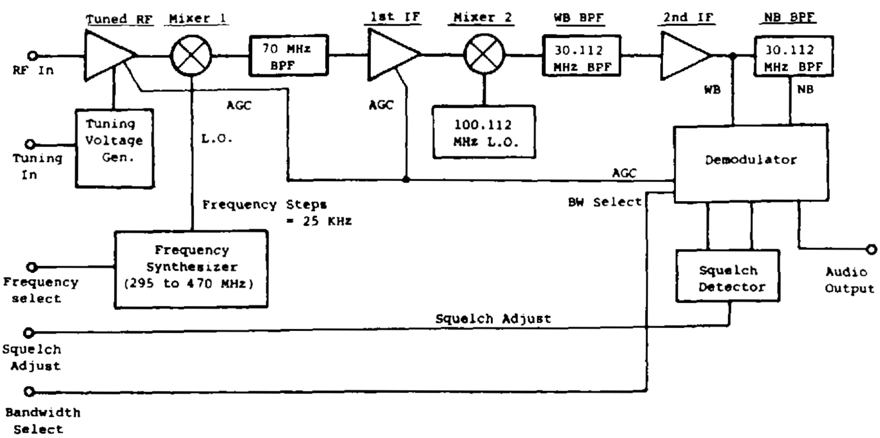


Figure 14.7 ARC-164 main receiver, simplified block diagram; 225 to 400 MHz, military aircraft radio receiver-transmitter.

ARC-164 development began in the late 1960s, which does not reflect on the receiver's performance but does show the time span over which military radios are designed and deployed. The ARC-164 receiver in fact has performance that is not equaled in many current designs. Table 14.3 lists the performance characteristics of the CA-663 version of the ARC-164, which is designed to transmit and receive both AM and FM signals within the 30- to 160-MHz band.

One of the unusual features of the ARC-164 receiver is that it includes a varactor-tuned RF preamplifier, which is not found in current higher frequency receiver designs. The RF amplifiers are dual-gate field effect transistors (FETs).

The ARC-164 receiver uses double-conversion superheterodyne architecture and includes three crystal filters for IF selectivity. The first IF is at 70 MHz

Table 14.3 AN/ARC-164, CA-663 version, 30 to 160 MHz AM-FM Receiver-Transmitter

CHARACTERISTICS

	FM MODE		AM MODE
Modulation Type	FM and Secure Voice		AM and Secure Voice
Frequency Range	30.000 to 79.975 MHz		108.000 to 159.975 MHz
Frequency Accuracy			
Standard Condition	±1.0 kHz		±1.0 kHz
Service Condition	±2.0 kHz		±1.5 kHz
Number Manual Channels	2000		2080
Number Preset Channels			
CA-663C	20 (Total AM & FM)		
CA-663R with CA-225 Control	32 (Total AM & FM)		
Channel Changing Time	0.25 Sec. Max.		
Input Voltage	18 to 33 VDC		
Input Power			
Receive	24 W (Nominal)		
Transmit	85 W (Nominal)		
Remote Indicator	ID-1994A/ARC, ID-2052/ARC		
Size (Inches): Panel Mount	4-7/8 H x 5-3/4 W x 6-1/2 D		
Remote	4-7/8 H x 5-3/4 W x 7 D		
Weight: Panel Mount	9 lbs Max.		
Remote	8.1 lbs Max.		
Environment			
MIL-E-5400	Class II Mod		Same as FM Mode
Temperature Range			
Operating	-54°C to +71°C		
Non-Operating	-62°C to +85°C		
Altitude	70,000		
Vibration	MIL-STD-810 (5 G's to 2000 Hz)		
Warm-Up Time	2 Sec. Max.		
Reliability (MTBF)	1000 Hours		
Transmitter			
Power Output			
Standard Conditions	10 W Min.		
Service Conditions	8 W Min.		
Duty Cycle			
Standard Conditions	Continuous Transmit @ +40°C		
Service Conditions	1:5, T:R at +71°C		
Receiver			
Sensitivity	0.5 μV for 10 dB (S+N)/N (Terminated)		3.0 μV for 10 dB (S+N)/N (Open Circuit)
Selectivity	NB	WB	
6 dB Points	22 kHz Min.	70 kHz Min.	Same as FM Mode
60 dB Points	44 kHz Max.	150 kHz Max.	
Audio Output	200 mW Min.		

SPECIFICATIONS FURNISHED ARE BELIEVED TO BE ACCURATE AND RELIABLE. HOWEVER, ALL SPECIFICATION DATA IS SUBJECT TO CHANGE WITHOUT NOTICE.

Magnavox

and employs a single 70-MHz crystal filter. The second IF, at 30.112 MHz, uses two crystal filters, and these provide for two separate modulated signal bandwidths. The first of the second IF filters has a 6-dB bandwidth of 70 kHz, while the second filter has a 22-kHz 6-dB bandwidth. This provides reception capability for AM and FM in normal and secure voice modes.

The same frequency synthesizer is employed in both receive and transmit modes. The synthesizer does use three VCOs to cover the required range, however. One VCO covers the 225- to 300-MHz range, a second covers 300 to 400 MHz, and a third is used for 400 to 470 MHz. (The ARC-164 receiver employs high-side injection, so the synthesizer must output 225 + 70 = 295 MHz to 400 + 70 = 470 MHz. In the transmit mode, the synthesizer outputs the operating frequency, 225 to 400 MHz.) The frequency synthesizer's output is typically in 25-kHz steps.

14.4 AN/ARC-182 RECEIVER (PART OF THE AN/ARC-182 RECEIVER-TRANSMITTER)

A receiver-transmitter similar to the ARC-164 but covering several bands between 30 and 400 MHz is the ARC-182, for which the specifications are shown in Table 14.4. The performance of this unit's receiver is comparable to that of the ARC-164.

Both the ARC-164 and the ARC-182 also include guard band receivers, as well as main receivers. The guard band receivers have only limited tuning capability, however, which is a function of the band to which the main receiver is tuned. Guard band frequency is

30 to 88 MHz band—40.5 MHz
108 to 160 MHz band—121.5 MHz
156 to 174 MHz band—156.8 MHz
225 to 400 MHz band—243.0 MHz

14.5 GLOBAL POSITIONING SYSTEM

Receivers employed in the Global Positioning System operate routinely with signals in the −135 dBm (approximately 0.035 μV at 50 ohms). The receivers work with spread spectrum signals whose bandwidth (3 dB) is either 0.9 or 9.0 MHz, which means that their effective input noise (kTB) is −114.5 to −104.5 dBm. This means that the input signal-to-noise ratio (given a noise figure of 2.0 dB) to such a receiver is −22.5 to −32.5 dB!

A serious question that must be considered at this point is, how can a GPS

Table 14.4 Rockwell AN/ARC-182 Specification for 30- to 400-MHz Receiver-Transmitter

SYSTEM SPECIFICATIONS

Frequency range	30.00 to 87.975 MHz FM. 108.00 to 117.975 MHz, receive only, AM. 118.00 to 155.975, AM. 156.000 to 173.975, FM. 225.000 to 399.975, AM/FM.
Channel spacing	25 kHz in all bands.
Guard receiver	40.50 MHz FM. 121.50 MHz AM. 156.80 MHz FM. 243.00 MHz AM. *The frequency of operation of the guard receiver is determined by the band of operation selected for the main receiver.*
Sensitivity	3 μV for 10 dB (s+n)/n ratio of 30-percent modulation AM. 1 μV (156-174 MHz) for 10 dB (s+n)/n ratio for 8-kHz deviation FM. 0.6 μV (30-88 MHz)
Transmitter power output	10 watts CW minimum - AM. 15 watts CW minimum - FM.
Duty cycle	5 minutes receive, 1 minute transmit.
Transmitter modulation capability	
AM	NLT 80 percent upward for 90 percent downward.
FM	Deviation capability of NLT 25 kHz, (5.6 kHz nominal - normal voice, secure voice).
Transmitter modulation response	
AM	Narrow band, 10 Hz to 12,000 Hz. Wide band, 10 Hz to 25,000 Hz.
FM	10 Hz to 12,000 Hz.
Receiver audio output	250 milliwatts into 150/600 ohms.
Receiver audio response	
Normal voice	400 to 3000 Hz AM/FM.
Wide band audio	10 to 25,000 Hz
Image	80 dB minimum.
Spurious rejection	70 dB minimum
Selectivity	
Narrow band	6 dB, 37 kHz minimum. 60 dB, 50 kHz maximum.
Wide band	6 dB, 69 kHz minimum. 60 dB, 170 kHz maximum.
Environmental	MIL-E-5400 class II with NMT 3-dB power reduction at 71 °C.
Input power	+28 V dc, in accordance with MIL-STD-704B.
Frequency stability	±1 PPM.

PHYSICAL CHARACTERISTICS

Weight	4.54 kg (10 lb).
Size	14.6 cm wide (5-3/4 in), 12.4 cm high (4-7/8 in), 16.5 cm deep (6-1/2 in), behind the mounting surface.

EQUIPMENT LIST

Options	Panel or remote mount transceivers, remote control, indicator.
ADF/homing	ADF in vhf-AM and uhf-AM, homing in vhf-FM (30-88 MHz).

Electronic Systems Group/Rockwell International
Cedar Rapids, Iowa 52406

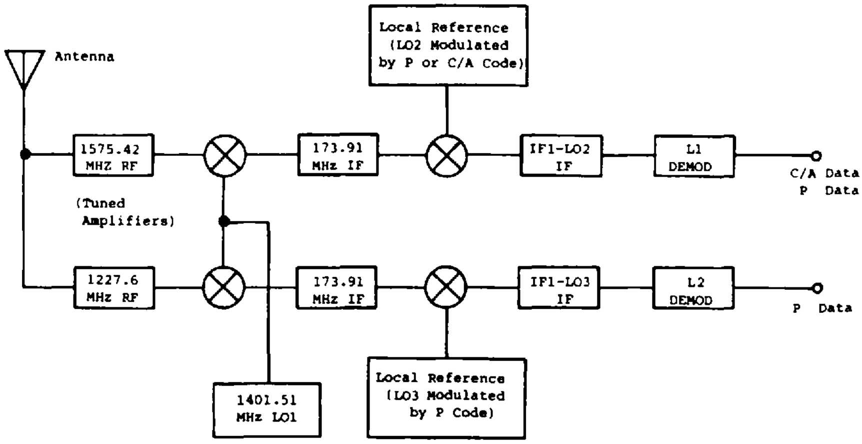


Figure 14.8 Simplified block diagram of L1/L2 (military-type) GPS receiver.

receiver work under such conditions? It works because the data rate being received is only 50 bps, and the spread spectrum bandwidth employed to send the signal is 1 to 10 MHz. Therefore the process gain in the receiver is approximately $10\log[(1 \times 10^6)/50] = 43$ dB to $10\log[(1 \times 10^7)/50] = 53$ dB, and this is the improvement in signal-to-noise ratio that is applied to the input signal-to-noise ratio, which brings it to the range of +20 to +30 dB in the data bandwidth. The signal, with its now high signal-to-noise ratio, is then demodulated with error rates similar to those of narrowband receivers with a high RF input signal-to-noise ratio.*

This improvement in signal-to-noise ratio is not the primary reason that GPS is a spread spectrum system, however. The primary reason is that the spread spectrum modulation/demodulation process used is much more robust in the presence of interference than any other technique available. The interference rejection capability of a GPS receiver is in the range of 43 dB − 10 dB to 53 dB − 10 dB, since the interference rejection capability of a spread spectrum receiver is approximately equal to its process gain in dB minus the signal-to-noise ratio required at the demodulator.

Figure 14.8 is a block diagram of a generic GPS receiver, such as might be used for reception of either of the two GPS downlink frequencies. The GPS system is employed widely today in both military and civil applications, and

*A mistake often made, because of the process gain improvement in GPS and other spread spectrum receivers, is that an improvement in sensitivity over narrowband receivers has been achieved. Not so. A narrowband receiver designed for a 50-bps data rate would have similar performance.

although the L2 signal (1227.6 MHz) is really intended for exclusive use of military and other "authorized" users, there are many civil applications of the L2 signal. (Only the P code restricts full use of the L2 signal by unauthorized users.)

In this block diagram, a single local oscillator is employed to down-convert both the L1 (1575.42 MHz) and the L2 (1227.6 MHz) signals to a single IF frequency (173.91 MHz) where the resulting signals are physically separated. (They could be amplified in the same IF, as two signals uniquely identified by having different codes.) Separate local reference signals, which consist of a local oscillator that is modulated by the particular code that is being received, are then used to correlate the received coded signals. When the input signals and the local code(s) are synchronized, a narrowband, 50-bps modulated signal is output to either or both of the second IF amplifiers, whose bandwidth is approximately 100 Hz, and then demodulated by the L1 and/or L2 demodulator(s). The data received at any given time are a function of whether the receiver is in the synch acquisition or postacquisition mode and whether it is a military or civil receiver. (Civil receivers cannot demodulate P-coded signals without authorization.)

This block diagram does not show the processor used to compute the receiver's position or to derive the system time, both of which are done from the range measurements carried out by the receiver, aided by the satellite downlink information carried within the received data stream.

14.6 PERSONAL COMMUNICATIONS SYSTEMS

Personal communications systems have rapidly gained recognition and have sprung into being very rapidly in the past few years. The first such systems were in fact the cellular telephone systems that came into widespread use recently and have become so popular that they have already reached the point where they cannot be readily accessed in many parts of the country, which led to the call for new frequency allocations and techniques that might provide more capacity. This in turn has led to the allocation of a band of frequencies from 1850 to 1990 MHz, in fact called the "PCS band," in which slices of the overall band have been auctioned to the highest bidder.

New frequencies have been allocated, but unfortunately the technology being used is warmed-over technology designed primarily for the cellular band. However, the best of integrated circuit technology is being applied to the construction of new equipment (including receivers) so that the new equipment being installed for PCS applications can service the new band as economically as possible.

A typical PCS receiver has a sensitivity of -102 to -104 dBm and is designed to handle digital voice at a rate of 9600 to 14,400 bits per second. The actual data rate employed in a typical time division multiple access (TDMA) receiver such as GSM (Global System for Mobility) is 272.3 kbps, which allows eight simultaneous users in a two-way link. Base stations receive in the 1930- to 1990-MHz range, and users receive in the 1850- to 1910-MHz range, which prevents interference from base station to base station and handset to handset but does nothing to prevent base-to-handset or handset-to-base interference.

As a result, every receiver can see multiple transmitters that are within 40 MHz of the frequency they are trying to receive, and this leads to the potential need for receivers with very high dynamic range and very low third-order products. A handset, for example, that can easily be within 10 m of a base station for another license, but whose frequency may be separated by as little as 1.2 MHz (two GSM licensees, each at their band edge), may transmit as much as 2 W of power (+33 dBm). The typical base may have as much as 13 DB of antenna gain, which means that the signal from a handset may appear at the base station at a level as high as

$$P_t + G_H + G_B - \text{propagation loss} = +33\,\text{dBm} + 13\,\text{dB} - 58\,\text{dB} = -12\,\text{dBm}$$

and the signal from the base appearing at handset may easily be

$$P_t + G_B + G_H - \text{propagation loss} = +43\,\text{dBm} + 13\,\text{dB} - 58\,\text{dB} = -2\,\text{dBm}$$

where P_t is 2 W (eirp) for the handset and 20 W (not including antenna gain) for the base; G_H is handset antenna gain (included in the 2 W eirp); G_B is the base antenna gain, 13 dB; and the propagation loss at 10 m is 58 dB.

These results show that the signal level that must be handled by both the base and the handset is very high, which means that both must be be designed with very high dynamic range. Even though it appears that the base station's input signals are less demanding, it should be realized that several handsets may be within a very short distance of the base station receiver.

The environment for PCS receivers may be one of the most difficult ever encountered. It remains to be seen how well the design challenges will be met.

This page intentionally left blank

Appendix 1
Glossary

A/D	Signifies a conversion from analog to digital. Analog signals are often converted to a digital format for processing by a computer or for transmission using a discrete form of carrier modulation.
Adjacent channel rejection	One of the most important characteristics of a receiver, adjacent channel rejection is the ability to reject signals in the channel closest to the channel that is desired.
AFC	Automatic frequency control, a technique applied in a receiver to maintain its IF frequency in the center of its demodulator's operating range. This is usually accomplished by feedback control of the receiver's local oscillator frequency. This locks the receiver to the incoming signal and neutralizes the effects of local oscillator drift. AFC was vital in the days of vacuum tube receivers.
AGC	Automatic gain control, applied in a receiver to provide the same performance from the receiver while its input RF signal is varying over a range of levels. AGC is usually applied to the receiver's RF and IF stages and serves the purpose of maintaining the demodulator input at a near-constant level.
AM	Amplitude modulation, which is one of the methods by which information can be impressed on a carrier signal.
Analog	Analog functions are those that can take on values that are continuous over a given range: given an $f(t)$ that is continuous over $f(t_1)$ to $f(t_2)$, there is an infinite number of values that $f(t)$ can assume as $\Delta t \rightarrow 0$.

ASK	Amplitude shift keying, a form of amplitude modulation in which the carrier is shifted to one of M discrete levels, each conveying $\log_2 M$ bits of information.
Automatic noise limiter	A type of limiter that prevents noise impulses from overdriving the stages in a receiver.
AVC	Automatic volume control, controls the audio stages of a receiver to keep the output at a constant level, by varying audio gain. The AVC time constant is designed to be slow compared with the audio rate of change, so that it does not compress the dynamic range of the information itself. Older receivers often employed AVC without AGC.
Bandwidth	The range of frequencies over which a receiver or some component of a receiver has a significant response. Usually characterized at a specific level, such as "3-dB bandwidth."
Baseband	The receiver's range of response for the information it conveys. For example, the baseband response of a voice receiver is typically 300 to 3000 Hz.
BER	Bit error rate; the rate at which errors can be expected to occur in a receiver, expressed as a fraction of the number of bits of information. BER = number of errors/number of bits sent.
BFO	Beat frequency oscillator, used in a receiver to enhance reception of unmodulated Morse code keyed transmitters or to enhance the intelligibility of single-sideband signals.
Binary	The most prevalent number system for representing digital information in modern communications and data processing systems. In such systems, information is represented as binary numbers, where the number of bits (binary digits) required to represent a quantity is $\log_2 N$. For example, if $N = 3$, 2 bits are required, for $N = 37$, 6 bits are needed, and if $N = 9732$, 14 bits are needed. The advantage of the binary system for communications is that on-off or yes-no decisions are practical, even though a wide range of values can be signified to a precision that is limited only by the number of bits that can be transmitted.
Bit	The smallest unit of information, a binary digit. Any N-level information function can be defined uniquely by a series of words, each composed of $\log_2 N$ bits.
BPSK	Biphase shift keying, wherein a carrier is shifted to one of two phases (usually $\pm$ 90°) to represent a one or a zero.
CDMA	Code division multiple access; any of a group of methods used to allow multiple users of a band of frequencies,

	wherein codes are employed to allow identification and separation of users. CDMA techniques are spread spectrum techniques.
Clippers/limiters	Devices employed in reducing amplitude variations on a signal; an alternative to using AGC. Limiting IFs are commonly used in FM receivers.
Costas loop	A type of carrier recovery and demodulation loop used to demodulate phase shift keyed and other double-sideband suppressed-carrier signals. First proposed by J. P. Costas.
Data	Plural of datum. Datum is "something used as a basis for calculating or measuring." Data transmission is usually in binary form.
Delayed AGC	A misnamed (delayed AGC has nothing to do with time or time delays) offset of the automatic gain control signal used to adjust the gain in a receiver, applied only to the receiver RF preamplifier. The purpose of delayed AGC is to prevent desensitization of the receiver due to small signals that are marginally greater than the minimum receiver input level. That is, delayed AGC prevents the AGC loop from reducing preamplifier gain unless the input signal is well above (10 to 20 dB minimum) the minimum input level.
Deviation ratio	The ratio of the peak frequency shift in an FM signal compared with the modulation rate. With FSK modulation, deviation ratio is the peak-to-peak frequency shift divided by the data rate.
Digital	"Of or relating to calculation by numerical methods or by discrete units." Digital communications include all of the techniques that send information in such a way that the signal employed varies in discrete steps. Typical of these are ASK, PSK, and FSK techniques.
Direct sequence	One of three basic spread spectrum modulation techniques, wherein a carrier is directly modulated by a high-speed (relative to the data rate) code sequence, which "spreads" the resulting information-bearing signal over a bandwidth that is much wider than the information itself requires.
Discriminator	One of a group of noncoherent FM demodulators. An example is a Foster-Seeley discriminator.
DPSK	Differential phase shift keying; identical to phase shift keying except that a phase shift occurs only when the bit that follows is a one. This technique solves the problem that a phase shift key demodulator can lock in either of two phases, and this in turn causes the demodulated data to be

	ambiguous with respect to which phase is a one and which is a zero.
Dynamic range	The range of RF input signal level over which a receiver will provide a specified output $(S + N)/N$ ratio or bit error rate. Dynamic range is also often specified for receiver stages such as the preamplifier or mixer(s).
EDAC	Error detection and correction.
Envelope detector	A noncoherent demodulator used for demodulation of amplitude-modulated or amplitude shift keyed signals.
FDD	Frequency division duplex; a technique for operation of a receiver and a transmitter in conjunction with one another wherein two frequencies are used, with the transmitter operating at one frequency and the receiver operating at the other.
FDM	Frequency division multiplex; the method of assigning different users to separate frequencies within a band to allow differentiation between them. This is the classic method upon which the commercial broadcast radio system is built.
FDMA	Frequency division multiple access; see FDM.
FEC	Forward error correction; forward error correction is the process by which codes that are properly designed allow error detection and correction, within the Hamming distance restrictions of the particular codes employed.
FIFO	First in–first out; one of the techniques for ordering the transmission of data, typical for time division systems.
FM	Frequency modulation, which is one of the methods by which information can be impressed on a carrier signal.
Frequency hopping	One of the three basic forms of spread spectrum signaling, in which a code is employed to determine the frequency of operation of the transmitter and receiver, and the frequency is therefore a pseudorandomly ordered set of short transmissions over the band of operation. This technique is also suitable for CDMA operation.
Frequency synthesizer	A frequency generator that develops receiver local oscillator signals from one or more reference frequencies.
FSK	Frequency shift keying; a modulation technique by which data can be represented through discrete shifts between two or more frequencies.
Hamming distance	The number of differences between two code words or symbols.
IF	Intermediate frequency.
Image frequency	A spurious response caused by the ambiguity that occurs because a signal either above or below the local oscillator

	in a superheterodyne receiver can cause a difference signal that falls at the IF frequency.
Image rejection	The degree to which a receiver's image response is less than its response to a signal at the frequency to which it is tuned.
Image response	The response of a receiver to a signal at the image frequency.
Information	The intelligence that a receiver is designed to demodulate and deliver to its user.
Intersymbol interference	The interference caused by energy intended to be in a given bit interval that encroaches on another (usually adjacent) bit interval. Manifested as increased BER.
LO	Local oscillator
Local oscillator	A signal internal to a receiver that is used to convert (by heterodyning) a received signal to another frequency for convenience in processing.
Manchester coding	A technique in which the data clock for a data signal being transmitted is modulo-two added to the data, which prevents the transmission of sequences of either ones or zeros that are longer than 1 bit and thereby relaxes a receiver's low-frequency response requirements. At the same time, however, the apparent information rate is doubled.
Modulation	The process by which information is conveyed in a communications system. Three basic forms of modulation are available for impressing information on a carrier. These are 1) amplitude changes, 2) frequency changes, and 3) phase changes.
MPSK	M-ary phase shift keying; a modulation technique in which the carrier phase is shifted to one of M phase positions, usually defined as a multiple of $360/M$ degrees. Each phase shift conveys $\log_2 M$ bits of information.
MSK	Minimum shift keying; a form of frequency shift keying in which the shift between frequencies is phase continuous, and the deviation ratio is 0.5. Many variations of this modulation form exist, such as GMSK (Gaussian MSK) and SFSK (sinusoidal FSK).
Multipath	A term used to refer to the action in which more than one signal appears at a receiver due to the signal having taken multiple paths on its way from the transmitter. The variable phase and amplitude relationship(s) of the multiple signals that results causes degradation in the signal presented to a receiver.
Multiple access	A general term used to refer to a class of techniques used to provide access to a band of frequencies by more than one user, where the users may be either dependent on or inde-

	pendent of one another. Techniques generally applied are frequency division, time division, and code division. (See FDMA, TDMA, and CDMA.)
Multiplex	A term that refers to combining of two or more signals in such a way that a single service channel may be employed to provide a transmission path that allows them to be demultiplexed at a receiver or receivers as if they were sent over independent paths. (Ideally, a multiplex/demultiplex process is transparent to the user.)
NF	Noise factor.
Noise factor	The factor by which a receiver increases its effective input noise, kTB.
Noise figure	10 log NF. Signified as F.
OOK	On/off keying; a form of ASK in which the carrier is turned on and off to signify a one or a zero. Also called 2-ASK.
OQPSK	Offset quadriphase shift keying; a form of QPSK in which the two modulating bit streams are offset in time by half a bit time, to reduce incidental amplitude modulation of the carrier.
PAM	Pulse amplitude modulation; an encoding technique in which a pulse is generated whose amplitude is a function of the input analog signal level at the sample time. This pulse is then used to amplitude modulate a carrier.
PCM	Pulse code modulation; an encoding technique in which an N-bit symbol is generated as a function of the level of the input analog signal at the sample time. Symbol size (in bits) determines resolution, where the number of steps to which the analog signal is quantized is $\log_2^{-1} N$.
PCS	Personal communications services (or systems).
Percent modulation	The percentage of modulation used by a transmitter as compared with the maximum that is allowable.
Phase lock loop	A type of feedback servo loop in which a voltage-controlled oscillator is caused to track an input signal's frequency or harmonic thereof. Used in both frequency synthesis and demodulation.
PLL	Phase lock loop.
PM	Phase modulation.
PSK	Phase shift keying; a modulation technique in which the carrier may be shifted to one of M phases, where the information signified by any particular phase shift is $\log_2 M$ bits.
QAM	Quadrature amplitude modulation; a modulation technique in which both phase and amplitude shifts are employed to represent data symbols.

QPR	Quadrature partial response; a modulation technique similar to QAM, but in which the bandwidth is restricted.
RF bandwidth	The 3-dB or half-power bandwidth (two-sided).
RSSI	Received signal strength indication.
Selectivity	The frequency difference required to produce a given amount of rejection of a signal by a receiver, compared with the receiver's response when the signal is at the frequency to which the receiver is tuned. This is the general case of which adjacent channel rejection is a specific instance.
Sensitivity	The minimum signal level required at a receiver's RF input to produce a given output $(S + N)/N$ ratio or BER, when standard modulation is employed.
SFSK	Sinusoidal frequency shift keying; a type of minimum shift keying in which the carrier is continuously changing frequency.
Shape factor	The ratio of the rejection bandwidth of a filter and its passband bandwidth.
Spread spectrum	A general term applied to one of three techniques (direct sequence, frequency hop, and chirp) employed to provide a process gain against interference in a properly designed receiver.
Squaring loop	A coherent demodulator used to demodulate double-sideband suppressed carrier signals, such as PSK.
Squelch	A circuit or signal used to prevent a receiver from outputting a signal as information when a valid RF signal is not present.
SSB	Single sideband; a type of modulation in which only one sideband is transmitted. The carrier is usually suppressed, although a vestigial carrier is sometimes sent.
Standard modulation	The modulation percentage, and rate, used for testing receivers. In AM, for example, standard modulation is 30 percent at 1.0 kHz.
Superheterodyne	The most prevalent type of receiver in use. A superheterodyne receiver consists of a tuner that converts any signal in the desired signal range to a fixed intermediate frequency and then demodulates it.
Superregenerative (including regenerative)	A type of receiver that employs regenerative feedback through a frequency-selective filter to improve selectivity. The superregenerative receiver adds a quenching circuit to improve stability.
TDD	Time division duplex; a technique used in systems that allows a transmitter and a receiver to operate at a single frequency by taking turns.

TDM	Time division multiplex; the method in which different users of the same frequency band are assigned different time periods for their transmissions. All users take turns using the entire bandwidth provided.
TDMA	Time division multiple access. See TDM.
Threshold	The minimum signal-to-noise ratio required for a demodulator to demodulate a signal while providing the minimum output signal-to-noise ratio or BER.
TRF	Tuned radio frequency; a type of receiver in which all amplification and demodulation are carried out at the RF input frequency.
VFO	Variable frequency oscillator. See BFO.

Appendix 2
Spread Spectrum Receivers

Spread spectrum receivers are separable into three different types, according to the three spread spectrum signal categories:

1. Direct sequence, in which the signal consists of a carrier that is modulated by a high-speed code sequence. The carrier modulation techniques themselves are usually either phase shift key or minimum shift key, and the information is in phase shift or frequency shift form after the spreading code is removed.
2. Frequency hopping, where the frequency is shifted periodically under the direction of a code sequence. Frequency hopping systems usually employ frequency shift keying of the frequency-hopped carrier to send their information, and the receiver employs a noncoherent FSK demodulator.
3. Chirp (also called pulse-FM), wherein the carrier is swept over range dF in time T and the receiver employs a dispersive filter to detect the signal.

Receivers for spread spectrum signals employ the same techniques that are used in non–spread spectrum receivers, and the same circuits, but they are often somewhat more complex. The reason is that spread spectrum receivers usually deal with signals for which it is necessary to carry out some form of demodulation on more than one level. For example, a spread spectrum receiver often has to despread the signal and then demodulate the information that was embedded within the spreading code. In addition, a spread spectrum receiver may have to perform a separate synch acquisition process before it can get down to the business of despreading the signal and demodulating it. Here we will examine the techniques that are peculiar to spread spectrum receivers.

A2.1 DIRECT SEQUENCE RECEIVERS

Direct sequence receivers are, like other receivers, almost always of superheterodyne architecture. Figure A2.1 is a block diagram of a typical direct sequence spread spectrum receiver, employing double conversion, with the despreading operation performed after the second conversion, at the second IF frequency.

In the case of a direct sequence receiver, this particular architecture is advantageous because it puts the local reference generator at a fixed frequency. The local reference generator consists of the second local oscillator (labeled "Fixed L O" in Figure A2.1), a mixer that is used to modulate the second local oscillator, the reference code generator, and a bandpass filter. Because it is more difficult to maintain good balance in the local reference generator over a wide band of frequencies than at a single frequency, it is the usual practice to modulate the second local oscillator with the reference code, rather than the first local oscillator (which is usually tuned). Thus the despreading operation, in which the transmitted code is removed from the received signal, occurs at the second downconversion.

If we follow the signal through the direct sequence receiver, the following sequence of events occurs:

1. First, the RF input signal is filtered to reject unwanted input signals that are outside the signal bandwidth. This is especially important in receivers that must operate in the presence of interference. The filtering process may be either a part of the amplification process or separate.

2. The amplified and filtered RF signal is then mixed with a local oscillator signal and converted to the first IF frequency, where it is amplified and filtered to match the local reference bandwidth characteristics. It is important that the

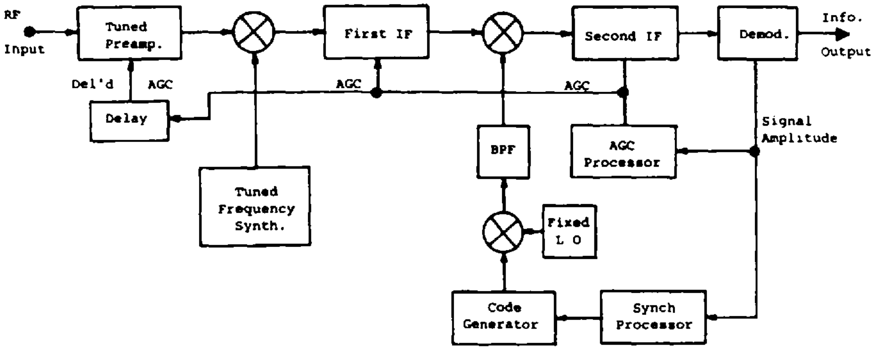


Figure A2.1 Block diagram of typical direct sequence spread spectrum receiver.

bandpass characteristics (bandwidth, rolloff, passband ripple, phase shift, etc.) of the first IF and the bandpass filter in the local reference are well matched, which helps to reduce losses in the synchronization process in the receiver.

The synchronization process in the receiver depends on a high degree of correlation between the first IF output signal and the local reference. The output signal from the second mixer is the recovered, information-modulated, despread signal. That is, it is no longer a spread spectrum signal, and if the correlation between the IF signal and the local reference has been properly carried out, it possesses latent process gain that can be used to advantage by the receiver.

3. The second mixer's output signal bandwidth, when its input signals are synchronized, has a bandwidth that is a function of the information bandwidth, and not the code rate, because the code is no longer present. Therefore, the second IF filters its input signal to the two-sided, information-modulated signal bandwidth. It is here that the process gain is realized, due to the trading of code-modulated RF bandwidth and information-modulated second IF bandwidth, and the process gain that is available is

$$10 \log \frac{\text{RF bandwidth}}{\text{data rate}} \quad \text{(in dB)}$$

which, in receivers such as those used in the GPS system (see chapter 14) can be as high as 50 dB.

4. Demodulation in a direct sequence receiver is usually performed by some form of double-sideband, suppressed carrier demodulator, such as a Costas loop or squaring loop. It is unusual for a direct sequence receiver's demodulator to encounter more than biphase or quadriphase modulated signals.

5. AGC in a direct sequence receiver is common but is often very difficult to implement, because such receivers must operate under widely varying interference conditions. This usually means that the AGC loop must have independent controls when desired signals dominate or when interference dominates the input. An offset is included in the AGC loop control to the preamplifier to accomplish "delayed" AGC and avoid desensitizing the receiver at small signal levels.

Direct sequence receivers that are intended to work in interference-laden environments must avoid limiting, on both desired signals and interference. Otherwise, any interference that is present can cause the signal to be suppressed, further enhancing the interference. Therefore, a direct sequence receiver has no choice but to either possess the ability to handle extremely large interfering or signal-plus-interfering levels linearly or provide automatic gain control that prevents the receiver from limiting.

6. Synchronization in a direct sequence receiver involves synchronizing the receiver's code generator with the incoming code and then tracking the incoming code rate as it varies due to frequency inaccuracy or Doppler shifts. This

is the task of the synch processor: to acquire code synchronization and maintain it as long as a link is required.

In many direct sequence receivers, carrier synchronization is also needed, in addition to code synchronization. This is accomplished by the demodulator in most receivers and is independent of the code synchronization process. In most cases, code synchronization must occur before carrier synchronization is possible, however.

A2.2 FREQUENCY HOPPING RECEIVERS

A typical frequency hopping receiver is shown in block diagram form in Figure A2.2. Frequency hopping receivers have some significant advantages over direct sequence receivers in that they may be simpler in some cases. Frequency hopping receivers can, under the right circumstances, work in a system without having automatic gain control. This is because limiting in a frequency hopping receiver does not have the same effect on a system's performance in the presence of interference as does limiting in a direct sequence receiver; limiting in the frequency hopping receiver does cause the desired signal to be suppressed, but the effect on the receiver's performance is not the same. In a direct sequence receiver, suppression of a desired signal's level in the presence of interference causes a direct loss in the receiver's output signal-to-noise ratio, but although a frequency hopping signal is suppressed the same amount by an interferer, the bit error rate is not increased by the signal being weaker in the frequency hopping receiver (unless the suppression causes the desired signal to drop below the receiver's threshold). That is, once a particular channel in a frequency hopping receiver is interfered with, adding more interference (or reducing the signal level in that channel) does

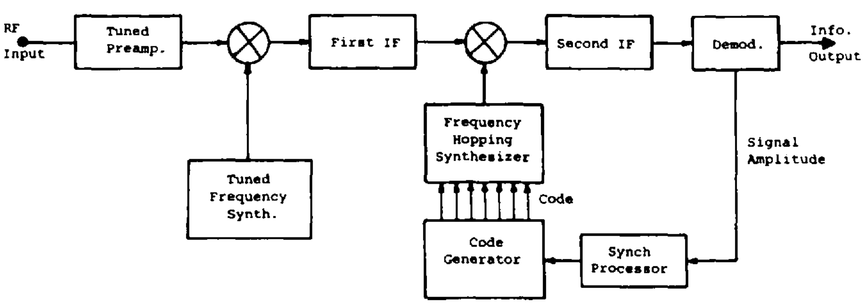


Figure A2.2 Block diagram of typical frequency hopping spread spectrum receiver.

not make the performance of that channel any worse, and that is the reason why limiting in a frequency hopping receiver is not catastrophic to its performance.

Following the signal through the frequency hopping receiver:

1. The input RF signal is filtered to reject unwanted signals. It is advantageous to employ a frequency hopping filter at the RF input of a frequency hopping receiver, but such a filter may not be practical, so it may be necessary to use a filter that simply passes the band of interest. As in the direct sequence receiver, the signal is amplified and passed on to the first mixer.

2. The received frequency hopping signal is converted to the first IF frequency by mixing the signal with a local oscillator signal whose frequency is at the center of the frequency hopping range, minus the first IF frequency (the LO does not hop).

At the first IF frequency, the signal hops with a pattern identical to that at the RF input, but shifted in frequency. Low-side injection is employed, because high-side injection would result in the frequency hopping pattern being inverted.

The signal at the first IF frequency is then applied to the second mixer to be dehopped. This process is similar to the despreading process employed in direct sequence receivers.

3. The local reference signal employed in a frequency hopping receiver is a frequency-hopped signal whose hopping pattern is identical to the first IF hopping pattern (which is identical to the RF input hopping pattern) except that the local reference signal is offset by an amount equal to the second IF frequency. That is, frequency one in the local reference pattern is offset from frequency one in the first IF pattern by exactly the same amount as every other corresponding pair of frequencies. Therefore, when the received pattern matches the local reference pattern, the output of the second mixer is a fixed frequency, with information impressed on it by the transmitter's modulation process.

The fixed output frequency of the second mixer (when the received and local reference codes are synchronized) is the second IF frequency, which is filtered, amplified, and input to the demodulator. As in the direct sequence receiver, the bandwidth of the second IF is designed to pass the information-modulated second IF carrier.

4. Modulation employed in most frequency hopping systems is frequency shift key, at the data rate plus whatever overhead is required to meet the system's bit error rate requirements. (It should be realized that frequency hopping systems are not viable without built-in error correction coding.) Therefore, the demodulator in a frequency hopping receiver is typically a noncoherent FSK demodulator, of which there are many from which one may choose. These demodulators do not require carrier synchronization.

5. Although the demodulator does not require carrier synchronization, code synchronization is required by the receiver. Once the receiver's code is synchronized with the received signal, it may or may not be necessary to track the

rate of the received code in the receiver. This is a function of the chip rate of the codes and the stability of the clocks used in generating code. If, for example, the hopping rate of a system were 100 hops per second and two clocks were accurate to one part in 10^6, then they would drift apart by a maximum of $2 \times 10^{-6} \times 100 = 0.0002$ code chips per second (once they were synchronized), so a clock tracking loop would not be necessary in many cases.

Some frequency hopping systems are designed to operate without any code rate tracking mechanism at all in the receiver. Either a high-stability clock or a GPS receiver for timing makes clock tracking unnecessary in a frequency hopping receiver. (Code synch acquisition is necessary, however.)

6. It should be noted that the receiver's code generator is shown with its code driving the receiver's frequency synthesizer in parallel, which is preferable to serial drive. With this (parallel) approach, the code clock rate is equal to the receiver hopping rate, no bus connection is necessary between the code generator and the frequency synthesizer, and latches are not required to hold the synthesizer's frequency for the duration of a hop time.

The synch processor in the case of the frequency hopping receiver has the task of telling the code generator to search when the receiver is not synchronized and to stop searching when synchronization has occurred. A short fine-synch process would be initiated to optimize synchronization once it has been achieved. It is common for both direct sequence and frequency hopping receivers to maintain synchronization to an accuracy of better than one tenth of a code chip time.

A2.3 CHIRP RECEIVERS

Chirp receivers have found their main application in the radar area, although they have been proposed for use in data transmission, and the basic chirp technique has been employed in "microscan" receivers for surveillance applications. Figure A2.3 shows a chirp receiver configured for data transmission. In this particular architecture, the signal is swept over the same range of frequencies, whether a one or a zero is intended, but the direction of the sweep (up or down in frequency) defines the detected output of the receiver.

A pair of frequency-dispersive filters are employed in this receiver, one of which is matched to a frequency sweep that goes from frequency f_1 to frequency f_2 in time T, while the other filter is matched to a frequency sweep that goes from frequency f_2 to frequency f_1 in the same amount of time (only one of the two is transmitted in a bit time). The filter that is matched to the particular sweep being sent puts out a pulse whose period is approximately $2/(f_1 - f_2) = t$, where the amplitude of the pulse is increased by a processing gain that is equal to the ratio T/t. (This ratio is also called the dispersion factor D.)

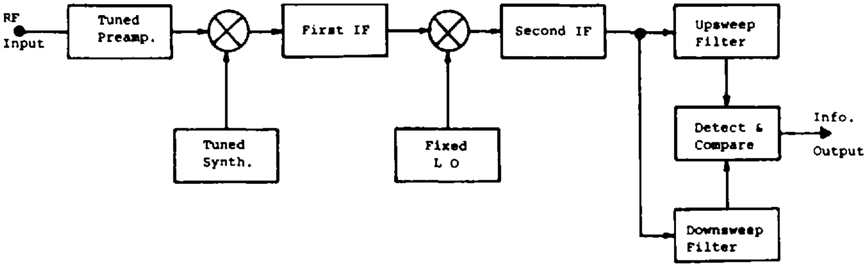


Figure A2.3 Block diagram of typical chirp spread spectrum receiver.

The decision as to whether a one or a zero is received is made on the basis of which of the filters (upsweep or downsweep) puts out the largest pulse.

Referring to Figure A2.3:

1. A received, swept RF input pulse is filtered and amplified in the receiver's preamplifier, whose bandwidth is designed to pass the entire swept bandwidth, $f_2 - f_1$.

2. The amplified and filtered signal is then mixed with a local oscillator signal that is tuned to the center of the swept band, minus the first IF frequency. Low-side injection is employed to prevent inversion of the frequency sweep. The mixer's desired output is at the first IF frequency.

3. At the first IF frequency, the signal is amplified and bandpass filtered again and applied to the second mixer for another down-conversion.

4. Unlike the situation with the direct sequence and frequency hopping receivers, the second conversion is not used to despread the desired signal. Here, the primary reason for a second conversion is to provide for the gain required and to improve the bandpass filtering applied to the signal. Adding a second conversion also permits more possible freedom in the design of the dispersive (chirp) filters necessary for making the one/zero decision in the receiver.

Despreading is performed by the dispersive filters themselves, in the process of correlating the energy across the swept frequency range. That is, the signal that represents a data bit is swept across the frequency range f_1 to f_2 or f_2 to f_1 and the filter that corresponds to the sweep that is sent assembles the energy sent over the bit time in such a way that it is compressed into a much narrower output pulse. The receiver then compares the output of both filters and chooses the output that is largest at the end of each bit time.

Separate synchronization is not required in a chirp receiver, although a bit synchronizer is usually used to recover the data clock from the data stream itself.

5. Automatic gain control is not required for typical chirp receiver applications.

A2.4 SYNCHRONIZATION IN SPREAD SPECTRUM RECEIVERS

Code synchronization is the most stringent requirement that must be met in a spread spectrum receiver (with the exception of chirp receivers, where spreading codes are not used). Synchronization of the spreading code is necessary to despread the received signal, make any processing gain available to the receiver, and recover data that are embedded in the spread signal. The synchronization problem is broken into two parts:

1. Code synch acquisition, where the receiver must bring its local code within one chip time of the received code.
2. Code tracking, where the receiver's code clock must be adjusted so that the transitions in the received code and the local code match one another and the receiver maintains the clock rate to track the incoming code.

Code synch acquisition is most often performed through a search process in which the receiver offsets its local code rate, so that the local code slips in time with respect to the received code rate. This allows the receiver to examine every possible time shift of the two codes with respect to one another, and the receiver must then only recognize the point at which they are in time synch with one another. The receiver then switches to the tracking mode.

Reduction of synch acquisition time is often aided by using codes for acquisition that are short enough that their entire length can be examined in a short (or at least reasonable) amount of time. GPS, for example, employs a synchronization code (the C/A code) that is 1023 chips long, and the P code has a period of 266 days at 10.23 million chips per second. Even so, the time for a GPS receiver to search chip by chip through the 1023-chip code is approximately 21 seconds, because the receiver bandwidth is optimized for a 50-bit-per-second data rate.

As a rule, spread spectrum receivers can search to acquire synchronization at a rate that is approximately equal to the data rate that the receiver is designed to demodulate. The reason for this is that the receiver's passband is optimized for the data bandwidth, and this pass bandwidth limits the rise time of the signal presented to the synch recognition circuitry in the receiver. Thus a GPS receiver with a 50-bps data rate can search through a 1023-chip C/A code in approximately 1023/50 = 20.46 seconds.

Faster synchronization methods are available, which of course require more complexity in the receiver (for example, multiple code correlators, offset in time). Nevertheless, code synch acquisition is still the most exacting and time-consuming process in a spread spectrum receiver.

After code synch acquisition, spread spectrum receivers usually employ some form of "delay lock" device that compares the received code with two replicas of the code that are offset in time from one another by an amount less than or equal to one chip time. This provides a pair of correlation outputs that, when compared, may be used to shift the receiver's code rate until the correlation at both is the same. Figure A2.4 shows a typical block diagram of a delay lock loop, as used in a spread spectrum receiver. A sampled form of delay lock loop, called a "tau-dither" loop, is also sometimes used.

As shown in Figure A2.4, there are three IF amplifiers, all at the same operating frequency; two of them (early and late) are used to develop the tracking signal. The third is necessary because neither of the other two tracks at the peak of code synchronization, which is halfway between the early and late codes. It should also be noted that the early and late channels employ noncoherent detectors (envelope detectors), whereas the on-time channel typically employs a coherent demodulator such as a Costas or squaring loop.

It is also practical to build spread spectrum receivers that do not employ tracking loops. One example already mentioned is frequency hopping receivers wherein stable code clocks and low hop rates make tracking loops unnecessary. Receivers in which coherence is not maintained from one bit time to the next, or in which such coherence is not vital to the demodulation process, may also be

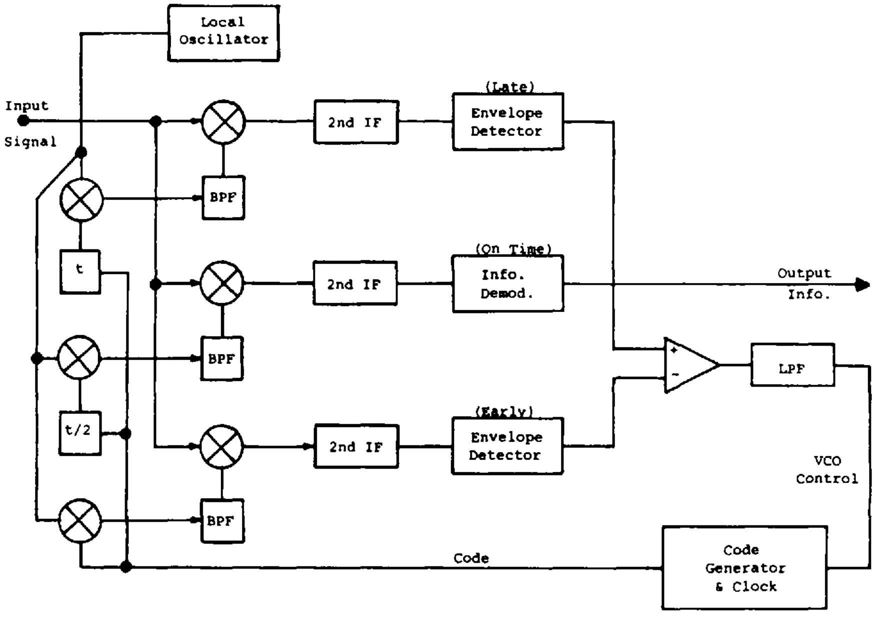


Figure A2.4 Delay lock loop block diagram for spread spectrum applications.

constructed without tracking loops. An example of such a receiver is a chirp receiver, in which coherence during a bit time is needed, but bit-to-bit coherence is not. In such receivers, there is no integration longer than one bit interval.

In all cases, the codes used for operation in spread spectrum systems must have adequate autocorrelation and cross-correlation performance to permit the receiver to unambiguously synchronize and maintain synchronization (where required). It should always be remembered that the received code must be synchronized to within one chip interval of the internal code before adequate correlation is achieved.

REFERENCES

1. Dixon, R. C., *Spread Spectrum Systems, with Commercial Applications*, John Wiley & Sons, New York, 1994.
2. Spilker, J. J., Jr., and Magill, D. T., The delay-lock discriminator—an optimum tracking device, *Proc. IRE*, Sept. 1961.

Appendix 3
Gray Codes

In systems that employ four or more symbols to convey information, it is advantageous to employ Gray coding to reduce the average bit error rate. Gray codes have the advantage that every symbol has a distance of one with respect to the symbols that are adjacent to it. This means that a 16-ary PSK system sending four Gray-coded bits per phase shift, as shown in Table A3.1, would have an advantage over a binary-coded approach that is significant, assuming that the system is working in a additive white Gaussian noise interference environment.

Observing Table A3.1, we see that the 4-bit Gray-coded symbols exhibit a difference of one for every symbol, with respect to its neighbors, while the binary-coded symbols have differences as great as four with their neighbors. This means that when Gray coding is employed, the most likely error, when any symbol is chosen, will produce only one bit error. Binary coding, on the other hand, will produce as many as four errors under the same conditions, using 16 four-bit symbols.

When an error is made, using 4-bit Gray coding, the receiver will see a single bit error, but the average number of bit errors made per symbol error with 4-bit binary coding would be 1.875, with a worst case of four.

Gray codes are derived from the binary codes, using two rules:

1. The most significant digit (the leftmost digit) is the same for both Gray and binary codes.
2. Counting from the right, $b_n \oplus b_{n+1} \oplus g_n = 0$. That is, if binary digits n and $n + 1$ are the same, then Gray digit n is a zero; but if binary digits n and $n + 1$ are different, then Gray digit n is a one.

For example, if we start with binary word 101101, then the Gray word that corresponds to it would be (using the $b_n \oplus b_{n+1} \oplus g_n = 0$ rule)

Table A3.1 Gray Coding Example for 16-ary PSK

Symbol number	Phase	Binary code	Gray code
1	11.25°	0000	0000
2	33.75°	0001	0001
3	56.25°	0010	0011
4	78.75°	0011	0010
5	101.25°	0100	0110
6	123.75°	0101	0111
7	146.25°	0110	0101
8	168.75°	0111	0100
9	191.25°	1000	1100
10	213.75°	1001	1101
11	236.25°	1010	1111
12	258.75°	1011	1110
13	281.25°	1100	1010
14	303.75°	1101	1011
15	326.25°	1110	1001
16	348.75°	1111	1000
1[a]	11.25°	0000	0000

[a]The phase map and codes are cyclic.

$$
\begin{array}{cccccccl}
1 & \multicolumn{7}{l}{\text{(the leftmost)}} \\
 & & & & & 1 & & (1 \oplus 1 \oplus 1 = 0) \\
 & & & & 1 & & & (1 \oplus 0 \oplus 1 = 0) \\
 & & & 0 & & & & (1 \oplus 1 \oplus 0 = 0) \\
 & & 1 & & & & & (0 \oplus 1 \oplus 1 = 0) \\
 & 1 & & & & & & (1 \oplus 0 \oplus 1 = 0) \\
\hline
1 & 1 & 1 & 0 & 1 & 1 & &
\end{array}
$$

The rightmost bits in the binary word are 01, so the rightmost bit of the Gray word must be 1, since $0 \oplus 1 \oplus 1 = 0$. Then proceeding across from right to left, modulo-two adding two binary bits in the manner shown produces the Gray-coded word in the bottom row.

Table A3.2 is a listing of Gray codes of length 3 through 6 bits, compared with their binary-coded equivalents. The 6-bit word table is not complete, however, since it goes through only 33 words. The remainder of the 6-bit table is one of those exercises that are left for the reader.

Table A3.2 Listing of Gray Codes Through Six Bits

3 bits (8-ary)	
Binary	Gray
000	000
001	001
010	011
011	010
100	110
101	111
110	101
111	110

4 bits (16-ary)	
Binary	Gray
0000	0000
0001	0001
0010	0011
0011	0010
0100	0110
0101	0111
0110	0101
0111	0100
1000	1100
1001	1101
1010	1111
1011	1110
1100	1010
1101	1011
1110	1001
1111	1000

5 bits (32-ary)	
Binary	Gray
00000	00000
00001	00001
00010	00011
00011	00010
00100	00110
00101	00111
00110	00101
00111	00100
01000	01100
01001	01101
01010	01111
01011	01110
01100	01010
01101	01011
01110	01001
01111	01000
10000	11000
10001	11001
10010	11011
10011	11010
10100	11110
10101	11111
10110	11101
10111	11100
11000	10100
10001	10101
11010	10111
11011	10110
11100	10010
11101	10011
11110	10001
11111	10000

6 bits (64-ary)		
Binary	Gray	
000000	000000	
000001	000001	
000010	000011	
000011	000010	
000100	000110	
000101	000111	
000110	000101	
000111	000100	
001000	001100	
001001	001101	
001010	001111	
001011	001110	
001100	001010	
001101	001011	
001110	001001	
001111	001000	
010000	011000	
010001	011001	
010010	011011	
010011	011010	
010100	011110	
010101	011111	
010110	011101	
010111	011100	
011000	010100	
011001	010101	
011010	010111	
011011	010110	
011100	010010	
011101	010011	
011110	010001	
011111	010000	
100000	110000	etc.

Appendix 4

Answers to Questions

CHAPTER 1

1. Wavelength $(\lambda) = \dfrac{300}{900} = 0.333$ m.

   Free space propagation loss $= 20 \log \left(\dfrac{4\pi R}{\lambda}\right)$

   $$= 20 \log \left(\frac{4 \times \pi \times 10{,}000}{\lambda}\right) = 111.54 \text{ dB}$$

2. Sensitivity, selectivity, adjacent channel rejection, output S/N ratio, and/or bit error rate.
3. UHF (although it does not cover the complete UHF band) or P band.
4. Wavelength = 300/824 = 0.364 m = 36.4 cm. 36.4 cm/2.54 = 14.33 in.
5. Transmit power, transmit antenna gain, propagation loss, receive antenna gain, and receiver sensitivity.
6. An impedance match consists of the process of ensuring that the source impedance of a device matches the impedance of the device that it is driving. (Example: A device such as an amplifier with 75-ohm output impedance should be connected to another device that has a 75-ohm input impedance.) Impedance matching is important because matching allows maximum power transfer between devices.
7. The loss (free space) at 99 miles and 2.3 GHz is

   $$36.6 + 20 \log 99 + 20 \log 2300 = 143.7 \text{ dB}$$

   No, it cannot be expected to transmit 99 miles without increasing the link budget (more transmitted power, more antenna gain, etc.).

8. Reduce the noise figure, reduce the receiver noise bandwidth, or improve the receiver's demodulator (reduce the signal-to-noise ratio the demodulator requires or increase its process gain, if it has such).
9. A receiver's selectivity is determined by its IF filter(s) and sometimes by a coherent demodulator together with the IF filter(s). RF filters do not usually contribute to a receiver's selectivity.
10. Effective input noise is KTB, which is

$$(1.38 \times 10^{-23}\ \text{W/Hz/K}^\circ) \times 290^\circ \times B_n\ \text{Hz}$$
$$= 1.38 \times 10^{-23} \times 290 \times 100{,}000 = 4.002 \times 10^{-16}\ \text{W}$$

In dBm, this is

$$-174 + 10 \log 100{,}000 = -124\ \text{dBm}$$

The receiver's sensitivity would be

$$-124\ \text{dBm} + \text{noise figure} + S/N\ \text{ratio required}$$

11. For receivers of analog information, output signal-to-noise ratio is specified at 10 to 12 dB, minimum. For receivers of digital information, bit error rate is specified at 1×10^{-3} to 1×10^{-6}.

CHAPTER 2

1. All modulation is analog modulation if it is broadcast, but digital information is conveyed by shifting a carrier in discrete increments. That is, the carrier and any modulation products are analog signals, but they are shifted from one frequency, phase, or amplitude to another in a discrete amount. For example, f_1 can be sent as a one and f_2 as a zero.
2. NCFSK is a modulation format in which data are transmitted as a series of frequency segments, in which one frequency is sent to signify a one, a second frequency to signify a zero, and no effort is made to control the phase relationship between the frequencies employed. This leads to phase discontinuities at the frequency shifts and increases the amount of bandwidth required to send a given data rate.

   CPFSK is a technique in which there is no discontinuity when shifting from one frequency to another, as the rate of change of phase is limited.

   NCFSK is noncoherent frequency shift key. CPFSK is continuous phase frequency shift key.
3. The deviation ratio used in "classic" MSK (minimum shift key) is 0.5, where deviation ratio for "digital" modulation is defined as

$$\frac{\text{frequency shift}}{\text{Data rate}} = \frac{\Delta f(\text{pk} - \text{pk})}{R_{\text{data}}}$$

Note that this definition is as opposed to analog modulation forms, where deviation ratio is defined as

$$\frac{\text{Frequency shift}}{\text{Modulating frequency}} = \frac{\Delta f(\text{pk})}{f_\text{m}}$$

4. For AM and DSBSC AM, the bandwidth required is two times the information bandwidth. SSB AM requires half the bandwidth of AM or DSBSC AM.
5. From a table of Bessel functions (section 2.3) the level is -0.3.
6. The FM gain is a function of the deviation ratio squared. Therefore, if the deviation ratio is increased from 1^2 to 3^2, the improvement should be nine times.
7. The null-to-null bandwidth of an on-off keyed signal is $2 \times R_\text{data}$. The null-to-null bandwidth of a BPSK-modulated signal is also $2 \times R_\text{data}$.
8. BPSK-modulated signals have 100 percent incidental AM. QPSK signals also have 100 percent AM but on only part of their phase shifts. The remainder of the QPSK phase shifts have 29.3 percent incidental AM. OQPSK-modulated signals have 29.3 percent AM at all phase shifts.

CHAPTER 3

1. A TRF (tuned radio frequency) receiver may amplify its desired signal and demodulate it, but there is no frequency conversion.
2. The primary difficulty with TRF receivers is that they have poor tuning accuracy, due to the fact that when multiple tuned stages are employed, each must be separately tuned.

   At higher frequencies, selectivity is poor, due to the limited Q available.

   For applications in which high gain is required, a TRF receiver is difficult to stabilize.
3. Stability is the primary problem with regenerative receivers. High gain is difficult to achieve.
4. Frequency synthesizers provide highly accurate superheterodyne receiver tuning.

   Superheterodyne receiver selectivity is provided by one or more fixed-frequency, highly selective filters.

   High gain is practical in superheterodyne receivers, since the gain can take place at two or more frequencies, and stability is easier to achieve.
5. Yes. Because a compressive receiver is more readily adaptable to superheterodyne architecture, and the compressive receiver has processing gain that is not available to the TRF receiver. A TRF receiver can receive AM or FM, however, whereas the compressive receiver cannot.
6. A receiver's noise bandwidth is the bandwidth over which the receiver's

response is equivalent to a filter with a rectangular response (that is, a filter with a 1:1 shape factor). For practical receivers that have filters with multipole response (crystal, ceramic, SAW, etc. filters) the 3-dB filter bandwidth is acceptable as the receiver's noise bandwidth.

The noise bandwidth determines the effective receiver input noise, KTB, and this is the noise floor upon which the receiver's performance is based.

7. The sensitivity of a receiver is the modulated input signal level that is required for the receiver to output a minimum signal-to-noise ratio (or bit error rate) when the input signal is modulated by the correct modulation at the design level specified for the receiver. For voice receivers, "standard" modulation is 1 kHz, at 30 percent, and the receiver must output a minimum of 10 dB *S/N* ratio.
8. Selectivity is the ability to reject signals that would otherwise be demodulated by the receiver but are at a frequency that is offset from the frequency to which the receiver is tuned. (Adjacent channel rejection is a special case of selectivity for a receiver, where a signal is input to the receiver at the closest undesired frequency and the sensitivity of the receiver is measured.) Selectivity is determined by the receiver's IF filter response.
9. The superheterodyne receiver is by far the most used receiver configuration. The reasons for this are:
   a. Wide frequency coverage, with similar performance at all frequencies.
   b. Highest sensitivity
   c. Best selectivity
   d. Stablest and most accurate

CHAPTER 4

1. The spurious responses that are created by the frequency conversion process(es).
2. Yes. F_{RF}/N and $F_1 \pm 2F_2$.
3. Because the amplitude of higher order responses is lower.
4. Passive mixer disadvantages:
   a. Cost in large quantities is relatively high, compared with active mixers.
   b. Local oscillator drive is relatively high.

   Passive mixer advantages:
   a. Good port-to-port isolation.
   b. Low development cost.
5. The primary purpose of automatic gain control (AGC) is to level receiver performance over a wide range of received signal input. Secondary reasons are to relieve the operator from the need to adjust the receiver under variable

signal conditions and to maintain the receiver's demodulator input level to as constant an amplitude as possible.

6. Improved:
   Selectivity
   Stability and accuracy
   Sensitivity
7. A superheterodyne receiver without a preamplifier would be most likely to suffer problems with:
   Sensitivity
   Image rejection
   LO radiation

   A superheterodyne receiver without a preamplifier would do best in the lower frequency bands (such as the 2- to 30-MHz "HF" band) because the noise in this band is such that the receiver noise contribution is low compared with the atmospheric noise that is present. On the other hand, desired signals that are present do not need amplification prior to the first mixer. (In fact, it is difficult to design an amplifier that is capable of handling the signal levels that are often present without becoming nonlinear.)
8. Multiple conversion is advantageous for the following reasons:
   a. To provide high gain
   b. To allow a high first IF frequency to aid in image rejection
   c. To allow a low final IF frequency to aid in demodulator design
   d. To improve selectivity

   Disadvantages of multiple conversion are
   a. More than one LO is required.
   b. New image frequencies are generated for each IF that is added.
   c. Receiver cost may be higher.
9. Delayed AGC is used for the purpose of preventing the AGC loop from desensitizing the receiver at input RF levels that are slightly larger than the smallest operating level but not large enough that a receiver preamp gain reduction can be implemented without the receiver's overall noise figure being degraded. Delayed AGC is typically set to begin operation at approximately a 50- to 100-μV RF input amplitude.
10. Contributors to the sensitivity of a superheterodyne receiver are
    a. Receiver noise figure
    b. Receiver gain
    c. Receiver noise bandwidth

    Noise figure is important, because it contributes to the noise seen by the demodulator, and a low noise figure aids in improving the demodulator's performance.

    Sufficient receiver gain is necessary for the demodulator input IF signal to reach the design level required to operate properly. Receiver noise bandwidth determines the effective receiver input noise and, together with the noise figure, the demodulator input noise level.

CHAPTER 5

1. Amplifier voltage gain is 15 V/0.1 V = 150.
   Input power is 0.1^2/50 ohm = 0.2 mW.
   Output power is 15^2/50 ohm = 4.5 W.
   Power gain is therefore 4.5 W/0.2 mW = 22,500.
   (In dB, the voltage gain is 20 log 150 = 43.52 dB, and power gain is 10 log 22,500 = 43.52 dB.)
2. Input power is as in question 1.
   Output power is 15^2/300 ohm = 0.75 W.
   Voltage gain is the same as in question 1.
   Power gain is 0.75 W/0.2 mW = 3750.
   (In dB, the voltage gain is still 20 log 150 = 43.52 dB, but power gain is 10 log 3750 = 35.74 dB.)

Note: These two examples show what happens when an amplifier's input and output impedances are not the same. Please note that if the output impedance were lower than the input impedance, the power gain would be *higher* than in the case of equal input and output impedances.

3. 4.5 W in question 1, and 0.75 W in question 2.
4. The difference between P1dB and IP3 for an amplifier is dependent on the amplifier, but typically the range of difference is 10 to 15 dB.
5. IIP3 is the IP3 (output intercept point) divided by the gain:

   $$\text{IIP3} = +15 \text{ dBm} - 21 \text{ dB} = -6 \text{ dBm}$$

6. P1dB is the output compression point, and IIP3 is the input third-order intercept.

   $$\text{P1dB} \cong \text{IIP3} + \text{gain} - 10 \text{ to } 15 \text{ dB}$$

7. Using the charts in section 5.2: −60 dB.
8. The amplifier will see a signal whose level is reduced by 1.2 dB. The receiver's sensitivity will be degraded by 1.2 dB.
9. Receiver noise figure will be approximately

$$\begin{aligned} F(\text{total}) &= 10\log\left(\text{NF}_1 + \frac{\text{NF}_2 - 1}{G_1} + \frac{\text{NF}_3 - 1}{G_1G_2} \cdots\right) \\ &= 10\log\left(1.58 + \frac{9}{63.1} + \frac{9}{252.4} \cdots\right) \\ &= 10\log 1.76 = 2.45 \text{ dB} \end{aligned}$$

Effective input noise = $-174 + 10 \log 100{,}000 = -124$ dBm

Noise input + noise figure = $-124 + 2.45 = -121.55$ dBm

Input signal-to-noise ratio at -107 dBm (1.0 μV)
= -121.55 dBm + 107 dBM = 14.55 dB

This *S/N* ratio is adequate for most types of modulation but does not provide sufficient margin if the receiver is to operate in a multipath environment. (Typically, 20 dB fading margin or more is provided, which would raise the input *S/N* ratio required to 30 dB, with 10-dB provision for *S/N* ratio during a 20-dB fade.)

10. At 10.7 MHz, an amplifier with 50 kHz bandwidth would have a Q of

$$\frac{10.7 \times 10^6}{50 \times 10^3} = 214$$

if the amplifier itself were tuned. A better approach would be to employ a separate amplifier and a crystal filter, although some tuning would still be useful in the amplifier itself.

11. The effect of adding a variable attenuator ahead of a preamplifier is that the receiver's sensitivity will be reduced by the amount of the attenuator's insertion loss. Adding a variable attenuator after the preamplifier serves the purpose of reducing the signal level to following stages but does not help the preamplifier when large signals are present.

12. A limiting IF amplifier serves the purpose of setting a receiver's demodulator input without having an AGC loop. The disadvantage of a limiting IF is that in the presence of interference, the desired signal may be suppressed. In broadcast FM receivers, where limiting IF amplifiers are standard practice, the largest signal present is the desired signal, and that allows any other smaller signal to be suppressed.

CHAPTER 6

1. A first mixer's noise figure affects a receiver's sensitivity to the extent that its noise factor (minus one)

$$\left[\text{Noise factor (NF)} = \log^{-1}\frac{\text{noise figure } (F)}{10}\right]$$

divided by the gain of the preamplifier does so. A second mixer affects the receiver's sensitivity even less, because its noise factor (minus one) is divided by the gain of the preamplifier, the first mixer, and the first IF amplifier.

2. A passive mixer typically requires higher local oscillator drive (typically

7.0 to 27 dBm). Passive mixers also tend to cost more in large-scale production than active mixers.

Active mixers tend to have lower dynamic range capability. Active mixers require power supply current. Port-to-port isolation is typically not as good as with passive mixers, which are usually double balanced.

3. If the signal applied to a mixer exceeds a level that is greater than approximately 7 dB below the local oscillator level, the mixer's transfer function is no longer linear.

   Note: This does not take into account any buffer amplifiers that may be used internal to an active mixer.
4. High-level mixers are mixers that are designed to have a linear transfer function when the signal applied is up to about 100 mW, although there is no concrete rule in this regard. The local oscillator for such a mixer must therefore be at level in the range of 500 mW or more.
5. Active mixers often have conversion gain, which relieves the need for IF gain somewhat in a receiver.

   Active mixers are typically very inexpensive in large quantity. Active mixers often employ buffer amplifiers internally, which may reduce the need for local oscillator power.
6. Low-side injection usually makes a tuned local oscillator more difficult to construct. Low-side injection does not invert the modulation sidebands on a signal being converted to another frequency (this may or may not be a disadvantage).
7. High-side injection is employed primarily because it reduces the percentage of bandwidth required of a tuned local oscillator.
8. The port-to-port isolation of the first mixer (specifically on the LO-to-input ports) directly affects the amount of local oscillator signal appearing at the receiver's antenna input.
9. A mixer's conversion gain, or loss, is the difference in the converted signal level between its RF input and its IF output. That is, the input desired signal and the output desired signal are at different frequencies, but conversion gain or loss is concerned only with their difference in amplitude.
10. Mismatch on one or more of the ports has the primary effect of increasing conversion loss. Mismatch, however, also implies that VSWR changes may significantly affect the devices that are attached to the mixer in other ways, even though they may not be attached directly to the port that is mismatched.

CHAPTER 7

1. The effect is to cause the receiver to respond to frequencies other than the desired signal frequency and to degrade the performance of the receiver at the desired frequency.

2. The advantages of employing a phase lock (indirect) synthesizer as a local oscillator are that where multiple frequencies are desired, the phase lock synthesizer is low in cost and provides multiple frequencies, all with the accuracy of the frequency reference used (a phase lock synthesizer can readily be referenced to an atomic standard).

   The primary disadvantage of a phase lock synthesizer is that it does not switch from frequency to frequency readily at a rate higher than a few hundred times per second.
3. Accuracy in a local oscillator defines the range within which the local oscillator signal can be expected to operate for all conditions, including temperature, power supply voltage, time since calibration, and G loading.

   Stability refers to the drift in frequency per unit time (usually per day) that is the maximum that can be expected after the oscillator is well aged and has been turned on for a period long enough to stabilize and reach its final aging rate.

   Such a specification means nothing, unless it also specifies the time period for which 1×10^{-6} is valid.
4. Direct synthesizers typically switch very rapidly from one frequency to another (as fast as 100 nsec switching time). They can also be designed to provide coherence between all of their output frequencies. Direct synthesizers tend to be complex and expensive, however. They also tend to consume a large amount of power. Indirect, or phase lock, synthesizers tend to be small, low in cost, and consume low power, but they are relatively slow in switching speed.

   Spurious output from both direct and indirect synthesizers can be held to the -80 dBc range.

   Both direct and indirect synthesizers are available in integrated circuit form, but direct synthesis ICs are not currently capable of operation to as high a frequency range as phase lock ICs.
5. Single-sideband phase noise is the noise output of an oscillator or frequency synthesizer in a 1-Hz bandwidth, at a given frequency offset from the desired output. The level of the power in the 1-Hz bandwidth is compared with the level of the desired output.

   Allen variance is a measure of the change in frequency of an oscillator in a given period, measured in a number of contiguous equal periods. A calculation of the variance of the frequency is made, based on the N consecutive measurements of frequency.
6. Digital phase detectors are employed in frequency synthesizers because:

   Their use decreases frequency-switching time.

   The divide-by-N counter used in a phase lock synthesizer has a digital output.

   The reference divider included in most ICs has a digital output.

   Frequency synthesizer ICs include digital phase detectors within themselves.

7. Choosing high-side injection inverts the modulation sidebands on the signal being converted but reduces the percentage of bandwidth required for the synthesizer to cover.

   There is no reason why both high- and low-side injection could not be used in a single receiver that converts more than once, as long as any other considerations are met.
8. The reference frequency in a phase lock synthesizer is chosen primarily to provide a particular amount of frequency shift per increment in the synthesizer's control input. (The synthesizer's output frequency is basically $N \times f_{ref}$, where N is the division ratio for the frequency divider.)

   The primary undesired output signals from a phase lock synthesizer are the reference-related sidebands and their harmonics. These are caused by frequency modulation of the VCO at the reference frequency, as it is transmitted through the loop filter.
9. Reduce the loop filter bandwidth. This reduces the sideband level by 20 dB per decade of reduction in the loop filter bandwidth.
10. Reduce the frequency shift required of the loop. This reduces lock-in time by a factor that is the cube of the frequency shift reduction.

    A second method is to increase the bandwidth of the loop filter, but this has the drawback that it increases the reference-related sideband level(s).

CHAPTER 8

1. In a coherent demodulator, a carrier that is used to demodulate the desired signal is derived from the signal itself. This carrier is a regenerated replica of the carrier found in the modulator before modulation occurred, and as such it has been regenerated as accurately as possible and with as little noise as possible. It is this carrier that is used to demodulate the desired signal. A noncoherent demodulator, on the other hand, typically demodulates a signal by a squaring process in which it multiplies the signal by itself, therefore multiplying the noise input from the channel along with the signal.

   Using an AM signal as an example, the squaring process produces

$$\left[A\cos(\omega_c + \phi_{n1})t + \frac{Am}{2}\cos(\omega_c + \omega_m + \phi_{n2})t \right.$$

$$\left. + \frac{Am}{2}\cos(\omega_c - \omega_m + \phi_{n3})t\right]^2$$

$$= \frac{A^2}{2}[\cos(\omega_c + \phi_{n1} + \omega_c + \phi_{n1})t + \cos(\omega_c + \phi_{n1} - \omega_c - \phi_{n1})t]$$

$$+ \frac{A^2 m}{4}[\cos(\omega_c + \phi_{n1} + \omega_c + \omega_m + \phi_{n2})t$$
$$+ \cos(\omega_c + \phi_{n1} - \omega_c - \omega_m - \phi_{n2})t]$$
$$+ \frac{A^2 m}{4}[\cos(\omega_c + \phi_{n1} + \omega_c - \omega_m + \phi_{n3})t$$
$$+ \cos(\omega_c + \phi_{n1} - \omega_c + \omega_m - \phi_{n3})t]$$
$$+ \frac{A^2 m}{4}[\cos(\omega_c + \omega_m + \phi_{n2} + \omega_c - \omega_m + \phi_{n3})t$$
$$+ \cos(\omega_c + \omega_m + \phi_{n2} - \omega_c + \omega_m - \phi_{n3})t]$$

which, after low-pass filtering, becomes

$$\frac{A^2}{2} + \frac{A^2 m}{4}[\cos(\omega_m + \phi_{n1} - \phi_{n2})t + \cos(\omega_m + \phi_{n1} - \phi_{n3})t$$
$$+ \cos(2\omega_m + \phi_{n2} - \phi_{n3})t]$$
$$= \frac{A^2}{2} + \frac{A^2 m}{4}[2\cos(\omega_m + 2\phi_{n1} - \phi_{n2} - \phi_{n3})t]$$
$$= \frac{A^2}{2} + \frac{A^2 m}{2}\cos(\omega_m + 4\phi_{\text{ntotal}})t$$

If we multiply the AM signal with a second carrier, rather than with itself, the answer is different. The demodulated output becomes (where the second carrier is $B \cos \omega_c t$, as recovered from the received signal)

$$\frac{ABm}{2}[\cos(\omega_c + \omega_c + \omega_m + \phi_{n2})t + \text{cus}(\omega_c - \omega_c - \omega_m - \phi_{n2})t]$$
$$+ \frac{ABm}{2}[\cos(\omega_c + \omega_c - \omega_m + \phi_{n3})t + \cos(\omega_c - \omega_c + \omega_m - \phi_{n3})t]$$
$$= \frac{ABm}{2}[\cos(\omega_m - \phi_{n2})t + \cos(\omega_m - \phi_{n3})t]$$
$$= A^2 \cos(\omega_m - 2\phi_{\text{ntotal}})t, \qquad \text{where } m = 1 \text{ and } A = B$$

and the signal output is twice the amplitude of the squared output with a signal-to-noise ratio twice that of the squared output.

This example is typical in showing the difference between coherent and noncoherent demodulation to be 3 dB, and this is due to the multiplication of a noisy signal by itself in the noncoherent case but multiplying the noisy signal with a clean carrier in the coherent case.

2. B_N is the noise bandwidth of a phase lock loop and is (from Hoffman) approximately 3.2 times the closed-loop 3-dB bandwidth, B_L. K is the open-loop gain and is the phase detector gain times the VCO gain times any amplification in the loop. Because phase detector gain is in volts per radian and VCO gain is in radians per volt, K is dimensionless.
3. The differences are that the phase lock demod typically employs an analog phase detector, whereas a phase lock synthesizer's phase detector is usually digital. Also, the synthesizer includes a variable divider between its VCO and phase detector, for frequency selection. Other differences are minor and are related to the frequency range and bandwidths that are to be covered.
4. An envelope detector designed for a 5.0-MHz carrier should have a time constant of not less than $10/(5 \times 10^6) = 2 \times 10^{-6}$ seconds to prevent its output from following the 5-MHz IF carrier but a time constant no greater than $1/[10(5 \times 10^3)] = 2 \times 10^{-5}$ seconds to allow the output to follow the modulation envelope. A good compromise would be halfway between the minimum and maximum values, or 11 microseconds.
5. The effect of overdamping would be to slow the transient response of the demodulator, including both the acquisition time and the response to a step function change in frequency.
6. The bandwidth required is approximately

$$B_L = \sqrt[3]{\frac{\Delta f^2}{t_{\text{lock}}}} = \frac{(1 \times 10^5)^2}{1 \times 10^{-2}} = 10 \text{ kHz}$$

7. Yes, a product detector can be employed for both single-sideband and double-sideband demodulation. In AM, a zero beat between the AM carrier and the product detector's reference carrier is difficult to avoid, however.
8. Any of the noncoherent FM demodulators would be acceptable. A coherent demodulator would be forced to reacquire the signal each time the modulation changed frequencies.
9. With a bit error rate of 1×10^{-3}, a PSK demodulator should theoretically operate at a 6.8-dB signal-to-noise ratio. Given that this Costas loop has a loss of 1.3 dB, the 1×10^{-3} bit error rate should be achieved at 8.1-dB demodulator input signal-to-noise ratio.

   At the input to the receiver, the signal-to-noise ratio required would be 8.1 dB, plus the receiver's noise figure, $8.1 + 2.5 = 10.6$ dB. Therefore, the receiver's sensitivity would be 10.6 dB above KTB, which is a function of the bandwidth required to pass the data-rate-related modulation, plus any frequency-uncertainty-related bandwidth allowance.
10. E_b/N_0 is the ratio of energy per bit received, compared with the noise density, at the demodulator.

    Energy per bit is the signal power, divided by the bit rate, and noise density is the noise power per hertz of bandwidth, also sometimes called KT.

CHAPTER 9

1. Receiver sensitivity measurements serve as the reference for a receiver's adjacent channel rejection measurement. First, a sensitivity measurement is made, with the signal source set to the receiver's desired input frequency. Then the frequency of either the receiver or the signal generator is changed to the next channel, and the measurement is repeated. Otherwise, there is no difference.
2. Image rejection, IF rejection, and adjacent channel rejection are the minimum set, but also important are IF/N tests, premixing tests, double and triple beat tests, and tests for birdies, as well as other tests.
3. The receiver's noise floor, or KTB, is $-174 + 10 \log 25{,}000 = -130$ dBm. Its sensitivity is 5.0 μV, or -93 dBm. The receiver's sensitivity is 37 dB above its noise floor. Receiver noise figure therefore is $37 - 10 = 27$ dB, minus any other possible losses. (This receiver evidently does not have a low-noise preamplifier.)
4. By counting the number of bits transmitted, comparing the bits received with the bits transmitted and counting the errors, and then dividing the number of errors by the number of bits transmitted.

   Note: It should be realized that multiple tests must be made to provide confidence in the bit error rate performance of a receiver. What is needed is to know the mean number of errors per given group length, where the number of groups is large.
5. With high-side injection, the image frequency for a receiver with a 455-kHz IF is RF + 910 kHz.

   With low-side injection, the image frequency for the same receiver is RF − 910 kHz.

   Since the broadcast AM radio band covers 560 to 1600 kHz, the image for 560 kHz is 560 + 910 = 1470 kHz; it falls inside the band (a higher IF frequency is needed). With the receiver tuned to the high end of the band, the image can fall at 1600 − 910 = 690 kHz.
6. Image response is measured in the same way as sensitivity, except that the signal generator is set to the image frequency for the receiver.
7. Not unless the receiver is to be operated with its AGC disabled.
8. A "hard" sensitivity measurement is made with a 6-dB attenuator inserted between the signal generator being used and the receiver being tested. "Soft" sensitivity measurements are made without the attenuator.
9. 1 GHz + 700 MHz = 1.7 GHz for high-side injection.
   1 GHz − 700 MHz = 300 MHz for low-side injection.
10. Signal-to-noise ratio is measured by alternately turning on and turning off the modulation of a signal input to a receiver and measuring the output of the receiver under both conditions. The difference between the two measure-

ments is the signal-to-noise ratio. (The measurement made is actually a signal-plus-noise-to-noise ratio measurement, since the noise is still present when the modulation is on.)

The difference between signal-to-noise ratio and signal-plus-noise-to-noise ratio is small once they reach a level of 10 dB or higher; therefore there is little concern about the difference.

E_b/N_0 measurements are not made directly. E_b/N_0 is normally calculated from other measurements or estimates.

CHAPTER 10

1. a. Horizontal antennas would need to be twice as long in many applications as vertical antennas.
   b. Mounting structures for horizontally polarized antennas would be difficult to implement and use (especially in mobile applications).
2. When the transmitting and receiving antennas are rotated 90 degrees, with respect to their polarization, up to 20-dB loss is typical. (This is compared to 0-dB loss when the two antennas are in the same orientation.)
3. If the receiving antenna is circularly polarized, a linearly polarized signal cannot be received without loss. However, the loss is only 3 dB over the loss that would occur if the two antennas were both linear and perfectly aligned, and the loss is never more than 3 dB, no matter what the orientation between the linear and circularly polarized antennas.
4. Directionality is the mechanism that produces antenna gain. If an antenna's power is reduced in one direction, it must be increased in another, given that power conservation holds. Therefore if an antenna transmits the same amount of power over 36 degrees instead of 360 degrees, it should have 10 log (360/36) = 10 dB gain, and if the same is true in both the vertical and horizontal planes, then the antenna will exhibit twice as much gain. Of course, this means the antenna must be pointed.
5. 10 dBi means that an antenna has 10 dB more gain than an "isotropic" antenna, which is an antenna that transmits equally in all directions, therefore has zero gain. (Such antennas do not really exist except as theoretical benchmarks.)

   10 dBd is the gain of an antenna compared with a dipole antenna's gain, which is theoretically 2.2 dB.

   Therefore 10 dBd is 2.2 dB greater than 10 dBi.
6. Antennas are designed to conform to their frequency of operation, as a function of the wavelength at that frequency. A typical antenna has physical dimensions that are in terms of wavelength, where commonly one quarter-wavelength and one half-wavelength are the most used for at least one of the antenna's dimensions.

7. 15 dB is considerably better, because 15-dB return loss means that only 1/31.6 of the power input to the antenna is being reflected to the source. On the other hand, 3-dB return loss means that half of power is being reflected back to the source from the antenna. The more power that is not returned to the source from the antenna, the better the antenna, unless the antenna itself is absorbing the power, in which case the antenna should get warm, because it can only radiate the input power as heat or electromagnetic energy.
8. Ground conductivity determines the efficiency with which an antenna can operate as a half-wavelength antenna, when its actual length is one quarter-wavelength. It also determines the strength and distance over which the ground wave will be effective.
9. E-plane (vertical) beamwidth would be 36 degrees. (See answer to question 4.)
10. If an antenna does not have enough bandwidth, its Q should be decreased. This implies that its capacity should be decreased (that is, it should appear more capacitive at its input). This is usually accomplished by making the antenna's length-to-width ratio smaller. (We make the antenna "fatter.")

CHAPTER 11

1. Twice the highest frequency in the band being sampled. 6.0 kHz.
2. 6.0 kHz.
3. Because the filters available do not roll off fast enough to prevent overlap of unwanted sampled signal components into the recovered information band.
4. $\text{Log}^{-1}(43/10) = 19{,}952$.
   $\text{Log}_2\ 19{,}552 \rightarrow 16$ bits, minimum.
5. A zero-IF receiver is one in which a down-conversion occurs in which the input signal is converted to a zero-hertz center frequency, which causes the upper and lower sidebands of the signal to fold over around zero, so that the lower sideband overlaps the upper sideband. If they are symmetrical and there is no frequency error in the conversion, then the upper and lower sidebands (which are coherent in many signals) add.

   If they are not symmetrical, then the problem is one of how to make use of them. (One solution is to filter one sideband away before down-conversion.)

   Sampling, on the other hand, does convert the signal to baseband (just as multiplication with a same-frequency local oscillator does) but the baseband signal exists as digital information, for which the task of processing may not be simpler but for which the processing can be carried out by a computer, and this may make the task tractable if not simpler.
6. To reduce the effects of aliasing.
7. To prevent aliasing due to the A/D converter seeing frequency components beyond those in the band of interest.

8. The number of levels to be resolved is $1.0/(1 \times 10^{-3}) = 1000$, which requires at least 10 bits per sample.

   At 100-kHz bandwidth, sampling at one third over the minimum, the sampling rate would be $2 \times 133{,}333 = 266{,}666$ samples per second. Therefore the serial data rate required would be $10 \times 266{,}666 = 2.66666$ Mbps.
9. Oversampling is any process in which the sampling rate is higher than that required by the Nyquist/Shannon theorem. Undersampling and subsampling are terms used to describe the process by which a modulated signal at a given carrier frequency is sampled at a rate that allows the information to be digitized and completely recovered, but the carrier is not. (The carrier is not wanted.)

   To recover the information that is modulated onto the carrier, one needs only to sample at twice the modulation rate, which is usually well below the rate necessary to recover the carrier. This means that, at the same time, one can undersample with respect to the carrier but oversample with respect to the modulation.
10. 266,666 10-bit bytes per second.

CHAPTER 12

1. $-90 + 5 = 85$ dB net gain.
2. Yes, because it is difficult to get this much gain (unless the package that contains it is large) with good stability.
3. Yes, but its cost may not be acceptable. It would also cause the net gain requirement to increase by 7 to 10 dB, because of its loss.
   The local oscillator's power might have to increase.
4. Yes, because the AM signal cannot be allowed to limit, and providing amplifiers with enough power-handling capability would not be acceptable. (With a -5 dBm signal and 95 dB gain, the final amplifier in the chain would output a 90-dBm signal, in a receiver having neither AGC nor limiting; 90 dBm is 1 million watts.) The control range should be at least 85 dB.
5. The noise figure is

$$10 \log\left(1.6 + 0 + \frac{3}{50} + \frac{9}{3.15 \times 10^5} \cdots\right) = 2.2 \text{ dB}$$

6. In the AM mode (assuming the receiver bandwidth is optimized) the noise input would be (including Mach 1 Doppler)

$$-174 + 10 \log 13{,}800 = -132.6 \text{ dBm}$$

   and in the BPSK mode

$$-174 + 10 \log 80{,}600 = -124.9 \text{ dBm}$$

Adding 10 dB for S/N ratio in the AM mode and 2.2-dB noise figure brings the AM mode sensitivity to

$$-132.6 + 12.2 = -120.4 \text{ dBm}$$

which provides $120.4 - 107 = 13.4$ dB margin in the AM mode. The BPSK demodulator, which requires an 8.3-dB S/N ratio (theoretical) with 1.7-dB degradation for implementation loss, would also work at 10 dB S/N in the receiver of section 12.4. This means that in the BPSK mode, with the same 2.2-dB noise figure, the receiver should have sensitivity of

$$-124.9 + 10.2 = -114.7 \text{ dBm}$$

which provides $114.7 - 107 = 7.7$ dB margin.

7. Because the entire FM band covers only 20 MHz (88 to 108 MHz) and the 10.7-MHz IF used is high enough that the image frequency falls ouside the band, yet the IF frequency is low enough that demodulators are easily built.
8. This is an example for which there is no single answer.
9. 5.0 μV is $-107 + 20 \log 5 = -93$ dBm.

   This input level would provide an S/N ratio at the demodulator of 21 dB. Yes, the receiver would operate at 5.0 μV, and the bit error rate would be extremely low except in fading environments.
10. See question 8.

CHAPTER 13

1.

$$\begin{array}{lr} (1) & 00101100 \\ (2) & \oplus\ \underline{10100110} \\ & 10001010 \end{array} \qquad \begin{array}{lr} (1) & 00101100 \\ (3) & \oplus\ \underline{11101101} \\ & 11000001 \end{array} \qquad \begin{array}{lr} (2) & 10100110 \\ (3) & \oplus\ \underline{11101101} \\ & 01001011 \end{array}$$

Pair	Distance
(1) (2)	3
(1) (3)	3
(2) (3)	4

2. The number of errors correctable is $(D - 1)/2 = 1$.
3. The Hamming weight of 110100100010101 is seven.
4. The Hamming encoding technique generates a specific symbol for any specific 4-bit input word (of which there are only 16 possible). Table 13.6 lists the 16 input words and the symbol that results from adding the Hamming-generated parity.

With input 0101, the symbol that would be generated is 0101101. To generate this symbol, the algorithm used is

$$X_1 \oplus X_2 \oplus X_3 = P_1$$
$$X_1 \oplus X_2 \oplus X_4 = P_2$$
$$X_1 \oplus X_3 \oplus X_4 = P_3$$

Where X_N designates a particular data bit in the input word and P_N designates a particular parity bit added to the input word. Specifically,

$$0 \oplus 1 \oplus 0 = 1$$
$$0 \oplus 1 \oplus 1 = 0$$
$$0 \oplus 0 \oplus 1 = 1$$

So we generate the transmitted symbol by adding 101 to the data word 0101, so that the completed symbol is 0101101.

5. Coding gain is the improvement (if any) seen in performance when comparing a system using an error correction code with the same system without the code. It is measured by operating a system and measuring its bit error rate without an error correction code. Then, after adding the error correction code, increase the noise (or decrease the signal) until the bit error rate is the same as before. The difference in the *S/N* ratio between the performance in the two conditions is the coding gain.
6. A (21, 6; 3) encoder is an encoder that outputs a 21-bit symbol for a 6-bit input data block. It would therefore be a rate 6/21 = 2/7 encoder. (The 3 signifies that 3 bits in the 21-bit symbol can be corrected.)
7. The encoded rate would be $2 \times 28{,}000 = 56$ kbps, and the BPSK modulated signal bandwidth would be 112 kHz null to null, or 49.2 kHz at the 3-dB points.
8. If the encoder were removed, the bandwidths would be reduced by half, or 56 kHz null to null and 24.6 kHz 3 dB.
9. For QPSK or OQPSK, the bandwidth would be half the bandwidth of a BPSK-modulated signal, so that with the rate one half code, the bandwidth would be the same as BPSK without the error correction code (56 kHz null to null and 24.6 kHz 3 dB).
10. The constraint length in a convolutional code is the length of the shift register used to encode the data. That is, constraint length seven means there are seven flip-flops in the shift register.

Appendix 5

Receiver Applications and Frequency Ranges

Wherever there is a need to send information, there is a need for a receiver. Today, receivers are designed to cover frequencies (or wavelengths, depending on one's point of view) from the very, very low (long wavelength) to the very, very high (extremely short wavelength). Some of these applications and their operating ranges (see Fig. A5.1) are:

Extremely low frequencies: Below 300 Hz, wavelength > 10^6 m. Very long distance, low data rate communications systems. Signals penetrate many solid and/or liquid objects such as large bodies of water. Primary problems are antenna size and bandwidth, and the bandwidth of the medium.*

Voice frequency: 300 to 3000 Hz, wavelength 10^6 to 10^5 m. Not assigned for radio applications.

Very low frequencies: 3000 Hz to 30 kHz, wavelength 10^5 to 10^4 m. Long-distance applications, similar to those for ELF frequencies, although penetration is not as good. Antennas are more practical, but signal bandwidth is still quite low.*

Low frequencies: 30 to 300 kHz, wavelength 10^4 to 10^3 m. Similar to VLF in application and bandwidth. Usually reserved for special applications such as navigation aids.

Medium frequencies: 300 kHz to 3 MHz, wavelength 10^3 to 10^2 m. A very sought-after frequency range. This range includes the AM broadcast band (535 kHz to 1.605 MHz) and the low end of what is usually called the HF radio band. Even though the HF band does not officially begin until 3

*Practical bandwidth for a signal in the earth's atmosphere is no more than 10 to 20 percent of the operating frequency. Also, for antennas that are shorter than approximately 1/6 wavelength, bandwidth tends to be very narrow.

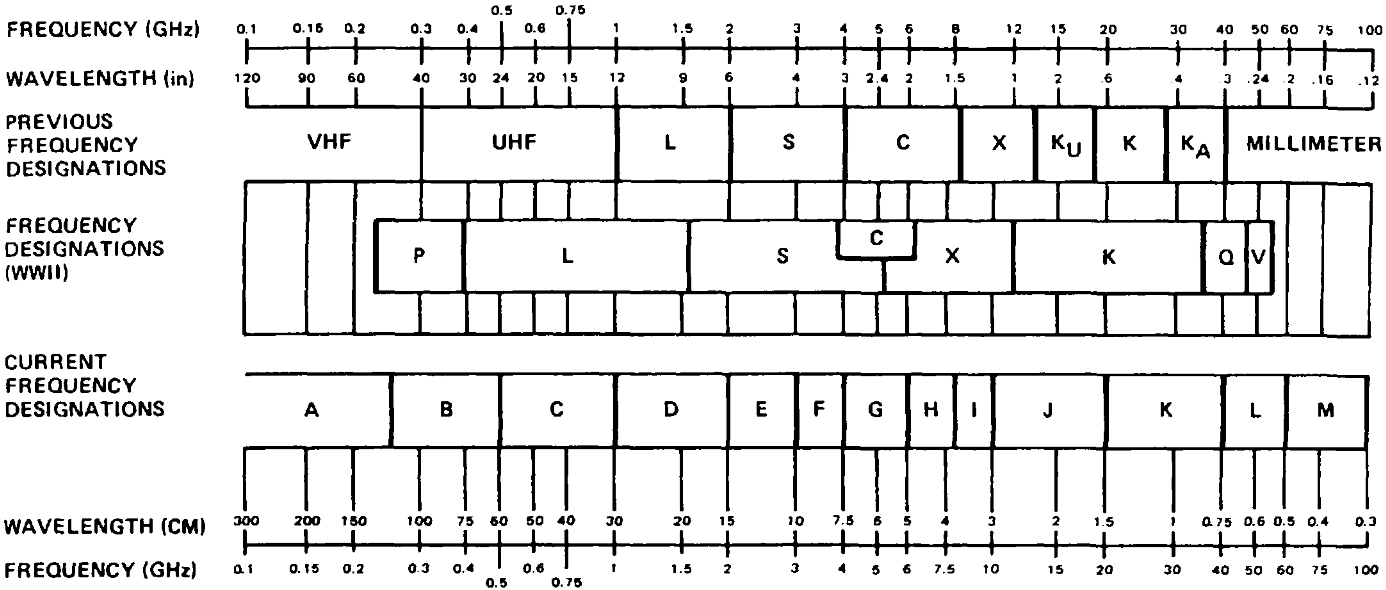


Figure A5.1 Frequency ranges by letter designation.

MHz, radios whose operating range begins at 2.0 MHz are usually defined as HF radios. This frequency range is the lowest at which voice-bandwidth signals are routinely sent. It is also routinely employed for long-distance Morse code transmissions.

High frequencies: 3.0 to 30 MHz, wavelength 10^2 to 10 m. This band of frequencies is among the most sought after of all, because it provides for long-distance, over-the-horizon communications and reasonably wide bandwidth (up to 3 MHz at the upper end of the band), and yet the antenna length for efficient transmission is reasonably short. That is, antennas in this band need be no longer than 2.5 to 25 m and are often shorter than 1/4 wavelength. The HF band is the lowest frequency range at which a mobile transmitter is practical for most applications, although mobile receiving antennas may be constructed at lower frequencies.

Very high frequencies: 30 to 300 MHz, wavelength 10 to 1.0 m. A part of this band, from 30 to 90 MHz, is often used for what are called VHF radios, and other radios called UHF radios employ the 225- to 300-MHz portion of the band. This frequency range is widely used for mobile applications, especially aircraft, including both military and civilian systems. For example, air traffic control signals for airports are found in this band, along with many other signals very important to everyday life, such as broadcast band television. This band, while not providing reliable propagation beyond the line of sight, is the lowest band that enables antennas that are small enough for many mobile applications, such as aircraft.

Ultrahigh frequencies: 300 to 3000 MHz, wavelength 1.0 to 0.1 m. The lowest end of this band, 300 to 400 MHz, is called the military UHF band (together with the 225- to 300-MHz portion of the VHF band). The UHF band is used for mobile applications of all kinds as well as many fixed point-to-point applications. It also contains the beginning of what is called the "microwave" frequency range, although "microwaves" are usually so loosely defined today that it is difficult to state that they begin at a particular frequency. Coincidentally, microwave ovens operate in this band of frequencies, at approximately 2440 MHz. Although propagation losses in this band are higher than in the lower frequency ranges, wavelength is such that relatively small high-gain antennas can readily be constructed, and their gain makes up for the higher propagation loss.

Superhigh frequencies: 3 to 30 GHz, wavelength 0.1 to 0.01 m. Employed primarily for fixed microwave applications,* because of the necessity for antenna pointing, due to a requirement for high antenna gain. (This is in turn necessary because of high propagation losses, together with the greater difficulty in generating RF power in this very high frequency band.)

*Including satellite systems.

Extremely high frequencies: 30 to 300 GHz, wavelength 0.01 to 0.001 m. Similar in application to the SHF band and including the beginning of what is termed the "millimeter wave"; range (in practice, lower frequencies are also sometimes included as millimeter waves). This band is currently the upper limit of the assigned frequency range. That is, the Federal Communications Commission does not currently regulate the usage of frequencies beyond 300 GHz.

Appendix 6
Wavelength (λ) and Frequency

The speed of light is 2.9979250×10^8 m/sec and a frequency of 299.7925 MHz gives exactly 1 m wavelength. Since wavelength is the distance that a signal sin $\omega_c t$ at a particular frequency travels before repeating,

$$\text{Wavelength} = VT = \frac{V}{f}$$

where T is the single cycle repetition period and V is the velocity of light. Therefore, wavelength λ in meters is

$$\lambda = \frac{2.99792500 \times 10^8}{\text{frequency in Hz}}$$

which is usually approximated by

$$\lambda = \frac{300}{\text{frequency in MHz}}$$

Table A6.1 lists the constant to be divided by frequency to obtain wavelength in units other than meters. The reciprocal of the velocity given in this table is, of course, the time required to travel the distance in question, which in itself is a useful quantity. For example, the time required to travel one statute mile at the speed of light is $1/(1.862824 \times 10^5) = 5.368$ μsec.

Table A6.1 Constants for Calculation of Wavelength ($\lambda = V/f$)

For λ in:	Velocity is[a]	Approximation[b]
Meters	2.997925×10^8 m/sec	300
Centimeters	2.997925×10^{10} cm/sec	30,000
Millimeters	2.997925×10^{11} mm/sec	300,000
Inches	1.180285×10^{10} in/sec	11,800
Feet	9.835712×10^8 ft/sec	984
Yards	3.278571×10^8 yd/sec	328
Miles (statute)	1.862824×10^5 mi/sec	0.186
Miles (nautical)	1.61875×10^5 nmi/sec	0.162

[a]Divide by frequency in Hz.
[b]Divide by frequency in MHz.

Appendix 7
Noise in a Receiver's Bandwidth

Receivers must contend with several different types of noise, each of which can degrade the receiver's performance. The noise type that is most effective in degrading a particular receiver's performance is a function of the receiver's bandwidth, the modulation used for the desired signal being received, and the signal frequency. The four types of noise of most concern to receivers are atmospheric noise, galactic noise, man-made noise, and KTB noise.

A7.1 ATMOSPHERIC NOISE

This noise is generated primarily by electrical storms with the paramount manifestation being caused by lightning strikes. Studies by the National Aeronautics and Space Administration (NASA) using satellite monitoring receivers have shown that there are on average 100 lightning strikes per second, worldwide, with a mean duration of 2 msec per strike. Therefore, one can expect to see lightning-generated noise for approximately 200 msec in every second, no matter where the receiver is located.

Even though a lightning strike is actually located 1000 miles away, it may produce millivolt-level signals at a receiver's input. This author's experience is that while observing a 50,000-W transmitter's signal (using a spectrum analyzer) at 170 kHz and a distance of approximately 20 miles, lightning-induced interference from a storm more than 200 miles away produced noise pulses that were 60 to 80 dB greater than the desired 50,000-W transmitter's signal.

We see then that when such conditions exist, there is no practical amount of transmitter power than can make up for the noise-to-signal ratio produced by lightning. (Incidentally, the majority of lightning strikes worldwide occur in the United States and the Caribbean area during the summer months.)

What can one do? Either 1) operate in a different frequency range or 2) employ a form of modulation that can tolerate interference 20 percent of the time. Fortunately, the effects of atmospheric noise fall off rapidly with frequency, so that above 30 to 40 MHz atmospheric interference is negligible. However, within the HF* band and lower frequencies (which include the AM broadcast band) atmospheric noise is as much as 40 dB greater than a receiver's internal noise.

A7.2 GALACTIC NOISE

Noise generated from outside the earth's atmosphere is designated galactic noise. Its effect is seen primarily within the VHF and UHF bands below approximately 400 MHz. Even at frequencies within the military VHF (30 to 88 MHz) and UHF (225 to 400 MHz) bands, galactic noise is not very much of a problem, primarily because of the antennas employed. That is, the antennas usually employed in such systems are vertical dipoles whose response is omnidirectional in azimuth but with a null overhead. This characteristic very effectively favors earthbound transmitters and receivers, which is enough to allow reliable operation.

A7.3 MAN-MADE NOISE

By far the worst of all sources of noise are human sources. Auto ignition, industrial processes, neon and fluorescent lights, microwave ovens, and many other electrical/electronic devices combine to raise the level of ambient noise. It is not unusual in a large city to find levels of ambient noise that are 20 dB or more above the level predicted from thermal noise analysis (KTB).

For this reason as well as others, including fading and poor signal propagation due to weather conditions etc., large link margins—as much as 40 to 50 dB—are often provided (i.e. link budgets that include more power than the minimum derived from analysis of the free space loss are said to have "link margin").

A7.4 KTB NOISE

Thermal noise, which is effectively the noise level seen at the input to a receiver, has a power level

$$\begin{aligned} \mathrm{KTB} &= \text{Boltzmann's constant } K \times 290 \text{ kelvin} \times \text{receiver bandwidth} \\ &= 1.38 \times 10^{-23} \times 290 \times B_{\mathrm{N}} \text{ (watts)} \end{aligned}$$

*High frequency is conventionally (but incorrectly) considered to be 2 to 30 MHz. See Appendix 5.

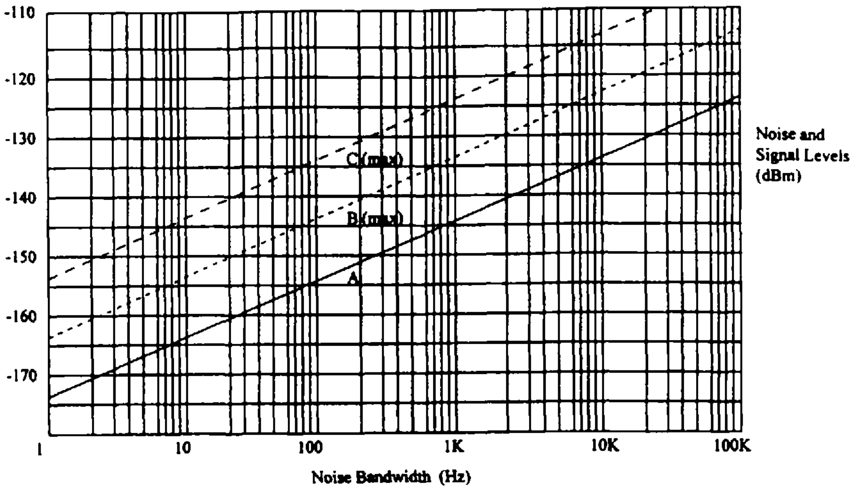


Figure A7.1 Noise bandwidth and sensitivity for receiver with 10-dB noise figure (including losses) and 10-dB $(S + N)/N$ ratio. A = KTB; B = KTB + noise figure (10 dB); C = sensitivity range for $(S + N)/N$ (10 dB).

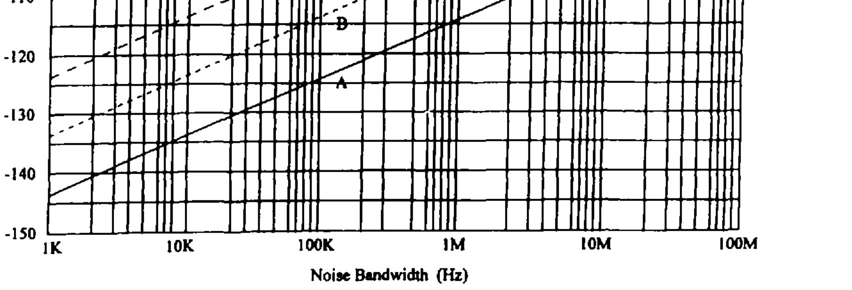


Figure A7.2 (Continuation of Fig. 1.) A = KTB; B = KTB + noise figure (10 dB); C = sensitivity range for $(S + N)/N$ (10 dB).

or in terms of dBm,

$$\text{KTB} = -174 + 10 \log B_{\text{N}} \text{ (dBm)}$$

where K is Boltzmann's constant, T is 290 kelvin (by agreement), and B_{N} is the noise bandwidth of the receiver. The temperature 290 degrees kelvin is agreed upon by the concerned engineering societies for the sake of uniformity. It corresponds to 16.82° C or 62.276° F.

Receiver noise bandwidth is

$$B_{\text{N}} = \frac{1}{gpk} \int_{-\infty}^{\infty} g(f)\, df$$

which is the equivalent rectangular bandwidth resulting from integrating the response of a receiver at all frequencies and dividing by the peak value. In practice, the 3-dB bandwidth of the filter used to determine receiver selectivity is assumed to be the receiver's noise bandwidth.

Figures A7.1 and A7.2 show KTB versus bandwidth for bandwidths from 1.0 Hz to 100 MHz. Also shown is the range of KTB plus noise figure for noise figures from 0 to 10 dB and sensitivity range for receivers requiring up to 10-dB $(S + N)/N$ ratio.

For example, if a receiver has 10-kHz bandwidth, then Fig. A7.2 shows that KTB = −134 dBm. A 10-dB noise figure would raise the effective input noise to −124 dBm, and a signal with 10-dB $(S + N)/N$ ratio would mean the receiver's sensitivity is −114 dBm.

Appendix 8
dBs, dBms, and Other Mysteries

Decibel measures are not, as some have said, a dialect used by engineers to mystify others, but a way to avoid mistakes in having to multiply and divide large numbers. The use of dBs is complicated, however, by the existence of what appear to be different kinds of dBs.

In fact, a dB measurement is a ratio, expressed in logarithmic form. Let us suppose that we have an amplifier whose input voltage is 0.1 V and whose output is 10,000 V. The amplifier's gain is simply output voltage/input voltage = 10,000/0.1 = 100,000. In dB notation, this amplifier would be said to have gain of

$$20 \log 100{,}000 = 100 \text{ dB}$$

(note that this is *voltage gain*).

If we are interested in the power gain, as in an audio amplifier, where sales are predicated on the ability of the amplifier to drive a speaker system to a level of hundreds of watts, we would express the gain as a *power gain*, which is

$$10 \log 100{,}000 = 50 \text{ dB}$$

This difference is due to the fact that power is a function of voltage. That is,

$$P \text{ (watts)} = \frac{E^2}{R}$$

and because of this a ratio expressed in voltage dB is given as a different number. In either case, the logarithmic form can be added or subtracted instead of multiplied or divided to reach the same answer. Throughout this book, unless otherwise noted, all dB measures are given in *power* dBs.

Although dB measures are actually ratios, dBs are also employed regularly to signify absolute levels. In receiver applications, for example, it is common to speak of receiver sensitivity in terms of dBm, which is simply the number of dBs above or below 1.0 mW, which is therefore a reference level of "zero dBm," and a

Table A8.1 Voltage and Power Ratios in Terms of dBV and dBm

RMS voltage	dBV	Power (W/dBm) 50 ohms	75 ohms	300 ohms
100	40	200/53	133/51.2	33.3/45.2
70.7	37	100/50	67/48.2	16.7/42.2
50	34	50/47	33/45.2	8.3/39.2
30	29.5	18/42.5	12/40.8	3/34.8
10	20	2/33	1.33/31.2	0.3/25.2
7.07	17	1/30	0.66/28.2	0.16/22.2
5.0	14	0.5/27	0.33/25.2	0.08/19.2
3.0	9.5	0.18/22.5	0.12/20.8	0.03/14.8
1.0	0.0	0.02/13	0.13/11.2	0.003/5.2
0.707	−3	0.01/10	0.007/8.2	0.0017/2.2
0.548	−5.2	0.006/7.8	0.004/6	1 mW/0
0.5	−6	0.005/7	0.003/5.2	0.0008/−0.9
0.3	−10.5	0.0018/2.5	0.0012/0.8	0.0003/−5.2
0.273	−11.25	1.5 mW/1.7	1 mW/0	0.24 mW/−6
0.223	−13	1 mW/0	0.66 mW/−1.8	0.16 mW/−7.8
0.1	−20	0.2 mW/−7	0.13 mW/−8.8	0.03 mW/−15.2
7.07 mV	−43	1 μW/−30	0.67 μW/−31.8	0.16 μW/−37.8
1.0 mV	−60	2×10^{-8}/−47	1.3×10^{-8}/−48.7	3.3×10^{-9}/−54.7
1.0 μV	−120	2×10^{-14}/−107	1.3×10^{-14}/−108.7	3.3×10^{-15}/−114.7

Table A8.2 Voltage and Power Ratios to the Nearest 1/10 dB

Ratio	Power (dB)	Voltage (dB)
1	0	0
2	3	6
3	4.8	9.5
4	6	12
5	7	14
6	7.8	15.6
7	8.5	16.9
9	9.5	19.1
10	10	20
100	20	40
1000	30	60

Table A8.3 Definitions of Various dB-Related Terms Commonly Used

Term	Definition
dB (power)	$10\log\dfrac{\text{power 1}}{\text{power 2}}$
dB (voltage)	$20\log\dfrac{\text{voltage 1}}{\text{voltage 2}}$
dBc	$10\log\dfrac{\text{trial signal power}}{\text{standard signal power}}$ (example—the level of a modulation sideband compared with its carrier)
dBd	$10\log\dfrac{\text{trial antenna gain}}{\text{dipole antenna gain}} = 10\log\dfrac{\text{trial antenna gain}}{1.66}$
dBi	$10\log\dfrac{\text{trial antenna gain}}{\text{isotropic antenna gain}} = 10\log\dfrac{\text{trial antenna gain}}{1.0}$
dBm	$10\log\dfrac{\text{trial signal power}}{1.0\ \text{mW}}$
dBμV	$20\log\dfrac{\text{trial signal power}}{1.0\ \mu\text{V}}$
dBmV	$20\log\dfrac{\text{trial signal power}}{1.0\ \text{mV}}$
dBV	$20\log\dfrac{\text{trial signal power}}{1.0\ \text{V}}$
dBW	$10\log\dfrac{\text{trial signal power}}{1.0\ \text{W}}$

receiver's sensitivity of 1.0 μV is considered to be -107 dBm at 50 ohms. That is, 1.0×10^{-6} V at 50 ohms is

$$1.0 \times 10^{-6} \text{ V at 50 ohms is } \frac{(1.0 \times 10^{-6})}{50} = 2 \times 10^{-14} \text{ W}$$

$$\text{and } 10\log\frac{1 \times 10^{-3}}{2 \times 10^{-14}} = 106.9897 \text{ dB}$$

One microvolt across 50 ohms is equivalent to a signal 107 dB below 0 dBm (1.0 mW) or -107 dBm.

Similarly, power is also expressed in terms of dBW (dB above or below 1 W), dBc (dB above or below a carrier level), and dBV (dB above or below 1 V).

Finally, it is necessary to realize that differences in dBm levels, for example, are expressed in dB and that adding or subtracting 10 dB to a level of -50 dBm produces -40 dBm or $-$ 60 dBm, respectively.

Tables A8.1 and A8.2 list common levels in terms of voltage, power, dBm, and dBV notation, for 50, 75, and 300 ohm impedances. Definitions of some of the measures that are expressed in terms of dBs are given in Table A8.3.

Appendix 9

Power Transfer and Impedance Match

At the input to a receiver, where the signal from an antenna (or other transducer) enters the receiver, an impedance match is very important, as the degree to which the receiver input impedance matches the antenna source impedance determines the energy transfer between the two. Maximum power transfer from one to the other occurs when the impedances are the same. In most cases, receivers and antennas, for example, are designed to have low impedance—preferably resistive as opposed to reactive. Typical impedances are 50, 75, and 300 ohms, with the receiver's application determining the particular impedance employed. For instance, television receivers typically employ 75 ohms with coax input or 300 ohms with twin-lead signal input transmission line.

To show that maximum power transfer does indeed occur when the impedance of a signal source is matched to the impedance of the load, let us assume a signal source and load as shown in Fig. A9.1.

Referring to Fig. A9.1,

$V_{in} \equiv$ source voltage (constant)
$Z_S \equiv$ source impedance
$Z_L \equiv$ load impedance

Current in the circuit, through Z_S and Z_L, is

$$\frac{V_{in}}{Z_S + Z_L} = i_L$$

and power dissipated in the load is

$$\left(\frac{V_{in}}{Z_S + Z_L}\right) Z_L = P_L$$

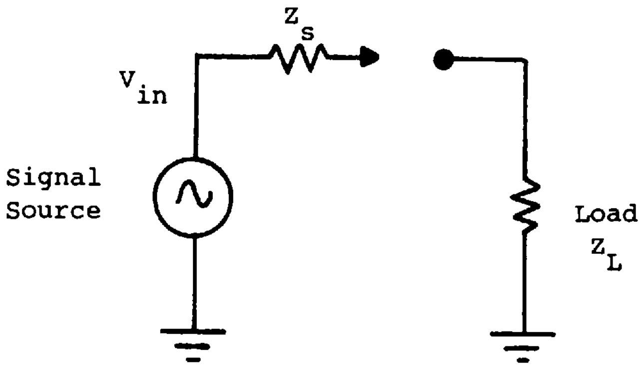


Figure A9.1 Illustration of signal source and load.

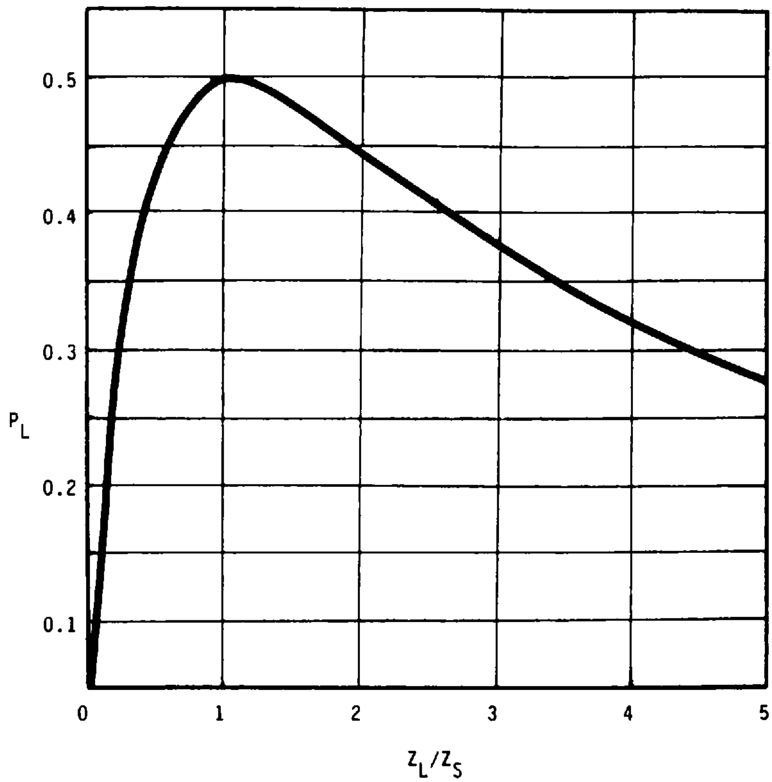


Figure A9.2 Power transfer as a function of impedance match Z_L to Z_S.

while power input is

$$\left(\frac{V_{in}}{Z_S + Z_L}\right)(Z_S + Z_L) = P_{in}$$

Figure A9.2 is a plot of P_L as a function of the ratio of Z_L to Z_S. Note that the power delivered to the load is maximum when $Z_L/Z_S = 1$ and that the maximum power transferred is $P_{in}/2$.

The same concern for a proper impedance match is just as important at every interface between circuits within a receiver as well as the interfaces that enter or leave the receiver. The antenna input is certainly important, because of the extremely small signal power available and the noise that is always present. (When the signal power is less than the noise power within the signal bandwidth, very few if any techniques exist that will allow recovery of the signal.) At this point it is sufficient to say that a proper impedance match is always important everywhere in a receiver because of the potential loss of signal.

This page intentionally left blank

Appendix 10
Grounding, Bypassing, and Shielding

Grounding, bypassing, and shielding are vital to the operation of receivers and cannot be overlooked. The way that all three are accomplished is specific to the operating frequency and the configuration of the particular receiver on which the grounding, bypassing, and/or shielding is applied, however. Therefore it is not practical to give detailed information that holds for every receiver design, although it is possible to give some general rules that always hold. These are:

1. Make ground lead impedance as low as possible. This means that the lead should be short and have as large a cross section as is practical. (Often, ground leads are made of braided wire, especially in high-current applications.) On a printed circuit board, the best practice is to employ one or more board layers as a ground plane(s) and to employ a plated through-hole placed as close to the point to be grounded as possible and connected directly to the ground plane. Where more than one ground layer is employed, it is also good practice to connect the ground layers together with plated through-holes at intervals much closer than 1/4 wavelength, all over the board.

2. Bypass capacitors should be placed as close to the supply voltage input of a circuit as possible and any other bias points as well. The type of capacitor and its value that should be used are a function of the frequency at which the circuit is employed. The function of a bypass capacitor is to lower the impedance of the point being bypassed, so that variations in current at the point being bypassed do not cause an unwanted voltage variation to be developed across the impedance, because such a variation could cause either a loss in signal or oscillation.

It is also important that the different stages in a receiver be well isolated from one another, to prevent interaction and instability. It would be advantageous from the standpoint of isolation to have a separate power supply for each stage in a receiver, but it would not be very practical and isolation would still be a problem. The approach usually employed is to provide the same supply voltage to every stage in a receiver but to isolate them with inductor-capacitor circuits. The

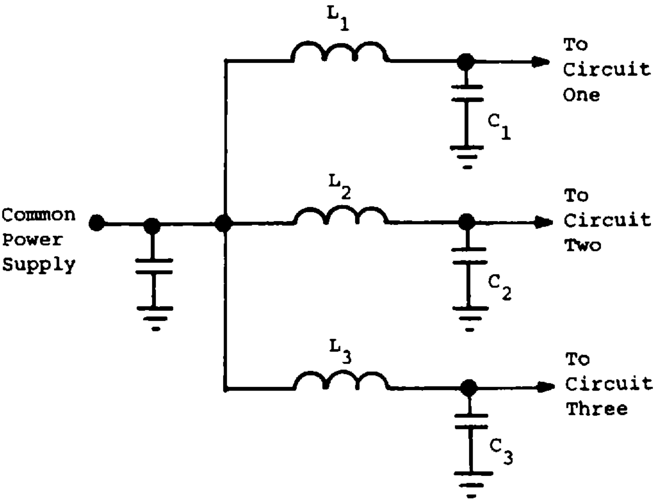


Figure A10.1 Typical circuit used for isolation of power input to several circuits from a single power source, using series inductors and bypass capacitors to ground.

capacitors serve as low-impedance bypasses, but the inductors act in series with the supply line to provide a high impedance at the operating frequency, in the supply line itself. This allows a signal on the supply line to be attenuated by

$$20 \log X_L/X_c$$

at any frequency of interest, where the inductor is in series with the supply line and the bypass capacitor is connected from the circuit input point to ground, which forms an "L" circuit as shown in Fig. A10.1.

The type and value of the capacitors used are very important, as the impedance characteristics of various types of capacitors are different, depending on the frequency range in which they are being used. In receivers, ceramic capacitors are commonly used for bypassing over wide bandwidths, but mica capacitors are also used, and at lower frequencies tantalum capacitors may be incorporated. Many other materials are employed in making capacitors, and it is not unusual in critical applications to see bypassing done in which several types of capacitors are placed in parallel with one another, so that the best characteristics of each are captured.

3. Shielding is also employed to isolate circuits from one another, to reduce unwanted emissions, and to inhibit reception of undesired signals. Shielding improves the stability of high-gain receivers and reduces local oscillator leakage.

It is not unusual in some countries to require that an entire receiver be enclosed in a metal enclosure for shielding purposes, but shields (where needed) are more often placed directly on a printed circuit board, around only the circuits that require them.

Shields are effective in reducing unwanted signal transmission or reception, typically by as much as 60 to 80 dB.

Inductors used for isolation should also be chosen carefully. Their use is also a function of the frequency range for the particular application, just as in the case of the capacitors used for bypassing. In general, inductors should be employed whose self-resonance is at a higher frequency than the operating frequency of the circuit for which they are providing an isolating impedance. (See Fig. A10.1.)

This page intentionally left blank

Appendix 11
Simulations

This appendix includes a pair of examples of simulated receiver operation from two different companies whose software for simulation both is affordable and will run on a personal computer. These are given for information and are not meant to endorse or critique the products. The cooperation of Tesoft, Inc. and Elanix, Inc. in providing these examples is greatly appreciated.

The first example, from Tesoft, is reprinted with permission from February 1994 *RF Design* magazine, copyright 1994 Intertec Publishing Corp., Overland Park, Kansas. Tesoft can be called at (800) 631-1113 for further information. The article was written by Don Miller of MCC Panasonic, with Tesla software used for simulation.

The second example was written specifically for this appendix by Dr. Morrie Schiff and his colleagues at Elanix. Elanix can be called at (818) 597-1414. This simulation was done with SystemView software and was performed in cooperation with Comquest Corp.

This page intentionally left blank

Appendix 11a
Simulating a QPSK Modem

Don Miller

Time domain simulation can provide insights into the operation of a QPSK modem, insights that are unobtainable with frequency domain tools. SPICE-type tools were not used in the simulations outlined here. TESLA, one of the available block level simulators, was used instead. Tools such as these animate the behavior of all the system blocks in parallel. The blocks interact just as they would in the laboratory. This is much faster than circuit simulation, which solves large matrices of circuit equations. Unlike steady-state nonlinear simulators, a tool like TESLA easily handles complex digital logic and long sequences of data.

RF engineers are comfortable analyzing the performance of their designs in the frequency domain. For many problems this makes sense because most RF components are characterized as a function of frequency. The effects of most filtering (and even some nonlinear blocks, such as mixers) are easier to understand in the frequency domain. Often, however, the baseband performance of a design is not so easily analyzed in the frequency domain.

Just as an oscilloscope will be favored over a spectrum analyzer for observing some aspects of system behavior in the laboratory, alternative software tools will be better suited for simulating some aspects of a design. If a system is linear and time invariant, then linear simulation will be the tool of choice. Nonlinear systems that need be analyzed only in the steady-state case are good candidates for harmonic balance analysis. Time domain simulation is a tool that is useful for simulating systems where transient behavior must be studied or where frequency domain analysis provides little intuitive insight. For example, spectral analysis may show distortion, but a time plot shows the actual clip point of a waveform. However, employing a time simulation tool does not prevent us from looking at frequency domain behavior. This is provided for by the fast Fourier transform.

A11a.1 MODULATOR

The block diagram of a QPSK RF modem is shown in Figures A11a.1 and A11a.2. Figure A11a.1 details the modulator section. Data and clock signals from the random bit generator go to a data stream splitter. It separates each pair of bits into two parallel data streams, each running at half the input data rate. Encoding is applied to the bit streams to determine the mapping of the bit pairs to the output carrier phase state. The resulting bit pairs, which define the current "symbol," are filtered with low-pass filters having carefully controlled time responses. The resulting analog waveforms are applied to a quadrature modulator to generate the QPSK signal.

A11a.2 DEMODULATOR

The demodulator circuit in Figure A11a.2 operates almost like a modulator in reverse. The quadrature demodulator drives filters that can be identical to the ones in the modulator. The filtered signals are sliced with a comparator, decoded, and serialized into the same data stream as was transmitted. If only it were that easy! In reality, the demodulator is much more complicated than the modulator. Carrier recovery is the problem. The local oscillator signal used in the quadrature demodulation process needs to be synchronized with the incoming carrier frequency (which is suppressed). Often, the demodulator is part of a special phase-locked loop, called a Costas loop, that recovers the carrier. The mixers in the demodulator are part of the phase detector, so the input signal level affects the phase detector sensitivity, (K_F). This means AGC is required! After the data is decoded, we must provide a bit rate clock to the device that is using the data we are delivering. This is yet another PLL! Luckily, we can "ship" clock and carrier to the demodulator from the modulator during our simulations until we are ready to add the carrier and clock recovery circuitry. This enables us to keep things simple at first and then add complexity as we are satisfied with the basics. Our ability to pick and choose between the real and ideal worlds enables us to simulate only as much of the

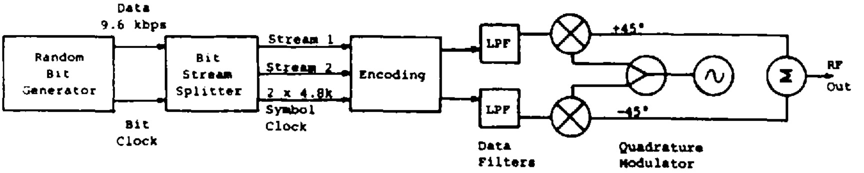


Figure A11a.1 Block diagram of a QPSK modulator.

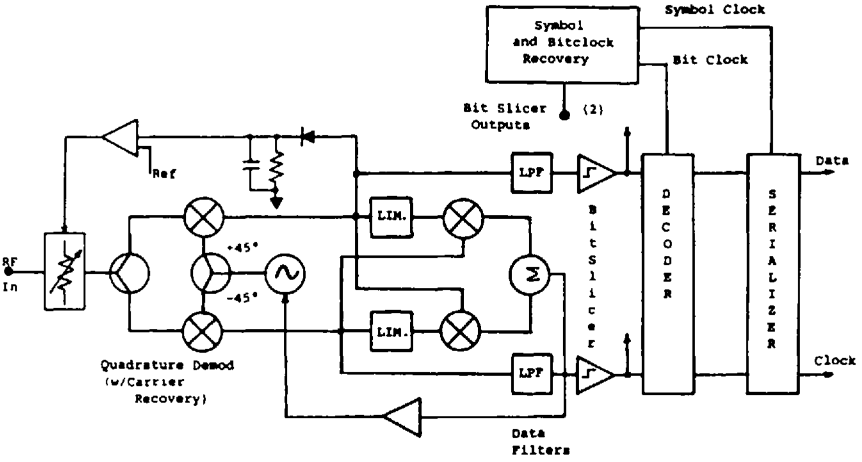


Figure A11a.2 Block diagram of a QPSK demodulator.

system as we care or need to. The simulations here will carefully ignore clock recovery and AGC, while illustrating how filter effects and carrier recovery may be analyzed.

A11a.3 FILTER PERFORMANCE

After confirming that the data stream splitter and coder operate properly, the first real analysis task on the modulator side will be to confirm proper operation of the low-pass data filters. If the filtering in a data link is optimum, the intersymbol interference (ISI) will be zero. Any ringing from the previous symbols or precursors from future symbols will pass through zero at each sample point. Figure A11a.3 shows the impulse response for the filtering in a system with zero ISI and one with significant ISI. Note that for the filter with zero ISI, the zero crossings of the impulse response coincide with the sampling points. If our filters have significant ISI, then the nonzero residual time response of the previous symbol will add to the pulse resulting from the next symbol. This interference adds to the uncertainty in interpreting the state of a particular symbol.

We can check the filtering in our simulated modem by doing a simulation with the baseband output of the transmit low-pass filters cascaded into the receive filters. A block diagram of this simulation is shown in Figure A11a.4. The output of the receive I or Q channel filter can be observed using a symbol rate time base in Lissajous fashion. This display is called an eye pattern. Wide eye openings are indicative of low ISI. The eye pattern shown in Figure A11a.5 was generated by

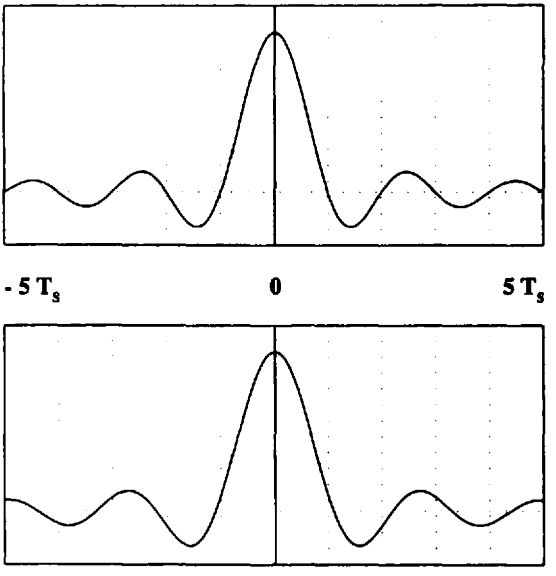


Figure A11a.3 Impulse responses of filters with and without intersymbol interference.

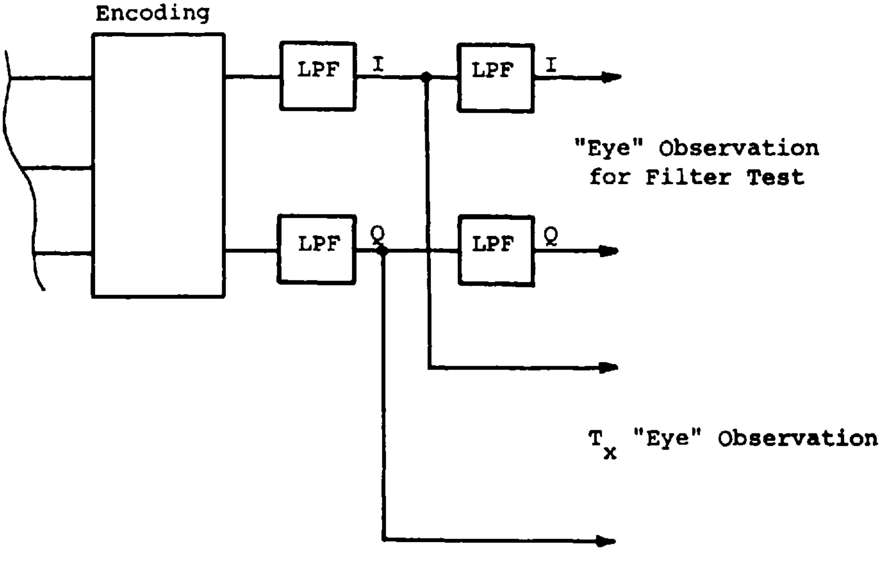


Figure A11a.4 Simulation block diagram.

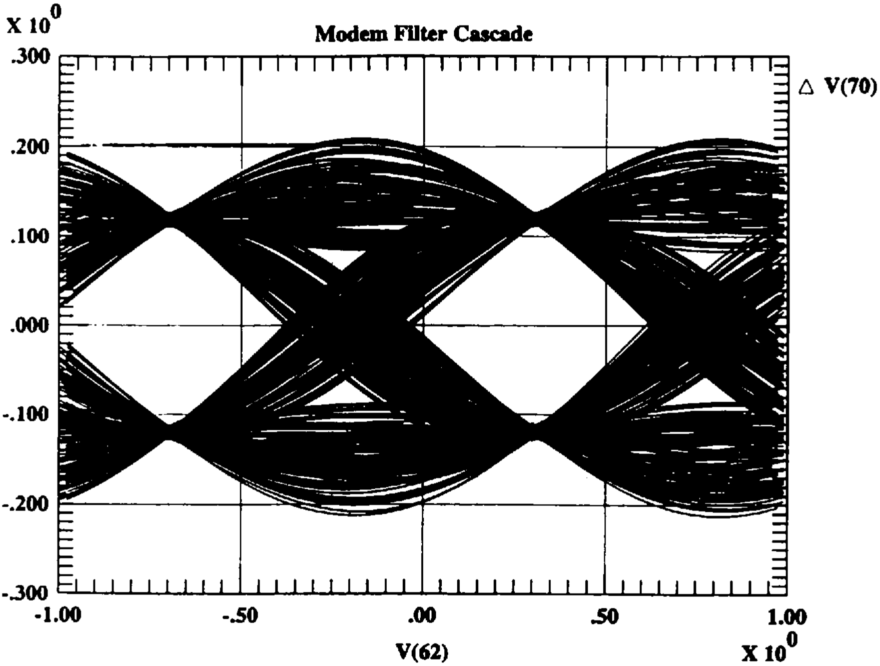


Figure A11a.5 Eye diagram of system in Fig. A11a.4.

simulating the system in Figure A11a.4 with TESLA. The simulation was done for 300 msec with a 1-msec time step. The data from 0 to 100 msec was omitted in the plot to keep start-up transients from corrupting the eye pattern. Significant ISI in the simulation output would have indicated that the filters needed work, but things look OK.

We can observe the actual QPSK signal (Figure A11a.6) by simulating the entire modulator as shown in Figure A11a.1. The transmit eye pattern can be observed as well. Significant ISI will be seen in the transmit eye, since only half of the system filtering is present (Figure A11a.7). This simulation was for 1 second at a 1-msec time step (1 million points). Only 200 msec of the data was plotted. The full second of simulation was necessary so that ample RF output simulation data could be stored to disk for the upcoming system bit error rate tests.

The quadrature demodulator can be tested quickly by using an imported local oscillator from the modulator in lieu of the carrier recovery system in Figure A11a.2. If the I and Q channels are displayed in an *XY* fashion, the phase and amplitude trajectory of the QPSK signal will result. Once the operation of the

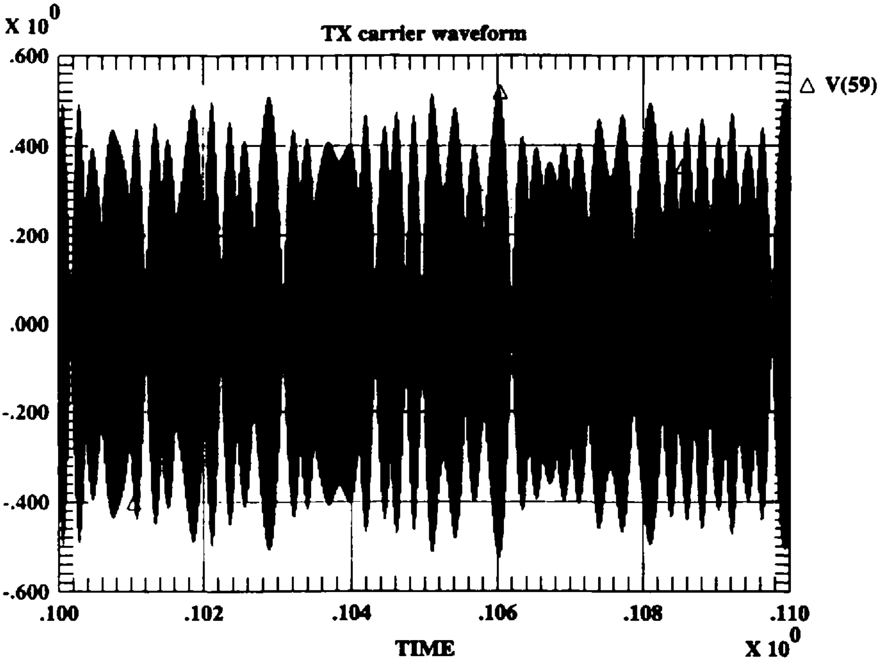


Figure A11a.6 Simulated QPSK signal in the time domain.

quadrature demodulator is confirmed, the carrier recovery circuitry (Costas loop) can be added and debugged.

A11a.4 CARRIER RECOVERY PERFORMANCE

This system is a second-order phase-locked loop, and the same methods as used on frequency synthesizers can be applied. The designer must be aware that the phase modulation on the carrier has spectral components that go quite low in frequency, so the bandwidth of the loop must be quite small. In a real system, switched time constants and/or sweep acquisition circuitry may be necessary for acceptable lock times. These types of systems can be simulated with TESLA, but the fast acquisition problem will be ignored here. The system simulator will let us set initial conditions so that we can simulate only the locked case or just the time period immediately surrounding the lock instant. Figure A11a.8 is a composite of four TESLA plots showing the receive constellation in the unlocked state approximately 1 second before acquisition, nearing the correct frequency as the constel-

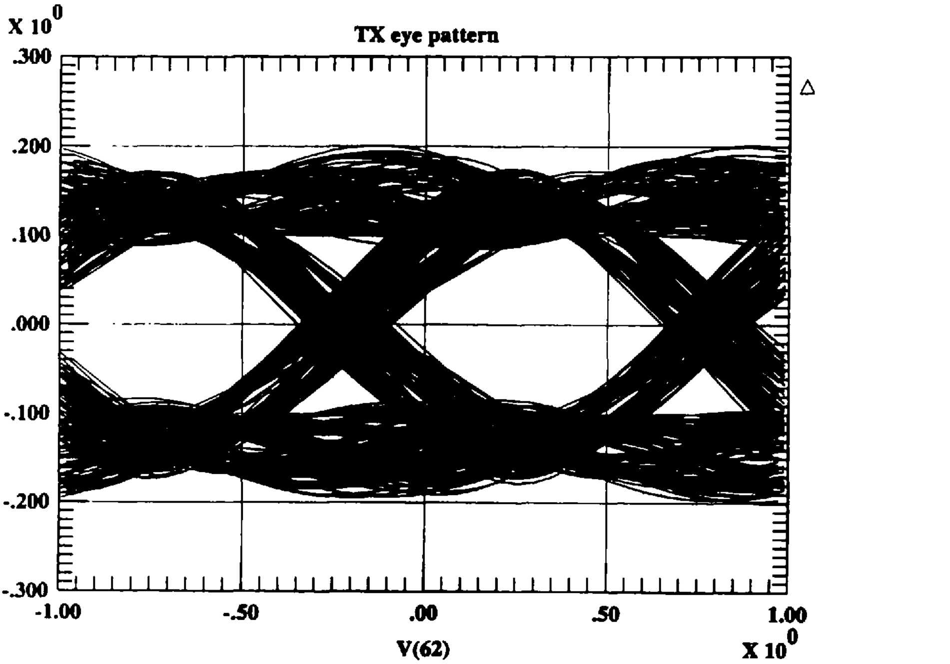
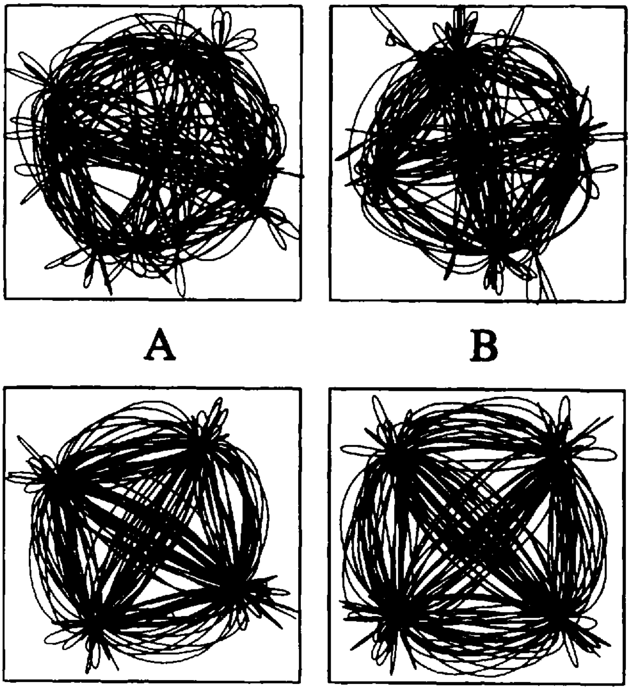


Figure A11a.7 Transmitter eye diagram.

lation spins slower a few tenths of a second later, the constellation on frequency just before phase lock, and finally the locked constellation.

A11a.5 BIT ERROR RATE PERFORMANCE

The real gauge of modem performance is the bit error rate (BER) in the presence of noise. The simulator has a BER tester model that can be used to this end. It is important to simulate the system at the highest error rate possible in order to keep the simulations short. In the case of QPSK, the variation of BER versus the signal-to-noise ratio, usually expressed as the energy per bit divided by the noise in a 1-Hz bandwidth, (E_b/N_0), is published and generally available. The BER of our demodulator with an input (E_b/N_0) of 5.7 dB is about 10^{-2} (Figure A11a.9). The difference in E_b/N_0 (for a given BER) for a real modem and an ideal one is referred to as the implementation margin. An ideal QPSK modem will have a BER of 10^{-2} with an input E_b/N_0 of 4.3 dB. Our radio therefore has an implementation margin of 1.4 dB. This degradation in performance could be due to the small amount of



A B

C D

Figure A11a.8 Receive constellation at various stages of signal acquisition and locking.

ISI resulting from the system filters, the nonzero bandwidth of the Costas loop, or a characteristic of the coding scheme. The coding scheme used here was differential QPSK. The phase of each symbol is interpreted on the basis of that of the previous one. Unfortunately, this means that any error more or less automatically causes an error in the next symbol. This doubles our bit error rate. Without this degradation, the E_b/N_0 would have been 5.2 dB. This leaves only 0.5 dB error that must be accounted for elsewhere (I feel better about my filters already!).

A11a.6 OCCUPIED BANDWIDTH

The tools at our disposal have been valuable for looking at things in the time domain, but sometimes we really need a frequency domain tool. For example, if we need to look at the transmitter occupied bandwidth, an oscilloscope-type display just would not do much for us. The simulator's FFT facility gives us the

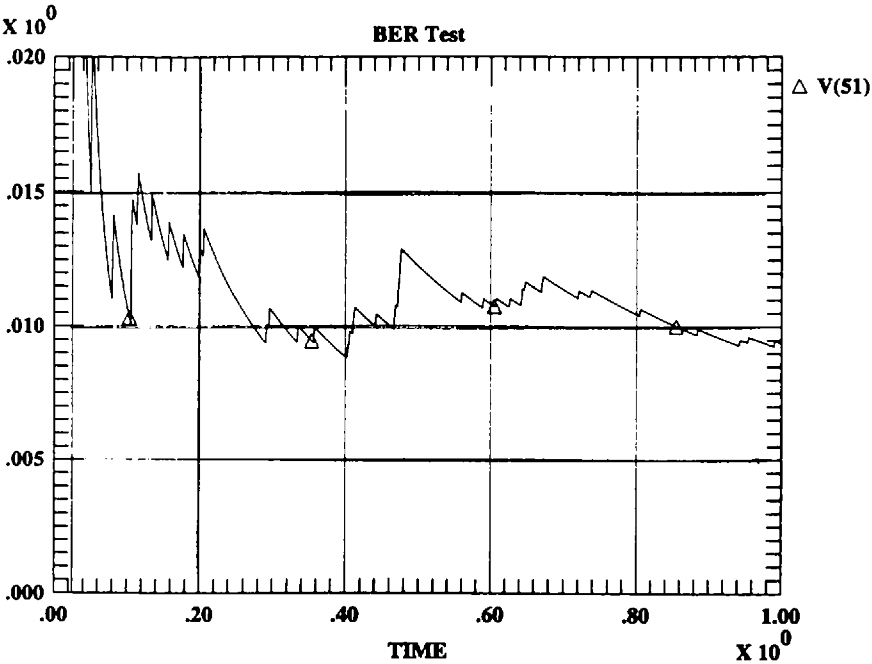


Figure A11a.9 Simulated bit error rate for the modem.

tool we need for the task. Figure A11a.10 shows the results of TESLA's FFT computation on the transmitter output time data. The result is the spectral output of the transmitter.

A11a.7 CONCLUSIONS

Outlined here is the sum of well less than a week of total engineering time. A simple QPSK modem was designed, and several aspects of the design were evaluated. Without touching a single electronic component, we have some idea of the BER performance we can reasonably expect and how long it would take to "lockup" with the loop filter chosen. If modifications need to be made to meet the requirements, we can do it on the simulation before we start burning up time in the laboratory.

Time domain simulation can be used as a valuable tool to study the effects of design decisions well before the hardware prototyping stage. The skeleton of a system can be simulated using ideal components to test concept, and then models

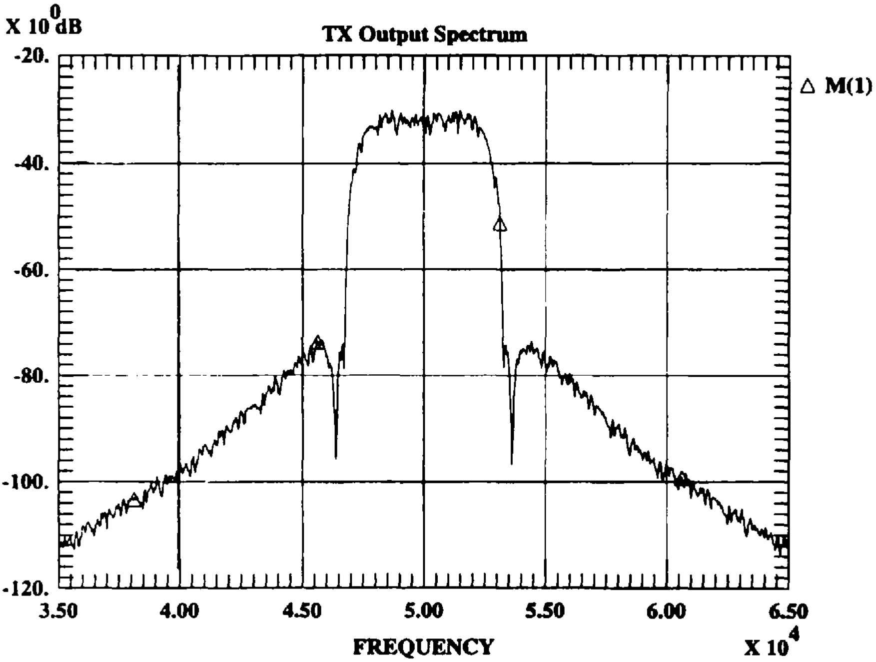


Figure A11a.10 QPSK spectrum.

of the real-world components can be substituted to study the overall effects on the system. With the cost of computer horsepower plummeting and the cost of laboratory time (and time to market) shooting through the roof, many designs can benefit from the initial substitution of keyboard for oscilloscope and spectrum analyzer.

REFERENCES

Manassewitsch, V., *Frequency Synthesizers Theory and Design*, Wiley, New York, 1976.

Poklemba, J. J., Pole-zero approximations for the raised cosine filter family, *Comsat Tech. Rev.*, Vol. 17, No. 1, Spring 1987, pp. 127–157.

Smith, D. R., *Digital Transmission Systems*, Van Nostrand Reinhold, New York, 1985.

Spilker, J. J., Jr., *Digital Communications by Satellite*, Prentice-Hall, Englewood Cliffs, NJ, 1977, pp. 302–312.

Appendix 11b
Simulating RF Systems Using System View by Elanix

Maurice L. Schiff and Stephen H. Kratzet

A11b.1 INTRODUCTION

SystemView by ELANIX is a PC-based simulation tool that provides the power to build quickly and easily accurate models of complete dynamic systems and subsystems through an intuitive block diagram interface. Linear and nonlinear, discrete and continuous time, analog, digital, and mixed mode (hybrid) systems can be created by selecting tokens from the SystemView Libraries and connecting them in the Design window with a few simple mouse clicks.

The RF/Analog Library enhances the SystemView core libraries with essential models required for system-level RF/analog design. This library includes models such as fixed- and variable-gain amplifiers, operational amplifier circuits (op-amps), active mixers, double-balanced mixers, power splitters, power combiners, couplers, diode circuits (including Zener diodes), resistor-capacitor circuits, resistor-inductor circuits, low-pass and high-pass RC/LC filters, PLL filters, LC tank and quadrature circuits, coupled-resonator pairs, and more. The RF/Analog Library tokens may be used to create complete transmitter/receiver systems including the propagated noise figure and intermodulation spurs.

A11b.2 SIMULATION EXAMPLE

The application example is the receiver of a GSM mobile unit. The receiver architecture was supplied by the CommQuest Technologies, Inc. It is modeled after the CommQuest CQT2030 RFIC transceiver chip, which is one of a series of chips devoted to the GSM system. The objective of the simulation is to verify signal levels, spurious products, and demodulation performance. The design example does not necessarily represent CommQuest's recommended chipset design. The

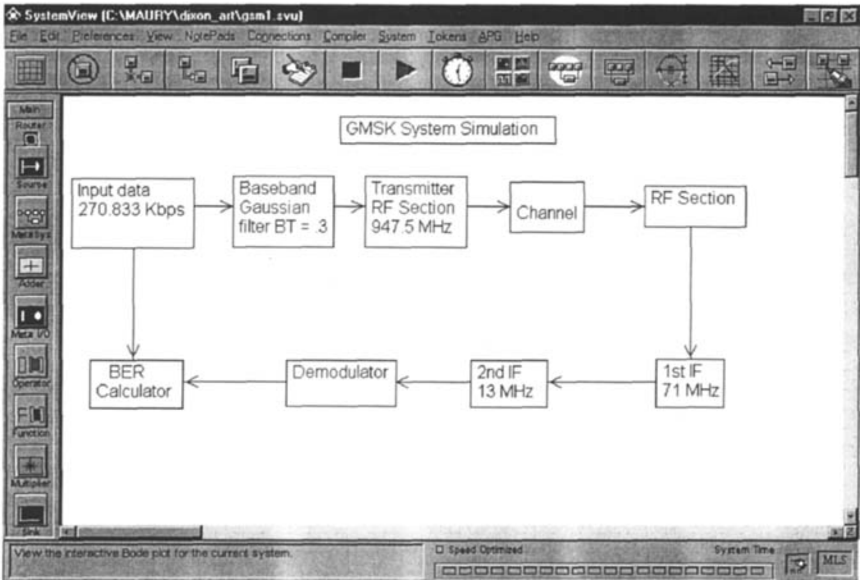


Figure A11b.1 Simulation system block diagram.

overall block diagram is shown in Figure A11b.1. The basic processing blocks are listed below:

Baseband Gaussian filter
Modulator/transmitter
Channel
RF section
1st IF section
2nd IF section
Demodulator
BER calculator

SystemView is a time-based simulator that operates from a master system sample rate. As in any computer simulation, the computations must be carried out in discrete time. The rules of sampling theory apply and must be taken into account. Figure A11b.2 shows the parameter entry form for the master system time window for this simulation. The system sample rate is set at $f_s = 4.096$ GHz. The value is slightly greater than four times the RF frequency of 947.5 MHz. This sets the sum frequency term $f_+ = 947.5 + 876.5$ MHz out of the 1st IF mixer to be

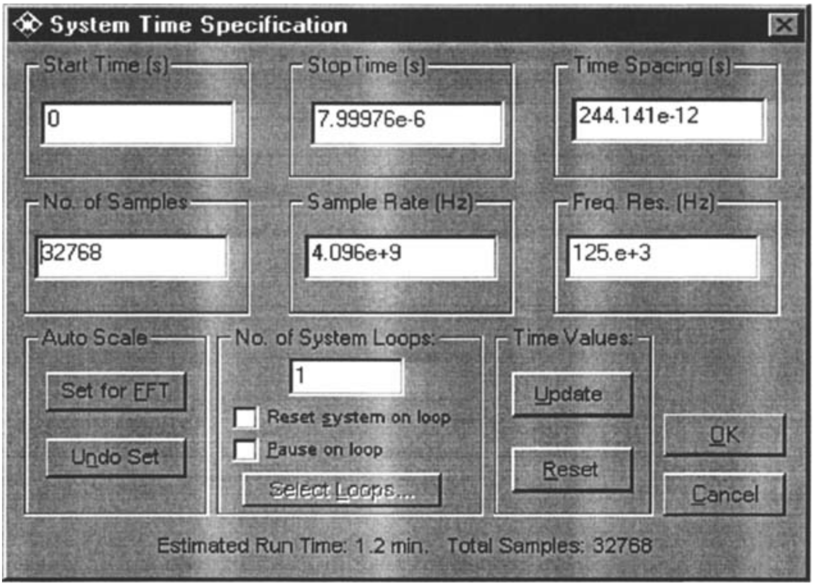


Figure A11b.2 SystemView Master time parameter form.

less than $f_s/2$, which prevents aliasing of this signal. Note that *any* unrelated data rate below 4.096 GHz is possible, as long as proper aliasing is taken into account.

The simulation is implemented by selecting the desired functional elements, or tokens, which reside in libraries at the left edge of Figure A11b.1. Each token element has an appropriate set of parameters where the desired numbers are entered. A typical parameter set for the RF Library amplifier is shown in Figure A11b.3. SystemView provides for a hierarchical structure where several tokens can be combined into one token known as a Metasystem. The processing elements of this simulation were implemented as a group of Metasystems.

The transmitter representing the base station is straightforward. For the purposes of this simulation, the transmitter is developed from individual component parts. A complete mobile transceiver architecture is provided by the CommQuest chips CQT2010, CQT2020, CQT2030.

A11b.3 BASEBAND GAUSSIAN FILTER

The baseband Gaussian filter processing is shown in Figure A11b.4. A binary data source with rate R = 270.833 kHz is passed through a Gaussian LPF with a BT

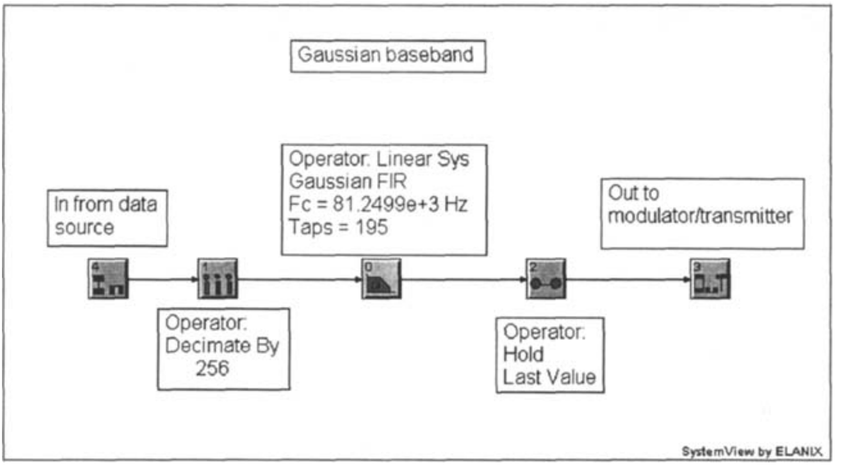


Figure A11b.3 Amplifier parameter input form.

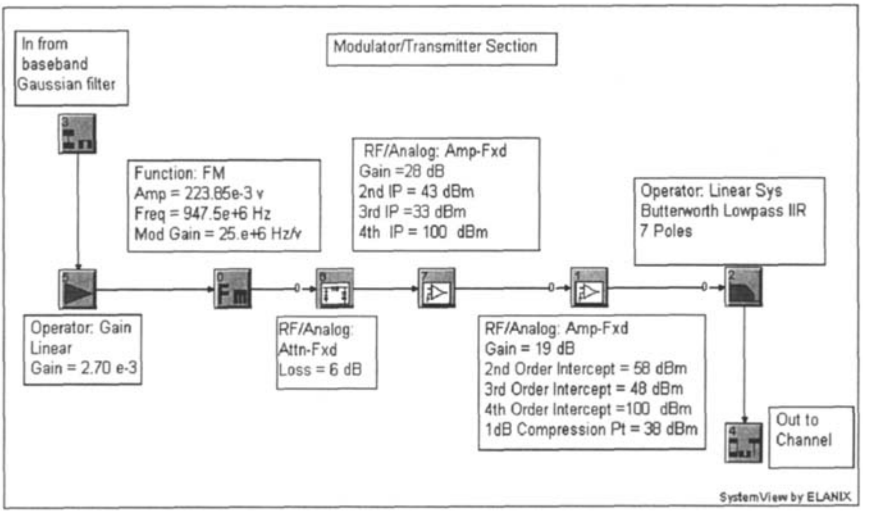


Figure A11b.4 Gaussian filter processing.

= 0.3 setting. This highly compacts the occupied bandwidth of the signal while introducing intersymbol interference. This portion of the transmitter can be run at a rate lower than the full system sample rate and eases the computational requirements on the Gaussian filter. The decimator rate chosen is 1/256 of the main system sample rate. After the filter, the rate is restored prior to the modulator/transmitter section.

A11b.4 MODULATOR/TRANSMITTER

The modulator/transmitter section is shown in Figure A11b.5. The frequency band covers 935–960 MHz. For the simulation, the midband frequency of 947.5 MHz was chosen. The operation must shift the 947.5 MHz carrier by $\pm R/4 = 67.71$ kHz. The VCO represents a Murata MQE001-902 modulator. The gain of the part is 25 MHz/v. Therefore, the output of the Gaussian filter is passed through a gain of $G = 67.71e3/25e6 = 2.71e-3$ (-25.7 dB). The nominal output power of the VCO is -3 dBm. The desired transmitted power is 5 W (37 dBm), which is representative of a base station. The power amplifier chosen is a Mini Circuits TIA-1000-4. The gain is set at 40 dB to provide the desired output power of $+37$ dBm. The final element of the transmitter is a low-pass filter used to eliminate spurious harmonics of the power amplifier.

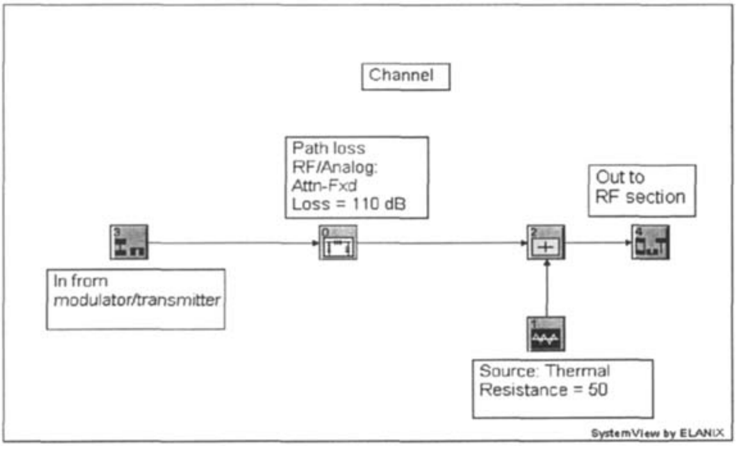


Figure A11b.5 Modulator transmitter section.

A11b.5 CHANNEL

The channel block shown in Figure A11b.6 consists of two parts. First, a gain (pad) token is used to reduce the 5 W transmit power by the path loss of the link (including antenna gains). The second element is the addition of thermal KT noise, which enters the receiver with the signal. It is possible to add any variety of fading phenomena at this point. It is also possible to add more transmitted signals at different carrier frequencies to simulate the effects of adjacent channel interference.

A11b.6 RF SECTION

The receiver is a dual-conversion architecture with a 1st IF frequency at 71.0 MHz and a 2nd IF frequency at 13.0 MHz.

The first section of the receiver is the RF section (see Figure A11b.7), which covers the 935–960 MHz band. The first element of the receiver, after the antenna, is the Murata DFY2R902CR947BGH duplexer. This part effectively acts as a bandpass filter with the specifications given. The next element is the HP MGA-87563 LNA amplifier. Figure A11b.7 shows the SystemView parameter input form used to implement its technical specifications. Note that all intercept points up to fourth order, as well as the 1-dB compression point, noise figure, and linear gain, can be specified. The parameters listed are with respect to the output

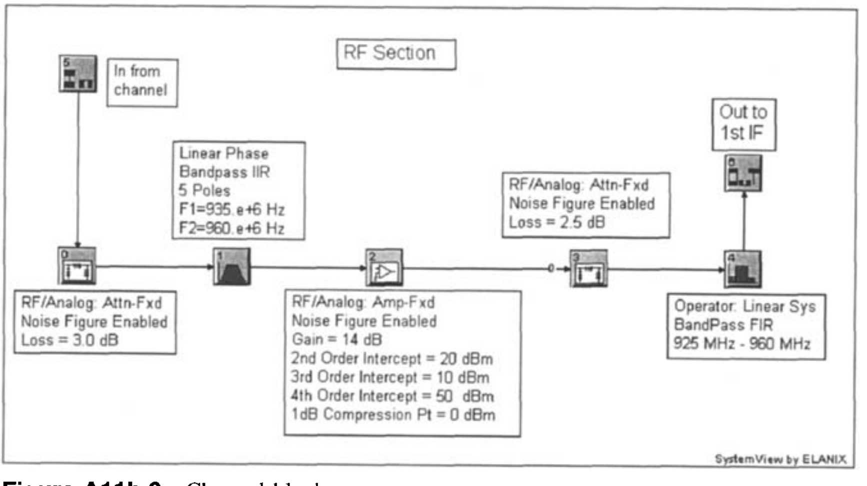


Figure A11b.6 Channel block.

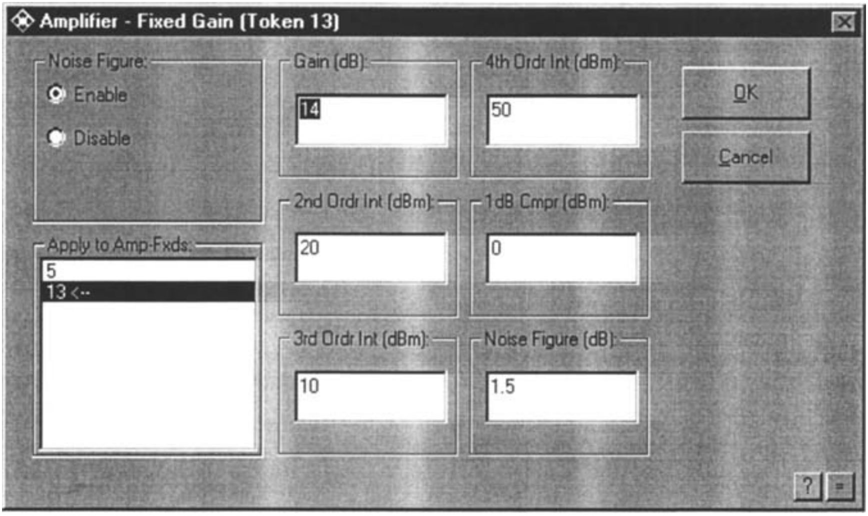


Figure A11b.7 RF section.

of the amplifier. The RF filter is a Fujitsu FAR-F5CH-947M50-L2EM surface acoustic wave (SAW) bandpass filter. This was implemented as a 319-tp FIR bandpass filter. The attentuator pads are used to simulate the filter loss and add the appropriate noise.

A11b.7 1ST IF SECTION

The 1st IF section is shown in Figure A11b.8. The first CommQuest CQT2030 on chip component is the 1st IF mixer. The mixer LO can be tuned over the 25.0-MHz range 864.0–889.0 MHz. The specific value is 876.5 MHz, which is required to produce the 71-MHz 1st IF frequency. The SystemView parameter input form for the mixer is shown in Figure A11b.9. These values are with reference to the input. In addition to the intercept points and other parameters, the LO leakage values can be specified.

The 1st IF filter is an off-chip SAWTEK 854252-1 SAW filter. For simplicity, a three-pole Butterworth filter was used. After the SAW filter, the highest frequency is 71 MHz as opposed to 947.5 MHz. It is therefore possible to decimate the filter output to a much lower sampling rate. In this case, a decimation rate of 10 was chosen. Thus, all tokens after this decimation operate at a rate no higher than 409.6 MHz. This greatly decreases the simulation time. The output of this filter

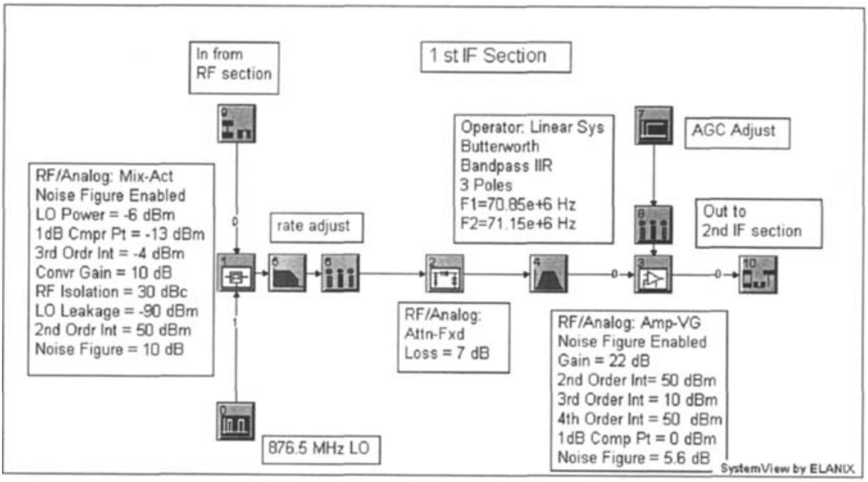


Figure A11b.8 1st IF section.

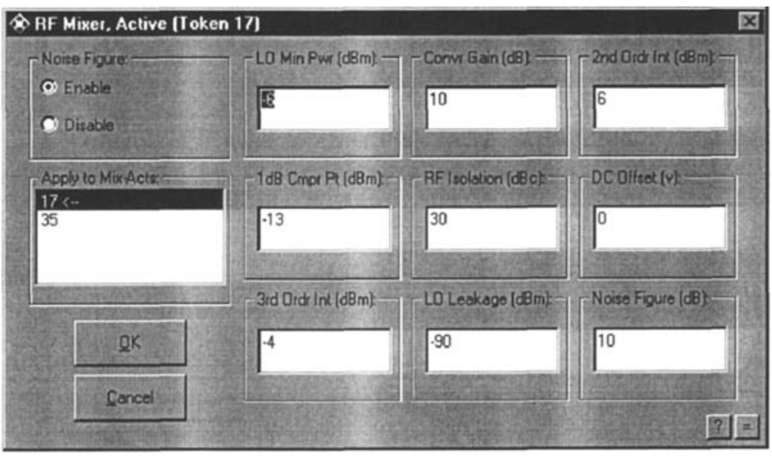


Figure A11b.9 Mixer input form.

(after decimation) is entered into an AGC amplifier/mixer with parameters as shown.

A11b.8 2ND IF SECTION

The 2nd IF section is shown in Figure A11b.10. The 2nd IF LO is a set 58 MHz and produces a 13.0-MHz 2nd IF frequency. This 2nd IF frequency signal is passed through a ceramic filter, modeled by a four-pole Bessel, and an AGC amplifier with nominal gain of 60 dB. The amplifier parameters are shown on the block diagram in Figure A11b.10. The output of this section corresponds to the output of the CQT2030.

A11b.9 DEMODULATOR

The CQT2020 and CQT2010 chips are designed for optimum demodulation and final voice recovery. For this simulation we are interested only in recovery of the digital data. A simple quadrature FM detector operating directly on the 13-MHz IF signal was chosen for this purpose. This allows a relative comparison of the effects of different RF components. A delay line is used to shift the 13-MHz carrier 90 degrees. The output of the quadrature mixer is filtered and amplified to recover the original data. Figure A11b.11. shows the SystemView block diagram.

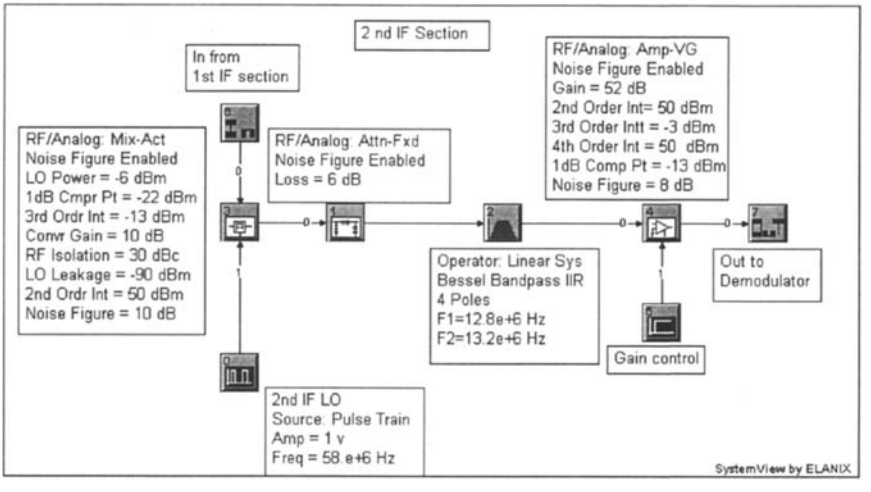


Figure A11b.10 2nd IF section.

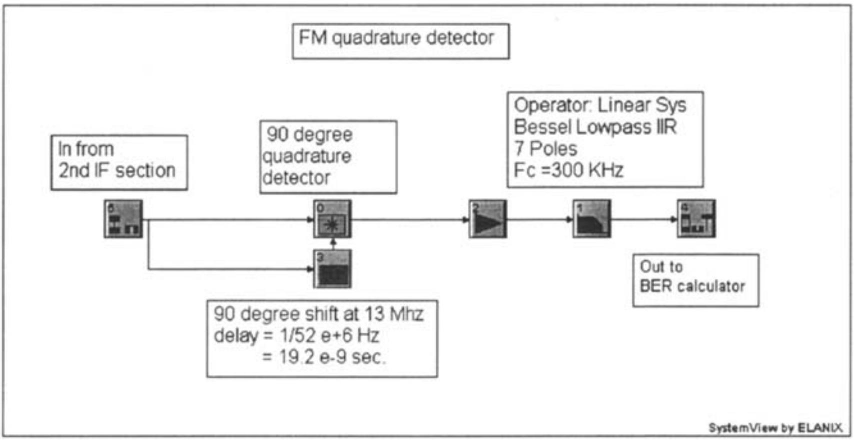


Figure A11b.11 FM demodulator.

Figure A11b.12 shows an overlay of the input to the FM modulator and the recovered signal at the output of the demodulator. The time shift is due to the real group delay through the various system filters.

A11b.10 BIT ERROR RATE (BER) CALCULATOR

The BER counter setup is shown in Figure A11b.13. The demodulated output of the quadrature detector is sampled at the maximum eye opening, i.e., at midbit. The hard decision is compared with the "truth" data in the BER token. The system

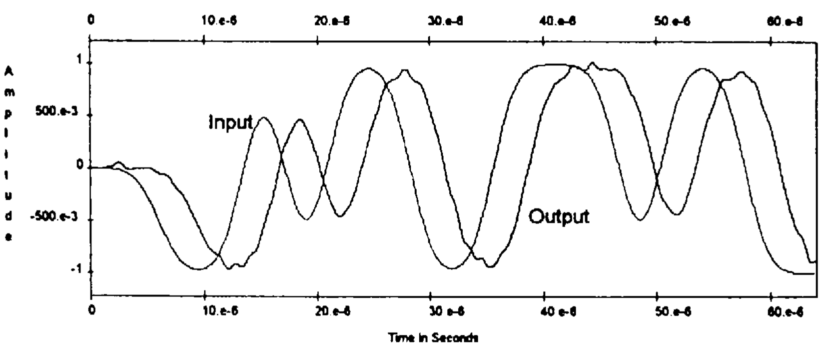


Figure A11b.12 Overlay of input and output signals.

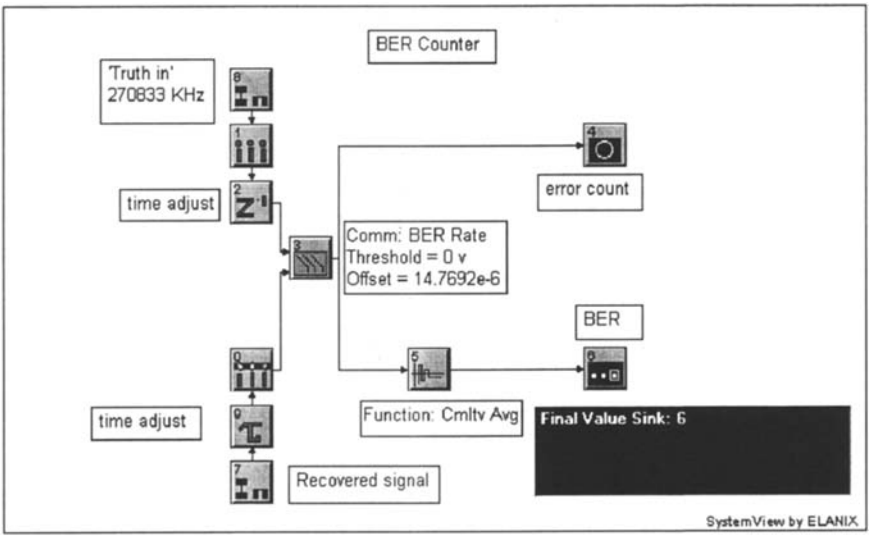


Figure A11b.13 BER calculator.

BER for this configuration is shown in Figure A11b.14. The BER is plotted against the received power as opposed to the usual E_b/N_0. The simulation can be set to adjust automatically the signal power plotted against the received power as opposed to the usual E_b/N_0 loop. The relation between the two parameters is

$$E_b/N_0 = P_r/FN_0R$$

$$E_b/N_0(\text{dB}) = P_r(\text{dBm}) - 5.4(\text{dB}) + 174(\text{dBm/Hz}) - 54(\text{dBHz})$$
$$= P_r + 114.6$$

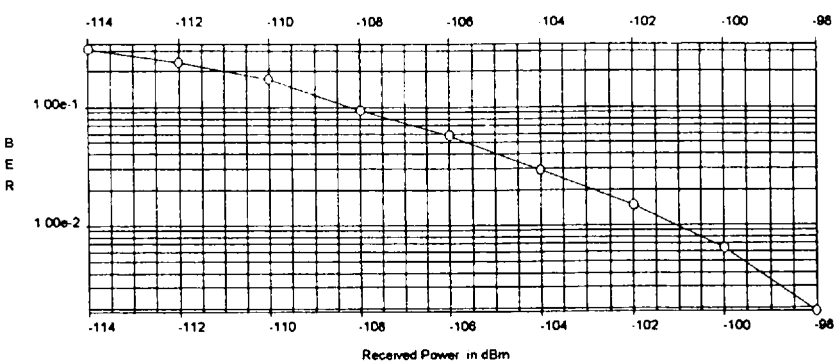


Figure A11b.14 BER vs. received signal power.

At each power level the loop runs long enough to detect a fixed number of errors. This optimizes the simulation processing time. The BER estimates of each loop have the same statistical significance.

For the purposes of obtaining the BER curve, the simulation block diagram in Figure A11b.1 was modified. The RF section and 1st IF mixer were eliminated. The VCO in the modulator/transmitter was changed to the 1st IF frequency of 71 MHz. The output of the channel section was sent to the modified 1st IF section. With these changes, the system master sample rate could be reduced from 4.096 GHz to 256 MHz, a savings of 16 in the run time. To compensate for the eliminated components, the noise figure of the individual components was disabled and the thermal noise was increased by the overall system noise figure of 5.4 dB, as discussed below. After the noise is inserted in the front, pure gain or loss elements do not effect the SNR. The wideband RF filter does not affect the final performance.

A11b.11 NOISE FIGURE

The noise figure of this receiver can be calculated by means of the analysis presented in chapters 1 and 5 of this book. The result of this calculation is NF = 5.4 dB. Figure A11b.15 shows a spectral overlay of the input noise source and a CW tone at 947.5 MHz. Figure A11b.16 shows a similar plot at the input to the quadrature detector after the 2nd IF section. The noise figure derived from this data is

$$\mathrm{NF} = 10\log_{10}\frac{(S/N)_{\mathrm{in}}}{(S/N)_{\mathrm{out}}} = -88 + 129 - 2 - 34 = 5 \text{ dB}$$

This now agrees with the theoretical value.

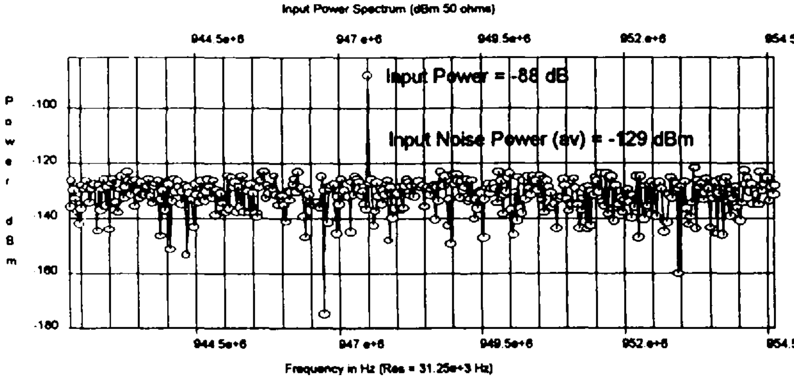


Figure A11b.15 PSD at 947.5 MHz input.

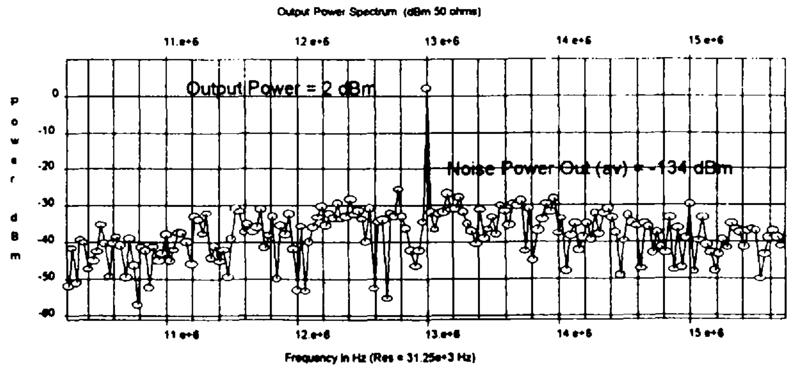


Figure A11b.16 PSD of signal at 13 MHz.

A11b.12 CONCLUSION

In this example, SystemView by ELANIX was used to simulate a complete GSM system, from bits in to bits out. The emphasis was on the receiver design and implementation. The parameters used were taken from commercially available components. All of the real-world effects, such as thermal noise and intermodulation products due to nonlinearities, are accurately taken into account.

This page intentionally left blank

Index