

PRINCIPLES OF GUIDED MISSILE DESIGN

WOODS HOLE
OCEANOGRAPHIC INSTITUTION

LABORATORY
BOOK COLLECTION



PURCHASE ORDER NO. 1625

D. O. KETCHUM 4/20/60



AIRBORNE
RADAR



Painting by Kirwan

THE AIR FORCE BOMARC INTERCEPTOR MISSILE

PRINCIPLES OF GUIDED MISSILE DESIGN

Editor of the Series

GRAYSON MERRILL, CAPTAIN, U.S.N. (Ret.)

*General Manager, Astrionics Division, Fairchild Engine
and Airplane Corporation*

GUIDANCE—by Arthur S. Locke and collaborators

AERODYNAMICS, PROPULSION, STRUCTURES AND
DESIGN PRACTICE—by E. A. Bonney, M. J. Zucrow,
and C. W. Besserer

OPERATIONS RESEARCH, ARMAMENT, LAUNCH-
ING—by G. Merrill, H. Goldberg, and R. H. Helmholtz

MISSILE ENGINEERING HANDBOOK—by C. W. Bes-
serer

DICTIONARY OF GUIDED MISSILES AND SPACE
FLIGHT—*Edited by Grayson Merrill*

SPACE FLIGHT—by K. A. Ehricke

Vol. I: ENVIRONMENT AND CELESTIAL
MECHANICS

Vol. II: DYNAMICS Vol. III: OPERATIONS

SYSTEMS PRELIMINARY DESIGN—by J. J. Jerger

AIRBORNE RADAR—by D. J. Povejsil, R. S. Raven, and
P. Waterman

RANGE TESTING—by J. J. Scavullo and Eric Burgess

FOUNDATIONS OF SEARCH THEORY—by N. S. Potter

PRINCIPLES OF INERTIAL NAVIGATION—by Richard
Parvin

PRINCIPLES OF GUIDED MISSILE DESIGN

Edited by

CAPTAIN GRAYSON MERRILL, U.S.N. (Ret.)

DONALD J. POVEJSIL

*Director, New Products Services;
Formerly Manager, Weapons Systems Engineering
Air Arm Division, Westinghouse Electric Corporation
Pittsburgh, Pa.*

ROBERT S. RAVEN

*Advisory Engineer, Weapons Systems Engineering
Westinghouse Electric Corporation
Air Arm Division, Baltimore, Md.*

PETER WATERMAN

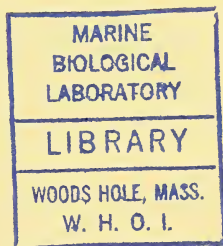
*Head, Naval Research Laboratories,
Radar Division, Washington, D. C.*



0

TL
694
P65

AIRBORNE RADAR



D. VAN NOSTRAND COMPANY, INC.
PRINCETON, NEW JERSEY · TORONTO · NEW YORK · LONDON

D. VAN NOSTRAND COMPANY, INC.
120 Alexander St., Princeton, New Jersey (*Principal office*)
24 West 40 Street, New York 18, New York

D. VAN NOSTRAND COMPANY, LTD.
358, Kensington High Street, London, W.14, England

D. VAN NOSTRAND COMPANY (Canada), LTD.
25 Hollinger Road, Toronto 16, Canada

COPYRIGHT © 1961, BY
D. VAN NOSTRAND COMPANY, INC.

Published simultaneously in Canada by
D. VAN NOSTRAND COMPANY (Canada), LTD.

Library of Congress Catalogue Card No. 61-8542

No reproduction in any form of this book, in whole or in part (except for brief quotation in critical articles or reviews), may be made without written authorization from the publishers.

CONTRIBUTORS

- P. J. ALLEN • *Naval Research Laboratory, Washington, D. C.*
(Coauthor, Chapter 10)
- G. S. AXELBY • *Westinghouse Electric Corp., Air Arm Division,
Baltimore, Md.*
(Coauthor, Chapters 8 and 9)
- B. L. CORDRY • *Bendix Aviation Corp., Radio Division, Baltimore,
Md.*
(Coauthor, Chapter 14)
- W. R. FRIED • *General Precision Laboratory, Pleasantville, N. Y.*
(Coauthor, Chapter 14)
- S. F. GEORGE • *Naval Research Laboratory, Washington, D. C.*
(Coauthor, Chapter 6)
- M. GOETZ • *Westinghouse Electric Corp., Central Laboratories,
Pittsburgh, Pa.*
(Coauthor, Chapter 13)
- D. J. HEALEY III • *Westinghouse Electric Corp., Air Arm Division,
Baltimore, Md.*
(Author, Chapter 7; Coauthor, Chapter 8)
- L. HOPKINS • *Raytheon Manufacturing Corp., Maynard Labora-
tory, Maynard, Massachusetts*
(Coauthor, Chapter 6)
- D. D. HOWARD • *Naval Research Laboratory, Washington, D. C.*
(Coauthor, Chapter 8)
- A. KAHN • *Westinghouse Electric Corp., Air Arm Division,
Baltimore, Md.*
(Coauthor, Chapter 12)
- M. KATZIN • *Electromagnetic Research Corp., Washington, D. C.*
(Author, Chapter 4)
- R. H. LAPRADE • *Westinghouse Electric Corp., Air Arm Division,
Baltimore, Md.*
(Coauthor, Chapter 14)
- T. MORENO • *Varian Associates, Palo Alto, California*
(Author, Chapter 11)
- R. M. PAGE • *Naval Research Laboratory, Washington, D. C.*
(Coauthor, Chapter 6)

- P. M. PAN • *Westinghouse Electric Corp., Air Arm Division, Baltimore, Md.*
(Coauthor, Chapter 10)
- D. J. POVEJSIL • *Westinghouse Electric Corp., New Products Services, Pittsburgh, Pa.*
(Editor and Coauthor, Chapters 1, 2, 6, and 12)
- R. S. RAVEN • *Westinghouse Electric Corp., Air Arm Division, Baltimore, Md.*
(Editor and Author, Chapters 3 and 5; Coauthor, Chapters 8 and 12)
- R. M. SANDO • *Westinghouse Electric Corp., Air Arm Division, Baltimore, Md.*
(Coauthor, Chapter 13)
- F. STAUFFER • *Westinghouse Electric Corp., Air Arm Division, Baltimore, Md.*
(Coauthor, Chapter 14)
- M. TAUBENSLAG • *Aeronica Manufacturing Corp., Aerospace Division, Baltimore, Md.*
(Coauthor, Chapter 12)
- J. W. TITUS • *Naval Research Laboratory, Washington, D. C.*
(Coauthor, Chapter 13)
- P. WATERMAN • *Naval Research Laboratory, Washington, D. C.*
(Editor and Coauthor, Chapters 1 and 2)
- M. S. WHEELER • *Westinghouse Electric Corp., Air Arm Division, Baltimore, Md.*
(Coauthor, Chapter 10)
- C. F. WHITE • *Naval Research Laboratory, Washington, D. C.*
(Coauthor, Chapters 8 and 9)
- H. YATES • *Barnes Engineering Co., Infrared Division, Stamford, Conn.*
(Coauthor, Chapter 6)

FOREWORD

"Airborne Radar" is the eighth volume in the series *Principles of Guided Missile Design*. Other titles in the series are *Guidance; Aerodynamics, Propulsion, Structures and Design Practice; Operations Research, Armament, Launching; Missile Engineering Handbook; Dictionary of Guided Missiles and Space Flight; Space Flight I—Environment and Celestial Mechanics; Space Flight II—Dynamics; Space Flight III—Operations; Preliminary System Design; and Range Testing*.

The purpose of the series as a whole is to give a basis for instruction to graduate students, professional engineers, and technical officers of the armed services so that they can become well grounded in the technology of guided missiles and space flight. This book concerns itself with one of the most important systems used now in missilery and to be used soon in space flight—airborne radar. As this is written the world is commencing a great search for a defense against intercontinental ballistic missiles; airborne radar will play a key role in this defense.

In the interests of brevity, this book presumes considerable knowledge on the reader's part—namely, knowledge of the basic principles of electronics and electromagnetic propagation and familiarity with the associated language. It also presumes a knowledge of the weapons systems employing airborne radars. Certain prior issues of the series will be found especially valuable as references; these are the *Dictionary of Guided Missiles and Space Flight; Guidance; and Operations Research, Armament, Launching*.

Criticisms and constructive suggestions are invited. With this aid and by keeping abreast of the state of the art we hope to make timely revisions to this volume.

Grateful acknowledgment is made to the many authors and publishers who kindly granted permission for the use of their material and to the Department of Defense, whose cooperation made possible a meaningful text without violation of security.

The opinions or assertions contained herein are the private ones of the authors and the editor and are not to be construed as official or reflecting the views of any government agency or department.

GRAYSON MERRILL
Editor

Wyandanch, Long Island, New York
November 1960

PREFACE

The basic purpose of this book is to present a balanced treatment of the airborne radar systems design problem. Primary emphasis is placed upon the interplay between radar techniques and components on one hand, and the types of weapons systems which employ airborne radars on the other. Radar design details have been eliminated for the most part except for illustrative examples which show how a design detail can exert an important influence on the operation of a complete weapons system.

Although the treatment is directed at airborne radars, this volume will be found extensively applicable to surface radars as well. Since the latter enjoy a less severe environment, especially with regard to relative target motion, stability of platform, and space and weight restrictions, the principles governing their design will implicitly be covered here.

Because this book attempts to bridge the gap between the abstractions of overall system design and the hard realities of hardware design, it contains material which will interest almost anyone involved in the study, design, or application of airborne radars. For example, although the receiver designer will not learn much that is new to him about circuitry design, he can learn a great deal about how the design of a receiver should be planned for optimum benefit to the overall system. Similarly, engineers and scientists charged with responsibility for monitoring the efforts of airborne radar subcontractors can find this book most useful in determining the type of direction they should give to subcontractors to ensure eventual compatibility of the airborne radar with the complete weapons system.

A particular effort has been made to present facts and combinations of facts which have not enjoyed prior publication in book form. This has been done at the expense of excluding a great deal of historical and background information already available in the printed literature.

This book possesses close ties with two previous volumes in the series: *GUIDANCE* by A. S. Locke et al. and *OPERATIONS RESEARCH, ARMAMENT, LAUNCHING* by Grayson Merrill, Harold Goldberg, and Robert H. Helmholtz. Radar techniques and problems are presented in greater detail than was possible in *GUIDANCE*; similarly the problem of translating operational studies into detailed airborne radar requirements is covered in greater detail than was possible in *OPERATIONS RESEARCH*. The basic theme of these earlier volumes—the importance of the systems approach—is continued in this volume.

Many of the authors have had previous association on team efforts aimed at the development and production of complex airborne radar equipments. The technical approaches presented thus represent tools forged on the anvil of experience—tools which have facilitated the solution of many difficult problems. It is the authors' hope that succeeding generations of system designers may use these tools to their advantage in designing the even more complex systems to come.

Acknowledgments

In addition to the authors, there are many individuals and organizations to whom acknowledgment must be made for an active part in the writing of this book. Members of the Naval Research Laboratories, the Naval Air Development Center, the Fairchild Astrionics Division and the Westinghouse Electric Corporation assisted with suggestions, criticisms, and technical readings. Deepest thanks must be extended to the government laboratories (NRL and WADC) and the corporations (Barnes Engineering, Bendix Aviation, General Precision Laboratory, Raytheon Manufacturing, Varian Associates, and Westinghouse Electric) who provided encouragement and assistance to the contributors.

The assistance of Mr. R. G. Clanton of Westinghouse was vital to the preparation of the examples employed in Chapter 2. In addition, Mr. Clanton's many helpful suggestions and detailed reviews of the remainder of this chapter are most gratefully acknowledged. Mr. R. H. Laprade had the responsibility of reviewing all the material relating to propagation in addition to his contribution to Chapter 14.

Mr. A. Stanley Higgins, Mr. Melvyn Goetz and Dr. J. F. Miner of Westinghouse rendered invaluable services in overseeing the myriad details involved in the editing and production of the final text. The Westinghouse Electric Corporation deserves special thanks for the assistance provided on drawings, typing, and the reproduction of the many drafts of the manuscripts.

D. J. POVEJSIL
R. S. RAVEN
P. WATERMAN

CONTENTS

LIST OF CONTRIBUTORS	vii
FOREWORD	ix
PREFACE	xi

1 ELEMENTS OF THE AIRBORNE RADAR SYSTEMS DESIGN PROBLEM

1-1 Introduction	1
1-2 Classifications of Radar Systems	2
1-3 Installation Environment	3
1-4 Functional Characteristics of Radar Systems	4
1-5 The Modulation of Radar Signals	16
1-6 Operating Carrier Frequency	26
1-7 The Airborne Radar Design Problem	27
1-8 The Systems Approach to Airborne Radar Design	30
1-9 Systems Environments	35
1-10 Weapons System Models	36
1-11 The Basic Statistical Character of Weapons System Models	39
1-12 Construction and Manipulation of Weapons System Models	41
1-13 Summary	44

2 THE DEVELOPMENT OF WEAPONS SYSTEM REQUIREMENTS

2-1 Introduction to the Problem	46
2-2 Formulating the System Study Plan	48
2-3 Aircraft Carrier Task Force Weapons System	50
2-4 The Target Complex	55
2-5 The Operational Requirement	57
2-6 The System Concept	57
2-7 The System Study Plan	59
2-8 Model Parameters	60
2-9 System Effectiveness Models	61
2-10 Preliminary Design of the Airborne Early Warning System	67
2-11 AEW System Logic and Fixed Elements	70
2-12 AEW Detection Range Requirements	73
2-13 AEW Target Resolution Requirements	75

2-14	Interrelations of the AEW System, the CIC System, the Interceptor System, and the Tactical Problem	79
2-15	Accuracy of the Provisional AEW System	80
2-16	Information-Handling Capacity of the Provisional AEW System	84
2-17	Velocity and Heading Estimates	85
2-18	AEW Radar Beamwidth as Dictated by the Tactical Problem	89
2-19	Factors Affecting Height-Finding Radar Requirements	92
2-20	Summary of AEW System Requirements	96
2-21	Evaluation of Tentative Design Parameters with Respect to the Tactical Problem	98
2-22	Interceptor System Study Model	100
2-23	Probability of Reliable Operation	103
2-24	Probability of Viewing Target—Vectoring Probability	104
2-25	Analysis of the Vectoring Phase of Interceptor System Operation	106
2-26	AI Radar Requirements Dictated by Vectoring Considerations	111
2-27	Analysis of the Conversion Problem	116
2-28	Lock-on Range and Look-Angle Requirements Dictated by the Conversion Problem	130
2-29	AI Radar Requirements Imposed by Missile Guidance Considerations	135
2-30	Summary of AI Requirements	136
2-31	Summary	137

3 THE CALCULATION OF RADAR DETECTION PROBABILITY AND ANGULAR RESOLUTION

3-1	General Remarks	138
3-2	The Radar Range Equation	138
3-3	The Calculation of Detection Probability for a Pulse Radar	141
3-4	The Effect of Scanning and the Cumulative Probability of Detection	156
3-5	The Calculation of Detection Probability for a Pulsed-Doppler Radar	162
3-6	Factors Affecting Angular Resolution	168

4 REFLECTION AND TRANSMISSION OF RADIO WAVES

4-1	Introduction	174
4-2	Reflection of Radar Waves	175
4-3	Effect of Polarization on Reflection	179
4-4	Modulation of Reflected Signal by Target Motion	180

4-5	Reflection of Plane Waves from the Ground	181
4-6	Effect of Earth's Curvature	190
4-7	Radar Cross Sections of Aircraft	192
4-8	Amplitude, Angle, and Range Noise	198
4-9	Prediction of Target Radar Characteristics	208
4-10	Sea Return	211
4-11	Sea Return in a Doppler System	217
4-12	Ground Return	219
4-13	Altitude Return	222
4-14	Solutions to the Clutter Problem	224
4-15	Attenuation in the Atmosphere	227
4-16	Attenuation and Back-scattering by Precipitation	230
4-17	Attenuation by Propellant Gases	231
4-18	Refraction Effects in the Atmosphere	233

5 TECHNIQUES FOR SIGNAL AND NOISE ANALYSIS

5-1	Introduction	238
5-2	Fourier Analysis	238
5-3	Impulse Functions	243
5-4	Random Noise Processes	245
5-5	The Power Density Spectrum	248
5-6	Nonlinear and Time-Dependent Operations	253
5-7	Narrow Band Noise	258
5-8	An Application to the Evaluation of Angle Tracking Noise	264
5-9	An Application to the Analysis of an MTI System	269
5-10	An Application to the Analysis of a Matched Filter Radar	272
5-11	Application to the Determination of a Signal's Time of Arrival	281

6 GENERIC TYPES OF RADAR SYSTEMS AND TECHNIQUES

6-1	Introduction	292
6-2	Basic Principles	293
6-3	Monopulse Angle Tracking Techniques	300
6-4	Correlation and Storage Radar Techniques	305
6-5	FM/CW Radar Systems	311
6-6	Pulsed-Doppler Radar Systems	320
6-7	High-Resolution Radar Systems	333
6-8	Infrared Systems	338

7 THE RADAR RECEIVER

7-1	General Design Principles	347
7-2	The Interdependence of Receiver Components	352

7-3	Receiver Noise Figure	353
7-4	Low-Noise Figure Devices for RF Amplification	356
7-5	Mixers	357
7-6	Coupling to the Mixer	361
7-7	IF Amplifier Design	362
7-8	Considerations of IF Preamplifier Design	368
7-9	Overall Amplifier Gain	373
7-10	Gain Variation and Gain Setting	375
7-11	Bandwidth and Dynamic Response	375
7-12	Sneak Circuits	377
7-13	Considerations Relating to AGC Design	379
7-14	Problems at High-Input Power Levels	380
7-15	The Second Detector (Envelope Detector)	382
7-16	Gating Circuits	386
7-17	Pulse Stretching	387
7-18	Connecting the Receiver to the Related Regulating and Tracking Circuits	388
7-19	Angle Demodulation	389
7-20	Some Problems in the Measurement of Receiver Characteristics	390

8 REGULATORY CIRCUITS

8-1	The Need for Regulatory Circuits	394
8-2	External and Internal Noise Inputs to the Radar System	395
8-3	Automatic Frequency Control	401
8-4	Variation of Transmitter Frequency with Environmental Conditions	402
8-5	Magnetron Pulling	403
8-6	Static and Dynamic Accuracy Requirements	405
8-7	Continuous-Correction AFC	407
8-8	Limit-Activated AFC	412
8-9	The Influences of Local Oscillator Characteristics	413
8-10	Relation to Receiver IF Characteristics	414
8-11	Discriminator Design	414
8-12	Instantaneous AFC	415
8-13	Problems of Frequency Search and Acquisition	416
8-14	Automatic Gain Control	416
8-15	Linear Analysis of AGC Loops	419
8-16	Static Regulation Requirements of AGC Loops	420
8-17	Dynamic Regulation Requirements of AGC Loops	422
8-18	AGC Transfer Characteristic Design Considerations	423
8-19	The Modulation Transmission Requirement	424
8-20	Design of an AGC Transfer Function	425

8-21	The IF Amplifier Control Characteristic	427
8-22	The Angle Measurement Stabilization Problem	429
8-23	AI Radar Angle Stabilization	433
8-24	Aircraft Motions	433
8-25	Stabilization Requirements	439
8-26	Search Pattern Stabilization	440
8-27	Search Stabilization Equations	440
8-28	Static and Dynamic Control Loop Errors	442
8-29	Search Loop Mechanization	448
8-30	Stabilization During Track	452
8-31	Possible System Configurations	453
8-32	Accuracy Requirements on the Angle Track Stabilization Loop	457
8-33	Dynamic Stability Requirements on Angle Track Stabilization	464
8-34	Stabilization Loop Mechanization	468

9 AUTOMATIC TRACKING CIRCUITS

9-1	General Problems of Automatic Tracking	471
9-2	Automatic Angle Tracking	474
9-3	External Inputs: Undesired and Desired	475
9-4	Requirements in Angle Tracking Accuracy	479
9-5	Angle Tracking System Organization	480
9-6	Tracking Loop Design	485
9-7	Angle Tracking Loop Rate Errors	486
9-8	Angle Tracking Loop Position Errors	489
9-9	Angle Tracking Loop Mechanization	492
9-10	Introduction to Range and Velocity Tracking.	498
9-11	Automatic Range Tracking	498
9-12	Servo System Transfer Function Relationship to Input Time Function for a Range Tracking System	502
9-13	Range Tracking Design Example	505
9-14	Practical Design Considerations.	508

10 ANTENNAS AND RF COMPONENTS

10-1	Antennas: Introduction to Radar Antennas	512
10-2	Some Fundamental Concepts Useful in the Development of Radar Antenna Requirements	513
10-3	The Paraboloidal Reflector as a Radar Tracking Antenna	515
10-4	System Requirements for Radar Antennas	518
10-5	Pattern Simulation as a Link Between System Requirements and Antenna Characteristics.	520

10-6	Several Anomalous Effects in Antennas for Tracking Systems	523
10-7	The Linear Array as a Fan Beam Antenna for Surveillance	524
10-8	Two-Arm Spiral Antennas	528
10-9	Radomes	531
10-10	Introduction to Transmission Lines and Modes of Propagation	535
10-11	Types of Transmission Lines and Modes of Propagation	536
10-12	Standing Waves and Impedance Matching	540
10-13	Broadband System Design	543
10-14	Pressurization	545
10-15	Miscellaneous Microwave Components	546
10-16	Microwave Ferrite Devices and Their Application	557
10-17	Microwave Dielectric, Magnetic, and Absorbent Materials	565
10-18	The Duplexing Problem	566
10-19	Duplexing Schemes	567
10-20	Special Problems of Coherent Systems	573
10-21	Solid-State Amplifiers.	574
11 THE GENERATION OF MICROWAVE POWER		
11-1	The Magnetron	580
11-2	The Klystron	590
11-3	Traveling Wave Tubes for High Power	597
11-4	Modulation Techniques for Beam-Type Amplifiers	599
11-5	A Typical Radar System Employing a High-Gain Amplifier	601
11-6	Backward Wave Oscillators—Carcinotrons	602
11-7	The Platinotron	603
12 DISPLAY SYSTEM DESIGN PROBLEMS		
12-1	Introduction	607
12-2	Uses of Display Information	608
12-3	Types of Displays	613
12-4	Types of Input Information	619
12-5	The Cathode Ray Tube	621
12-6	Important Characteristics of Electrical-to-Light Transducers	627
12-7	Important Characteristics of the Human Operator	634
12-8	Development of Requirements for a Display System.	651
12-9	Special Display Devices	655
12-10	Special Displays	673
13 MECHANICAL DESIGN AND PACKAGING		
13-1	The Influence of Environment on Design	680
13-2	Military Specifications	682
13-3	Temperature	683

13-4	Solar Radiation	692
13-5	Nuclear Radiation	692
13-6	Vibration and Shock	694
13-7	Acoustic Noise	704
13-8	Acceleration	707
13-9	Moisture	708
13-10	Static Electricity and Explosion	710
13-11	Pressure	711
13-12	Maintenance and Installation	712
13-13	Transportation and Supply	714
13-14	Potential Growth	715
13-15	Reliability	715

14 AIRBORNE NAVIGATION AND GROUND SURVEILLANCE RADAR SYSTEMS

14-1	Introduction to Doppler Navigation Systems	726
14-2	Basic Principles of Doppler Radar Navigation	728
14-3	System Considerations	733
14-4	Major Characteristics and Components of a Doppler Radar Navigation System	736
14-5	Doppler Navigation System Errors Caused by Interactions with the Ground and Water	746
14-6	Modifying the Radar Range Equation for the Doppler Navi- gation Problem	749
14-7	Low Altitude Performance and the "Altitude Hole"	752
14-8	Doppler Navigation System Performance Data	755
14-9	Introduction to Weather Radar	759
14-10	Meteorological Effects at Microwave Frequencies	760
14-11	Designing Airborne Radar Systems Explicitly for Weather Mapping	764
14-12	Modifying the Radar Range Equation for the Weather Problem	764
14-13	Relative Importance of Design Variables in Airborne Weather Radar	766
14-14	Design Features	769
14-15	Introduction to Active Airborne Ground Mapping Systems	772
14-16	Basic Principles	772
14-17	System Considerations	774
14-18	Major Characteristics and Components	778
14-19	Modifying the Radar Range Equation for the Active Ground Mapping Problem	782
14-20	Resolution Limits in Ground Mapping Systems	784

14-21	Future Possibilities in Airborne Active Ground Mapping Systems	787
14-22	Introduction to Infrared Reconnaissance	787
14-23	Basic Principles Concerning IR Ground Mapping	788
14-24	System Considerations	791
14-25	Major Systems Features	798
14-26	New Developments	801
	INDEX	805

CHAPTER 1

ELEMENTS OF THE AIRBORNE RADAR
SYSTEMS DESIGN PROBLEM

1-1 INTRODUCTION

This book presents, and illustrates by examples, the basic information and procedural techniques required to plan and execute the design of an integrated airborne radar system. Basically, this design problem has three parts: (1) the development of radar system performance requirements based on the operational requirements of the overall weapons system; (2) the development and application of specific radar techniques that will meet the performance requirements within the limitations imposed by laws of nature and the state of the art; (3) the evaluation of the proposed radar system to determine whether or not it meets the requirements of the overall weapons system.

In each part of the design problem, the systems concept is employed; i.e., the airborne radar system is viewed as an integral part of a complete weapons system rather than as a separate entity.

The systems concept will be developed by the case study method. A hypothetical weapons system model will be constructed. This model will then be analyzed in relation to the operational requirements in order to derive the specific characteristics of the various system environments which have an important bearing on the airborne radar system design. Those areas which sensitively affect the overall system capability will then be developed.

By using the derived characteristics, it will be shown how airborne radar systems may be selected and designed to fulfill the overall system requirements and be compatible with the system environments.

As an example, the air defense system of a naval carrier task force will be considered. Two types of airborne radar systems are included in this weapons system.

1. An airborne early warning (AEW) system for alerting the air defense of a fast carrier task force.
2. An interceptor defense system, utilizing the primary information generated by the AEW system.

The analysis of these examples presents a basic method of approach which involves the concept of balancing various system elements — a procedure that can and should be used in the design of any airborne radar system.

Perhaps the most important concept that must be grasped by the radar designer is this: A radar is usually a small but vitally important part of a *dynamic* system, i.e., a system whose basic characteristics and parameters are constantly changing functions of time. Because of its role as the “eyes” of the system, the dynamic performance of the radar must be related to — and to a large extent, governed by — the dynamic performance required of the entire system. For this reason, the radar designer must possess the capability for understanding and analyzing the overall weapons system in addition to his specialized knowledge of the details of radar systems analysis and design.

1-2 CLASSIFICATIONS OF RADAR SYSTEMS

In order to provide background for the discussion of the systems aspects of airborne radar design, the basic characteristics and uses of radar systems are described. Many of the descriptive terms commonly used in radar system technology are defined. Some of the simpler mathematical expressions that arise in radar work are presented.

Radar is a word derived from the function performed by early radar systems — *R*ADIO *D*ETECTION *A*ND *R*ANGING. The word was meant to denote systems that transmitted and received radio signals. Today the meaning of the word has been extended to include a wide variety of systems that employ microwave techniques. It encompasses systems using received energy originating in the system (active systems), systems using received energy originating at the target (passive systems), systems using received energy originating at a transmitter separate from the receiver or target (semiactive systems), and systems emitting electromagnetic radiations for various purposes (transmitting or illuminating systems). Many complex weapon systems include combinations of these basic types. For example, an electronic countermeasures system may be composed of a passive radar system that detects the presence of hostile electromagnetic radiation and utilizes this intelligence to control the action of a jamming system to combat the enemy radiation.

The profusion of radar systems in use today requires that some logical means of classification be employed. One such means that has achieved general acceptance classifies a radar system according to the four characteristics:

1. Installation environment
2. Function(s)

3. Types of modulation intelligence carried on the transmitted and received radiations and the types of demodulation processes used to extract information from the received signals
4. Operating carrier frequency

Reference to these four characteristics is usually made in any general qualitative description of a radar — e.g. an (1) *airborne* (2) *intercept search and track* (3) *conical-scan pulse radar* (4) operating at *X Band*.

1-3 INSTALLATION ENVIRONMENT

The most common types of radar system installations are:

1. Ground-based
2. Ship-based
3. Airborne (piloted aircraft)
4. Airborne (missile)

Procurement agencies, in general, have been divided into groups according to installation environment in order to simplify their diversity of interest. Such a division facilitates the proper treatment of the complex problems associated with the development and design of a radar system for a particular installation environment, but does not always provide the cross fertilization of experience needed to take advantage of progress in any one particular line.

1-4 FUNCTIONAL CHARACTERISTICS OF RADAR SYSTEMS

Some basic functions which may be performed by radar systems are:

1. Search and detection
2. Identification
3. Tracking
4. Mapping
5. Navigation
6. Communication
7. Radiation detection (Ferret)
8. Illumination
9. Information relay
10. Jamming
11. Scientific research (e.g. radio astronomy)

A given radar system may perform only one of these functions. More frequently it will perform two or more. Multimode operation is particularly characteristic of airborne radar systems where space, size, and weight limitations dictate that maximum capability and flexibility be obtained from each pound of radar equipment.

The specification of the functions that must be performed by the radar systems equipments is a major product of the system study that must precede equipment design. This system study must also produce quantitative performance goals for each of the required functions.

In cases where multimode operation is required, the system study must set up a definitive specification of *primary* and *secondary* modes. This

definition can serve as the basis for arbitrating conflicting design requirements resulting from the multimode requirement.

The specific functions performed by a radar system are outlined below in somewhat more detail.

Search and Detection. An important function of a radar system is to interrogate a given volume of space for the presence (or absence) of a target of tactical interest. One very common method by which a radar system may be used to perform this function is shown in Fig. 1-1. In this

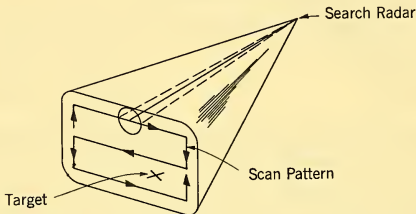


FIG. 1-1 Radar Search and Detection.

example, RF (radio-frequency) energy is generated in the radar system (active system). This energy is focused into a highly directional beam by an antenna and propagated through space. Should there be an object of appropriate characteristics within the radar beam, a portion of the electromagnetic energy impinging on the object will be scattered away from it. A portion of this scattered energy finds its way back to the point of transmission where it may be detected by a receiver.

In order to extend the *space coverage* of the radar system, it is customary to *scan* a predetermined volume of space in a cyclic manner by changing the direction of propagation as indicated in Fig. 1-1.

Identification. The system may be required to operate in an area where both friendly and unfriendly aircraft or targets possibly exist. A requirement will then arise to search the area and identify any targets as friend or foe (IFF). When it is performing the search and detection function, the radar system generates answers to a specific question: Is there — or is there not — a target of tactical interest within a given volume of space? The basic characteristics of a radar — or any detection device — are such that both correct and incorrect answers to this question may be generated. There are, in fact, four possible sets of circumstances:

1. There is a target within the searched volume and its presence is detected by the radar.

2. There is a target within the searched volume, but for one reason or another its presence is not detected by the radar.
3. There is, in fact, no target within the searched volume and none is indicated by the radar.
4. There is no target within the searched volume; however, the presence of a target is indicated by the radar.

In cases (1) and (3) the radar provides the proper answer to the question. In case (2) the radar *fails* to provide the proper answer by failing to provide any information whatsoever. In case (4) the radar provides the wrong answer by providing *spurious* information.

The manner in which the identification function is performed varies widely according to the type of radar and the tactical use to which it is put. In some cases, the detection and identification functions may be combined by a logical nonmechanical process which uses a suitable choice of a detection criterion and a prior knowledge of the probable target characteristics. For example, in the search and detection system, Fig. 1-1, one might specify that the appearance of a target indication on each of three successive scan cycles constitutes a detection — the assumption being that it is not likely that a spurious indication would be repeated on three successive scan cycles. One might further stipulate that any target thus detected shall be considered an *enemy* target if it is approaching at predetermined altitudes, speeds, or courses.

The identification function is sometimes performed by a completely separate radar system designed specifically to accomplish some part of the identification problem. Many forms of identification-friend-or-foe (IFF) systems fall into this category: e.g., in Fig. 1-2 a presumably friendly

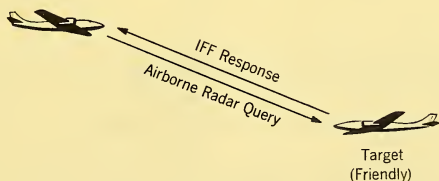


FIG. 1-2 IFF System.

aircraft is equipped with a passive receiver that detects the search radar signals. These signals are used to initiate transmission of a coded signal back to the search radar location. This coded signal is correlated with the search radar target return signal to establish and define the presence of a "friendly" aircraft.

The foregoing discussion of the search, detection, and identification functions points out an important characteristic that affects the performance of these functions. This characteristic is the implied uncertainty that the desired result will be obtained in a given case. The element of uncertainty requires a statistical approach to the problem of understanding and analyzing the detection and identification characteristics of a radar system.

Tracking. A radar system may be designed with the capability of measuring the relative range, range rate, and bearing of any object which scatters microwave energy impinging on it. When a radar makes any or all of these measurements on a more or less continuous basis (depending upon whether it is also searching), it is said to be *tracking* the target. The tracking function can provide information for:

1. A continuous display or record of relative *target position* as a function of time
2. Calculation of relative *target motion*
3. Prediction of *future relative target position*

The range measurement is achieved by measuring the elapsed time between a transmitted signal and the reception of the portion of the transmitted energy that is scattered by the target back along the direction of transmission and multiplying it by a constant representative of the average propagation velocity. The radar energy is propagated at the speed of light ($c = 328 \text{ yd}/\mu\text{sec}$). Thus the time required for the radar energy to travel from the transmitter to the target and back to the transmitter location is

$$t = \frac{2R}{c} \quad (1-1)$$

The range to the target may be expressed

$$R = \frac{ct}{2} = 164t \text{ yards} \quad (1-2)$$

where R = range to target in yards

t = time in microseconds between transmission and reception

c = propagation velocity in yards per microsecond.

The closing velocity along a line from the radar to the target (range rate) can be measured by means of the frequency difference between the transmitted and received signals caused by the relative target motion. This doppler effect will be discussed in Paragraph 1-5.

Angular bearing of the target is measured by utilizing a directive beam like that shown in Fig. 1-1. With this arrangement a target return is obtained only when the beam is pointed in the direction of the target. Thus by measuring the angular position of the beam with respect to some reference axis when a target return is present, a measure of relative target bearing from the radar system is obtained. The accuracy of this measurement depends to a large extent on the parameters associated with the detailed design. The nature of this dependence and the means that may be used to improve the accuracy of angular measurement will be developed in later portions of this book.

Target motion relative to the tracking radar platform may be computed with measured range information and the time derivatives of the measured range and angle information. Analysis of the two-dimensional case displayed in Fig. 1-3 illustrates the basic principle.

The relative velocity of the target, V_{TR} can be represented by two components — one parallel to the line-of-sight, V_{TRp} , and the other normal to the line-of-sight, V_{TRn} . These quantities may in turn be expressed

$$V_{TRp} = \frac{d}{dt}(R) = \dot{R} \quad (1-3)$$

$$V_{TRn} = R\dot{\phi} \quad (1-4)$$

where \dot{R} = range rate along the line-of-sight and $\dot{\phi}$ = space angular rotation of the line-of-sight.

Range-rate information can be obtained by differentiation of the radar range measured. It can also be measured directly by doppler frequency shift as previously indicated.

Commonly, the space angular rate of the line-of-sight is measured by an angular-rate gyroscope mounted on the antenna of a tracking radar.

The relative target velocity information may be utilized in several ways. For example, this information coupled with a knowledge of the tactical situation can provide a means for *identifying* targets of tactical interest. In addition, the computation of the components of relative target velocity makes it possible to *predict* the future target position relative to the radar

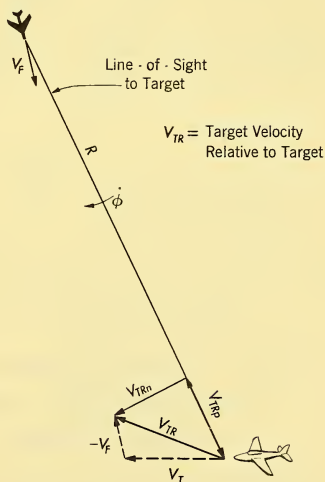


FIG. 1-3 Relative Target Motion: Two-Dimensional Case.

platform. This capability is essential to the solution of the fire-control problem.

An analysis of the two-dimensional fire-control problem (Fig. 1-4) illustrates the basic principles. If an aircraft is armed with a weapon which is fired along the aircraft flight line, the weapon position relative to the interceptor t_f seconds after firing can be expressed

$$R_{fw} = V_0 t_f \tag{1-5}$$

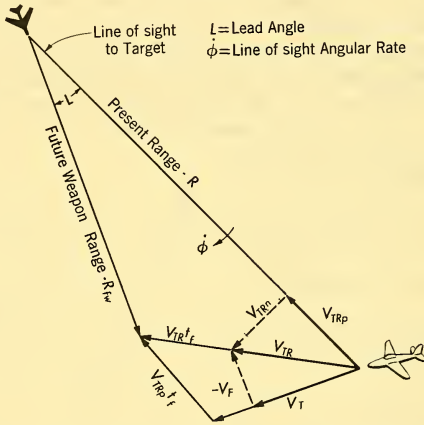


FIG. 1-4 Air-to-Air Fighter-Bomber Duel Fire-Control Problem in Two-Dimensional Coordinates Relative to the Weapon Firing Aircraft.

where R_{fw} = future relative weapon position

V_0 = average velocity of weapon relative to fighter velocity

t_f = weapon time of flight (i.e. time elapsed after weapon firing).

The fire-control problem is solved when the future relative range of the weapon coincides with the future relative range of the target R_{ft} , i.e. $R_{fw} = R_{ft}$ (at some value of t_f).

The predicted future relative target range may be expressed in terms of its components relative to the line of sight

$$R_{ftp} = R - V_{TR} t_f = R - \dot{R} t_f \tag{1-6}$$

$$R_{ftn} = V_{TRn} t_f = R \dot{\phi} t_f. \tag{1-7}$$

Similarly, relative weapon range may be expressed

$$R_{fwp} = V_{of} \cos L \quad (1-8)$$

$$R_{fwn} = V_{of} \sin L \quad (1-9)$$

Equating components, we obtain the basic fire-control equations

$$V_{of} \cos L = R - \dot{R}t_f \quad (\text{time of flight equation}) \quad (1-10)$$

$$\sin L = \frac{R\dot{\phi}}{V_0} \quad (\text{lead angle equation}). \quad (1-11)$$

Mapping. The microwave energy scattering characteristics of physical objects provide a wide range of characteristic returns. The differences between these returns make it possible to use a radar system to obtain a map of a given area and permit the interpretation of the results through an understanding of the characteristic returns. The mapping function is accomplished by "painting" (scanning) a designated area with a radar beam of appropriate characteristics. Two common means for performing this function are shown in Fig. 1-5.

In the first method, Fig. 1-5A, the picture is "painted" by rotating the antenna beam around an axis perpendicular to the area to be mapped. The resulting picture is a circular map whose center, disregarding trans-

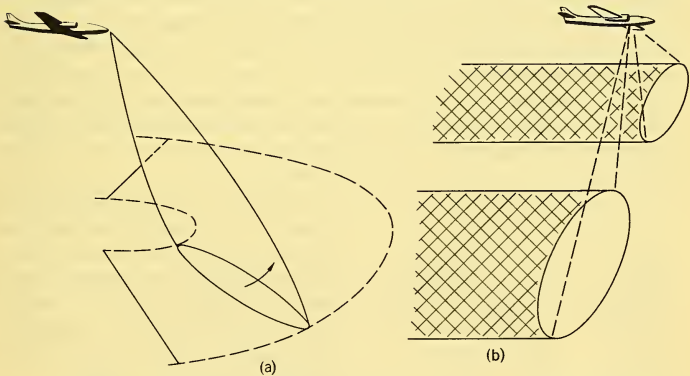


FIG. 1-5 Radar Mapping: (a) Forward-Look System, Variant of the Plan Position System. (b) Side-Look System.

lational motion, is the radar's position. The coordinates of the display are conveniently in terms of angle and range. The title "Plan Position" is applied to this type of map. A variant of this scheme would be a system that mapped only a sector of the circle — for example, a sector just forward of the radar aircraft (Forward-Look System).

In the second method, Fig. 1-5B, fixed antennas are mounted on each side of the aircraft. The motion of the aircraft with respect to the ground provides the scanning means. Thus the picture obtained by this radar is a continuous map of two strips on either side of the aircraft flight path.

In each case the detail is very different than that obtainable from photographs of the same terrain under conditions of good visibility. Nevertheless, a considerable amount of potentially useful tactical information can be obtained from such pictures. The distinction between land and water areas is particularly striking, and prominent targets — large ships, airfields, and cities — can also be clearly distinguished.

The basic capabilities of radar provide several attractive features in the performance of the mapping function. The range to the target is directly measurable. Smoke, haze, darkness, clouds, and rain do not prohibit taking useful radar pictures (depending on the radar parameters chosen). A camouflaged target that might be exceedingly difficult to distinguish by visual means is often readily unmasked by a radar picture. Finally, a radar picture does not necessarily have the same problems of perspective that tend to distort a visual picture.

The change of target characteristics with frequency can be employed to provide increased contrast. The basic principle is illustrated in Fig. 1-6, which shows hypothetical backscattering curves for the sea and a target. If the mapping is performed at two frequencies, f_1 and f_2 , and if the returns at these frequencies are transformed into green and blue, respectively, on a visual display, then the target will appear green and the sea blue. This color transposition utilizes the human eye's ability to discern color differences (see Paragraph 12-7), thereby improving the contrast in cases where a relationship similar to Fig. 1-6 exists.

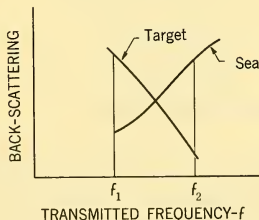


FIG. 1-6 Utilizing the Change of Target Characteristics with Frequency to Enhance Mapping.

By the use of the doppler (velocity discrimination) effect, a mapping system may also be provided with the capability for distinguishing moving targets that have a component of velocity along the sight-line of the radar. This is known as *moving target indication* (MTI).

Another type of radar mapping does not involve the generation and transmission of microwave energy by the radar. Rather, it utilizes the fact that all bodies — as a consequence of their temperature and emissivity characteristics — emit energy in the microwave spectrum. By using highly directional antenna and a receiver that is sensitive to these radi-

tions, a given area may be mapped by scanning the area and correlating the signals received with the antenna position. This method — often referred to as microwave thermal mapping (MTR) — is similar in concept to the various forms of infrared mapping. The only difference is the frequency spectrum covered. The use of microwave frequencies sometimes alleviates the severe weather limitations of the much higher-frequency infrared spectrum. Counterbalancing this advantage is the inherently poorer resolution obtained at microwave frequencies and the vastly smaller amounts of thermal radiation energy at these lower frequencies.

Navigation. The mapping capability can be used to perform a portion of the navigation function, particularly under conditions of poor visibility. Prominent land masses, land-water boundaries, and objects located in a relatively featureless background such as an aircraft carrier at sea are usually readily distinguishable — even on a radar picture obtained from a radar system not specially designed to perform the mapping function.

By a proper choice of radar parameters, cloud formations that represent a potential flight hazard can readily be detected by a radar of appropriate design. Radar systems designed specially to perform this function have become standard equipment on many transport and military aircraft. A typical radar picture obtained from such a system is shown in Fig. 14-15. Information such as this represents a valuable navigational aid. It can permit the successful completion of many missions that might otherwise be aborted because of weather uncertainty. Radars designed for other purposes can provide this information as an auxiliary function.

Another radar navigational aid is the radar beacon system (Fig. 1-7). In this system an airborne radar transmits microwave energy at a specified beacon frequency. When some of the energy is received by a beacon station tuned to this frequency, this energy is, in effect, amplified greatly and transmitted back to the interrogating aircraft. There is preset, fixed time delay t_b between the reception and the transmission in the beacon. Thus if the total time between interrogation of the beacon and the reception of the beacon reply is t_i μ sec, the range to the beacon is

$$R = \frac{c}{2} (t_i - t_b)$$

or

$$R = 164(t_i - t_b) \text{ yards} \quad (1-12)$$

where $c/2 = \frac{1}{2}$ propagation speed of light in yd/ μ sec

t_i = propagation transit time

t_b = beacon delay time.

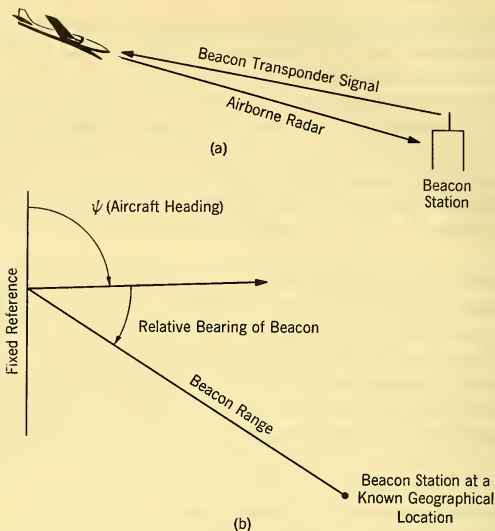


FIG. 1-7 Radar Beacon System.

The angular position of the beacon relative to the aircraft is measured by the airborne radar. Since the pilot knows his own heading in space and the geographical position of the beacon, the knowledge of relative range and bearing of the beacon permits him to determine his own geographical location.

It is quite common for an airborne radar to have a *beacon mode* as an auxiliary function. Despite the apparent simplicity of the mode, the proper integration of this function into an airborne radar system is often difficult, particularly if early systems planning neglects to include the cooperative beacon itself. Variations of the beacon mode of operation are also quite common in guided missile applications.

An airborne radar possesses an inherent capability for providing still another type of navigational information — true ground speed — achieved through the use of the doppler effect mentioned above in the discussion of the tracking function. This application will be discussed in detail in Chapter 6.

Communications. The transmitted radar signal may also be used as a carrier for the transmission of communications intelligence. While such transmission is limited essentially to line-of-sight because of the inherent

nature of microwave propagation (see Chapter 4), it has a number of potential advantages: (1) high directivity, increasing the security of the communications link; (2) dual utilization of the same antenna and carrier power source; and (3) relative predictability of the transmission characteristics.

Radiation Detection. The radiation detection or passive listening function that may be performed by a radar system has already been mentioned in the preceding discussions of IFF, ECM, beacon, and communications systems.

A passive radar system consists of only a receiving channel or channels designed to detect and — in some applications — to track microwave energy that is emitted or scattered by a separate source. Passive radars cannot measure range without auxiliary devices.

There is a variety of means for obtaining range measurements from a passive system — e.g., triangulation using several passive tracking systems at different locations; but all these methods are complicated and inaccurate when compared with the convenience of range measurement in an active radar system.

Several important functions may be performed by passive radar systems in addition to those already discussed. In the *Ferret* application, radar receivers tuned to cover a wide band of frequencies are used to detect enemy radiations, thereby providing intelligence data on the characteristics and capabilities of enemy radar systems. Such information is of great value in determining the tactics and countermeasures to be employed in subsequent operations.

A variation of the above application is one in which the enemy radiation is used as a source upon which a guided missile homes — a system known colloquially as a “radar buster.” Despite their simplicity of concept, such systems may present formidable systems engineering and design problems. The multiplicity of enemy signal sources, the intermittency of transmission from a scanning source, and the importance of having a stand-by mode of operation in the event that the enemy ceases to radiate for extensive periods of time, all contribute to the difficulties.

A special case of the “radar buster” passive radar homing system is the “home-on-jam” system. This system might be used as an alternative mode of operation for an active radar system. When the active radar is jammed, the jamming source could be detected and tracked by the passive system.

A passive radar system also forms a vital part of a *semiactive* guidance system. This application is discussed later in this chapter.

Illumination. A common form of radar system is the semiactive system. The functional operation of such a system is shown in Fig. 1-8

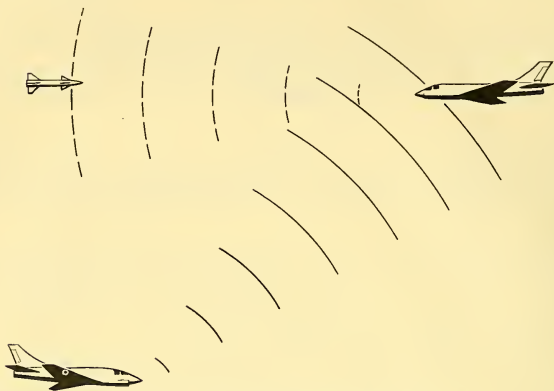


FIG. 1-8 Semiactive Guidance System.

and is described elsewhere in greater detail.¹ In this system, the target is illuminated by a source of microwave energy. A portion of this energy is scattered by the target and may be detected and tracked by a passive receiver located at some distance from the transmitting source.

Semiactive systems find their greatest use in guided missile systems, where it is often desirable to retain the basic advantages of an active system without incurring the weight penalty and transmitting antenna size restrictions that would result from placement of the transmitter in the guided missile.

It is possible to obtain a crude measurement of range in a semiactive system if the missile is illuminated by the same energy transmission as the target. The accuracy of this range measurement is greatest when the illuminator, missile, and target are in line as shown in Fig. 1-9. In this

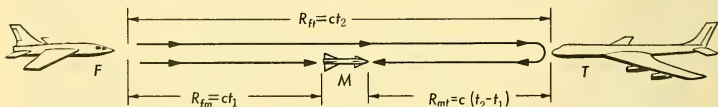


FIG. 1-9 Range Measurement in a Semiactive system.

case, the target receives energy from the interceptor-borne radar t_2 μ sec following transmission. The illuminating energy is also received directly

¹A. S. Locke, *Guidance* (Principles of Guided Missile Design Series), D. Van Nostrand Co., Princeton, N. J., 1955.

by a rearward-looking antenna on the missile t_1 μ sec after transmission. The missile, by measuring the time difference between these two signals, can obtain the range to the target; thus

$$R_{ft} = ct_2$$

$$R_{fm} = ct_1.$$

Since

$$R_{mt} = R_{ft} - R_{fm}$$

then

$$R_{mt} = c(t_2 - t_1) \quad (1-13)$$

where c = speed of propagation in $\text{yd}/\mu\text{sec} = 328 \text{ yd}/\mu\text{sec}$.

The relative velocity between the missile and the target can be obtained by analogous means, using the frequency difference between the direct and reflected signals. This frequency difference is caused by the doppler effect.

Information Relay. From a systems standpoint, it is often desirable to display and utilize radar information at a different location from the point of collection of the information. Typical of such an application is the air surveillance system shown in Fig. 2-15. Data are collected by a number of airborne early warning (search radar) systems located in such a manner as to provide the required coverage. It is desirable to assemble, correlate, and assess the data at a central location (Fleet Center) in order to provide a complete picture of the tactical situation. From this analysis, instructions and data can be relayed to the operating elements. This type of operation is typical of airborne, ground, or ship-based combat information centers (CIC).

Jamming. Radars may also be used to transmit microwave energy with the object of confusing or obscuring the information that other radars are attempting to gather. Jamming is of two fundamental types: (1) "brute force" and (2) deceptive.

Brute force jamming attempts to obscure as completely as possible the information contained in other radar signals by overpowering these signals.

Deceptive jamming, on the other hand, endeavors to create mutations in the information contained in other radar signals to render them less useful tactically.

Both types of jamming are aided by their one-way transmission characteristic as contrasted with the two-way transmission characteristic of active radar. This feature allows a jammer to operate successfully with a few watts of transmitted power against a radar transmitting hundreds of thousands of watts of peak power.

Despite this formidable advantage, the design of a jamming radar system can be one of the most perplexing of all radar systems problems — from the points of view of both systems engineering and hardware design. This arises from the vast multitude of possibilities with which a jamming system must cope.

Scientific Research. Airborne radars are frequently utilized to gather basic scientific data such as atmospheric transmission characteristics, target reflectivity, and ground reflectivity and emission characteristics.

The coming space age opens up several interesting possibilities. It is very probable that the first glimpse of the surface characteristics of the planet Venus will be provided by a radar picture taken from an interplanetary vehicle. The use of radar techniques would permit the penetration of the optically opaque atmosphere which completely obscures this enigmatic planet, as well as provide a quantitative evaluation of its atmospheric components. This could be accomplished by measurement of the attenuation of the radar energy as a function of frequency. As will be discussed in Chapter 4, water vapor, oxygen, and carbon dioxide exhibit a marked effect upon radar energy transmission characteristics at certain frequencies. Passive radar techniques (microwave thermal mapping) could be employed to ascertain the surface temperature distributions and the heat balance. This type of scientific data would be invaluable for the determination and prediction of weather conditions.

1-5 THE MODULATION OF RADAR SIGNALS

A radar system may perform a number of functions (Paragraph 1-4) that involve the collection or transmission of intelligence for some defined tactical objective. The intelligence is carried by modulations of the radar microwave signal. The means used to create these modulations and the means employed to extract information from them (demodulation) form a convenient and mathematically useful way to describe and classify radar systems.

As will be seen in later portions of this book, the key to the understanding and proper design of a radar system is a knowledge of the modulation processes that can take place. The various processes of modulation and demodulation are conveniently explained by the use of simple generic representations of the three basic elements of a radar system: (1) the transmitter, (2) the target, and (3) the receiving system.

A simple transmitting system is shown in Fig. 1-10. It consists of a means for generating alternating current power, a means for carrying this power to an antenna, and an antenna that radiates some portion of this power into the surrounding space.

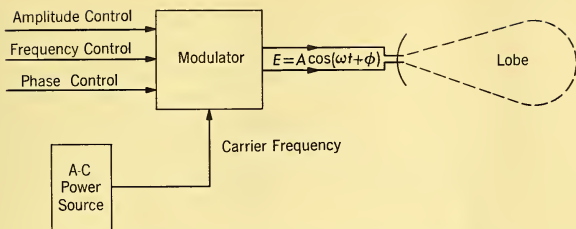


FIG. 1-10 Simple Transmitting System.

The generating device may be visualized as producing a sine wave output of constant amplitude and frequency.

$$E(t) = A \cos(\omega_0 t + \phi). \quad (1-14)$$

If this power is in turn applied to an antenna which radiates a portion equally in all directions (omnidirectional), we have the simplest sort of radar transmitter. We may proceed to refine the system by modulating the radiation in different ways.

Space Modulation. The radiated energy may be *space-modulated* by an antenna possessing directivity. Such a characteristic is shown in Fig. 1-10; the radiated energy is concentrated into a *lobe* by means of a parabolic reflector.

Three other types of modulation — *amplitude*, *frequency*, and *phase* — may be introduced by suitable operations upon the power generator.

Amplitude Modulation. If the output of the transmitter is amplitude-modulated at an angular frequency ω_1 with fractional modulation m , it then has the form

$$\begin{aligned} E(t) &= A_0 (1 + m \cos \omega_1 t) \cos(\omega_0 t + \phi) \\ &= A_0 \cos(\omega_0 t + \phi) + \frac{mA_0}{2} \cos[(\omega_0 - \omega_1)t + \phi] \\ &\quad + \frac{mA_0}{2} \cos[(\omega_0 + \omega_1)t + \phi]. \end{aligned} \quad (1-15)$$

Note that this type of modulation produces *sidebands* in the generated voltage; i.e., the generated voltage has frequency components which differ from the carrier angular frequency ω_0 by plus-or-minus the modulating angular frequency ω_1 . The transmitted spectrum for the case of 100 per cent modulation ($m = 1$) has the form shown in Fig. 1-11. The voltage amplitude of each sideband in this case is one-half that of the carrier, and the power in each sideband is one-quarter of the carrier power. Obviously

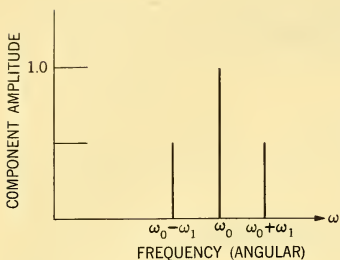


FIG. 1-11 Generated Frequency Spectrum for 100 Per Cent Sinusoidal Amplitude Modulation of Carrier.

as m decreases the power in the sidebands decreases and becomes a lesser fraction of the carrier power.

A common type of amplitude modulation arises from a modulating signal of the form shown in Fig. 1-12. Essentially, this signal turns the transmitter on and off on a periodic basis. Accordingly the output is a train of pulses of the carrier frequency. Since this modulating signal is periodic, it may be expressed as a Fourier series with a fundamental frequency equal to the pulse

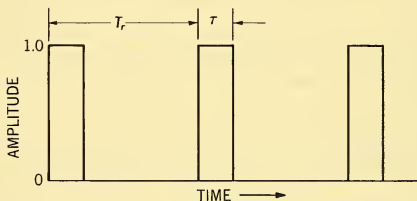


FIG. 1-12 Pulse Modulation.

repetition frequency (PRF — $1/T_r$), where T_r is the time between successive pulses.² Thus, this type of modulation gives rise to a large number of sidebands separated from the carrier frequency by multiples of the pulse repetition frequency. The amplitude spectrum of such a modulated wave is shown in Fig. 1-13. As can be seen, the pulse width τ determines the amplitude of each of the sidebands.

Radar systems employing the type of amplitude modulation just described are known as *pulse-type radars*. Pulse radars, however, are not limited to this type of modulation, as will be described in later paragraphs.

Frequency Modulation. Another major type of modulation is *frequency modulation*. In this case, the argument of the cosine function in Equation 1-14 is varied in such a manner as to cause the instantaneous frequency to be altered in accordance with the modulating signal. When

²Actually, the pulse amplitude modulated AF wave can be represented by a Fourier series with a fundamental frequency equal to the pulse repetition frequency only when the carrier frequency ω_0 is an integral multiple of the PRF.

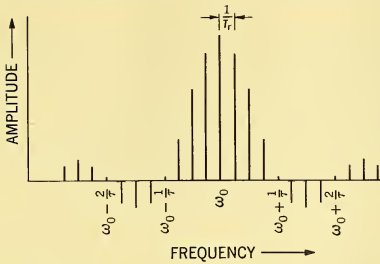
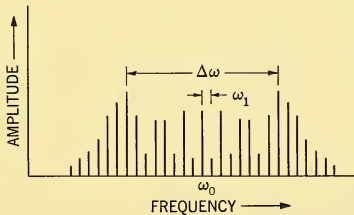


FIG. 1-13 Amplitude Spectrum of a Pulse Train.

the latter is a cosine wave of angular frequency ω_1 and the peak excursion of the modulated transmitting angular frequency is $\Delta\omega$, the transmitter output is

$$E(t) = A \cos \left(\omega_0 t + \frac{\Delta\omega}{\omega_1} \sin \omega_1 t + \phi \right) \quad (1-16)$$

whose envelope has a constant value. A typical frequency-modulated wave is shown in Fig. 1-14.

FIG. 1-14 Typical FM Spectrum for High-Modulation Index ($\Delta\omega/\omega_1 > 10$).

A key parameter in an FM system is the ratio

$$\frac{\Delta\omega}{\omega_1} = \text{modulation index.} \quad (1-17)$$

If this index is relatively high — say 10 or greater — the output spectrum has the form shown in Fig. 1-14. As can be seen, a single modulating frequency gives rise to a large number of sidebands separated from the carrier frequency by harmonics of the modulating frequency ω_1 . The sidebands of primary importance lie within a bandwidth $\Delta\omega$ centered about the carrier frequency ω_0 .

Various types of transmitter frequency modulations are commonly employed. *Pulse-width modulation* and *pulse-time modulation* are used to transmit information on a train of pulses. In Chapter 6 it will be seen how range can be obtained from a continuous-wave (CW) radar by frequency modulation of the transmitted frequency.

Frequency modulation of the transmitted signal often occurs inadvertently owing to the characteristics of the transmitter. The magnitude of this effect must be carefully controlled by the designer.

Phase Modulation. Phase modulation is similar to frequency modulation in that the instantaneous phase angle is varied from some mean value. With phase modulation by a single cosinusoid of frequency ω_1 and phase deviation ϕ , the transmitter output is

$$E(t) = A \cos (\omega_0 t + \Delta\phi \cos \omega_1 t). \quad (1-18)$$

The difference in the arguments of the cosine functions of Equations 1-16 and 1-18, while not important for audio systems, is important elsewhere where the waveshape must be controlled.

Subcarriers. The foregoing discussion has shown that it is possible to modulate the transmitted radar signal in four basic ways — space, amplitude, frequency, and phase. At this juncture, it is appropriate to consider just why one would want to modulate the transmitted signal. The purpose of these modulations is to create *information subcarriers*, i.e., an angle information subcarrier, a range information subcarrier, etc. The target information is contained in modulations of these subcarriers (and also the carrier frequency) that are created by the target itself and is derived upon return of the signal to the receiver by correlation with the transmitted subcarriers.

Target Modulations. In order to understand the basic processes involved, it is now appropriate to investigate the modulations of the main carrier and its associated subcarriers that are created by the target. First of all, the amplitudes of the transmitted radar signals that are reflected back to the transmitting location are vastly reduced — perhaps by a factor of 10^{20} on a power basis. Moreover, the reflecting characteristics of the target are, in general, a function of frequency. Thus, the amplitudes of each carrier frequency in the reflected wave may not be modulated by equal amounts.

Additional amplitude modulations are created by characteristic time variations of the target reflective characteristics. Chapter 4 will cover this phenomenon in detail. It will suffice for the moment to state that this effect introduces additional modulation which broadens each of the returned

sidebands to an extent depending upon the rate of target reflection characteristic fluctuations.

The target reflection entails *phase* changes with reference to the transmitted signal incident to the finite time required for propagation of microwave energy to and from the target. These phase changes occur in all the frequencies of the transmitted wave. The phase changes are linear with frequency and have a proportionality constant which depends upon the distance to the target. The phase modulation that occurs in the portion of the transmitted signal that is reflected back from the target provides the basic means for measuring range to the target. Pulse radars, for example, measure the phase (or time) difference between transmitted and received pulse trains.

Phase modulations of a somewhat different sort may result from the motion of the target in conjunction with the space-modulation characteristic of the radar. As an example of this process, consider a radar which scans a directional beam through an angle of 360° once each second. If there is a stationary target at an angle of θ_T with respect to the reference axis, a return from the target will be obtained as the radar beam sweeps past this point. The amplitude of the return signal will have the general shape of the radar beam resulting in a return signal having the envelope shown in Fig. 1-15. Thus, the scanning process gives rise to an angle in-

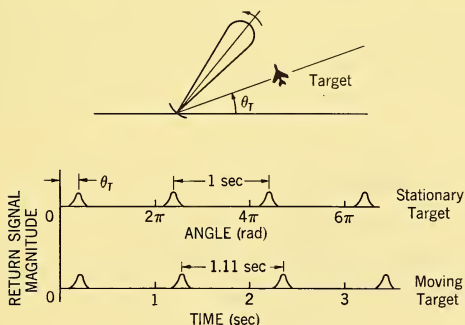


FIG. 1-15 Effect of Target Motion.

formation *subcarrier* which has a fundamental frequency of 1 cps. The angle information is carried on the phase angle of this subcarrier fundamental.

Now let us assume that the target flies in a circle around the radar station in the same direction the beam is revolving, at a speed of 1 revolution

every 10 seconds (i.e., one-tenth of the scanning velocity). The return signals now have the form shown in the lower diagram. The effect of the target motion has been to shift the fundamental frequency of the angle information subcarrier by 10 per cent to 0.9 cps for a target moving in the same direction as the scan. This process might also be viewed as introducing a time-varying phase shift in the 1-cps subcarrier (phase modulation). How the target modulates the information subcarriers is one of the most important problems of radar design. The choice of frequency and bandwidth for the subcarrier frequency information channels is largely governed by these characteristics.

One other target modulation — the doppler frequency shift mentioned in preceding paragraphs — is of fundamental importance.

Motion of the target along the direction of propagation (see Fig. 1-16)

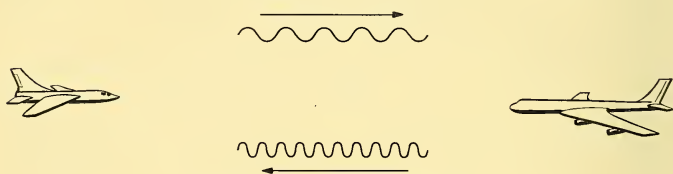


FIG. 1-16 The Doppler Effect.

causes each frequency component of the transmitted wave that strikes the target to be shifted by an amount

$$f_D = (V_T/c) f \quad (1-19)$$

where V_T = the velocity of the target

c = the velocity of light

f = the radio frequency.

When this signal is reflected or reradiated back to the radar, the total frequency shift of each component is

$$f_D = (2V_T/c) f. \quad (1-20)$$

The frequency modulation caused by target motion is important; an entire family of radars known as *doppler* radars has been developed to exploit this characteristic. However, whether use is made of this characteristic or not, the doppler shift occurs in all signals reflected from objects that possess relative radial motion.

Thus, it can be seen that the target generates a large number of amplitude, phase, and frequency modulations of the transmitted signal.

These modulations create information sidebands about the carrier and subcarrier frequencies. The designer's problem is to determine how this information may be extracted from the target return signal.

Extraction of Target Intelligence from Radar Signals. One thing is common to all the many techniques for extracting target information from a radar return signal. This is the concept of taking a *product* between the target return signal and another quantity which serves as the reference for the particular piece of information being extracted from the target return. Thus, the generic building block for a radar receiving system is a product-taking device, as shown in Fig. 1-17.

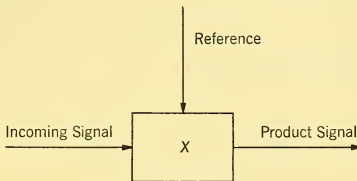


FIG. 1-17 Generic (Product) Building Block for a Radar Receiver.

Conceptually, the simplest product-taking device is a network — or filter — composed of linear impedances which can be characterized by a transfer function $F(j\omega)$. Each frequency component of the incoming signal is multiplied by the vector transfer function of the network corresponding to the frequency (see Fig. 1-18). The output product is a signal containing

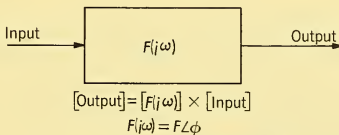


FIG. 1-18 Impedance Products.

the same frequencies as the input; however, the amplitude and phase of each frequency component may be changed with respect to the input. In this type of product device, the references are the characteristics built into the filter.

The second type of product-taking device is the nonlinear impedance. A simple example of such a device is shown in Fig. 1-19. The operation of the device is such that positive inputs are faithfully reproduced at the output while negative inputs are completely suppressed. Thus, for an

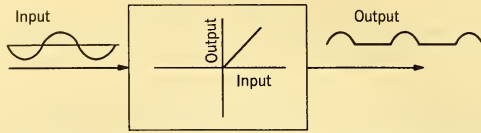


FIG. 1-19 Product Obtained from a Nonlinear Impedance.

input sine-wave, the output consists of only the positive half-cycles as shown.

It is interesting to observe that this process may be put into the form of the generic device of Fig. 1-18 merely by considering the output to be the product of the input and a reference square wave of the same frequency and phase as the input. Such a representation is shown in Fig. 1-20.

Thus, the input may be defined as

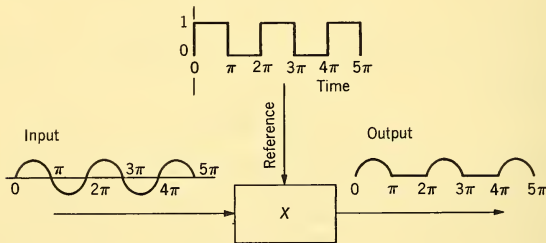


FIG. 1-20 Nonlinear Impedance as a Product-Taking Device.

$$E_i = A \sin \omega t. \tag{1-21}$$

The reference signal may be expressed by a Fourier series

$$E_r = \frac{1}{2} + \frac{2}{\pi}(\sin \omega t + \frac{\sin 3\omega t}{3} + \frac{\sin 5\omega t}{5} + \dots) \tag{1-22}$$

and the product has the Fourier series form

$$E_i \times E_r = \frac{A}{\pi} + \frac{A}{2} \sin \omega t - \frac{2A}{\pi} \sum_{n=1}^{\infty} \frac{\cos 2n\omega t}{4n^2 - 1} \tag{1-23}$$

The consequences of this product process are quite evident from Equation 1-23. Although the input contains only one frequency, the output has a d-c component, an input-frequency component, and components at all the even harmonic frequencies of the input.

Now, if the amplitude of the incoming wave A , instead of being constant as implied, were amplitude-modulated at a frequency ω_m , such that

$$A = A_0 (1 + m \cos \omega_m t) \quad (1-24)$$

where $\omega_m =$ modulating frequency ($\omega_m \ll \omega$) and $m =$ modulation ratio ($m < 1$), then, each of the terms of the product ($E_i \times E_r$) would contain modulation sidebands. For example, the d-c term would now become

$$\frac{A_0}{\pi} + \frac{mA_0}{\pi} \cos \omega_m t \quad (1-25)$$

and the fundamental frequency term would become

$$\frac{A_0}{2} \sin \omega t + \frac{mA_0}{4} [\sin(\omega + \omega_m)t - \sin(\omega - \omega_m)t] \quad (1-26)$$

and so on for the higher harmonics.

If this ($E_i \times E_r$) product were then passed through a filter, $F(j\omega)$, which eliminated the d-c term and the fundamental frequency ω and all its harmonics, the final output would be

$$(E_i \times E_r) \times F(j\omega) = \frac{mA_0}{\pi} \cos \omega_m t. \quad (1-27)$$

Thus, we observe, the frequency and the magnitude of the modulation intelligence are recovered from the incoming wave by the product-taking procedures. The procedures just described are often referred to as *demodulation* or *detection*.

A third type of product-taking device closely resembles the basic model of Fig. 1-17. The incoming signal is multiplied by a reference signal generated within the radar receiver. One form of this process is known as *mixing* or *heterodyning*. In this process, a cross-product is taken between the incoming signal and a locally generated signal. This process converts the microwave signal to a much lower frequency, which may be filtered and amplified by relatively simple electronic techniques.

Two general forms of microwave mixing are commonly used, *noncoherent* mixing and *coherent* mixing. In coherent mixing, the phase of the locally generated signal is made to have a known relationship to the phase of the transmitted signal. This type of mixing makes it possible to detect the phase and frequency modulations introduced by target motion.

The extraction of angle and range information from the received signals is almost always accomplished by a cross-product of the received intelligence and an internally generated reference signal.

The detailed analysis of the various means for extracting target intelligence from radar signals — and the problems that arise in these processes — forms a major portion of this book. Chapter 3 and Chapters 5 through 9 are all concerned with various phases of these problems.

1-6 OPERATING CARRIER FREQUENCY

The operating frequencies for radar systems cover an extremely wide band, ranging from below 100 to above 10,000 Mc. This range is divided up into bands designated P, L, S, X, K, Q, V, and W as shown in Fig. 1-21.

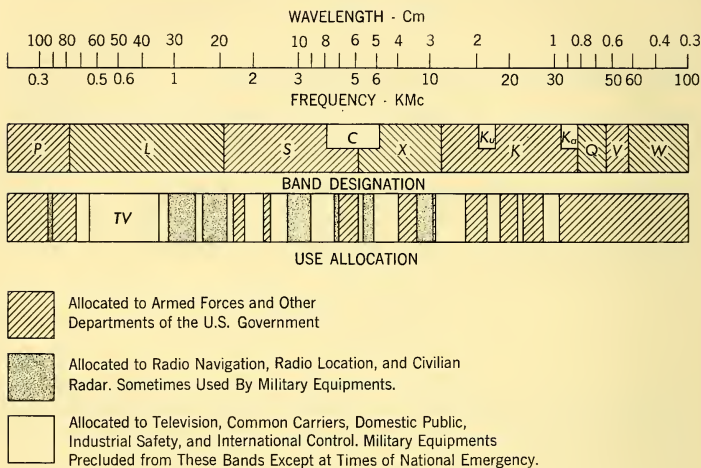


FIG. 1-21 Operating Radar Frequency.

The specific frequencies available for airborne radar systems are, in general, regulated by the Federal Communications Commission during times of comparative peace.

The operating carrier frequency has a profound effect on the following characteristics of a radar system:

1. Size, weight, and power-handling capabilities of the RF components (see Chapter 11)
2. Propagation of RF energy (see Chapter 4)
3. Scattering of RF energy (see Chapter 4)
4. Doppler frequency shift from a target moving relative to the radar direction of propagation.

These characteristics vary quite radically over the range of radar operating frequencies — enough, in fact, that it becomes convenient to classify a radar according to its operating carrier frequency. This method of classification is commonly used by microwave component designers

because the design problems and the techniques used to solve them are strongly dependent upon the operating frequency.

This method of classification is also important to the system designer because the operating frequency determines certain of the radar's reactions to its physical and tactical environment. For example, an atmosphere heavily laden with moisture is more or less opaque in some bands to the highest radar operating frequencies, whereas the transmission of the lower frequencies is little affected.

In airborne applications, the smaller size of the higher-frequency radar components has favored the use of S, X, and K bands despite their limitations with respect to weather and moving target indication, as discussed in Chapters 4 and 6.

1-7 THE AIRBORNE RADAR DESIGN PROBLEM

Preceding sections discussed general radar characteristics. The following problem is of paramount importance: How does the radar designer select and employ the right combination of these characteristics to achieve an acceptable performance level in a given weapons system application?

The design problem may be divided into two basic parts, problem definition and problem solution.

Problem Definition. The airborne radar design problem is defined by the weapons system application. In such applications, an airborne radar combines with other system elements — human operators and the airborne vehicle and its associated propulsion, navigation, armament, flight control, support, and data processing systems — to form a closely integrated weapons system designed to perform a specific mission. To achieve a given performance level, the weapons system *requires* certain performance characteristics from the airborne radar.

The radar designer's first task is to examine the requirements and characteristics of the complete weapons system. From this analysis, the nature of the airborne radar's contributions to overall weapons system performance (mission accomplishment) may be obtained. Typical examples of the parametric relationships developed in such a study are shown in Fig. 1-22a. From such curves, the radar requirements for a desired level of mission accomplishment may be obtained. In addition, the *sensitivity* of mission accomplishment to changes in radar performance is displayed, thereby providing the designer knowledge of the relative importance of each performance characteristic.

The derivation of such relationships must be relatively uninhibited by known limitations in the radar state of the art. That is to say, the range of values considered for each of the radar's performance capabilities need

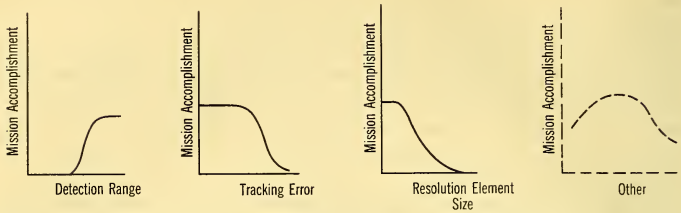


FIG. 1-22a Typical Relations Between Mission Accomplishment and Radar Performance in the Operating Environment.

not bear any relation to a presently realizable radar system. The purpose of this analysis is to define the radar problem solely as it is dictated by the weapons system problem. Whether the radar problem thus defined is technically reasonable for a given era is determined in the next major step of the design process.

Problem Solution. The systems analysis *defined* the required radar performance. Now, the designer must attempt to solve the defined problem by (1) hypothesizing a radar system of given general characteristics and (2) examining the interrelationships between radar parameters and radar performance. Typical examples of the interrelationships developed by such a study are shown in Fig. 1-22b for the case of a single radar param-

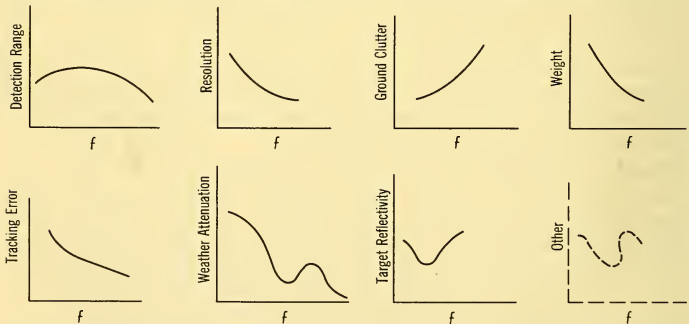


FIG. 1-22b Typical Interrelations of Radar Frequency and Radar Performance Parameters.

eter — operating frequency. Similar parametric relations are derived for each radar parameter that exercises important influences on radar performance.

The information thus derived is examined and correlated to find — if possible — the combinations of radar parameters which fulfill the previously derived radar performance requirements. Then and only then can the designer proceed in an intelligent manner to design the radar hardware for fabrication, evaluation, and service use. Often the proper combinations cannot be found. State-of-the-art limitations, laws of nature, and other factors may conspire to prevent a successful problem solution using the assumed radar concept. In these cases, the parametric information generated for the problem definition and the problem solution provide readily available means for ascertaining the most promising course of action — whether it be a change in radar concept, the initiation of a new component development, or a change in the overall weapons system concept. In extreme cases, a failure to find a radar solution may justify abandonment of a weapons system concept; in other cases an early display of seemingly irreconcilable deficiency may provide the spur for the generation of a bold new radar concept that performs as required.

Summary and Discussion. Airborne radar performance usually exercises a decisive influence on overall weapons system performance. The approach to the design problem must therefore be an overall systems approach, even though the radar is only a weapons system component.

The two basic steps in the design process are problem definition and problem solution as illustrated in Fig. 1-23. The first step derives the radar

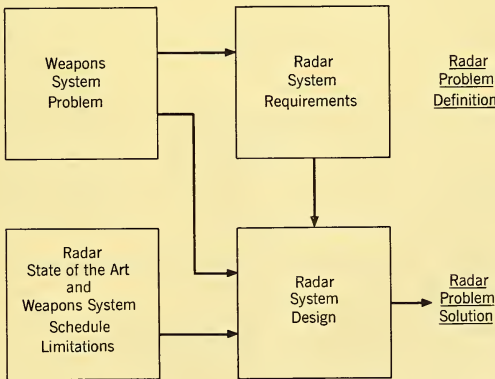


FIG. 1-23 The Airborne Radar System Design Problem Approach.

requirements imposed by the complete weapons system and *neglects* possible limitations of radar techniques. The second step is concerned with

fulfillment of the requirements *considering* limitations of radar state of the art, schedules, and other factors germane to the problem of achieving a useful operational capability for the overall weapons system.

In actual practice, the design process is long enough and complicated enough to justify subdivision of the two basic phases described. The next paragraph will discuss the complete design cycle of a typical weapons system development with particular emphasis on the role of the radar designer as a vital part of the designers' team.

1-8 THE SYSTEMS APPROACH TO AIRBORNE RADAR DESIGN

A representation of a typical weapons system design cycle is shown in Fig. 1-24. Each step represents a subdivision of the problem definition

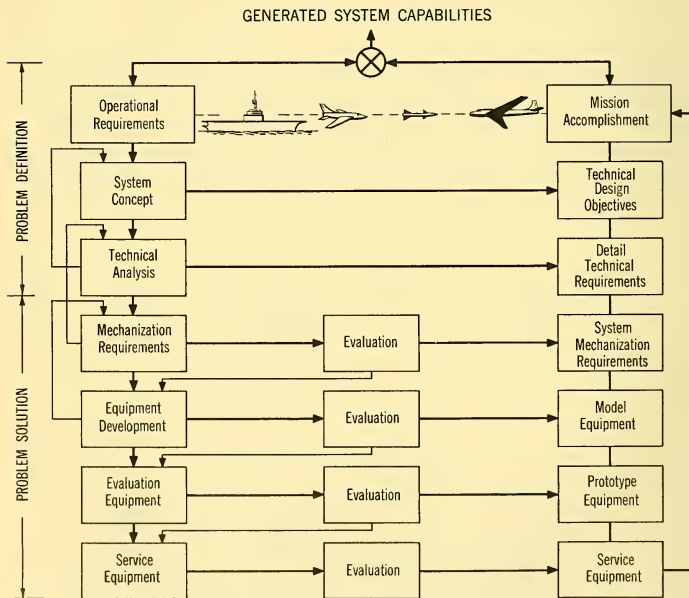


FIG. 1-24 Weapons System Design Cycle.

or a problem-solving step just discussed. On the left side of the figure are displayed the sequential steps of an orderly development process from the

initiation of an operational requirement to the fabrication of service equipment designed to fulfill the requirement. On the right of the figure are the definitive outputs or accomplishments resulting from this development. In the middle of the figure are the evaluation processes which measure the level of accomplishment attained in the problem-solving phases of the development. The outputs of these evaluations may also be used to modify succeeding phases of the development process.

The diagram also indicates feedbacks from the various development phases into preceding phases. These reflect the fact that as more is learned about the system, prior concepts must be modified and expanded to ensure that the system development objectives are current and realistic.

Viewed in its entirety, the indicated procedure provides a basis for playing current accomplishment against the requirement to obtain a continuous rating factor representing the generated system capability.

The step-by-step processes for executing the system development plan may be summarized as follows.

Operational Requirement. The overall system objective is set forth in an *operational requirement*. This requirement usually outlines the *military task(s)* which the weapons system must perform. It will also specify — or at least indicate — the *level of mission accomplishment* which the system must achieve to accomplish the desired military objective. The mission accomplishment requirement often has the general form displayed in Fig. 1-25. The weapons system must be operative in a given time

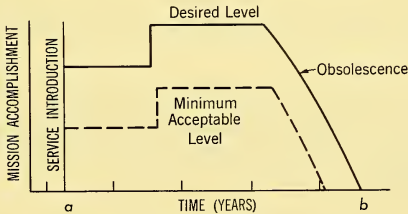


FIG. 1-25 Operational Requirement.

period *a-b*. The *desired level* of mission accomplishment represents the best estimate of what the military planners believe is necessary to achieve unquestioned military superiority in a given area of interest. This goal may be variable over the expected operational use cycle (as shown) by reason of anticipated introduction of new techniques by the enemy.

The *minimum acceptable level* of mission accomplishment represents a capability which the military planners believe is still useful enough to

justify the weapons system cost. Thus, although the system design will endeavor to meet the desired goal, some degradation may be acceptable if such degradation can be shown to be unavoidable.

Unexpected developments in technology or in the long-range strategic situation can cause radical changes in the operational requirement during the design cycle. For this reason, the radar designer must constantly monitor the operational requirement to ensure quick reaction to such changes.

System Concept. The operational requirement defines a military problem. The next step is to define a system concept which provides bases to presume a weapons system potential capability compatible with the operational requirement. This step usually is implemented in the following way. Various possible systems are postulated. Technical military agencies examine these in the light of available or projected technical capabilities to determine which provides the best foundation for a subsequent development. Weapons system contractors may assist this study phase by providing new ideas, state-of-the-art evaluations, etc.; however, the basic responsibility for decision and action invariably rests with the military.

Once a decision is made on the type of system desired, the basic features of the selected system are set forth in the form of *technical design objectives*. These comprise the performance specification of the overall weapons system and

1. The system effectiveness goal related to the operational requirement
2. The basic system philosophy, i.e., mode of operation
3. The system environment as defined by tactics, logistics, climate, etc.
4. The characteristics of major system elements
5. The system design, development, and evaluation program
6. Fundamental state-of-the-art limitation in various portions of the system

Unless he has already participated in the definition of the system concept, the radar designer's *first* task is to become familiar with these conceptual characteristics of the overall weapons system. They define the elements of his problem which are relatively fixed and with which his design must be compatible.

Technical Analysis. The systems problem and its boundary conditions having been defined and understood, the radar designer now is ready for the next step — the construction of a weapons system model that will *define* the radar problem. This model is used to determine the quan-

titative interrelationships of the radar and other system elements. From manipulation of these interrelationships the designer must obtain the true technical requirements of the radar necessary to attain a proper balance between mission accomplishment and the operational requirement (see Fig. 1-24).

The detailed technical requirements include specification of

1. Functional capabilities
2. Radar range and angle coverage requirements
3. Information handling, transfer, and display requirements
4. Radar information accuracy requirements
5. Radar environmental requirements
6. Radar system reliability requirements
7. Radar maintenance, stowage, and handling requirements

In this stage of the design, the emphasis is placed on the job the radar must do and the environment in which it must operate. As previously mentioned, this analysis should not be inhibited by the introduction of state-of-the-art radar limitations. The problem in this phase is to ascertain what requirements the radar must satisfy to allow the defined system concept to demonstrate an adequate system capability.

This type of analysis is not popular with radar designers. Often it results in establishing technical requirements beyond the scope of known radar technology. The radar designer is forced to admit a set of requirements he does not believe he can meet.

However, the wise weapons systems contractor will *demand* that such an analysis be performed and demonstrated by the radar designer for two reasons:

1. It will ensure that the radar designer really understands the problem *before* he tries to solve it.
2. The earlier a potential source of system degradation is known, the easier it is to correct by invention or by modification of the development program.

Because this step is so important to the radar designer, Chapter 2 is devoted to a detailed discussion and example of the processes involved in the derivation of technical requirements for a radar system.

Mechanization Requirements. The next step in the design process is to synthesize a realizable radar system to meet the derived technical requirements within the limitations imposed by development time, state of the art, and delivery considerations. Various radar systems are designed on paper and analyzed in detail to demonstrate their performance relative

to the previously established requirements. Often, it is found that the requirements are not compatible with realizable radar systems. This discrepancy may be corrected in some cases by using the interrelationships of the radar and the overall system derived in the previous phase to find a different balance of radar requirements that will still permit mission accomplishment. In other cases, a known degradation in performance — relative to the initial weapons system goal — may have to be accepted. A somewhat happier situation arises when it is found that a realizable radar system can provide *greater* capabilities than those required by the initial weapons system concept and operational requirement. In this case, the initial system concept could be enlarged and improved or, conversely, the development objectives could be revised with a resulting economy of design.

The indicated processes of evaluation and feedback are shown in Fig. 1-24. The flexibility achieved by the feedback process is the real strength of the systems approach. There are few weapons systems that cannot benefit from modification of the originally established concepts and technical requirements. Circumstances change — often in a highly unpredictable manner — over the five-to-ten-year development period of a weapons system. That is why the radar designer must continue to treat this problem on an overall weapons system basis throughout the life of the project.

Equipment Development, Evaluation, and Use. Similar comments concerning the value of the systems approach apply to the vitally important task of building, evaluating, and using the equipment in accordance with the requirements derived in the first three phases. The problems in the latter phases can be formidable. For some perverse reason the actual equipment in certain critical areas may not perform in the manner predicted, particularly with respect to reliability. A vital part of the system approach is the process of rapid isolation and correction of system deficiencies in these phases and the anticipation of potentially critical areas. Since it is often difficult to distinguish between a genuine system deficiency and a temporary bottleneck, the judgment and experience of radar systems engineers who have also participated in the requirements derivation phase is most important. There are so many problems in these phases that it is easy to concentrate effort on the wrong ones.

One problem — *reliability* — dominates these last phases. This is the most vexing, most difficult, and most important problem in the design of a radar system. Radar systems have never been simple; in the future, their complexity may be expected to increase. The most common failure of the systems approach to the radar design problem has been the tendency to maximize system capability by specifying unnecessarily exotic radar requirements which lead to reliability problems in the mechanization phases.

In the technical requirements analysis phase, all possible ingenuity should be employed to minimize radar complexity for a given weapons system capability. This is one of the most important reasons why the radar system designer must analyze the radar problem as a weapons system problem. Important performance benefits can be achieved by making the proper reliability-complexity trade-off early in the game.

Summary. The systems approach to radar design implies that the radar is considered in its relation to the construction and objectives of the entire weapons system during all phases of its conception, design, construction, and use. The radar systems designer must participate in all phases of the development; he must demonstrate a good understanding of the overall system and the characteristics of its other components prior to laying out a radar design.

To increase the reader's understanding of the basic features of the systems approach to airborne radar design, several of the points presented will be amplified in succeeding paragraphs. These include a more precise definition of the system environment and its effects upon the problem, and a brief discussion of the concepts and processes involved in the construction of a weapons system model.

1-9 SYSTEMS ENVIRONMENTS

In this book, the "expected tactical conditions of operation" include all of the following environments.

1. *Tactical Environment* — The salient elements of this environment are the speed, altitude, operating characteristics, and mission profile of the airborne portion of the weapons system; the composition, operating characteristics, and relative position of the ground-based portion of the weapons system; and the characteristics (speed, course, altitude, number, physical size) of the target complex.

2. *Physical Environment* — The salient elements involved are temperature, pressure, humidity, precipitation, fog, salt spray, wind, clouds, sand, and dust. In systems requiring a human operator, the physical environment will include factors affecting his ability to operate the system. Among these are habitability, ease of operation, length of attention span required, and the physical readiness and mental acumen required.

3. *Airframe Environment* — The salient elements involved are volume and configuration of allotted space within the airframe, weight limitations, vibration, and shock.

4. *Electronic Environment* — This includes all the external sources of electromagnetic radiations and electromagnetic radiation distortions and anomalies. Examples are ground, sea, and cloud clutter; radiation from

other systems; electronic countermeasures; propagation anomalies, atmospheric attenuation; and target radar reflective characteristics.

5. *Logistics Environment* — This includes all salient considerations of the parts of the weapons system that affect reliability, maintenance, handling, stowage, supply, replacement, and transport.

6. *Weapons System Integration Environment* — This is the environment formed by other systems with which the system under consideration must be compatible. An example would be the environment formed by a ground-to-air missile weapons system operating in the vicinity of a proposed interceptor weapons system.

These environments should be looked upon as boundary conditions imposed upon the systems design problem. The concept is shown diagrammatically in Fig. 1-26. Each element of the system design must be com-

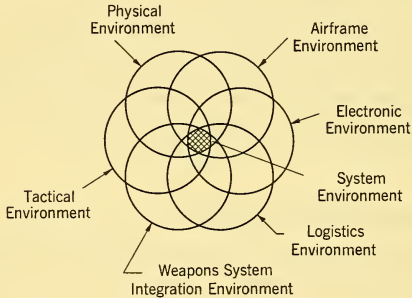


FIG. 1-26 The System Environment.

patible with the requirements and limitations imposed by *all* of these environments.

1-10 WEAPONS SYSTEM MODELS

A *weapons system model* is a simplified representation of the actual system which can be used to predict the changes in system performance when one or more of the components which make up the system are changed. For example, a common problem in radar system design is to determine the effect of radar detection range on the performance of the overall system. Such a problem would be solved — as will be shown in Chapter 2 — by constructing a model containing radar detection range as a variable parameter. The radar detection range would be related by appropriate means to the pertinent characteristics of the overlapping environments which make up the system complex. Within the model,

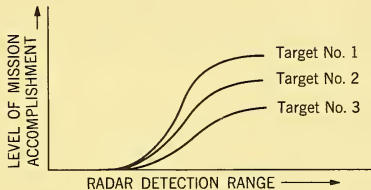
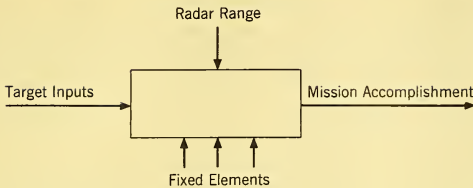


FIG. 1-27 Diagrammatic Representation of a Model Used to Determine Sensitivity of Mission Accomplishment to Variations in Radar Detection Range.

the elements of the system model interact to provide a generated mission accomplishment as shown in Fig. 1-27. From such a model, the effect of radar range upon system effectiveness could be obtained for various inputs (targets), yielding the characteristic type curves shown.

It is far more convenient to conduct an experiment with a model of a system than with the system itself. This is particularly true with a military weapons system where the testing of actual hardware is enormously expensive. Moreover, system hardware usually is not available until long after the original concepts. Theoretical models are required to predict expected system performance.

There are three classes of models which can be used in systems analysis work: iconic, analogue, and symbolic. Briefly, we may define the characteristics of these model types in the following ways:

1. An iconic model represents certain characteristics of a system by visual or pictorial means.
2. An analogue model replaces certain characteristics of the system it represents by *analogous* characteristics.
3. A symbolic model represents certain characteristics of a system by mathematical or logical expressions.

Iconic Models. An iconic model is the most literal. It “looks like” the system it represents. Iconic models can quickly portray the role that each subsystem plays in the operation of the overall system. It is therefore particularly well adapted to illustrating the *qualitative* aspects of system performance, such as information flow and functional characteristics of various portions of the system.

The iconic model is not well adapted to the representation of dynamic characteristics of the system because it does not reveal the quantitative relationships between various elements of the system. For the same reason, it is not very useful for studying the effects of changes in the system. Because of its pictorial value, most system analyses usually begin with the construction of an iconic model (block diagram) in order to establish the characteristics of the system and to provide the investigator with a realistic frame of reference for subsequent studies. This process will be illustrated by examples later in this book.

Analogue Models. Analogue models are made by transforming certain properties of a system into analogous properties, the object being to transform a complicated phenomenon into a similar form that is more easily analyzed and manipulated to reveal at any early time the initial elements of system performance. For example, fluid flow through pipes can be replaced by the flow of electrical current through wires. A slightly more abstract example would derive from the problem of calculating the probability of a mid-air collision in a situation where only the laws of chance were operative — i.e., where no special equipment or techniques were used to prevent collisions. The imaginative investigator might perceive that this problem bears a striking similarity to the problem of calculating the mean free path of a gas molecule. Having established the validity of this insight, we would then be free to make appropriate transformations between the two problems and apply the kinetic theory of gases to his problem.

Unlike the iconic model, the analogue model is very effective in representing dynamic situations. In addition, it is usually a relatively simple matter to investigate the effects of changes in the system with an analogue model. For these reasons, analogue models form very powerful tools for the solution of complex system problems — particularly problems involving many nonlinearities.

The great utility of analogue models is evidenced by the large-scale analogue computer installations that form a part of almost every major weapons system engineering organization.

Symbolic Models. The symbolic model represents the components of a system and their interrelationships by mathematical or logical symbols.

This type of model is the most abstract. When such a model is formulated and used without incurring prohibitive mathematical complexity, it is the most useful model for obtaining quantitative answers to systems problems.

Many problems may be solved by either analogue or symbolic models. Where a choice exists, it is preferable to employ the symbolic model, for it allows one to examine the effect of changes by a few steps of mathematical deduction. This process was implied in the example of the mid-air collision analogy, just cited. Here the problem was transformed into an equivalent gas dynamics analogue. However, for such a problem we should not construct a complex instrument and make measurements — it is far simpler to use the symbolic models already established for the kinetic behavior of gases. Further, we would gain greater insight into the basic nature of the problem in this way than would be obtained by empirical methods.

The utility of symbolic models is particularly evident for problems involving probability concepts. Often, answers may be obtained in closed form for problems that would otherwise require many repeated tests of an analogue model.

The primary disadvantage of symbolic models arises from limitations in available mathematical and computational techniques for obtaining answers from the model. State-of-the-art improvements in applied mathematics and large digital computers are relieving this problem. Despite these advances, however, there will always be a great premium on the ability to construct symbolic models that strike at the heart of a problem and eliminate nonessentials that merely increase complexity.

1-11 THE BASIC STATISTICAL CHARACTER OF WEAPONS SYSTEM MODELS

The model approach consists of abstracting from a complex system certain persistent and discernible relations and using these relations to construct a system model. Frequently, owing either to the inherent nature of the process being examined or to the complex nature of the process, the relations must be expressed in a statistical form. That is to say, certain portions of the system — and as a result of this, the system itself — will not possess a unique output for a given input. Rather, the output must be expressed as a spectrum of possible events where each event has a certain *probability* of occurrence.

Two simple examples may serve to illustrate the nature of the phenomena involved:

Example 1—Measurement Uncertainties. Measurements of time, distance, temperature, etc., always possess a certain error tolerance. For example, a large number of distance measurements made with the same

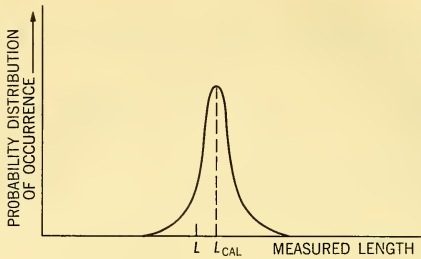


FIG. 1-28 Measurement Errors Obtained in Determining Length.

instrument measuring the same distance might give rise to a distribution of values about some mean value as shown in Fig. 1-28. If these measurements are compared with a standard, we see that two sources of error exist: (1) a calibration or bias error, and (2) a random error. The calibration error — so long as it remains fixed or if its variations can be predicted — is obviously a correctable source of inaccuracy. However, the random error is just that — random. Any given measurement may be in error by an amount determined by the character — usually Gaussian — of the randomness.

Measurement uncertainties are a vitally important problem in any weapons system analysis. Unlike many engineering problems, where measurements may be made with whatever degree of preciseness is necessary to render inconsequential the measurement error, weapons systems habitually are required to work with measurement uncertainties that exercise a profound and usually decisive influence upon their performance.

Example 2 — Dice Throwing. The cast of a die is an example of a process that is, theoretically, completely predictable; however, because of the extreme complexity of the mechanisms that govern its behavior, the whole process is more easily handled by probability concepts. For example, if we knew the exact orientation of the die, its velocity, direction of motion, physical size, shape and weight distribution, condition, characteristics and orientation of the die table, etc., we could predict with certainty which of the die faces would appear on top. However, the amount of information that must be obtained and processed to arrive at this result is usually prohibitive. It is much easier — and, also, as in the case of representations of this type, less profitable — to characterize such a process by saying that for any input (legal throw) the system (balanced die cube) may produce any output from 1 to 6 with equal probability.

Weapons systems contain many processes of similar brand. Ballistic trajectories are a prime example. The behavior of a human being in a

control loop is another. Thus the basic parameters of the model usually take the form of distribution functions.

Attention is invited to the Operations Research volume³ of this series for a detailed treatment of the theory of probability as applied to weapons system evaluation.

1-12 CONSTRUCTION AND MANIPULATION OF WEAPONS SYSTEM MODELS

Before analyzing the structure of a model, let us review some of the peculiar characteristics of a weapons system.

1. A weapons system is an organization of men and equipment designed for operation and use against specific classes of enemy targets. To carry out its overall function — usually the destruction of the enemy target — it must carry out many complex subfunctions. Each functional activity converts certain quantitative *inputs* into *outputs*. The entire weapons system is merely a series — or series-parallel — arrangement of these subfunctions connected in such a manner as to permit achievement of the overall system objective. As an example of a typical organization, an air-to-air intercept system might be characterized by the sequence of operational functions shown in Fig. 1-29. Also shown are the major equipments that are involved in the performance of each function.

The input to the system is an enemy target — the output is the destruction of the enemy target. Similarly, each operational function can be viewed as an input-output device. The subject of input-output relations brings us quite naturally to a consideration of another distinguishing characteristic of a weapons system.

2. A weapons system is a dynamic or *time response* system. Both the system inputs and outputs have time variables. This fact makes it necessary to treat a weapons system in terms of the *time delays* that it introduces between the input (enemy target) and the output (action against the enemy target). The likelihood of mission accomplishment usually is strongly dependent upon the ability of the system to respond to an input within a specified period of time.

Each operational function can contribute to the overall time response characteristic. For example, a finite time is required to process target intelligence and tactical situation information for the purpose of assigning a weapon to the target. Upon being assigned, the interceptor aircraft requires a certain amount of time to take off and fly to the target location, etc. Thus, the concept of partitioning the system into subfunctions —

³Grayson Merrill, Harold Greenberg, and Robert H. Helmholtz, *Operations Research, Armament, Launching* (Principles of Guided Missile Design Series), D. Van Nostrand Co., Inc., Princeton, N. J., 1956.

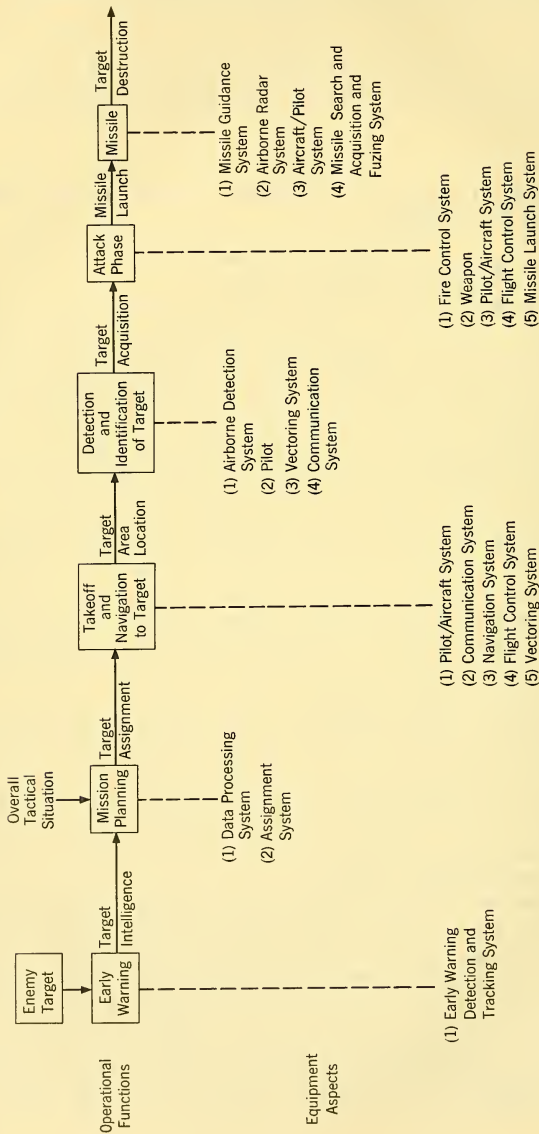


FIG. 1-29 Sequence of Operational Functions for an Air-to-Air Intercept System.

each characterized by a time delay response characteristic — is basic to our modeling approach.

Bearing these observations in mind, we may outline the structural composition of a weapons system model as follows:

1. *Input Variables* — The input variables include all the characteristics of the enemy target complex — size, number, location, speed, defense capability, etc. — that are pertinent to the operation of the weapons system. Also included are the elements of the fixed environment — the physical environment, the environment provided by other weapons systems, etc. — that affect the operation of the weapons system.

2. *Mission Accomplishment Goals* — This is a quantitative expression of the desired system output. Usually, it derives from the operational requirement. This quantity and the input variables define the problem that the weapons system must solve.

3. *System Logic* — The system logic describes the system organization and the flow of information through the system; i.e. how the system operates on input data, what sequence of operations takes place, what the pre-established tactical doctrine is for a given set of input variables, etc. This structural element of the mathematical model provides the means for breaking up the overall system function into logically consistent subfunctions.

4. *System Configuration Parameters* — These include the basic elements and characteristics of the system needed to implement the system logic — the geometry of the system, the number of weapons, the weapon characteristics, and the capabilities and characteristics of each of the system elements such as aircraft, radars, etc.

5. *Model Parameters* — The basic model parameters are the time delays that are defined for each of the system subfunctions on the basis of the input parameters, the system logic, and the system configuration parameters. For reasons that were discussed in Paragraph 1-11, the time interval associated with the performance of each function must usually be expressed as a *probability distribution* of time delays rather than as fixed time delay.

Often, range to the target is used instead of time as the means for expressing the basic parameters of the model. This is merely another way of expressing the time delay, since range and time are related through the relative velocity between the target and the weapon.

Suboptimization. Most weapons systems are so complex that it is not possible to construct a single model that includes all possible parameters and variables. Instead, many different models must be constructed, each designed to explore a certain facet of the system operating and its relation

to the whole. In the example of the interceptor system previously referred to, we might have the following models.

1. *System Flow Model* illustrating the qualitative aspects of system operation.
2. *Overall System Effectiveness Model* indicating in gross terms the defense level that could be provided against a multiplane attack.
3. *Early Warning Detection and Tracking Model* indicating the quantitative aspects of the problem of early-warning detection, identification, tracking, and target assignment.
4. *Interceptor-Bomber Duel Model*, including (a) Vectoring Model, (b) Airborne Radar Detection Model, (c) Attack Phase Model, and (d) Target Destruction Model.

Each model must be logically consistent with the others, to maintain a unified systems approach. Constructing these models shows the same stages as constructing the overall system, namely,

1. Define input variables.
2. Define performance criteria ("mission accomplishment goals").
3. Outline system logic.
4. Define the configuration parameters of the part of the system being analyzed.
5. Derive the quantitative relationships on the basis of the input parameters, system logic, and configuration parameters.

This process of analyzing only a small portion of the system problem at any one time is known as *suboptimization*. Successive suboptimizations of various portions of the system can form a step-by-step approximation which converges to the results that would be obtained if one could analyze the entire system with one model.

Counterbalancing the obvious analytical advantages of the suboptimization technique is the fact that it gives rise to a serious bookkeeping problem. The results of each suboptimization process must be logically consistent with the rest. As we successively suboptimize various portions of the system, the assumptions of previous suboptimizing routines may be changed. We must recognize such changes and modify previous suboptimization routines to be consistent with these changes. Summary tables, information flow diagrams, and functional block diagrams form indispensable tools for the bookkeeping process.

1-13 SUMMARY

In this chapter, we have covered the general characteristics of radar systems; the environments in which they operate; the functional capabilities

they provide; the basic means by which intelligence is carried on and extracted from microwave radiations.

In addition, we have emphasized the role of the radar as a device intimately connected with other devices to form an overall *system*. The requirements for the radar must be derived from a logical study of the performance of the complete weapons system in its expected tactical conditions of operation. A knowledge of broad weapons system modeling techniques, technical competence in the diverse aircraft and guided missiles arts, and a capacity for logical reasoning are required for this process in addition to a detailed knowledge of radar systems.

The next chapter will demonstrate how the systems approach is applied to typical examples of airborne radar system design. Succeeding chapters will cover the detailed problems of radar design and their relationship to the system problem.

CHAPTER 2

THE DEVELOPMENT OF WEAPONS SYSTEM
REQUIREMENTS

2-1 INTRODUCTION TO THE PROBLEM

Assume that a radar design group is presented with the following problem.

Specify, design, and build the following radar systems for use in a fast attack carrier task force environment in the time period 19xy to 19xx.

1. Airborne Intercept (AI) radar and fire-control system for all-weather transonic interceptors, and
2. Airborne Early Warning (AEW) radar system for installation in a 35,000 lb gross weight AEW aircraft.

This chapter will demonstrate how the first part of this problem — specification of radar requirements — can be solved using the general approach outlined in Chapter 1. (Later chapters will discuss the additional problems involved in mechanizing a radar system to meet a set of derived requirements.)

Although a specific problem is treated, the method of approach has applicability to all radar specification problems. Particular emphasis is placed on the processes of obtaining a good understanding of the overall weapons system problem which dictates the radar requirements.

Methods for formulating a system study plan and for constructing system models are illustrated by examples. The type of information which must be collected and analyzed by the radar design group *prior* to attempting the radar design also receives attention and comment.

The hypothetical problem used as a vehicle for this discussion has been greatly simplified; however, it is still a complicated problem and the reader will be required to perform the same sort of rechecking, cross-referencing, and backtracking that is required to understand and follow an actual systems analysis problem. To facilitate this process, much of the reference data is displayed on charts and diagrams which typify the pictorial (iconic) representations that should be constructed for an actual systems problem.

Wherever possible, mathematical complexities have been eliminated, minimized, or referenced. However, the reader will find that knowledge of simple probability theory, operational calculus, and feedback control theory provides a key to better understanding of a weapons system study.

Readers having some familiarity with the radar design problem may be somewhat surprised — even appalled — at the amount of study which must be done prior to considering the radar design problem itself. They may be tempted to say: “Surely, you do not expect a radar designer to concern himself with such matters? These are the responsibility of the agency that prepares our specification!”

The authors’ reply to these remarks is based on their own not always pleasant personal experiences and close observation of the experiences of many other radar system designers.

Seldom, if ever, does the basic specification given to a radar designer provide the necessary input information for a successful system design. It is a starting point — nothing more. Until the radar designer understands the nature of his contribution to the solution of the overall system problem he has little chance of designing a radar system that will operate satisfactorily as part of a complete weapons system.

In recognition of this fact, the weapons system contractor not only should encourage, he should require the radar designer to demonstrate his understanding of the problem by preparing a formal document embodying the type of analysis and development to be demonstrated in this chapter. This will tend to ensure that the purposes and objectives of the weapons system are not made subservient to the preconceived biases of the radar designer.

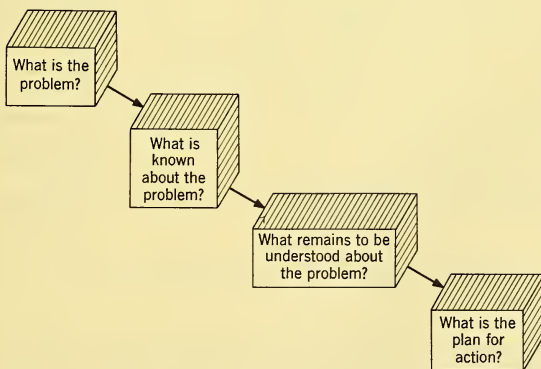


FIG. 2-1 Steps Leading to the Formulation of the System Study Plan.

2-2 FORMULATING THE SYSTEM STUDY PLAN

Reduced to its barest essentials, any system may be visualized as the logical process of asking and answering the sequence of four questions outlined in Fig. 2-1. This basic pattern is repeated — again and again — throughout the course of the study.

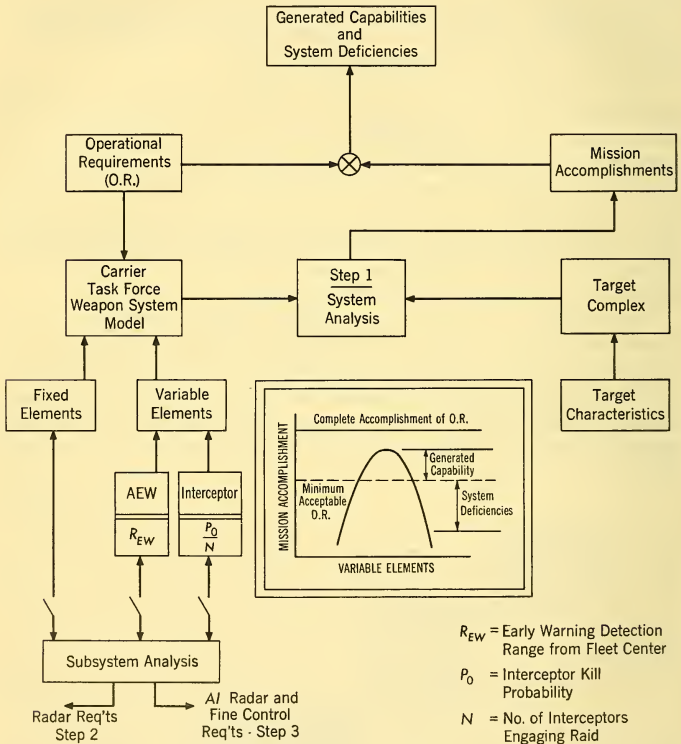


FIG. 2-2 Master Plan for Weapons System Analysis.

First, the problem is defined. For the hypothetical carrier task force air defense system, this problem definition takes the form of the following question:

“What does the weapons system require of the airborne radars (AI and AEW) to achieve a satisfactory level of mission accomplishment?”¹

The basic elements of this problem are shown in Fig. 2-2. The *operational requirement* defines a weapons system problem. By the processes described in Paragraph 1-8, the operational requirement leads to the establishment of the concept of a *system* which depends for its operation upon the characteristics of a number of subsystems — aircraft, missiles, detection devices, and shipboard installations, for example.

As is apparent from this figure, the solution to the radar requirements problems will involve consideration of many complex characteristics and relationships *external* to the airborne radars. Hence, the middle two questions (Fig. 2-1) and their answers are fundamental to further progress. For the hypothetical problem, the known and unknown elements of the problem are displayed in Figs. 2-2 through 2-8. These will be discussed in greater detail in subsequent paragraphs.

The *plan for action* will develop quite naturally from the indicated sequence of questions and answers. Elements of the problem that are not known or understood must be investigated in greater detail and related to the known elements. In some cases, adequate information may not be available on the unknown elements of the problem — or the path to understanding may be blocked by the inherent difficulty of the problem. These cases will require that arbitrary assumptions be made in order that the analysis can proceed.

When a sufficient understanding of the overall problem is obtained by analysis (or assumption), weapons system models are constructed. These models have as variable parameters the performance characteristics of the airborne radar known to be important to weapons system operation.

Using the techniques of Operations Research and Systems Analysis, these models are “played” against the target inputs (Fig. 2-2). The level of system performance (Mission Accomplishment) obtained is compared with the operational requirement, thereby generating a measure of the system capabilities (or deficiencies). By such processes, mission accomplishment may be related to the radars’ performance characteristics (see inset, Fig. 2-2) thereby providing a means for obtaining the true requirements of the airborne radars as dictated by the weapons system requirements.

It is important to emphasize once again that the derivation of requirements should not be affected by state-of-the-art considerations in radar technology. The purpose of the analysis is to define the radar problem, not to solve it. Only after this task is completed is the radar designer free to turn his attention to the job of designing and building a specific radar system to meet the requirements imposed by the weapons system problem.

¹The problem of defending a carrier task force is quite analogous to the defense of a city or important military base.

2-3 AIRCRAFT CARRIER TASK FORCE AIR DEFENSE SYSTEM

The fixed elements of the carrier task force environment, the target complex, and the operational requirement provide the basis for the derivation of radar requirements. We shall assume that all elements of the task force air defense system are fixed, except the following:

1. The Airborne Early Warning Aircraft Radar and Data Processing System,
2. The Airborne Interceptor Radar and Fire-Control System.

Referring to the system study plan of Fig. 2-2, the first task is to determine: (1) what is known about this weapons system, and (2) what remains to be known or understood.

The overall weapons system is broken down into four major operating elements: (1) Ship Weapons System, (2) AEW Aircraft System, (3) Interceptor Aircraft System, and (4) Air-to-Air Missile System. The characteristics of each element — and the interrelationship between elements — which exert a sensitive influence on the overall system performance are shown in Figs. 2-3 through 2-6. The known characteristics are checked; graphical or tabular description of these appear in the same figures.

The unknown (unchecked) characteristics constitute the items a knowledge of which must be gained from the system study. These include the characteristics of the AEW and AI radar systems; they also include important interrelationships among the various system elements, known and unknown. For example, the individual characteristics of the interceptor aircraft and the air-to-air missile are known; the manner in which these two elements combine and interrelate with the interceptor fire-control system to produce a weapons system capability is not known and must be derived by study.

The list of sensitive parameters is not complete. At the outset of a system study it is not possible to name all the parameters that may be important. As more is learned about the system through study, these must be added and considered in their proper relationship with other parameters.

Carrier Task Force Weapons System (Fig. 2-3). This system includes two aircraft carriers and three missile-firing cruisers in the disposition shown. (This configuration is, of course, fictitious and is used to illustrate the possible elements of this problem. Elements of the task force not germane to this example are excluded.)

Two large carriers constitute the main offensive and defensive elements. The carriers are separated (5-20 n. mi.) for tactical reasons. Mass attacks

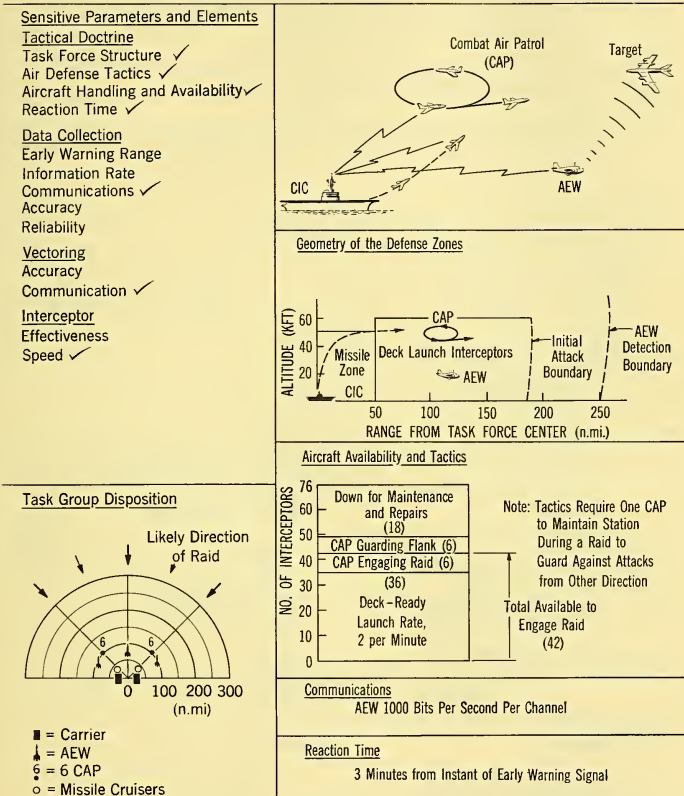


FIG. 2-3 Carrier Task Force Weapons System Characteristics.

are assumed to occur only on the side of the task force which is exposed to enemy territory.

Local defense of the task force is augmented by missile-firing cruisers flanking, and somewhat forward of each carrier. These provide a point-defense system with an altitude capability of 50,000 ft extending out to about 40-50 n. mi. in front of the carriers. A third missile-firing cruiser

guards against sneak attacks from the rear as well as turn-around and reattack tactics.

Early warning detection and interceptor vectoring information is provided by airborne early warning aircraft (AEW). The task force is assumed to have a capability for maintaining three AEW aircraft aloft on a 24-hour basis. The primary functions of the AEW system are:

- (1) To provide detection of enemy aircraft at sufficient range forward of fleet center to permit interception by piloted aircraft.

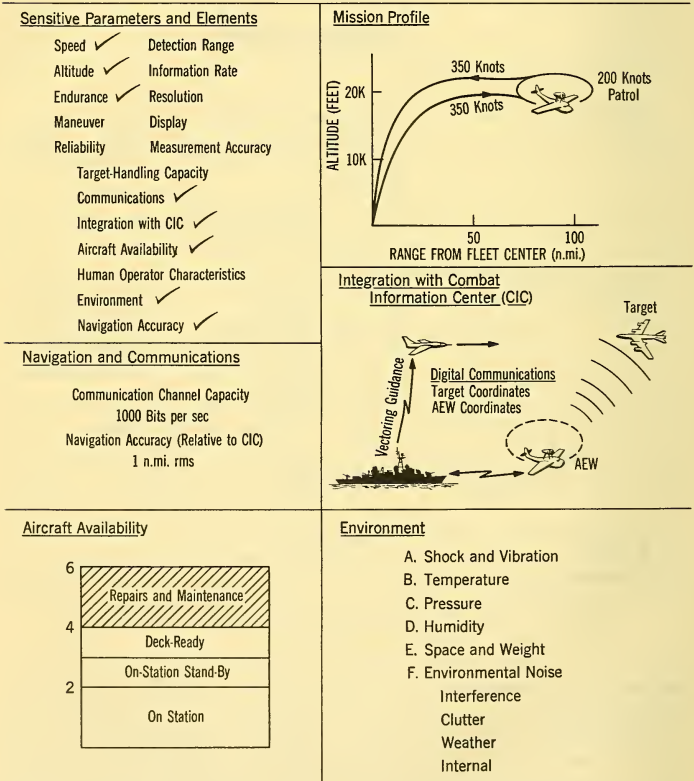


FIG. 2-4 Airborne Early Warning System. See Fig. 2-18 for Tactical Deployment.

- (2) To provide target data for the ship-based combat information center (CIC), which supplies vectoring information to piloted interceptors and guidance information to ship-based missile radars.

Two combat air patrols (CAP) are maintained. Each CAP contains 6 all-weather interceptor aircraft. In addition, interceptors may be launched at a maximum rate of 1 per minute from each of the two carriers during attack conditions. Aircraft availability limits the total number of deck-launched interceptors to 36.

During an attack, only one CAP (6 aircraft) engages the raid; the other is held in reserve to guard against attacks from other directions. Thus, a maximum of 42 interceptors can be used to engage a raid.

The interceptors are armed with air-to-air guided missiles and are required to perform the interception function at altitudes from sea level to 60,000 ft.

An optimum battle-control and communications system is assumed. This is to say, the deployment of interceptors by CIC is such that any interceptor which enters the interceptor zone is able to make an attack so long as there are targets within the zone. As will be shown later, system performance is sensitively affected by this assumption, which represents a condition most difficult to realize in practice.

Airborne Early Warning System (Fig. 2-4). The basic functions of the AEW system have been described.

The carrier-based aircraft available for this purpose is assumed to be capable of housing an antenna with a maximum dimension of 12 ft in the mushroomlike appendage shown.

The exact disposition of the AEW aircraft and the AEW radar and data processing requirements will be determined by study.

A major unknown is the contribution of AEW target information accuracy to interceptor effectiveness.

Interceptor Aircraft System (Fig. 2-5). The known and unknown characteristics are defined as shown. The determination of detailed radar requirements will require an analysis of the dynamics of the closed-loop system formed by the target, the interceptor, the pilot, and the AI radar and fire-control system.

The interrelationships among the aircraft system, the air-to-air missile system, and attack tactics are also unknown and must be analyzed as a prelude to the ascertainment of radar requirements.

A major unknown, to be determined by the system study, is the contribution of vectoring accuracy to interceptor effectiveness.

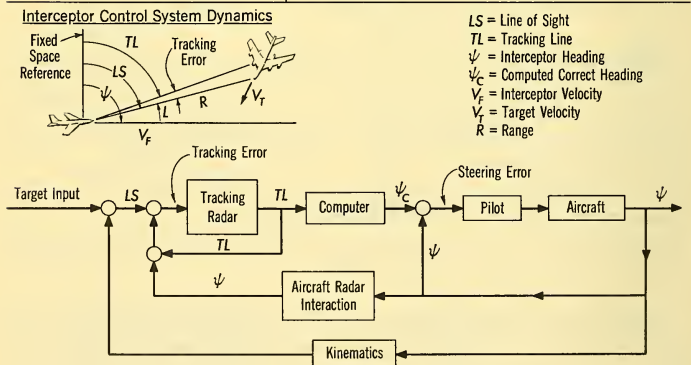
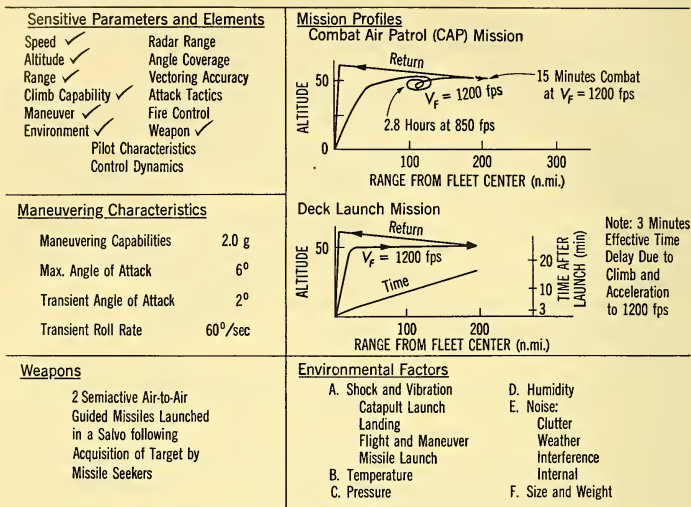


FIG. 2-5 Interceptor System Characteristics.

Air-to-Air Missile System (Fig. 2-6). The sensitive parameters of the air-to-air missile system are shown. The major unknown parameters involve interrelationships of the missile with the aircraft and the radar and fire-control system.

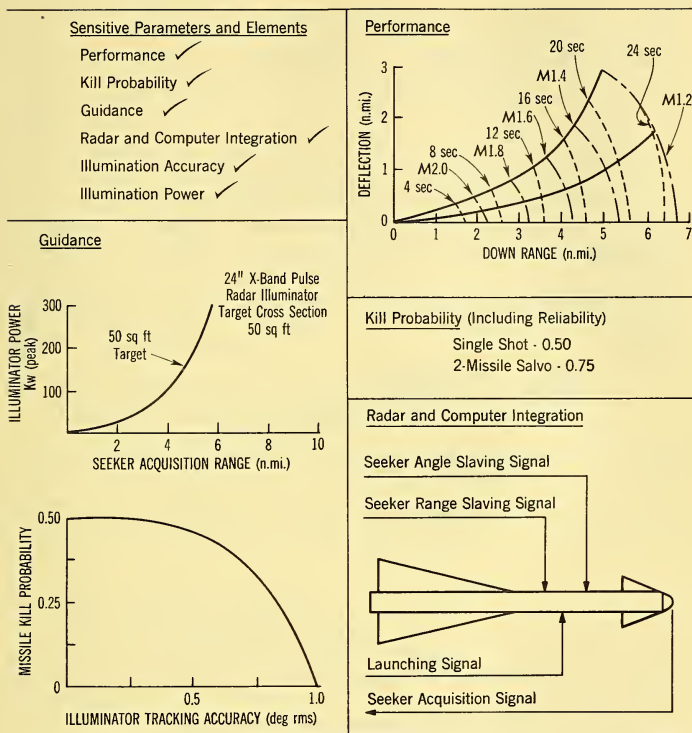


FIG. 2-6 Air-to-Air Missile System Characteristics.

2-4 THE TARGET COMPLEX (FIG. 2-7)

Target Complex. The characteristics of the target complex include its parameters and its mission as shown in Fig. 2-7.

The 50-nautical mile (n. mi.) air-to-surface missile (ASM) carried by the hostile aircraft requires that interception take place *outside* of a 50-n. mi. circle around the aircraft carriers.

The target's 2-g maneuver capability will exercise an important influence on the radar and fire-control system design. It will be assumed for this example that the enemy aircraft is provided with the capability for optimum timing of this maneuver. Also, it is assumed that the target does not employ electronic countermeasures (ECM).

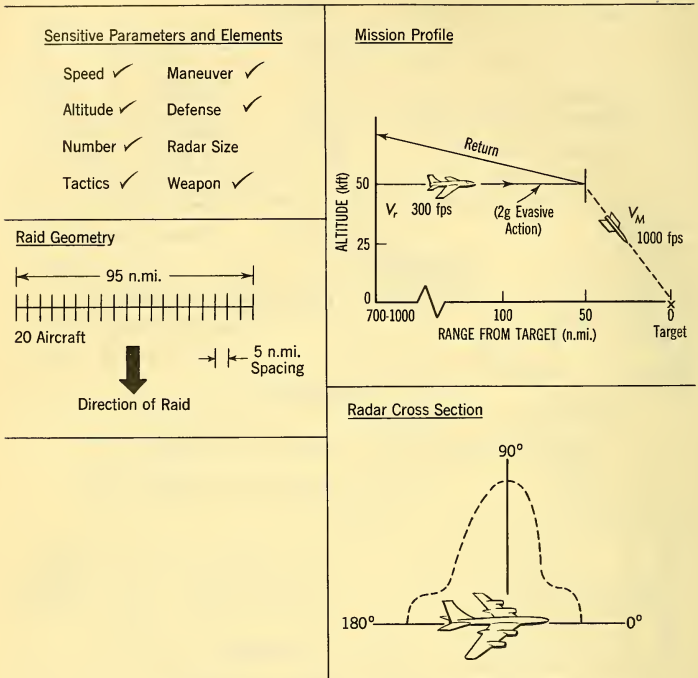


FIG. 2-7 Target Complex Characteristics.

The radar cross-section characteristics of the hypothetical target are only generally known and are shown in Fig. 2-2. Paragraph 4-7 explains the factors contributing to characteristics of this type. Paragraph 4-9 discusses how the target radar area may be estimated for purposes of preliminary design.

The turboprop propulsion system of the enemy aircraft was chosen to introduce into the model the effects of the modulation characteristics of the reflected radar energy (Paragraphs 4-7 and 4-8). This can be an important radar design consideration.

The target is assumed to carry a high-yield nuclear weapon. Destruction resulting from impact on the target aircraft is assumed to cause a detonation capable of producing a destructive overpressure within a 1000-ft radius sphere around the target. Ignoring time effects, 1000 ft thus defines the

point of allowable minimum approach of the interceptor to the target aircraft.

The number of target aircraft (20) and their spacing (5 n. mi.) is characteristic of a raid designed to present a difficult problem to the model air defense system.

In an actual problem, a number of different target complexes would have to be defined in this way. The behavior of the system would be analyzed for the several inputs and the design parameters chosen on the basis of the response to all expected target complexes with emphasis on the most effective configuration. For simplicity in this example, we will confine our attention to the single problem defined; however, the *sensitivity* of system performance to *changes* in this input (i.e. target speed and number) will be examined.

2-5 THE OPERATIONAL REQUIREMENT: MISSION ACCOMPLISHMENT GOALS (FIG. 2-8)

The operational requirement defines a military problem which must be solved by the combination of known and unknown weapons system elements previously described. Bases for judging the military usefulness of any system proposed as an answer to the operational requirement are also shown.

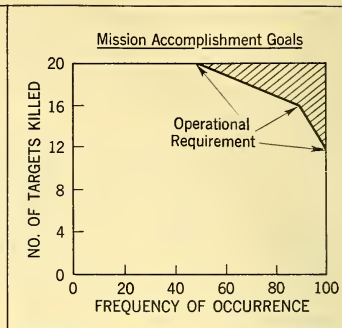
2-6 THE SYSTEM CONCEPT

The operational requirement defined a weapons system problem. The procedures for solution of this problem are determined by the system concept or logic.

Within the framework of the system elements already defined, the system logic for the interceptor system may be developed by listing the sequence of events which lead to the interception of the target by the missile-armed interceptor. The following events would normally be expected to occur in sequence:

- a. Early warning detection
- b. Identification
- c. Threat evaluation
- d. Weapon assignment
- e. Interceptor direction or vectoring
- f. AI radar search and detection
- g. AI radar acquisition
- h. Airborne weapons system tracking control and missile launching
- i. Air-to-air missile guidance
- j. Missile detonation and target destruction (without self-destruction)

<u>Function</u>	High-Attrition Air Defense Against Medium Bombers
<u>Features</u>	(1) Compatibility with Surface-to-Air Missile System (2) Compatibility with Fleet Elements, Logistics, and Tactics (3) Compatibility with Transonic Interceptor Aircraft -- 30,000 Lb Gross Weight



Mission Accomplishment Goals (See Inset Figure)

- (1) A 20-Plane Raid at 50,000 ft and 800 fps shall be Employed as the Input for Judging System Performance. System Performance shall be Judged as Satisfactory if the System can be Demonstrated to Attain or Exceed the Following Performance Levels (As Indicated by the Shaded Area in the Figure)
 - (a) 50 per cent Probability of Killing all 20 Targets
 - (b) 90 per cent Probability of Killing 16 Targets
 - (c) 99 per cent Probability of Killing 12 Targets
- (2) All Kills are to be Accomplished at a Minimum Distance of 50 n.mi. from Fleet Center

FIG. 2-8 Operational Requirements, Attack Carrier Task Force Interceptor Air Defense.

k. Return to base

l. Transfer of residual target elements to the surface-to-air missile support defense system.

A diagrammatic summary representation of the overall tactical situation is shown in Fig. 2-9.

The statistical nature of the system operation is shown by this diagram. For various reasons — interceptor availability, time limitations, system failures, system inaccuracies — a certain percentage of the interceptors fail to complete each of the successive steps required for interception.

Thus, any interceptor chosen at random from the total complement has a certain *probability* that it will kill a target. This probability is the *product* of the individual probabilities that it will pass successfully through each successive stage of the interception.

This line of reasoning points out the necessity for obtaining a proper balance between the performance of various elements of the system. A

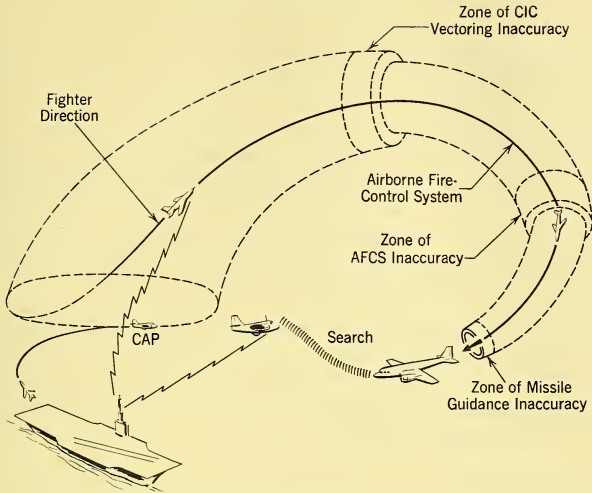


FIG. 2-9 A Tactical Situation.

very low probability of success for any phase of the intercept mission can render pointless any efforts to achieve very high probabilities in other phases and thus would serve as a guide to more effective development emphasis.

2-7 THE SYSTEM STUDY PLAN

The known (fixed) and unknown (variable) elements of the problem have now been defined. Referring to Fig. 2-2, it is seen that the next step is to analyze the interrelationships between the fixed and variable elements to determine the contribution of each variable element to mission accomplishment. From such analyses, a quantitative understanding of system operation will be obtained and — eventually — radar requirements will evolve.

The unknown or variable elements may be broken into two basic categories: (1) weapons system variables, and (2) subsystem variables.

The primary weapons system variables are:

1. The early warning detection range measured from fleet center.
2. The number of interceptors which may engage the specified target complex
3. The effectiveness (kill probability) of each interceptor which engages the target complex

All of the other unknown parameters (subsystem variables) affect one or more of these three basic weapons system variables. This observation makes it possible to organize the study plan on a step-by-step basis as follows:

Step 1: Construct a model of the overall weapons system using the three primary weapons system variables as adjustable parameters. *Assume* values for each of these adjustable parameters and *calculate* the resulting system performance. Compare this performance with the desired level of mission accomplishment; use any discrepancy between the two to adjust parameter values for another tentative design. Testing of the model continues until the following information is derived.

1. All combinations of the adjustable weapons system parameter values that will allow achievement of the mission accomplishment goal.
2. The sensitivity of system performance to changes in the values of the adjustable parameters.

Additional information — useful for obtaining a good understanding of the overall problem — is obtained by ascertaining the sensitivity of system performance to changes in the fixed elements.

Step 2: Assume fixed values for the three weapons system variables of Step 1 that permit the system to achieve the desired level of mission accomplishment. *Construct* a model (or models) which expresses the relationships between the adjustable (unknown) AEW parameters (beam width, information rate, radar detection range, etc.) and the assumed weapon system parameters. *Test* this model for various assumed combinations of AEW parameters. *Establish* acceptable combinations of AEW parameter values and the sensitivity of system performance to parameter changes. *Derive* a specific set of AEW requirements.

Step 3: Using the values for the unknown system variables derived in Step 1 and 2, repeat Step 2 for the adjustable parameters of the interceptor weapons system. Derive a specific set of requirements for the AI radar and fire-control system. The suggested order of Steps 2 and 3 is somewhat arbitrary; a reasonable case might be made for reversing this order. As a general rule, where a choice exists, it is wise to select an order which places the most difficult subsystems first, since this will maximize the number of adjustable parameters available for its preliminary design.

2-8 MODEL PARAMETERS

The interrelations between major system parameters and the contributions of each parameter to overall effectiveness may be developed through

the use of mathematical models. These interrelationships form the quantitative bases for the choice of basic system parameters.

The following paragraphs will develop a number of models designed to expose some of the more important aspects of the task force air defense problems. The techniques used to develop these models are illustrative of the means by which any complex system problem may be broken down to forms that can be handled by analytical means.

2-9 SYSTEM EFFECTIVENESS MODELS

The operational requirement (Fig. 2-8) specified the effectiveness of the interceptor air defense system in terms of the degree of success which must be achieved with a required reliability. For example, the probability of destroying at least 16 of the 20 targets should exceed 90 per cent.

The first step of the systems analysis must determine the nature of the relationships between system effectiveness and the fixed and variable elements of the defense system (Fig. 2-2). The following examples demonstrate how such an analysis may be carried out.

Assume that 40 interceptors may be brought to bear against the 20-target raid previously assumed as the threat ($N = 40$). Each interceptor can make only one attack with its two-missile salvo. Thus, when one attack against a target fails, another interceptor will be assigned to that target until either 40 attacks have been made or all the targets have been destroyed.

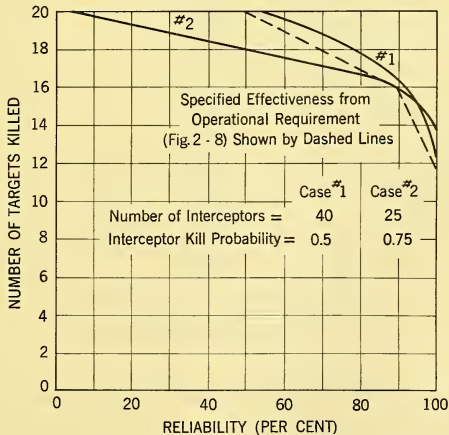


FIG. 2-10 System Effectiveness Operating Characteristics.

If the effectiveness of each interceptor is assumed to be $P = 0.5$, the system operating characteristic shown in Fig. 2-10 may be calculated by the application of simple probability theory.² As can be seen, the assumed parameters allow the operational requirement to be met.

To obtain a complete picture, other possibilities may be assumed and analyzed in the same fashion. For example, the effectiveness level provided by 25 interceptors, each with a kill probability of 0.7, is also shown in the figure. This combination of parameter values fulfills only part of the operational requirement. The probability that more than 16 targets will be destroyed is less than that for the previous assumptions; thus, this system would impose increased requirements on the back-up surface-to-air missile system.

Continuing in this fashion, trade-off curves between the number of interceptors and the interceptor kill probability can be determined for each point of the operational requirement. Such a curve is shown in Fig. 2-11. Here all the combinations of the interceptor kill probability and number of interceptions are shown which will kill at least 16 out of 20 targets with a 90 per cent reliability.

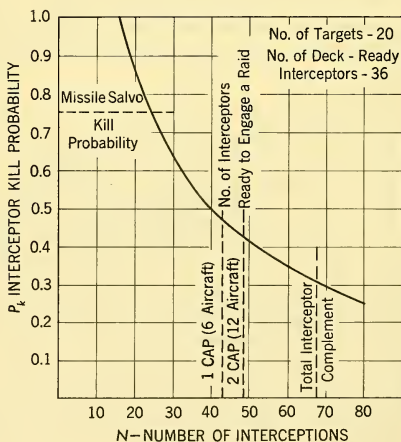


FIG. 2-11 Interceptor Kill Probability vs. Number of Interceptions Required to Kill 16 or More Targets with 90 Per Cent Probability.

²Grayson Merrill, Harold Goldberg, and Robert H. Helmholtz, *Operations Research, Armament, Launching* (Principles of Guided Missile Design Series), D. Van Nostrand Co., Inc., 1956. Chapters 6 and 7 provide an excellent discussion of the mathematical techniques involved.

Also shown are the limiting effects of the fixed problem elements previously outlined in Figs. 2-3 through 2-6. For example, the missile salvo kill probability is 0.75; obviously the interceptor kill probability cannot exceed this value. If the raid is engaged by six CAP interceptors and all 36 of the deck-ready interceptors, an interceptor kill probability of 0.48 is required. If the tactics are changed to allow *both* CAP patrols to engage the raid in addition to the 36 deck-ready interceptors, the individual interceptor kill probability required drops to 0.42. If the total complement of 66 interceptors could be used, a kill probability of only 0.3 would be required.

The basic parametric relationships between the number of interceptors and system effectiveness now are established. The next phase of the systems analysis must determine the relationships between the number of interceptors and the other fixed and variable elements. Completion of this phase will provide the basic parametric data which will, in turn, allow intelligent selection of the following system parameters (see Fig. 2-2).

1. Number of interceptors (N)
2. Interceptor effectiveness (P_0)
3. Early-warning range (R_{AEW}).

The number of interceptors which can be used to defend a given raid, and thus the required interceptor kill probability, is a function of initial interceptor deployment, detection ranges, reaction times, and target and interceptor speeds. These factors can be conveniently summarized in a diagram similar to Fig. 2-12, which shows the sequence of events in a typical raid. The interceptor and target performance characteristics were given in Figs. 2-5 and 2-7. We assume, as an illustrative case, that the AEW detection range is 250 n. mi. from the fleet center. Since the target has a speed of 800 fps (474 knots), it will arrive at the fleet center 32 minutes after detection. The target track is shown in Fig. 2-12 as the straight line connecting 250 n. mi. at zero time to 32 minutes at zero range.

The CAP interceptors stationed 100 n.mi. from the fleet center are vectored to intercept the raid following a 3-minute time delay consumed by the process of identification, acquisition, and assignment. The track of the CAP aircraft is constructed as a line with a slope equal to the reciprocal of their speeds (1200 fps or 710 knots). We observe that the intersection of the two tracks occurs at 175 n.mi., the maximum range at which the raid can be engaged.

In accordance with defined tactical doctrine (Fig. 2-3), only one combat air patrol (6 aircraft) is committed to the raid. The remaining CAP maintains its station to guard against the possibility of attacks from other directions.

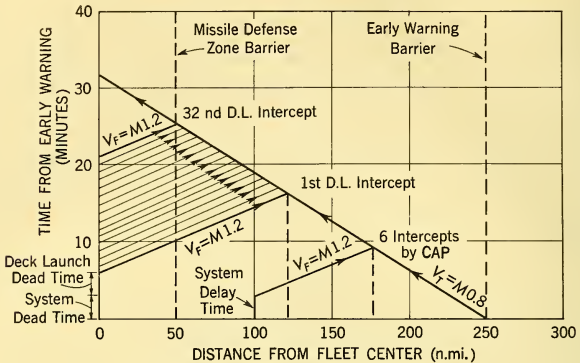


FIG. 2-12 An Attack Diagram.

From Fig. 2-5 it is seen that an additional 3-minute delay is created by the interceptor climb and acceleration characteristics. This makes the total effective reaction time of the deck-launched aircraft equal 6 minutes. After the initial reaction, interceptors will be launched at a rate of 1 per minute from each of the two carriers until either no more interceptors are left or it is obvious that the interceptors will not be able to intercept the targets outside of the surface-to-air missile zone corresponding to a 50-n.mi. radius from the fleet center. In our example, this latter consideration is the limiting factor, and it is possible to launch only 32 interceptors from the carriers. Thus, a total of 38 attacks can be made against the raid with the assumed deployment, tactical doctrine, and equipment performance. This is close to our previously assumed case with 40 interceptors, and the required interceptor kill probability will be slightly greater than 0.5. The air battle takes place during a 16-minute time period to enemy penetration of the missile defense zone barrier. The maximum time that any interceptor must fly at Mach 1.2 is 11 minutes (for the first two deck-launched interceptors), which is well within the interceptor performance capabilities as displayed in Fig. 2-5.

This model may be used to examine the effect of variations in the system parameters. The results of such an analysis are shown in Fig. 2-13, where trade-off curves relating pertinent factors are given. If the early warning range is increased to 300 n.mi., 50 interceptors can engage the raid. With 50 interceptor attacks, the required interceptor kill probability will be reduced to 0.42. However, with the tactical doctrine assumed, the maximum interceptor complement available to counter an attack is limited to 42. Thus as early warning range increases, aircraft availability in this

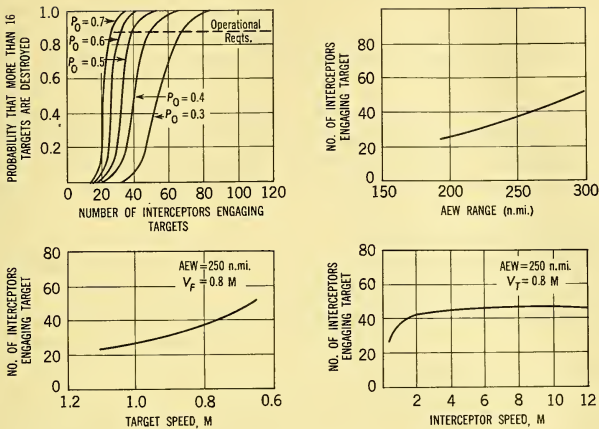


FIG. 2-13 Sensitivity of System Effectiveness to Number of Interceptors, AEW Range, and Interceptor Kill Probability P_0 .

model becomes the limiting factor. This limitation might indicate that a trade-off of parameter values elsewhere in the problem should be examined to exploit the potential advantage which might accrue from a range increase.³ Conversely, a 50-n.mi. decrease in early warning range would require that interceptor kill probability be increased to 0.70 — a value that is almost equal to the kill probability of the missile salvo alone — in order to maintain the system effectiveness required.

Increases in target velocity have much the same effect as decreases in early warning range. If the target velocity were to increase by 10 per cent, only 34 interceptions could be made. Thus, to maintain the same defense level under these conditions, interceptor kill probability would have to be raised to 0.58 or early warning range would have to be increased by about 30 n.mi. Increases in interceptor velocity have the same general effect as increases in early warning range; i.e., aircraft availability limits the usefulness of such increases. System sensitivity to this change is relatively small, however, for extremely high interceptor speeds.

The time delays defined for the model made no allowance for any time delay introduced by the vectoring process. The assumption is that the vectoring system guides each interceptor on a straight-line path to the

³For example, we might explore the possibility of using the other CAP aircraft which were assumed to maintain their stations.

earliest possible interception point. Other types of vectoring guidance — for example, a tactic whereby it is attempted to guide the interceptor on a tail-chase attack — introduce additional time delays, and these reduce the total number of interceptions which can be made with a given early warning range.

System tactics also exercise other important influences. For example, a 3-minute dead time delay was assumed between early warning detection and assignment of the first interceptors to specific targets. This type of operation places a high value on the time delay. Each minute of time delay requires 8 additional miles of early warning detection range to maintain a fixed number of interceptions.

The effect of this time delay would be different if the system tactics called for launching of interceptors to begin *before* evaluation was completed. This operation, however, incurs a risk that interceptors may be launched unnecessarily. In this latter case, it would be necessary to evaluate the consequences of a false alarm as a function of threat evaluation time; i.e., the penalties of launching interceptors when the threat does not materialize following an early warning detection in terms of fuel loss, vulnerability to attacks from other directions, etc. Some of these considerations may seem to go a little far afield, but the answers to such questions are of great importance to the radar designer because they affect what his equipment must do. To simplify our example, we assume that *no* interceptors are launched until evaluation of the threat is completed.

As a second example of the effect of tactics, we might consider the target assignment procedure. In our example, we assume that an optimum assignment procedure could be used. That is to say, each of the 40 interceptors was able to make an attack during the course of the air battle — except in the cases where all 20 targets were destroyed by less than 40 attacks. This assumption assumes a very sophisticated battle control and communications system. Another method of assignment could be as follows: the first 20 interceptors are assigned — one-on-one — to the first 20 targets. The following interceptors are assigned as back-up interceptors on the same basis — i.e. interceptor 21 to target 1, interceptor 22 to target 2, etc. For 40 interceptions, this would mean each target could be attacked twice. In some cases, however, the target would be killed by the first interceptor thereby leaving the back-up interceptor without a target to attack, resulting in a potential inefficiency. On the other hand, two attacks may not suffice to kill the target since each attack has less than unity success probability.

With these alternate tactics, a substantially greater interceptor kill probability would be required to meet the operational requirement for the case of 40 interceptors reaching the attack zone. This value has been determined to be 0.7 as compared with 0.5 when optimum target assign-

ments are made. The advantages of the optimum assignment tactic are apparent.

In this paragraph we have shown how the effectiveness of an interceptor system may be analyzed for an assumed mode of operation and assumed values for system parameters. We have also illustrated the concept of obtaining the "trade-off" between various system parameters and the effects of changes in system logic. The examples chosen are merely illustrative of the information that must be generated for an actual problem to enable the radar system designer to understand his part of the overall problem.

Using the assumed or derived values for the overall system parameters, and the defined system logic, we shall now derive the requirements for the AEW and AI radar systems. The first phase of these analyses (Steps 2 and 3 of Master Plan of Fig. 2-2) is to establish the allocation of responsibility between these two systems.

2-10 PRELIMINARY DESIGN OF THE AIRBORNE EARLY WARNING SYSTEM

The AEW system must contribute to the solution of the air defense problem in several ways as may be seen from the operational sequence given in Paragraph 2-3.

1. The targets must be *detected* at sufficient range from task force center to permit fulfillment of the required system kill probability.

2. The targets must be identified and evaluated. This means that their identity, number, position, heading, speed, and altitude must be obtained; this information must be evaluated in terms of the implied threat to the task force; and weapons must be assigned, if necessary. This process must be completed within a delay time that is compatible with early warning detection range and the characteristics of the interceptor defense system.

3. The AEW system must provide information which can be used to vector the interceptors toward their assigned targets so that the interceptors may detect and acquire the targets with their own AI radars. The *type* of vectoring guidance employed must be compatible with system response times permitted by the early warning detection range. The *accuracy* of vectoring guidance must be compatible with the input accuracy requirements of the interceptor aircraft, AI radar, and fire-control system.

4. The AEW system must provide information that may be used for overall battle control and surveillance.

The basic plan to be used for the AEW system analysis is shown in Fig. 2-14. Also shown are the interrelations between: (1) the AEW system and the overall problem, and (2) the AEW system and interceptor system.

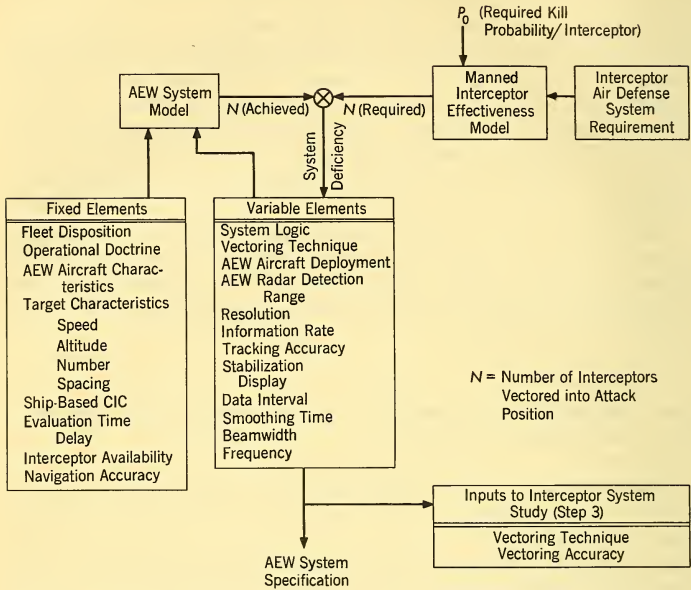


FIG. 2-14 Plan for Analysis of AEW System Requirements (Step 2 of Master Plan — Fig. 2-2).

The number of interceptors N which can be directed against the separate elements (single surviving targets) of the 20-plane raid during the air battle is selected as the criterion for judging AEW system performance.

As already shown (Paragraph 2-9) the *required* number of interceptions depends upon the kill probability per interceptor. If we assume a value of 0.50, the required number of interceptions is 40 (Fig. 2-11). The *task force* early warning range needed to meet this requirement is 255 n.mi.⁴ This is one possible combination of parameters satisfying the operational requirement and the interceptor availability limitation. We may select this combination as a design point — keeping in mind that it is possible to trade off interceptor effectiveness and number of interceptions should subsequent analysis indicate this to be desirable. For purposes of

⁴More correctly, the specification of detection range should include the required minimum probability of obtaining that range, e.g. 90 per cent probability of detection when the target has closed to 255 n.mi. This point is covered in more detail in Paragraph 2-12.

ready reference, the selected system parameters and the predicted system performance compared with the operational requirement are shown in Table 2-1.

Now, the problem is to find the combination of variable elements which in combination with the fixed system elements will allow the desired value of N to be achieved.

The first phase of this process is to hypothesize a specific AEW system that provides the required functional capabilities by techniques that experienced judgment deems reasonable. The specific parameters of the assumed system are then derived from the overall problem requirements.

State-of-the-art and schedule limitations are not considered in this analysis (see Paragraph 1-8). The only restrictions arise from the fixed problem elements, laws of nature, and the basic nature of the assumed AEW system concept. The latter element is variable. In an actual design study, a number of possible AEW system concepts would be examined in this manner with the object of determining which provided the best solution to the system problem. We shall investigate only one possibility to illustrate the nature of the analysis problem. The AEW system selected as an example is not intended to be an optimum solution to the AEW problem presented by the hypothetical air defense system being examined — or to any other AEW problem. It is presented only to illustrate the *types of problems* that must be considered in any AEW system design; the *form of the specification* for an AEW system; and the *nature of the interrelationships of AEW* parameters and other system elements.

TABLE 2-1 SUMMARY OF SYSTEMS ANALYSIS

Selected System Parameters	Predicted System Performance		Operational Requirement	
	Minimum Number of Targets Killed	Minimum Probability, %	Minimum Number of Targets Killed	Minimum Probability, %
Number of interceptions $N = 40$	20	55	20	55
Kill probability per interceptor $P_0 = 0.5$	16	90	16	90
Early warning detection range $R_{AEW} = 255$ n.mi.	14	100	12	100
Fixed system parameters as defined in Figs. 2-3 to 2-6	See Figs. 2-10 and 2-11		See Fig. 2-8	
Target parameters as defined in Fig. 2-7				

NOTE. Selected parameters allow the operational requirements to be met or exceeded. Sensitivity of the system performance to parameter changes are shown in Figs. 2-11 and 2-13.

2-11 AEW SYSTEM LOGIC AND FIXED ELEMENTS

A hypothetical AEW system that represents a possible answer to the air defense problem being considered is shown in Figs. 2-15 and 2-16.

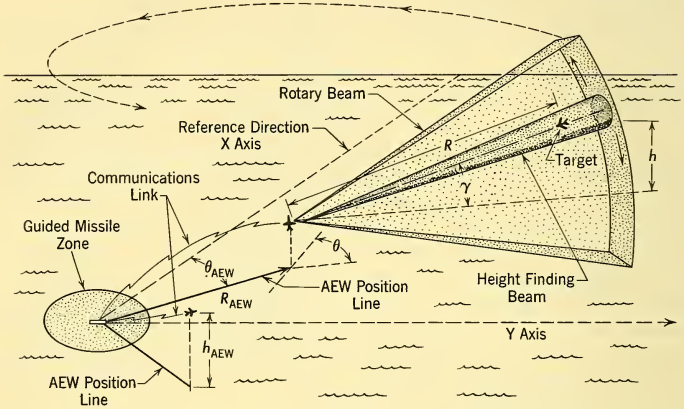


FIG. 2-15 AEW Operation Illustrating Azimuth Location and Height-Finding Means and AEW Aircraft Relations to Fleet Center.

Two interrelated airborne radars are employed in each AEW aircraft: (1) a fan beam which is rotated through 360° , and (2) a pencil beam which is nodded up and down past the target to measure height.

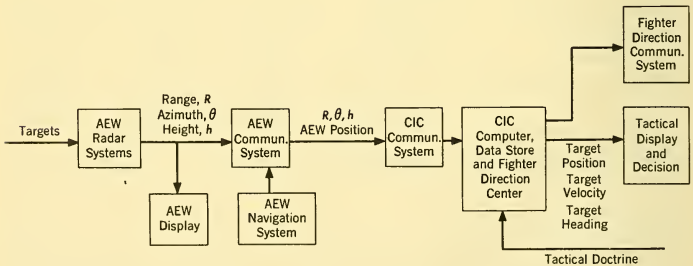


FIG. 2-16 Early Warning and Vectoring System Information Flow.

Initial detection of the target is provided by the fan-beam radar. This equipment also measures slant range to the target, R , and target azimuth position θ with respect to a reference direction.

The height-finding radar is positioned in azimuth with information obtained from the fan-beam radar. It measures the elevation angle of the target γ with respect to the horizontal. The sine of the target elevation angle, multiplied by the range R and modified by AEW aircraft altitude and an earth curvature correction, provides a measure of target altitude h . The measured target data are displayed in the AEW aircraft for monitoring purposes.

The AEW system encodes the measured range, azimuth, and height information and transmits this intelligence to the CIC in the form of a digital message. AEW aircraft position — as obtained from the navigation system — is also transmitted via the digital communications link.

The task force is provided with means for ascertaining AEW aircraft position relative to the combat information center (CIC) but with an error dependent upon the specific defense problem. A standard deviation of 1 n.mi. in both the rectangular coordinates is assumed for our analysis, as defined in Fig. 2-3.

Several AEW aircraft are employed — the number and disposition will be derived in the succeeding paragraph. The information from all AEW aircraft is presented on a master tactical display in CIC to permit overall battle control and surveillance.

Each AEW aircraft measures range and azimuth of all aircraft within its zone of surveillance. Height measurements are made only on the designated targets; the interceptors are commanded to climb to target altitude, so there is no reason (in this example) for measuring the interceptor altitude.

CIC System Information Processing. The polar coordinate (R, θ) information gathered by the AEW radars is transformed into a common rectangular (cartesian) coordinate system by the CIC computer to facilitate the generation of target heading and velocity information. Rectangular coordinates have an advantage over polar coordinates because constant-velocity, straight-line flight paths can be represented by x and y velocity components which also remain constant. Thus, if the position, $\bar{P}(t)$, of a constant velocity straight-line target at any time t is designated in rectangular coordinates, then

$$\bar{P}(t) = \bar{i}_0 x + \bar{j}_0 y \quad (2-1)$$

where x, y = target position in rectangular coordinates at time t

\bar{i}_0, \bar{j}_0 = unity vectors along the x_0 and y_0 axes of the stationary rectangular coordinate system.

The instantaneous velocity of this target may be expressed as the time derivative of $P(t)$, or

$$\dot{\bar{P}}(t) = \bar{V}(t) = \dot{\bar{i}}_0\dot{x} + \dot{\bar{j}}_0\dot{y} = \dot{\bar{i}}_0V_x + \dot{\bar{j}}_0V_y = \text{constant} \quad (2-2)$$

and the position of the target at any time τ seconds later can be written

$$\bar{P}(t + \tau) = \bar{P}(t) + \tau\bar{V}(t) = \dot{\bar{i}}_0(x + \tau V_x) = \dot{\bar{j}}_0(y + \tau V_y). \quad (2-3)$$

Thus, the computation of the velocity components and the *prediction* of future target position can be done by relatively simple means once the present position information has been transformed to rectangular coordinates.⁵

CIC Command Functions. The position and velocity information computed by the CIC is first used for purposes of assessing the threat on the basis of numbers, position, and velocity. Then, it is employed to compute a vectoring guidance course for each interceptor assigned to engage specific target aircraft. The guidance information is transmitted to the interceptor and displayed there by appropriate means.

Overall battle control is maintained by CIC using a master tactical display in combination with a pre-established operating doctrine. The tactical doctrine — target assignment, force deployment, etc. — applicable to the threat situation is formulated by the CIC officer and is used to monitor and adjust the processing of information in CIC.

The CIC computer also generates commands which are transmitted to the AEW for the purpose of designating targets for the height-finding radars.

Vectoring Guidance (Fighter Direction). The type of vectoring guidance employed is dictated by the requirements of the tactical problem and should be uniquely controlled by the weapons system requirements. In the hypothetical example, a high premium was placed on the ability to bring the interceptors into a position to fire their missiles as quickly as possible. In fact, the calculation of the number of interceptors that could engage the threat (40 for 255 n.mi. AEW range) was based on the implied assumption that each interceptor flew in a straight line from fleet center to a point where it could engage its assigned target (see Fig. 2-12).

The type of fighter direction best fulfilling this requirement is *collision* vectoring. Its basic principle is shown in Fig. 2-17. For a target at P_1 traveling with velocity V_T and an interceptor at P_2 traveling with velocity

⁵This advantage does not always lead one to choose rectangular coordinates for the processing of radar information. For example, in Paragraph 1-4 and Fig. 1-4 the use of polar coordinates is indicated.

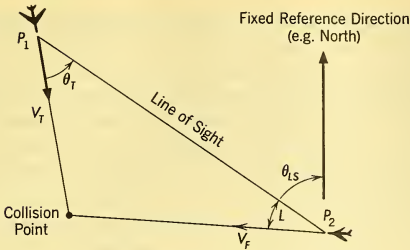


FIG. 2-17 Collision Vectoring Geometry.

V_F , the interceptor will close with the target in the shortest possible time if the following relationship is satisfied:

$$\sin L = (V_T/V_F) \sin \theta_T \quad (2-4)$$

where each of the variables may be defined from the figure.

The quantities, θ_T , V_T , and θ_{LS} are obtained from the AEW and CIC systems and transmitted to the interceptor. A computer in the interceptor uses this information along with its measurement of its own velocity V_F to calculate the proper lead angle L . Then, it adds this angle to the space line-of-sight direction, θ_{LS} , to obtain the desired space heading of the interceptor. The pilot flies the aircraft to maintain this heading and commences to search for the target with his AI radar oriented along the line of sight.

Target altitude also is transmitted to the interceptor. The interceptor climbs to this altitude using his altimeter as a reference. In the hypothetical system, vectoring guidance information is transmitted at a rate equal to the scanning rate of the AEW fan-beam radars. Vectoring guidance is continued until the interceptor acquires the target with its own AI radar.

The choice of this vectoring technique has a profound effect on the inter-related requirements of the interceptor AI radar and the AEW and CIC systems. A further description of the vectoring problem and the manner by which vectoring errors affect AI radar requirements is contained in Paragraph 2-25.

2-12 AEW DETECTION RANGE REQUIREMENTS

The foregoing discussion has established that an early warning range of 255 n.mi. is compatible with the system effectiveness goal and the assumed fixed elements of the problem.

It has been assumed that the early warning coverage need be provided only in the area of most likely attack. The operational doctrine established that a calculated risk would be taken that the defined mass attack would

not, in this case, approach the task force from the side farthest from enemy bases.

Carrier deck space and AEW aircraft cycle time limitations dictate that the required coverage be provided by a maximum of 3 AEW aircraft. Another systems consideration governs the choice of detection characteristics of the early warning radars: back-up or overlapping coverage where the loss of an AEW aircraft due to enemy action or equipment failure leaves the task force undefended.

The required detection range and AEW aircraft spacing for the assumed system may be analyzed by the simple geometrical model of Fig. 2-18.

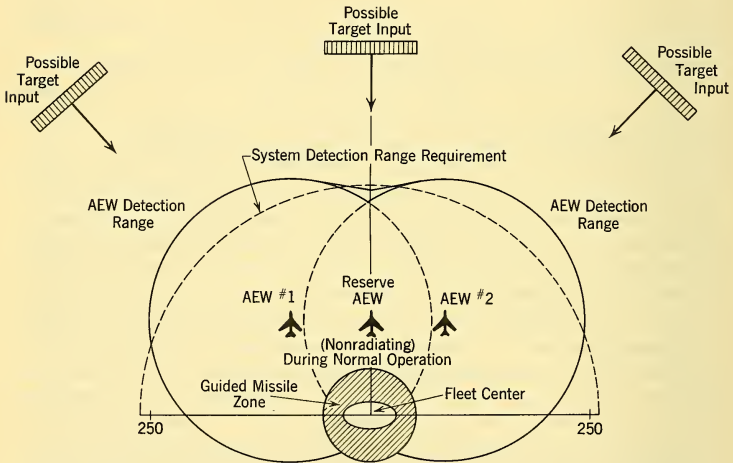


FIG. 2-18 Possible AEW Aircraft Detection Range, Coverage, and Disposition to Provide 255-n.mi. Early Warning Range.

The arrangement shown represents one possible answer to the hypothetical system requirements. This deployment shows 2 AEW aircraft, each capable of detecting enemy targets at ranges of 150 n.mi. with a 360° search sector. The two operating AEW aircraft are positioned with respect to task force center so that detection occurs at a distance of 255 n.mi. or more from task force center in the directions from which enemy raids are expected.

A third AEW aircraft is positioned as shown for use as a back-up or ready replacement for either of the other two aircraft. This aircraft does not radiate during normal operation, in order to make its detection and destruction by the enemy more difficult.

The configuration developed in this manner is one of many that could be developed as a possible problem solution. An actual study would examine a number of such configurations. This example is chosen to illustrate some of the quantitative and qualitative aspects of the system problems that must be considered in an AEW design.

The specification of detection ranges must take account of the uncertainty attending the detection process. For identical tactical situations, detection by radar will take place within a band of possible ranges, such as are shown by the distribution density function in Fig. 2-19. The proba-

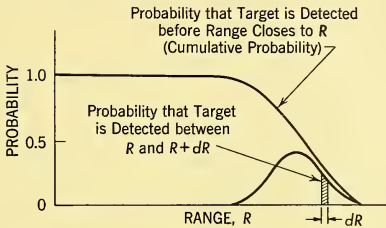


FIG. 2-19 Characteristic Forms of Radar Detection Probability Distributions and Cumulative Detection Probability Curves.

bility that the target will be at some time detected *before* it closes to a given range R —customarily called the *cumulative probability* of detection—is the integral of the distribution density function taken from R to ∞ ; its usual form is also shown in Fig. 2-19.

A preliminary requirements study such as we are performing generally expresses the radar detection requirement in terms of the range for 90 per cent cumulative probability of detection. Accordingly, the detection requirements for the AEW radars may be defined as:

Search Sector — 360° azimuth — Sea level to 50,000 ft.

Detection Range — 90 per cent cumulative probability of detection of the specified enemy targets at 150 n.mi.

2-13 AEW TARGET RESOLUTION REQUIREMENTS

A primary function of the AEW system is to provide an early description of potential targets. The description might include range, bearing, elevation, and number of targets. This information is employed in the threat evaluation phase and, later, to vector interceptors against specific targets. In both phases, the ability of the AEW system to resolve separate target elements is of fundamental importance.

Threat assessment and tactical decision must occur within 3 minutes following initial detection at 255 n.mi. from fleet center. During this time, enough information must be obtained to allow the CIC system to compute an estimate of the position, speed, altitude, heading, and number composition of a potential target complex. The specified targets can travel about 25 n.mi. in 3 minutes; therefore this information must be gathered and processed when the target is at ranges of 125 to 150 n.mi. from the AEW aircraft in order to provide sufficient problem lead time.

As a first step, the *sensitivity* of system performance to the resolution quality of airborne early warning information must be examined. Then the problem of providing the necessary resolution by appropriate AEW radar design parameters may be treated.

The defined target complex consists of 20 targets spaced 5 n.mi. apart (see Fig. 2-7). Now, consider the following problem: Suppose that during the threat assessment phase only 10 separate targets are indicated by the AEW radar information (such a condition could be caused by insufficient resolution in the AEW system — i.e. a circumstance which could cause two or more targets to appear as only one target on the radar display). What effect does this condition have upon overall system operation?

This question may be answered by considering the effect of this condition upon each phase of the air-defense operation.

First of all, the 6 CAP aircraft would be directed to engage the threat elements. Simultaneously, deck-ready interceptors would be launched at the rate of 2 per minute. To ensure high target attrition, tactical doctrine might dictate that at least 2 interceptors be employed for every potential target. This would require launching at least 14 deck-ready aircraft in response to a 10-target threat. Thus, for the first 10 minutes following initial detection (3 minutes delay time plus 7 minutes for launching 14 deck-ready interceptors), the conduct of the air battle would be in no wise different from what would have taken place if all 20 targets had been indicated initially.

During this 10-minute interval, the threat will have closed to about 175 n.mi. from fleet center. At this range the 6 CAP aircraft will engage separate elements of the raid (see Fig. 2-12). For these interceptions to be vectored successfully, at least 6 of the separate target elements must be resolved and tracked by this time.

In addition, if we assume that the number of deck-ready aircraft launched is a direct function of the number of known targets, it is necessary to begin to distinguish more than 10 objects by the time the threat has reached 175 n.mi. from fleet center (or 75 n.mi. from the AEW aircraft). In fact, to prevent delay in deck-ready aircraft launchings, the number of targets counted must increase at a minimum rate of 1 per minute until all 20 are separately resolved.

Thus, it is seen that the fact that only 10 of the 20 targets were resolved initially does not in itself degrade system performance. So long as resolution is sufficient to resolve additional targets faster than the interceptor launching rate, system performance will not be affected for the assumed tactical doctrine.⁶ A further examination would disclose that as few as 5 targets could be indicated by the initial early warning information provided that the subsequent "break-up" of targets was sufficient to keep pace with deck-ready interceptor launch rate.

The vectoring phase imposes additional resolution requirements. The assumed tactical doctrine requires that individual interceptors be directed against individual targets. Thus, both targets and interceptors must be separately resolved and tracked in this phase.

An inspection of the tactical geometry of Figs. 2-12 and 2-18 discloses that contact between targets and interceptors will take place at ranges that are seldom greater than 75 n.mi. from an AEW aircraft. The majority of contacts will be less than 50 n.mi. from an AEW aircraft. Thus, if the AEW radar resolution and the interceptor tactics are chosen to ensure that substantially all the targets and all the interceptors can be separately resolved at ranges of 75 n.mi. or less from the AEW aircraft, little or no degradation in system performance will result if at least 5 separate targets are indicated at the early warning range (150 n.mi.).

Now the foregoing tactical requirements may be translated into radar performance requirements. With a radar, it is possible to measure three quantities directly (see Paragraph 1-4) — range, angle, and velocity along sight-line to target. Resolution between targets may be done on the basis of any or all of these.

Fig. 2-20 shows a particularly difficult case that could exist for the hypothetical threat. The target threat complex is approaching along a radial line which passes through the AEW aircraft and fleet center. The location of each threat element relative to the AEW aircraft is shown in the expanded view. As can be seen, the angular differences between adjacent threat elements are of the order of 4° . The range difference between adjacent elements varies from about 2.5 n.mi. for the extreme outer threat elements to less than 1 n.mi. for the central elements. In the case of the two center elements, the range difference is zero. The differences in radial velocity components of adjacent elements vary from about 20 fps for the outer elements to 0 fps for the central elements.

From the diagram, it is seen that an angular resolution capability of 4° or less will provide the stipulated tactical capability. However, this is not the only means for meeting the requirement. A range resolution ca-

⁶This analysis does not consider the possible benefits of finer resolution to the assignment procedure. These might be significant in a practical case and should be taken into account. The analysis of this problem is too complex for consideration in this example.

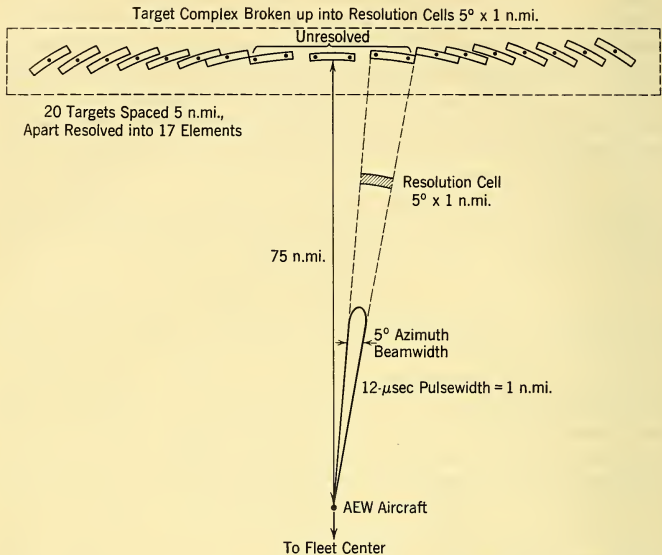


FIG. 2-20 AEW Radar Resolution Capacity at 75 n.mi.

pability of 0.5 n.mi. would permit resolution of 19 out of the 20 targets at 75 n.mi. range. (The two center targets would appear as one if range resolution were used.) At 150 n.mi., 15 targets would be seen.

Similarly, various combinations of range and angular resolution capabilities may be employed. For example, the diagram shows that an angular resolution capability of 5° coupled with a range resolution of 1 n.mi. will allow resolution of 18 of the 20 targets at 75 n.mi. range. At 150 n.mi., 15 separate targets will be indicated.

Several other factors must be considered in a practical treatment of the resolution problem. The individual threat elements will be unable to maintain perfect station-keeping with respect to each other. Errors in relative heading, velocity, and position will exist at any given time. This will cause the actual target positions and velocities to be distributed around the values shown in Fig. 2-20. We shall assume that these errors are small relative to the size of a resolution element for the example. However, if these errors were of the order of magnitude of a resolution element, they could cause substantial modification of the tactical resolution capability.

Resolution of the interceptors may be accomplished by a combination of tactical doctrine and AEW radar resolution capability. For example, the deck-launched interceptors are launched at the rate of 2 per minute. Thus, between successive pairs of interceptors there is a range difference of about 12 miles. Each pair of interceptors may be instructed to maintain a given relative spacing (e.g., 5 mi. or more). Assignment doctrine, in turn, must be adjusted to be compatible with these tactics. If these steps are taken, the AEW radar resolution capability dictated by the threat will also be adequate for resolution of the separate interceptors.

For an actual AEW design problem, many combinations of range and angular resolution would be examined for a number of different threat configurations and approach geometries. Such analysis would serve to place upper bounds on the required resolution capability and would establish the allowable trade-offs between range and angular resolution for the particular system problem. The principles and types of reasoning used for the single case examined in this paragraph could be employed for the more comprehensive analysis required for an actual design. In the example problem, it was seen that angular resolution capabilities of less than 5° and range resolutions of 0.5 to 1.0 n.mi. represented potentially useful ranges of values. The final choice will depend upon the influence of other functions and problems of the AEW radar system design.

2-14 INTERRELATIONS OF THE AEW SYSTEM, THE CIC SYSTEM, THE INTERCEPTOR SYSTEM, AND THE TACTICAL PROBLEM

Target tracking will follow detection. The tracking information is first utilized for threat assessment; later, tracking of both targets and interceptors provides the information needed for fighter direction.

Three interrelated characteristics of the AEW/CIC system are of fundamental importance in determining the contributions of this system to overall mission accomplishment.

- (1) Detection range
- (2) Accuracy
- (3) Information handling capacity (number of separate tracks, etc.)

In Paragraph 2-9, the detection range was found to be one of the critical factors in determining the level of mission accomplishment along with the individual interceptor effectiveness and the number of interceptors available for defense. Implicit in the analysis, however, were the assumptions that AEW/CIC system accuracy or data-handling capacity did not limit overall system performance. We must now determine the specific characteristics that the AEW/CIC system must possess to make these assumptions valid.

2-15 ACCURACY OF THE PROVISIONAL AEW SYSTEM

For the hypothetical problem under consideration, there are 20 targets and 40 interceptors — all of which could conceivably be in the zone of coverage of a single AEW aircraft. Thus each AEW aircraft must be capable of keeping track of 60 objects. Height measurements must be made on a maximum of 20 objects (targets only).

One facet of the accuracy problem — resolution — and its relation to the overall system problem has already been discussed in Paragraph 2-13. In addition to separating the 60 objects, the AEW/CIC system must also track each object, i.e. determine its position relative to some reference coordinate system and — for each of the attacking aircraft — its heading, velocity, and altitude. As has been described, this information is utilized to direct specific interceptors on collision courses with specific targets. The required accuracy of this guidance depends upon the characteristics of the interceptor system, particularly upon the AI radar and fire-control system. The accuracy of the AEW/CIC system determines the accuracy with which the interceptors can be vectored, and the vectoring error in turn determines the required lock-on range of the AI radar. This last factor is a critical item and may be severely limited by fixed elements of the problem and use environment. Thus the AEW accuracy can only be firmly specified *after* a study of the vectoring problem has determined the trade-off relation between vectoring error and the required AI lock-on range.⁷

Unfortunately, because of the complex interrelations between AEW/CIC system errors and vectoring errors, the analysis in Paragraphs 2-22 to 2-28 cannot be made abstractly but will require, as inputs, provisional assumptions of the AEW/CIC system design and accuracy. Thus, in this and some of the following paragraphs, we shall assume tentative values for the required AEW/CIC system accuracy and carry on our study of the preliminary design of the AEW radar on the basis of these assumptions. We should bear in mind, however, that these provisional values may lead to an unacceptable requirement for the AI lock-on range, in which case the analysis would have to be repeated for a modified AEW/CIC system design.

The AEW radar measures the relative position — azimuth and range — of a target with respect to itself once per revolution of the fan beam. The accuracy of each measurement, as it is seen in CIC, is limited by a number of factors. The most significant of these are:

1. Beamwidth
2. Range aperture (pulse length)
3. Data quantization
4. Data stabilization
5. Time delay errors

⁷An illustrative analysis of this kind has been carried out in Paragraphs 4-6 and 4-7 of Merrill, Greenberg, and Helmholtz, *op. cit.*

Measurement Errors due to Beamwidth and Range Aperture.

Referring to Fig. 2-21, if a large number of individual measurements were made on a target at point T , the measured values could be plotted as probability density distributions of azimuth and range values about the point T . For any *single* measurement, the indicated target position might lie anywhere in the region encompassed by these distributions, e.g., the point A indicated in the figure.

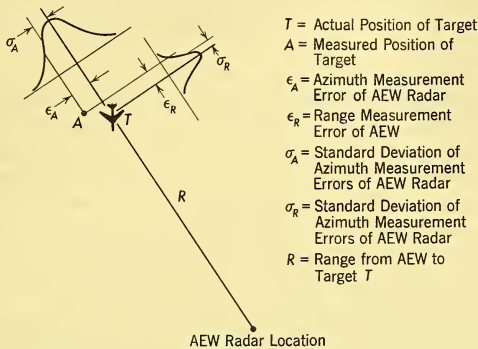


FIG. 2-21 Representation of AEW Measurement Errors.

It is convenient to describe the measurement errors by their *standard deviations* or root mean square (rms) errors. The magnitudes of these rms errors are closely related to the resolution capabilities previously discussed; however, the reader should be careful not to confuse the resolution of two targets and the accuracy in tracking one.

Accuracy describes the radar's ability to measure the position of a single target; as a rough approximation the standard deviation of the errors of a single measurement of target position may be considered to be about one-quarter of the resolution capability. Actually, besides being related to the beamwidth, the measurement error is a function of the signal-to-noise ratio and the number of hits per beamwidth. An analysis of these relations is given in Paragraph 5-11. As an example, with a 5° azimuth beamwidth and a pulse width of $12 \mu\text{sec}$ corresponding to 1 n.mi. (i.e. resolution capabilities shown to be adequate in the preceding paragraph), the approximate accuracy of a position measurement for a single scan will have rms values of

$$\sigma_A = \frac{(5)(0.25)(150)}{57.3} = 3.3 \text{ n.mi. (at 150 n.mi. range)}$$

$$\sigma_A = 1.65 \text{ n.mi. (at 75 n.mi. range)} \quad (2-5)$$

$$\begin{aligned}\sigma_R &= (0.25)(164t) \\ &= (0.25)(164)(12) = 494 \text{ yd} = 0.25 \text{ n.mi. (independent of range).}\end{aligned}$$

Several factors act to make the effective errors somewhat larger than the basic errors given in Equation 2-5.

Quantization Errors. Errors may be introduced by the process of quantizing or "rounding off" measured values of range and angle. This is done both at the AEW radar and in the CIC computer in order to minimize the amount of data which must be processed and transmitted over the associated data links.

The process may be visualized as follows. The space around a reference point (in this case either the AEW aircraft or the CIC) is broken into "cells" of arbitrary size and shape. An object anywhere in one of the cells is assigned a position description corresponding to the position of the center of the cell. This process introduces rms errors which are approximately one-quarter of the cell dimensions. These errors are independent of the measurement errors. Thus, if the quantization level (cell size) of the AEW system is chosen to be equal to the rms measurement errors (1.25° and 0.25 n.mi. in our case), the equivalent rms error in the AEW data will be increased by only about 4 per cent.

The insensitivity of the equivalent errors to the quantization level of the foregoing example shows that coarser range quantization could be employed if desired. For example, if the space around CIC were broken into cells 1 mi. on a side, the equivalent rms range error would be increased by about 40 per cent to a value of 0.35 n.mi. when the polar data from the AEW system were transformed to rectangular coordinates in CIC.

Because the angle measurement is considerably coarser than the range measurement, the 1-n.mi. CIC data quantization cells would make almost no contribution to the rms angular error data received from the AEW aircraft. Accordingly, the following quantization levels may be chosen as reasonable:

AEW: Azimuth	1.25°
Range	0.25 n.mi.
CIC: X coordinate	1 n.mi.
Y coordinate	1 n.mi.

These levels are compatible with a range resolution requirement of 1 n.mi. (Paragraph 2-13). If finer range resolution were to be employed, the CIC quantization levels would have to be reduced accordingly.

Stabilization Errors. Another important possible source of error is the rolling and pitching motion of the AEW aircraft due to maneuvers and wind gusts. It was required (Paragraph 2-11 and Fig. 2-15) that AEW

measurements be referenced to a fixed angular coordinate reference system. Errors may be induced in the azimuth measurements by aircraft motions if the AEW radar is not space stabilized. Therefore, the preliminary design must consider how AEW aircraft motions affect the AEW radar system measurements. If the effects are substantial, means must be provided for correcting the errors thus introduced. For the purpose of this analysis it will be initially assumed that stabilization errors do not degrade the angular accuracy by more than 10 per cent. The effect of this assumption may be examined in more detail when the operation of the system has been more completely analyzed and understood.

Time Delay Errors. A possible source of additional position error on a moving target is the fact that time may elapse between the measurement of target position by the AEW radar and the registration and use of this information in the CIC. If the time delay is τ_D , the amount of position error is simply

$$\epsilon = V\tau_D \text{ n.mi.}$$

where V = velocity of object being tracked.

Since the data-handling system must process information at least as fast as it is coming into the system, the maximum possible value of the time delay would be approximately equal to the time, t_{sc} , for the AEW fan beam to make a 360° scan. For example, if the scan time were 6 seconds, the maximum error against an 800 fps target caused by time delay would be approximately 0.8 n.mi. For 1200 fps interceptors, this error would be 50 per cent larger or about 1.2 n.mi.

Three courses of action are open to the designer with respect to this error. (1) The error may be tolerated if it does not appreciably affect system performance. (2) Scan speed and data processing speed may be increased to reduce the error magnitude to an acceptable level. (3) The position information may be up-dated by using estimates of velocity and heading of the object being tracked along with a knowledge of the time delay, to produce a term which cancels the time-delay error.

For preliminary design of the overall AEW system, it will be assumed that time-delay error does not increase the total position error by more than 10 per cent. The effects of this assumption upon system operation and the detailed requirements of AEW radar can be examined when more is understood about the interrelationships among various parameters of the air defense system. At that time a decision can be made about the course of action to be taken to correct the time-delay error.

Summary of Assumed Accuracy Characteristics. For purposes of analysis, the AEW radar is assumed to have the following characteristics:

Beamwidth.....	5°
Pulse length.....	12 μ sec
Scan time.....	10 sec

These characteristics coupled with assumed values for quantization levels in the AEW and the CIC systems and the assumed limits for stabilization errors and time-delay errors lead to the following accuracy characteristics for the provisional AEW system.

$$\begin{aligned}\sigma_A &= (3.3) \times (1.04) \times (1.10) \times (1.10) = 4.15 \text{ n. mi. rms} \\ &\quad \text{(at 150 n.mi. range)} \\ \sigma_A &= (1.65) \times (1.04) \times (1.10) \times (1.10) = 2.07 \text{ n.mi. rms} \\ &\quad \text{(at 75 n.mi. range)} \\ \sigma_R &= (0.25) \times (1.04) \times (1.41) \times (1.10) = 0.41 \text{ n. mi. rms} \\ &\quad \text{(unaffected by maneu-} \\ &\quad \text{vers and range).}\end{aligned}\tag{2-7}$$

The total rms position errors may be expressed as the vector sum of the range and azimuth errors or

$$\begin{aligned}\sigma_T &= (\sigma_A^2 + \sigma_R^2)^{1/2} = 4.17 \text{ n. mi. (150 n.mi. range)} \\ &= 2.09 \text{ n. mi. (75 n.mi. range).}\end{aligned}\tag{2-8}$$

One source of error — the 1 n.mi. rms navigation error of the AEW aircraft (see Paragraph 2-11) — has not been included in this analysis. This error is not significant when all of the target and interceptor tracking data come from a single AEW aircraft and when the navigation error changes vary slowly with time. When these conditions prevail, each piece of data in CIC will be biased by the same error. The relative errors between pieces of data are therefore unaffected. As we shall see, it is these relative errors that determine tracking and vectoring accuracy.

The navigation error does become important for targets and interceptors which are tracked by both AEW radar aircraft. Such an overlapping zone is shown in Fig. 2-18. A large navigational error would complicate the problem of correlating data from the same target. However, since the navigation error is less than the measurement errors and less than aircraft separation distances, no great amount of difficulty can be expected for this hypothetical case.

2-16 INFORMATION-HANDLING CAPACITY OF THE PROVISIONAL AEW SYSTEM

An important aspect of system design relates to its data-handling capabilities. Both the data link for transmitting information between the

AEW aircraft and the CIC and the data-processing computer at the CIC will have limited capacities for handling data. The maximum per channel capacity of the data link has been specified (Fig. 2-3) as 1000 bits⁸ per second.

There will be 60 objects (40 interceptors and 20 targets) in the field of the AEW radar. To identify each object requires 6 bits as shown in the footnote. The azimuth location of each target is determined to the nearest multiple of 1.25°. This will require 9 bits per object. Range information to the nearest 0.25 n.mi. from zero to 150 n.mi. requires 10 bits per object. If we add these items and multiply by 60, we find that the amount of information needed to specify the range and azimuth of the 60 objects on a single scan of the radar is 1500 bits. In addition, the elevation of the 20 targets must be determined. The accuracy and quantization level of the target altitude has not yet been specified. Here, we shall assume that target altitude is determined to the nearest 0.25 n.mi. = 1500 ft, the same as in range. To specify a target altitude from zero to 50,000 ft, then, requires 6 bits, and all the altitude data for 20 targets comprise 120 bits. The total information load on one scan, then, is $1500 + 120 = 1620$ bits. In order to incorporate self-checking codes and message redundancy in the data link to increase reliability, this figure should be about doubled. Thus, in a round figure, the data link must transmit about 3000 bits per scan to the CIC.

The actual information rate will, of course, depend upon the scan time. It is generally desirable to make the scan time relatively short in order to increase the accuracy of the heading and velocity estimates. A study in Chapter 3 indicates that the cumulative detection range tends to be relatively independent of scan time, although a broad optimum may exist. Yet the scan time cannot be made indefinitely small, because of the limitations of mechanical design and the increase in the data rate. We have chosen a provisional scan time of 6 seconds for the basic AEW radar. This radar then scans at a rate of 60°/sec. The information rate which the data link must handle is 500 bits/sec. This figure is well within the capacity of the defined data link system.

2-17 VELOCITY AND HEADING ESTIMATES

The position data are used in the CIC to compute estimates of target heading and velocity. This may be done in a variety of ways. One of the simplest can be illustrated with the aid of Fig. 2-22.

⁸A "bit" represents a binary digit, i.e. either zero or one. Transformations from decimal to binary are made in the following manner: the number 60, for example, may be expressed

$$60 = (1 \times 2^5) + (1 \times 2^4) + (1 \times 2^3) + (1 \times 2^2) + (0 \times 2^1) + (0 \times 2^0)$$

In binary form, the number 60 is the six-digit number formed by the multipliers of the powers of 2:

$$60 \text{ decimal} = 111100 \text{ binary.}$$

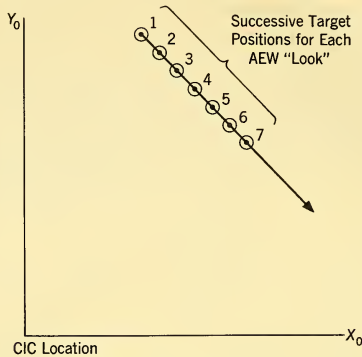


FIG. 2-22 Position Data Used in Computing Estimates of Target Heading and Velocity.

Assume that the specified 800-fps target is initially detected at point 1. Since the assumed AEW radar scanning time is 6 seconds, the next look at the target will occur when it reaches position 2. It will be at position 3 on the third look, and so on. At the end of seven looks, the target will be at position 7. The position measurement on each look is characterized by an assumed radial error with a standard deviation of σ_T . This error may be broken into components parallel and normal to the target path, where the standard deviation of each component is⁹

$$\sigma_P = \sigma_T / \sqrt{2} \quad \text{n.mi.} \quad (2-9)$$

$$\sigma_N = \sigma_T / \sqrt{2} \quad \text{n.mi.} \quad (2-10)$$

If the errors of each position measurement are assumed to be independent, the relative errors between any two measurements have the standard deviations

$$\sigma_{P_1P_2} = \sqrt{\left(\frac{\sigma_T}{\sqrt{2}}\right)^2 + \left(\frac{\sigma_T}{\sqrt{2}}\right)^2} = \sigma_T \quad \text{n.mi.} \quad (2-11)$$

$$\sigma_{N_1N_2} = \sigma_T \quad \text{n.mi.} \quad (2-12)$$

A very simple procedure for determining the target velocity and heading can be based on the extreme position measurements. The estimated

⁹Breaking the errors into equal components ignores the influence of the fact that the range and azimuth errors for a given target measurement are markedly different (Equations 2-7 and 2-8). If this factor is considered, the mathematical complexity of the problem is greatly increased; the final answer expressing the probable position errors for any randomly chosen target position relative to the CIC is not changed substantially.

velocity will be the difference of the extreme measurements divided by the total time, nt_{sc} , where n is the number of scans and t_{sc} is the scan time. The estimated direction is simply that determined by the two extreme measurements. The velocity and heading errors are expressed in terms of the parallel and normal components of the relative position errors. Thus

$$\sigma_V \doteq \frac{\sigma_{P_1P_2}}{nt_{sc}} = \frac{\sigma_T}{nt_{sc}} \quad (2-13)$$

$$\sigma_\psi \doteq \frac{\sigma_{N_1N_2}}{Vt_{sc}} = \frac{\sigma_T}{nt_{sc}} \quad (2-14)$$

where V = true velocity of object being tracked.

For example, assume that position measurements are made on an 800-fps target 75 n.mi. from the AEW aircraft. This range represents a likely value of the maximum range at which accurate tracking will be required for the generation of vectoring information as explained in Paragraph 2-13.

With the assumed parameters of the provisional AEW system, the total position error of a single measurement was derived to be Equation 2-8:

$$\sigma_T = 2.09 \text{ n.mi.} \doteq 2 \text{ n.mi.} \quad (2-15)$$

Thus, the standard deviations of the two components of the relative error between two measurements are from Equations 2-11 and 2-12:

$$\sigma_{P_1P_2} \doteq 2 \text{ n.mi. (rms)} \doteq 12,000 \text{ ft (rms)} \quad (2-16)$$

$$\sigma_{N_1N_2} \doteq 2 \text{ n.mi. (rms)} \doteq 12,000 \text{ ft (rms)}. \quad (2-17)$$

Thus the rms errors in the estimated velocity and heading are calculated by Equations 2-13 and 2-14 to be

$$\sigma_V \approx 12,000 / (6) (10) = 200 \text{ fps (rms)} \quad (2-18)$$

$$\sigma_\psi \approx 12,000 / (800) (6) (10) = 0.25 \text{ rad} = 14.5^\circ \text{ (rms)}. \quad (2-19)$$

Accordingly, we see that the accuracy of the velocity and heading estimates depends upon the following factors:

1. Accuracy of each position measurement
2. Number of position measurements used for estimates of velocity and heading
3. Elapsed time between position measurements
4. Velocity of object being tracked

In addition, accelerations of the object being tracked during the time interval, nt_{sc} , also give rise to additional errors in the position and velocity estimates.

There are several means for improving the tracking accuracy. Each of these involves a trade-off between improved accuracy and greater information-handling complexity. For instance, in the simple case we have just been discussing, when only the extreme position measurements are used all the information associated with interior measurements is lost. If all the position measurements were used to determine the velocity and heading estimates, the errors would be substantially smaller. Thus, the error estimates in Equations 2-18 and 2-19 are somewhat pessimistic. In some cases, a trade-off between maneuvering and nonmaneuvering target tracking accuracy also is required.

The velocity and heading estimates may be used in several ways. First of all, this information is used to compute vectoring guidance for the interceptors. The collision vectoring equation (Equation 2-4) illustrates a typical application. Part of the vectoring problem involves prediction of the future positions of the targets and interceptors. Prediction for a single scan is also used to update the position information. An example of such a prediction process is shown in Fig. 2-23. The AEW/CIC system indicates target position as point A. The velocity and heading estimates are used to generate a track AA^1 . On the next scan, target position is indicated as point B. The position data are corrected to this point and a new continuous track BB^1 is estimated, etc. Thus, at any time between measurements a

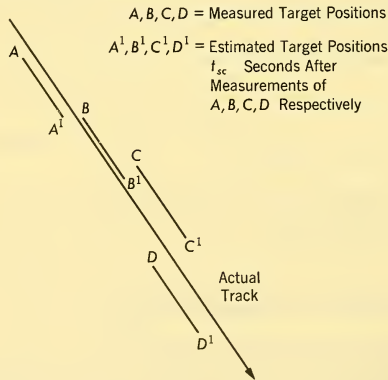


FIG. 2-23 Prediction Process Employing Scan-to-Scan Correction of Position Data.

position measurement is available which accounts for the change in target position since the last measurement was made. This type of information processing (updating) greatly reduces the time-delay error discussed in Paragraph 2-15.

The updating process has another advantage. It produces an *estimate* of target position on the next scan (see Fig. 2-23). This estimate greatly facilitates the problem of maintaining the identity of a target from scan to scan because it provides a better idea of where each target is going to be the next time the AEW radar looks at it. The heading and velocity information is used to obtain these predictions.

Prediction must be paid for, though, and the longer the prediction time, the larger the errors in the predicted position. Fig. 2-24 illustrates how the

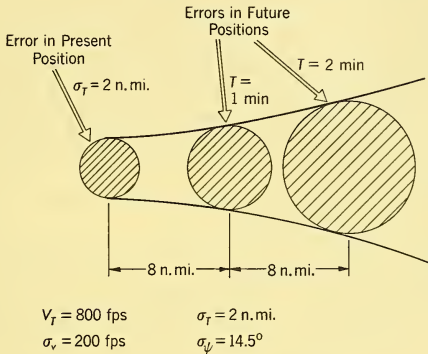


FIG. 2-24 Growth of Position Error with Prediction Time.

indeterminacy volume of the predicted position expands with the prediction time. This figure was determined on the basis of the following expressions for the future parallel and normal rms errors σ_{PF} and σ_{NF} in terms of prediction time T and the present position, velocity, and heading errors.

$$\sigma_{PF} = \sqrt{\sigma_P^2 + \sigma_V^2 T^2} \quad (2-20)$$

$$\sigma_{NF} = \sqrt{\sigma_N^2 + \sigma_\psi^2 V^2 T^2}. \quad (2-21)$$

With the same target velocity and system characteristics used previously, the position error expands from 2 n.mi. rms to 4.46 n.mi. rms with a prediction time of 2 minutes.

2-18 AEW RADAR BEAMWIDTH AS DICTATED BY THE TACTICAL PROBLEM

On the basis of target resolution requirements (Paragraph 2-13) a value of 5° was selected for the fan beamwidth of the provisional AEW radar design. Subsequent estimates of accuracy and information handling characteristics were based on this value (Paragraphs 2-15 to 2-17).

The selected value of 5° was representative of a likely allowable upper limit for the AEW radar design on the basis of resolution. The use of larger beamwidths would complicate the problem of target resolution since the angular beamwidth would then be appreciably larger than the angular separation of the targets (Fig. 2-20) at the maximum vectoring range (75 n.mi.).

Since the system accuracy is almost a direct linear function of beamwidth, the estimated accuracy of the provisional AEW design represents the poorest that might be obtained from a potentially suitable AEW design. Thus the accuracy performance characteristics of the provisional AEW design will tend to place the most severe requirements on the interceptor system. If the interceptor system can be built to meet these requirements, the same interceptor system will be more than adequate for smaller values of AEW radar beamwidth. On the other hand, if the selected value of 5° makes AI radar requirements unreasonable, the maximum permissible AEW radar beamwidth may have to be reduced. The objection to a reduced beamwidth is the larger antenna which it entails and the penalty thus imposed upon the AEW aircraft.

In this chapter, only the interrelationships of AEW radar beamwidth with the tactical problem are discussed. As will be seen in Chapter 3, radar beamwidth also enjoys close interrelationships with other parameters and performance characteristics of the radar system. Among these are (1) detection range, (2) information rate, (3) operating frequency, (4) antenna size, and (5) stabilization requirements.

In addition, AEW radar beamwidth affects the response of the radar system to electromagnetic disturbances arising in the tactical operating environment. Enemy countermeasures, radar returns from clouds and ground, and radiations from other AEW aircraft are representative of such phenomena. Strictly speaking, the consideration of these factors should be made at the same time as the resolution and accuracy requirements studies since they are an important part of the AEW radar's relationship with the overall tactical problem. For simplicity, the discussion of these factors is deferred until Chapter 14 because a knowledge of radar techniques and propagation phenomena is necessary to make such a discussion meaningful.

To summarize, then, AEW radar fan beamwidth is dictated by three primary tactical considerations: resolution, vectoring accuracy, and interaction with electromagnetic disturbances. Resolution considerations have been shown to dictate a value of about 5° or less. Vectoring accuracy requirements are unknown at the present time. In order to proceed with the problem, the vectoring accuracy obtainable with a 5° beam will be used. Subsequent analysis of the AEW and AI systems will disclose whether vectoring accuracy dictates a narrower beam. Electromagnetic disturbance

effects are unknown. The details of this problem will be largely neglected in the development of system requirements in this chapter.

Vertical Beamwidth. Vertical beamwidth also is an important factor. The AEW radar must detect and track the specified 50,000 ft altitude targets. It should also have a capability for detecting and tracking targets at all other reasonable values of altitude, since the specified threat could not be considered realistic if there were significant holes in the early warning coverage at other altitudes which could be exploited by the enemy.

The characteristics of the threat determine the required vertical coverage. If it is assumed that the primary threat (Mach 0.8, 50,000 ft) could also attack from lower altitudes — for example, 10,000 to 50,000 ft — then, AEW coverage must be provided over this range of altitudes. The coverage must be sufficient that targets are not lost for appreciable periods of time. For example, Fig. 2-25 shows that vertical coverage of 45° upward and 18.3°

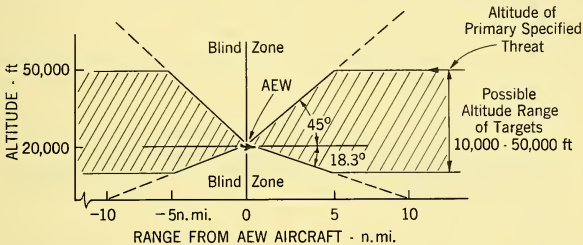


FIG. 2-25 AEW Vertical Coverage Diagram — Example.

downward can create a zone 10 n.mi. in diameter where the primary target (50,000 ft, 800 fps) can be lost from view. In the worst case, this would involve loss of the 800-fps target for a period of slightly greater than 1 minute. With the assumed target spacing — 5 n.mi. — a maximum of two targets would be within this zone at any one time.

By the time targets enter this zone, the estimates of their velocity and heading have been obtained quite accurately since they have been under surveillance for almost 150 n.mi. These estimates may be used to update the target position during the blind time, thereby reducing the effect of the blind zone on system performance. Moreover, the tracking of objects entering the zone is being done at very short ranges, and this greatly improves the position accuracy of the data obtained just before the target enters the zone. On these bases, it is reasonable to assume that dead zones of the order of 10 n.mi. do not sensitively affect system performance, since surveillance is lost for a relatively short time. Thus, vertical coverages

of the order of that indicated in Fig. 2-25 should be adequate for the tactical problem.

One other problem related to vertical beamwidth is of great importance to the AEW problem, namely ground and sea return. Fig. 2-25 shows that the fan beam intersects the surface of the water — or land, as the case may be — at all ranges greater than 10 n.mi. Thus reflections from the ground will compete with target signal reflections at all ranges greater than 10 n.mi. This fact requires that means be provided in the AEW system to distinguish between returns from the surface of the earth and target returns.

Interactions between the radar, the target, and the ground constitute a very complicated problem. The polarization of the radar transmission, surface characteristics, and AEW and target altitudes all interrelate to produce nulls and reinforcements which influence the system capability. These factors are discussed in some detail in Chapter 4.

In an actual design study, the quantitative aspects of this problem should be carefully studied and set forth at this point of the systems requirements development. The interrelations of the tactical geometry and propagation and scattering characteristics must be ascertained to define the magnitude of the problem implied by the requirements for distinguishing between ground and target returns.

2-19 FACTORS AFFECTING HEIGHT-FINDING RADAR REQUIREMENTS

The height-finding radar for the example problem is positioned in the aircraft nose. It is directed to point in a given azimuth direction at a target located by the fan-beam radar. It is then nodded up and down to determine the elevation of the target with respect to a horizontal reference in the AEW aircraft. The nodding action causes the target return to vary as a function of the space (or angle) modulation characteristic (elevation) of the height-finding beam. The particular type of space modulation characteristic that is used depends upon the accuracy requirements of the height finder.

The requirements of the height-finding radar are dictated primarily by the following tactical considerations:

1. The characteristics of the expected threat including possible variations from the specified values. These characteristics include speed, altitude, and number of aircraft.
2. Height-finding requirements during threat evaluation.
3. Height-finding requirements during vectoring.
4. Height-finding requirements dictated by the need to supply early information to the ground-to-air missile system.

To meet these requirements within the limitations of the hypothesized

system logic, the height finder must operate with azimuth input commands obtained from the AEW fan-beam radar.

The possible dimensions of the height-finding radar are limited by the dimensions of the AEW aircraft nose. For purposes of specification, it will be assumed that the aircraft in the example may accommodate an antenna with a maximum dimension of 3 ft. It is also assumed that such an antenna may be gimballed so as to be capable of performing the height-finding function on objects within a $\pm 80^\circ$ horizontal zone around the AEW aircraft's nose. The vertical coverage of the height finder must be matched to the primary vertical pattern coverage of the fan-beam azimuth search system.

The assumed placement of the height-finder places a limitation on tactical usage. To evaluate a target the AEW aircraft must be pointed within $\pm 80^\circ$ of the line of sight to the object whose height is being measured. Thus, by virtue of the assumed system logic, the AEW aircraft must maneuver to perform its mission. The required maneuver must be within the performance characteristics of the AEW aircraft. In addition, the *effect* of the maneuver upon the stabilization problem must be evaluated.

Requirements Dictated by Threat Evaluation. Height-finding information need not be obtained at the same rate as position information for the specified threat, since its altitude does not change during the attack. In fact it need be measured only once during the specified attack. Once again, the possibility of other attack situations must be considered. If the enemy aircraft were capable of making an abrupt altitude change during the attack, the height-finding system must be able to detect such a change in time for appropriate defensive measures to be taken.

Immediately following detection, the height-finding radar is required to begin measuring target altitudes for purposes of raid evaluation. Ideally, the evaluation of target altitudes should take place within the time allowed for threat evaluation (assumed to be 3 minutes in the example). If we allow an average time of 1 minute¹⁰ for the AEW aircraft turning to face the raid, a total time of 2 minutes is available to measure target altitudes

¹⁰The AEW aircraft speed is 200 knots. At this speed, and at a bank angle of 10° , the AEW aircraft can turn at the rate of about $1^\circ/\text{sec}$.

$$\dot{\psi} = 109 \frac{\tan \phi}{V}$$

where V = velocity in knots, ϕ = roll angle (degrees), and $\dot{\psi}$ = horizontal turning rate ($^\circ/\text{sec}$). Since the height-finding radar coverage is $\pm 80^\circ$ or 160° , the maximum turn required to bring a target under height-finder surveillance is 100° . Thus, 100 seconds would be required in the worst case for a 10° bank angle. On the average less than half this time would be required, since the orientation of the AEW aircraft relative to the raid is random. Thus, the assumption of 1 minute does not imply extreme maneuvers by the AEW aircraft.

and transmit this information to CIC for decision and target assignment. Since there are 20 targets, a maximum time of 6 seconds per target is permissible. During this phase, it is sufficient to know whether the targets are high, medium, or low altitude.

Requirements Dictated by Vectoring. During vectoring, interceptors are vectored to the measured altitudes of the assigned targets (for the assumed system logic of the hypothetical example). A height-finding error can limit system performance in the following ways.

1. It can cause the interceptor to fly at an unnecessarily high altitude, thereby degrading speed and maneuvering capability.
2. It can cause the interceptor to approach the target with an altitude differential which its weapon (assumed to be a guided missile) cannot overcome.
3. It increases the zone of probable target positions which must be searched by the AI radar.

The first limitation can be attenuated somewhat by the use of tactical doctrine based on a prior knowledge of threat characteristics. For example, if the probable threats are known to have a performance ceiling of 50,000 ft, there would be little point in directing the interceptor to fly at 60,000 ft even though the height finder indicated such an altitude.

The second limitation must be related to weapon characteristics and aircraft and fire-control system characteristics. An inspection of the missile performance (Fig. 2-6) shows that the weapon can itself correct substantial altitude errors by its maneuvering capability. For a weapon travel of 3.2 n.mi. or more, altitude errors 0 to 2 n.mi. can be corrected if the weapon is fired horizontally. A further attenuation of the effects of altitude can be obtained from the fire-control system. Following AI radar lock-on, the pilot obtains a reasonably precise measurement of relative target elevation. This may be used by the fire-control system to point the aircraft up or down as required to eliminate an elevation error. Of course, the required climb or dive angle must be compatible with aircraft performance characteristics.

The third limitation — the required AI radar search zone needed to encompass the height-finding inaccuracies — is also most important. As will be demonstrated later, the range performance of a radar system is strongly influenced by the volume it must search.

On the basis of these considerations, a height-finding error of approximately 0.5 n.mi. (3000 ft) standard deviation at a range of 75 n.mi. represents a reasonable first approximation to the height-finding accuracy requirement. This corresponds to a maximum error of about 1.5 n.mi. — a value which is still within the guided missile performance capabilities.

It will be assumed that the same height-finding information rate (one measurement on each target aircraft every 2 minutes) will be maintained during vectoring.

Requirements Dictated by the Surface-to-Air Missile System.

Height-finding data can be used to direct the search and tracking system associated with the surface-to-air missiles to those regions of the airspace where targets are most likely. For the system of the example, such information can be most useful, since the primary target for the ground-to-air missile system is a missile launched from the hostile aircraft at a range of about 50 n.mi. The relatively smaller size of the missile makes knowledge of where to look for it most desirable. For such an operation, provision must be made for the proper transfer of data within the CIC system. It is not likely, however, that the requirements of this function are more severe than the interceptor vectoring height-finding requirements. For the purposes of the example, this will be assumed to be the case. Once again, this is an area which deserves more detailed scrutiny in an actual design study.

Requirements Dictated by the Stabilization Problem. Height-finding — even when the requirements are as coarse as indicated for the hypothetical problem — involves measuring rather small angles. The assumed system logic requires that elevation angle of the target be measured with respect to the horizontal plane. In addition, the height-finder must be commanded to the measured space azimuth position of a particular target.

Some idea of the problem may be obtained by translating the derived 0.5-n.mi. rms interceptor vectoring height-finding requirement into an equivalent angle for 75 n.mi. range. This angle may be expressed

$$\sigma_h = \frac{0.5}{75} = 0.067 \text{ rad} = 0.38^\circ \text{ (rms)}. \quad (2-22)$$

This is a *total* error — including the accuracy of the radar, the stabilization errors, mechanization errors, and quantization errors. If the latter two errors are assumed negligible and if the stabilization and radar measurement rms errors (σ_{hs} and σ_{hm} respectively) are assumed equal, normally distributed, and independent then

$$\sigma_{hm} = \sigma_{hs} = 0.38 / \sqrt{2} = 0.27^\circ \text{ rms}. \quad (2-23)$$

Thus, to meet the height-finding requirement of 75 n.mi., the height-finding system must be stabilized to within 0.27° of true vertical. This accuracy must be maintained despite aircraft pitching or rolling motions.

In addition, the azimuth beamwidth of the height-finding radar must be large enough to include the uncertainty of the azimuth fan beam. It should

not be appreciably larger than that of the fan beam or it will have difficulty resolving between adjacent targets that are resolved by the fan beam. Thus, as a first approximation, the azimuth beamwidths of the height finder and the fan-beam radar may be made approximately equal.

The elevation beamwidth depends upon the required accuracy. An approximation may be obtained by using the same expression employed for fan-beam azimuth accuracy (Paragraph 2-15).

$$\sigma_n \doteq \Theta_n/4 \text{ degrees (rms) for a single measurement} \quad (2-24)$$

$$\sigma_n \doteq \Theta_n/4\sqrt{n} \text{ degrees (rms) for } n \text{ measurements} \quad (2-25)$$

where n = number of measurements averaged to obtain a single estimate

Θ_n = elevation beamwidth of height-finding radar.

Since 6 seconds can be taken for the height-finding measurement, it is reasonable to assume that five to ten separate measurements could be made. If, e.g., nine measurements are made, the required beamwidth is found by a manipulation of the above equation as

$$\Theta_n = 4\sqrt{n} \sigma_n = (4) (3) (0.27) = 3.2^\circ. \quad (2-26)$$

Actually, techniques known as *beam splitting* can be employed to obtain greater angular accuracy than is implied by Equation 2-26. Accordingly, the derived result is only one of the possible solutions to the height-finding problem.

2-20 SUMMARY OF AEW SYSTEM REQUIREMENTS

The preceding discussions have shown some major considerations involved in the design of a typical AEW system. Numerical examples illustrated the various interrelations and were chosen in such a manner as to be applicable to the solution of the hypothetical air defense problem we have been considering.

We may now use all of this information to compile an estimate of the basic characteristics of an AEW system which represents a reasonable answer to the overall system problem. These estimated characteristics may then be employed to provide the basic input data needed to specify the AI radar and fire-control system. All during this process, we estimate — as best we can — the overall system performance to ensure that we do not depart from the mission accomplishment objectives. As already mentioned, in an actual overall systems study, we would repeat this process several times to obtain a better feeling for the trade-offs between the AEW system and the AI system. However, for all cases, the basic considerations and the method of attack on the problem would remain very much the same; only the assumed system logic and specific parameter values would undergo appreciable change.

One important characteristic of a systems problem has been *implied* by the foregoing discussion; it is worth mentioning explicitly at this point to impress the reader with its importance. In a systems study designed to derive basic system requirements, it is often necessary to make arbitrary decisions on the basis of incomplete quantitative results. The system logic described in Paragraph 2-11 for the hypothetical AEW system is an example of such a decision. Many choices could have been made; however, in order to get on with the problem *one* choice had to be made and then followed to its logical conclusion. Conceivably, it would develop that this was the wrong choice, in which case we should have to repeat the entire process for a more satisfactory initial hypothesis.

Keeping in mind the provisional nature of a system specification at this stage of the analysis, we may specify the basic parameters of the AEW system as in Table 2-2. The number of the paragraph which discusses each parameter is included for convenience.

TABLE 2-2 TENTATIVE AEW RADAR PARAMETERS

<i>Detection Range</i>	90 per cent probability of detection at 150 n.mi. (Paragraph 2-12)
<i>Number of Targets</i>	20 hostile targets, 40 interceptors (Paragraph 2-13)
<i>Threat Evaluation Range</i>	125-150 n.mi. (Paragraph 2-11)
<i>Nominal Vectoring Range</i>	75 n.mi. (Paragraph 2-13)
<i>Azimuth Coverage</i>	360° (Paragraph 2-11)
<i>Elevation Coverage</i>	45° up, 18.3° down. Operation at 20,000 feet (Paragraph 2-18).
<i>Range Resolution</i>	1 n.mi. (Paragraph 2-13)
<i>Angle Resolution</i>	5° maximum (Paragraph 2-13)
<i>Range Accuracy</i>	$\sigma_A = 0.25$ n.mi., rms (Paragraph 2-15)
<i>Angular Accuracy</i>	$\sigma_A = 1.25^\circ$ rms (Paragraph 2-15)
<i>Quantization Levels</i>	(Paragraph 2-15)
AEW Azimuth	1.25°
AEW Range	1 n.mi.
CIC System	1.0 n.mi.
<i>Stabilization Errors</i>	Less than 10 per cent of measurement error (Paragraph 2-15)
<i>Time Delay Errors</i>	Less than 10 per cent of measurement error (Paragraph 2-15)

Total System Position Error (Paragraph 2-15, Equation 2-8)

At 75 n.mi.

$$\sigma_T \doteq 2 \text{ n.mi. rms } (5^\circ \text{ beam}).$$

At 150 n.mi.

$$\sigma_T \doteq 4.2 \text{ n.mi.}$$

Height-Finding Radar (Paragraph 2-19)

Threat Evaluation Range 125-150 n.mi.

Nominal Vectoring Range 75 n.mi.

Azimuth Coverage $\pm 80^\circ$ from aircraft nose

Elevation Coverage 45° up, 18.3° down

Beamwidth — Elevation 3.2°

Height Finding Error 0.5 n.mi. rms or 3000 ft rms

Beamwidth — Azimuth Approx. 5° , to match fan beamwidth

Stabilization Data stabilized to within 0.27° rms of true vertical

2-21 EVALUATION OF TENTATIVE DESIGN PARAMETERS WITH RESPECT TO THE TACTICAL PROBLEM

We have discussed the general problems of AEW radar design; we also have hypothesized an AEW System which provides answers to certain of these problems (detection range, resolution, target counting, etc.). Now we shall hypothesize reasonable means for processing the radar information to provide heading and velocity information in the tactical environment of the example. The accuracy of the heading and velocity estimates obtained — coupled with the position accuracy — form inputs for the study of the interceptor system effectiveness.

Three minutes (180 seconds) are available to evaluate the threat following detection. From Equations 2-13 and 2-14 we see that the standard deviations of the heading and velocity measurements obtained by using the position measurements made at the beginning and end of the 3-minute interval are

$$\sigma_{VT} \doteq \frac{4.2 \text{ n.mi.}}{180 \text{ sec}} = 0.0233 \text{ n.mi./sec} = 142 \text{ fps rms} \quad (2-27)$$

$$\sigma_{\psi T} = \frac{(4.2)(6080)(57.3)}{(800)(180)} = 10.1^\circ \text{ rms} \quad (2-28)$$

where 6080 = conversion factor between knots and fps

57.3 = conversion factor between radians and degrees.

This accuracy is sufficient to provide a basis for evaluating the threat within the 3-minute period. Actually, the accuracy is somewhat better than is indicated by these figures. As already mentioned, the range resolution capability of the radar allows fifteen of the twenty targets to be resolved at

the range of required detection. The remaining five would appear as one or two large targets until they reached a range close enough to the AEW radar to be resolved. Tracking can begin on each of the targets indicated by the AEW radar. The average standard deviations of the raid considered as a whole would tend to approach the standard deviations of one track divided by the square root of the number of separate target tracks.

By the time the targets have closed to 75 n.mi., it is possible to make further refinements in the measurements of target velocities. Two additional 3-minute intervals are available for this purpose. Neglecting the decrease in position error for each interval and considering that the measurement made in each interval is independent of the previous measurement, the error can be reduced by the square root of 3 by averaging the three readings taken over a 9-minute period. This process yields an error of

$$\sigma_{VT} = \frac{142 \text{ fps error}}{\sqrt{\text{No. of velocity measurements}}} = \frac{142}{\sqrt{3}} = 82 \text{ fps} = 49 \text{ knots.} \quad (2-29)$$

Smoothing times consistent with this magnitude are allowable for velocity measurements because it is not reasonable to expect large changes in target velocity.

A somewhat different situation attends the measurement of heading. The target can make heading changes at a maximum rate of 3° per second. Thus, it is not desirable to use long smoothing times for heading information. In fact, a major problem in the design of the data-processing system is to choose an observation time and smoothing technique for heading information that provide a satisfactory compromise between maneuvering and nonmaneuvering targets. This is a complicated problem which cannot be considered here in detail. However, the basic nature of the problem will be indicated.

The development so far has considered the very simplest type of heading measurement; the target position is measured at two different times, t and $t + nt_{sc}$, and the heading is determined by the direction of the straight-line passing through these points (Fig. 2-22).

At a range of 75 n.mi., with an observation time nt_{sc} equal to 60 seconds, this technique gives rise to an error (Equation 2-19) in measured heading equal to

$$\sigma_{\psi_T} = \frac{(2)(6080)(57.3)}{(800)(60)} = 14.5^\circ \quad (2-30)$$

where the constants 6080 and 57.3 have been previously defined.

Now, let us assume a scan time t_{sc} of 6 seconds. The heading of the previous expression was calculated on the basis of information obtained from two scans, which we may relate to each other by calling the first scan number 1, and the second scan, occurring 60 seconds later, number 11. A

similar computation may be made using scan number 2 and scan number 12. If the errors for the two computations are independent, we may improve the heading approximation by $\sqrt{2}$ by simply averaging the two computations to yield for a straight-line target:

$$\sigma_{\psi T} = 14.5/\sqrt{2} = 10.2^\circ. \quad (2-31)$$

Such improvement is obtained at the expense of increased dynamic lags when the target maneuvers.

Many other smoothing schemes could be used. However, for present purposes, it is reasonable to assume that the hypothetical AEW system can provide heading information with an accuracy of the order of 10° (standard deviation). The suitability of this figure will depend upon the sensitivity of interceptor performance to this figure.

From the foregoing analysis, the accuracies of the AEW radar system with which the interceptor system must be compatible are approximately

$$\text{Position Error:} \quad \sigma_T \doteq 2 \text{ n.mi. radial error, rms} \quad (2-32)$$

$$\text{Velocity Measurement Error: } \sigma_{VT} \doteq 50 \text{ knots rms} \quad (2-33)$$

$$\text{Heading Measurement Error: } \sigma_{\psi T} \doteq 10^\circ \text{ rms} \quad (2-34)$$

For the vectoring problem we are interested in the relative position inaccuracy between the interceptor and the target. The total relative radial position error, σ_{RT} , between two objects is

$$\sigma_{RT} = \sqrt{2} \sigma_T = 2.8 \text{ n. mi.} \quad (2-35)$$

As will be shown later (Paragraph 2-25) it is convenient to express the total relative position error in terms of two components: (1) a component σ_{RR} along the line of sight between target and interceptor and (2) a component, σ_{Ra} , normal to the target sight-line,

$$\text{where} \quad \sigma_{RR} \doteq \sigma_{RT}/\sqrt{2} \doteq 2 \text{ n.mi.} \quad (2-36)$$

$$\sigma_{Ra} \doteq \sigma_{RT}/\sqrt{2} \doteq 2 \text{ n.mi.} \quad (2-37)$$

The position, velocity, and heading information is employed to vector interceptors on a collision course with assigned targets (Paragraph 2-11 and Equation 2-4).

2-22 INTERCEPTOR SYSTEM STUDY MODEL

The design goal for the kill probability of a single interceptor has been derived as 0.5. We shall now study the problem of specifying the requirements of an airborne intercept (AI) radar and fire-control system that will allow the interceptor to achieve this goal within the limitations imposed by other system elements and the operational environment (Step 3, Fig. 2-2).

As before, the first step is to formulate a master plan for the analysis. This master plan shows the fixed and variable elements of the interceptor system problem; it must also show the method by which the problem can be handled on a step by step (suboptimization) basis without losing the relation of each step to the overall problem. Such a master plan is shown in Fig. 2-26. It is merely a variant of the plans for steps 1 and 2 showing the details of the interceptor weapons system analysis.

The system effectiveness goal P_0 , and the fixed elements of the system, including AEW and vectoring system characteristics, have been derived or defined in preceding analyses. These are shown in Fig. 2-26 as providing the effectiveness criteria and inputs for the interceptor system analysis.

The output of the system model is P_0 (achieved). The variable elements are manipulated in such a manner as to make P_0 (achieved) equal P_0 (required). The combinations of variable element values for which this condition is realized form the basis for the interceptor system specification.

The separate steps of the interceptor system analysis can be derived from the basic system logic and a careful consideration of the factors affecting each phase of interceptor system performance. The interceptor reaching the defense zone goes through three discrete phases in attacking a target (see Fig. 2-9): (1) a *vectoring phase* which terminates in AI radar lock-on, (2) a *tracking phase* which terminates in weapon launch, and (3) a *missile guidance phase* which terminates in the destruction of the target.

The performance in each phase of operation may be characterized by the probability that — for a given set of fixed and variable elements — the phase will be completely successful. These probabilities and the factors¹¹ which determine their values are shown in Fig. 2-20 as:

P_m = probability that the two-missile salvo will kill the specified target (already specified as 0.75)

P_c = probability that the interceptor will proceed from the point of AI radar lock-on to a point where the missile salvo may be launched with a kill probability of 0.75

P_v = probability that the vectoring system will operate to bring the interceptor to a position and orientation where it may detect, identify, and lock on the target with its own AI radar.

¹¹Only the most significant factors are shown for this hypothetical example. An actual analysis might include many more. However, the same basic model would be applicable and the approach to the problem — though more complicated mathematically — could be much the same as will be used for this hypothetical example. Another important fact: Often it is difficult to establish all of the vital factors affecting a given phase of system operation — some of these are products of the analysis itself. This model has considerable flexibility in that such additions can be made by simply reanalyzing the phase(s) affected.

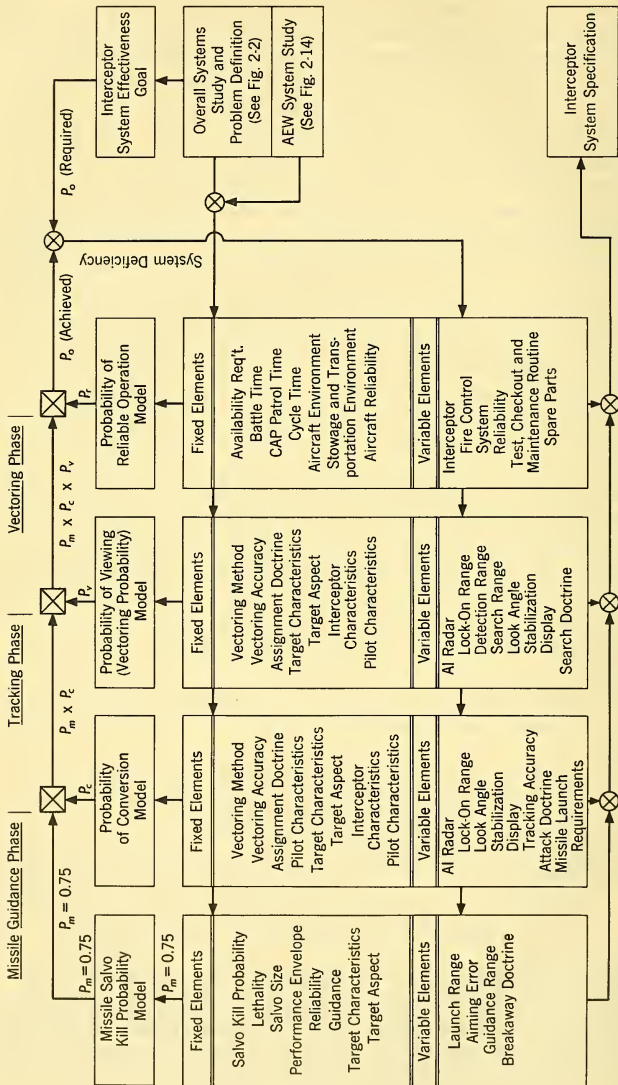


FIG. 2-26 Master Plan for Interceptor System Analysis.

Throughout all of the phases of operation, an equipment failure can cause the interception to fail. To account for this, we define a fourth probability:

P_r = probability that the interceptor system equipment (fire-control system, aircraft, communications, etc.) will operate satisfactorily until weapon impact on the target.

The interceptor kill probability P_0 may be defined as the likelihood that the *complete sequence* of events will be completed successfully for any interceptor operating under the expected tactical conditions. Mathematically, this statement has the form:

$$P_0 \text{ (achieved)} = P_m \times P_c \times P_v \times P_r. \quad (2-38)$$

Thus, the basic model is established. We will now demonstrate how quantitative models may be derived and manipulated for each phase of operation to produce specifications for the variable elements of the interceptor weapons system (AI Radar, Computer, Display, and Missile Guidance Tie-in). First, we make an *estimate* of the expected contribution of each phase of system operation to the overall kill probability. For instance, in the hypothetical example, we may substitute specified input values in Equation 2-38 and write

$$P_0 = 0.50 = 0.75P_cP_vP_r \quad (2-39)$$

or

$$0.667 = P_cP_vP_r.$$

Any combination of P_c , P_v , and P_r which yields this result will satisfy the requirement. For preliminary design purposes, we shall select one of the possible combinations to provide a criterion for the performance of each phase:

$$\begin{aligned} P_r &= 0.85 \\ P_v &= 0.95 \\ P_c &= 0.825 \end{aligned} \quad (2-40)$$

Note that $P_0 = (0.85)(0.95)(0.825) = 0.50$.

This choice is somewhat arbitrary. In an actual analysis a number of different combinations might be assumed to establish trade-offs between the contributions of each phase of system operation.

2-23 PROBABILITY OF RELIABLE OPERATION

The analysis of the tactical situation showed that the total air battle lasted less than one-half hour. The total number of interceptors in the air

during a battle is 48 — 12 combat air patrol and 36 deck launch — out of a total complement of 66. The combat air patrol interceptors maintain station for 2.8 hours.

These considerations coupled with the interceptor kill-probability goal are the principal factors that determine required interceptor system reliability during the attack operational situation.

The AI radar and fire-control system may be expected to be the primary contributors to interceptor system unreliability — recognizing that the guided missiles' reliability has already been included in the specified missile kill probability. On this basis, we shall assume for purposes of specification that failures in the AI radar and fire control will cause two-thirds of the aborts due to equipment failure. Since we specified the overall reliability of the interceptor system as 0.85, the reliability requirement for the AI radar and fire-control system is 0.90.

A reliability requirement has little meaning unless the element of time is included. Based on the large number of CAP interceptors that must be kept continuously aloft, it is specified that the reliability requirements shall be met for any 3-hour operating period. Chapter 13 will discuss the implications of this requirement, the type of design techniques that must be employed to meet it for the defined environmental conditions, and the means for determining whether a given radar can meet such a requirement.

2-24 PROBABILITY OF VIEWING TARGET — VECTORING PROBABILITY

The study plan — as extracted from the master plan of Fig. 2-26 — is shown in Fig. 2-27. The object of the study is to derive the combinations of the variable elements that will permit achievement of the assumed performance goal and to ascertain the sensitivity of vectoring probability performance to changes in the system parameters.

From Fig. 2-26 it can be seen that several variable factors — notably lock-on range and look-angle (maximum gimbal angle) — are common to conversion and vectoring probability. Accordingly, we cannot develop firm requirements for these in this phase of the study. Rather, the results will be expressed as a spectrum of possibilities, all of which satisfy the viewing probability requirement. Later we shall determine the portion of these possibilities which also satisfy the conversion probability requirements.

Search for the target and its detection obviously must precede AI radar lock-on. Thus, these factors are functions of the lock-on range and cannot be specified until lock-on range is specified.

AI radar search data display and stabilization and search doctrine are dictated almost entirely by vectoring phase considerations. Thus, these may be specified by the analysis of the viewing probability problem.

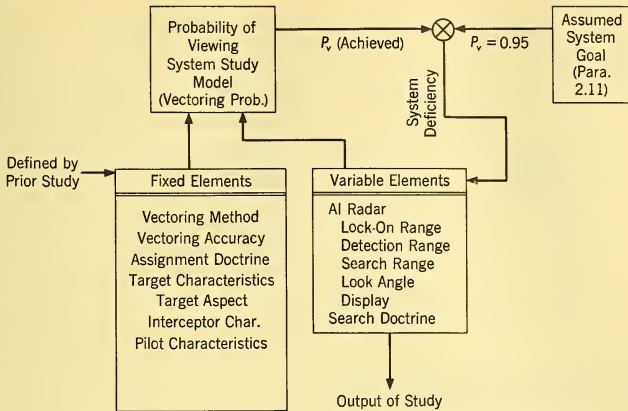


FIG. 2-27 Plan for the Study of Viewing (Vectoring) Probability.

One of the fixed elements of the problem, target aspect, deserves some discussion preparatory to the systems analysis. The target aspect or angle off the target's nose at the beginning of vectoring is a function of the geometry of the attack situation. Primary emphasis is placed on forward hemisphere attacks; the first twenty interceptors are vectored into such attacks on the twenty targets. The remaining interceptors are sent to back up the first twenty. Some of these will be initially vectored to targets that are destroyed by earlier interceptors. In such cases, the interceptor will be assigned to a new target in order to utilize fully the total interceptor fire power. These attacks may require the interceptor to approach the target on the beam or from the rear hemisphere. In addition, some of the forward hemisphere attacks will be aborted before missile launching because of a failure to see the target or to make the proper conversion. In such cases, the interceptor can turn around and employ its speed advantage to attack one of the targets from the rear. These considerations indicate that *all* initial angles off the target's nose must be considered. The interceptor kill probability should be realized or exceeded for all possible approach angles; i.e., the interceptor should have "around the clock" capability.

Of paramount importance to both the vectoring and conversion phases is the manner in which the fixed and variable problem elements combine to produce distributions of possible aircraft headings at any point in space. We may visualize this problem from Fig. 2-28. At any selected point (R, θ) relative to the target, the uncertainties of the vectoring system may cause the heading of an interceptor passing through that point to assume any value within the bounds shown. The spectrum of possible headings usually

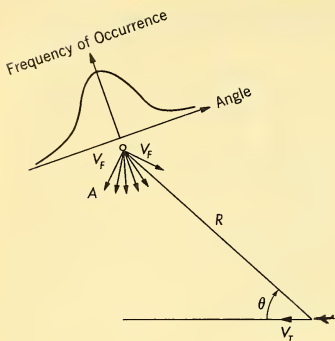


FIG. 2-28 Distribution of Interceptor Headings Due to Vectoring Errors.

2-28 were along line OA , a look-angle of approximately 90° would be required for the AI radar to "see" the target.

2-25 ANALYSIS OF THE VECTORING PHASE OF INTERCEPTOR SYSTEM OPERATION

Analysis of the vectoring phase must yield the following information:

- (1) The distributions of aircraft headings as functions of lock-on range and angle off the target's nose
- (2) The AI radar characteristics required for compatibility with the operation of the vectoring phase; i.e. display requirements, stabilization requirements, look-angle requirements.

Vectoring System Logic. The flow of information and allocation of function for the vectoring system are shown in Fig. 2-29. This diagram expresses the system logic outlined in Paragraph 2-11 for the assumed AEW/CIC system.

System Configuration Parameters. The basic factors governing the operation of the vectoring system may be ascertained from preceding definitions of target inputs and fixed parameters and the design objectives established for the vectoring system. These factors are summarized in Table 2-2 and Paragraph 2-21.

Collision vectoring was specified to minimize the average target penetration. The equation defining this vectoring method was derived as

$$\sin L_D = (V_T/V_F) \sin \theta \quad (2-4)$$

enjoys approximately a normal distribution about some mean heading as indicated. If the transition from the vectoring phase to the tracking phase is made at this point (AI radar lock-on), this distribution defines the range of initial conditions for the conversion phase. In addition, for any point in space the distribution defines the likely angular positions of the target with respect to the interceptor flight path. The maximum look-angle required for the AI radar is largely determined by this consideration coupled with the viewing probability requirement. For example, if the interceptor heading in Fig.

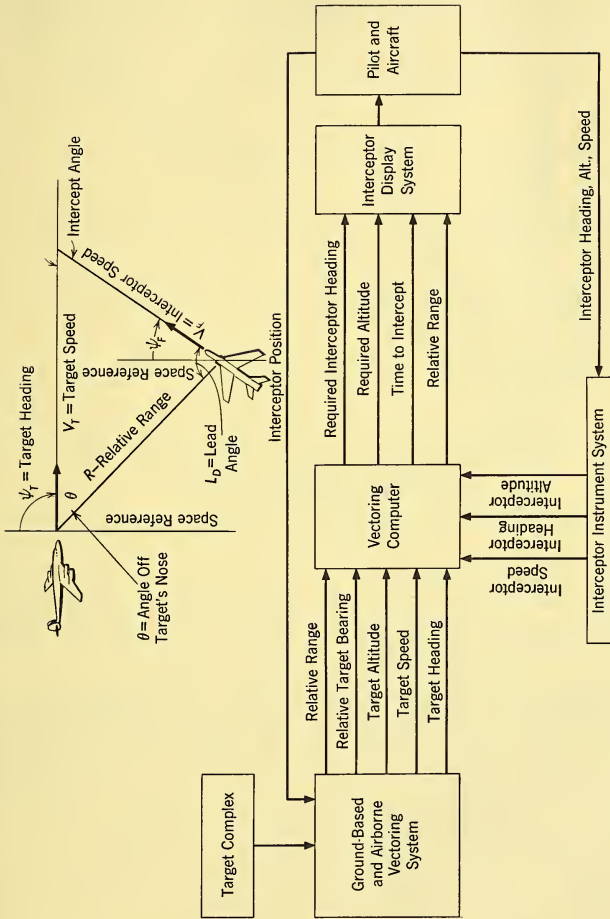


Fig. 2-29 Block Diagram of Vectoring System Logic.

where L_D = the collision course lead angle for perfect collision vectoring

θ = the angle off the target's nose.

A plot of the required collision course lead angle versus *angle* off target's nose is shown in Fig. 2-30. The vectoring system computes this angle from

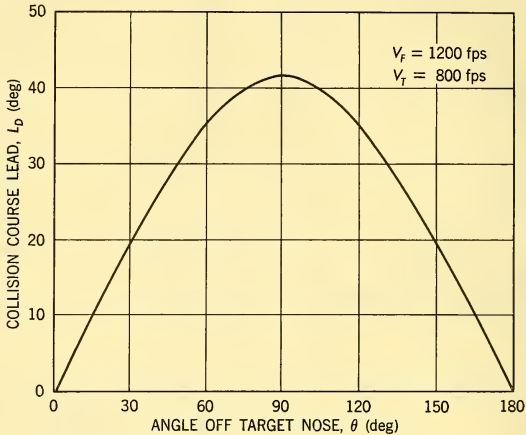


FIG. 2-30 Collision Course Lead Angles Versus Angle off Target's Nose.

the AEW radar measurements. It transforms this lead angle into a space heading command which is transmitted to the interceptor. The pilot flies the aircraft so that the heading as measured by the aircraft compass corresponds to the vectoring system heading command.

Distribution of Aircraft Headings due to Vectoring Errors.

Because of errors in the vectoring system measurements, the commanded heading does not always correspond to the correct collision-course lead angle. In addition, the ability of the pilot to follow the commanded heading is limited by the resolution of his display, compass accuracy, the aircraft stability and control characteristics, and the distracting effects of the search and acquisition functions he must perform just prior to lock-on.

The diagram of Fig. 2-31 may be used for an analysis of the heading error distributions. The uncertainties of the vectoring system cause errors to develop in a sequence that may be examined as follows:

The interceptor-target sight line established by the vectoring system may differ from the true sight-target line by an amount which can be expressed approximately as

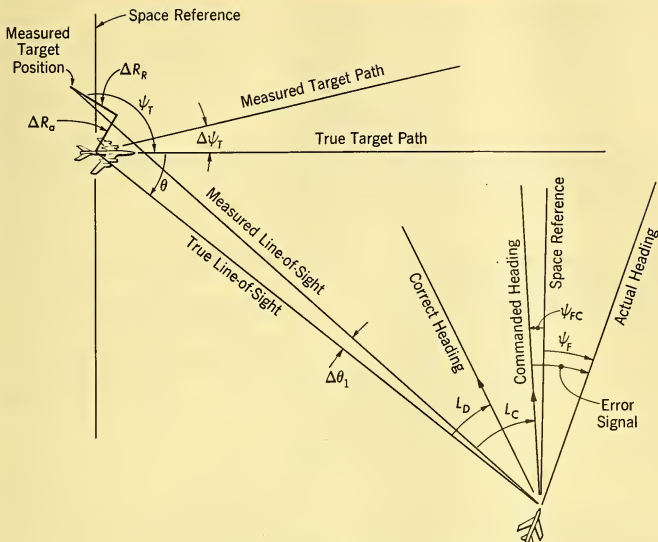


FIG. 2-31 Vectoring Error Geometry.

$$\Delta\theta_1 \approx \Delta R_a / R \quad (2-41)$$

for

$$\Delta R_a \ll R$$

where ΔR_a = component of the relative position error between target and interceptor which is normal to the sight line.

The vectoring system computes a desired interceptor lead angle L_C with respect to the erroneous sight-line angle. From Fig. 2-31 the computed lead angle is

$$L_C = L_D + \frac{\partial L_D}{\partial \theta} \Delta\theta_1 + \frac{\partial L_D}{\partial \theta} \Delta\psi_T + \frac{\partial L_D}{\partial V_T} \Delta V_T \quad (2-42)$$

since

$$\sin L_D = V_T / V_F \sin \theta. \quad (2-4)$$

Then,

$$L_C = L_D + \left[\frac{V_T \cos \theta}{V_F \cos L_D} \right] \Delta\theta_1 + \left[\frac{V_T \cos \theta}{V_F \cos L_D} \right] \Delta\psi_T + \left[\frac{\sin \theta}{V_F \cos L_D} \right] \Delta V_T. \quad (2-43)$$

The commanded heading differs from the correct heading by

$$\epsilon_C = L_C + \theta_1 - L_D = \left[1 + \frac{V_T \cos \theta}{V_F \cos L_D} \right] \Delta\theta_1 + \left[\frac{V_T \cos \theta}{V_F \cos L_D} \right] \Delta\psi_T + \left[\frac{\sin \theta}{V_F \cos L_D} \right] \Delta V_T. \quad (2-44)$$

The error signal presented to the pilot is the difference between the commanded heading and the actual heading,

$$\Delta\psi_F = \psi_{FC} - \psi_F. \quad (2-45)$$

It is assumed that the pilot follows the commanded heading with an error whose standard deviation is 5° .

The total heading error with respect to the correct heading is then

$$\epsilon_T = \Delta\psi_F + \left[\frac{-R}{R V_F \cos L_D} \right] \Delta R_a + \left[\frac{V_T \cos \theta}{V_F \cos L_D} \right] \Delta\psi_T + \left[\frac{\sin \theta}{V_F \cos L_D} \right] \Delta V_T \quad (2-46)$$

since the closing rate, \dot{R} , may be expressed

$$-\dot{R} = V_T \cos \theta + V_F \cos L_D. \quad (2-47)$$

If the vectoring errors are assumed to be independent, we may write the standard deviation of the collision course heading error as

$$\sigma_{\epsilon,T} = \left[\left(\frac{\dot{R} \sigma_a}{V \cos L_D} \right)^2 + \left(\frac{V_T \cos \theta}{V_F \cos L_D} \sigma_{\psi,T} \right)^2 + \left(\frac{\sin \theta}{V_F \cos L_D} \sigma_{V,T} \right)^2 + \sigma_{\psi,F} \right]^{1/2} \quad (2-48)$$

where $\sigma_a = (\Delta R_a)/R$.

The evaluation of this expression for various values of lock-on range from 8 to 30 n.mi. is given in Fig. 2-32 for the estimates of measurement uncertainty derived for the AEW system (Paragraph 2-21). The curves may be interpreted in the following manner. For range to the target R and an angle off the target's nose θ : if the proper collision-course lead angle for this condition is L_D (Fig. 2-30) then the vectoring errors will cause the interceptor lead angles to be normally distributed about the value L_D with a standard deviation of $\sigma_{\epsilon,T}$ degrees. The magnitude of the heading error increases very rapidly as the range decreases. This will be shown to have detrimental effects on the AI radar gimbal angle requirements for short-range lock-ons and on the ability to convert a short-range lock-on into a successful attack.

The large magnitude of the heading errors for forward-hemisphere attacks is characteristic of any guidance system employing "prediction". Collision vectoring is such a system; it attempts to guide the interceptor towards a point in space where the target will be at some future time.

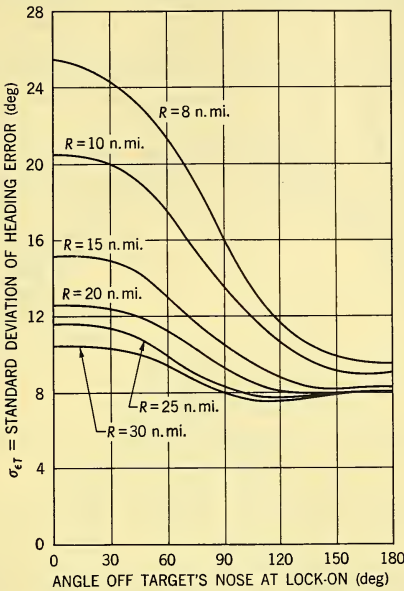


FIG. 2-32 Standard Deviation of Heading Error vs. Angle off Target's Nose at Lock-on.

Prediction guidance systems require the use of velocity as well as position information. For this reason they are most sensitive to closing speed. This phenomena was indicated in Fig. 2-24. The reader might satisfy himself on this point by analyzing the errors for a pursuit vectoring system, i.e., a system where the interceptor is commanded to point at the target. This analysis would disclose that the heading error distributions for all angles are about equal to the tail-chase distributions for collision vectoring. Thus the tactical advantage of collision vectoring is bought at the price of increased AI radar and vectoring system requirements.

2-26 AI RADAR REQUIREMENTS DICTATED BY VECTORING CONSIDERATIONS

The vectoring situation gives rise to several requirements that must be fulfilled by the AI radar.

Look-Angle Requirements. If lock-on is to occur at any selected range in the 8–30 n.mi. interval, the radar must be able to “look” at the target. That is to say, the maximum look-angle of the AI radar antenna must be sufficient to encompass the distributions of probable target angular positions relative to the interceptor heading. (*Look-angle* is often referred to as *gimbal angle* or *train angle*).

A typical situation is shown by Fig. 2-33. The possible positions of the target relative to the interceptor are shown as a distribution of angular positions around the lead angle L_D that would exist for perfect vectoring.

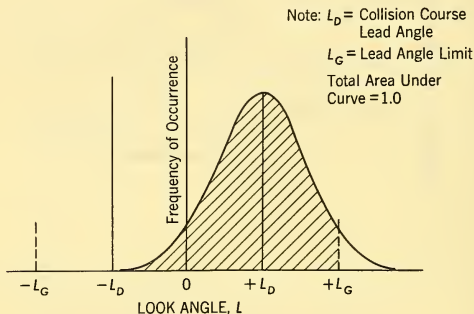


Fig. 2-33 Probability Density Distribution of Target Angular Positions Relative to Interceptor Heading.

For any range and angle off the nose, such a figure could be formulated from the data in Figs. 2-28 and 2-32 in the preceding paragraph. The probability that the AI radar can look at the target at this range and angle is simply the area under the curve that lies between the AI radar look angle limits L_G .

The look-angles required to ensure that 95 per cent of the targets are within the AI radar field of view are displayed in Fig. 2-34. The prices of short-range lock-ons and “around the clock” attack capability are evidenced by the large radar gimbal angles required to maintain 95 per cent probability. When the lock-on range satisfying the conversion probability requirement is found, Fig. 2-34 may be used to determine the AI radar gimbal angle dictated by vectoring considerations.

Display Requirements. From Fig. 2-29, we see that the vectoring system transmits required attack altitude, time to collision, and range relative to the interceptor — in addition to the heading commands already discussed — to provide tactical situation information to the pilot. All the vectoring information plus pitch and roll information must be presented on

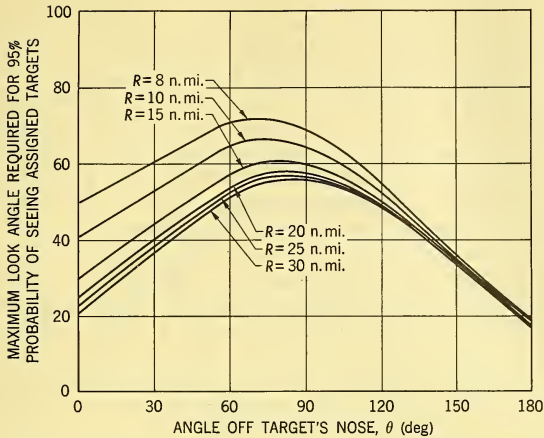


FIG. 2-34 Maximum Look Angles Required for 95 Per Cent Probability of Seeing Assigned Target with Collision Vectoring.

an integrated display which allows the pilot to fly the aircraft in response to the vectoring commands.

During the last part of the vectoring phase, the pilot must detect and acquire the target. The displays required for these functions must also be integrated with the other vectoring displays to permit proper utilization of the information. The considerations governing the design of a display system to meet such requirements are treated in Chapter 12. This is one of the most difficult design problems for any radar system; it is particularly so for an AI radar because of the limited space and multiplicity of functions the pilot must perform. Display integration, like reliability, is easier to specify than to achieve.

Search Volume Requirements. Radar search is accomplished by scanning a prescribed volume of space as was shown in Chapter 1 (Fig. 1-1). Target position uncertainty relative to the interceptor determines the required dimensions of this volume.

For a given lock-on range, the azimuth look-angle needed to accommodate 95 per cent of the expected tactical situations is shown in Fig. 2-34. Since the search and acquisition procedures precede AI radar lock-on, it is necessary to ascertain whether a target that is within the field of view at a given range would also have been continuously within the field of view at greater ranges. An inspection of Fig. 2-32 shows this to be the case. The largest gimbal angles are required by the shortest ranges. Thus, only the

required lock-on range need be known to specify a maximum gimbal angle satisfying search, acquisition, and lock-on requirements.

The sensitivity of viewing probability to gimbal angle may be obtained by examining similar curves for viewing probabilities of 90, 80, and 70 per cent (Figs. 2-35 to 2-37). For example, a 67° maximum look-angle is required to achieve 95 per cent viewing probability at 10 n.mi. range and 75° off the target's nose (Fig. 2-34). For a 90 per cent probability under the same conditions, a 60° maximum look-angle is required (Fig. 2-35). This heavy price suggests that a different allocation of viewing and conversion probabilities might yield a result nearer the optimum.

The required elevation angular coverage is determined by the elevation uncertainty of the vectoring system. As already derived (Paragraph 2-19), the elevation measurement error has a standard deviation of 0.5 n.mi. Thus the probability is virtually unity that the target height is within three standard deviations (1.5 n.mi.) of the vectoring radar system measurement. At a range of 10 n.mi. an AI radar elevation coverage of 17° (0.3 radian) is required to encompass this uncertainty. This requirement varies inversely with the required lock-on range and may be expressed

$$\text{Search pattern elevation coverage} = \frac{6\sigma_H(57.3)}{R_1} \text{ degrees.} \quad (2-49)$$

The maximum range dimension R_1 of the search volume is the range at which search begins. Its value depends on the required lock-on range and

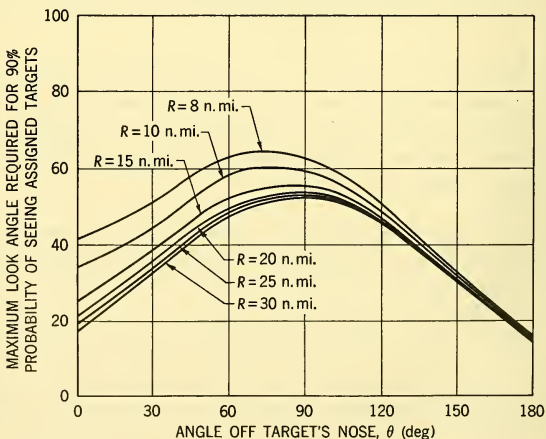


FIG. 2-35 Maximum Look Angles Required for 90 Per Cent Probability of Seeing Assigned Target with Collision Vectoring.

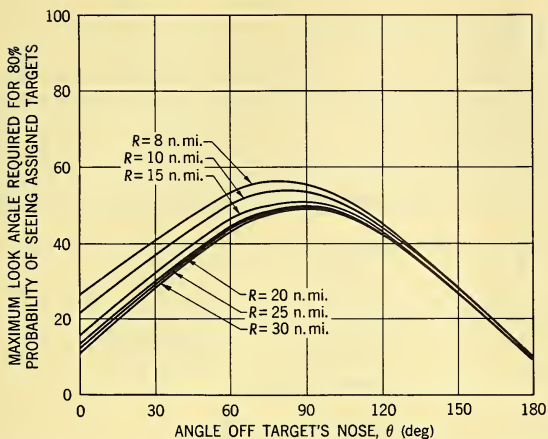


FIG. 2-36 Maximum Look Angles Required for 80 Per Cent Probability of Seeing Assigned Target with Collision Vectoring.

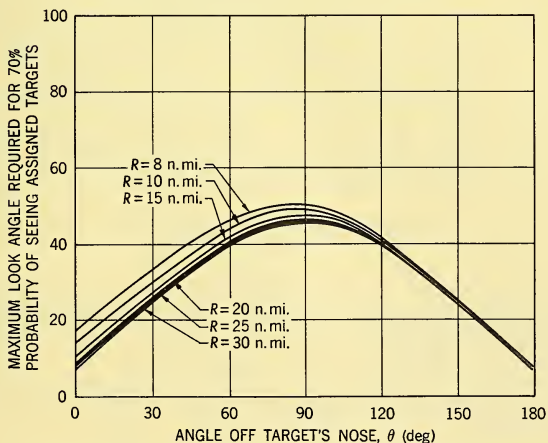


FIG. 2-37 Maximum Look Angles Required for 70 Per Cent Probability of Seeing Assigned Target with Collision Vectoring.

the radar characteristics. For most radars, a value of two times the required lock-on range is adequate.

Stabilization Requirements. The angular requirements for the search pattern were derived with the tacit assumption that the search pattern was space-stabilized in roll and pitch about the aircraft flight line. That is to say, the volume of space illuminated by the radar is independent of aircraft angles of attack and roll. This assumption results in a considerably smaller search pattern than would be the case if these motions were allowed to displace the search pattern. This effect is illustrated by Fig. 2-38. Search pattern stabilization also makes the radar search display

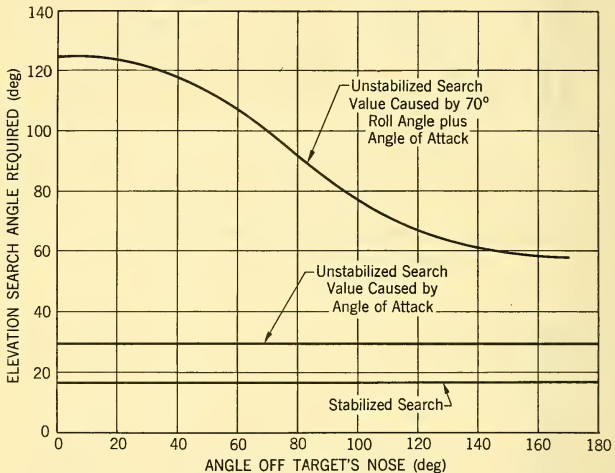


FIG. 2-38 Elevation Search Angle Requirements for 10 n.mi. Lock-on (Stabilized and Unstabilized Search).

problem easier to solve, as will be shown in Chapter 8. For these reasons it is required that AI radar search pattern be stabilized in roll and pitch about the aircraft flight line.

Summary. A summary of the AI radar requirements dictated by vectoring considerations is shown in the overall requirements summary, Paragraph 2-30.

2-27 ANALYSIS OF THE CONVERSION PROBLEM

The plan for analyzing the conversion problem is shown in Fig. 2-39. Analysis of the conversion phase must yield the following information relevant to the AI radar design:

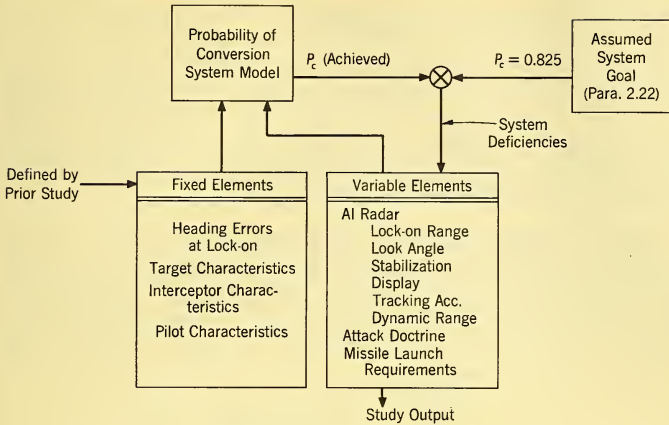


FIG. 2-39 Plan for the Analysis of Conversion Probability.

1. The minimum required AI radar lock-on range
2. The fire-control computer requirements
3. The attack display requirements
4. The missile launching and illumination requirements
5. The radar tracking and stabilization requirements

Attack Phase System Logic. The flow of information and the allocation of function during the attack phase are shown in Fig. 2-40. Following AI radar lock-on, the AI radar measures target range, lead angle,

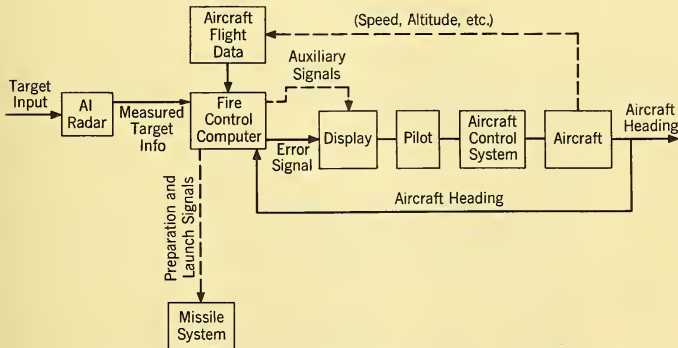


FIG. 2-40 Interceptor System Logic Diagram During Attack Phase.

angular velocity, and range rate along the line-of-sight. This information is utilized in conjunction with aircraft flight data (speed, altitude, etc.) to compute an attack course that permits the weapons to be launched with a high kill probability (see Paragraph 1-4 and Figs. 1-3 and 1-4). Deviations between the computed attack course and the actual interceptor flight path are presented to the pilot as a steering error signal. The pilot — or autopilot — flies the aircraft to reduce the steering error to within the limits required by the weapon characteristics.

Guided Missile Launching Zone Parameters. The allowable launching ranges and angular error launching tolerances for the interceptor's guided missile may be obtained from a graphical representation of the launching problem. This analogue model — shown schematically in Fig. 2-41 — utilizes the fixed parameters of the target, interceptor, and

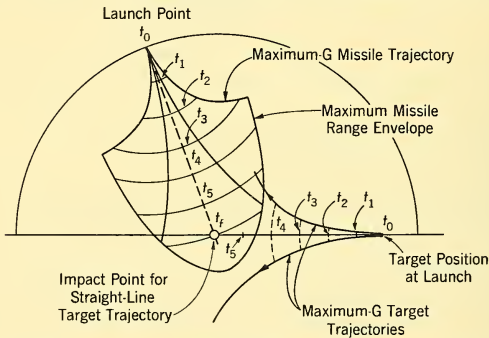


FIG. 2-41 Graphical Determination of Launching Zones.

guided missile as defined in Figs. 2-5 to 2-7. The launching tolerances calculated from this model represent the permissible deviations from perfect solutions to the fire-control problem.

We may construct and employ this model to analyze the problem in the following manner:

- STEP 1. A value of range to impact point that lies between the maximum and minimum missile downranges is chosen. A semicircle with a radius equal to this range is drawn around the impact point.
- STEP 2. The missile time of flight corresponding to the chosen downrange is read from the missile performance diagram. The target position at weapon launch can then be plotted V_{Tt_f} units back of the impact point.

- STEP 3. The target is assumed to have two possible types of trajectories during the weapon time of flight: (a) a straight line, and (b) a maneuver at the maximum permissible target aircraft load factor. These trajectories can be plotted as functions of time after weapon launch.
- STEP 4. Now we may plot the missile performance diagram on a transparent sheet, using the same scale as the target and firing circle diagram. The origin of the missile performance diagram is made to coincide with a point on the firing circle. The missile performance diagram overlay can then be rotated with respect to the target and firing circle diagram to determine the maximum aiming errors that would still permit interception of the target by the guided missile. An interception is defined as any point within the missile performance contour where a missile time-of-flight line and the time marker on the target trajectory are equal. This procedure may be repeated for a number of points on the firing circle.
- STEP 5. The foregoing steps may be repeated for a number of assumed ranges-to-impact and for all the assumed altitude and speed conditions. Using maximum allowable aiming error as a parameter, we may plot range against angle off the target's nose

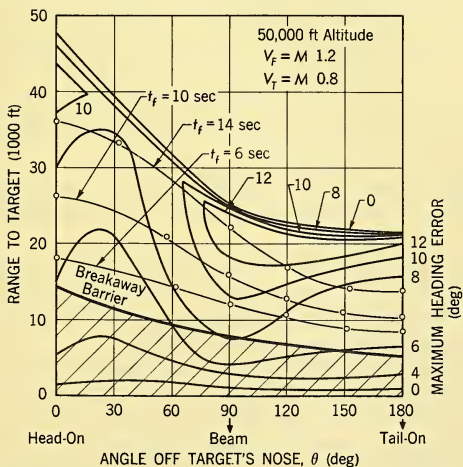


Fig. 2-42 Missile Launch Zones and Launching Tolerances; 50,000-Ft Altitude.

at launch. The results of such a process as applied to our fictitious problem are shown in Fig. 2-42.

This analysis shows that the allowable launching tolerance varies quite widely, depending upon the launching range. The tolerances on heading at launch are quite tight for very large or very small ranges. They are comparatively liberal for intermediate ranges. For example, if missile firing occurs from 20,000 ft range at 90° off the target's nose, an error of 12° is permitted. The missile time of flight for this instance is 12 seconds.

The usable minimum missile launching range is determined by the requirement that the interceptor not pass closer than 1000 ft to either the impact point or the target in order to preclude self-destruction. Using the defined maneuvering capabilities of the interceptor, the minimum launching range or breakaway barrier dictated by this requirement may be calculated by graphical techniques similar to those used for the launching tolerance determination. The result of such an analysis (for a nonmaneuvering target) is shown superimposed on the missile launch zone diagram (Fig. 2-42).

Thus the allowable missile launching ranges and angular aiming errors — as limited by the characteristics of the target, interceptor, and guided missile and the target avoidance problem — are determined for each angle off the target. Note that the allowable angular launching tolerances are appreciably smaller than an inspection of only the missile performance diagram would indicate — 5° to 10° compared with 10° to 30° for the missile itself (Fig. 2-6). This is a typical result of a study which examines the guided missile performance in its expected tactical environment. It can be seen that the allowable launching tolerances determine the required accuracy of the AI radar and fire-control system. This is why the AI radar designer must be certain the missile performance is defined for *operation in the expected tactical environment*.

Fire-Control System Parameters — Attack Doctrine. All the basic information needed for fire-control system specification is now available.

The fire-control system must be compatible with five requirements or limitations: (1) minimum average penetration distance; (2) “around-the-clock” launching capability; (3) collision vectoring; (4) missile launching tolerances; and (5) interceptor maneuver limits.

A modified form of collision guidance — known as *lead collision* — provides a reasonable answer. For any tactical situation this guidance system attempts to direct the interceptor on a *straight-line course* to a point where the missiles may be fired with high kill probability. The straight-line characteristic reduces penetration, reduces interceptor maneuver requirements, and allows missile launching to take place at any angle off the

target's nose. In a lead-collision system, missile launching occurs automatically at such a range that the missile time of flight to the impact point equals a preset constant. The value of this constant may be chosen to utilize the best characteristics within the allowable launching zones. For a given angle off the target, the lead angles required in a lead-collision system correspond closely to the collision vectoring lead angles — a fact which is helpful in solving the conversion problem.

Lead-collision geometry is shown in Fig. 2-43. Solution of the fire-control triangle yields

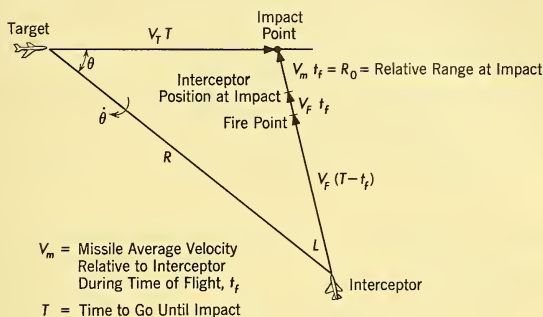


FIG. 2-43 Lead-Collision Geometry: Two-Dimensional.

$$R = V_T T \cos \theta + V_T T \cos L + V_m t_f \cos L \quad (2-50)$$

$$V_T T \sin \theta = (V_T T + V_m t_f) \sin L. \quad (2-51)$$

The component of relative velocity along the line-of-sight is

$$\dot{R} = -V_T \cos \theta - V_F \cos L. \quad (2-52)$$

The component of relative velocity perpendicular to the line-of-sight is

$$R\dot{\theta} = V_T \sin \theta - V_F \sin L. \quad (2-53)$$

By definition of a lead collision course

$$t_f = \text{a preset constant.} \quad (2-54)$$

From the definition of missile characteristics for straight-line flight (Fig. 2-6)

$$V_m = f(V_F, \text{altitude}, t_f). \quad (2-55)$$

Thus, for a fixed time of flight and known speed and altitude conditions

$$V_m t_f = R_0 = \text{constant.} \quad (2-56)$$

Using Equations 2-52, 2-53, 2-56 to eliminate velocity terms in Equations 2-50 and 2-51 and rearranging terms, we obtain

$$R + \dot{R}T - R_0 \cos L = 0 \quad (2-57)$$

$$\sin L = (RT/R_0) \dot{\theta}. \quad (2-58)$$

The fire-control system must solve two problems. It must provide (1) a signal for automatically firing the missiles at the correct point, and (2) an aiming error signal for the pilot or autopilot.

The AI radar measures range, range rate, lead angle, and space angular velocity of the line of sight ($R_M, \dot{R}_M, L_M, \dot{\theta}_M$).¹² Aircraft speed and altitude may be combined with known missile performance at the preset time of flight to obtain R_0 (see Equations 2-54 to 2-56).

The measured target inputs and the computed missile characteristics may be substituted into Equations 2-57 and 2-58 to obtain

$$T_c = \frac{-R_M + R_0 \cos L_M}{\dot{R}_M} \quad (\text{computed time-to-go until impact}) \quad (2-59)$$

$$\sin L_c = \left(R_M \frac{T_c}{R_0} \right) \dot{\theta}_M \quad (\text{computed correct lead angle}). \quad (2-60)$$

Firing occurs when

$$T_c = t_f \quad (\text{preset}). \quad (2-61)$$

A steering error signal is obtained by taking the differences between the sines of computed and measured lead angles and multiplying this difference by a sensitivity factor $(R_0 \cos L)/(R_0 + V_f T)$. This factor causes the computed angular error signal to be a close approximation of the actual angular aiming error. Thus, the computed steering error is

$$\epsilon_{Hc} = [R_0 \cos L_M / (R_0 + V_f T_c)] [\sin L_c - \sin L_M]. \quad (2-62)$$

Both the azimuth and elevation error signals are computed from an expression of this form.

Equations 2-59 to 2-62 define the fire-control and tracking problems that are to be solved by the AI radar and fire-control system. The precision required of this solution is determined by the angular aiming tolerances corresponding to the selected value of preset time-of-flight.

For the purpose of developing a representative set of accuracy specifications we shall select 10 seconds for the preset time of flight t_f , which corresponds to a relative displacement at impact of $R_0 = 6800$ ft. An inspection of Fig. 2-42 shows this is a reasonable choice since firing will occur near the center of the allowable launch zone for all angles off the nose at

¹²The subscript M denotes a measured quantity.

firing. The maximum allowable launching error for a 10-second time of flight may be plotted from the data of Fig. 2-42 as shown in Fig. 2-44. As

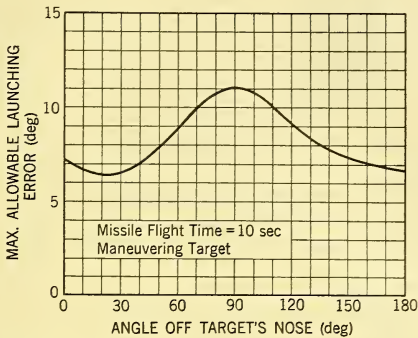


FIG. 2-44 Maximum Allowable Launching Errors.

can be seen, head-on and tail-on attacks impose the most severe requirements upon overall aiming accuracy.

The functions and overall accuracy required from the fire control system have now been defined. The next problem is to specify how this error is to be divided among the possible sources of error in the system.

Fire-Control System Error Specification. The sources of system error can be listed as follows:

- (1) AI radar measurements ($R_M, \dot{R}_M, \dot{\theta}_M, L_M$)
- (2) Flight-data measurements (altitude, speed)
- (3) Fire-control computation
- (4) Pilot-airframe-display interaction

There are two general types of errors — predictable bias errors and random errors.

Predictable bias errors arise from the dynamic response characteristics of the measuring device. For example, in the lead-collision system specified, the variables R , \dot{R} , L , and $\dot{\theta}$ can change rapidly as the launching point is approached (Fig. 2-45).¹³ Dynamic lags in the measuring devices will cause measurement errors whose values may be predicted from a knowledge of the input parameters and the dynamics of the measuring device. The

¹³In this application, the system must continue to track after missile launch to provide illumination for the missile seeker. Thus, dynamic lags are also important after missile launch ($T < 10$ seconds); it will be noticed in this connection, that the dynamic inputs are quite severe for this case. This point will be discussed further in Paragraph 2-29.

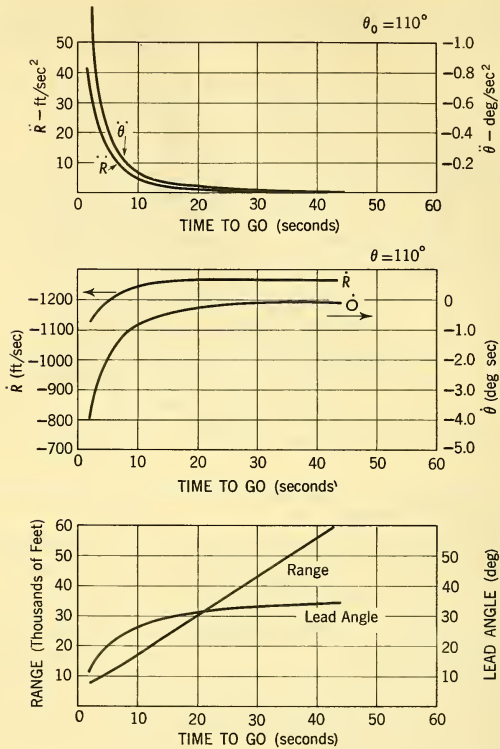


FIG. 2-45 Dynamic Variation of Lead Collision Fire Control Parameters.

measurement errors affect the firing time and the steering error as calculated in the fire-control computer by Equations 2-59 through 2-62.

Because dynamic lag errors *are* predictable, it is theoretically possible to eliminate them entirely by suitably clever design. However, the more usual approach is to limit the magnitude of these errors to some finite value. As a general rule of thumb, it is desirable that the total value of the predictable bias error contribution obey the following inequality:

$$B < \sqrt{2} \sigma \tag{2-63}$$

where B = total predictable bias error

σ = standard deviation of the total random error.

and computed quantities in this expression were correct and if the pilot flew the aircraft in such a manner as to reduce the computed error to zero, then

Random errors arise from several main sources. First of all are the measurement uncertainties caused by the basic limitations of the measuring device. The angular measuring accuracy of a radar, for example, is limited by beamwidth as was indicated in the discussion of AEW radar requirements.¹⁴ Mechanical and electrical component tolerances also contribute to errors of this type.

The system noise sources also contribute to random errors. For example, the finite dimensions of a radar target introduce time-dependent uncertainties into the measurements of range and angle (see Paragraph 4-8). Similarly the vagaries of airflow past the aircraft may introduce random noise errors into flight data measurements. These latter would affect the computation of R_0 .

Random aiming errors also are caused by the pilot's inability to guide the aircraft on exactly the course indicated by the displayed error information. Paragraph 12-7 will discuss this problem in some detail. Generally speaking, however, if the pilot is presented with an error signal which is band-limited to about 0.25 rad/sec¹⁵ and if the error signal is contaminated by random noise which is bandlimited to about 1 rad/sec, then the pilot can steer the aircraft with a random error which has a standard deviation approximately equal to the standard deviation of the noise. Thus the pilot's contribution to the total aiming error may be written:

$$\sigma_{Pf} = \sigma_N \quad (2-64)$$

where σ_{Pf} = standard deviation of the pilot's flyability error

σ_N = rms value of the noise on the error signal display.

To illustrate how the error specification might be developed we shall consider two cases: (1) a head-on attack and (2) an attack which begins at an angle off the target's nose at launch of 80°.

The method for attacking the problem can be outlined as follows. As already mentioned, Equation 2-62 is designed to provide a reasonable approximation of the actual heading error. In fact, if all of the measured

¹⁴Actually, as will be indicated in Chapter 5, the problem is a good deal more complicated than is indicated by this statement. Signal-to-noise ratio and observation time also strongly affect the angular accuracy. However, for fixed values of these latter parameters, the statement is substantially correct.

¹⁵The bandwidth of the error signal depends upon the type of attack trajectory flown. A lead-collision course is a straight line; hence the effective bandwidth of the input guidance signals is very low. Curved-course trajectories such as lead-pursuit have higher effective guidance signal bandwidths. Chap. 12 of the "Guidance" volume of this series presents an excellent discussion of the concept of treating a guidance trajectory in terms of its frequency spectrum.

perfect aiming would result. Practically, however, the measured and computed quantities (R_0, L_M, T_c, L_c) are not correct for reasons previously discussed. Thus the computed error $\epsilon_{H,C}$ differs from the actual aiming error. The contribution of each source of error to this difference may be expressed.

$$\Delta\epsilon_H \doteq (\partial\epsilon_{H,C}/\partial x)\Delta x \quad (2-65)$$

where $\Delta\epsilon_H$ = the steering error due to error in the quantity x

$\partial\epsilon_{H,C}/\partial x$ = partial derivative of the steering error with respect to the quantity x

Δx = error in the measurement of the quantity x .

As an example, the sensitivity of the steering error to an error in measured lead angle may be derived from Equations 2-62 and 2-59 as:

$$\begin{aligned} \partial\epsilon_{H,C}/\partial L_M &= (\partial\epsilon_{H,C}/\partial L_M) + (\partial\epsilon_{H,C}/\partial T_c)(\partial T_c/\partial L_M) \\ &= \left[\frac{R_0 \cos L_M}{V_F T + R_0} \right] \left[-\cos L_M - \frac{R_M \dot{\theta}}{R_M} \sin L_M \right]. \end{aligned} \quad (2-66)$$

It should be noted that the sensitivity is a variable quantity during an attack course; it also varies from one course to another. Consequently the sensitivities must be examined for the range of attack courses. In this discussion we will confine our attention to the two courses assumed (head-on and 80° off the nose).

The values of the input variables and their derivatives are shown in Fig. 2-45. The error sensitivity factors for each of the assumed attack courses are shown in Fig. 2-46. It will be noted that dynamic variations of the input quantities are greatest for the attack which terminates near the target's beam; thus, predictable bias errors arising from dynamic lags will be greatest for this course. On the other hand, the effect of errors in angular rate and lead angle is greatest for head-on attacks. This fact is particularly significant because angular rate errors tend to be the most important source of system errors.

Using the foregoing error data, an error specification may be derived in the following manner. For a head-on attack, the total system aiming error must be held below 7° to ensure that the missile will hit a maneuvering target (see Fig. 2-44). For purposes of deriving a tentative specification, we may split this error among the various error sources by appropriate manipulation of the following expression:

$$\begin{aligned} \text{Total system error} &\doteq \text{pilot requirement} + \Sigma(\partial\epsilon_{H,C}/\partial x_i)\Delta x_i \\ &\quad + 2\sqrt{\Sigma[(\partial\epsilon_{H,C}/\partial x_i)\sigma x_i]^2} \end{aligned} \quad (2-67)$$

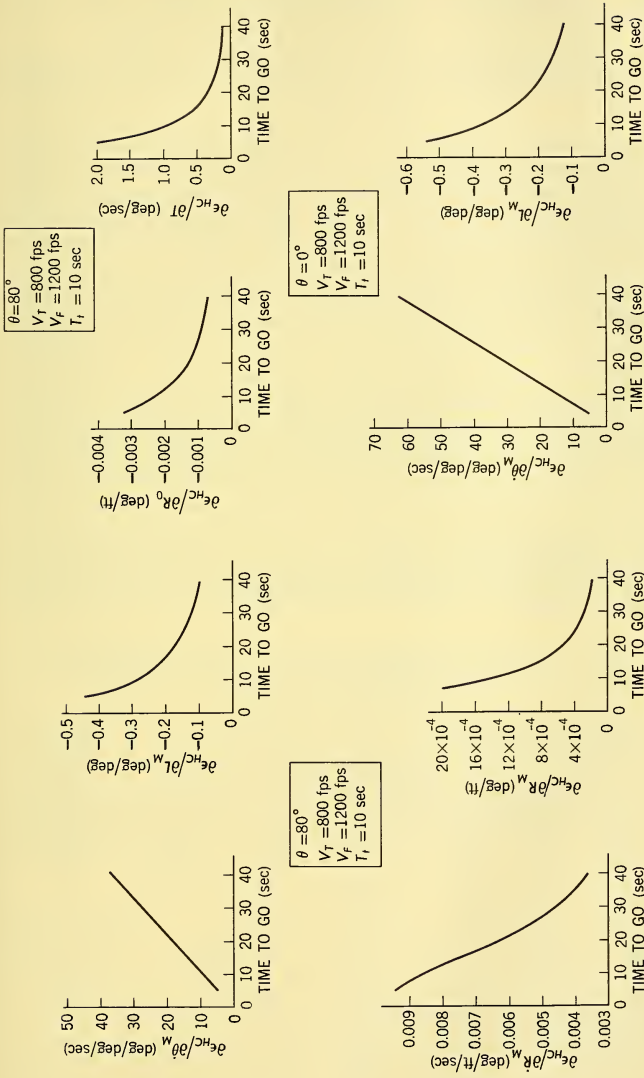


FIG. 2-46 Error Sensitivity Factors.

where pilot requirement = maximum allowable indicated error at firing

$\Sigma(\partial\epsilon_{H,C}/\partial x_i)\Delta x_i$ = summation of predictable bias errors

$2\sqrt{\Sigma[(\partial\epsilon_{H,C}/\partial x_i)\sigma x_i]^2}$ = twice the standard deviation of the total random error.

We will assign a value of 2° to the pilot requirement; i.e., the pilot is required only to bring the indicated error within a value of 2° to ensure successful missile launching. This error is, in effect, treated as an allowable predictable bias error and it is desirable that its allowable value be made as large as possible because this will reduce the total time needed to reduce an initial steering error at lock-on (see Fig. 2-49 below).

For the head-on case, predictable bias errors due to dynamic lags present no problem because the input quantities (\dot{R} , $\dot{\theta}$, L_M) are relatively constant over the entire attack course and the system is relatively insensitive to mechanization approximations used in the computation of R_0 (relative range of the guided missile at impact). Thus, predictable bias errors (other than pilot bias) can be assigned a value of zero for the head-on case. The remaining error tolerance (5°) can be split up among the sources of random error as shown in Table 2-3. It will be noted that no tolerances are given for range and time-to-go quantities; their effect on the head-on attack problem is too insignificant to provide a satisfactory basis for specification.

The allowable random angular errors ($\dot{\theta}_M$, L_M , and pilot steering) are equally divided between the azimuth and elevation channels by dividing the total allowable error by $\sqrt{2}$. This analysis shows that the radar must

TABLE 2-3 MEASUREMENT ACCURACY REQUIREMENTS FOR HEAD-ON ATTACKS

Source of Error (x_i)	Sensitivity ($\partial\epsilon_{H,C}/\partial x_i$)	Allowable Steering Error Contribution ($\partial\epsilon_{H,C}/\partial x_i$) $\times 2\sigma x_i$	Error Specification σx_i	Allowable rms Error per Channel (Azimuth and Elevation)
R_M	0	0		
\dot{R}_M	0	0		
R_0	0	0		
T	0	0		
$\dot{\theta}_M$	14.2	4.25	$0.15^\circ/\text{sec}$	$0.11^\circ/\text{sec}$
L_M	0.36	0.11	0.15°	0.11°
Pilot	≈ 1	2.8	1.4°	1°
Total random error		$\approx 5^\circ$		
Pilot bias		2°		
Total error		7°		

provide angular rate and angle information which has rms errors in each channel of about 2 mils/sec (0.11°/sec) and 2 mils (0.11°) respectively. Referring to Equation 2-64 and the accompanying discussion, it is also seen that the computer filtering system must be designed to limit the rms noise on the indicator to a value of about 1.0° rms in order to meet the pilot steering accuracy requirement.

The other attack course (80° off the nose at lock-on) may be analyzed in a similar fashion. For this case the maximum allowable error is about 10.7° (80° off the nose at lock-on will result in about 90° off the target's nose at time of firing). Using the allowable errors already established for the head-on case, the values of the allowable predictable bias errors and the values of the random range and time errors may be established as shown in Table 2-4. It should be emphasized that this allocation can be adjusted to suit the designer's convenience, provided the total error allowance is not exceeded.

Chapter 9 will present a discussion of how error specifications and dynamic input requirements derived in this manner can be used to dictate the detailed requirements of the range and angle tracking loops of the radar.

TABLE 2-4 MEASUREMENT ACCURACY REQUIREMENTS FOR BEAM ATTACKS

Source of Error	Sensitivity $\partial\epsilon_{H,C}/\partial x_i$	Allowable Predictable Bias Error Contribution	Allowable Random Bias Error Contribution $(\partial\epsilon_{H,C}/\partial x_i) \times 2\sigma_{x_i}$	Predictable Bias Error Specification	Random Bias Error Specification σ_{x_i}	
R_M	0.0014	0.25	1.85	179	662	
\dot{R}_M	0.0086	0.25	1.85	29	108	
R_0	0.0023	0.25	1.85	109	400	
T						
(computation)	1	0.30	1.85	0.30	0.925	
					Total	Per Channel
$\dot{\theta}_M$	10	2	3.0	0.2	0.15	0.11
L_M	0.285	0.15	0.086	0.525	0.15	0.11
Pilot	≈ 1	2	2.8	2	1.4	1.0

UNITS: degrees, seconds, feet. $\Sigma = 5.2$. $\sqrt{\Sigma(\)^2} = 5.5$.

2-28 LOCK-ON RANGE AND LOOK-ANGLE REQUIREMENTS DICTATED BY THE CONVERSION PROBLEM

The establishment of the weapon firing range as a function of target aspect angle completes the information needed to calculate lock-on range requirements for the conversion probability of 0.825. Fig. 2-47 displays

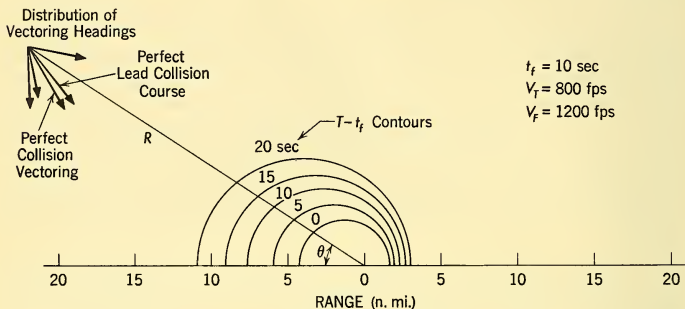


FIG. 2-47 Interceptor System Model for Conversion Problem.

the essential elements of the problem. If lock-on occurs at (R, θ) the heading error that must be corrected has two components: (1) the vectoring uncertainty and (2) the difference between the correct collision-course lead angle at (R, θ) and the correct lead collision-course lead angle at (R, θ) . We shall assume that the distribution of vectoring errors is centered about the correct collision lead angle for point (R, θ) . The error which is present at lock-on must be reduced below the allowable 2° pilot bias error prior to reaching the missile firing range at 10 seconds time-to-go. We shall assume that the time available for reduction of the steering error at lock-on is equal to the time available to an interceptor passing through the point (R, θ) on a lead collision course prior to reaching the missile launch range. Fig. 2-47 illustrates the situation. Contours indicating the time from (R, θ) to missile release are shown, as well as a typical heading distribution at (R, θ) which will arise at lock-on. (The distribution of aircraft headings relative to a perfect collision vectoring course is defined in Fig. 2-32.)

The lead collision-course lead angle is a function of both time-to-go and aspect angle. Fig. 2-48 illustrates the variation in lead-collision lead angle as a function of time-to-go and aspect angle. For a given (R, θ) value, the lead collision-course lead angle always is less than the correct collision-course lead angle.

The time required to reduce an initial steering error is shown in Fig. 2-49 for various initial values of steering error. The primary factor contributing

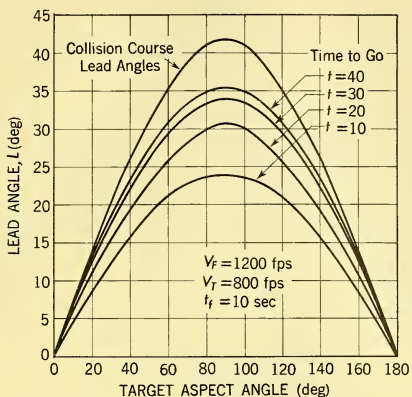


FIG. 2-48 Lead Collision Lead Angles.

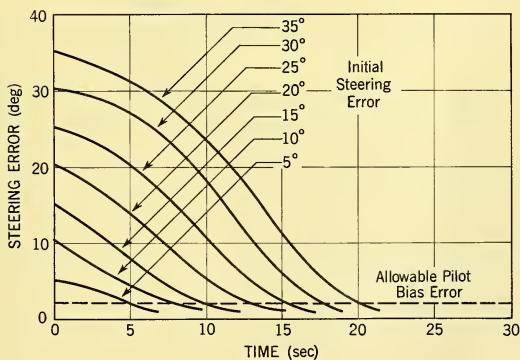


FIG. 2-49 Time for Example Interceptor to Reduce an Initial Steering Error.

to this time is the limitation on aircraft maneuverability (2 g's). Data of this type are usually obtained from simulation studies of aircraft-pilot-display performance.

The analysis procedure to determine the probability of conversion is indicated by the flow diagram shown in Fig. 2-50. As an example of how such a calculation might be made, we may consider the following case.

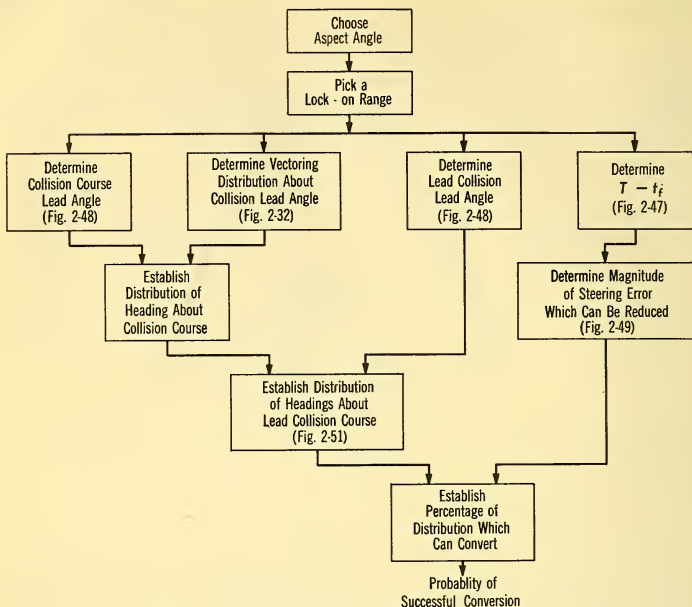


FIG. 2-50 Conversion Probability Analysis Plan.

Aspect angle at lock-on = 60°

Lock-on range = 8 n.mi.

Collision-course lead angle = 35°

Vectoring distribution, $\sigma_{\epsilon, T} = 21.5^\circ$

Lead collision-course lead angle, $L = 26.5^\circ$

$T - t_f = 20.5$ sec

Maximum correctable steering error = 36°

Fig. 2-51 shows the distribution of heading errors relative to the correct collision and lead-collision courses. Since an error of 36° may be corrected, any initial heading which results in a lead angle between 62.5° and -9.5° may be converted into a successful missile launch. Thus, the probability of conversion is equal to the shaded area shown, which may be determined as 88.5 per cent.

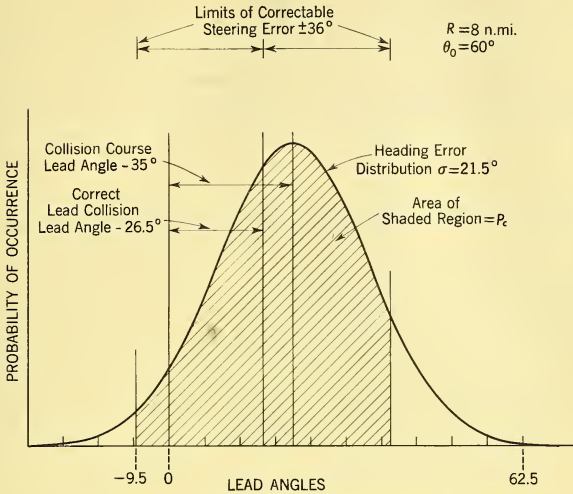


FIG. 2-51 Method for Calculating Conversion Probability.

The calculation of conversion probability by this technique is approximate. Certain kinematic effects such as the change of collision-course lead-angle with time-to-go and the effects of initial steering error on the ultimate attack course flown by the pilot are neglected. Evaluation of these effects requires elaborate simulation programs. In a practical case, it is usually desirable to investigate these areas by more elaborate techniques.

This simplified analysis, repeated for many values of lock-on range and aspect angle, culminates in curves like those in Fig. 2-52. Notice that as one would expect, the head-on attack provides the most stringent requirements for lock-on range. The assumed system requirement stated that the conversion probability must be at least 0.825 for any aspect angle. The corresponding viewing probability requirement was 0.95. Thus for this hypothetical system approximately 10 n.mi. lock-on range is required to achieve the requisite conversion probability.

The look-angle requirements are dictated by vectoring considerations, since the collision-course lead-angle is greater than the lead-collision-course lead-angle for the same range and aspect angle (see Fig. 2-48). From Fig. 2-34 we may determine that a 10 n.mi. lock-on range and a vectoring probability requirement of 0.95 combine to dictate a look-angle capability of 67° .

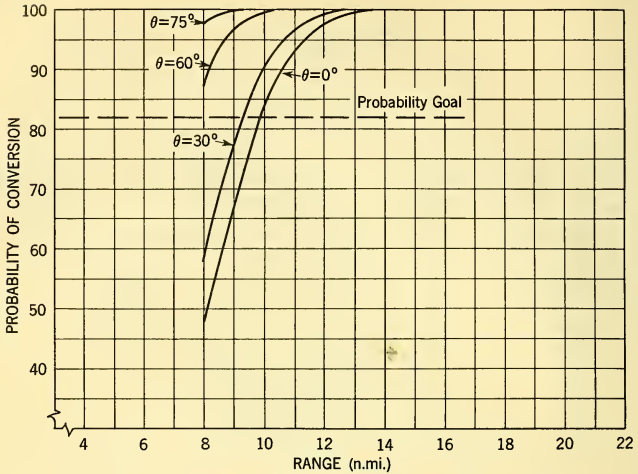


FIG. 2-52 The Probability of Conversion After Lock-on.

The discerning reader will note that a trade-off analysis could be made between lock-on range and look-angle. For example, a longer lock-on range would allow a smaller look-angle. When space in the aircraft nose is at a premium, it may be easier to increase lock-on range than to provide large look-angles. In addition, the derived look-angle specification (67°) is pessimistic. At the aspect angle at which look-angle is critical ($\theta = 75^\circ$) a lock-on range of 10 n.mi. yields a conversion probability of 100 per cent. This reduces the vectoring probability requirement for this attack from 95 per cent to 78.5 per cent. The look-angle requirement corresponding to this vectoring probability may be read from Fig. 2-36 as 53°. This is quite a significant relaxation of requirements and illustrates the advantages to be gained by examining the interrelationships among the system factors.

In summary, the lock-on requirements are established by the head-on attack situation, and the look-angle requirements are established by the beam aspect approach situation. These requirements are:

- Required lock-on range 10 n.mi. with 90 per cent cumulative probability
- Required look-angle $\pm 53^\circ$ in azimuth and elevation

The look-angle capability must be provided in both azimuth and elevation because the aircraft will roll to angles approaching 90° during the conversion and vectoring phases.

The required detection range is found by specifying a mean lock-on time and adding the range closed between the target and the interceptor during this time. For example, the closure rate in a head-on attack is 2000 fps (0.33 n.mi./sec). For mean lock-on times of 6 and 12 seconds the required detection ranges are therefore 12 and 14 n.mi. respectively. In each of these cases, the required cumulative probability of detection is defined as 90 per cent.¹⁶

2-29 AI RADAR REQUIREMENTS IMPOSED BY MISSILE GUIDANCE CONSIDERATIONS

The requirements dictated by missile guidance considerations can be derived from Fig. 2-6 and the previous analysis of the tactical situation.

The AI radar illuminated the target continuously during the missile flight time; the missile seeker tracks the reflected signal and homes on the target on a proportional navigation course. From Fig. 2-6 it is seen that if the AI radar tracking accuracy is better than 0.35° rms, the AI radar will not cause degradation of missile performance. The specified tracking accuracy of 0.15° rms is well within these limits. However, dynamic lag errors pose an additional complication. The data of Fig. 2-45 show that very rapid changes in angular rate and range occur near the end of missile flight ($T \rightarrow 0$). The dynamic responses of the range and angle tracking loops must be sufficient to maintain AI radar range lock-on and limit the angle lag error. The exact determination of the allowable lag error would require a more detailed study of interrelations between the missile seeker and the AI radar. However, a value of about 0.25° would represent a reasonable estimate.

The maximum range to the target for which illumination must be provided is obtained on the head-on attack (4.4 n.mi.). Fig. 4-6 shows that 120 kw of peak pulse power is required to ensure seeker lock-on at this range with a 24-inch antenna. A larger antenna would reduce the power requirement and vice versa.

The frequency of the seeker (X band) and the type of seeker (pulse radar semiactive) are major factors governing the choice of AI radar frequency and type, since a separate illuminating system would have to be provided if the two were different. As will be indicated in Chapter 6, the choice of a

¹⁶The derivation of the radar detection and lock-on requirements made no mention of probability. As discussed in Paragraph 2-12, it is customary to express ranges so derived as the range at which the radar should have 90 per cent cumulative probability of detection (on lock-on). This assumption puts a safety factor into the analysis, since a radar which meets this requirement will yield a slightly better probability of conversion than a radar which always locked on at exactly 10 n.mi. This comes about because 90 per cent of the lock-ons occur at ranges greater than 10 n.mi.; the resulting improvement in conversion probability for these cases more than offsets the decreased conversion probability of the 10 per cent which occur at ranges less than 10 n.mi.

pulse radar is entirely reasonable for the high-altitude (i.e. clutter-free) operation required in this tactical application.

To assist lock-on of the missile seeker, the AI radar also is required to provide range and angle slaving signals to the missile seeker. The angular accuracy is not particularly critical, since the missile seeker beamwidth is relatively wide, perhaps of the order of 10° to 12° . Range accuracy, on the other hand, can be fairly critical. If the missile seeker is assumed to operate with a pulsewidth of $0.5 \mu\text{sec}$ (250 ft) and a $1\text{-}\mu\text{sec}$ (500-ft) range gate, then range errors in excess of about 150-200 ft can begin to affect seeker lock-on capability. The range error specification previously derived (Table 2-4) dictated allowable errors of about this magnitude (bias error plus 2σ value of random error). In a practical case, this condition would dictate a more comprehensive analysis of seeker AI radar interrelations.

2-30 SUMMARY OF AI REQUIREMENTS

Reliability: 90 per cent for 3-hour operation

Search Pattern: 60° azimuth;
 17° elevation;
 Stabilized in roll and pitch

Search Range: 20 n.mi.

Search Display: Vectoring heading command
 Time to collision
 Attack altitude
 Range to target
 Interceptor roll and pitch
 AI radar target detection information

Detection Range: 12 n.mi. (90 per cent probability) with 6-second lock-on time
 14 n.mi. (90 per cent probability) with 12-second lock-on time

Lock-on Range: 10 n.mi. at a closing speed of 2000 fps with 90 per cent probability

Look-Angle: $\pm 60^\circ$ in azimuth and elevation

Required Computation: Lead collision (see Equations 2-59 to 2-62)

Required Accuracies: See Table 2-4, Paragraph 2-28

Dynamic Inputs: See Fig. 2-45

Maximum Allowable Angle Tracking Lag: 0.25° during missile guidance phase only

Stabilization: Compatible with accuracy requirements and maneuvering characteristics listed in Table 2-5

Display: Steering error signal display

Aircraft roll and pitch (see Chap. 12)

Additional tactical information as shown to be necessary (see Chap. 12)

Noise filtered to 1° rms

Maximum signal information delay 0.5–1.0 second

Frequency: X band

Power: Greater than 120 kw peak with a 24-inch diameter antenna

Radar Type: Pulse

2-31 SUMMARY

The foregoing analyses have demonstrated the vast amount of systems analysis that must precede the design of a successful airborne radar system for a particular application. The length of this chapter is in itself testimony to the possible complexities of such analyses. The drawing together and rationalization of the important factors of an airborne radar application problem is as difficult as it is necessary to proper system design. Moreover, work of this nature should continue in parallel with the radar system design to ensure that the radar design problem is always viewed in the light of the most advanced understanding of the overall weapons system problem.

Succeeding chapters of this book will break the radar design problem into its component parts, with the general objective of showing how each element of the radar — transmitter, propagation path, target, receiver, data processing, and display — may be related to the overall functions and requirements of the system. Where appropriate, the examples developed in this chapter will be employed to develop further examples.

CHAPTER 3

THE CALCULATION OF RADAR DETECTION
PROBABILITY AND ANGULAR RESOLUTION

3-1 GENERAL REMARKS

In establishing the preliminary design of a radar subsystem to meet overall weapons system requirements, the designer must first choose the basic radar organization or configuration. He then endeavors to select the radar parameters so as to provide the required performance with practical equipments. In order to do this rationally, he must have reliable methods for estimating the performance of hypothetical systems. In this chapter, calculations in the critical areas of detection performance and angular resolution will be discussed. The former is a particularly complicated area of analysis because of the statistical problems introduced by receiver noise and target fluctuations. The effects of multiple looks at a target and operator performance further complicate the situation. Techniques for taking these factors into account for a conventional pulse radar and a pulsed doppler radar will be developed.

The definition of angular resolution and the factors which might act to degrade it will be discussed briefly. These factors include the effects of unequal target sizes, signal-to-noise ratio, receiver saturation, pulsing, and system bandwidth.

3-2 THE RADAR RANGE EQUATION

A primary basis for the choice of radar system parameters is the radar range equation. In one form or another, this relation gives the power received from a radar target or the ratio of this signal power to the power of competing noise or other interference from which the signal must be distinguished. We shall briefly consider the origin of the range equation.

We suppose that a radar transmitter radiates power denoted by P_t isotropically (uniformly in all directions). At a range R , then, the power density or power per unit area will be

$$\text{Power density of an isotropic radiator} = \frac{P_t}{4\pi R^2} \quad (3-1)$$

Normally, the transmitter is not an isotropic radiator but possesses a directivity or power gain due to the influence of an antenna. The power gain on transmission is denoted by G_t and the resulting power density at the range R is

$$\text{Power density with an antenna} = D = \frac{P_t G_t}{4\pi R^2}. \quad (3-2)$$

This power is incident upon some sort of target which reflects a portion to the receiver. The target will be characterized by an idealized or effective cross-sectional area σ . This area is defined to reradiate isotropically all the incident energy collected. The target cross section will be dependent on the radar frequency being used and the aspect from which the target is viewed. It is normally determined experimentally and often represents a large unknown factor in radar system calculations. By definition, the power collected by the target is $D\sigma$. When this power is reradiated isotropically, the power density at the receiver, which is assumed to be located near or at the transmitter, is simply

$$\text{Power density at the receiver} = \frac{D\sigma}{4\pi R^2} = \frac{P_t G_t \sigma}{(4\pi)^2 R^4}. \quad (3-3)$$

The effective area of the receiving aperture is denoted by A_r . The power intercepted by the effective area of the receiving antenna is simply the product of this area and the power density. The receiving area is related to the receiving gain G_r and the wavelength λ by the following relation.¹

$$A_r = \frac{G_r \lambda^2}{4\pi}. \quad (3-4)$$

We shall assume, as is normally the case, that the same antenna is used for reception and for transmission. In this event, the receiving gain will equal the transmission gain or $G_r = G_t = G$.

The power received by the receiver will be simply the product of the power density at the receiver and the receiving area. Combining Equations 3-3 and 3-4, the received signal power will be

$$\text{Received signal power} = S = \frac{P_t G^2 \lambda^2 \sigma}{(4\pi)^2 R^4}. \quad (3-5)$$

This expression represents one version of the radar range equation. It shows how the received power varies with target range and size and with the wavelength and power gain of the antenna. The received power can represent either average power or peak power, depending upon what the transmitted power P_t represents.

¹See Paragraph 10-1 for a further discussion and references.

An extensive discussion of target cross section is given in Chap. 4. The radar cross section of aircraft targets is discussed in Paragraph 4-7 and some typical examples are shown in Figs. 4-20, 4-21, and 4-22. The effective cross sections of sea and ground surface reflections are discussed in Paragraphs 4-10 through 4-13. In this connection, a normalized cross section is defined as the radar cross-sectional area per unit surface area. This quantity is denoted by σ^0 and is usually referred to as *sigma zero*. With the illuminated surface area denoted by A , the radar cross section and sigma zero are related by

$$\sigma = \sigma^0 A. \quad (3-6)$$

The area of the resolution element on the ground is a function of the pulse length, depression angle, and antenna beamwidths and is given by Eq. 4-60a and b. Examples showing the variation of sigma zero with environmental conditions and radar frequency are given in Figs. 4-34 through 4-43.

The radar range equation is often expressed as the ratio of the received power reflected from the target to the power of some interfering signal. Most commonly, the interfering signal is random noise generated within the receiver; it might also be ground or sea clutter, atmospheric reflections or anomalies, or some sort of jamming. Internal receiver noise is often referred to as *thermal noise*, not necessarily because it arises physically from electronic agitation but because in characterizing it a comparison is made with noise which does arise from this source. Normally, internal receiver noise determines the maximum range of the radar system; and even when other sources of interference predominate, it provides a useful reference point. The equivalent input noise power of a receiver is normally expressed in the following form.²

$$\begin{aligned} \text{Equivalent input noise power} &= N = FkTB \quad \text{watts} \\ &= 4 \times 10^{-21} FB \quad \text{watts} \end{aligned} \quad (3-7)$$

where F = noise figure — the factor by which the equivalent input noise of the actual receiver exceeds that of an ideal reference

$$k = 1.37 \times 10^{-23} \text{ joule/}^\circ\text{K} = \text{Boltzmann's constant}$$

$$T = \text{absolute temperature of noise source — arbitrarily, } 290^\circ \text{ K}$$

$$B = \text{equivalent rectangular bandwidth of the receiver in cycles per second.}$$

The ratio of the signal and noise powers as given by Equations 3-5 and 3-7 yields the signal-to-noise ratio, S/N .

$$\text{Signal-to-noise ratio} = S/N = \frac{P_t G^2 \lambda^3 \sigma}{(4\pi)^3 F k B T R^4} \quad (3-8)$$

²See Paragraph 7-3 for a further discussion of receiver noise and the origin of this expression.

This expression is also referred to as the *radar range equation*. The receiver bandwidth B is normally determined by the IF amplifier in pulse radar systems, although in some cases subsequent filtering or integration is interpreted as equivalent to a narrowing of the noise bandwidth.

Another convention is to solve Equation 3-8 for the range when the signal-to-noise ratio is unity. This range is called the *idealized radar range* and will be denoted by R_0 :

$$\text{Idealized radar range} = R_0 = \sqrt[4]{\frac{P_t G^2 \lambda^2 \sigma}{(4\pi)^3 k T B F}} \quad (3-9)$$

With this definition, the expression for the signal-to-noise ratio given in Equation 3-8 takes the following simple and useful form:

$$\text{Signal-to-noise ratio} = S/N = (R_0/R)^4. \quad (3-10)$$

To provide an illustration of the use of Equation 3-9, let us suppose that an airborne radar system possesses the following parameter values:

$$P_t = \text{peak power} = 200 \text{ kw} \quad \sigma = \text{target cross section} = 1.0 \text{ m}^2$$

$$G = \text{antenna gain} = 1000 = 30 \text{ db} \quad F = \text{noise figure} = 10 \text{ db}$$

$$\lambda = \text{wavelength} = 3 \text{ cm} - 0.03 \text{ m} \quad B = \text{IF bandwidth} = 1 \text{ Mc/sec}$$

It is convenient to express each parameter in decibels relative to a convenient set of units and then simply to add these figures with appropriate signs to obtain the logarithm of the idealized range, thus:

$$P_t = 83.0 \text{ db (milliwatts)}$$

$$F = -10.0 \text{ db (unity)}$$

$$G^2 = 60.0 \text{ db (unity)}$$

$$kTB = \underline{114.0 \text{ db (milliwatts)}}$$

$$\lambda^2 = -30.5 \text{ db (meters}^2\text{)}$$

$$R_0^4 = 183.6 \text{ db (meters}^4\text{)}$$

$$\sigma = 0 \text{ db (meters}^2\text{)}$$

$$R_0 = 45.0 \text{ db (meters)} = 3.89 \times$$

$$(4\pi)^3 = -32.9 \text{ db (unity)}$$

$$10^4 \text{ meters}$$

$$= 20.4 \text{ n. mi.}$$

3-3 THE CALCULATION OF DETECTION PROBABILITY FOR A PULSE RADAR

Target detection is a radar function of primary importance and a necessary preliminary to other important functions such as tracking, resolution, and discrimination. In this paragraph, we shall discuss the detection process and describe methods for estimating its reliability as a function of the radar system and target parameters. Although the developments in this paragraph will relate primarily to pulse radar systems, the principles apply generally to any type of radar system used for detection.

Factors to be considered include the model assumed to represent system operation, the effect of the operator, the effect of the target's closing velocity, and the effect of fluctuations in target size.

Model of System Operation. The notation which we shall adopt in this paragraph is listed below.

a = received signal voltage (peak)	S/N = signal to noise power ratio
f_r = pulse repetition frequency	t_{sc} = scan time
N = noise power	u = video voltage at pulse integrator output
n = number of pulses illuminating a target during scan	v = video voltage at square law detector output
R_0 = idealized range	η = false alarm number
R = actual range	Θ = antenna beamwidth
\dot{R} = range rate	τ = pulse length
ΔR = range decrement between scans	ω_c = angular carrier frequency (RF or IF)
S = received signal power (peak)	ω_s = scan speed

The radar system model providing the basis for our analysis of the detection process is shown in Fig. 3-1. The target is assumed to be an aircraft at a range R closing on the radar system at a constant range rate \dot{R} .

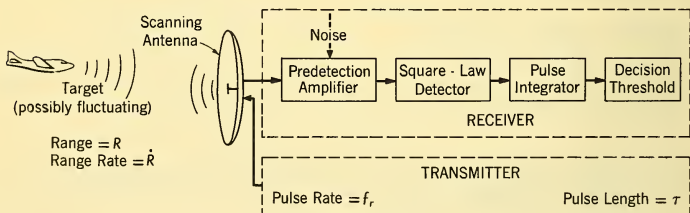


FIG. 3-1 Radar System Model Assumed for Analysis of Detection Process.

Two cases are distinguished: (1) a target of constant size and (2) a target whose size fluctuates in accord with a Rayleigh distribution as is discussed in Paragraphs 4-7 and 4-8.

A pulse radar with a small duty cycle (on the order of a thousandth or less) is assumed. The target is illuminated periodically by a scanning

antenna. The antenna pattern is approximated by a constant gain over the antenna beamwidth θ , and zero gain outside of this region. The received signal on a single scan will consist of n pulses. In the absence of target size fluctuations, these pulses will all be of the same size. The number of pulses is given by the product of the repetition frequency f_r and the beamwidth, divided by the scan velocity ω_s :

$$\text{Number of pulses in a scan} = n = f_r \theta / \omega_s. \quad (3-11)$$

The received signal is assumed to be a pulsed sinusoid. The signal power during a pulse is denoted by S , and the internal noise power referred to the same point in the system is denoted by N . In the case of a fluctuating target an average signal power \bar{S} will be defined.

The essential parts of the receiver for this analysis consist of a predetection amplifier, a square law detector, a pulse integrator, and a decision threshold.

The *predetection amplifier* is normally the intermediate frequency (IF) amplifier, and it is assumed to be matched to the envelope of the pulse shape. That is, the bandwidth of this amplifier is approximately equal to the reciprocal of the pulse length. Noise with a uniform power density is assumed to be introduced into the system at the input to this amplifier. The power spectrum of the noise at the amplifier output will thus be equal to the power transfer function of the amplifier. The peak signal-to-noise ratio at the output of the predetection amplifier is S/N , as was indicated above.

A *square-law detector* is assumed to generate a video voltage equal to the square of the envelope of the predetection signal plus noise. In this case, the development in Paragraph 5-7 is applicable and can be used to establish the amplitude distribution and power-density spectrum of the video signal plus noise. The assumption of a square-law detector rather than a linear detector is primarily for mathematical convenience. It does not represent a serious restriction because the basic results are only slightly dependent on the detector law.

The *pulse integrator* combines the n pulses received during a scan over the target. In Paragraph 5-10 it will be shown that the linear operation which gives the greatest signal-to-noise ratio for a signal consisting of n pulses corresponds to the addition of these pulses to form a sum signal. Accordingly, in order to provide the greatest possible signal-to-noise ratio at the decision threshold — and thus the greatest reliability of detection — these n pulses are assumed to be added together by a pulse integrator.³

³In many practical systems, integration is provided by the memory of the human operators or by retention of the signal on the face of the cathode ray display tube. In such cases, the integration is not a perfect summing process, and degradation in the S/N ratio is experienced. This degradation is discussed later in this chapter and also in Chapter 12, on radar displays.

The *decision element* in the radar system is assumed to be simply a threshold or bias. When the integrated video voltage exceeds this threshold, detection is said to have occurred. When this voltage fails to exceed the threshold, no detection occurs. The bias may be exceeded for one of two reasons. (1) The integrated video signal-plus-noise may exceed the bias; in this case, "target detection" takes place. (2) The integrated noise alone may exceed the bias; in this case, a "false alarm" takes place. Fig. 5-16 shows how a decision threshold or bias is used to distinguish between the distribution of signal plus noise and noise alone. The selection of the threshold level thus will represent a compromise between the desire for maximum sensitivity to integrated signal plus noise and the system penalties incurred by false alarms.

Method of Analysis. Using the radar system model already described and the mathematical theory presented in Chapter 5, we will trace the progress of noise and signal plus noise through the elements of the receiver. *The objectives of this analysis are to derive the target detection and false-alarm probabilities as functions of S/N ratio, threshold level, and the amount of integration.*

The analysis will be performed for both constant and fluctuating radar targets to determine the *probability of detection on a single scan*.⁴

Finally, the concept of single-scan probability of detection will be employed in Paragraph 3-4 to develop the *multiple-scan probability of detection* for a moving target. This quantity — also called the *cumulative probability of detection* — is the one most directly related to system performance in the tactical-use environment.⁵ For example, the detection ranges specified for the examples in Chapter 2 were expressed in terms of the cumulative probability of detection.

From the standpoint of clear exposition, it is rather unfortunate that a true understanding of the radar detection problem is wrapped in complexities of statistical theory which do not convey to the practicing designer a real feel for the problem. The author has attempted to alleviate this problem by confining some of the more detailed mathematical derivation to Chapter 5; the analysis that follows herein applies some of the results of these derivations as they pertain to the assumed model.

Signal Analysis. As previously mentioned, the input to the square-law detector consists of noise with a power spectrum equal to the power transfer function of the amplifier and — when a signal is present — a signal

⁴This quantity is often called the "blip-scan" ratio or the "single glimpse" detection probability.

⁵See Paragraph 2-12 and Fig. 2-19.

with a peak power of S . We are now interested in finding analytical expressions for the square-law detector output under the conditions of noise-only inputs and signal-plus-noise inputs.

It is conventional and convenient to approximate the video voltage in the absence of the signal by a series of independent samples which are spaced at intervals equal to the reciprocal of the predetection bandwidth. Such an approximation is shown in Fig. 3-2. This approximation is based upon

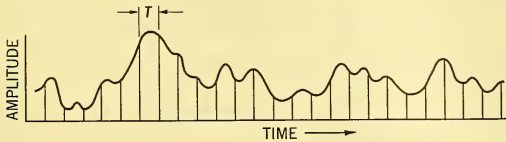


FIG. 3-2 Representation of Continuous Video Voltage by a Sequence of Samples.

the famous sampling theorem which states: *If a function $f(t)$ contains no frequencies higher than $W/2$ cps, it is completely determined by its ordinates at a series of points spaced $1/W$ seconds apart.*⁶ In this connection, the envelope of the noise in a predetection band of width W cps can be shown to be equivalent to a low-frequency function limited to frequencies less than $W/2$ cps, and it can be represented by a series of samples spaced by $T = 1/W$ seconds.

Since the spectrum of the predetection filter is matched to the spectrum of the pulse envelope, it will have a width approximately equal to the reciprocal of the pulse length, τ . In this case, the samples will be spaced by intervals equal to τ . It can also be shown by an appropriate application of the material in Chapter 5 that the statistical fluctuations in these samples are independent. In this case, each sample can be considered a separate detection trial, and the input to the decision element during an observation period can be regarded as a series of independent trials for which methods of analysis are well known. For instance, if the probability of exceeding the threshold is p , the probability of exceeding the threshold at least once in m trials is

$$\text{Probability of at least one success in } m \text{ trials} = 1 - (1 - p)^m. \quad (3-12)$$

Further, the average number of trials between successes is the same as the average number of trials per success, which is equal to the reciprocal of the probability on a single trial, $1/p$. When there is no signal present so that any exceeding of the threshold represents a false alarm, the number $1/p$ is called the *false-alarm number*.

⁶C. E. Shannon, "Communication in the Presence of Noise," *Proc. IRE* 37, 10-21 (1949).

The video signal and noise samples are statistical variables. Their probability density functions are determined in Paragraph 5-7. We denote the video voltage out of the square law detector by v . This voltage is equal to the square of the video envelope r in Equations 5-78 and 5-79 in Paragraph 5-7. Making the transformations $v = r^2$, $S = a^2/2$, and $N = \sigma^2$ in these equations provides the probability density functions of the video voltage for signal plus noise and noise alone.

Probability density video voltage signal plus noise =

$$P_{S+N}(v) = \frac{1}{2N} \exp \left[\frac{-v}{2N} - \frac{S}{N} \right] I_0(\sqrt{2vS/N^2}) \quad (3-13)$$

Probability density video voltage noise alone =

$$P_N(v) = \frac{1}{2N} \exp \left[\frac{-v}{2N} \right]. \quad (3-14)$$

The video voltage when no signal is present is thus represented by a series of independent samples at intervals of $\tau = 1/W$ which are chosen from a statistical population with the probability density of Equation 3-14. When the signal is present, the sample is chosen from a population with the probability density of Equation 3-13.

An interpretation of these expressions may be given as follows. For a given value of noise power N and a given value of signal power S the probability that the video voltage will have a value between v and $v + dv$ may be expressed as $P_{S+N}(v)dv$.

Next we examine the effects of integration. We denote the sum signal at the pulse integrator output by u .

$$u = v_1 + v_2 + v_3 + \dots + v_n \quad (3-15)$$

The components of the sum v_k are independent because they are separated in time by the repetition period while the correlation time of the video voltage is approximately the pulse length τ , which is on the order of microseconds.

Probability density functions giving the distribution of the signal plus noise and noise alone of the sum signal out of the integrator are required in order to determine whether a decision threshold will be exceeded. These probability density functions are denoted by

$$\text{Probability density integrator output, signal plus noise} = P_{S+N}(u) \quad (3-16)$$

$$\text{Probability density integrator output, noise alone} = P_N(u). \quad (3-17)$$

The probability density function of the integrator output when a signal is present is quite complicated, and we will not attempt a detailed study of this function here.⁷ Some calculations are greatly simplified, however, by

⁷For such a study see J. I. Marcum, *A Statistical Theory of Target Detection by Pulsed Radar*, RM-754; and *A Statistical Theory of Target Detection by Pulsed Radar: Mathematical Appendix*, RM-753, The RAND Corp., Santa Monica, Calif.

the adoption of a suitable approximation to $P_{S+N}(u)$. Such an approximation can be based on the assumption that $n(S/N) \gg 1$. In this case, the distribution of u is very nearly normal. This is so because when $S/N \gg 1$, v itself tends to be normally distributed (see Equation 5-80), while with $n \gg 1$, the distribution of the sum u tends to normality by the *central limit theorem*.⁸

The mean and standard deviation of the video voltage v can be found from Equation 5-81 or Fig. 5-12.

$$\text{Video d-c voltage} = \bar{v} = 2(N + S) \quad (3-18)$$

$$\text{Video rms a-c voltage} = \sigma_v = 2\sqrt{N(N + 2S)}. \quad (3-19)$$

The mean and standard deviation of the sum signal u will be larger by n , the number of components in the sum, and by the square root of n , respectively.

$$\text{Integrator output, d-c voltage} = \bar{u} = 2n(N + S) \quad (3-20)$$

$$\text{Integrator output, rms a-c voltage} = \sigma_u = 2\sqrt{nN(N + 2S)}. \quad (3-21)$$

The probability density function of the integrator output for noise alone can be established by standard statistical procedures to have the following form.

$$P_N(u) = \frac{1}{2N(n-1)!} \left(\frac{u}{2N} \right)^{n-1} e^{-u/2N}. \quad (3-22)$$

For the statistics-minded, we may note that each variable v_k is given by the sum of squares of two independent normal variables x_k and y_k . Thus, u will be the sum of the squares of $2n$ normal variates, and $P_N(u)$ will be the probability density function of a chi-squared distribution with $2n$ degrees of freedom.⁹

The Decision Element. The threshold type of decision element assumed for this system corresponds closely to the detection operations which would be performed by an automatic system such as might be employed in the terminal seeker of a guided missile. In many important cases, however, the human operator is the decision element. It is postulated that the human operator does something very similar to the threshold type of decision element. The functions of the human operator, though, would probably deviate somewhat from those performed by an ideal decision threshold. For instance, the threshold of human operators appears to vary

⁸J. L. Lawson and G. E. Uhlenbeck, *Threshold Signals*, pp. 46-52, McGraw-Hill Book Co., Inc., New York, 1950.

⁹See P. G. Hoel, *Introduction to Mathematical Statistics*, pp. 134-136, John Wiley & Sons, Inc., New York.

randomly from look to look because of their inability to judge accurately the signal strength or to remember exactly the threshold level. Their average threshold would also tend to increase with fatigue and inattention. The net result of these deviations generally seems to be a loss in detection efficiency of the human operator in comparison with that of a mathematical threshold. This degradation is often introduced through an "operator factor" or efficiency factor, p_0 . The probability of detection obtained on the basis of some threshold assumption is simply multiplied by p_0 to give the "realistic" probability of detection. Values ranging all the way from 0.05 to 0.8 have been specified for this factor at one time or another.

It is quite possible that a degradation of this kind represents certain detection operations quite well where the operators become fatigued or bored. On the other hand there are many detection situations where the use of an "operator factor" is very dubious. One such situation is that of the operator of an AI radar on a vectored, 10-minute interception mission. It is somewhat ridiculous to suppose that an operator on such a mission would completely miss, say, 50 per cent of all targets no matter how brightly they are painted on his scope. Another situation where the "operator factor" concept is obviously not applicable is in connection with automatic equipments. Here, the detection is directly accomplished through the use of a threshold.

In this chapter the "operator factor" concept will be abandoned in favor of simply introducing an operator degradation of the signal-to-noise ratio. This procedure is a standard one,¹⁰ and it leads to a somewhat simpler formulation of the cumulative probability of detection. A typical value for the degradation is given in the footnote reference as 2 db.

The decision threshold is chosen to give a false-alarm probability, or probability of detecting a target when none is present, which is compatible with the cost of committing the radar or weapon system to such an alarm. When such commitment costs can be established numerically, a selection of false-alarm time can be made on the basis of minimizing total costs. Most commonly, though, such costs cannot be established and the false-alarm time is arbitrarily fixed after a thorough but subjective study of its effect on the operational performance of the system.

The number of independent samples of signal-plus-noise in the false-alarm time is called the false-alarm number and is denoted by η . With false-alarm times varying from seconds to hours and pulse lengths varying from fractions of a microsecond up to milliseconds, the false-alarm number might have approximate upper and lower bounds of 10^{12} to 10^4 . The probability of having a false alarm on a single trial is the reciprocal of the false-alarm number. This probability, the probability that a noise sample

¹⁰W. M. Hall, "Prediction of Pulse Radar Performance," *Proc. IRE* (Feb. 1956) 234-231.

will exceed a threshold b , is given by the integral of the probability density function of the sum voltage u of noise alone.

$$\text{False alarm probability} = \frac{1}{\eta} = \int_b^{\infty} P_N(u) du. \quad (3-23)$$

This integral has been evaluated, and the result is shown in Fig. 3-3 for a useful range of parameters.

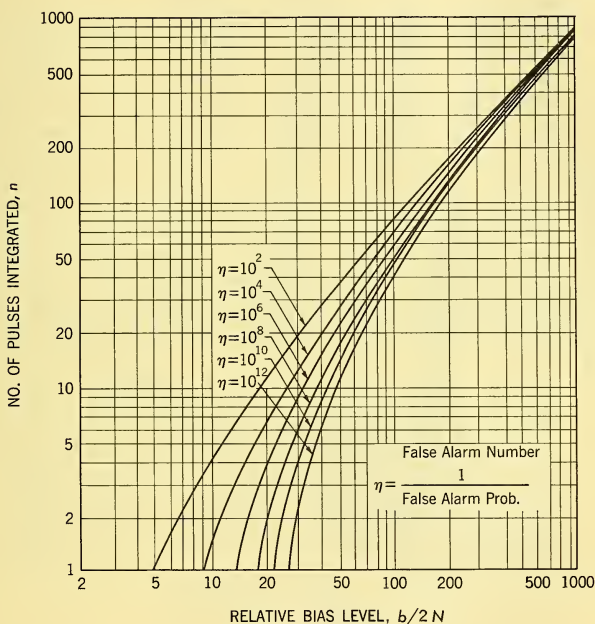


FIG. 3-3 Relation Between False Alarm Number and Bias Level with a Square-Law Detector.

Single-Scan Probability of Detection. Having chosen the threshold level on the basis of a required false-alarm time, the probability of detecting a target P_d is found by integrating the probability density function of the signal-plus-noise sum voltage

$$\text{Probability of detection} = P_d = \int_b^{\infty} P_{S+N}(u) du. \quad (3-24)$$

This integral gives the probability of detection of a nonfluctuating target on a single scan. It has been evaluated numerically for a useful range of the false-alarm number and the number of pulses integrated. The results of this calculation are shown in Fig. 3-4.¹¹ In this figure P_d is plotted as a function of the relative range R/R_0 whose reciprocal is equal to the fourth root of the signal-to-noise ratio.

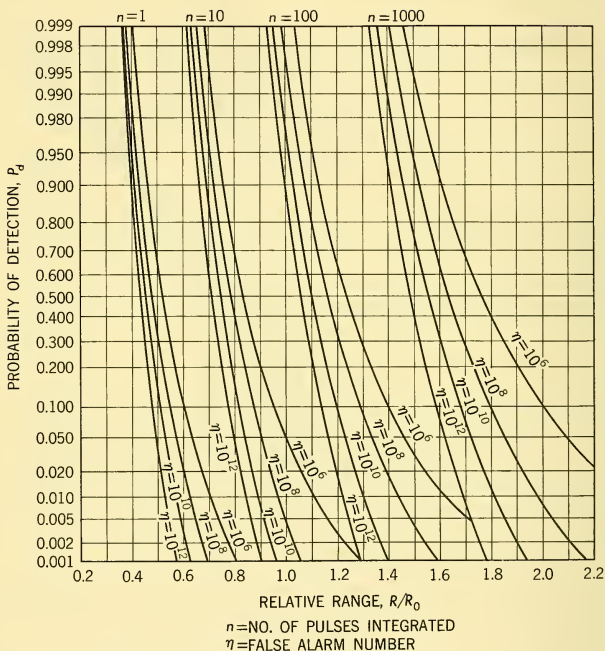


FIG. 3-4 Single-Scan Probability of Detection of a Nonfluctuating Target.

We refer to P_d as the *single-scan* or *single-glimpse probability* to distinguish it from the probability of detection when multiple looks are considered. Another term which is commonly used in this connection is *blip-scan ratio* which refers to the fraction of the time that an operator will see a blip in a single scan over the target.

¹¹For a wider range of such curves and also a discussion of a method for calculating them, see J. I. Marcum, *op. cit.*

To fix ideas, we consider an example of an AI (airborne intercept) radar. The idealized range R_0 has been determined in Paragraph 3-3 to be 20.4 n.mi. The radar parameters of interest are assumed as follows:

Pulse length = 1 μ sec

Beamwidth = 4.15°

Azimuth scan = 120°

Scan time = 3 sec

Elevation scan = 17°

PRF = 500 pps

The scan area is approximately $17^\circ \times 120^\circ = 2040^\circ$ squared. The beam area is approximately $(\pi/4) 4.15^2 = 13.6^\circ$ squared. The number of beam areas within a scan area is $2040/13.6 = 150$. With a scan time of 3 seconds and a PRF of 500 pps, the number of pulses received in a scan over the target is $3 \times 500/150 = 10$. We suppose that a false-alarm time of 100 seconds is chosen. With the IF bandwidth matched to the pulse width (approximately equal to its reciprocal) there will be 10^6 independent noise samples per second, and the false-alarm number will be $100 \times 10^6 = 10^8$.

Referring to Fig. 3-4, the relative range at which $P_d = 0.9$ for $n = 10$ and $\eta = 10^8$ is determined to be 0.72. The actual range giving 90 per cent probability of detection is thus $0.72 \times 20.4 = 14.7$ n.mi. A similar calculation gives the range corresponding to a detection probability of 10 per cent as 17.5 n.mi. The complete probability of detection curve for this example is shown in Fig. 3-5 along with the single scan and cumulative

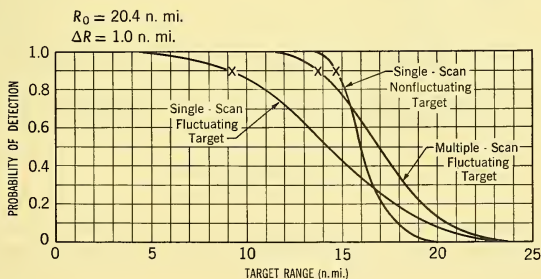


FIG. 3-5 Single-Scan and Cumulative Probability of Detection for Text Example.

probability of detection curves for a fluctuating target when the same basic radar parameters are used. The derivation of methods for the calculation of these other curves is discussed in the rest of this paragraph and in Paragraph 3-4.

The Effect of Target Fluctuations. The discussion thus far and the curves in Fig. 3-4 refer only to the case where the magnitude of the received

signal is constant. The energy reflected from an aircraft in flight is not generally constant. Such an object is a complex reflector of electromagnetic waves. As it moves in flight, it vibrates and turns relative to the radar system, and various parts of the aircraft reflect signals with more or less random amplitudes and with slightly different doppler shifts. As a result, the signal reflected from the aircraft fluctuates and can be represented as a noise process.¹² The power in the signal will be distributed similarly to the square of the envelope of narrow-band noise with a probability density function of the same form as that in Equation 3-14. This distribution is often called a Rayleigh distribution. The probability density function of the signal power S will thus have the following form:

Probability density of signal power from fluctuating target =

$$P(S) = \frac{1}{\bar{S}} e^{-S/\bar{S}}. \quad (3-25)$$

The factor \bar{S} in this expression represents the average signal power.

Because the rate of turn of aircraft is relatively slow, the spectrum of the fluctuations in S is normally less than about 3 cps in width,¹³ and S is reasonably well correlated over intervals of less than about 50 msec. We shall suppose that the observation time is less than this and assume that S is constant during a look, but independent from look to look.

With the signal-to-noise ratio fluctuating from scan to scan, the probability of detection will also fluctuate, and an average value of P_d must be calculated by weighting the various values of P_d by their probability of occurrence. Thus the average probability of detection is of the form

Average probability of detection for a fluctuating target =

$$\bar{P}_d = \int_0^{\infty} P_d(u, S) P(S) dS. \quad (3-26)$$

This integral can be evaluated approximately by making use of the previously noted Gaussian approximation to the distribution of integrated video signal plus noise. Using the values given in Equations 3-20 and 3-21 for the mean and variance of the integrated video, the approximate probability of detection is

$$P_d(S) \approx \frac{1}{\sqrt{8\pi nN(N+2S)}} \int_b^{\infty} \exp \frac{-[u - 2n(N+S)]}{8nN(N+2S)} du$$

or, with an appropriate change of variable, (3-27)

¹²A detailed discussion of the fluctuations in apparent size of aircraft is given in Paragraph 4-8.

¹³See Fig. 4-24 for an example.

$$P_d(S) \approx \frac{1}{\sqrt{2\pi}} \int_{b-2n(N+S)}^{\infty} \exp(-x^2/2) dx.$$

$$\frac{2\sqrt{\pi N(N+2S)}}{2\sqrt{\pi N(N+2S)}}$$

In Fig. 3-6, $P_d(S)$ is plotted as a function of S . For small values of S , $P_d(S)$ is very small, while for large S , $P_d(S)$ is approximately unity. A transition

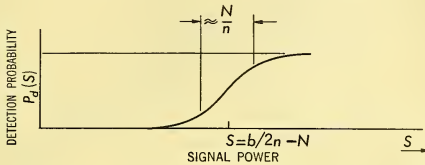


FIG. 3-6 Probability of Detection as a Function of the Signal Amplitude.

occurs when $S = b/2n - N$. The width of the transition region is inversely proportional to n . When n is large, then, it is reasonable to approximate $P_d(S)$ as zero for $S < b/2n - N$ and unity for S larger than this transition point. We shall subsequently indicate that this is a fair approximation even when $n = 1$. With this approximation, the integral in Equation 3-26 is easily evaluated.

$$\bar{P}_d \approx (1/\bar{S}) \int_{b/2n-N}^{\infty} \exp(-S/\bar{S}) dS = \exp \frac{-(b/2nN - 1)}{\bar{S}/N}. \quad (3-28)$$

It is convenient for some developments to work directly with the range to the target instead of the average signal to noise ratio. The expression in Equation 3-10 giving the signal to noise ratio as the fourth power of the ratio of an ideal range to the actual range provides this relation:

$$\bar{S}/N = (R_0/R)^4 \quad (3-29)$$

In addition, a factor $K(n,\eta)$ is defined

$$K = (b/2nN - 1). \quad (3-30)$$

With this notation, the average detection probability can be represented in the following very simple form.

$$\bar{P}_d \approx e^{-K(R/R_0)^4}. \quad (3-31)$$

The K factor can be evaluated from the data in Fig. 3-3 giving the relation between the relative bias, the number of pulses integrated, and the false-alarm number. The results of such an evaluation are shown in Fig. 3-7, where the K factor is plotted versus the number of pulses integrated for representative values of the false-alarm number.

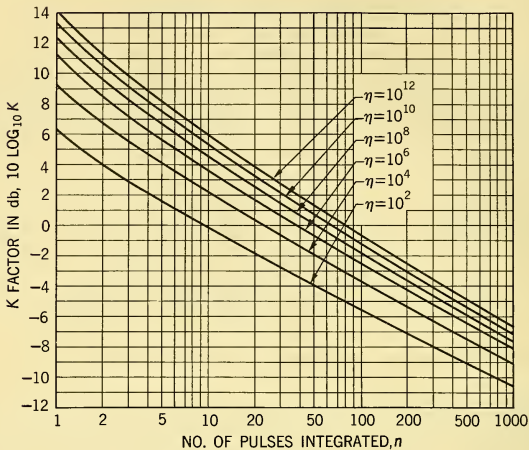


FIG. 3-7 The Factor $K(n, \eta)$ as a Function of n .

It is of interest to note that we can infer from the slopes of the curves in this figure the trade-off of signal-to-noise ratio with n , the number of pulses integrated. From Fig. 3-7, a typical slope is about -6 db for a factor of 10 in the number of pulses integrated. This is equivalent to a variation of K with n of the following form:

$$K \sim n^{-0.6}. \quad (3-32)$$

Because the average probability of detection is a function of the ratio $K/(\bar{S}/N)$, a variation in K is equivalent to an inverse variation in the average signal to noise ratio. The trade-off between signal to noise ratio and n , then, is simply

$$\bar{S}/N \sim n^{0.6}. \quad (3-33)$$

Because of the rather gross approximation which had to be made to obtain Equation 3-31, there is a question about its range of validity. A reasonable validation of this equation is obtained by comparing it with some examples of exact calculations and observing the error. This is done in Fig. 3-8 where the average detection probability as given by Equation 3-31 has been plotted as a function of the normalized range $K^{1/4}(R/R_0)$. It is approximately a straight line on the normal probability coordinates used in that figure. Also plotted in Fig. 3-8 are the exact values of \bar{P}_d for $n = 1$. This is the case when the approximations made introduce the greatest error.

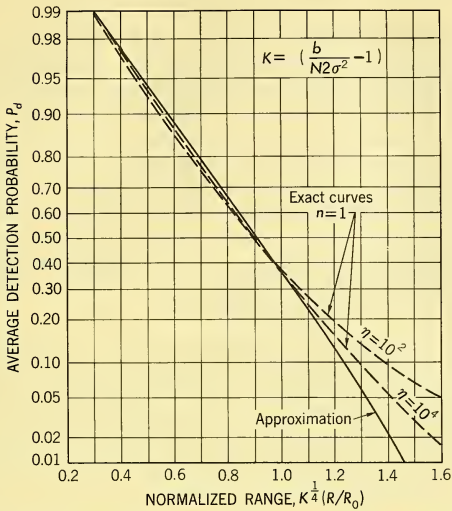


FIG. 3-8 Average Detection Probability as a Function of Normalized Range.

When $n = 1$, the probability density function departs from normality to the greatest degree, and the width of the transition region in Fig. 3-8 is largest. The approximation will also be poorest when the false alarm number is small. Curves are plotted in Fig. 3-8 for the two values $\eta = 10^2$ and $\eta = 10^4$. These are considered small values for this parameter. With the false alarm number η equal to 10^2 , the approximation is already quite good for values of P_d greater than 20 per cent. For larger values of η or n , the approximation becomes very good.¹⁴

It is of interest to compare the average detection probability on a fluctuating target with that obtained on a constant target of the same size. The detection probability on a nonfluctuating target was previously determined in the case of $R_0 = 20.4$ n.mi., $n = 10$, and $\eta = 10^3$. For the fluctuating case, we first determine the value of K from Fig. 3-7 for these parameters. This value is found to be 4.5 db or a factor of 2.82. The fourth root of K is then 1.3. In Fig. 3-8, the normalized range for $P_d = 0.9$ is found to be 0.58. The actual range giving an average 90 per cent probability of detection is thus

¹⁴More detailed development of these ideas can be found in P. Swerling, "Probability of Detection for Fluctuating Targets," *Research Memorandum 1217*, The RAND Corporation, Santa Monica, Calif. (17 March 1954).

$$R_{90\%} = 0.58 \times 20.4/1.3 = 9.1 \text{ n.mi.} \quad (3-34)$$

This range is substantially less than the range (14.7 n.mi.) which gives $P_d = 0.9$ in the case of a nonfluctuating target. From Fig. 3-8, the normalized range giving $P_d = 0.1$ is 1.23, which yields an actual range corresponding to an average detection probability of 10 per cent of 19.3 n.mi. This value corresponds to 17.5 n.mi. in the nonfluctuating case. Thus, while the fluctuations degrade the performance at high probabilities, they enhance the performance on small and distant targets. The complete curve of detection probability versus range in this case is plotted in Fig. 3-5.

3-4 THE EFFECT OF SCANNING AND THE CUMULATIVE PROBABILITY OF DETECTION

Because the beamwidth of a high-gain antenna is normally much smaller than the search area within which a target might appear, the beam must be made to scan over the area. For AEW or ground-mapping systems where the beam is narrow in only one dimension, this motion is generally very simple, either a wigwag or a complete rotation. For systems where the beam is narrow in both azimuth and elevation, the motion of the beam can become quite complex.

The effect of scanning is to provide multiple looks at the target, giving multiple chances for detection. In this case, it is the *cumulative probability of detection* which is most significantly related to the tactical use of the system. Complex scans can produce a nonuniform coverage of the scan area, with holes in the pattern and undesired modulation of the received pulse packet.

Multiple-Scan Probability of Detection. In a typical detection situation, the radar will periodically scan the target and there will be a number of looks at the target when a detection can be made. Moreover, since the target will normally move during the scan time, the average signal-to-noise ratio and thus the average single glimpse probability will vary from scan to scan. This situation is conveniently described by the cumulative probability of detection. When the target is closing on the radar, the cumulative probability of detection at a given range is defined as the probability that the target is detected on or before reaching that range.

We shall assume that the radar closes on the target at a constant rate $-\dot{R}$, and the scan time t_{sc} is also constant. Thus, the range interval which is closed during a scan is given by

$$\text{Range decrement} = \Delta R = -\dot{R}t_{sc} \quad (3-35)$$

If the first look occurs at the range R_1 , then the k -th look will occur at the range

$$R_k = R_1 - (k - 1)\Delta R. \quad (3-36)$$

We shall limit our discussion to a consideration of fluctuating targets which can be handled rather generally thanks to the simplicity of the expression for the average probability of detection (Equation 3-31). At each look, the average probability of detection is given by Equation 3-37 below.

$$P_d(R_k) = e^{-K(R_k/R_0)^4}. \quad (3-37)$$

The cumulative probability that a target is detected at the range R_k or before is denoted by $P_c(R_k)$ and is given by the well-known expression for the probability of at least one success in a sequence of k trials:

$$P_c(R_k) = 1 - \prod_{i=1}^k [1 - P_d(R_i)]. \quad (3-38)$$

An additional refinement needs to be introduced. Equation 3-39 implicitly assumes that the last look occurred at R_k . Actually, the last look may occur anywhere between R_k and $R_{k-1} = R_k + \Delta R$ with equal probability. That is, there will be a random phase between the antenna scan and the relative motion of the target. To take this effect into account, an average value of the cumulative probability of detection must be computed:

$$\bar{P}_c(R_k) = \frac{1}{\Delta R} \int_0^{\Delta R} P_c(R_k + x) dx. \quad (3-39)$$

Fortunately, the calculations shown in Equations 3-38 and 3-39 do not have to be carried out every time the cumulative probability of detection is desired. With properly normalized variables, a universally applicable series of detection curves can be derived.

In order to do this, a normalized range denoted by ρ is defined:

$$\rho = K^{1/4}(R/R_0). \quad (3-40)$$

The normalized range decrement is defined similarly:

$$\Delta\rho = K^{1/4}(\Delta R/R_0). \quad (3-41)$$

With these definitions, the average single-glimpse probability of detection takes the following form:

$$\bar{P}_d = e^{-\rho^4} \quad (3-42)$$

With this form for the single-glimpse probability of detection, universal curves of the average cumulative probability of detection have been calculated on the basis of Equations 3-38 and 3-39. These curves are plotted in Fig. 3-9.

From the appearance of these curves, it would seem desirable to make the normalized decrement $\Delta\rho$ as small as possible in order to obtain the

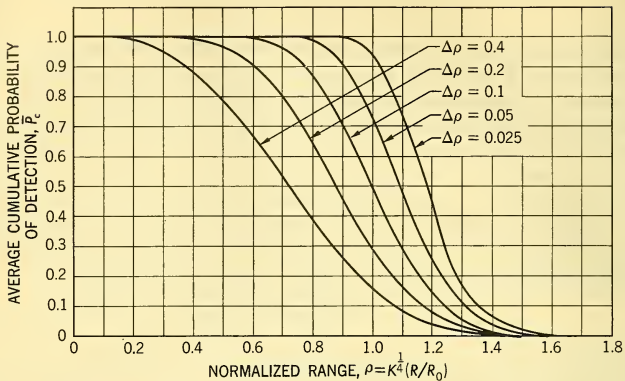


FIG. 3-9 Universal Curves of Average Cumulative Probability of Detection.

maximum range. This is only part of the story, though. Normally, ΔR and thus $\Delta \rho$ would be made small by decreasing the scan time, which in turn is obtained by speeding up the scan. With a higher scan speed, the number of pulses returned on a scan over the target is reduced. This reduces the factor K approximately through the relation in Equation 3-32. The net result is to give an optimum value of scan speed or scan time which maximizes the range at which a given value of cumulative detection probability is obtained. With a slower scan than this optimum, the target closes too much between scans and there will be too few chances to detect it. With a faster scan, there are not enough hits per scan. The determination of this optimum scan time will be illustrated as part of the following example.

To illustrate the use of the curves in Fig. 3-9, we shall continue with the AI radar example which we have previously used in Paragraph 3-3 to illustrate the calculation of the single-scan probability of detection for both constant and fluctuating targets. We assume that the target closes on the radar at 2000 ft/sec or about Mach 2. The scan time was assumed to be 3 seconds. The range decrement is $3 \times 2000 = 6000$ ft or 1.0 n.mi. The value of $K^{1/4}$ was previously determined to be 1.3 while the idealized range is 20.4 n.mi. The normalized range decrement is thus

$$\Delta \rho = 1.3 \times \frac{1}{20.4} = 0.0635. \quad (3-43)$$

Referring to Fig. 3-9, the normalized range giving a cumulative probability of detection of 90 per cent for $\Delta \rho = 0.0635$ is $\rho = 0.87$. The equivalent actual range is

$$R_{90\%} = 0.87 \times \frac{20.4}{1.3} = 13.7 \text{ n.mi.} \quad (3-44)$$

The complete curve of cumulative probability of detection versus range is given in Fig. 3-5 along comparable single-scan curves for both a constant and fluctuating target.

Finding an Optimum Scan Time. In order to demonstrate how an optimum scan time would be determined, the calculations made above for a scan time of 3 seconds will be repeated for scan times of 1, 2, 5, and 10 seconds as well. The values of N , K , $K^{1/4}$, $\Delta\rho$, $\rho_{90\%}$ and $R_{90\%}$ are given in Table 3-1.

TABLE 3-1

Scan Time	n	K	$K^{1/4}$	$\Delta\rho$	$\rho_{90\%}$	$R_{90\%, n. mi.}$
1 sec	3.33	6.02	1.57	0.0256	1.025	13.3
2 sec	6.67	3.63	1.38	0.0451	0.93	13.7
3 sec	10.0	2.82	1.295	0.0635	0.87	13.7
5 sec	16.67	2.04	1.197	0.0976	0.785	13.4
10 sec	33.33	1.305	1.058	0.1725	0.655	12.65

The detection ranges given in Table 3-1 are plotted in Fig. 3-10. From this figure and the table, it is apparent that the optimum scan time in this

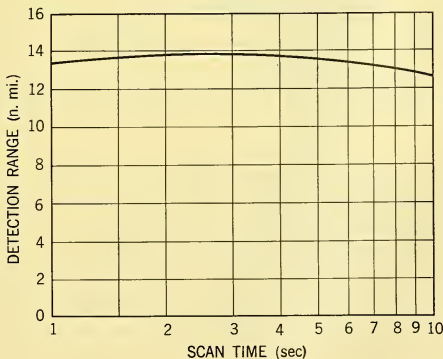


FIG. 3-10 Example of Detection Range vs. Scan Time.

case is the original choice of 3 seconds. Another observation which can be made in Fig. 3-10 is that the optimum is very broad, and it actually does not make a great deal of difference whether a scan time of 2 seconds or 5 seconds is selected if the rest of the system is made compatible. This kind of

situation is often true in matters of this nature and is often not generally recognized until after elaborate studies have been made, if at all.

A question which often comes up in connection with discussions of beamwidth and scanning is, why scan at all? Why not simply use a wider beam and a fixed antenna? This thought has a good deal of merit to it. The loss in gain due to the use of a wide beam can be offset by the integration of a much larger number of pulses, and the actual detection ranges might very well be comparable. A narrow beam, though, has other advantages which make its use desirable. One of these is that upon detection, the location of the target is known at once so that tracking can commence immediately. Further, the resolution which can only be provided by a narrow beam is often a basic tactical requirement of the system (see Paragraph 2-13). In addition, a narrow beam is often required to give sufficient accuracy during track or to provide a means for narrowing the scan area and "search-lighting" a suspected target.

Types of Scans. Fig. 3-11 shows some scan patterns which have been used with pencil beam systems. The most common type of scan is a simple

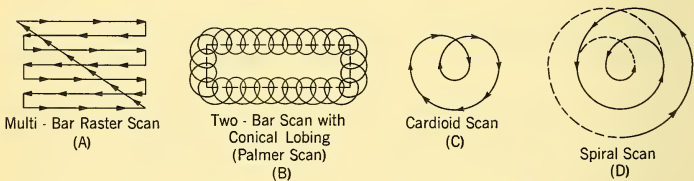


FIG. 3-11 Some Possible Scan Patterns with a Pencil Beam System.

constant-velocity raster scan with a fly-back at the bottom of the pattern. With a large area to be covered, up to seven or eight bars might be required. Very often the basic scan is modified by a lobing motion. Conical or circular lobing may be used during track to generate angular error signals. During search the lobing may be left on, either because there is no convenient way in which to stop and start the lobing motor or because the larger equivalent beamwidth can be utilized to cut down on the number of scan bars. When this is done, the circular lobing motion combined with the constant-velocity azimuth motion produces a cycloidal scan of the beam centers (Fig. 3-11b). This type of scanning motion is often referred to as a *Palmer scan* because of its resemblance to a pennmanship exercise. The cardioid and spiral scans shown in Fig. 3-11 represent attempts to minimize the fly-back or dead time. They are not generally regarded as normal designs, but may be required for some applications.

The Number of Pulses per Scan. In the system model adopted in Paragraph 3-3, to develop an analytical method for calculating the probability of detection, it was assumed that n equal-amplitude pulses were received on a scan over a target. With a complex scan and realistic beam shapes, the pulses received are not all of the same amplitude; neither is it clear just what n should be in many situations. For instance, with the Palmer scan illustrated in Fig. 3-11b, the pulses received on a single scan over a target may be grouped into several separate packets by the cyclical motion imposed on the basic scan. The grouping and number of pulses in the individual packets can then change with the location of the target in the scan pattern.

In order to analyze situations of this nature correctly and in detail, extensive analytical investigations are often required. More commonly, it is quite adequate to make reasonable approximations which will allow the methods developed in Paragraph 3-3 to be applied. This is what we shall do here.

We consider first the problem of estimating the effect of the antenna beamshape in a linear scan over a target. We suppose that the antenna pattern has a Gaussian shape similar to that defined in Equation 3-45:

$$\text{Two-way power pattern of antenna} \sim e^{-\theta^2/0.18\Theta^2} \quad (3-45)$$

where θ = angular position of the antenna

Θ = antenna beamwidth (half-power, one-way).

We wish to approximate this antenna pattern by a uniform pattern so that the results of the preceding paragraph are applicable. This type of approximation is indicated in Fig. 3-12. In making this approximation, the

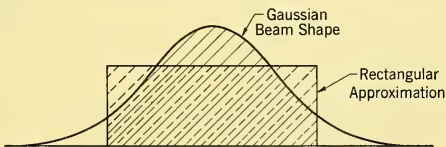


FIG. 3-12 Rectangular Approximation to Gaussian Beam Shape (Equal Area Approximation).

total integrated power will be maintained constant. That is, the integral of the uniform approximation will be made equal to the integral of the antenna power pattern between the effective limits of integration. This will result in an equivalent loss in signal power for pulses in the uniform pattern in comparison with pulses in the center of the more realistic pattern. This loss is referred to as the *scan loss*. Following current practice, we suppose that

the effective number of pulses integrated n are those contained within the antenna beamwidth. The optimum number of pulses to integrate will differ slightly from this.¹⁵

The effective power within the antenna beamwidth will be proportional to the following integral.

$$\text{Total received power} = \int_{-\theta/2}^{\theta/2} e^{-\theta^2/0.18\theta^2} d\theta = 0.68\theta. \quad (3-46)$$

Thus, where the maximum power of the pulses in a Gaussian beam is unity, the equivalent power of uniform pulses is only 0.68, giving a scan loss of 1.7 db.¹⁶

A second problem concerns the number of pulses integrated when the scan is complex. Where it is probable that there is a substantial non-uniformity in the pulse distribution, a pulse count should be carried out. That is, the actual number of pulses returned from typical target locations for a sample scan would be determined by counting them. More usually, it is adequate simply to use the average number of pulses per scan as was done in Paragraph 3-3 for the example illustrating the calculation of the single-scan probability of detection. The beam area was divided into the scan area to give the number of beams per scan. This number was in turn divided into the total number of pulses per scan to yield the received pulses per scan.

3-5 THE CALCULATION OF DETECTION PROBABILITY FOR A PULSED DOPPLER RADAR

With proper interpretation, the methods developed in Paragraphs 3-3 and 3-4 are applicable to a variety of types of radar systems. To illustrate how this can be carried out, we shall develop some of the details of such an application to the gated pulsed doppler radar described in Paragraph 6-6, whose functional block diagram is given in Fig. 6-25. This type of radar transmits pulses at a very high repetition rate in order to avoid doppler frequency ambiguities. The duty ratio is also considerably greater than in a conventional pulse radar. All the possible target ranges (ambiguous) are gated into separate filter banks which cover the spectrum of possible doppler frequencies. The filters respond to the fundamental component of the gated doppler signal which is received.

Single-Scan Probability of Detection. The idealized range for this type of system is essentially given by Equation 3-9. This is restated in Chapter 6 as Equation 6-39 with the effects of the signal and gating duty

¹⁵L. V. Blake, "The Number of Pulses per Beamwidth in a Scanning Radar," *Proc. IRE*, June, 1953.

¹⁶A scan loss of 1.6 db was obtained by L. V. Blake in the paper cited in footnote 15.

cycles specifically incorporated. For a gated pulse doppler system the noise is gated with the same duty factor as the signal so that $d_g = d_s$ in Equation 6-39.

Detecting only the fundamental doppler signal in a filter output corresponds to the case of detecting a single pulse in Paragraph 3-3. Thus the basic single-scan probability of detecting a fluctuating target should be given by the expression in Equation 3-31 with the factor K corresponding to the integration of a single pulse.

In order to account for certain features of the pulsed doppler system, this basic probability must be modified somewhat. This modification is due to the straddling of a pulse by contiguous range gates and the eclipsing of part of the received pulse by the transmitted pulse. These effects act to decrease the single-scan probability of detection from its basic value. This reduction is denoted by the factor $F(R)$. Thus the probability of detection of a fluctuating target by a pulsed doppler radar can be represented by

$$\bar{P}_d \approx F(R)e^{-K(R/R_0)^4}. \quad (3-47)$$

The Straddling and Eclipsing Factor. *Range gate straddling* refers to the situation when the received signal simultaneously falls within two range channels. This situation is illustrated in Fig. 3-13, where the received

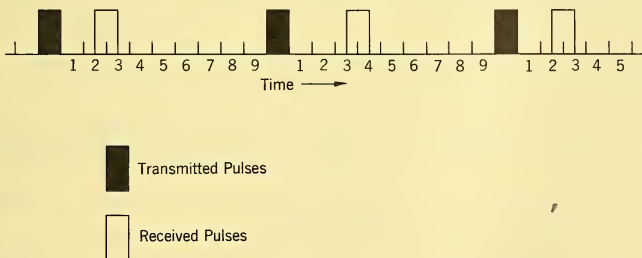


FIG. 3-13 Range Gate Straddling.

pulse lies partly in channel 2 and partly in channel 3. Since the gating of only a fraction of the received pulse into a given channel is equivalent to decreasing the duty ratio by this factor and since the noise in that channel is undiminished, the signal power in the channel will be proportional to the square of the fraction of the pulse within the channel gate. Thus, if a fraction α of the received pulse falls within gate k and the fraction $1 - \alpha$ falls within gate $k + 1$, the signal power in the first gate will be proportional to α^2 while that in the second channel will be proportional to $(1 - \alpha)^2$. When half the pulse lies in each gate, there will be a loss of 6 db in each

channel. Of course, there will be two chances to detect the target. If the probability of detecting the target in the first channel is denoted by P_1 and that for the second channel by P_2 , the probability of detecting the target in at least one of the channels will be

$$P_{d2} = 1 - (1 - P_1)(1 - P_2) = P_1 + P_2 - P_1P_2. \quad (3-48)$$

The straddling factor will be periodic in range, with a period equal to the pulse length in range units. In the example illustrated in Fig. 3-14, the

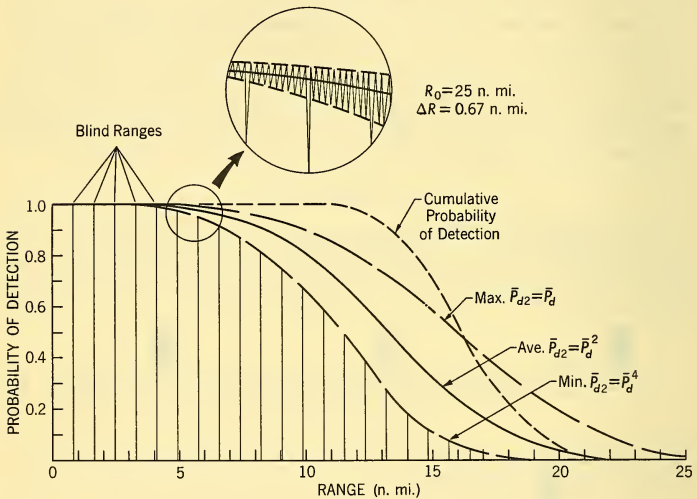


FIG. 3-14 Single Scan and Cumulative Probabilities of Detection for a Pulsed Doppler Radar.

oscillations of the detection probability with a period of about $1/12$ n.mi., which are shown in the expanded view, represent the effects of straddling.

When the received pulse straddles the transmitted pulse, eclipsing is produced because the receiver is gated off when the transmitter is on to prevent feed-through. The effect of eclipsing is much more severe than that of gate straddling. When the received pulse is centered on the transmitted pulse, the signal received and the resulting probability of detection become zero and produce a blind range at which the system is completely insensitive. These blind ranges are periodically spaced at intervals equal to the repetition period measured in range units. The nulls in the curve in Fig. 3-14 at intervals of slightly less than a nautical mile represent the effects of eclipsing.

It is not completely correct to substitute the average values of the detection probability in each channel into Equation 2-48 when considering a fluctuating target because the signal will fluctuate similarly in the two channels. Instead, we should use the procedure previously used in Equation 3-28 for finding the approximate average value of a single channel to determine the average of the two-channel expression given in Equation 3-48.

Equation 3-28 was derived on the basis of an approximation to the curve in Fig. 3-6 that P_d was zero out to the value $S = KN$ and unity for higher values of S . With this approximation, P_1 and P_2 become

$$P_1(\alpha^2 S) \approx \begin{cases} 0, & S < KN/\alpha^2 \\ 1, & S > KN/\alpha^2 \end{cases} \quad (3-49)$$

$$P_2[(1 - \alpha)^2 S] \approx \begin{cases} 0, & S < KN/(1 - \alpha)^2 \\ 1, & S > KN/(1 - \alpha)^2 \end{cases}$$

Adopting these approximate expressions, it is apparent that the product $P_1 P_2$ is equal to P_1 when $\alpha < 1/2$, and to P_2 when $\alpha > 1/2$. This observation materially simplifies Equation 3-48, since only one term is retained:

$$P_{d2} \approx \begin{cases} P_2, & \alpha < \frac{1}{2} \\ P_1, & \alpha > \frac{1}{2} \end{cases} \quad (3-50)$$

The average value of P_{d2} in each case will be of the exponential form first given in Equation 3-31:

$$\bar{P}_{d2} \approx \begin{cases} \exp - K(R/R_0)^4 [1/(1 - \alpha)^2] = (P_d)^{1/(1 - \alpha)^2}, & \alpha < \frac{1}{2} \\ \exp - K(R/R_0)^4 (1/\alpha^2) = (P_d)^{1/\alpha^2}, & \alpha > \frac{1}{2} \end{cases} \quad (3-51)$$

The minimum value is attained when $\alpha = \frac{1}{2}$, and the received pulse is divided equally between the two channels:

$$\min \bar{P}_{d2} = \bar{P}_d^4. \quad (3-52)$$

Of more interest is the average value, which can be used as a smooth replacement for oscillatory curves similar to that in Fig. 3-14 in many cases. The function in Equation 3-51 could be integrated by numerical means. It is more expedient, though, and a good approximation to simply use the average of $1/\alpha^2$, which = 2 in Equation 3-51. Thus,

$$\text{ave } \bar{P}_{d2} \approx \bar{P}_d^2. \quad (3-53)$$

On the average, then, the effect of straddling can be interpreted as a 3-db loss in signal-to-noise ratio.

It should also be noted that \bar{P}_{d2} is of the same general form as \bar{P}_d itself. Thus, if the effects of eclipsing can be neglected, the methods developed in Paragraph 3-4 for determining the cumulative probability of detection and the normalized curves in Fig. 3-9 are quite applicable.

An approach similar to that taken in this paragraph should be applicable to many similar problems. For instance, a multiple-PRF method of determining range in a high-PRF pulse doppler system is described in Paragraph 6-6. In order to determine range on a given scan over the target, it must be detected in all PRF's, and the return must not be eclipsed nor can there be interference with a return from another target with the same doppler shift but at a different range. Calculating the probability of measuring range in such a situation is quite complicated, but should be possible with the methods indicated.

An Example. The following system parameters of a gated pulsed doppler radar are assumed to provide an illustration of the methods under discussion.

$$\begin{aligned}
 R_0 &= \text{idealized range} &= 25 \text{ n. mi.} \\
 \eta &= \text{false-alarm number} &= 10^2 \\
 \tau &= \text{pulse width} &= 1\mu\text{sec} \\
 f_r &= \text{pulse repetition rate} &= 100 \text{ kc/sec} \\
 d &= \text{duty ratio} &= 0.1 \\
 n &= \text{pulses integrated} &= 1 \\
 \dot{R} &= \text{closing rate} &= 0.33 \text{ n. mi./sec} \\
 t_{sc} &= \text{scan time} &= 2 \text{ sec} \\
 \Delta R &= \text{range decrement} &= 0.67 \text{ n.mi.}
 \end{aligned}$$

For $\eta = 10^2$ and $n = 1$, the value of K is found from Fig. 3-7 to be 6 db, or $K = 4$. The basic single-scan probability of detection of a fluctuating target is thus

$$\bar{P}_d = e^{-4(R/25)^4}, \quad R \text{ in n.mi.} \quad (3-54)$$

This probability has been plotted in Fig. 3-14 as the maximum value of \bar{P}_{d2} . Also plotted in that figure are \bar{P}_d^4 and P_d^2 corresponding to the minimum and average values of \bar{P}_{d2} . The shaded area between the minimum and maximum values of \bar{P}_{d2} is composed of many oscillations with a period of about 1/12 n.mi. This is illustrated in the expanded view. At intervals of slightly less than a mile, one of these oscillations deepens into a complete null due to the eclipsing to give a narrow blind region.

When the effect of the eclipsing is neglected, the cumulative probability is easily determined from Fig. 3-8. Remembering that straddling has the effect of doubling the effective value of K , the normalized range corresponding to \bar{P}_d^2 is defined by

$$\rho = 8^{1/4} \left(\frac{R}{25} \right) = \frac{R}{14.9}. \quad (3-55)$$

The normalized range decrement is thus

$$\Delta\rho = \frac{\Delta R}{14.9} = \frac{0.67}{14.9} = 0.045. \quad (3-56)$$

The resulting cumulative probability is also plotted in Fig 3-14.

Postdetection Filtering. It is not uncommon in pulsed doppler systems to use a predetection doppler filter which is considerably wider than the reciprocal of the observation time of the signal. Subsequent postdetection filtering is matched to the signal observation time to provide the maximum output signal-to-noise ratio. In this manner the number of doppler filters required can be materially reduced at the expense, of course, of velocity resolution. The filtering or integration is also somewhat less efficient because it is noncoherent representing an operation on the detected signal plus noise.

An exact analysis of postdetection filtering is not possible in general, and we shall look for reasonable approximations. Postdetection filtering is essentially similar to video pulse integration, whose effect on detection was discussed in some detail in Paragraph 3-3, and it is natural to use this approach in establishing the approximate effect of this operation. What we shall do is to derive an equivalent predetection bandwidth which provides approximately the same detection performance as the combination of pre- and postdetection filters which it represents. It is assumed that the target fluctuates from scan to scan but has a constant size during the observation time.

The following notation is adopted:

B = predetection bandwidth (band pass)

b = postdetection bandwidth (low pass)

B' = equivalent predetection bandwidth (band pass)

n = equivalent number of signal samples integrated

The output of the bandpass predetection filter can be represented by a series of samples separated by $1/B$ (seconds) as was indicated in Paragraph 3-3 where the sampling theorem is quoted. Similarly, the output of the low-pass, postdetection filter can be represented by a series of samples spaced by $1/2b$ (seconds). In order to provide signal integration the postdetection sampling time will be longer than that of the predetection signal. The ratio of these sampling times gives the number of predetection samples which are integrated in the postdetection filter:

$$\text{Equivalent number of samples integrated} = n = B/2b. \quad (3-57)$$

Now in general the predetection signal-to-noise ratio is proportional to the reciprocal of the predetection bandwidth:

$$S/N \sim \frac{1}{B} \quad (3-58)$$

Also, the equivalent signal-to-noise ratio of a fluctuating target is proportional to a power of the number of video pulses integrated as in Equation 3-33:

$$S/N \sim n^\gamma \quad (3-59)$$

The appropriate power γ corresponds to the slopes of the curves in Fig. 3-7.

The equivalent gain in signal-to-noise ratio obtained through postdetection integration can now be expressed either as the ratio of the actual and equivalent predetection bandwidths or simply as n^γ :

$$\text{Equivalent gain in } S/N \approx \frac{B}{B'} \approx n^\gamma = \left(\frac{B}{2b}\right)^\gamma \quad (3-60)$$

The equivalent predetection bandwidth thus is given approximately by

$$B' \approx (2b)^\gamma B^{1-\gamma} \quad (3-61)$$

In Equation 3-33, the value of γ was found to be 0.6. If this is stretched a point and assumed to be 0.5, the following simple expression is obtained:

$$B' \approx \sqrt{2bB} \quad (3-62)$$

This approximation is often used for estimating performance where post-detection filtering is involved.

3-6 FACTORS AFFECTING ANGULAR RESOLUTION

In many applications, it is required that a radar system be capable of separating or distinguishing closely spaced targets. This capability is referred to as the *resolution* of the system. Targets may be resolved on the basis of any of their characteristics. Thus they may be distinguished in range, velocity, or angular position. This paragraph discusses angular resolution.¹⁷ In ground mapping, the radar's angular resolution provides a primary means of target discrimination. In AEW radar systems, the angular resolution of the system breaks up multiple target complexes into individual components to provide an estimate of the threat. In fire-control radar, the angular resolution must be sufficient to separate desired targets from interfering targets and clutter.

¹⁷A similar discussion can be found in J. Freedman, "Resolution in Radar Systems," *Proc. IRE* 39, 813-1818 (1951), upon which parts of this section are based.

Antenna Pattern Characteristics. Angular resolution is provided by the directive properties of the radar antenna. The greater the directivity, the better the resolution.

There is an enormous variety of types of microwave antennas in use today. The most widely used in airborne radar systems are those employing parabolic reflectors. The discussion will center about this type of antenna although many of the observations are applicable to a much wider class.

Parabolic reflectors can be constructed whose characteristics closely approximate those of a uniformly illuminated aperture. The relative voltage pattern radiated (or received) by a uniformly illuminated circular aperture will have the following form¹⁸ at long ranges.

$$\text{One-way voltage pattern, circular aperture} = \frac{2J_1[(\pi D/\lambda) \sin \theta]}{(\pi D/\lambda) \sin \theta} \quad (3-63)$$

where D = aperture diameter

λ = wavelength

θ = angle relative to aperture normal

$J_1(\)$ = first-order Bessel function.

For convenience, we represent the argument of this expression by x so that the one-way relative voltage pattern is $2J_1(x)/x$.

The received voltage reflected from a point target to a uniformly illuminated circular aperture used both for transmission and reception will be given by the square of the function in Equation 3-63 or $(2J_1(x)/x)^2$. This is also equal to the one-way relative power pattern of such an antenna. This pattern is illustrated in Fig. 3-15 where it is referred to as the two-way voltage envelope generated by a scan over a single target.

The antenna beamwidth is normally defined as the width between the half-power points of the one-way antenna pattern. This is indicated in Fig. 3-15. For a uniformly illuminated circular aperture the beamwidth is related to the diameter and wavelength by

$$\text{Beamwidth, circular aperture} = 58\lambda/D \text{ degrees.} \quad (3-64)$$

The envelope of the received power on the two-way power pattern is probably most significant for defining resolution. This is given by the square of the envelope plotted in Fig. 3-15 or $(2J_1(x)/x)^4$.

The antenna pattern and the beamwidth can be modified by illuminating the aperture in a nonuniform manner. A uniform illumination yields one of the narrowest beams, but the sidelobe level is relatively high. The sidelobes of the one-way power pattern in Fig. 3-15 are down 17.6 db from the peak. When the illumination is tapered or stronger in the center of the

¹⁸J. D. Kraus, *Antennas*, p. 344, McGraw-Hill Book Co., Inc., New York, 1950.

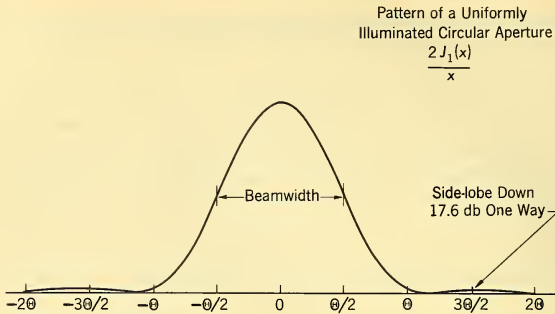


FIG. 3-15 Two-Way Voltage Envelope Generated by a Scan over a Single Target.

aperture than at the edge, the sidelobe level can be minimized, but at the expense of a wider beamwidth. In actual practice it is customary to taper the illumination so that the effective beamwidth is about 20 per cent greater than indicated by Equation 3-64; i.e., the multiplying factor becomes 70 rather than 58.

Resolution Criteria. When two targets are separated sufficiently, they can be identified as two distinct targets. When they are brought together, their returns merge into a single unresolved return. There are a number of criteria for deciding just when there are two returns and when there is only one. Fundamentally, resolution should be defined relative to the discrimination abilities of the human operator in the particular system involved. In general, though, this is much too complex an approach because of the many factors affecting human performance, and it is more convenient to adopt an arbitrary definition of resolution. In some cases, this will lead to a situation where targets which are defined to be unresolved can actually be observed as separate entities. Most of the definitions which have been suggested for angular resolution lead approximately to the same result: targets separated by about 1 beamwidth can be resolved. A *beamwidth* is normally defined as the width between half-power points of the main lobe. We shall adopt a very similar definition of resolution which has the convenient virtue of yielding a resolution of 1 beamwidth for a uniformly illuminated circular aperture. We shall say that *two point targets are resolved when the average minimum of the received power envelope in a scan over them is less than half the power from the maximum of the smaller of the two.*

This definition is illustrated in Fig. 3-16, the two-way voltage envelope received from two point targets which are just resolved. As indicated in the figure, the voltage pattern fluctuates markedly depending upon whether the returns are in phase or out of phase. When the received reflections are

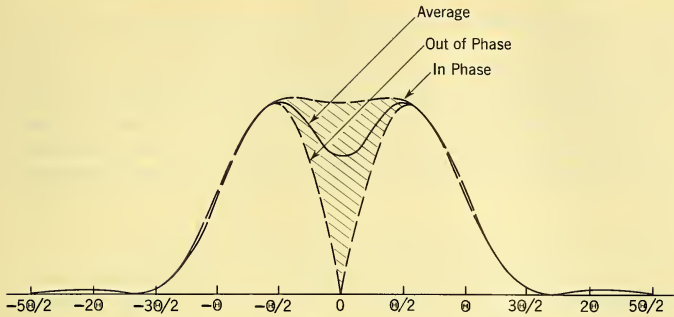


FIG. 3-16 Two-Way Voltage Envelope Generated by a Scan over Two Targets Separated by One Beamwidth.

in phase, only a small notch separates the two targets — they have merged in a single return. When the received signals are out of phase, there is a sharp null midway between the two targets. An average envelope can be determined for a random phase between the two reflections. This average two-way voltage envelope is also shown in Fig. 3-16. The minimum of this average curve is 0.707 of the maxima corresponding to half of the maximum received power. Consequently, the case illustrated in Fig. 3-16 shows the envelope of two targets which are just barely resolved. These targets are separated by a single beamwidth. Thus the definition of resolution adopted conveniently yields one beamwidth for two targets of *equal* size.

Degrading Influences. In most practical situations the resolution will be degraded somewhat by a variety of factors. One such factor is unequal strength of the targets. In Fig. 3-17, the two-way voltage envelope

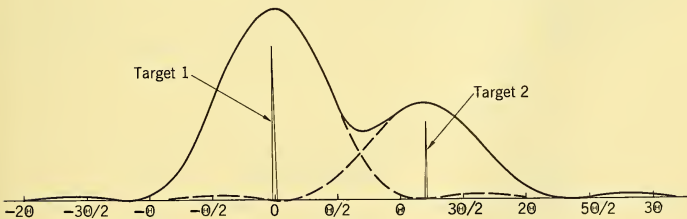


FIG. 3-17 Average Two-Way Voltage Envelope Generated by a Scan over Two Separated Targets of Unequal Size (4 : 1 Power Ratio).

is shown of two targets whose maximum received voltages have a 2-to-1 ratio. The radar size of these two targets is normally expressed in terms of

the ratio of their reflected powers, which is 4-to-1 or 6 db. The minimum separating the two targets in Fig. 3-17 is 0.707 of the smallest maximum, so that these targets are just barely resolved. The target separation required to achieve this resolution is 1.21 beamwidths. Thus, with a 4-to-1 size ratio for targets, the resolution is 21 per cent greater than for targets of equal size. This can become important when the target's size fluctuates randomly. Fig. 3-18 shows how the effective resolution angle varies with target power ratio.

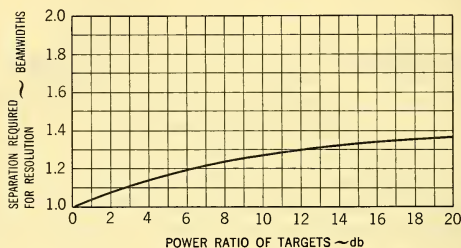


FIG. 3-18 Resolution as a Function of Target Power Ratio.

Another factor which can affect resolution is the signal-to-noise ratio. The simplest way to account for this factor is to apply the already adopted definition for resolution to the received signal-plus-noise power envelope. The deterioration of angular resolution with signal-to-noise ratio which can be determined in this manner is shown in Fig. 3-19.

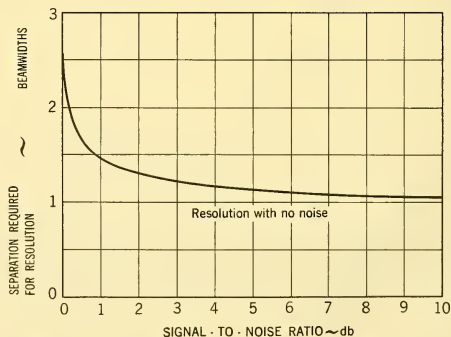


FIG. 3-19 Resolution of Two Equal Targets as a Function of Signal-to-Noise Ratio.

Very large degradations of resolution can often be attributed to nonlinearities in the receiving system. The dynamic range of many search radar systems is less than 10 db above the average noise level, and 20 db is rare. The apparent beamwidth when scanning a very strong target with a system which has limited dynamic range can be as great as twice the normal beamwidth. In such cases, it is quite possible for large targets to completely blank out smaller adjacent targets which might have been resolved with a linear system.

Two other minor factors might be noted, the effects of pulsing and the system bandwidth. When only a limited number of pulses compose the envelope generated by a scan over the target, the exact form of the continuous envelope is somewhat indeterminate. As an extreme example, if only two pulses are received during a scan over a target, the question arises as to whether these are two pulses from a single strong target or from two weaker targets. The effect of pulsing can be regarded as a widening of the effective beamwidth. Equation 3-65 gives a simple and useful approximation for the equivalent effective beamwidth in terms of the actual beamwidth and the angular interval between pulses:

$$\text{Effective beamwidth} \approx \sqrt{\theta^2 + \Delta\theta^2} \quad (3-65)$$

where θ = antenna beamwidth

$\Delta\theta$ = angular interval between pulses.

The antenna pattern described by Equation 3-63 and illustrated in Fig. 3-15 assumed monochromatic radiation. In some applications where very wide bandwidths are required, the antenna beamwidth will be modified. Such an application might be the use of microwave radiometers for mapping. When there is no chromatic aberration (approximately true when a parabolic reflector is used) and the average frequency is maintained constant, the increase in beamwidth with bandwidth is small. A maximum beamwidth increase of about 5 per cent is given for a bandwidth of 15 per cent of the average frequency.

CHAPTER 4

REFLECTION AND TRANSMISSION OF RADIO WAVES

4-1 INTRODUCTION

In the propagation of radio waves between a transmitter and receiver, we are interested in the problems associated with power transfer between two terminals. This involves an antenna problem at each terminal (that is, the transformation of electrical power into electromagnetic waves or vice versa) and the problem of determining how the waves propagate to the receiver. In the case of airborne radar, the receiving antenna is replaced by the target, and interest is centered in reradiation by the target in the reverse direction, back toward the transmitter. This reradiation phenomenon is usually called *scattering*. The radar case with which we shall be primarily concerned is a special case of scattering in which the angle between the propagation directions of incident and scattered fields is 180° . Scattering may be viewed as an antenna problem, too, for the incident field sets up in the target currents whose distribution depends on the target material and configuration and on the distribution of the incident field. If this current distribution is known, then the field reradiated by it can be determined just as though that current distribution were set up in an antenna. In propagation back from the target to the radar, the scattered wave is involved with the same factors as in propagation from the radar to the target: the radar problem involves (1) two-way propagation, and (2) back-scattering by the target. Thus, in order to predict the strength of the echo received from a target it is necessary to determine the characteristics of the propagation mechanism and also the back-scattering properties of the target.

The frequencies normally used for radar operation range from about 100 Mc/sec on up, or wavelengths of 3 meters down to less than 1 cm (see Fig. 1-21). Consequently most targets are many wavelengths in dimension. An antenna of corresponding size would have an extremely sharp radiation pattern, so that the target, considered as an antenna, has a correspondingly sharp scattering pattern. It follows that in general the field scattered backward is very sensitive to target orientation. Targets which move, therefore, usually give a radar echo which varies with time. Since a

differential radial movement of a half-wavelength of the target or a portion of it is sufficient in many cases to produce a profound variation of echo amplitude, even such targets as trees, towers, and buildings, normally considered stationary, frequently give fluctuating echoes. For a given target, the rate of fluctuation usually will be proportional to radar frequency.

The current distribution set up in the target depends on the distribution of the incident field. In many common situations, the incident field is rather uniformly distributed over the target aperture, so that the target may be considered to be illuminated by a uniform plane wave. Then the scattering characteristics of the target may be analyzed independently of the propagation factors. This is permissible in the case of most airborne or *elevated targets*. More generally, however, the incident field may be distributed nonuniformly over the target, because of the nature of the propagation phenomena obtaining between the transmitter and various portions of the target. A ship is an example of a target in which the incident field varies over the target aperture because of the interference between direct and surface-reflected rays, which gives a resultant amplitude that varies with height. In such cases the scattering properties of the target cannot be separated from the propagation factors, so that a specification of the target properties becomes more complicated and involves the propagation factors. This same type of complication is also involved in sea and ground return.

The principal propagation factors which affect airborne radar are the following.

1. Reflection from the ground
2. Attenuation by liquid water drops in the air
3. Absorption by atmospheric gases
4. Refraction in the atmosphere

This chapter will be devoted to a discussion of these factors and to a description of the characteristics of the principal radar targets of interest in airborne applications; viz., aircraft, sea return, ground return, and rain.

4-2 REFLECTION OF RADAR WAVES

The radar equation for free space is derived in Chapter 3 (Equation 3-9). It may be modified to account for the effect of obstacles such as the earth's surface or an inhomogeneous atmosphere by introduction of a quantity called the *propagation factor*, which is the factor by which the free-space field is to be multiplied to obtain the actual field. This factor, which we denote by F , is a complex quantity, or *phasor*, representing the modification in amplitude and phase of the free-space field by the actual propagation

process. F may be a function of the range and other parameters of the particular situation at hand. Thus, the radar equation becomes

$$P_r = \frac{PG^2\lambda^2\sigma}{(4\pi)^3 R^4} |F|^4. \quad (4-1)$$

The quantity σ is variously called the *radar area*, *radar cross section*, *echoing area*, and *back-scattering cross section*.

It is sometimes useful to relate σ to another quantity known as the *radar length*, designated by l . This is a phasor which represents the ratio, in amplitude and phase, of the back-scattered field-at-unit-distance to the incident field strength. Its relation to σ is

$$\sigma = 4\pi |l|^2. \quad (4-2)$$

The radar length bears a relation to the received field strength similar to that of radar area to received power in Equation 4-1. Thus, the received field strength E_r is given by

$$E_r = lE_0 F^2 \frac{e^{-j2KR}}{R^2} \quad (4-3)$$

where E_0 = the transmitted field at unit distance (the far field extrapolated to unit distance from the transmitting antenna)

$K = 2\pi/\lambda$ = phase constant which expresses the relationship between distance and the phase angle of a transmission of wavelength λ .

The radar area σ may be very much larger than the actual projected area of the target. This may be shown in the following way. If the target is large relative to the wavelength, then it is essentially correct to consider that it intercepts a power P' equal to the product of its projected area A' and the incident power density W_i ,

$$P' = A'W_i. \quad (4-4)$$

The currents set up in the target by the intercepted field will produce a far field which has a certain directive characteristic, just as if the target were an antenna with such a current distribution. Hence the target will have a directive gain which is a function of angle. If we call the directive gain in the radar direction G' , then the effective power reradiated backwards will be

$$P'G' = A'G'W_i = \sigma W_i. \quad (4-5)$$

Hence

$$\sigma = A'G'. \quad (4-6)$$

It is obvious that if G' is large, then σ will be large relative to the actual projected area A' . As an example, consider a target in the form of a flat

metallic sheet perpendicular to the direction of the incoming wave. If we neglect edge effects, the current density is of constant amplitude and phase throughout the sheet. Accordingly, the sheet reradiates like an antenna of aperture A' with a uniform amplitude and phase distribution. Since the gain of such an antenna is

$$G' = 4\pi A'/\lambda^2 \quad (4-7)$$

(which is large if A'/λ^2 is large) we obtain from Equation 4-6

$$\sigma = 4\pi(A'/\lambda^2)^2, \text{ for } A'/\lambda \gg 1 \quad (4-8)$$

One of the conditions assumed in deriving Equation 4-1 is that the variation in range R over the target results in a negligible variation of the phase of the incident field. In order to obtain a numerical estimate of the significance of this limitation, we may consider, as an example, an airborne search radar viewing, in free space, a rectilinear target of length $2L$ at a range R , as illustrated in Fig. 4-1. The difference in range between a point at x on the target and the nearest point of the target is

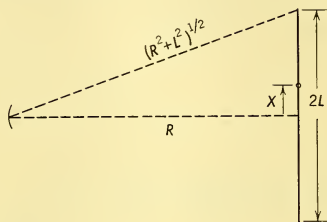


FIG. 4-1 Geometry for Limitations of Plane-Wave Conditions.

$$\Delta R = (R^2 + x^2)^{1/2} - R \approx x^2/2R. \quad (4-9)$$

Assuming that the antenna may be treated as a point source, the round-trip phase difference between the fields reflected back to the source from these two points is

$$\Delta\phi = 2k\Delta R = 2\pi x^2/\lambda R. \quad (4-10)$$

From Equation 4-3 the contribution of a differential length of the target to the received signal in free space is

$$dE_r = \frac{E_0}{R^2} e^{-i2k(R+\Delta R)} dl \quad (4-11)$$

where dl = differential radar length of the differential target element dx located at a distance x from the center of the target.

If we denote the plane-wave radar length of the target by l and assume for simplicity that the radar length per unit length of the target is constant,

$$\frac{dl}{dx} = \frac{l}{2L} \quad (4-12)$$

then, neglecting the slight effect of variations in the $\frac{1}{R^2}$ term, the total field received will be

$$\begin{aligned} E_r &= \frac{E_0}{R^2} e^{-i2kR} \int_{X=-L}^{X=L} e^{-i\Delta\phi} dl \\ &= \frac{lE_0}{2LR^2} e^{-i2kR} \int_{-L}^L e^{-i2\pi x^2/\lambda R} dx \\ &= lE_0 \frac{e^{-i2kR}}{R^2} \left[\frac{C(u)}{u} - j \frac{S(u)}{u} \right] \end{aligned} \quad (4-13)$$

where $u = 2L/(\lambda R)^{1/2}$, and $C(u)$ and $S(u)$ are the Fresnel integrals

$$C(u) = \int_0^u \cos(\pi z^2/2) dz,$$

$$S(u) = \int_0^u \sin(\pi z^2/2) dz.$$

Hence the effective radar length is

$$l' = l \left[\frac{C(u)}{u} - j \frac{S(u)}{u} \right]. \quad (4-14)$$

From Equations 4-2 and 4-14 the effective radar area σ' may be derived as

$$\sigma' = \sigma \left[\frac{C^2(u) + S^2(u)}{u^2} \right]. \quad (4-15)$$

Thus the radar length (and radar area σ) becomes a function of range, especially for targets of great width at short range, the measurements being

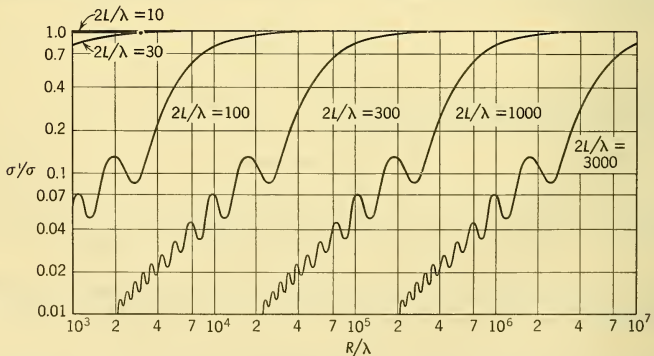


FIG. 4-2 Radar Cross Section as a Function of Range.

in terms of the wavelength. Fig. 4-2 shows a plot of σ'/σ versus R/λ for various values of $2L/\lambda$.

4-3 EFFECT OF POLARIZATION ON REFLECTION

Although the majority of radars utilize linear polarization, for certain purposes other polarizations are found to be advantageous. The use of circular polarization, for example, reduces rain clutter considerably. Since any state of polarization may be described in terms of two orthogonal polarizations (for example, horizontal and vertical, or right-hand and left-hand circular), we may denote an arbitrarily polarized incident wave by the matrix

$$E_i = \begin{pmatrix} E_1 \\ E_2 \end{pmatrix} \quad (4-16)$$

in which the orthogonal components E_1 , E_2 are complex quantities, or *phasors*.

The radar area of a target depends on the polarization of the incident wave. A long thin (in comparison to λ) wire is a good example, since its reflection is very small when the incident field is linearly polarized at right angles to the wire axis, and maximum when parallel to the axis. It is evident, therefore, that the radar area and radar length are dependent on the polarization of the incident field.

For targets of complex shape, the total field strength incident at a given point of the target is the resultant of the primary field from the radar and the reradiated fields from other parts of the target. Especially in the case of targets of large size which are in part inclined to the wave front, some of the latter fields have a component of polarization orthogonal to that of the primary field. For targets of symmetrical shape (as viewed from the radar) this *cross-polarized* component balances out in back-scattering, but otherwise it usually does not. Hence, in general, the back-scattered field has a different polarization from the incident field. The coupling between the incident and scattered polarizations depends on the incident polarization itself. As a result, the radar length is a tensor quantity, which may be written in matrix form as

$$l = \begin{pmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{pmatrix} \quad (4-17)$$

in which each of the components l_{11} , etc. is a phasor. For example, if the 1-polarization is horizontal and 2-polarization is vertical, l_{11} represents the radar length of the horizontally polarized echo from a horizontally polarized radar, l_{12} is the radar length of the horizontally polarized echo from a vertically polarized radar, etc. The reflected field-at-unit-distance is then given by

$$E_{0r} = \begin{pmatrix} E_{0r1} \\ E_{0r2} \end{pmatrix} = lE_i = \begin{pmatrix} l_{11} E_1 + l_{12} E_2 \\ l_{21} E_1 + l_{22} E_2 \end{pmatrix}. \quad (4-18)$$

By the reciprocity theorem, $l_{21} = l_{12}$.

An interesting theorem follows from Equation 4-18: *For any given target and aspect, there is a polarization of incident field which gives maximum echo, and another which gives zero echo.* This can be seen readily as follows. By adjustments of the radar antenna system, the ratio E_2/E_1 may be adjusted (in magnitude and phase) until the received polarization is orthogonal to that of the receiving system, so no signal will then be received from the target.¹ Similarly a polarization may be chosen such that the polarization of the echo coincides with that of the receiver, so that a *maximum echo* will be received.

The radar area σ also may be written in the form of a matrix by replacing the quantities l_{mn} in Equation 4-18 by

$$\sigma_{mn} = 4\pi |l_{mn}|^2. \quad (4-19)$$

Then

$$\sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}. \quad (4-20)$$

However, one could not deduce the polarization theorem above from this, since the radar area is a scalar.

4-4 MODULATION OF REFLECTED SIGNAL BY TARGET MOTION

The radar area of a complex target such as an aircraft depends on its orientation, or aspect relative to the radar. An aircraft is subjected to roll, pitch, and yaw motions by atmospheric turbulence. In addition, it may have internal motions due to rotating propellers and surface vibrations. Its gross aspect will vary with time if the target aircraft is on a noncollision or maneuvering course. All of these factors will affect the instantaneous radar area, so that the radar echo will have corresponding time variations. Some of these effects will be considered in greater detail in Paragraphs 4-7 and 4-8.

Another important effect produced by target motion is the change in frequency due to the doppler effect which was discussed in Paragraph 1-5. If the radar and the target have a relative approach velocity V , and the transmitter frequency is f_0 , the echo frequency is (see Equation 1-19)

$$f = f_0(1 + 2V/C) = f_0 + 2V/\lambda_0 = f_0 + f_D. \quad (4-21)$$

¹It is possible to build a radar which transmits one polarization and receives, on two separate receivers, the transmitted polarization and its orthogonal. For such a system, the theorem applies to only one received polarization at a time.

In ordinary (non-doppler) radar, this shift in frequency due to the average approach velocity of the target is not noticed in the case of a point target. For extended targets, such as rain clouds and the ground or sea, for which various portions of the target area fill all or an appreciable part of the radar beam, the approach velocity varies over the beam, so that the composite echo has a spectrum of frequencies. In a doppler radar this spectrum will be properly discernible as frequency shifts relative to the radar frequency. In a non-doppler radar, beats between the various frequencies will be produced in the final detector, so that an echo spectrum will also be obtained.

In the case of an aircraft, a turn, pitch, or yaw will also introduce doppler beats which are discernible in a non-doppler radar. For example, consider the effect of a turn, which imparts an angular velocity $\bar{\omega}$ of the target about its center of gravity. Two fixed points on the target a distance D apart will then have a relative radial velocity toward the radar of

$$\Delta V = \bar{\omega} D \cos \theta \quad (4-22)$$

as can be seen from Fig. 4-3. Hence by Equation 4-21 the difference in doppler frequency between these two reflection points is

$$\Delta F = \frac{2\Delta V}{\lambda} = \frac{2\bar{\omega} D \cos \theta}{\lambda} \quad (4-23)$$

Thus the doppler frequency will be proportional to radar frequency, to the angular velocity of the target, and to the gross aspect of the target.

These and other effects which result in fluctuations of the target echo will be discussed further in later sections.

4-5 REFLECTION OF PLANE WAVES FROM THE GROUND

The reflection of radar waves from the ground or sea surface is an important factor in a number of phenomena associated with airborne radar. Among these may be cited the lobe structure which is encountered in tracking low-altitude targets, height-finding errors for such targets, and the dependence of sea and ground clutter upon polarization and depression angle. In all these cases, an understanding of the basic phenomena can be obtained from a consideration of the reflection of plane waves from a plane homogeneous surface.

The reflection of a plane wave from flat ground depends on the frequency, polarization, and angle of incidence of the wave, and on the electrical properties of the ground (dielectric constant and conductivity). A wave of

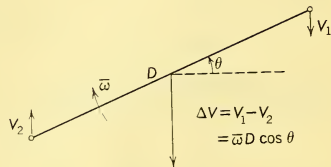


FIG. 4-3 Differential Doppler Effect Due to Turning of Target.

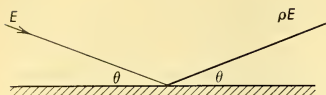


FIG. 4-4 Reflection at the Ground.

complex polarization customarily is resolved into its orthogonal linearly polarized components parallel and perpendicular to the surface, which, in the case of reflection from the ground, are horizontally and vertically polarized components, respectively.

These components can be treated separately and recombined after determining the change in amplitude and phase of each on reflection. The reflection coefficients for horizontal and vertical polarizations are given by the well-known Fresnel equations²

$$\rho_H = \frac{\sin \theta - (\epsilon - \cos^2 \theta)^{1/2}}{\sin \theta + (\epsilon - \cos^2 \theta)^{1/2}} = |\rho_H| e^{-j\phi_H} \quad (4-24)$$

$$\rho_V = \frac{\epsilon \sin \theta - (\epsilon - \cos^2 \theta)^{1/2}}{\epsilon \sin \theta + (\epsilon - \cos^2 \theta)^{1/2}} = |\rho_V| e^{-j\phi_V} \quad (4-25)$$

where θ = depression angle of the radar (see Fig. 4-4)

ϵ = complex dielectric constant of the surface.

The complex dielectric constant ϵ is given in terms of the permittivity and conductivity of the ground κ and σ by

$$\epsilon = \frac{\kappa}{\epsilon_0} - j \frac{\sigma}{\omega \epsilon_0} = \epsilon' - j\epsilon'' \quad (4-26)$$

where ϵ_0 = permittivity of free space.

Values of typical ground constants and reflection coefficients are readily available in the literature.³⁻⁵

A dependence of the reflection coefficient on frequency enters Equations 4-24 and 4-25 through the dependence of ϵ'' on frequency. In addition, however, the ground "constants" κ and σ themselves are functions of frequency, by virtue of the dispersion of water. This dispersion takes place just in the frequency region most used for airborne radar. The resulting dispersion of ground thus depends on its water content. For airborne radar this is particularly important for water surfaces. Figs. 4-5 and 4-6 show the variation of the dielectric properties of pure, fresh, and sea water with

²See J. A. Stratton, *Electromagnetic Theory*, Secs. 9.4 and 9.9, McGraw-Hill Book Co., Inc., New York, 1941.

³F. E. Terman, *Radio Engineers' Handbook*, pp. 700-709, McGraw-Hill Book Co., Inc., New York, 1943.

⁴C. R. Burrows, "Radio Propagation over Plane Earth-Field Strength Curves," *Bell System Tech. J.* **16**, 45-75 (1937).

⁵R. S. Kirby, J. C. Harman, F. M. Capps, and R. N. Jones, *Effective Radio Ground-Conductivity Measurements in the United States*, National Bureau of Standards Circular 546.

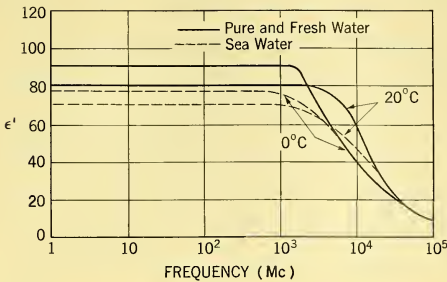


FIG. 4-5 Dielectric Properties of Pure, Fresh, and Sea Water.

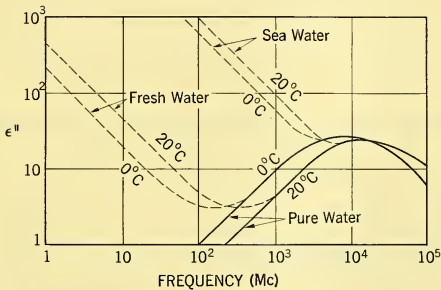


FIG. 4-6 Dielectric Properties of Pure, Fresh, and Sea Water.

frequency, taken from Saxton.⁶ The curves for temperatures of 0° and 20°C bring out a dependence on temperature as well.

Figs. 4-7 through 4-10 show the magnitude and phase angle of the reflection coefficient of sea water for a temperature of 10°C at several wavelengths. Similarly, Figs. 4-11 and 4-12 show the reflection coefficients for two different types of ground. For most airborne radar work, solid ground may be treated as a pure dielectric. These figures bring out clearly the difference between horizontal and vertical polarization. For horizontal polarization, there is only a slight variation in magnitude and phase of the reflection coefficient with depression angle. For vertical polarization, however, there is a marked variation, caused by a partial impedance match of the two mediums which occurs at the Brewster angle. The reflection coefficient reaches a minimum magnitude and has a phase angle of 90° at this angle (the Brewster angle itself depends on frequency).

⁶J. A. Saxton, "Electrical Properties of Sea Water," *Wireless Engineer* 29, 269-275 (1952).

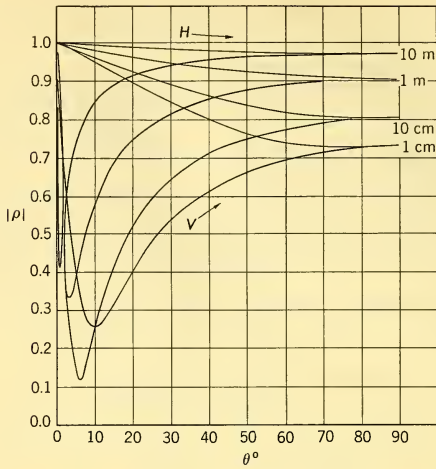


FIG. 4-7 Magnitude of Reflection Coefficient for Sea Water (Temperature = 10°C) as a Function of Depression Angle.

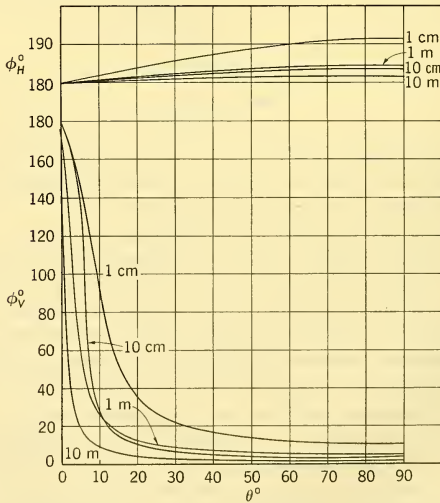


FIG. 4-8 Phase of Reflection Coefficient for Sea Water (Temperature = 10°C) as a Function of Depression Angle.

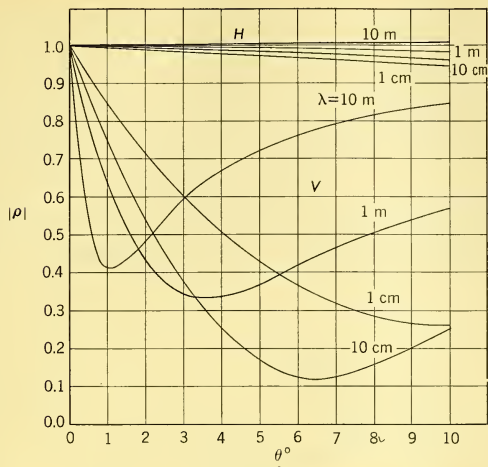


FIG. 4-9 Expanded Plot of Fig. 4-7 for Depression Angles Between 0 and 10°.

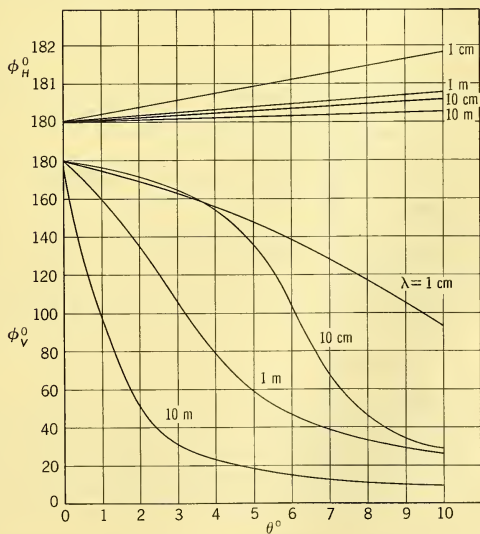


FIG. 4-10 Expanded Plot of Fig. 4-8 for Depression Angles Between 0° and 10°.

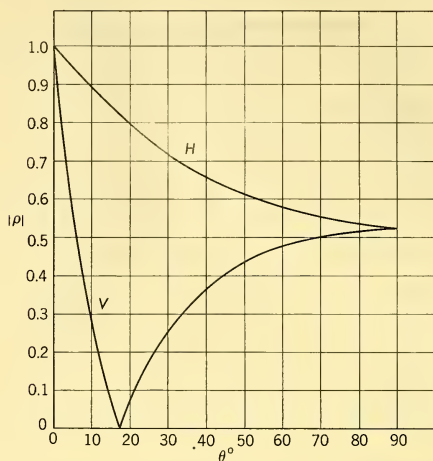


FIG. 4-11 Magnitude of Reflection Coefficient for Average Land ($\epsilon' = 10$, $\sigma = 1.6 \times 10^{-3}$ mho/m) as a Function of Depression Angle.

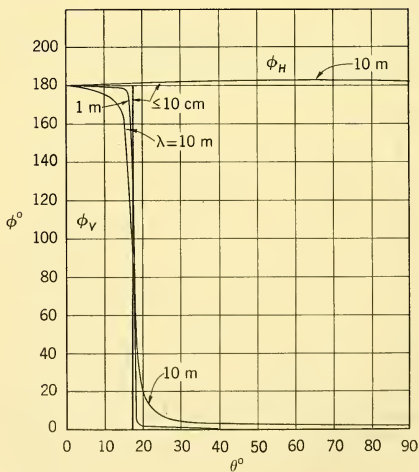


Fig. 4-12 Phase of Reflection Coefficient for Average Land ($\epsilon' = 10$, $\sigma = 1.6 \times 10^{-3}$ mho/m) as a Function of Depression Angle.

The behavior is more complicated when the ground is stratified. The cases which are important to airborne radar are that of a layer of ice on top of a water surface, and that of a layer of snow on land. Then multiple reflections can occur between the surface and subsurface boundaries, with a resulting modification of the effective reflection coefficients⁷; the effective reflection coefficient then becomes an oscillating function of the electrical thickness of the ice or snow covering.

When the radar target is at a low altitude, a variety of phenomena are generated by the interference of direct and reflected waves. Referring to Fig. 4-13, if both the direct and indirect paths are illuminated equally by the radar antenna, the resultant field at the target is

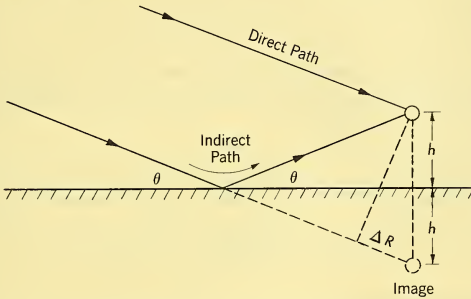


FIG. 4-13 Path Difference Between Direct and Indirect Paths.

$$E = E_d(1 + \rho e^{-iK\Delta R}) \tag{4-27}$$

where E_d is the field due to the direct wave, and $\Delta R = 2h \sin \theta$ is the path difference.

The ratio E/E_d , obtained from Equation 4-27, thus is the propagation factor F due to the presence of the ground. In Paragraph 4-2, where this factor was defined, it was pointed out that the received power from a radar target is modified by the factor $|F|^4$. Introducing the values for K and ΔR in Equation 4-27, we obtain

$$F = E/E_d = 1 + \rho e^{-i4\pi h \sin \theta / \lambda} \tag{4-28}$$

The most significant and striking phenomena resulting from the interference of direct and reflected rays are the *lobe structure* and the *polarization dependence* below the first lobe. The formation of a set of lobes is easily seen from Equation 4-28. With fixed θ and continuous increase of h , the resultant field will pass through alternate maximums and minimums when the phase

⁷J. A. Saxton, "Reflection Coefficient of Snow and Ice at V. H. F.," *Wireless Engineer* 27, 17-25 (1950).

lag of the reflected wave is an even or odd multiple, respectively, of 180° . The height interval between an adjacent maximum and minimum is

$$\Delta h = \lambda/4 \sin \theta. \quad (4-29)$$

This succession of maximums and minimums of the resultant field gives rise to the lobe structure in the vertical coverage of the radar, which is especially important for search radars. The location of a given maximum or minimum is different for vertical and horizontal polarization because of the phase of the reflection coefficient ρ . For airborne radar work with pencil beam antennas, the lobe structure usually is of importance only for targets at small depression angles, since otherwise the narrow beamwidth of the antenna would not illuminate the indirect path strongly. The lobe structure is pronounced only if the value of Δh is large compared with the vertical extent of the target. If the target covers more than one lobe, it effectively averages out the field variation over the lobe. This actually produces a net increase of gain over the free-space field acting alone, which is due to the field reflected from the surface.

A similar oscillation in the propagation factor is observed with fixed radar and target altitudes and a continuously varying range as the target passes through the lobe pattern. In this case the angle θ can be expressed as

$$\sin \theta = \frac{H + h}{\sqrt{R^2 + 4hH}} \approx \frac{H + h}{R} \quad (4-30)$$

where H = radar altitude

h = target altitude

R = target range.

Neglecting the change in the phase angle of the reflection coefficient ρ , the range interval between an adjacent maximum and minimum is

$$\Delta R \approx \frac{\lambda R^2}{4h(H + h)} \quad (4-31)$$

where R is the mean range. Thus for a target flying at a constant height, the lobes become packed more densely as the range is decreased. The oscillations of received power caused by the lobes are superimposed on a free-space variation which is proportional to the inverse fourth power of range as indicated in Equation 4-1.

The situation is somewhat different when the target lies below the first lobe. In this case, the angle θ will be small and an expansion of F in powers of $\sin \theta$ can be used. To obtain this expansion, we note first that ρ_H and ρ_V , which are given in Equations 4-24 and 4-25, can be approximated as

$$\rho_H \approx -1 + 2(\epsilon - 1)^{-1/2} \sin \theta \quad (4-32)$$

$$\rho_V \approx -1 + 2\epsilon(\epsilon - 1)^{-1/2} \sin \theta. \quad (4-33)$$

Similarly, the exponential term in Equation 4-27 may be approximated as

$$e^{-j2Kh \sin \theta} \approx 1 - j2Kh \sin \theta. \quad (4-34)$$

Substituting these approximations into Equation 4-28, and retaining only the first-order terms in $\sin \theta$, we obtain

$$F_H = 2[(\epsilon - 1)^{-1/2} + jKh] \sin \theta \quad (4-35)$$

$$F_V = 2[\epsilon(\epsilon - 1)^{-1/2} + jKh] \sin \theta. \quad (4-36)$$

Thus, for sufficiently small θ , both F_H and F_V are proportional to $\sin \theta$. But $\sin \theta$, as shown by Equation 4-30, is inversely proportional to range. Hence below the first lobe it follows that F is also inversely proportional to range:

$$F \propto R^{-1}. \quad (4-37)$$

Therefore the received echo power, which is given by Equation 4-1, will be inversely proportional to the eighth power of the range in the region below the first lobe:

$$P_R \propto R^{-8}. \quad (4-38)$$

This is in contrast to the inverse fourth power of the range which holds for free space. The range at which the transition occurs from a fourth-power law to an eighth-power law for a target which spans more than one lobe will be discussed in Paragraph 4-10.

Equations 4-35 and 4-36 show how the resultant (one-way) field varies with height below the first lobe. Very close to the surface, where the term is small in magnitude compared with the other term in the brackets, F_H and F_V become

$$F_H \approx \frac{2}{(\epsilon - 1)^{1/2}} \sin \theta \quad (4-39)$$

$$F_V \approx \frac{2\epsilon}{(\epsilon - 1)^{1/2}} \sin \theta. \quad (4-40)$$

Hence the ratio of the fields at the target with vertical and horizontal polarization will be

$$F_V/F_H = \epsilon. \quad (4-41)$$

If the radar area of the target is the same for these two polarizations, then the ratio of the received echo powers will approach $|\epsilon|^4$. This difference is important in the case of sea clutter.

As the height is increased, the term $j2Kh$ eventually will become large relative to the other term in the brackets in Equations 4-35 and 4-36. Then the field at the target will be approximately proportional to height and will be almost the same for either polarization.

The first-order expansions of F in powers of $\sin \theta$ are limited in their ranges of validity. This can be seen, for example, in Fig. 4-7. For vertical polarization, the range of validity is limited to angles smaller than the Brewster angle, while for horizontal polarization, the angular range is much greater. For airborne radar frequencies, the range is about $\theta < 30^\circ$ for horizontal polarization, and $\theta < 4^\circ$ for vertical polarization.

The results deduced above are based on the properties of plane waves. In the case of the spherical waves radiated by an antenna, there is a surface wave which should be added to the direct and reflected waves. For airborne radar frequencies, however, this generally is unimportant.

4-6 EFFECT OF EARTH'S CURVATURE

The effect of the earth's curvature is twofold. First, it alters the geometry so that the path difference between the direct and reflected waves is decreased, and second, it decreases the amplitude of the reflected wave.

The change from the plane to the spherical geometry is equivalent to a reduction in the heights of radar and target, as illustrated in Fig. 4-14.

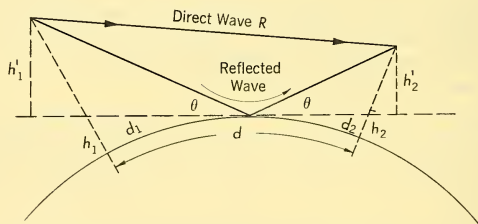


FIG. 4-14 Curved Earth Geometry.

The second effect of the earth's curvature is to decrease the amplitude of the reflected wave, because the incident waves within a small range of vertical angles are spread out, or *diverged*, into a larger range of vertical angles on reflection from the convex surface of the earth.

For all distances encountered in airborne radar work, the reduced heights h'_1 , h'_2 may be calculated from

$$\begin{cases} h'_1 = h_1 - \Delta h_1 = h_1 - d_1^2/2a & (4-42a) \\ h'_2 = h_2 - \Delta h_2 = h_2 - d_2^2/2a & (4-42b) \end{cases}$$

where a is the earth's radius, and $d_1(d_2)$ is the distance from the reflection point to the transmitting (receiving) point. As will be shown in Paragraph 4-18, the effect of average or "standard" atmospheric refraction may be

allowed for by increasing the earth's radius by a factor $4/3$ to an effective earth's radius

$$a_e = 4/3 a. \quad (4-43)$$

With this factor, and expressing heights in feet and distances in statute miles, the height reductions due to curvature take the simple form

$$\Delta h_{1,2} = d_{1,2}^2/2. \quad (4-44)$$

For a given value of range R , which is practically the same as the total distance $d = d_1 + d_2$ measured along the earth's surface, the determination of d_1 (or d_2) leads to the cubic equation

$$2d_1^3 - 3dd_1^2 - [2a_e(h_1 + h_2) - d^2]d_1 + 2a_e h_1 d = 0. \quad (4-45)$$

Once this is solved for d_1 , h_1' and h_2' may be calculated from Equation 4-42 and the remainder of the geometry handled like a plane-earth problem. Since the solution of the cubic is laborious, it is usually simpler to employ a graphical solution by plotting h_1'/d_1 and h_2'/d_2 versus d_1 . The proper value of d_1 occurs where these two quantities are equal, since this gives equal values of θ before and after reflection.

As mentioned above, reflection at a spherical surface reduces the reflection coefficient from the plane earth value ρ to

$$\rho' = \rho D \quad (4-46)$$

where D is the *divergence factor*. This is given by

$$\begin{aligned} D &= \left(1 + \frac{2d_1 d_2}{a_e d \sin \theta}\right)^{-1/2} \\ &= \left(1 + \frac{2d_1^2 d_2^2}{a_e d h_1'}\right)^{-1/2}. \end{aligned} \quad (4-47)$$

For very small values of θ the divergence factor causes reduction of the effective reflection coefficient ρ' given by Equation 4-46 to a small value. In fact, at the horizon ($\theta = 0$) $D = 0$, so that there $\rho' = 0$. However, the representation of the propagation process in terms of only a direct and a reflected ray breaks down as the horizon is approached. Norton⁸ gives as the limit to which Equation 4-46 is restricted:

$$\frac{h_1'}{d_1} > \left(\frac{\lambda}{2\pi a_e}\right)^{1/2}. \quad (4-48)$$

Practically all airborne radar ranges will be within this limit as long as atmospheric refraction does not depart greatly from the standard condition.

⁸K. A. Norton, "The Calculation of Ground-Wave Field Intensity over a Finitely Conducting Spherical Earth," *Proc. IRE* **29**, 623-639 (1941).

4-7 RADAR CROSS SECTIONS OF AIRCRAFT

Because all aircraft have dimensions large in terms of the wavelengths used in airborne radar, the radar area of an aircraft target is very sensitive to its instantaneous aspect. Because of air turbulence, the aspect is subject to statistical variations of roll, pitch, and yaw. Consequently the radar area is a statistically fluctuating quantity and it is not possible to give a single number for the radar area of such a target. The quantities of chief interest are the probability distribution of amplitudes, the aspect and frequency dependencies, and the time characteristics, or spectra, of the fluctuations.

The amplitude distributions and aspect and frequency dependencies of certain aircraft will be presented in this paragraph, while the fluctuations and their effect on tracking systems will be discussed in Paragraph 4-8. A summation of the echo characteristics and their association with the physical structure and dynamic behavior of the aircraft will then be presented in Paragraph 4-9. This should make it possible to predict, with useful accuracy, the main features of the radar properties of a new or unmeasured target aircraft.

An appreciation of the complicated nature of the radar area of an aircraft and its association with the physical structure of the aircraft can be obtained from some of the results of basic investigations into the properties of radar echoes from aircraft carried out by the Naval Research Laboratory, and recently made available.⁹⁻¹² Pulse-to-pulse measurements were made of both fighter and bomber categories, with propeller-driven and jet-propelled models in each category. The measurements were made on three frequencies, 1250, 2810, and 9380 Mc/sec, with the radars searchlighted on the target by an optical tracker, and pulsed simultaneously. No antenna scanning was used, so that the observed fluctuations were all attributable to the target. The data were analyzed to determine amplitude distributions, median radar area versus aspect, and frequency spectrum of the amplitude fluctuations. The particular series of measurements to be discussed was made at elevation angles less than 15°. These measurements will be discussed in some detail, since comparable data have not been published before. Many of the characteristics observed can be explained in terms of physical processes, so that from these it should be possible to predict the principal characteristics to be expected in other situations.

⁹F. C. MacDonald, *Quantitative Measurements of Radar Echoes from Aircraft III: B-16 Amplitude Distributions and Aspect Dependence*, NRL Report C-3460-94A/51, 19 June 1953.

¹⁰W. S. Ament, M. Katzin, F. C. MacDonald, H. J. Passerini, P. L. Watkins, *Quantitative Measurements of Radar Echoes from Aircraft V: Correction of X-Band Values*, NRL Report C-3460-132A/52, 24 Oct. 1952.

¹¹W. S. Ament, F. C. MacDonald, H. J. Passerini, *Quantitative Measurements of Radar Echoes from Aircraft VIII: B-45*, NRL Memorandum Report No. 116, 28 Jan. 1953.

¹²W. S. Ament, F. C. MacDonald, H. J. Passerini, *Quantitative Measurements of Radar Echoes from Aircraft IX: F-51*, NRL Memorandum Report No. 127, 4 March 1953.

Fig. 4-15 shows the cumulative amplitude distribution of a 2-second sample of echoes from the B-36, plotted on so-called Rayleigh coordinates.

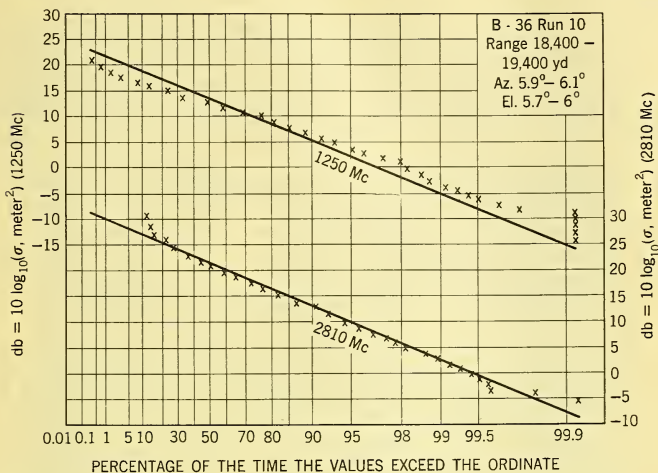


FIG. 4-15 Cumulative Amplitude Distribution of B-36 Echo, Approach Aspect.

The straight lines through the points represent the Rayleigh distribution.¹³ Even with such a short sample (in this case, of only 240 pulses), the fit to a Rayleigh distribution is quite good. From the data obtained, it was concluded that for a 2-second sample the echo amplitude (and thus the radar area) is Rayleigh distributed for most aspects, except at broadside aspect. At broadside the amplitudes were compressed into a rather narrow range.

The Rayleigh distribution signifies that the target consists of a large number of elements whose relative phases are independent and vary randomly during the time of the observation. The number of independent elements which constitute a "large" number, however, need be only about four or five if their amplitudes are comparable. Thus the conclusion to be drawn from the B-36 amplitudes distribution is that, except at broadside, the target consists of just such a "large" number of independent scatterers, and that in 2 seconds their relative phases pass through substantially all possible combinations.¹⁴ At broadside, however, the echo from the flat

¹³J. L. Lawson and G. E. Uhlenbeck, "Threshold Signals" *Mass. Inst. Technol., Laboratory Series 24*, 53, McGraw-Hill Book Co., Inc., New York, 1950.

¹⁴Practically, "all possible combinations" probably is satisfied if the phases vary over one or two times 360°.

fuselage is so large relative to the echoes from other parts of the aircraft that it predominates over them and a relatively small amplitude variation results.

Fig. 4-16 shows a 5-second sample for the B-45 twin jet bomber, taken for an approach run in which the aspect varied by $2\frac{1}{2}^\circ$. Here the approach to the Rayleigh distribution is poor, the range of amplitude variation being much more compressed. However, a 5-second sample at another aspect, in which the aspect angle varied $4\frac{1}{2}^\circ$ (Fig. 4-17) shows a much closer

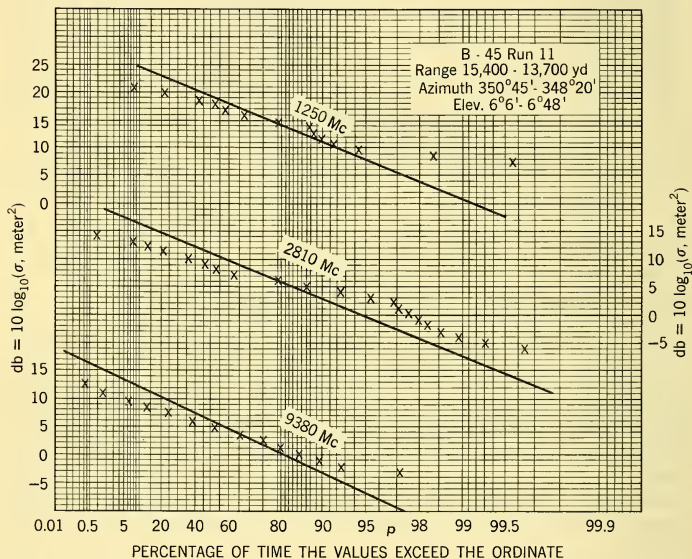


FIG. 4-16 Cumulative Amplitude Distribution of B-45 Echo, Approach Aspect; Small Range of Aspect Angle.

approach to the Rayleigh distribution. From an examination of data taken over a wide range of aspects, it was concluded that samples in which the azimuth of the B-45 varied by more than 4° gave a satisfactory fit to the Rayleigh distribution.

Fig. 4-18 shows a set of distributions for the F-51 single-engine propeller-driven fighter. Although the lower amplitudes follow the Rayleigh distribution quite well (on 9380 Mc/sec the lower levels were lost in the noise at the range of the measurements plotted in this figure), there is a pronounced upswing at high levels above the values expected from the extension of the

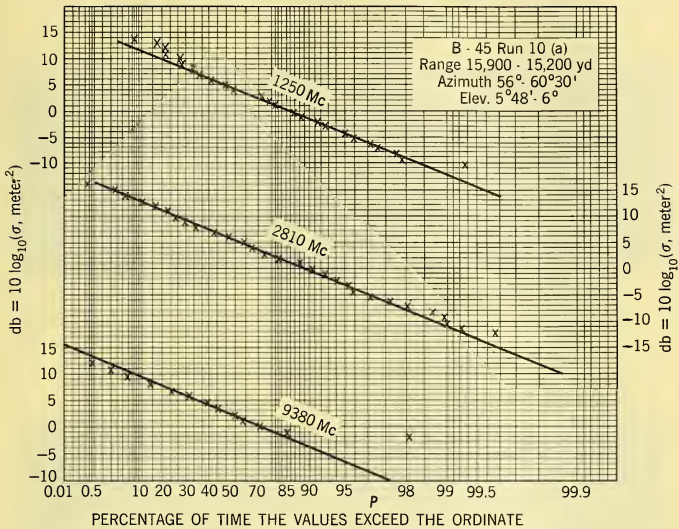


FIG. 4-17 Cumulative Amplitude Distribution of B-45 Echo; Larger Range of Aspect Angle.

Rayleigh line fitted to the lower levels. This is attributed to reflections from the propeller, which, for a rather large range of angles, are stronger than from the remainder of the aircraft. This is shown by the original pulse-to-pulse photographs shown in Fig. 4-19. Here every fifth pulse (repetition rate 120 cps) is much larger (on all three frequencies) than the intervening ones. The dominance of the propeller echo was found to be especially marked at oblique aspects of the aircraft between head-on and broadside, corresponding to the region where a portion of the blade is nearly normal to the line from the radar.

To depict the gross aspect variation of σ , the median values over roughly 5° of azimuth were plotted against azimuth angle. Figs. 4-20 to 4-22 show the results for the B-36, B-45, and F-51, respectively. In averaging over an angular range of this amount, sharp peaks of the aspect dependence are largely smoothed out. In all cases, however, a prominent and rather broad maximum occurs in the neighborhood of the broadside aspect. This is especially true in the case of the B-36 (which has a rather flat fuselage) as shown by the 9380-Mc plot in Fig. 4-20 (broadside data for 1250 and 2810 Mc were saturated and so are absent from this figure). The F-51 has its broadside maximum at an azimuth of about 98° , probably owing to the tapered tail section of the fuselage.

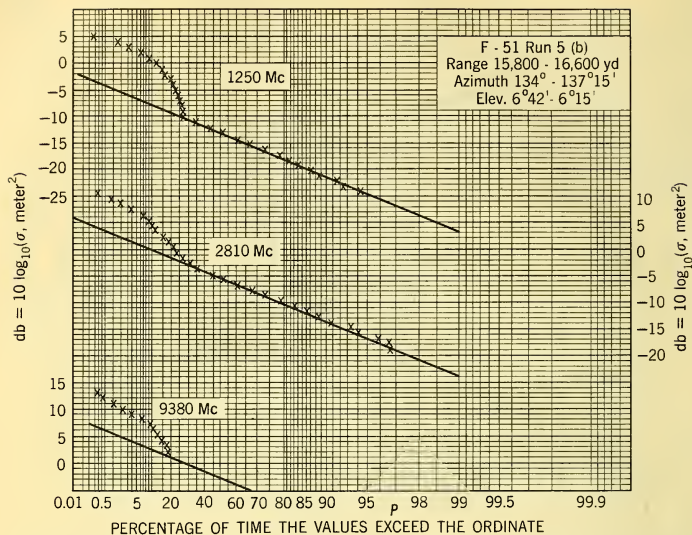


FIG. 4-18 Amplitude Distribution of F-51 Echo, Showing Effect of Propeller Reflection.

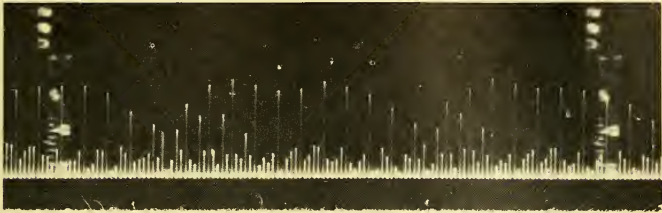
The data in Figs. 4-20 to 4-22 also give the frequency dependencies of the radar area. A single number for the average radar area was obtained for each frequency by averaging all the values of σ (in square meters) plotted in each figure. These averages are controlled by the large peak in the neighborhood of the broadside aspect. Similar numbers were obtained for all aspects measured outside of the broadside region. These latter numbers give a measure of the frequency dependencies of the aircraft for most tactical applications. The results are shown in Table 4-1. The B-36 and B-45 averages are roughly independent of frequency, but the F-51 average σ increases approximately proportional to frequency.

TABLE 4-1 COMPARISON OF B-36, B-45, AND F-51 AVERAGE RADAR AREAS

Frequency, Mc	Average σ (in $db > 1m^2$) Excluding Broadside Region			Average σ (in $db > 1m^2$) Including Broadside Region		
	B-36	B-45	F-51	B-36	B-45	F-51
1250	13.2	10.7	-0.5	^a	14.9	-0.2
2810	16.4	10.6	3.6	^a	14.5	5.6
9380	10.	10.2	6.6	^a	17.3	9.2

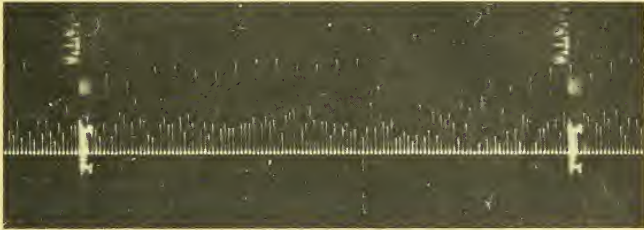
^aBroadside region saturated.

1250 MC/S



← 1-SEC →

2810 MC/S



9380 MC/S



FIG. 4-19 Pulse-to-Pulse Records of F-51 Echo, Showing Strong Propeller Echo Every Fifth Pulse.

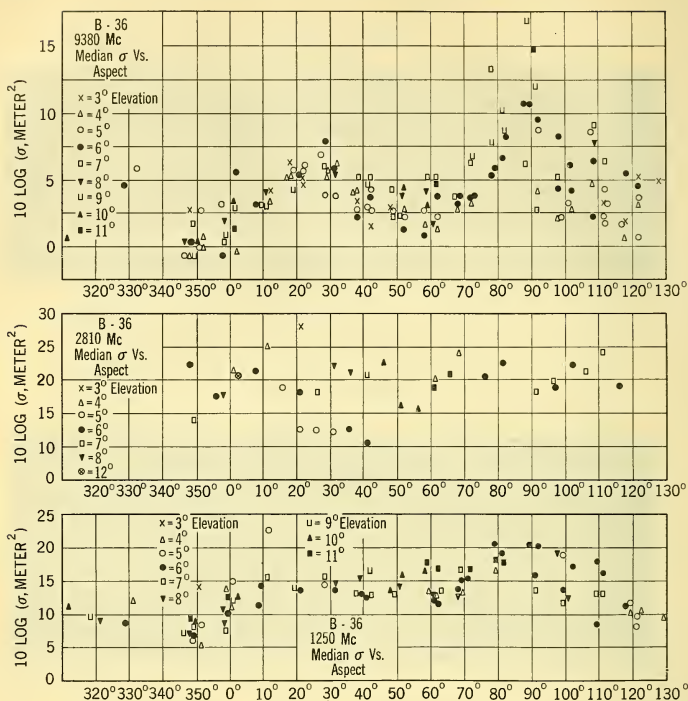


FIG. 4-20 Plot of Median Echo of B-36 Averaged over 5° of Azimuth.

4-8 AMPLITUDE, ANGLE, AND RANGE NOISE¹⁵

In a tracking radar, rapid variations in target aspect can affect the smoothness of tracking and hence its accuracy. The variations in the target

¹⁵Most of the material in this Paragraph has been derived from the following NRL reports, and from references in footnotes 11-14, which can be consulted for further details:

J. W. Meade, A. E. Hastings, and H. L. Gerwin, *Noise in Tracking Radars*, NRL Report 3759, Nov. 15, 1950.

A. E. Hastings, J. E. Meade, and H. L. Gerwin, *Noise in Tracking Radars, Part II: Distribution Functions and Further Power Spectra*, Jan. 16, 1952.

D. D. Howard and B. L. Lewis, *Tracking Radar External Range Noise Measurements and Analysis*, NRL Report 4602, Aug. 31, 1955.

A. J. Stecca, N. V. O'Neal, and J. J. Freeman, *A Target Simulator*, NRL Report 4694, Feb. 9, 1956.

A. J. Stecca and N. V. O'Neal, *Target Noise Simulator — Closed-Loop Tracking*, NRL Report 4770, July 27, 1956.

B. L. Lewis, A. J. Stecca, and D. D. Howard, *The Effect of an Automatic Gain Control on the Tracking Performance of a Monopulse Radar*, NRL Report 4796, July 31, 1956.

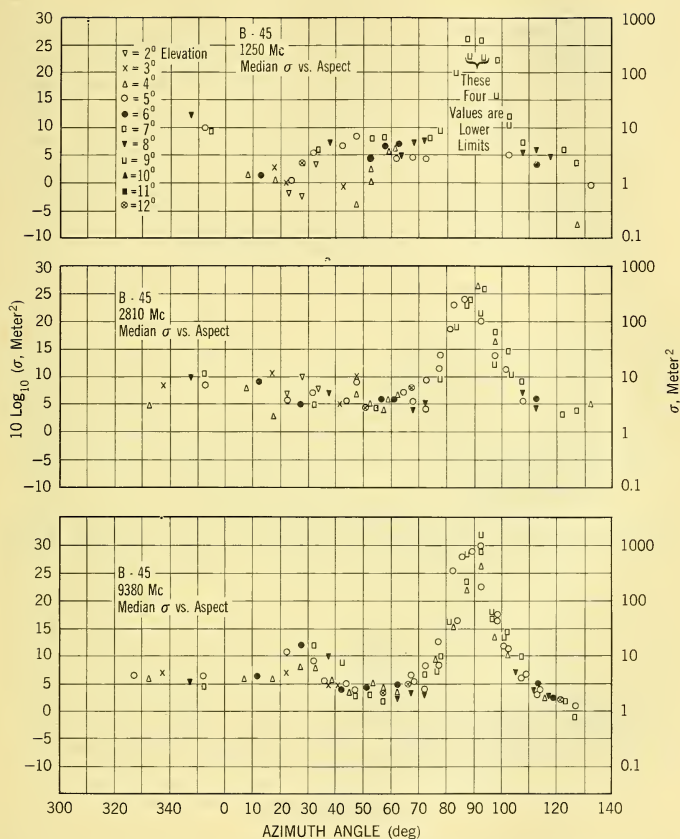


FIG. 4-21 Plot of Median Echo of B-45 Averaged over 5° of Azimuth.

characteristics are referred to as *target noise* (or *scintillation*) and have an effect similar to noise originating in the tracking system. Target noise is subdivided into *amplitude noise*, *angle noise* (or *glint*), and *range noise*.

Amplitude Noise. In a sequential lobing radar, the antenna beam is scanned over a small range of angles, and comparison is made of the signals received with the beam swung to opposite sides of the boresight axis. The difference between these signals is used to drive the antenna toward a

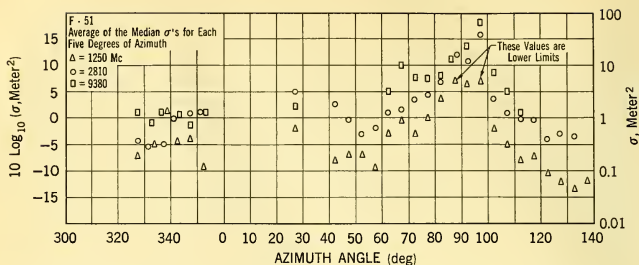


FIG. 4-22 Plot of Median Echo of F-51 Averaged over 5° of Azimuth.

position where the signals are equal. Since the two samples are not received at the same instant, any change in signal amplitude during the scanning cycle, caused by target fluctuation, will lead to an angle error indication even if the antenna is pointed correctly at the target. In order to make an optimum choice of the parameters of a sequential lobing system, therefore, it is necessary to have information on the amplitude fluctuation characteristics of target aircraft.

Some of the causes of amplitude fluctuations already have been mentioned in Paragraph 4-7, e.g. propeller modulation. A clear example of this is shown in Figure 4-23, which is the spectrum of amplitude fluctuations in

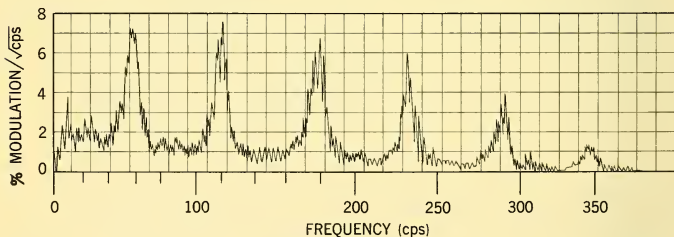
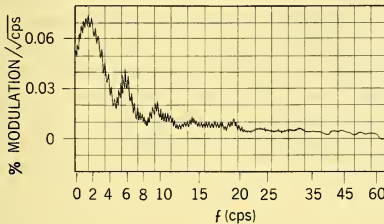


FIG. 4-23 Spectrum of Amplitude Noise of SNB (Two-Engine Transport) Aircraft in an Approach Run, Showing Spectral Lines Due to Propeller Modulation (≈ 9400 Mc).

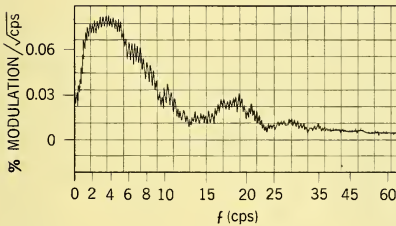
an approach run of an SNB (two-engine transport) taken at a radar frequency around 9400 Mc. The repetition frequency of the radar was 1000 cycles, so spectral information up to 500 cycles is derivable. Here peaks occur at multiples of about 58 cps. The remainder of the spectrum consists of a continuous band¹⁶ whose amplitude decreases with increasing

¹⁶The small ripples or scintillations are due to incomplete smoothing in the spectrum analyzer.

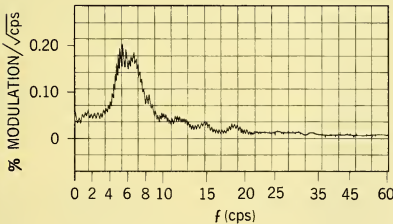
frequency. The propeller modulation occurs at the blade frequency (engine rps \times number of propeller blades), so the fundamental modulation frequency should be the same for all radar frequencies. This was seen to be the case for the F-51, as illustrated in Fig. 4-19.¹⁷



B - 45 RUN 10 1250 Mc/sec
 RANGE 13,600 YD
 AZIMUTH $96^{\circ}20'$ - $98^{\circ}42'$
 ELEVATION $6^{\circ}56'$ - $6^{\circ}55'$



B - 45 RUN 10 2810 Mc/sec
 RANGE 13,300 YD
 AZIMUTH 91° - 94°
 ELEVATION $6^{\circ}55'$



B - 45 RUN 10 9380 Mc/sec
 RANGE 13,600 YD
 AZIMUTH $96^{\circ}20'$ - $98^{\circ}42'$
 ELEVATION $6^{\circ}56'$ - $6^{\circ}55'$

FIG. 4-24 Spectrum of Amplitude Noise of B-45.

For a jet aircraft, the mechanical vibrations of the salient reflection surfaces would be expected to be of high frequency and noiselike in nature

¹⁷Because the radar samples the instantaneous radar area of the target at discrete intervals r times per second (r = the pulse repetition frequency), a frequency of s cps in the spectrum could result from beats between the actual frequency and the repetition frequency. Hence the actual frequency of the target radar area spectrum may be any value given by $nr \pm s$, where n is an integer (including 0). For the F-51, the observed propeller modulation frequency was found to correspond to $n = 7$ or 8.

so that a line frequency spectrum similar to the propeller modulation spectrum would not be expected. The amplitude modulation spectrum observed in such cases would be due chiefly to beats between the doppler frequencies of the salient reflection centers, as described in Paragraph 4-4. The frequency of this type of modulation, therefore, would be proportional to the radar frequency. Whether a continuous spectrum is obtained, or one with spectral lines superposed on a continuous spectrum, depends on the nature of the perturbations of the aircraft for straight-line flight and on the duration of the observation (that is, on the length of sample).

Figure 4-24 shows a spectrum¹⁸ of the amplitude noise of the B-45 for which the observation time was 5 seconds. The voltage-time plots from which the spectrum was prepared are shown in Fig. 4-25. The 9380-Mc

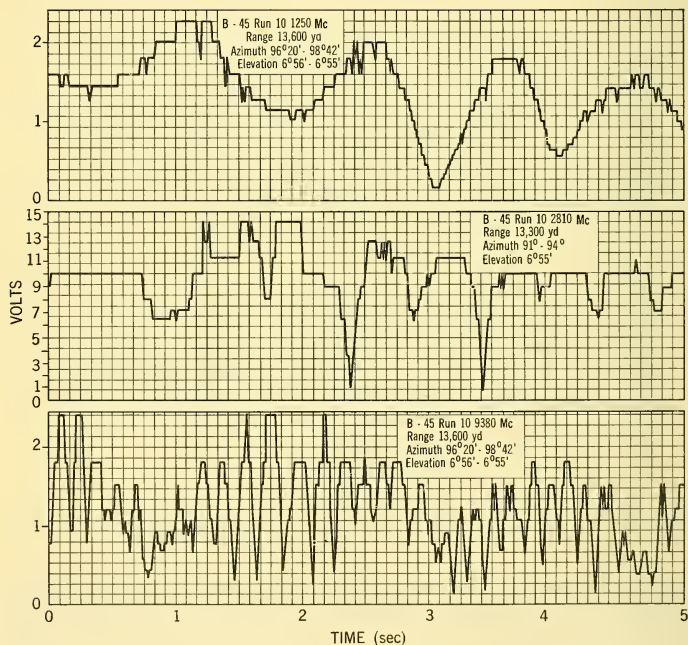


FIG. 4-25 Voltage-Time Plots of B-45 Echo, Showing Low-Frequency Modulation.

¹⁸The resolving power of the spectrum analyzer used was about 1.5 cps. Also, its response begins to drop off below about 6 cps.

spectrum in Fig. 4-24 shows two peaks, not quite resolved completely, at 5 and 6.6 cps. In Fig. 4-25 the presence of a fundamental period in this region can be seen clearly; and similar fundamental periods, but of progressively lower frequency, can be seen in the 2810- and 1250-Mc plots. From an examination of the drawings of the B-45, it was concluded that the observed spectrum could be explained as the doppler beats between reflections from the engine nacelle, the wing tank, and a portion of the fuselage near the tail which was broadside at this aspect, all caused by a yaw rate of $0.14^\circ/\text{sec}$.

For a longer sample (i.e., longer period of observation) during which the yaw rate varied, the discrete frequencies varied with time so that the spectrum over such a time of observation was smeared out into a more or less continuous band. Whether one should be concerned about a continuous band or discrete frequencies in the design of a tracking radar depends, therefore, on the time constant of the system — in other words, on the passband of the servo loop. This problem will be discussed in more detail in Chapter 9.

Angle Noise. In a simultaneous lobing system (to be discussed in more detail in Paragraph 6-2) the signals which are compared to obtain angle information arrive simultaneously; thus amplitude fluctuations of the target echo do not generate angle error signals. If the angle tracking servo loop of a simultaneous lobing system is opened and the target is tracked optically, it is found that error signals still occur. These must be caused, therefore, by wandering of the effective center of reflection of the target.

The principle involved in the generation of angle noise may be explained in terms of a target which consists of two point reflection centers, whose

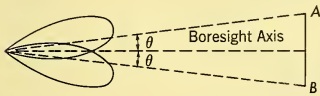


FIG. 4-26 Physical Arrangement for Illustrating the Origin of Angle Noise.

relative amplitude and phase vary as the target aspect changes. Fig. 4-26 illustrates this model. Consider the following type of tracking system.

A dual-feed antenna produces lobes on each side of the boresight axis, the gains being equal along that axis.

The voltage from each lobe is passed through an amplifier and square-law detector, and their difference is used to derive angular information. The sum of the detector outputs is used for the AGC voltage of the receiver, so that the angular deviation of the arriving signal is determined by the difference divided by the sum of the detector outputs.

Let the boresight axis be directed at the center of the target (midway between A and B) and let the angle of A (and of B) to the boresight axis be θ . For small θ , the slope of the antenna lobes can be considered constant, and will be denoted by g . Denoting the received RF voltages due to A alone

and B alone when $\theta = 0$ by E_A and E_B , and their phase difference by ϕ , the total RF voltage received by the upper lobe is

$$E_U = E_A(1 + g\theta) + E_B(1 - g\theta)e^{i\phi} \quad (4-49)$$

and by the lower lobe

$$E_L = E_A(1 - g\theta) + E_B(1 + g\theta)e^{i\phi}. \quad (4-50)$$

The difference channel voltage is then

$$\begin{aligned} E_D &= k(|E_U|^2 - |E_L|^2) \\ &= 4kg\theta(E_A^2 - E_B^2) \end{aligned} \quad (4-51)$$

where k = amplifier gain, and the sum channel voltage is

$$\begin{aligned} E_S &= k(|E_U|^2 + |E_L|^2) \\ &= 2k[(E_A^2 + E_B^2 + 2E_A E_B \cos \phi) + (g\theta)^2(E_A^2 + E_B^2 - 2E_A E_B \cos \phi)] \\ &\approx 2k(E_A^2 + E_B^2 + 2E_A E_B \cos \phi). \end{aligned} \quad (4-52)$$

The error voltage is

$$\frac{E_D}{E_S} \approx 2g\theta \frac{E_A^2 - E_B^2}{E_A^2 + E_B^2 + 2E_A E_B \cos \phi}. \quad (4-53)$$

In the presence of only a single target, say at A , the error voltage would be

$$\left(\frac{E_D}{E_S}\right)' = 2g\theta. \quad (4-54)$$

Comparing this with Equation 4-53 we see that the *apparent reflection* center of the dual-reflector target lies at an angle θ' to the boresight axis

$$\theta' = \theta \left[\frac{1 - (E_B/E_A)^2}{1 + (E_B/E_A)^2 + 2(E_B/E_A) \cos \phi} \right]. \quad (4-55)$$

Thus, θ' depends on the ratio

$$r = \frac{1 - (E_B/E_A)^2}{1 + (E_B/E_A)^2 + 2(E_B/E_A) \cos \phi} \quad (4-56)$$

and therefore on the relative amplitude and phase of the two reflections. This ratio can be less than, equal to, or greater than unity in absolute value, and may be positive or negative. In other words, *the apparent reflection center can lie anywhere within the target, or even completely outside it.*

From Equation 4-56, $r = 0$ when $E_A = E_B$, so that equal reflectors have an apparent center midway between them, regardless of their relative phase. For values of E_B/E_A other than unity, the value of r depends on the relative phase, ϕ . The apparent reflection center lies outside the target when $|r| > 1$, which requires that

$$-\cos \phi = \cos(\pi - \phi) > \frac{E_B}{E_A} \quad \text{or} \quad \frac{E_A}{E_B}. \quad (4-57)$$

Thus this phenomenon occurs in the region of destructive interference between the two reflections, as may be seen from the circle diagram in Fig. 4-27, which is drawn for $E_B > E_A$.

In general, a target may have a number of reflection centers whose relative amplitudes and phases vary with the instantaneous target aspect. Delano¹⁹ has developed the theory for a target composed of an infinite number of statistically independent point sources and has determined the statistical properties of the

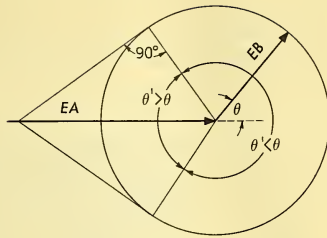


FIG. 4-27 Vector Diagram for Two-Target Example.

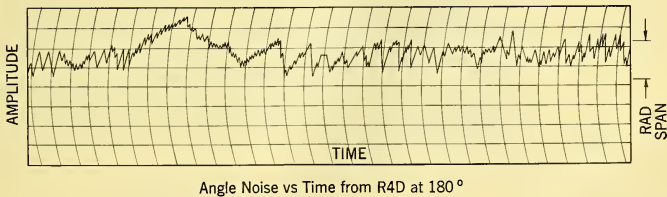
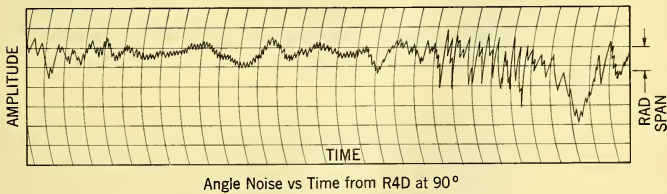
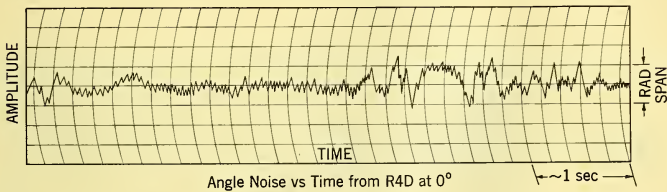


FIG. 4-28 Angle Noise Samples for R4D.

¹⁹R. H. Delano, "A Theory of Target Glint or Angular Scintillation in Radar Tracking," *Proc. IRE* **41**, 1778-1784 (1953).

apparent center of reflection. For a row of reflection centers uniformly spaced along a length L perpendicular to the line-of-sight from the radar, for example, the fraction of time that a conically scanning radar points off

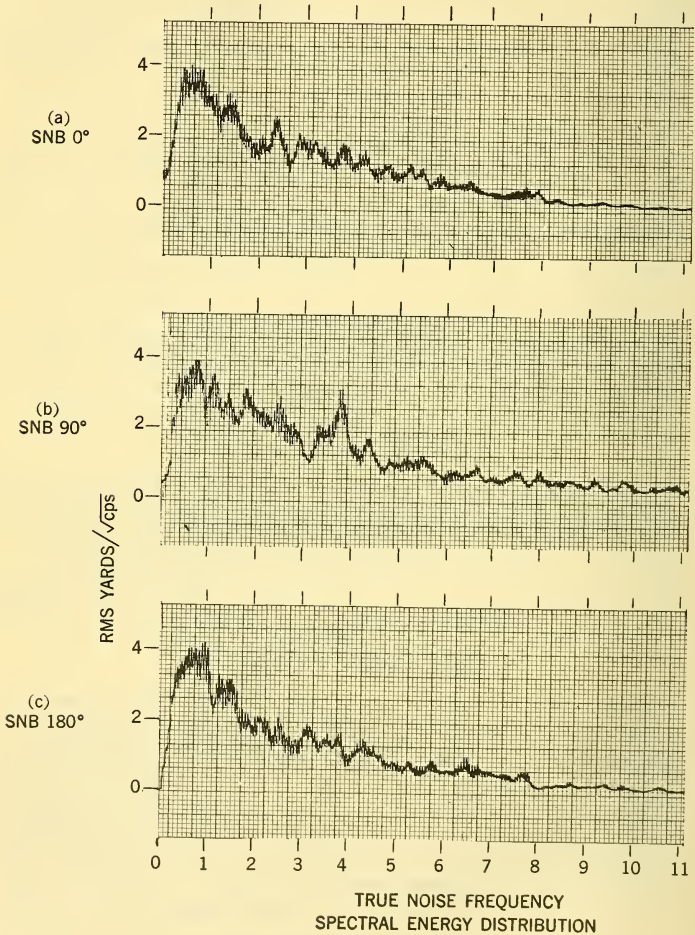


FIG. 4-29 Spectra of Angle Noise for SNB Aircraft.

the target is 0.134. For equal reflection centers uniformly spaced over a circular area, this fraction becomes 0.2.

Data on angle noise have been collected at the U. S. Naval Research Laboratory in connection with investigations of tracking noise. This work has been done at a frequency of about 9400 Mc, so that it is applicable to airborne radar problems. Examples of angle fluctuations of the type discussed above are shown for an R4D in Fig. 4-28. The variations in apparent reflection center are seen to be greater than the linear dimensions of the aircraft. Also, the deviations at the 90° aspect are larger than at 0° and 180° .

Fig. 4-29 shows the spectrum²⁰ of angle noise for several runs of an SNB. The angular noise in this spectrum has been multiplied by the target range so that the spectral density is independent of range and expressed in yards per $\sqrt{\text{cps}}$. The amplitude decreases fairly regularly with increasing frequency, so that the total noise power is finite. In many cases the spectrum can be fitted satisfactorily by a curve of the form

$$A = A_0(1 + f^2/f_0^2)^{-1/2} \quad (4-58)$$

which corresponds to the transfer characteristic of a single-section RC low-pass filter.

Since the vertical span of an aircraft is much less than its horizontal span, angle noise of a single aircraft is much less in elevation than in azimuth. In low-angle tracking, however, reflection from the ground or sea has the effect of creating an image aircraft at an equal distance below the surface (see Fig. 4-13). If the angle between the target and its image is less than the elevation beamwidth, the two will not be resolved. Variations in the phase difference between direct and reflected rays then will cause the effective reflection center to wander between target and image or beyond them. This has been observed to be the case. As the range decreases, the angular fluctuations increase until target and image can be resolved and the tracking system locks on. However, it is possible for lock-on to occur on the image instead of the target!

Multiple targets have a similar effect on azimuthal variations. Thus, multiple targets which are not resolved give rise to a much higher level of tracking noise than a single target.

Range Noise. In addition to causing angle noise, fluctuations of the effective center of reflection of the target can give rise to fluctuations in range, or range noise. Fig. 4-30 shows typical time plots of range noise for several classes of target. Fig. 4-31 shows the range noise of a single SNB at

²⁰The observation time included in this spectrum is about 80 seconds, so that spectral frequencies below about $\frac{1}{2}$ cps are cut off by analyzer limitations.

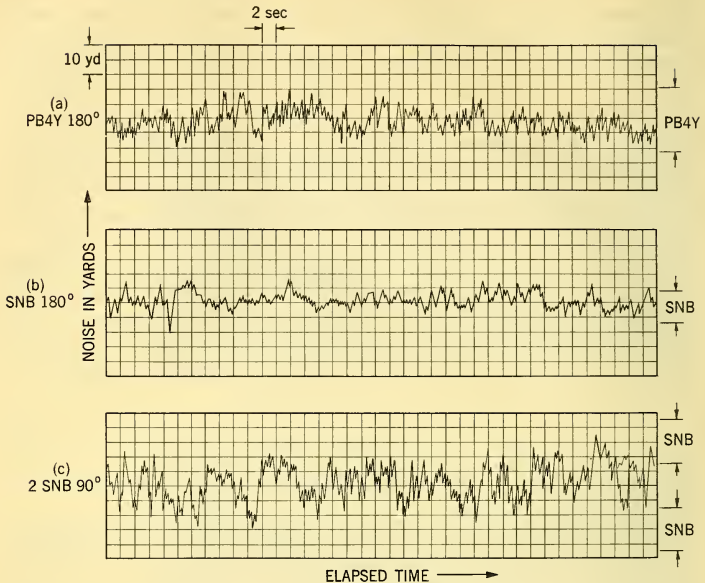


FIG. 4-30 Sample Time Function Plots of Range Noise from the (a) PB4Y (Four-Engine Bomber) at 180° Target Angle, (b) ANB (Two-Engine Transport) at 180° Target Angle, and (c) SNB pair.

aspect angles of 0° , 90° , and 180° , and of two SNB's at 90° . Both the range noise spectrum and its probability distribution are shown.

The distribution of the apparent reflection center in range generally lies wholly within the target, as can be seen from the curves at the right in Fig. 4-31. In this respect, range noise differs from angle noise. The reason lies in the different methods used for error detection in angle and in range.

As in the case of angle noise, multiple targets which are not resolved will give rise to much higher noise levels than a single target.

4-9 PREDICTION OF TARGET RADAR CHARACTERISTICS

Quantitative measurements of radar characteristics require special instrumentation which is not widely available and is costly. Furthermore, targets of interest may not be available for measurement. For example in a problem of the type outlined in Chapter 2, it is highly unlikely that such definitive target information will exist. Hence great importance attaches to methods whereby the characteristics of interest may be calculated. As

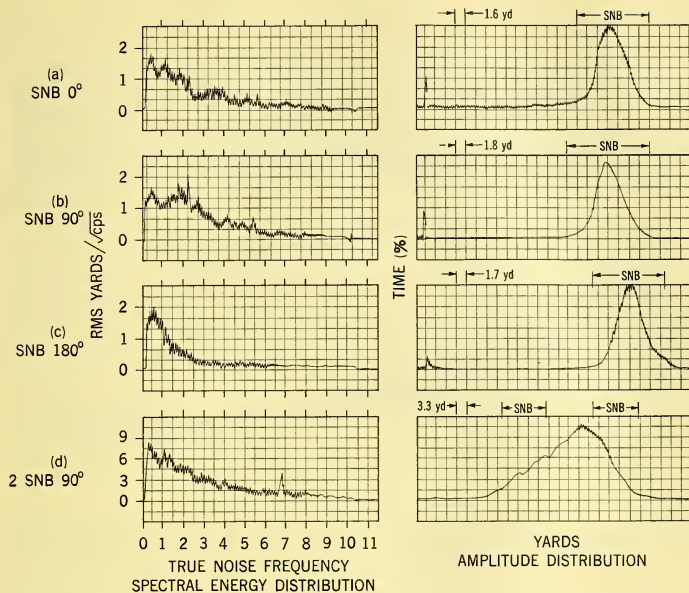


FIG. 4-31 Sample Spectra Obtained for the SNB at Target Angles of (a) 0° , (b) 90° , (c) 180° , and for (d) the SNB pair.

a result of theoretical studies,²¹ and measurement programs such as those referred to in Paragraphs 4-7 and 4-8, techniques have been developed with which rather good success can be expected in predicting the radar characteristics of a target or target complex if their basic configurations are known. A brief discussion will be given of methods which have been used for aircraft targets.

Since the dimensions of aircraft are many wavelengths for airborne radar frequencies, the methods of geometrical and physical optics are sufficiently accurate for most purposes. The principal reflections, therefore, come from surfaces which have portions parallel to the wavefront. The aircraft then can be approximated by a small number of bodies of simple shapes, for which the radar lengths can be calculated. Over a small range of angles about any given aspect, the contributions from the individual bodies will pass through substantially all values of relative phase, so that the average

²¹See, for example, *Studies in Radar Cross-Sections, XV*, University of Michigan, Engineering Research Institute, Report 2260-1-T, Appendix A, and further references therein.

radar area and its rms spread may be calculated quite easily. In this way, these quantities may be determined over the range of aspect angles of interest.

In addition to the average areas, it is possible to predict the target noise. After having replaced the aircraft by a finite number of simple shapes at fixed locations, the doppler frequencies generated by a known variation of angular velocity can be calculated in the manner described in Paragraph 4-4, and the spectrum can be determined. This calculation can be expedited by the use of a target simulator.²² As an example, Fig. 4-32 shows the

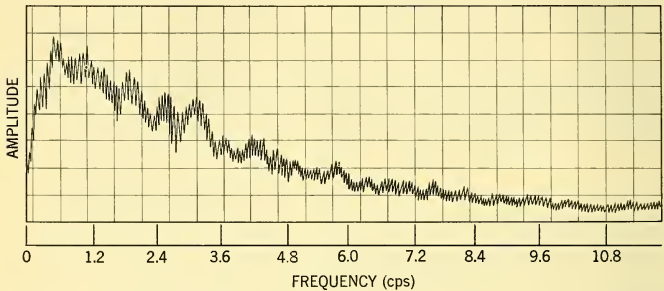


FIG. 4-32 Spectral Distribution of Angle Noise from a B-17 (Four-Engine Bomber) Aircraft (Actual Measurement).

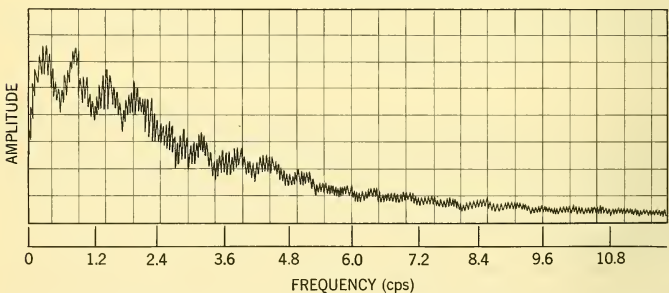


FIG. 4-33 Spectral Distribution of Simulated Angle Noise from a B-17 Target Simulator.

measured angle noise spectrum of a B-17, while Fig. 4-33 shows the simulated spectrum. The latter was obtained with a "target motion" composed of a random oscillation having a Gaussian distribution of velocities about zero mean, with a standard deviation of $0.3^\circ/\text{sec}$.

²²See footnote 17, NRL Reports 4694, 4770, 4796.

4-10 SEA RETURN²³

In detection or tracking of targets on or near the surface, it is necessary to be able to distinguish the target from the background clutter due to reflections from the surface itself. For example, the AEW system discussed in Chapter 2 was required to distinguish between sea return and target echoes. In order to design a radar for such an application, the mechanism of sea return and its relation to radar and tactical parameters must be understood.

In operations over water, this *sea return* or *sea clutter* is caused by all elements of the surface within a resolution element of the radar. Since this surface area is a function of pulse length and antenna beamwidth, the radar area of such a target complex is a function of range and some of the radar parameters. If the area of surface illuminated by a resolution element is not too small, then the return can be considered to be from scatterers uniformly distributed over this area, or *area extensive*. It is then convenient to use a quantity, the *radar area per unit area of sea surface*, usually denoted by σ^0 , which is independent of the radar parameters. Then the radar area σ is

$$\sigma = \sigma^0 A, \quad (4-59)$$

where A is the physical area of a resolution element. For pulse radar with pulse length τ and azimuth and elevation beamwidths Φ and Θ , respectively, where A is the smaller of the two values:

$$A = R\Phi \frac{\tau c \sec \theta}{2} \approx R\Phi L \quad (\theta \text{ small}) \quad (4-60a)$$

$$A = R^2\Phi\Theta/\sin \theta \quad (\theta \text{ large}) \quad (4-60b)$$

with $L = \tau c/2$.

The characteristics of sea return depend on a number of parameters. These are the depression angle, polarization, frequency, and the condition (or "state") of the sea. The last quantity includes the many factors which affect the contour of the surface, such as the wind (its speed, direction, and duration), swell (wave systems generated by distant storms), currents, shoals, breakers, and others. Large waves in themselves do not necessarily produce strong clutter, since a heavy swell, with little or no wind blowing, does not produce a high level of clutter. On the other hand, clutter springs up suddenly with a sudden onset of wind, even before waves of appreciable height are built up. Thus clutter seems to be more intimately connected with the secondary wave structure due to the local wind than with the primary wave structure. Although the many factors which influence sea

²³For a thorough discussion of the World War II investigations of sea return, see the account by H. Goldstein in D. E. Kerr (Ed.), *Propagation of Short Radio Waves*, pp. 481-581, McGraw-Hill Book Co., Inc., New York, 1951.

clutter make the phenomenon a complicated one, by now most of these appear to be understood.

The radar parameters which control sea clutter are the depression angle (θ), polarization, frequency, antenna beamwidths, and pulse length. The last two have been discussed already and may be eliminated, when the return is area extensive. The other three are interrelated.

Since sea return is back-scattering from the surface itself, factors which affect the illumination of the surface elements responsible for the back scattering have an important effect on σ^0 . Polarization, frequency, and wave height are such factors, and their effects are intertwined. Katzin^{24,25} showed that a number of their effects could be explained on the basis of an illumination of the scattering elements which is the combination of direct and reflected waves, similar to that above a plane reflecting surface.

As was shown in Paragraph 4-5, interference between the direct and reflected rays creates a lobe structure above the surface. Below the lowest lobe, $P_R \propto R^{-8}$ for a single target. For an extended target distributed in height from the surface upward, this relation still holds if the top of the target is below the first lobe. At nearer ranges, where the target subtends one or more lobes, the target in effect integrates the varying illumination over it, so that the deep ripples of the lobe pattern are smoothed out and $P_R \propto R^{-4}$. For pulsed radar, the illuminated area of the sea surface at small depression angles is proportional to range, in accordance with Equation 4-60a above, so that the reflection mechanism just described should

give an R^{-3} range variation at short ranges, and an R^{-7} range variation at long ranges.

Fig. 4-34 shows a composite plot of sea-clutter power measured with horizontal polarization at various frequencies and altitudes to test the interference mechanism. Here the measured points clearly define a region where $P_R \propto R^{-3}$, which changes into one where $P_R \propto R^{-7}$. The agreement with the type of behavior just discussed lends strong support to the reflection mechanism. Presumably reflection takes place from the region ahead of the wave crests in the manner indicated in Fig. 4-35.

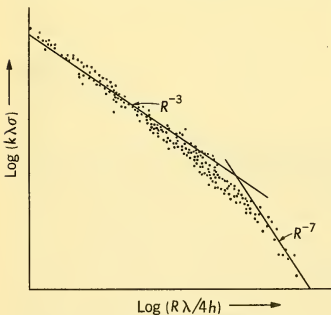


FIG. 4-34 Composite Plot of Sea-Clutter Power at Three Frequencies and Six Altitudes from 200 to 10,000 ft: Coordinates Normalized to Test Interference Mechanism.

²⁴M. Katzin, "Back Scattering From the Sea Surface," *IRE Convention Record*, 3, (1), 72-77 (1955).

²⁵M. Katzin, *On the Mechanisms of Radar Sea Clutter*, *Proc. IRE* 45, 44-54 (1957).

As was shown in Paragraph 4-5, vertical polarization produces a much stronger field on and just above a reflecting surface than does horizontal polarization. Hence the scattering elements of the sea surface are more strongly illuminated if vertical polarization is used so that sea clutter at low angles is much stronger with vertical than with horizontal polarization, assuming that the same for both polarizations.

Because of the presence of reflected waves, the appropriate radar equation for sea clutter is obtained from Equations 4-1 and 4-59:

$$P_r = \frac{P_t G^2 \lambda^2 \sigma^0 A \bar{F}}{(4\pi)^2 R^4} \quad (4-61)$$

where \bar{F} is a suitable average value of F^4 . For a uniform distribution of σ with height above the surface, Katzin²⁵ gives

$$\bar{F} = 6, \quad R < R_t \quad (4-62a)$$

$$\bar{F} = 6(R_t/R)^4, \quad R > R_t \quad (4-62b)$$

where R_t is the transition range between the R^{-3} and R^{-7} regions. The simple plane surface reflection theory for a surface with a reflection coefficient of -1 gives

$$R_t \approx 5hH/\lambda \quad (4-63)$$

where h is the radar height and H the height of the top of the target, which here is to be interpreted as the height of the wave tops above the equivalent reflecting plane. Since wave heights themselves are distributed in a statistical manner, and the location of the equivalent reflecting plane is not known, an empirical relation must be deduced from experiment. A limited amount of experimental evidence suggests the relation

$$R_t \approx 2hH_{1/10}/\lambda \quad (4-64)$$

in which $H_{1/10}$ is the crest-to-trough wave height exceeded by 10 per cent of the waves (a unit frequently used by oceanographers).

A further consequence of the reflection interference phenomenon at very small depression angles is that the return no longer remains "area extensive." The appearance of the sea clutter on an A scope then breaks up into a series of discrete echoes or "spikes" which appear much like individual targets. These can persist at fixed ranges for periods of a number of seconds. Fig. 4-36 shows an example of this. Spikiness is explainable by the com-

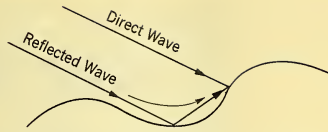


FIG. 4-35 Possible Geometry of Reflected Wave from Sea Surface.

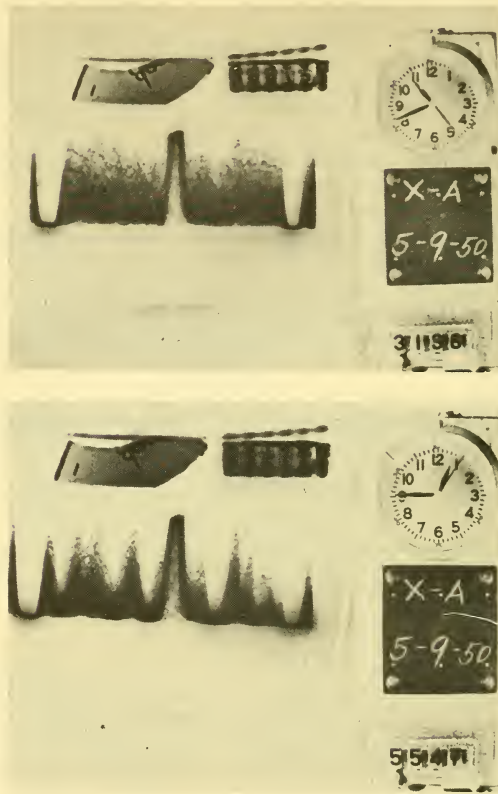


FIG. 4-36 Expanded A-Scope Photographs of Sea Clutter. The Saturated Echo in the Center of the Sweep Is from a Stationary Ship. Blanking Gates Near Both Ends of the Sweep Define the Base Line. Wavelength, 3.2 cm.

bined effects of destructive interference below the first lobe and a statistical variation of wave heights. Because of the statistical distribution of wave heights, there are relatively few waves which exceed the average height, and these thus appear as isolated "targets."^{25,26}

²⁶For a summary of a series of observations of spiky clutter, see F. C. MacDonald, *Characteristics of Radar Sea Clutter, Part I: Persistent Target-like Echoes in Sea Clutter*, NRI Report 4902, March 19, 1957.

The variation of σ^0 with depression angle is a function of wind speed. Fig. 4-37 shows measurements on vertical and horizontal polarization at 24 cm as reported by MacDonald,²⁷ while Figs. 4-38 and 4-39 show measure-

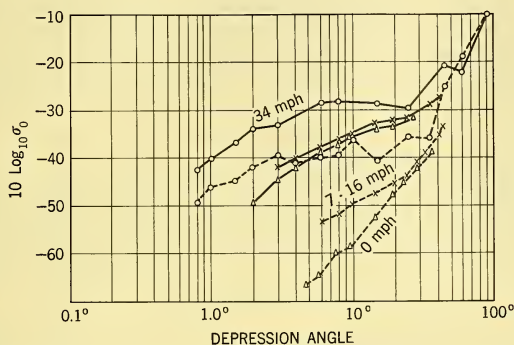


FIG. 4-37 Sea Clutter, 1250 Mc: Solid Line = Transmitted and Received Vertical Polarization. Dotted Line = Transmitted and Received Horizontal Polarization.

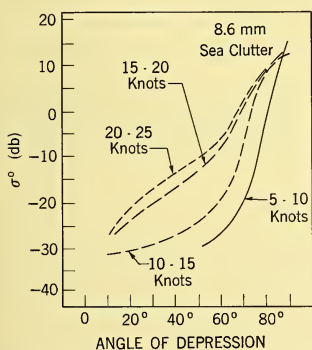


FIG. 4-38 σ^0 as a Function of Wind Velocity; $\lambda = 8.6$ mm.

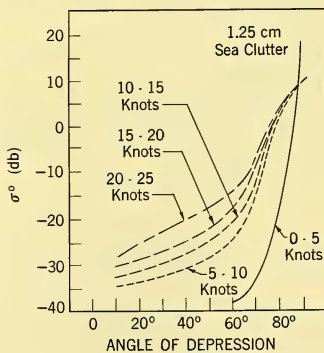


FIG. 4-39 σ^0 as a Function of Wind Velocity, $\lambda = 1.25$ cm.

ments on vertical polarization at $\lambda = 8.6$ mm and 1.25 cm by Grant and Yaplee.²⁸ At small depression angles, σ^0 increases with wind speed, but near vertical incidence this trend is reversed and σ^0 decreases with increasing wind

²⁷F. C. MacDonald, "Correlation of Radar Sea Clutter on Vertical and Horizontal Polarization with Wave Height and Slope," *IRE Convention Record* 4 (1), 29-32 (1956).

²⁸C. R. Grant and B. S. Yaplee, "Back-Scattering from Water and Land at Centimeter and Millimeter Wavelengths," *Proc. IRE* 45, 976-982 (1957).

speed. It should be noted that near vertical incidence, σ^0 rises to as high as +15 db.

These and other characteristics of sea clutter have been explained by a theory developed by Katzin.²⁵ This theory is based on scattering by the small facets of the sea surface as the basic scattering elements. At small depression angles, where none of the facets is viewed broadside, the facets which back-scatter most effectively are those with perimeters of about a half-wavelength. The back-scattering of a facet increases with its slope, so that those near the wave crests contribute most strongly, even if the illumination is constant with height.

Although at small depression angles the back-scattering is at angles far removed from the facet normals, at large depression angles some of the facets are viewed broadside, so that these contribute most strongly in this region. The larger the facet the greater is the contribution. The angular dependence at large depression angles then is governed by the slope distribution of the facets. At airborne radar frequencies, the angular dependence of σ^0 should follow the slope distribution rather closely. This distribution is approximately Gaussian, but is more peaked and is skewed in the upwind-downwind direction.²⁹

At small depression angles, the theory shows that σ^0 is directly proportional to wind speed, but at high angles it is inversely proportional to wind speed. These features of the theory seem to be in accord with available measurements.

The evidence regarding the frequency dependence of σ^0 is not uniform. Katzin²⁵ stated that σ^0 (at small θ) was roughly proportional to frequency in the frequency range 1.25–9.4 kMc, and gave the formula for σ^0 upwind at small depression angles,

$$\sigma^0 = (2.6 \times 10^{-4})W^{3/4}\lambda^{-1} \quad (4-65)$$

where W is the wind speed in knots and λ the wavelength in cm. (In this formula, the illumination factor \bar{F} is included in σ^0 .) Wiltse, Schlesinger, and Johnson³⁰ found σ^0 to be substantially constant in the frequency range 10–50 kMc. Grant and Yaplee²⁸ found σ^0 to increase with frequency range 9.4–35 kMc, the increase being about as the square of frequency at vertical incidence and about as the first power or less at 10° depression angle. Grant and Yaplee's measurements on the different frequencies used were made on different occasions, however, so that their results on the frequency dependence are subject to wider variations due to different surface conditions. It is quite possible that the frequency dependence of σ^0

²⁵C. Cox and W. Munk, "Measurement of the Roughness of the Sea Surface from Photographs of the Sun's Glitter," *J. Opt. Soc. Am.* **44**, 838-850 (1954).

³⁰J. C. Wiltse, S. P. Schlesinger, and C. M. Johnson, "Back-Scattering Characteristics of the Sea in the Region from 10 to 50 kmc, *Proc. IRE* **45**, 220-228 (1957).

may vary somewhat from time to time, depending on the condition of the sea surface.

4-11 SEA RETURN IN A DOPPLER SYSTEM

The doppler shift due to relative motion of radar and target was discussed in Paragraph 4-4, and the echo frequency due to a transmitted frequency f_0 was given as

$$f = f_0 + 2V/\lambda \quad (4-66)$$

where V is the line-of-sight component of the approach velocity of radar and target. If both radar and target are in motion, then with respect to fixed coordinates V may be divided into two parts, one due to the radar velocity V_r , the other to the target velocity V_t . Equation 4-66 correspondingly may be written as

$$f = f_0 + f_r + f_t. \quad (4-67)$$

If the angle of the target from the ground track of the radar is χ , then the doppler frequency due to the radar motion is

$$f_r = (2V_r/\lambda) \cos \chi = f_1 \cos \chi \quad (4-68)$$

$$f_1 = 2V_r/\lambda. \quad (4-69)$$

If the target is the surface of the sea, then the angle χ will vary over the portion of the surface which is illuminated by the radar, owing to the finite width of the antenna beam. Hence there will be induced by the motion of the radar a corresponding band, or spectrum, of doppler frequencies f_r . This may be called an *induced doppler spectrum*.

Similarly, if the various portions of the surface are in relative motion, then even if the radar is stationary or the radar beam is so narrow that no appreciable variation in $\cos \chi$ takes place over the illuminated area, a range of doppler frequencies f_t will result from the intrinsic motion of the surface. This may be called the *intrinsic doppler spectrum*.

The relative importance of the induced and intrinsic doppler components depends on the relative velocities and the geometry, as well as on the antenna beamwidth. Referring to Fig. 4-40,

$$\cos \chi = \cos \theta_0 \cos \phi_0$$

where θ_0 is the depression angle and ϕ_0 the azimuth angle of the surface target relative to the aircraft motion.

Hence for a small azimuth deviation $\pm \Delta\phi$ from the mean value ϕ_0 , we have

$$\begin{aligned} \cos \chi &= \cos \theta_0 (\cos \phi_0 \cos \Delta\phi \pm \sin \phi_0 \sin \Delta\phi) \\ &\approx \cos \theta_0 \{ [1 - (\Delta\phi)^2/2] \cos \phi_0 \pm \Delta\phi \sin \phi_0 \}. \end{aligned} \quad (4-70)$$

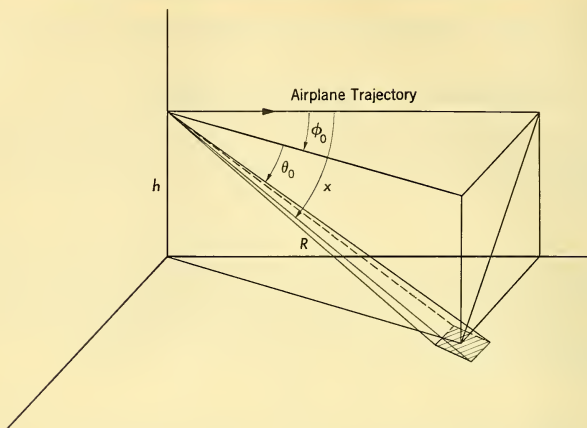


FIG. 4-40 Geometrical Relations for Doppler Spectrum of Sea Return.

Thus it is evident that the spread in $\cos \chi$, and hence the width of the induced doppler spectrum, will be minimum along the ground track ($\phi_0 = 0$). As an example, we consider an airborne X-band pulse radar (9400 Mc/sec) with a horizontal beamwidth of 3° ($\Delta\phi = 1.5^\circ$), and an aircraft speed of 200 knots. Then from Equation 4-69, $f_1 = 6.44$ kc. At grazing depression angles along the ground track ($\theta_0 = 0$), the induced doppler spectrum has a half-power width of $0.000343 f_1 = 2.2$ cps, while at 45° to the ground track the half-power width is $0.0037 f_1 = 238$ cps.

In principle, the induced spectrum is known from information available at the radar and thus may be compensated, in part, by appropriate (though complicated) circuitry. There still remains the intrinsic spectrum, and a knowledge of this is necessary in order to determine the capabilities and limitations of doppler radar in target detection and tracking through clutter.

Measurements of the intrinsic spectrum of sea clutter have been made by the Control Systems Laboratory of the University of Illinois.³¹ These were made with a coherent airborne radar operating on a wavelength of 3.2 cm. By making measurements along the ground track, the width of the induced spectrum was made small relative to that of the measured spectrum, so that the measurements yielded the intrinsic spectrum directly. By multiplying the frequencies by $\lambda/2$ (see Equation 4-69) the results were converted to a *velocity spectrum*.

³¹The information on the intrinsic doppler spectrum of sea clutter was furnished through the courtesy of the Control Systems Laboratory, University of Illinois.

Frequency spectrums were obtained for 15-second samples of recorded data (corresponding to 3750 ft along the sea surface), and also frequency B-scope records of the spectrum as a function of position of the illuminated patch on the sea surface (250 ft long). These will be referred to as the A display and the B display, respectively.

For low sea states, the average spectrum had a Gaussian shape, and width between half-power points of 2 to 3 knots (60–100 cps at X band). The corresponding B display was generally smooth on upwind and downwind edges for all ranges. Fig. 4-41(a) shows a sample of the A and B displays for a low sea condition (wave height 2 ft, wind 9 knots). The 3-db bandwidth of 82 cps in this sample corresponds to a velocity spread of 2.55 knots.

As the wind increased and white caps became evident, the A display broadened asymmetrically to 5 knots or more. The B display then was broadened on the downwind edge to an extent which varied irregularly with range, but the upwind edge remained smooth. Fig. 4-41(b) shows a sample of the A and B displays for a medium sea (wave height 5 ft, wind 16 knots). Here the 3-db bandwidth is 172 cps, corresponding to a velocity spread of 5.35 knots. These characteristics suggest that the irregular downwind broadening was due to spray filaments or patches associated with the white caps, blown off the wave crests and moving downwind more rapidly than the crests.

4-12 GROUND RETURN

The applications of airborne radar over land cover an even broader field than operations over sea. As in the case of sea clutter, reflections from a land surface form a clutter background which tends to obscure the desired echo, e.g. from a target aircraft flying at low altitude. At small depression angles, ground return generally is considerably larger than sea return. Hence the problem of detecting ground targets obscured by ground clutter is correspondingly more difficult.

In another type of radar application — ground mapping — the most important characteristic is the contrast obtainable between objects and their immediate surroundings as determined by the nonuniformity of the return. This characteristic governs the type of ground map which may be obtained by radar techniques, as was discussed in Paragraph 1-4.

The ground return which competes with or obscures the target echo is confined to the return from ground elements at the same apparent range as the target. Such returns can be received either on the main beam or the sidelobes of the antenna pattern. A special form of sidelobe clutter — the altitude line — will be discussed in the next paragraph.

In a pulsed radar the returns which arrive at times precisely separated by the interpulse period appear at the same apparent range. This gives rise to

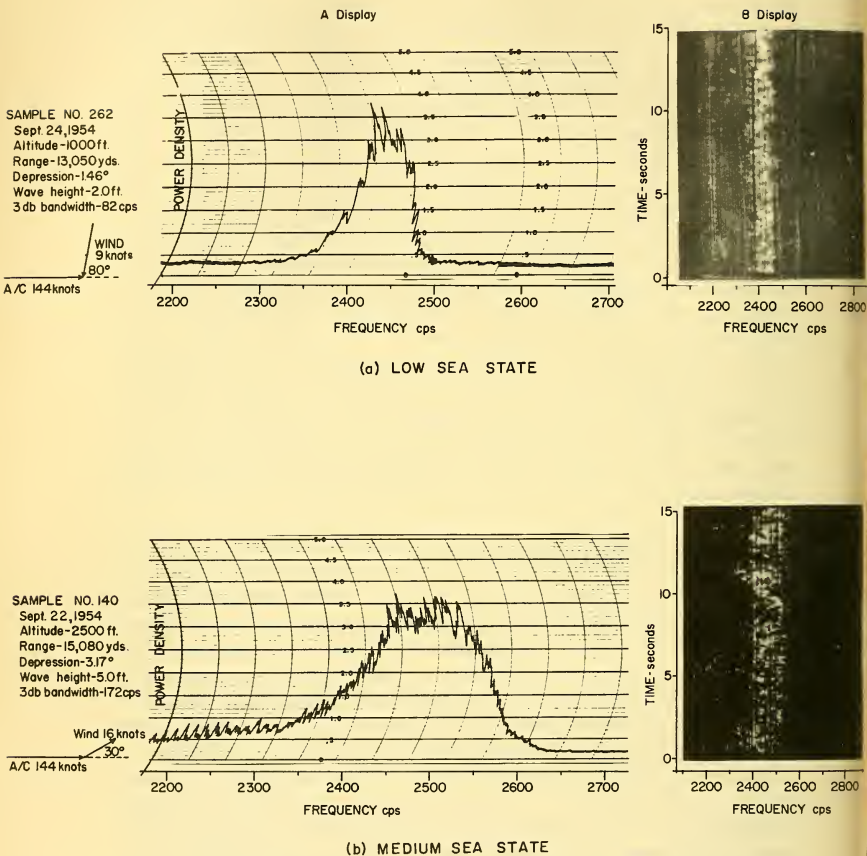


FIG. 4-41 Doppler Spectrum of Sea Clutter, Showing both the Spectrum Averaging over the Sample (a) and the Spectrum vs. Time or Range. A/C velocity refers to Ground Track.

a form of clutter known as *multiple-time-around echo* (MTAE). This clutter is therefore important for ground targets whose range is given by

$$R_{\theta} = R_t + nR_{prf} \quad (4-71)$$

where R_g = range of ground reflector

R_t = target range

R_{prf} = range corresponding to an interpulse period.

Echoes from objects in the range interval $R_t + R_{prf}$ ($n = 1$) are known as *second-time-around echoes* (STAE). It is not unusual for STAE to be comparable to or stronger than the desired target echo. The range of angles for which STAE may be troublesome depends upon the geometry and radar parameters of the particular system under consideration. Obviously, a knowledge of the characteristics of ground return is of importance in this and in other applications.

Some measurements of ground return at wavelengths of 0.86, 1.25, and 3.2 cm are given in a paper by Grant and Yaplee,²⁸ who used vertical polarization. Fig. 4-42 shows their results for a tree-covered terrain with the trees in full foliage. It will be noted that σ^0 is very roughly independent of the angle of incidence. σ^0 also increases with the frequency, but even at $\lambda = 8.6$ mm does not exceed -13 db at any angle. Thus this type of terrain absorbs most of the incident energy.

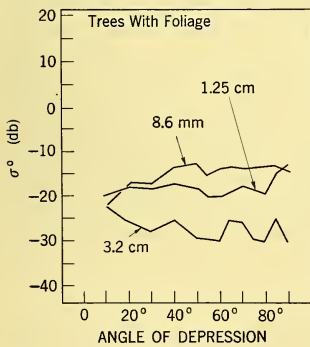


FIG. 4-42 σ^0 for a Tree-Covered Terrain.

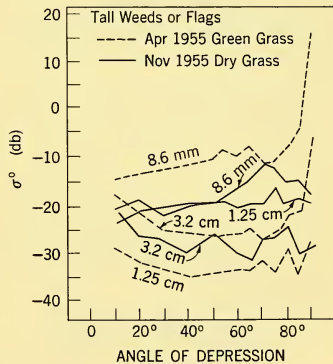


FIG. 4-43 Comparison of σ^0 for Green Grass and Dry Grass.

Fig. 4-43 shows the results for ground covered with tall weeds and grass in the spring when the grass was green and the ground wet and marshy, and in the fall when the grass and ground were dry. Two effects are clearly evident from this figure. (1) There is a very large and rapid rise in σ^0 near vertical incidence, amounting to 15–20 db, when the ground is wet. (2) Although σ^0 increases steadily with frequency under dry conditions, when

the ground is wet the curve for 1.25 cm falls much below the other two. Grant and Yaplee state that this behavior was always found on 1.25 cm when the ground was wet, and suggest that this anomaly may be associated with the water vapor absorption peak near this wavelength (see Paragraph 4-16). Aside from the large rise near vertical incidence, the remainder of the curve lies approximately 5 db higher when the ground is wet than when it is dry.

The large increase of σ^0 near vertical incidence when the ground is wet can be explained as caused by patches of water which are viewed broadside, as in the case of the facets which have been proposed as the scattering elements for sea clutter. This emphasizes the importance of plane surfaces whose dimensions are comparable to or large compared with the wavelength when they are viewed broadside. Hence, in attempting to generalize on the basis of the rather meager experimental results which have been reported in the literature, this characteristic should be kept in mind.

An important example of this is the case of cultural areas, especially cities. Here, in addition to the presence of large flat surfaces, such as building walls, windows, roofs, and streets, there are many possibilities for corner reflectors. Since corner reflectors have a large radar area over a wide range of angles, they have a very large effect on the radar return. For example, observations of ground painting by airborne radar³² show that the signals from man-made structures are often too strong to be fully explained in terms of their size, and that a certain amount of corner-reflector action ("retrodirectivity") in the targets must be present. This action is present principally at long ranges (small depression angles) and is responsible for sharp contrast in the return from arrays of buildings at long ranges. At short ranges, where the depression angle is outside the range of corner-reflector action, this contrast tends to fade. These principles have to be kept in mind, for example, in estimating the effect of STAE from a city on the performance of the radar in an AI, an AEW, or a target-seeking missile application.

4-13 ALTITUDE RETURN

In Paragraphs 4-10 to 4-12, we have discussed the back-scattering properties of the sea and ground in terms of the scattering parameter σ^0 . This has been done in order that the properties could be applied to radars with a wide range of parameters. In order to determine the response for a particular radar, one needs to consider the radar parameters in connection with the scattering characteristics of the surface. One case which is of some importance is that of the altitude return in pulse radar. This is the signal received from the ground directly beneath the aircraft. On a PPI display

³²L. E. Ridenour (Ed.), *Radar System Engineering*, Vol. 1, pp. 100-101, McGraw-Hill Book Co., Inc., New York, 1947.

it gives rise to the "altitude circle," while on an A display it is referred to as the "altitude line". In many cases this return is prominent because of the marked increase of σ^0 which occurs for depression angles near 90° (see Figs. 4-37, 4-38, 4-39, and 4-43).

To a radar altimeter the altitude return is the desired signal, while to target detection and tracking radars it is a source of interference or "clutter." Since the antennas of these two classes of radars have widely different beam patterns, the illumination of the ground as a function of angle may vary widely between different applications. A full discussion of the problem, therefore, is beyond the scope of the present treatment, so that only some of the principal factors will be discussed.

The expressions (Equation 4-60) were given for the area of a resolution element on the surface. For small depression angles this area is proportional to range, while for large depression angles it is proportional to range squared. The distinction between these two in the case of the altitude line is actually a function of altitude. For example, if both the antenna beam and the pulse shapes are rectangular, and if σ^0 is a slowly varying function of angle near vertical incidence (as in the case of Fig. 4-42, for example), then the illuminated area is beamwidth limited if the leading edge of the transmitted pulse passes the outer edge of the antenna beam before the trailing edge of the pulse reaches the ground.

The received power of the altitude line then will vary as the inverse square of altitude in accordance with Equation 4-60b. Because of the inverse square relationship (as contrasted with an inverse fourth power relationship for a point target) the altitude line return can be very strong. This is particularly true for altitude line return from a flat calm sea which tends to act as a perfect reflector (see Figs. 4-37 and 4-38.) However, if the altitude or beamwidth is great enough that the trailing edge of the pulse reaches the ground before the leading edge passes out of the antenna beam, then the return is pulse-length limited, and the received power of the altitude line will vary as the inverse cube of altitude in accordance with Equation 4-60a (see Fig. 4-44).

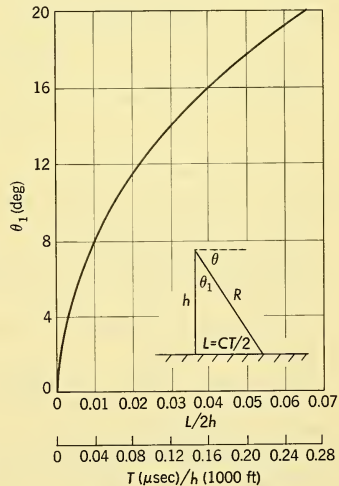


FIG. 4-44 Angular Extent of Altitude Line vs. Pulse Length.

Actually, neither the antenna beam nor the pulse shape is rectangular, and the scattering properties of the ground, even if they are area-extensive, may vary with angle, so that a continuous transition from an inverse square to an inverse cube relation takes place. More complicated situations occur when one or more large individual scatterers are located within the illuminated area. A more detailed discussion of this problem can be found in a paper by Moore and Williams.³³

4-14 SOLUTIONS TO THE CLUTTER PROBLEM

Having considered the characteristics of radar targets and of sea and ground clutter, we can now examine these together in order to find the most favorable solution to the clutter problem. There is no unique solution, since the factors involved depend on the operational problem and the limitations placed on the radar parameters. A full discussion of all the considerations and possible solutions is beyond the scope of this chapter, since the problem involves the overall system design and operational philosophy. We shall restrict ourselves to a consideration of certain features of the antisubmarine warfare (ASW) problem, in order to bring out some interesting possibilities based on sea clutter characteristics discussed in Paragraphs 4-10 and 4-11.

In the first place, an early decision can be made regarding the polarization of the antenna. Both theory and experiment show that sea clutter levels are much lower on horizontal polarization than on vertical polarization. From Fig. 4-37 it is seen that this can amount to 10 db or more. Hence, unless the target shows a preference for vertical polarization by more than this amount, horizontal polarization clearly is to be chosen. Furthermore the discrimination based on target height, which will be discussed below, will be achievable only with horizontal polarization.

The following discussion will be based on a flat earth and will illustrate the principles involved. The modifications necessary to take into account the effect of the earth's curvature have been described in Paragraph 4-6. These will affect the answer only quantitatively and will not change the nature of the results.

The primary mission of airborne radar in ASW is search; tracking is a secondary mission. The object of system design and operation is to choose the radar parameters so that the probability of detection is optimized. Inevitably practical limitations will arise which restrict the ranges of certain of the parameters. Ordinary (non-doppler) pulse radar will be considered first, and then the additional improvement due to doppler radar will be discussed briefly.

³³R. K. Moore and C. S. Williams, Jr., Radar Terrain Return at Near-Vertical Incidence, *Proc. IRE* 45, 228-238 (1957).

Since in search it is desirable to sweep out a large area, the problem is concerned primarily with small depression angles. The variation of received clutter power with range will then be of the form shown in Fig. 4-34, and will be given by Equation 4-61:

$$P_C = \frac{P_t G^2 \lambda^2 \sigma^0 A \bar{F}}{(4\pi)^3 R^4} \quad (4-72)$$

In this we may insert the values of A and \bar{F} given by Equations 4-60a and 4-62, respectively. The horizontal beamwidth Φ and the gain of the radar antenna may be expressed by

$$\Phi = \frac{k_a \lambda}{l_w} \quad (4-73)$$

$$G = \frac{k_b l_w l_h}{\lambda^2} \quad (4-74)$$

where l_w and l_h are the horizontal and vertical antenna apertures, respectively, and k_a and k_b are constants of the antenna design. If we adopt the form of relation given in Equation 4-65 for σ^0 ,

$$\sigma^0 = k_0 / \lambda \quad (4-75)$$

then Equation 4-61 becomes for the received clutter power

$$P_C = \frac{k_C P_t l_w l_h^2 L \bar{F}_C}{(\lambda^2 R^3)} \quad (4-76)$$

where

$$k_C = k_a k_b^2 k_0 / (4\pi)^3$$

$$\bar{F}_C = 6, \quad R < R_C$$

$$\bar{F}_C = 6(R_C/R)^4, \quad R > R_C$$

$$R_C = 2hH_{1/10}/\lambda = \text{transition range for clutter}$$

as in Equations 4-62 and 4-64. k_C is primarily a function of local wind speed, while $H_{1/10}$ is dependent rather on wind history, but may be forecast with reasonably good accuracy.³⁴

Similarly, for the power P_T of the target echo, we have from Equation 4-1

$$P_T = \frac{P_t G^2 \lambda^2 \sigma_T F^4}{(4\pi)^3 R^4} \quad (4-77)$$

If the target is a surface target of uniform section and height H_T , then F^4 is to be replaced by \bar{F} of Equation 4-62, with its transition range, R_T , given by Equation 4-63

$$R_T = \frac{5hH_T}{\lambda} \quad (4-78)$$

³⁴W. J. Pierson, Jr., G. Neumann, and R. W. James, *Practical Methods for Observing and Forecasting Ocean Waves*, H. O. Pub. No. 603, U. S. Navy Hydrographic Office, 1955.

Then P_T may be written as

$$P_T = \frac{k_T P_{l_w}^2 l_h^2 \sigma_T \bar{F}_T}{\lambda^2 R^4}, \tag{4-79}$$

in which $k_T = k_b^2 / (4\pi)^3$.

Plots of Equation 4-76 for a specific sea condition and of Equation 4-79 will then be as shown in Fig. 4-45. From this, it follows that the range scale

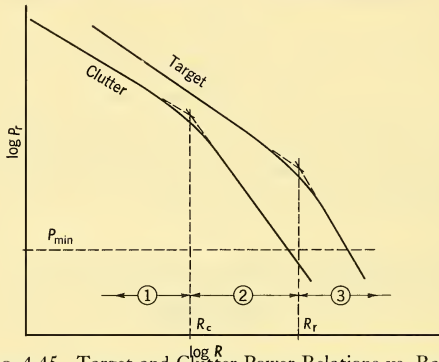


FIG. 4-45 Target and Clutter Power Relations vs. Range.

may be divided into three regions, in which the target-to-clutter ratio may be expressed as follows:

$$\begin{aligned} \text{REGION 1: } \frac{P_T}{P_C} &= \frac{\sigma_T l_w}{k_a k_0 L R} \\ \text{REGION 2: } \frac{P_T}{P_C} &= \frac{\sigma_T l_w}{k_a k_0 L R} \left(\frac{R \lambda}{2h H_{1/10}} \right)^4 \\ \text{REGION 3: } \frac{P_T}{P_C} &= \frac{\sigma_T l_w}{k_a k_0 L R} \left(\frac{5H_T}{2H_{1/10}} \right)^4. \end{aligned} \tag{4-80}$$

Obviously, a large antenna width and a short pulse length will increase the target-to-clutter ratio in all three regions. Furthermore, if σ_T is independent of frequency, then so is P_T/P_C in regions 1 and 3. The locations of the transition ranges R_T and R_C can be controlled by the height h of the radar. It is evident from Fig. 4-45 that the largest target-to-clutter ratios generally will be obtained in region 3.

Since region 3 is one in which destructive interference operates on both the target and clutter signals, this is a region of relatively low signal

strength. Depending on the transmitted power and other radar parameters, therefore, the useful limit of region 3 will be set by the minimum power required to produce a signal detectable above the noise. This minimum power level is indicated by the horizontal dashed line labeled P_{\min} in Fig. 4-45. Since P_{\min} depends on receiver bandwidth, effective antenna scanning rate and beamwidth, and other factors, changes which are made in l_w/L in order to increase P_T/P_C will also increase P_{\min} . Although P_T/P_C (in regions 1 and 3) does not contain an explicit frequency factor, both P_T and P_C contain the factor λ^{-2} , and so increase with frequency.

Fig. 4-45 relates to a specific target area and sea condition. Obviously, one must consider a whole family of such curves, relating to various possible combinations of interest, in order to arrive at the optimum choice of parameters. Some of these parameters depend on the operational philosophy (e.g. barrier patrol, hunt-and-kill). In addition, the effect of the earth's curvature, which will steepen the rates of signal decrease in region 3, will have to be taken into account.

The above discussion refers to non-doppler radar. Doppler radar offers the additional possibility of increasing the target-to-clutter ratio by exploiting differences in the target and clutter spectrums. In order to achieve a gain in target-to-clutter ratio, it is necessary that the target doppler frequency spectrum lie outside the range of the induced doppler spectrum of the clutter. For the example given in Paragraph 4-11 ($V_r = 200$ knots, $\Delta\phi = 1.5^\circ$) each doppler component of the intrinsic doppler spectrum would be broadened by about 2 cps along the ground track and about 350 cps at right angles to the ground track. The corresponding effective velocity broadening would be about 0.1 and 11 knots, respectively. Thus, no significant improvement will be obtained at large angles to the ground track unless the radial component of target velocity exceeds 10-15 knots, for the 3° beamwidth assumed. Smaller beamwidths would reduce this figure proportionately.

In principle it is possible to improve the target-to-clutter ratio by exploiting the difference between the widths of the received target and doppler spectrums. This requires a "velocity" filter (or a set of them). A system employing such techniques is described in Paragraph 6-6, below.

4-15 ATTENUATION IN THE ATMOSPHERE

The atmosphere is almost perfectly transparent to radio waves until frequencies in the microwave region are reached. Attenuation of radio waves in the atmosphere is due to absorption by gases (oxygen and water vapor) and absorption and scattering by suspended particles (precipitation, dust). The first effect will be discussed here, and the second in Paragraph 4-16.

The theory of microwave absorption by oxygen and water vapor has been developed by Van Vleck.³⁵ The oxygen absorption is due to a large number of overlapping resonance lines, resulting in peaks centered at wavelengths of 5 and 2.5 mm, while water vapor has an absorption peak at 1.35 cm. Fig. 4-46 shows the theoretical attenuation due to oxygen for paths at sea

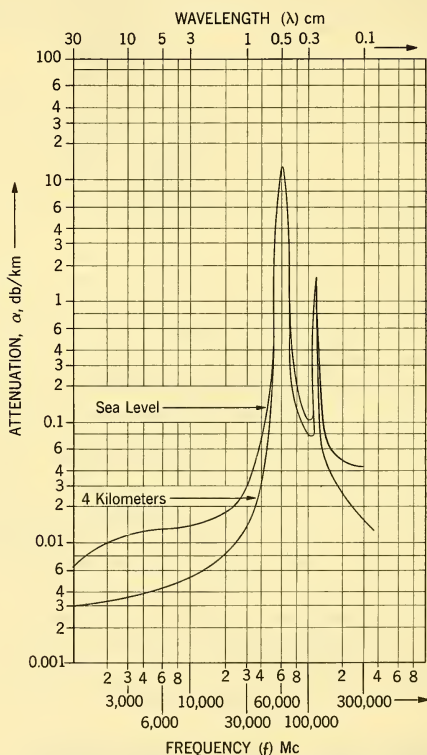


FIG. 4-46 Atmospheric Attenuation Due to Oxygen.

level and at 4 km. Fig. 4-47 shows the theoretical water vapor attenuation in an atmosphere containing 1 per cent water vapor. The attenuation is closely proportional to the water vapor concentration. Experimental points

³⁵See reference of footnote 25, pp. 646-664.

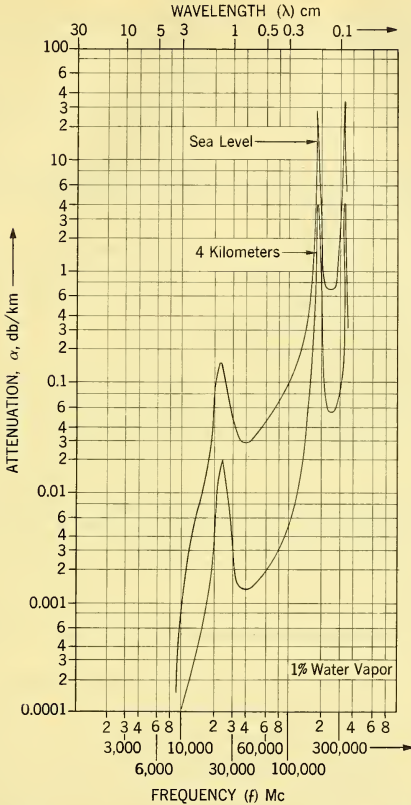


FIG. 4-47 Atmospheric Attenuation Due to Water Vapor.

due to Tolbert and Straiton³⁶ show general agreement with the oxygen attenuation, but for water vapor the measured values are 2.5 to 4 times the theoretical values. Theissing and Caplan³⁷ also found that the water vapor absorption between the peaks was higher than Van Vleck's theoretical curve by a factor 2.7. The reason for this disagreement with the theory is not known.

³⁶C. W. Tolbert and A. W. Straiton, "Attenuation and Fluctuation of Millimeter Waves," *IRE National Convention Record* 5 (1), 12-18 (1957).

³⁷H. H. Theissing and P. J. Caplan, "Atmospheric Attenuation of Solar Millimeter Radiation," *J. App. Phys.* 27, 538-543 (1956).

In the presence of absorption, an additional factor is required in the radar equation. This factor is

$$10^{-0.2\alpha_{db}R} \quad (4-81)$$

where α_{db} is the one-way attenuation in db per unit distance.

4-16 ATTENUATION AND BACK-SCATTERING BY PRECIPITATION

Solid particles suspended in or falling through the air can affect radar operation both by the attenuation to waves passing them, and by the clutter due to back-scattering from them. The attenuation is a combination of absorption by the particles and scattering out of the forward beam. The particles which are most frequently encountered are those due to precipitation — viz., water, snow, and ice (hail). Of these, only water absorbs strongly, so that its attenuation is caused mainly by absorption. We shall give here only some salient features of the attenuation and back-scattering by precipitation, since rather complete summaries have been given in the literature.^{38,39}

For liquid water drops, the attenuation caused by absorption is much larger than that caused by scattering. For small drops ($\pi D/\lambda \ll 1$), the absorption is proportional to D^3 while the back-scattering is proportional to D^6 . Hence the attenuation through small rain drops is proportional to the total liquid water content, but the back-scattering is proportional to ΣD^6 . Thus the larger drops are much more effective in back-scattering than the smaller ones.

Because of the dispersion of water in the microwave region (see Paragraph 4-15) the attenuation varies in a complicated way with frequency, and also with drop size. The total attenuation is the integrated effect of all the drops in the beam between the radar and target, and thus depends on the drop size distribution, the drop density (number of drops per unit volume), and the length of the path through the precipitation. Drop size distribution is known only imperfectly, since most measurements have been made by catching rain drops *at the ground*. The distributions are then usually related to the *precipitation rate*. These may not be the same as the distribution and drop density encountered aloft. A further complication is that the precipitation density usually is not uniform for any great distance through the precipitation region. Hence the calculations made on the basis of such measurements necessarily must be considered as only approximate estimates of the actual effects which may be experienced.

³⁸The wartime research is summarized on pp. 671-692 of the reference of footnote 33 above.

³⁹K. L. S. Gunn and T. W. R. East, "The Microwave Properties of Precipitation Particles," *Quart. J. Roy. Meteorol. Soc.* **80**, 522-545 (1954).

Calculations of attenuation and back-scattering (radar area) for spherical drops have been made by Haddock⁴⁰ on the basis of the drop size distributions of Laws and Parsons.⁴¹ These are reproduced in Figs. 4-48 and 4-49. The total radar area is found by multiplying the value found in Fig. 4-49 by the volume of precipitation illuminated by a pulse length. If the entire antenna beam is filled with precipitation, then this volume is $R^2\Phi\theta L$. The curves in these figures may be extended to longer wavelengths by assuming a dependence as λ^{-4} .

Snow is a mixture of air and ice. Since the refractive index of ice is much smaller than that of water, the scattering and attenuation due to snow are less than those of a corresponding mass of water. However, when a snow flake begins to melt, it becomes coated with a thin film of water. The scattering and absorption then become almost the same as a water particle of the same size and shape and thus increase greatly. This effect has been advanced as the explanation for the radar "bright band" observed at or near the freezing level.

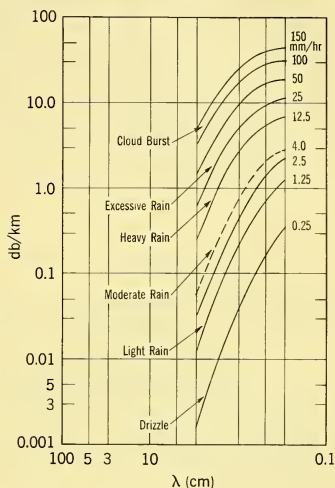


FIG. 4-48 The Variation of Attenuation with Wavelength for Various Rainfall Rates.

4-17 ATTENUATION BY PROPELLANT GASES

In the transmission of information between a missile and its ground control station, the flame of the propellant gases lies in or near the path between the missile antenna and the ground station antenna. Attenuation, reflection, and refraction of the radio waves by the flame then are an important factor in determining the performance of the radio channel. A discussion of the nature of this problem appears in the *Guidance* volume of this series.⁴²

⁴⁰F. T. Haddock, *Scattering and Attenuation of Microwave Radiation Through Rain*, Report of NRL Progress, June 1956.

⁴¹J. O. Laws and D. A. Parsons, "The Relation of Raindrop Size to Intensity," *Trans. Am. Geophys. Union* **24**, 452-460 (1943).

⁴²A. S. Locke (Ed.), *Guidance*, pp. 118-124, D. Van Nostrand Co., Inc., Princeton, N. J., 1955.

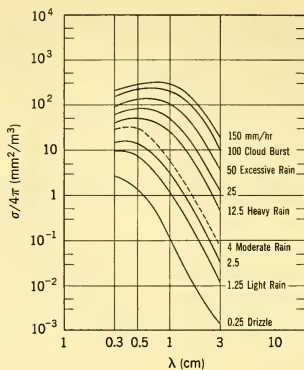


FIG. 4-49 The Variation of Radar Cross Section of Actual Rain-Filled Space with Wavelength, for Various Rainfall Rates.

The ionization processes which render flames conducting are still not completely understood. A recent summary of the subject by Calcote⁴³ presents the status of the understanding of these mechanisms. He cites experimental evidence from the older literature of ion concentrations of 10^{12} cm^{-3} . From the standpoint of radio wave attenuation, only the electron density is of importance, since the conductivity due to a constituent ion of a highly ionized gas is approximately inversely proportional to the mass of the ion (see, for example, *Guidance*,⁴² p. 121, Equation 4-19).

The ion density varies greatly with the type of fuel. Furthermore, the ion density is influenced markedly by small quantities of low-ionization potential contaminants. For example, only trace quantities of the alkali metals such as potassium and sodium are sufficient to increase greatly the ion densities.

Information on the quantitative attenuations to be expected from jet flames can be pieced together from the literature. Adler⁴⁴ measured the attenuation in acid-aniline jet flames in a waveguide and found an attenuation of 0.033 db/m at 200 Mc. Since attenuation is approximately proportional to $f^{1/2}$, this is equivalent to 0.25 db/m at X band. Adler also observed that the addition of slight amounts of sodium caused large and erratic increases in the attenuation. It is probable that much higher attenuations would occur in modern high-energy fuels.

⁴³H. F. Calcote, "Mechanisms for the Formation of Ions in Flames," *Combustion and Flame* 1, 385-403 (1957).

⁴⁴F. P. Adler, "Measurement of the Conductivity of a Jet Flame," *J. Appl. Phys.* 25, 903-906 (1954).

Andrew, Axford, and Sugden⁴⁵ measured the attenuation at X band in the flame of a rifle flash. They found values in the brightest part of the flash of 0.6 db/cm.

The results quoted show that the attenuation in the flames of propellant gases can be serious whenever the geometry is such that the flame is a large obstacle in the path between transmitter and receiver. For example, a flame length of 1 meter in the path could introduce an attenuation at X band in the order of 50-60 db. The effects of the flame are likely to be most serious as the missile ascends into rarefied air and the size of the flame grows. This indicates that special thought should be given to the location and design of the antenna on the missile in order to avoid placing the flame directly in the propagation path.

4-18 REFRACTION EFFECTS IN THE ATMOSPHERE

In computing the power received from a target by means of the radar Equation 4-1, allowance was made for a process other than free-space propagation by means of the propagation factor F . A process which can produce profound modifications is refraction in the atmosphere.

The atmosphere is a nonhomogeneous dielectric because of the variation of its pressure, temperature, and humidity. The variations actually are three-dimensional, but the most pronounced refraction effects are caused by variations in a vertical direction.

In a homogeneous atmosphere, it is convenient to plot rays as straight lines and to show the earth's surface (assumed to be smooth) as a curve. If the atmosphere is not homogeneous it is then more convenient to use the earth's surface as a frame of reference. Rays which are straight lines in space then appear as curves when referred to the earth's surface as the abscissa. This is equivalent to the situation where the earth is *flat* and the (homogeneous) atmosphere has a constant *positive gradient* of refractive index. This is known as the *earth-flattening procedure*, in which the actual refractivity of the atmosphere is replaced by a *modified refractive index*. The modified index is denoted by M and is determined by the equation

$$M = (n - 1 + h/a) \times 10^6 = N + \frac{10^6 h}{a} \quad (4-82)$$

where h = height above the earth

a = radius of the earth.

Its unit of measurement is called the M unit. N is called the *refractivity*, and is the excess of the refractive index over unity, measured in parts per

⁴⁵E. R. Andrew, W. E. Axford, and T. M. Sugden, "The Measurement of Ionization in a Transient Flame, *Trans. Faraday Soc.* **44**, 427-437 (1948).

million. Numerically, $10^6 h/a$ amounts to 0.048 *M unit* per foot. From Equation 4-82

$$\frac{dM}{dh} = \left(\frac{dn}{dh} + \frac{1}{a} \right) \times 10^6. \quad (4-83)$$

It follows from this that a homogeneous atmosphere ($dn/dh = 0$) has an *M curve* with dM/dh equal to $10^6/a$, and that an atmosphere with a constant gradient of refractive index is equivalent to a homogeneous atmosphere of effective radius a_e , where

$$\frac{1}{a_e} = \frac{1}{a} + \frac{dn}{dh}. \quad (4-84)$$

In temperate climates an average value of dn/dh is about $-1/(4a)$. Hence from Equation 4-84

$$a_e = \frac{4}{3}a \quad (4-85)$$

which is the so-called "four-thirds earth." Such an atmosphere is known as the *standard atmosphere*, and the corresponding *M curve*, which is a straight line of slope 0.036 *M unit* per foot, as the *standard M curve*. Actually the *M curve* is rarely a straight line except in a restricted height range.

The *M curve* is useful in ray tracing, since a one-to-one correspondence exists between the change in slope of a ray over a height interval and the change in *M*. In fact, if θ represents the elevation angle, measured in mils (1 mil = 10^{-3} radian = 3.44 minutes of arc), at height h where the modified index has the value *M*, and θ_0 , M_0 are the corresponding quantities at a reference height h_0 (such as the ground), then

$$\theta = \sqrt{\theta_0^2 + 2(M - M_0)}. \quad (4-86)$$

It can be seen from this that a height interval over which $M - M_0$ is negative will give rise to a decrease in the absolute value of the elevation angle. Also, if the *M curve* has a sufficiently large negative excursion ($M_0 - M_{min} > \theta_0^2/2$), then the ray will become horizontal at a certain height, and then curve back to earth. Assuming no loss in reflection at the earth's surface, the process will be repeated over and over, and the ray will go through a succession of hops along the surface. The ray is then *trapped* between the earth's surface and the height at which it becomes horizontal. A region of the atmosphere within which certain rays are trapped is called an *atmospheric duct*. The multi-hop trajectory resembles somewhat the crisscross path between the walls in waveguide propagation; and like a waveguide, an atmospheric duct can trap only waves of frequency higher than a lower limit. For effective utilization of the duct, both the radar and the target should be within the duct.

Since $2(M_{min} - M)$ seldom exceeds 100 M units, trapping occurs only for rays with maximum elevation angle (which may or may not occur at the ground) of the order of 10 mils, or about $\frac{1}{2}^\circ$. Hence trapping is a phenomenon which occurs only in almost horizontal propagation.

The refractivity of the atmosphere for radio frequencies under about 4×10^4 Mc⁴⁶ is given by the formula

$$N = \frac{77.6}{T} \left(P + 4.81 \times 10^3 \frac{e}{T} \right) \quad (4-87)$$

where T is the absolute temperature ($^\circ\text{K}$), p the total pressure, and e the partial pressure of water vapor, both in millibars. The refractivity decreases with an increase in temperature, but increases with pressure, and is especially sensitive to variations in vapor pressure.

The refractivity at a given point usually fluctuates with time, so that average values are used for drawing an M curve. The principal types of M curves observed are illustrated in Fig. 4-50. Curve (a) is the standard

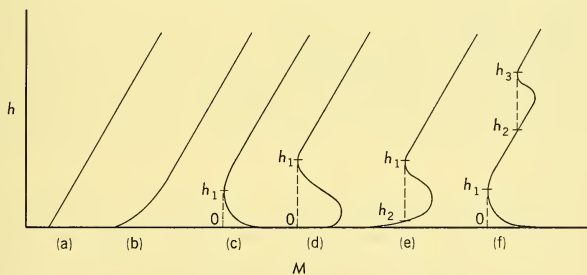


FIG. 4-50 Various Classes of M Curves.

M curve already referred to. The *substandard* M curve, shown in (b), is so called because the rays are refracted less than in the standard case, and it generally results in lower field strengths. Curves (c) and (d) are types associated with surface ducts. The duct extends from the surface to the height h_1 , the "nose" of the M curve. In (e) the value of M at the surface is less than that at the nose, so that the duct then extends from h_1 to h_2 . This is called an *elevated duct*. Various combinations of types can take place, such as a surface duct (0 to h_1) with an elevated duct (h_2 to h_3) shown in (f).

From Equation 4-87, situations where the temperature increases with height together with a simultaneous decrease of vapor pressure lead to a strong decrease of M with height. Such situations are favorable for duct formation. Just such conditions occur at subsidence inversions. These

⁴⁶See Essen and Froome, *Proc. Phys. Soc. London*, B64, 873 (1951).

usually give rise to elevated ducts, since inversion levels commonly occur at 5000 to 10,000 ft. In some cases subsidence inversions descend low enough to form a strong surface duct of the type shown in Fig. 4-50d.

Inversions can also be produced by cooling of the ground at night through the process of radiation. In the absence of wind, the radiation inversion grows upward as the night progresses, forming a surface duct of the type shown in Fig. 4-50c.

Strong surface ducts are formed when warm air from a large land mass moves out over water. The air in contact with the water is cooled and moistened. This cooling and pickup of moisture works its way upward with time by eddy diffusion. As a result, during the formative process an inclined duct usually results, which can extend 200 miles or more out to sea. Duct heights can extend up to 1000 feet or so, and hence can influence airborne radar operation.

Weak surface ducts are formed over the open oceans in the trade-wind regions. Here the air is colder than the water, so that an *increase* of temperature with height is accompanied by a *decrease* of vapor pressure with height. Their effects on the refractivity thus oppose, as can be seen from Equation 4-87, but the influence of the moisture predominates. These ducts are very persistent, lasting almost all year round, incident to the persistence of the trade winds. The duct height is about 50-75 ft, so that they are not very important for airborne radar, except possibly in unusual situations.

An adverse effect on airborne radar can occur when an elevated layer lies below the radar and the target. Then, in addition to a direct ray, a ray refracted by the layer can be received. At certain ranges, well within the horizon, the two rays can interfere destructively, resulting in a decrease in field. This is referred to as a *radio hole*. Radio holes have been observed in which the field strength falls by as much as 15 db over a one-way path, which would mean a 30-db drop for a radar path. Radio holes extend in range for 20 to 50 miles, and so can seriously decrease the range of an airborne radar.

Radio holes have been shown⁴⁷ to be caused by only small departures of the *M* curve from a straight line. A layer in which the slope changes by as little as 10 per cent of the slope in adjoining regions can produce a radio hole. It has been estimated that layers of this kind are present at altitudes between about 5000 and 10,000 ft between 50 and 95 per cent of the time. Thus this phenomenon can have a profound effect on airborne radar.

Many of the effects of the varying refractivity of the atmosphere can be deduced, and to a certain extent predicted, from climatological considerations. However, most of the propagation measurements which have been

⁴⁷*Investigation of Air-to-Air and Air-to-Ground Experimental Data*, Final Report Part III, Contract AF33(038)-1091, School of Electrical Engineering, Cornell University, 10 Dec. 1951.

made to evaluate the effects of atmospheric refraction have been either between two ground stations, or between an aircraft and a ground station. Thus the situations which are encountered in the use of airborne radar have not been explored sufficiently to yield a quantitative understanding of the meteorological effects which may be encountered.

CHAPTER 5

TECHNIQUES FOR SIGNAL AND NOISE ANALYSIS

5-1 INTRODUCTION

The performance of radar systems can often be determined only by tracing the received signal and corrupting noise in detail through the individual system components in order to establish the cumulative effect of each operation. In this chapter, some of the mathematical methods of signal and noise analysis which are appropriate for studies of this kind will be developed and their application illustrated with several examples. These examples will include a discussion of the characteristics of signal plus noise after undergoing some common nonlinear operations, the erratic performance of an angle tracking system in response to internally generated noise, the clutter cancellation which can be achieved with a moving target indicator (MTI system) and the characteristics of a matched filter radar.

Noise analysis embodies a generalization of classical Fourier methods which recognizes the statistical properties of random noise. Much of this material will be presented briefly. More detailed discussions can be found in the referenced literature.¹⁻³

5-2 FOURIER ANALYSIS

To develop the theory and methods of noise analysis several basic ideas relating to the representation of functions in terms of their frequency components as Fourier integrals are required. This paragraph explains and illustrates the concepts of:

1. Fourier integrals or transforms and inverse transforms.
2. Energy density spectra.
3. Transfer functions and impulse responses.

¹S. O. Rice, "Mathematical Analysis of Random Noise," *Bell System Tech. J.* **23**.

²J. L. Lawson and G. E. Uhlenbeck, "Threshold Signals," Chap. 2 (Radiation Laboratory Series) McGraw-Hill Book Co., Inc., New York, 1950.

³P. M. Woodward, *Probability and Information Theory with Applications to Radar*, McGraw-Hill Book Co., Inc., New York, 1953.

We are familiar with the representation of periodic functions by Fourier series. A Fourier integral is a limiting case of such a series where the period becomes indefinitely long. The separation between components becomes indefinitely small as do their magnitudes. For properly restricted functions, however, the magnitude density possesses well-defined values and a Fourier integral exists. The restrictions on a function $f(t)$ in order that it have a Fourier integral are that the integrals of both its square and its absolute value have finite values and that it possess only a finite number of discontinuities in any finite interval. When these conditions are met, a function $F(\omega)$ can be defined by the relation

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt. \quad (5-1)$$

If we suppose that $f(t)$ is a function of time, then $F(\omega)$ is the *spectrum* of $f(t)$ and gives the density of its differential frequency components in much the same way that a Fourier series gives the resolution of a periodic function into finite frequency components. The variable ω is the angular frequency equal to 2π times the cyclical frequency. In general, $F(\omega)$ may be complex.

The time function $f(t)$ is given by the integral of all the differential Fourier components in a manner very similar to the way in which the sum of all the components of a Fourier series represents a periodic function. Thus, $f(t)$ can be represented in terms of $F(\omega)$ by the integral

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{j\omega t} d\omega. \quad (5-2)$$

The functions $f(t)$ and $F(\omega)$ are often regarded as constituting a *Fourier transform* pair which are mutually related by Equations 5-1 and 5-2. With this terminology, Equation 5-1 is said to transform $f(t)$ into the frequency domain, while the operation indicated in Equation 5-2 constitutes the inverse transformation. The symmetry of these transforming operations is striking.

As a concrete illustration of such a pair of functions, suppose that $f(t)$ is zero for negative values of time while for positive values it is a decaying exponential:

$$f(t) = e^{-t}, \quad 0 < t < \infty. \quad (5-3)$$

The spectrum is easily calculated:

$$F(\omega) = \int_0^{\infty} (e^{-t}) e^{-j\omega t} dt = \frac{1}{1 + j\omega}. \quad (5-4)$$

In this case, the spectrum is complex. Upon performing the inverse operation indicated by Equation 5-2, the exponential function given by Equation 5-3 will again be obtained. We shall not carry out the details of

this calculation, which involves treating ω as a complex variable and integrating around a semicircular contour in the complex plane.

The square of the absolute value of the spectrum is important in the development of techniques for analyzing noise processes. This is the *energy density spectrum* giving the distribution of signal energy with frequency. This terminology is adopted because the function $f(t)$ will normally be a voltage or its equivalent, and its square will be proportional to power. The integral of the square of $f(t)$, then, will be proportional to the total energy. In the development to be given below, it will be shown that (in this sense) the square of the absolute value of the spectrum of $f(t)$ gives a resolution of the energy into frequency components. This development is obtained by manipulating a general definition of the energy density spectrum as it is derived from Equation 5-1:

$$\begin{aligned} |F(\omega)|^2 &= F(\omega) F^*(\omega) \\ &= \int_{-\infty}^{\infty} f(t_1) e^{-j\omega t_1} dt_1 \int_{-\infty}^{\infty} f(t_2) e^{j\omega t_2} dt_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(t_1) f(t_2) e^{-j\omega(t_1 - t_2)} dt_1 dt_2. \end{aligned} \quad (5-5)$$

Making the substitution $\tau = t_1 - t_2$ and $d\tau = dt_1$ and interchanging the order of integration

$$\begin{aligned} |F(\omega)|^2 &= \int_{-\infty}^{\infty} e^{-j\omega t} d\tau \int_{-\infty}^{\infty} f(t_2 + \tau) f(t_2) dt_2 \\ &= \int_{-\infty}^{\infty} e^{-j\omega t} \varphi(\tau) d\tau. \end{aligned} \quad (5-6)$$

The right-hand side of Equation 5-6 is of exactly the same form as Equation 5-1; that is $|F(\omega)|^2$ is expressed as the spectrum of the function $\varphi(\tau)$ or its Fourier transform. If $\varphi(\tau)$ satisfies the conditions prescribed for the existence of a Fourier integral, then the inverse operation given by Equation 5-2 is applicable, and $\varphi(\tau)$ can be expressed by

$$\varphi(\tau) = \int_{-\infty}^{\infty} f(t + \tau) f(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |F(\omega)|^2 e^{j\omega\tau} d\omega. \quad (5-7)$$

When τ is set equal to zero, the following important special case is obtained.

$$\varphi(0) = \int_{-\infty}^{\infty} f^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |F(\omega)|^2 d\omega. \quad (5-8)$$

This relation is often referred to as *Parseval's equality*. It expresses the idea that the total energy of $f(t)$ is equal to the sum of the energies of each component of the frequency representation of $f(t)$.

Continuing with the example adopted in Equation 5-3, the form of $\varphi(\tau)$ should be easy enough to find in this case.

$$\begin{aligned}\varphi(\tau) &= \int_0^{\infty} (e^{-(t+\tau)})(e^{-t})dt, \quad \tau > 0 \\ \varphi(\tau) &= \int_{-\tau}^{\infty} (e^{-(t+\tau)})(e^{-t})dt, \quad \tau < 0 \\ \varphi(\tau) &= e^{-|\tau|} \int_0^{\infty} e^{-2t} dt = \frac{1}{2} e^{-|\tau|} \\ \varphi(0) &= \frac{1}{2}.\end{aligned}\tag{5-9}$$

Also, the absolute value of the spectrum is easily obtained from Equation 5-4 in this case:

$$|F(\omega)|^2 = \frac{1}{1 + \omega^2}.\tag{5-10}$$

By virtue of the relationship indicated in Equations 5-6 and 5-7, the functions given by the two equations above must constitute a Fourier transform pair, and the total energy in the signal is $\frac{1}{2}$.

We consider next the effect of transmission through a linear network on the time history and spectrum of a signal. Linear networks are conveniently characterized in terms of either their impulse response or their transfer function. The *impulse response*, sometimes called the *network weighting function*, is simply the transient output of the network for a unit impulse⁴ at the time $t = 0$. The *transfer function* is most commonly defined as the complex ratio of the network output to an input of the form $\exp(i\omega t)$. These two functions are closely related. In fact, the transfer function is the Fourier transform of the impulse response. This relation is made more understandable by noting that an impulse function has a uniform spectrum (see Paragraph 5-5) and so represents an input of the required form where all the frequency components occur simultaneously with differential amplitudes. As an example, consider the single-section, low-pass, RC filter shown in Fig. 5-1. Suppose that the driving point impedance is zero and the load impedance is very large. Then the transfer function of this network is readily recognized as

$$\text{Transfer function} = \frac{1/(jC\omega)}{1/(jC\omega) + R} = \frac{1}{1 + jRC\omega}.\tag{5-11}$$

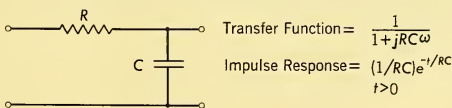


FIG. 5-1 RC Filter.

⁴See Paragraph 5-3 for a definition of an impulse function and a discussion of its properties.

Similarly, the impulse response of the network is recognized as a decaying exponential. If the RC time constant is assumed to be unity, the impulse response and transfer function are identical with the functions given as an example in Equations 5-3 and 5-4 which make up a Fourier transform pair.

We denote the transfer function of a network and the input and output spectra by $Y(\omega)$, $F_i(\omega)$, and $F_o(\omega)$, respectively. Since the input spectrum gives the resolution of the input into components of the form $\exp(j\omega t)$ and the transfer function indicates how each such component is modified by transmission through the network, it is clear that the output spectrum should be given by the product of these two functions. This can be rigorously demonstrated.⁵

$$F_o(\omega) = Y(\omega) F_i(\omega). \quad (5-12)$$

The relation between the input and output energy density spectra is easily found by multiplying each side of this equation by its conjugate:

$$|F_o(\omega)|^2 = |Y(\omega)|^2 |F_i(\omega)|^2. \quad (5-13)$$

Thus the input and output energy density spectra are related by the absolute square of the transfer function, which might appropriately be called the *energy transfer function* or, if power spectra are being considered, the *power transfer function*.

It is often convenient to express the time history of the output of a network purely in terms of the time history of the input and the impulse response of the network. This relation is easily determined by substituting for $Y(\omega)$ and $F_i(\omega)$ in Equation 5-12 their expressions as Fourier transforms of $y(t)$ and $f_i(t)$, the filter impulse response and the input to the filter:

$$F_o(\omega) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y(t_1) f_i(t_2) e^{-j\omega(t_1+t_2)} dt_1 dt_2. \quad (5-14)$$

Substituting $\tau = t_1 + t_2$ and $d\tau = dt_1$, and interchanging the order of integration,

$$F_o(\omega) = \int_{-\infty}^{\infty} e^{-j\omega\tau} d\tau \int_{-\infty}^{\infty} f_i(t_2) y(\tau - t_2) dt_2. \quad (5-15)$$

The right-hand side of this expression is again in the form of Equation 5-1; that is, $F_o(\omega)$ is given as a Fourier transform. Thus we can formally make an inverse transformation of both sides to obtain the desired relation between the input and output time histories:

$$f_o(\tau) = \int_{-\infty}^{\infty} f_i(t) y(\tau - t) dt. \quad (5-16)$$

⁵See M. F. Gardner and J. F. Barnes, *Transients in Linear Systems*, Vol. 1, pp. 233-236, John Wiley & Sons, Inc., New York, 1942.

5-3 IMPULSE FUNCTIONS

Impulse or *delta functions* (so called because they are often denoted by the symbol δ) provide a most useful mathematical device in signal and noise studies. These functions can be visualized as the limiting form of a function whose integral is unity but which is concentrated at a particular value of its argument. Specific representations of impulse functions may take a number of forms. One such form is shown in Fig. 5-2. In this figure a rectangular function of height A and width $1/A$ is shown centered at the point t_1 . As A becomes large, the function becomes very highly concentrated at the point t_1 . For any finite value of A , though, the integral of this function will be unity and independent of A . Thus, the limit of this integral as $A \rightarrow \infty$ exists and is equal to the value of the integral. In physical problems, it is conventional to suppose that these operations are interchanged and that an impulse function denoted by $\delta(t - t_1)$ whose integral is unity is given by the limit of the function pictured in Fig. 5-2 as $A \rightarrow \infty$. This certainly seems reasonable in view of the fact that for any finite A , no matter how large, the integral is unity. Unfortunately, though, integration over the singularity produced when $A \rightarrow \infty$ cannot be justified in a mathematical sense, and these operations cannot correctly be interchanged. Thus, although we shall formally regard impulse functions conventionally as being infinite in height with unit integrals, there is an implicit understanding that the limiting operation must, in actuality, be carried out after the finite function has been integrated. In this connection we note that impulse functions acquire physical significance only after being integrated and do not in themselves represent the end product of any calculation.

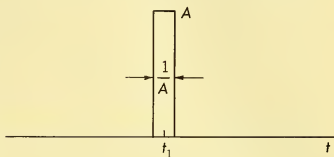


FIG. 5-2 Representation of an Impulse Function.

With these provisos, we proceed to a discussion of some of the properties of impulse functions. Probably their most important characteristic is their sampling property. The integral of the product of a continuous function and an impulse is simply the value of the continuous function at the location of the impulse. We can establish this relation with the aid of the representation pictured in Fig. 5-2:

$$\int_{-\infty}^{\infty} f(t)\delta(t - t_1)dt = \lim_{A \rightarrow \infty} A \int_{t_1 - 1/2A}^{t_1 + 1/2A} f(t)dt = f(t_1). \quad (5-17)$$

Additional properties can be established by finding the Fourier transform of an impulse function:

$$\int_{-\infty}^{\infty} \delta(t) e^{-j\omega t} dt = 1. \quad (5-18)$$

Thus the spectrum of an impulse is constant or uniform. Such a spectrum is often referred to as "white noise" in view of the fact that all frequencies are equally represented. If the spectrum of an impulse (unity) is multiplied by the transfer function of a network, the spectrum of the network output is seen to be simply the transfer function itself. Thus, formally at least, the transfer function of a network is the Fourier transform of the transient response of the network to an impulse function input as was noted in Paragraph 5-2.

The constant spectrum given by Equation 5-18 does not have a finite integral and so does not properly have an inverse Fourier transform. We can, however, approximate this spectrum by one which is unity for $|\omega| < A$ and zero for $|\omega| > A$, where A is very large but finite, and this approximation will have an inverse transform. This inverse transform should have characteristics very similar to the finite impulse pictured in Fig. 5-2 and should approach an impulse function as $A \rightarrow \infty$. This turns out to be true and gives us a second representation for impulse functions:

$$\delta(t) = \lim_{A \rightarrow \infty} \frac{1}{2\pi} \int_{-A}^A e^{j\omega t} d\omega = \lim_{A \rightarrow \infty} \frac{\sin At}{\pi t}. \quad (5-19)$$

This expression occurs often in signal and noise studies. Many important functions cannot be transformed from the time to the frequency domain because the Fourier integral Equation 5-1 does not converge with time. Approximations to this integral, however, can often be derived on the same basis as for the uniform spectrum. When this is done, the resulting expression often contains expressions which can be interpreted as impulse functions in the limit.

For example, it was just established that the Fourier transform of a constant $f(t) = 1$ is an impulse, $2\pi\delta(\omega)$, which denotes concentration of the frequency spectrum at zero frequency.

Similarly, a sinusoid will have a spectrum which may be derived as

$$\begin{aligned} \lim_{A \rightarrow \infty} \int_{-A}^A \cos \omega_1 t e^{-j\omega t} dt &= \lim_{A \rightarrow \infty} d \left[\frac{\sin(\omega + \omega_1)A}{\omega + \omega_1} + \frac{\sin(\omega - \omega_1)A}{\omega - \omega_1} \right] \\ &= d\pi[\delta(\omega + \omega_1) + \delta(\omega - \omega_1)]. \end{aligned} \quad (5-20)$$

This expression indicates a spectrum which is concentrated at the positive and negative values at the frequency of the sinusoid.

In deriving the impulse function representation given in Equation 5-19, a constant over the entire range of ω was approximated by a truncated function which approached the constant function in the limit. Other

approximations to the constant function will give different impulse function representations. A third representation of an impulse function can be obtained in this way by using the triangular approximation shown in Fig. 5-3. As $A \rightarrow \infty$, this triangular function obviously approaches a constant.

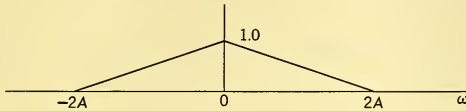


FIG. 5-3 Triangular Approximation to a Constant Spectrum.

The limit of the Fourier transform of this function will give the desired representation:

$$\begin{aligned} \delta(t) &= \lim_{A \rightarrow \infty} \frac{1}{2\pi} \int_{-2A}^{2A} \left(1 - \frac{|\omega|}{2A}\right) e^{j\omega t} d\omega \\ &= \lim_{A \rightarrow \infty} \frac{1}{\pi A} \left(\frac{\sin At}{t}\right)^2. \end{aligned} \quad (5-21)$$

It is apparent that there are a variety of specific representations of impulse functions. A familiarity with the forms of the representations, so that they may be recognized when they arise during the course of an analysis, is useful. A case of this kind occurs in Paragraph 5-5, where in an example of a noise process the expression in Equations 5-21 turns up as part of the power density spectrum (Equation 5-40).

5-4 RANDOM NOISE PROCESSES

In describing noise mathematically, it is useful to visualize a very large group or ensemble of noise generators with outputs $x(t)$, $x'(t)$, $x''(t)$, The output of a specific noise generator may be any one of the ensemble functions with equal probability. The totality of all possible noise functions is referred to as a *random process*. Such processes are described in terms of their statistical characteristics over the ensemble. Fig. 5-4 shows a few of the elements of a noise process. At any time t the mean value, the variance, or other

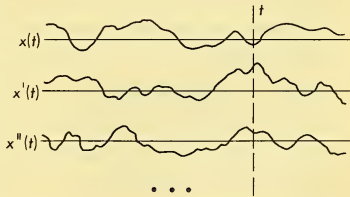


FIG. 5-4 Elements of a Noise Process.

statistical parameters can be determined. These parameters can all be derived from the *probability density function* of the process at that time which describes the distribution of values of the elements of the process.

The integral of the probability density function between any two values will give the fraction of the elements of the process which lie between those values. As an example, we consider the most common type of noise process, a *Gaussian process*, so called because the probability density is Gaussian or normal in form:

$$\text{Probability density function of a Gaussian noise process} = \frac{1}{\sqrt{2\pi}\sigma} \exp \frac{x^2}{2\sigma^2} \quad (5-22)$$

This process has an average value of zero and a variance or mean square value of σ^2 . Most important, the probability density is independent of time. For most of the noise processes which are of importance in engineering applications the statistical parameters are independent of time, and such processes are therefore called *stationary processes*.

Fig. 5-5 shows the probability density function for a Gaussian process. Most of the elements of the process have values in the neighborhood of the

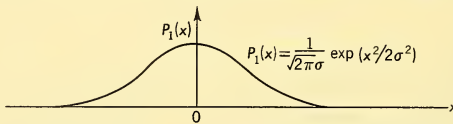


FIG. 5-5 Gaussian Probability Density Function.

origin. Only a very few of the noise functions will be very large or very small at any particular time. If the process is stationary, the values of the component functions will have the same distribution at any time.

As previously noted, Gaussian noise processes are very common in physical applications. They can be generated by the superposition of a large number of time functions with random time origins. An example is the *shot noise* generated in an electron tube. The random times of arrival of electrons at the plate produce the shot noise fluctuations in the plate current, which has the properties of Gaussian noise. A mathematical example of Gaussian noise is produced by the superposition of a large number of sinusoids of different frequencies and random phases.

Also very useful and important is the joint probability density function of values of the process at two different times. For a stationary Gaussian process with zero mean, this joint probability density will have the following form.

Second-order probability density function of a Gaussian noise process

$$= \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp \left[\frac{-x_1^2 + 2\rho x_1 x_2 - x_2^2}{2\sigma^2(1-\rho^2)} \right] \quad (5-23)$$

In this expression x_1 and x_2 are values of the noise process at times t_1 and t_2 , σ^2 is the variance of x_1 and x_2 , and ρ is a factor indicating the degree of correlation between x_1 and x_2 . This factor is called the *normalized autocorrelation function*. It is defined in this case, where the mean is zero, as the average value of the product x_1x_2 divided by the average value of x^2 which normalizes it so that its range is from $+1$ to -1 . When t_1 and t_2 are close together so that x_1 and x_2 have about the same values, the value of ρ will be close to unity, indicating a high degree of correlation. That is, when x_1 is high, x_2 is also very likely high; and when x_1 is low, x_2 is probably low. On the other hand, when t_1 and t_2 are sufficiently far apart for several oscillations of the noise functions to occur between them, x_1 and x_2 will tend to be uncorrelated and ρ will be close to zero. When the process is stationary, the autocorrelation function will be independent of the particular times t_1 and t_2 and depend only upon their difference, which we denote by $\tau = t_1 - t_2$.

The significant and meaningful attributes of noise processes must be expressed as average values. The notation we shall adopt to indicate the average value of some function of the process is to simply bar that function. Thus the average value of the process itself is denoted by \bar{x} . If the process represents voltages or currents, then \bar{x} can be interpreted as the d-c level. For the process whose probability density is given by Equation 5-22, the average value corresponding to the d-c is zero. The mean square value of the process about the mean or the variance can similarly be regarded as the average power in a unit resistance. As previously noted, this quantity is denoted by σ^2 .

$$\sigma^2 = \overline{(x - \bar{x})^2} = \overline{x^2} - \bar{x}^2. \quad (5-24)$$

Actually, the term *power* will often be used very generally to refer to the square of arbitrarily measured variables so that sometimes it cannot be identified with physical power, although the electrical terminology has been retained. As an example, suppose that an angle θ is found to be oscillating with an amplitude A and a frequency ω or $\theta = A \cos \omega t$. In this case, we might say that the angle θ has a power of $A^2/2$ although the dimensions of this quantity are certainly not watts.

The average value of the product x_1x_2 is very significant in signal and noise studies. This quantity is called the *autocorrelation function*, and we shall denote it by $\varphi(\tau)$, where τ is the time difference $t_2 - t_1$. For a stationary process, x_1 and x_2 are uncorrelated when τ is very large except for the average value or d-c component:

$$\varphi(\pm\infty) = \bar{x}^2. \quad (5-25)$$

When $\tau = 0$, the autocorrelation function simply equals the average value of x^2 . Thus, the variance of the process is given by

$$\sigma^2 = \varphi(0) - \varphi(\infty). \quad (5-26)$$

The normalized autocorrelation function can be defined in terms of the function $\varphi(\tau)$ by subtracting the d-c term and dividing by the variance:

$$\rho(\tau) = \frac{\varphi(\tau) - \varphi(\infty)}{\varphi(0) - \varphi(\infty)}. \quad (5-27)$$

For a Gaussian process with the joint probability density given in Equation 5-28, the autocorrelation would be computed in the following manner:

$$\begin{aligned} \overline{x_1 x_2} = \varphi(\tau) &= \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 \\ &\exp\left[\frac{-x_1^2 + 2\rho x_1 x_2 - x_2^2}{2\sigma^2(1-\rho^2)}\right] dx_1 dx_2 \quad (5-28) \\ &= \sigma^2 \rho(\tau). \end{aligned}$$

This integral can be evaluated by completing the square in the exponent for one of the variables and transforming to a standard form. In the next paragraph, it will be shown that the autocorrelation function is very closely related to the power spectrum of the process.

5-5 THE POWER DENSITY SPECTRUM

It is possible to decompose random processes into frequency components in a certain sense, and this will provide a powerful analytic technique. For instance, it was previously mentioned that a Gaussian random process could be constructed by the superposition of a large number of sinusoids of varying frequency and random phase. This sort of a process can certainly be decomposed into frequency components. Of course, the average values of the in-phase and quadrature components at a given frequency will be zero because of the introduction of a random phase angle. The power at a given frequency, though, will be independent of phase and in general have a non-zero value. Thus a frequency decomposition could be carried out on a power basis. This possibility turns out to be valid for more general random processes and leads to the useful concept of the *power density spectrum*. Physically, the power density spectrum of a noise process corresponds to the average power outputs of a bank of narrow filters covering the frequency range of the process.

To develop this idea, consider a stationary random process $x(t)$. Subject to the restrictions noted in Paragraph 5-2, the portions of the elements of the process between $-T$ and T possess Fourier transform spectra.

By limiting the range we can ensure that the integrals of the squares of the elements of the process are finite. Over an infinite range these integrals

would not be finite and Fourier transforms could not be defined. Thus we have the spectra $X_T(\omega)$:

$$X_T(\omega) = \int_{-T}^T x(t) e^{-j\omega t} dt. \quad (5-29)$$

Energy spectra will be given by expressions similar to Equation 5-6. If these energy spectra are divided by the observation time $2T$, power spectra will be obtained which we denote by $N_T(\omega)$:

$$N_T(\omega) = \frac{1}{2T} |X_T(\omega)|^2 = \frac{1}{2T} \int_{-T}^T e^{-j\omega\tau} d\tau \int_R x(t + \tau)x(t) dt. \quad (5-30)$$

The range of the last integral has been denoted by R . Because the elements of the process $x(t)$ are in effect zero for $|t| > T$, the limits of integration will be from $-T + \tau$ to T for $\tau > 0$ and from $-T$ to $T + \tau$ for $\tau < 0$. In either case, the total range is $2T - |\tau|$.

We are, of course, primarily interested in the statistical average of the power spectrum since only average values represent meaningful and measurable attributes of the process. To compute the average value of $N_T(\omega)$, we average the product $x(t + \tau)x(t)$ in the expression for $N_T(\omega)$ given by Equation 5-30. The average of this product is the autocorrelation function of the process which will depend only upon the time difference τ if the process is stationary:

$$\begin{aligned} \overline{N_T(\omega)} &= \frac{1}{2T} \int_{-\infty}^{\infty} e^{-j\omega\tau} d\tau \int_{R=2T-|\tau|}^{\infty} \varphi(\tau) dt \\ &= \int_{-\infty}^{\infty} e^{-j\omega\tau} \varphi(\tau) \left(1 - \frac{|\tau|}{2T}\right) d\tau, \quad 2T > |\tau|. \end{aligned} \quad (5-31)$$

Letting $T \rightarrow \infty$, the factor involving T in the integrand approaches unity, and we obtain the following expression for the average power density spectrum of the process:

$$\overline{N_T(\omega)} = \int_{-\infty}^{\infty} \varphi(\tau) e^{-j\omega\tau} d\tau. \quad (5-32)$$

This expression gives the power density spectrum as the Fourier transform of the autocorrelation function. These two functions form a Fourier transform pair and the knowledge of one is, at least in theory, equivalent to a knowledge of the other. The inverse of the relation in Equation 5-32 gives

$$\varphi(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \overline{N(\omega)} e^{j\omega\tau} d\omega. \quad (5-33)$$

When τ is set equal to zero in this relation

$$\varphi(0) = \sigma^2 + \bar{x}^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \overline{N(\omega)} d\omega. \quad (5-34)$$

Thus, the noise power or mean square value is equal to the sum of the power components at all frequencies. Equation 5-34 can be regarded as a generalization of Parseval's equality given in Equation 5-8.

At the end of Paragraph 5-2 it was pointed out that the absolute square of the transfer function of a network acts as a transfer function relating the input and output energy spectra. We have just defined power density spectra as the average of the energy spectra of the elements of the process divided by the observation time to give power. Thus the same relation must hold between the input and output power density spectra, $N_i(\omega)$ and $N_o(\omega)$, of a noise process being transmitted through a network with a transfer function $Y(\omega)$:

$$\overline{N_o(\omega)} = |Y(\omega)|^2 \overline{N_i(\omega)}. \quad (5-35)$$

We might note at this time that it is normal practice not to use a bar to indicate specifically that the power density spectrum of a noise process is an average value unless the averaging takes place explicitly in the derivation of the power spectrum. Thus, the power density spectrum of a noise process would normally be denoted by $\overline{N(\omega)}$ rather than $\overline{N(\omega)}$.

In order to illustrate some of these ideas, we shall make up a noise process and compute its power spectrum and autocorrelation function. We suppose the process to be composed of the sum of identical functions $h(t)$ which occur at random times. Initially, we consider only functions which originate in the finite range $-T$ to $+T$. We denote the average density of these functions by γ and suppose that there are $2T\gamma = n$ functions in the finite range of interest. Denoting the origin of the k -th function by t_k , our approximation to a random process is given by the following expression.

$$f_n(t) = \sum_1^n h(t - t_k). \quad (5-36)$$

Denoting the Fourier transform of $h(t)$ by $H(\omega)$, the Fourier transform of $f_n(t)$ is given by

$$F_n(\omega) = \int_{-\infty}^{\infty} h(t - t_k) e^{-j\omega t} dt = H(\omega) \sum_1^n e^{-j\omega t_k}. \quad (5-37)$$

The power spectrum is simply the absolute square of $F(\omega)$ divided by the observation time:

$$\begin{aligned} \frac{1}{2T} |F_n(\omega)|^2 &= \frac{1}{2T} [H(\omega) \sum_1^n e^{-j\omega t_k}] [H^*(\omega) \sum_1^n e^{j\omega t_l}] \\ &= \frac{1}{2T} |H(\omega)|^2 \sum_1^n \sum_1^n e^{j\omega(t_k - t_l)}. \end{aligned} \quad (5-38)$$

To find the average power spectrum, we must average over each variable t_k supposing it to be uniformly distributed between $-T$ and $+T$:

$$\frac{1}{2T} \overline{|F_n(\omega)|^2} = \frac{1}{2T} |H(\omega)|^2 \frac{1}{(2T)^n} \int_{-T}^T \dots \int_{-T}^T \sum_1^n \sum_1^n e^{-j\omega(t_k - t_l)} dt_1 \dots dt_n. \quad (5-39)$$

The integrals of the terms in this sum will have two forms, depending upon whether $k = l$ or not. When $k = l$, the average value of each term is unity. There are n such terms. When $k \neq l$, the average value of each term is $(\sin \omega T / \omega T)^2$. There are $n(n-1)$ of these terms. Thus, the average power spectrum has the following form:

$$\frac{1}{2T} \overline{|F_n(\omega)|^2} = |H(\omega)|^2 \left[\frac{n}{2T} + \frac{n(n-1)}{(2T)^2} \left(\frac{2}{T} \right) \left(\frac{\sin \omega T}{\omega} \right)^2 \right]. \quad (5-40)$$

As $T \rightarrow \infty$, we note that the term involving the factor $\sin^2 \omega T$ is of the same form as the definition of an impulse function given by Equation 5-21. The power density spectrum over all time, then, will have the following form:

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \overline{|F(\omega)|^2} = |H(\omega)|^2 [\gamma + 2\pi\gamma^2 \delta(\omega)]. \quad (5-41)$$

The singular part of this spectrum corresponds to a concentration of power at zero frequency or the d-c component. If $h(t)$ has no such d-c component, then $H(0)$ will be zero and the impulse has no significance. The continuous portion of the power spectrum is seen to be proportional to the energy spectrum of $h(t)$. The remarkable thing about this is that the form of the spectrum is independent of the average number of functions per unit time γ .

As a concrete illustration, suppose that $h(t)$ is given by the decaying exponential defined in Equation 5-3. An element of such a noise process might then look like the example shown in Fig. 5-6. The energy spectrum of the exponential function has already been computed in Equation 5-10.

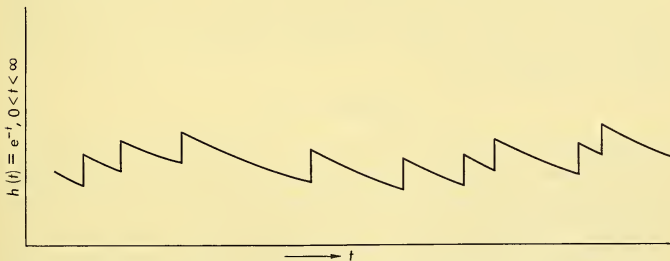


FIG. 5-6 Element of a Noise Process Composed of Identical Exponential Functions with Random Time Origins.

The power density spectrum of this process will thus have the following form:

$$N(\omega) = \frac{1}{1 + \omega^2} [\gamma + 2\pi\gamma^2\delta(\omega)]. \quad (5-42)$$

The autocorrelation function, which is just the Fourier transform of the power spectrum, has already been partially computed in Equation 5-9 and will have the form

$$\varphi(\tau) = \frac{\gamma}{2} e^{-|\tau|} + \gamma^2. \quad (5-43)$$

We may note that since we have used an $h(t)$ corresponding to the impulse response of the RC filter pictured in Fig. 5-1, the noise process that has been defined can be generated approximately by short pulses occurring at random times which are modified by this filter.

A physical interpretation of our model of a noise process is provided by shot noise, fluctuations in the number of electrons arriving at the plate of a vacuum tube per unit time. We shall use our model to show that the mean square fluctuation in electron current, ΔI^2 , incident to the shot effect is given by

$$\overline{(\Delta I)^2} = 2\epsilon I \Delta F \quad (5-44)$$

where $\epsilon =$ electronic charge

$I =$ average current (d-c)

$\Delta F =$ observation bandwidth.

We suppose that each function $h(t - t_k)$ in the sum in Equation 5-36 represents the arrival of one electron at the plate. In this case, the integral of $h(t)$ should equal the electronic charge ϵ , and we assume this, or what is equivalent, that $H(0) = \epsilon$. The magnitude of both the square of the direct current and average of the square of the fluctuation or noise currents can be determined from Equation 5-41. The square of the direct current corresponds to the magnitude of the impulse function at zero frequency in that expression and is given by

$$I^2 = |H(0)|^2 \gamma^2 = \epsilon^2 \gamma^2. \quad (5-45)$$

The mean square value of the noise currents corresponds to the integral of the nonsingular term $|H(\omega)|^2 \gamma$, in Equation 5-41. We are unable to determine this exactly without knowing the form of the spectrum of a current pulse, $H(\omega)$. If, however, we are interested in the output of a filter which is narrow compared to $H(\omega)$ we can approximate the mean square current in the output of the narrow filter by the product of twice the filter bandwidth $2\Delta F$ and the low-frequency power density of the electronic pulse

power spectrum. The factor 2 is introduced to account for contributions from negative frequencies. Forming this product and substituting I/ϵ for γ in Equation 5-41, yields the following expression for the mean square noise current:

$$\overline{(\Delta I)^2} = |H(0)|^2 \gamma 2\Delta F = 2\epsilon^2 I \Delta F / \epsilon = 2\epsilon I \Delta F. \quad (5-46)$$

Comparison with Equation 5-44 indicates that the noise process model used does indeed give the correct expression for shot noise. The forms of the functions $h(t)$ are not significant in this derivation as long as their spectra are wide compared with ΔF . Similar discussions can be made in connection with many physical phenomena which generate noise by means of some random mechanism.

5-6 NONLINEAR AND TIME-DEPENDENT OPERATIONS

In tracing signals and noise through radar systems, we find that the operations of many components are either nonlinear or time-dependent. Examples of such operations are rectification by second detectors, automatic gain control, time and frequency discrimination, phase demodulation, and sampling or gating. In this paragraph, procedures which can be used in the analysis of such operations will be discussed briefly and illustrated with a few examples.

A basic case is provided by a nonlinear device which has no energy storing capacity; that is, it is assumed to operate instantaneously. We suppose that the input to this device is a Gaussian noise process denoted by x ; the output noise process is denoted by y . The functional relation between these processes is denoted by

$$y = f(x) \quad (5-47)$$

The process y will be random but not in general Gaussian. The average values of y and y^2 can be found as the weighted averages of $f(x)$ and $f^2(x)$:

$$\bar{y} = \overline{f(x)} = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} f(x) e^{-x^2/2\sigma^2} dx \quad (5-48)$$

$$\overline{y^2} = \sigma_y^2 + \bar{y}^2 = \overline{f^2(x)} = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} f^2(x) e^{-x^2/2\sigma^2} dx. \quad (5-49)$$

The power spectrum of the y process can be found by first finding its autocorrelation function and then computing the Fourier transform of this function. The autocorrelation of y is the average value of the product $y_1 y_2 = f(x_1) f(x_2)$. The average will have to be computed relative to the joint Gaussian probability density function expressed by Equation 5-23. If this probability density function is denoted by $P_2(x_1, x_2)$, the autocorrelation of y is given by

$$\varphi(\tau) = \overline{y_1 y_2} = \iint f(x_1) f(x_2) P_2(x_1, x_2) dx_1 dx_2. \quad (5-50)$$

The power spectrum of y is simply the Fourier transform of $\varphi(\tau)$.

A Square Law Device. As a specific example, suppose that the nonlinear operation is provided by a square law device:

$$y = x^2 \quad (5-51)$$

This type of nonlinearity is often assumed to approximate the rectifying action of second detectors in radar receivers. The mean and variance of y are found by carrying out the operations indicated in Equations 5-48 and 5-49:

$$\begin{aligned} \bar{y} &= \overline{x^2} = \sigma^2 \\ \overline{y^2} &= \overline{x^4} = 3\sigma^4 \\ \sigma_y^2 &= \overline{x^4} - (\overline{x^2})^2 = 2\sigma^4. \end{aligned} \quad (5-52)$$

The autocorrelation function is found by evaluating the following integral:

$$\begin{aligned} \varphi(\tau) = \overline{y_1 y_2} &= \frac{1}{2\pi\sqrt{1-\rho^2}\sigma^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1^2 x_2^2 \\ &\exp\left[\frac{-x_1^2 + 2\rho x_1 x_2 - x_2^2}{2\sigma^2(1-\rho^2)}\right] dx_1 dx_2 \\ &= \sigma^4(1 + 2\rho^2). \end{aligned} \quad (5-53)$$

This integral is evaluated by completing the square of one of the variables in the exponent and transforming to standard forms. The constant term in $\varphi(\tau)$ corresponds to the square of the average value of y and will contribute an impulse function at zero frequency to the power spectrum of y .

In general, the squaring operation will provide a widening of the continuous noise spectrum as the various frequency components beat with themselves to produce sum and difference frequencies. To show this and to illustrate this type of analysis generally, suppose the x process is similar to the one defined in Paragraph 5-5 (Fig. 5-6) by a sum of exponential functions. For simplicity, we assume that on the average only half of the exponential functions are positive while the other half are negative, so that the average value of the x process is zero. We assume further that the variance is unity. The power spectrum and autocorrelation of the x process will be given by Equations 5-42 and 5-43. There will be no d-c term, and in order to have unit variance $\gamma = 2$:

$$\varphi(\tau) = \sigma^2 \rho(\tau) = e^{-|\tau|} \quad (5-54)$$

$$N(\omega) = \frac{2}{1 + \omega^2}. \quad (5-55)$$

From Equation 5-50, the autocorrelation function of the y process will be

$$\varphi_y(\tau) = \overline{y_1 y_2} = 1 + 2e^{-2|\tau|}. \quad (5-56)$$

The Fourier transform of this expression gives the power spectrum of the y process:

$$N_y(\omega) = 2\pi\delta(\omega) + \frac{4}{4 + \omega^2}. \quad (5-57)$$

Thus, in this case, the form of the continuous spectrum remained the same, but the bandwidth was doubled.

Another case which is very common in radar applications corresponds to the assumption of a uniform spectrum of finite bandwidth for the x process. Such an assumption normally represents a simplifying approximation to the more complicated forms which actual spectra might take. Such a

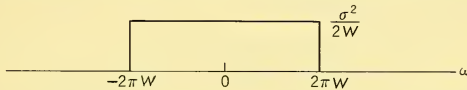


FIG. 5-7a Uniform Spectrum (x Process).

spectrum is shown in Fig. 5-7a. The autocorrelation function corresponding to this spectrum will be

$$\varphi(\tau) = \frac{\sigma^2}{2W} \frac{1}{2\pi} \int_{-2\pi W}^{2\pi W} e^{j\omega\tau} d\omega = \sigma^2 \frac{\sin 2\pi W\tau}{2\pi W\tau}. \quad (5-58)$$

The autocorrelation function of the y process will now be

$$\varphi_y(\tau) = \overline{y_1 y_2} = \sigma^4 + 2\sigma^4 \left(\frac{\sin 2\pi W\tau}{2\pi W\tau} \right)^2. \quad (5-59)$$

At the end of Paragraph 5-3 it was indicated that the Fourier transform of a triangular function is of the same form as the trigonometric term above. Thus the continuous part of the power spectrum of y will be triangular. This spectrum is pictured in Fig. 5-7b and it is represented symbolically by

$$N_y(\omega) = 2\pi\sigma^4\delta(\omega) + (\sigma^4/W)(1 - |\omega|/4\pi W), \quad |\omega| < 4\pi W. \quad (5-60)$$

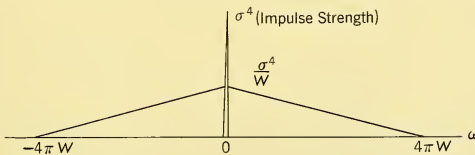


FIG. 5-7b Triangular Spectrum ($y = x^2$ Process).

An application of this result is made in Paragraph 5-7 in course of a discussion of the effect of the second detector in a pulse radar.

A Synchronous Detector. Another example which is of interest is that of a product demodulator or synchronous detector. Such a device or an approximation to such a device is a common component in many types of radar systems. It will provide an example of a time-dependent operator. In operation, a *product demodulator* simply multiplies the signal or noise by a sinusoid. Thus if the input is $x(t)$, the output would be $x(t) \cos \omega_m t$. When $x(t)$ possesses a component at the angular frequency ω_m , the dc in the output gives a measure of the phase between the input component and the reference. We again assume that x is a Gaussian noise process with zero mean, autocorrelation $\varphi(\tau)$, and power density spectrum $N(\omega)$. The autocorrelation of the output is given by

$$\begin{aligned} y_1 y_2 &= \overline{x_1 x_2} \cos \omega_m t \cos \omega_m (t + \tau) \\ &= \left(\frac{1}{2}\right) \varphi(\tau) [\cos \omega_m \tau + \cos \omega_m (2t + \tau)]. \end{aligned} \quad (5-61)$$

The autocorrelation of the output evidently varies with time periodically at the angular frequency $2\omega_m$. The spectrum of the output will likewise vary periodically. In most cases, however, the angular frequency $2\omega_m$ is outside the range of practical interest, and we can use the time average of the autocorrelation or spectrum for our purposes. On taking the time average, the periodic component disappears:

$$\overline{\varphi_y(\tau)} = \overline{y_1 y_2} = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \overline{y_1 y_2} dt = \left(\frac{1}{2}\right) \varphi(\tau) \cos \omega_m \tau. \quad (5-62)$$

The wavy bar is used to indicate a time average. Bearing in mind that the autocorrelation function and power density spectrum $\varphi(\tau)$ and $N(\omega)$ of the input noise are Fourier transforms, the Fourier transform of the expression above is easily computed to give the output power density spectrum in terms of that of the input:

$$\begin{aligned} N_y(\omega) &= \frac{1}{2} \int_{-\infty}^{\infty} \varphi(\tau) \cos \omega_m \tau e^{-j\omega\tau} d\tau \\ &= \frac{1}{4} \int_{-\infty}^{\infty} \varphi(\tau) [e^{-j(\omega-\omega_m)\tau} + e^{-j(\omega+\omega_m)\tau}] d\tau \\ &= \frac{1}{4} [N(\omega - \omega_m) + N(\omega + \omega_m)]. \end{aligned} \quad (5-63)$$

A product demodulation, then, operates to shift the input power density spectrum $N(\omega)$ into sidebands about the modulating frequency ω_m and the image of the modulating frequency $-\omega_m$.

A Clamping Circuit. *Clamping circuits*, sometimes called *pulse stretchers* or *boxcar detectors*, are another common component of radar systems. They also provide an example of an operation with a periodic time dependency. Such a circuit clamps the output to a sampled value of the input for a fixed period of time; at the end of this period, the output is clamped to a new value of the input. The operation of such a circuit is shown in Fig. 5-8. Symbolically, the output of this device can be repre-

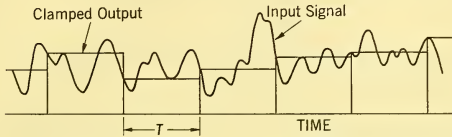


FIG. 5-8 Operation of a Clamping Circuit.

sented by

$$y(t) = x(t_k), t_k < t < t_{k+1} = t_k + T. \quad (5-64)$$

Clearly the autocorrelation of $y(t)$ is dependent upon time. As with the case of the product demodulator, however, the time average of the autocorrelation function and power density spectrum yield results which can be used for almost all applications. To determine the average autocorrelation of $y(t)$, consider that when the delay $t_2 - t_1 = \tau$, used in computing the autocorrelation functions, is a multiple of the sampling interval T , the average value of the product $y_1 y_2$ of the sampled and stretched process must be the same as the average value of the product $x_1 x_2$ because at the sample points $y_1 = x_1$ and $y_2 = x_2$. Thus, for $\tau = kT$,

$$\varphi_y(kT) = \varphi(kT). \quad (5-65)$$

When the time delay is intermediate between these isolated points, say $kT < t < (k+1)T$, the autocorrelation function of y will sometimes be $\varphi(kT)$ and sometimes $\varphi(kT + T)$ depending upon the value of t . The fraction of the time during which $\varphi_y(\tau)$ takes one of the other of these values is proportional to the relative values of $\tau - kT$ and $(k+1)T - \tau$. Thus, the average value of $\varphi_y(\tau)$ should vary linearly between its values at the discrete points where $\tau = kT$, and it will be composed of these points connected by straight lines.

A limiting case of special interest occurs when the sampling frequency is much smaller than the width of the input spectrum. In this case, the autocorrelation function of the input is narrow compared with the sampling period. That is, values of the process which are separated by more than the sampling period are very nearly independent. Since in this case

$\phi_y(T) \approx 0$, the autocorrelation of the output is very nearly a triangle as is indicated in Fig. 5-9a. The Fourier transform of a triangular function has

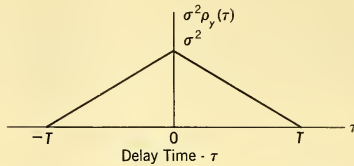


FIG. 5-9a Autocorrelation Function of Pulse Stretcher Output with Wide-Band Noise Input.

already been determined in Equation 5-21. On using that result, the power spectrum of the stretched process will be

$$N_y(\omega) = \sigma^2 T \left(\frac{\sin \omega T/2}{\omega T/2} \right)^2. \quad (5-66)$$

The autocorrelation and the power spectrum of the pulse stretcher output in this case are both shown in Fig. 5-9. We might note that one of the basic features of this sort of operation is to concentrate the noise in a wide input spectrum in a low-frequency spectrum of width approximately $1/T$ cps.

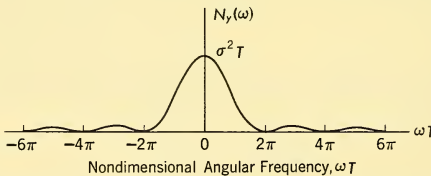


FIG. 5-9b Power Density Spectrum of Pulse Stretcher Output with Wide-Band Noise Input.

5-7 NARROW BAND NOISE

Signals in radar systems normally have the form of a radio-frequency carrier modulated by a low-frequency envelope which contains the essential intelligence. Such signals are filtered and amplified by tuned circuits with bandwidths just sufficient to pass the modulation sidebands. Noise associated with signals of this form or originating in circuits designed to amplify such signals will have a narrow spectrum centered about the carrier. In this paragraph, we shall develop some of the properties of narrow band noise and signal plus noise.

We suppose that the noise power is concentrated in the neighborhood of a carrier frequency ω_c . Such a noise process can be constructed by modulating a relatively low-frequency noise process by the carrier frequency. The carrier frequency signal can be represented by either the in-phase or quadrature component, and, in general, the narrow band noise will be composed of both components. Denoting the low-frequency noise processes corresponding to the in-phase and quadrature components about the carrier by $x(t)$ and $y(t)$, the narrow band noise process denoted by $z(t)$ can be represented by

$$z(t) = x(t) \cos \omega_c t + y(t) \sin \omega_c t. \quad (5-67)$$

In general, $x(t)$ and $y(t)$ could be correlated and also might have dissimilar features. But in most problems of practical interest they will be independent and have identical spectra and other statistical characteristics. If the x and y processes did not have the same spectra and autocorrelation functions, the narrow band process would depend upon time, as is apparent in Equation 5-68 below. Requiring x and y to be independent makes the spectrum of the narrow band process symmetrical about the carrier frequency ω_c . We assume that x and y are independent and have identical spectra. The autocorrelation function of the z process is computed as follows:

$$\begin{aligned} \varphi_z(\tau) &= \overline{[x_1 \cos \omega_c t_1 + y_1 \sin \omega_c t_1][x_2 \cos \omega_c t_2 + y_2 \sin \omega_c t_2]} \\ &= \left(\frac{1}{2}\right) \overline{x_1 x_2} [\cos \omega_c \tau + \cos \omega_c (2t + \tau)] \\ &\quad + \left(\frac{1}{2}\right) \overline{y_1 y_2} [\cos \omega_c \tau - \cos \omega_c (2t + \tau)] \\ &\quad + \left(\frac{1}{2}\right) \overline{x_1 y_2} [\sin \omega_c \tau + \sin \omega_c (2t + \tau)] \\ &\quad - \left(\frac{1}{2}\right) \overline{y_1 x_2} [\sin \omega_c \tau - \sin \omega_c (2t + \tau)] \\ &= \varphi(\tau) \cos \omega_c \tau \end{aligned} \quad (5-68)$$

where $\varphi(\tau)$ denotes the autocorrelation function of the x and y processes. The autocorrelation function $\varphi_z(\tau)$ is of exactly the same form as that of the output of a product demodulator discussed in the preceding paragraph and given in Equation 5-62. Thus the Fourier transform of $\varphi_z(\tau)$ giving the power spectrum of the z process will be related to the spectrum of the x and y processes, $N(\omega)$, in a manner similar to that indicated in Equation 5-63:

$$N_z(\omega) = \left(\frac{1}{2}\right)[N(\omega - \omega_c) + N(\omega + \omega_c)]. \quad (5-69)$$

From this expression, we see that the spectrum of narrow band symmetric noise has the same form as the low-frequency modulating functions, but is shifted to the vicinity of the carrier frequency.

In a large class of radar systems, the transmitted and received signals have the form of an RF carrier amplitude modulated by a low-frequency waveform. In the majority of these systems, the modulation consists of a

periodic pulsing of the carrier. In such systems, the signal, when it is present, is a constant amplitude sinusoid. Noise will normally be present with a spectrum centered about the carrier frequency and a width determined by the amplifiers which are designed to transmit the modulation sidebands. The ratio of the bandwidth to the carrier frequency is normally very small. Thus the noise can be considered narrow band noise with the representation and characteristics described above.

When a signal is present, it is assumed to be of the form

$$\text{Signal} = a \cos \omega_c t. \quad (5-70)$$

The peak signal power is denoted by $S = a^2/2$. The noise power is denoted by $N = \sigma^2$ so that the signal-to-noise power ratio when the signal is present is

$$S/N = a^2/2\sigma^2. \quad (5-71)$$

With the narrow band noise represented as in Equation 5-67, the signal plus noise has the form

$$\text{Signal plus noise} = (a + x) \cos \omega_c t + y \sin \omega_c t. \quad (5-72)$$

A typical power spectrum of a c-w signal plus noise is shown in Fig. 5-10.

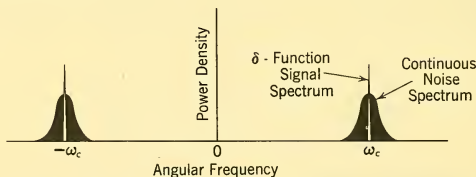


FIG. 5-10 Power Density Spectrum of CW Signal Plus Narrow-Band Noise.

General operations upon radar signals to extract desired information or to transform the signals into a more useable form are often referred to as *demodulation* or *detection* operations as discussed in Chapter 1. The simplest and most common such operation consists in the generation of the envelope of a narrow band signal by means of a rectifier. In superheterodyne receivers, this operation corresponds to the action of the second detector. The envelope output of the second detector is most often referred to as the *video signal* since it is commonly used as an input of some sort of visual display. In the following brief analysis, we shall develop some of the more important features of video signals and noise.

The envelope of narrow band signal plus noise can be exhibited by rewriting the expression in Equation 5-72 in the following form:

$$\text{Signal plus noise} = \sqrt{(a+x)^2 + y^2} \cos \left[\omega_c t + \tan^{-1} \frac{y}{a+x} \right]. \quad (5-73)$$

Here an envelope function modulates a carrier frequency with random phase modulation depending upon x and y . We note in passing that a frequency discriminator would be sensitive to this phase modulation and that studies similar to those which we shall make of the video envelope can also be made of a discriminator output.

We first determine the probability density function of the envelope which is denoted by r :

$$\text{Envelope} = r = \sqrt{(a+x)^2 + y^2}. \quad (5-74)$$

The random variables x and y are assumed to represent independent Gaussian noise processes with zero means and equal variances. The differential probability that they will be found in the differential area $dx dy$ is given by their joint probability times this differential area:

$$dp = P_1(x)P_1(y)dx dy = \frac{1}{2\pi\sigma^2} \exp \left[\frac{-x^2 - y^2}{2\sigma^2} \right] dx dy. \quad (5-75)$$

In order to determine the probability density function of the video envelope, this expression will be transformed to polar coordinates and the average value for all angles found. This transformation is represented as follows:

$$\begin{aligned} a+x &= r \cos \theta \\ y &= r \sin \theta \\ dx dy &= r dr d\theta. \end{aligned} \quad (5-76)$$

Substituting these relations into the expression in Equation 5-75 and integrating over the variable θ gives

$$dp = P_1(r)dr = \frac{1}{\sigma^2} \exp \left[\frac{-r^2 - a^2}{2\sigma^2} \right] r dr \frac{1}{2\pi} \int_{D_0}^{2\pi} \exp [ar \cos \theta / \sigma^2] d\theta. \quad (5-77)$$

The integral in this expression can be recognized as a representation of a zero-order Bessel function of the first kind with imaginary argument⁶ denoted by $I_0(ar/\sigma^2)$. The probability density function of r is thus of the following form:

$$P_1(r) = \frac{r}{\sigma^2} \exp \left[\frac{-r^2 - a^2}{2\sigma^2} \right] I_0(ar/\sigma^2). \quad (5-78)$$

A curve showing $P_1(r)$ for some representative values of S/N is given in Fig. 5-11. In the two extremes of very small and very large values of the signal-to-noise ratio, $P_1(r)$ approaches the following forms:

⁶J. L. Lawson and G. E. Uhlenbeck, *op. cit.*, p. 173.

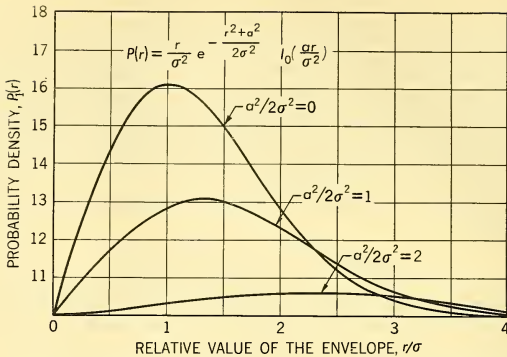


FIG. 5-11 Probability Density Functions of the Envelope of Narrow-Band Signal Plus Noise.

$$P_1(r) \approx \frac{r}{\sigma^2} \exp\left[-\frac{r^2}{2\sigma^2}\right], \quad a^2/2\sigma^2 \ll 1 \quad (5-79)$$

$$P_1(r) \approx \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(r-a)^2}{2\sigma^2}\right], \quad a^2/2\sigma^2 \gg 1. \quad (5-80)$$

The first of these forms is often called a *Rayleigh probability density* and corresponds to the case of noise alone. When the signal-to-noise ratio is large, the envelope has approximately a normal distribution as is indicated by Equation 5-80.

As might be expected from the form of the probability density function of r , its basic statistical properties such as its autocorrelation or spectrum cannot be expressed simply in terms of elementary functions. Approximate expressions valid for either large or small values of the signal-to-noise ratio have been developed.⁷ Instead of becoming involved with such approximations, however, it is often either more convenient analytically or more realistic in a physical sense to assume that the second detector is a square law rectifier producing the square of the envelope rather than the envelope itself. In most problems where such an assumption is made, the variations of many phenomena with parameters of interest are relatively independent of the detector law. The statistical properties of the square of the envelope can be expressed in much simpler forms than those of the envelope itself because r^2 is a simple second-degree polynomial function of x and y . Thus, the autocorrelation function of r^2 will involve the average values of products of the form $x_1^2 x_2^2$ and $y_1^2 y_2^2$ which have already been evaluated in Para-

⁷*Ibid.*, Chap. 7.

graph 5-6 in connection with the discussion of a square-law device. Using the results of that paragraph, the autocorrelation of r^2 is computed as follows:

$$\begin{aligned}
 \overline{r_1^2 r_2^2} &= \overline{[(a + x_1)^2 + y_1^2][(a + x_2)^2 + y_2^2]} \\
 &= a^4 + 2a^3\overline{x_1} + 2a^3\overline{x_2} + a^2\overline{x_1^2} + a^2\overline{x_2^2} + a^2\overline{y_1^2} + a^2\overline{y_2^2} + 2a\overline{x_1 x_2} \\
 &\quad + 2a\overline{x_1 y_2} + 2a\overline{x_2 y_1} + 2a\overline{x_2 y_2} + \overline{x_1^2 x_2^2} + \overline{x_1^2 y_2^2} + \overline{x_2^2 y_1^2} \\
 &\quad + \overline{y_1^2 y_2^2} + 4a^2\overline{x_1 x_2} \tag{5-81} \\
 &= (a^2 + 2\sigma^2)^2 + 4\sigma^2(\sigma^2\rho^2 + a^2\rho) \\
 &= (2\sigma^2)^2(1 + S/N)^2 + (2\sigma^2)^2[\rho^2 + 2(S/N)\rho].
 \end{aligned}$$

The spectrum of the video signal plus noise has three components: (a) an impulse at zero frequency representing the d-c, (b) a continuous portion of the same shape as the spectrum of the component x and y processes representing beats between the signal and the noise, (c) a continuous portion somewhat wider than the spectrum of the x and y processes representing beats between various parts of the noise spectrum. Fig. 5-12 illustrates the

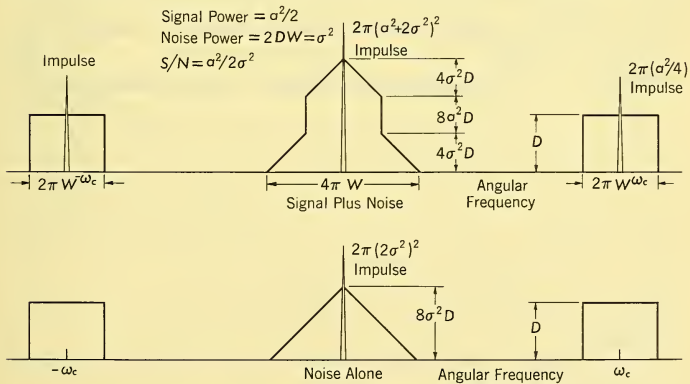


FIG. 5-12 Power Density Spectra of the Square of the Envelope of a Sinusoidal Signal Plus Narrow-Band Noise.

forms of the various spectra in a typical case. The spectrum of the x and y processes is assumed rectangular with bandwidth W . The density of the positive and negative portions of the narrow band spectrum is denoted by D . The d-c level is equal to twice the sum of the signal and noise powers. The portion of the continuous spectrum corresponding to (b) is rectangular, of half the width of the narrow band spectrum (considering only positive

frequencies), and has a power density equal to the product of 8 times the signal power and D . The portion of the continuous spectrum corresponding to (c) is triangular, with a width equal to that of the narrow band spectrum and with a power density at zero frequency equal to the product of 8 times the noise power and D .

5-8 AN APPLICATION TO THE EVALUATION OF ANGLE TRACKING NOISE

In this paragraph, the techniques developed for tracing signals and noise through radar systems will be illustrated by a discussion of the performance of an angle tracking loop in a pulse radar as a function of the signal-to-noise ratio. A block diagram showing the elements of the receiver composing this angle tracking loop is given in Fig. 5-13. This diagram represents a pulse

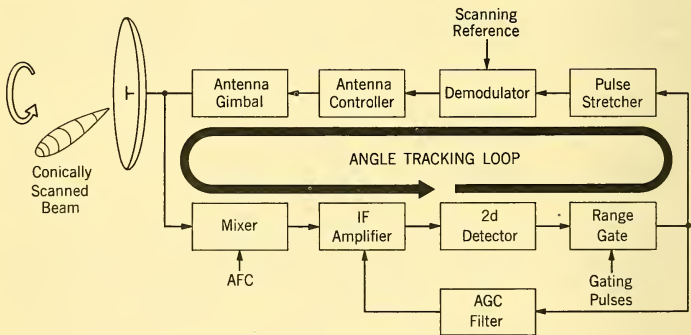


FIG. 5-13 Block Diagram of Angle Tracking Loop Employing Conical Scanning.

radar with a pencil beam which is conically scanned to generate an angular error signal.

A signal received from a target which is being tracked will have the following form:

$$\text{Received signal} = a[1 + k\epsilon \cos(\omega_s t + \varphi)] \cos \omega_c t \quad (\text{pulse modulation})$$

$$\text{where } a = \text{signal amplitude} \quad (5-82)$$

k = modulation constant of the antenna

ϵ = angular error magnitude

φ = angular error direction

ω_s = scan frequency (rad/sec)

ω_c = carrier frequency (RF or IF, rad/sec).

The angular error information is contained in amplitude modulation at the scanning frequency. We shall refer to this modulation as the *a-c error signal*. Its amplitude is proportional to the error amplitude, while its phase gives the error direction. The RF carrier of the received signal is transformed to an intermediate frequency in the mixer or first detector. The IF amplifier then provides the necessary gain and maintains the average level of the signal at a convenient constant value in response to the feedback signal from the AGC (automatic gain control) filter. The envelope of the IF signal is developed by the second detector, which is basically a rectifier. For our purposes we shall assume the second detector to be a square-law device whose characteristics have already been discussed to some extent in the preceding paragraph. A range gate selects only pulses occurring at the proper radar time for use in deriving the angle error. The range gate is positioned by an auxiliary range tracking loop which is not shown in Fig. 5-13. The AGC loop maintains the d-c value of the video signal during a pulse at a constant value so as to preserve a fixed relation between per cent modulation at the scanning frequency and angular error, independently of the received signal strength.

A pulse stretcher generates a continuous signal suitable for use in the low-frequency control circuits from the pulsed signal delivered by the range gate. The output of the pulse stretcher is delivered to a product demodulator or synchronous detector which develops a servo control signal from the a-c error.

Internally generated noise arises primarily within the mixer and the first stages of the IF amplifier. The noise may be represented exactly as in Paragraph 5-7 (Equation 5-67). That is, in-phase and quadrature components at the carrier frequency are modulated by independent low-frequency noise processes which we denote by x and y . The noise power is denoted by σ^2 so that the average signal-to-noise ratio will be

$$\text{Signal-to-noise ratio} = S/N = a^2/2\sigma^2. \quad (5-83)$$

This is an average signal-to-noise ratio because, on a short term basis, the signal power is modulated by the a-c error signal.

The signal plus noise during a pulse will be of the following form:

$$\text{Signal plus noise} = [a(1 + k\epsilon \cos(\omega_s t + \varphi)) + x] \cos \omega_c t + y \sin \omega_c t. \quad (5-84)$$

The video envelope from the square law detector during a pulse consists of the sum of the squares of the in-phase and quadrature components:

$$\begin{aligned} \text{Video signal plus noise} = r^2 = & a^2[1 + 2k\epsilon \cos(\omega_s t + \varphi) \\ & + k^2\epsilon^2 \cos^2(\omega_s t + \varphi)] \\ & + 2ax[1 + k\epsilon \cos(\omega_s t + \varphi)] \\ & + x^2 + y^2. \end{aligned} \quad (5-85)$$

In analyzing the effects of the remaining circuits in the loop on this signal, it is convenient to make certain simplifying approximations. First, it is assumed that the fractional modulation $k\epsilon$ is small enough that its square may be neglected. Second, when determining the spectrum of the video noise, the target is assumed centered in the beam so that $k\epsilon$ is zero. This reduces the expression above to the case already considered in Paragraph 5-7 (Equation 5-74).

The average value of the video signal plus noise during a pulse will consist of a d-c term and the a-c error signal:

$$\text{Average video} = \overline{r^2} \cong a^2 + 2\sigma^2 + 2a^2k\epsilon \cos(\omega_s t + \varphi). \quad (5-86)$$

The AGC loop will act to maintain the value of the d-c part of the video at a constant level which we may conveniently assume to be unity. Thus, ideally, the effect of the AGC is to divide the video by its d-c level. We assume that the AGC loop does indeed operate in this manner, although in an actual system only an approximate quotient would be formed. This assumption is sufficiently accurate for our purposes. In this case the effective a-c error signal during a pulse becomes

$$A-C \text{ error signal} = \left(\frac{S/N}{1 + S/N} \right) 2k\epsilon \cos(\omega_s t + \varphi). \quad (5-87)$$

One effect of the noise is to introduce a factor depending upon the signal-to-noise ratio which attenuates the a-c error at low values of this ratio. The net result of this suppression of the signal by the noise is to decrease the gain around the angle tracking loop.

A pulse stretcher is used to generate a signal suitable for use in the low-frequency control circuits from the pulsed signal delivered by the range gate. The pulse stretching operation will introduce some distortion of the angular error modulation, but because the scanning frequency is normally much smaller than the pulse repetition frequency, this distortion can be neglected and the pulse stretcher assumed to generate the fundamental component of the pulsed signal. Thus the a-c error signal delivered to the phase-sensitive demodulator is essentially of the form given in Equation 5-87.

We suppose the demodulator to be a simple product type consisting of a multiplication of the modulated error signal by a sinusoidal reference, $(1/k) \cos \omega_s t$. The factor $1/k$ is incorporated in order that the output may be equal to the angular error. The properties of such a device with noise inputs were established in Paragraph 5-6. The demodulator output is filtered so that only the very low frequencies are retained (components of ω_s and above are eliminated) as the angular error signal. The development of the error signal in the demodulator can be represented by the following operations:

$$\begin{aligned} \text{Error signal} &= \left(\frac{S/N}{1 + S/N} \right) \overline{[1 + 2k\epsilon \cos(\omega_s t + \varphi)](1/K) \cos \omega_s t} \\ &= \left(\frac{S/N}{1 + S/N} \right) \epsilon \cos \varphi. \end{aligned} \quad (5-88)$$

In this expression the wavy bar indicates the time average, which eliminates the fluctuating terms. The factor $\cos \varphi$ indicates that the error derived is the projection of the total error on the axis represented by one of the angle tracking loops. Complete directional control of the antenna requires it to be controlled in two directions, normally azimuth and elevation. The error signal for the other loop is obtained from a demodulator with a reference $\sin \omega_s t$.

This error represents the input to the antenna controller which moves the antenna in order to null the error and track the target. In order to arrive at a definite result in this example, we shall assume that the antenna controller is composed of a single integrator, although in a practical system the dynamic response of the angle tracking loop might be quite complicated. With this assumption, the response of the whole loop becomes the same as that of a low-pass RC filter, and the power transfer function has the following familiar form.

$$\text{Angle tracking loop power transfer function} = \frac{K^2}{\omega^2 + K^2} \quad (5-89)$$

where $K = \text{gain around the tracking loop} = \text{bandwidth (rad/sec)}$

As noted above, the gain K will be attenuated by a factor depending on the signal-to-noise ratio. Thus we shall express K as the product of this factor and a design bandwidth β achieved at high signal-to-noise ratios:

$$K = \left(\frac{S/N}{1 + S/N} \right) \beta. \quad (5-90)$$

Our primary interest in this example is to determine the response of the loop to internally generated noise. It will turn out that the spectrum of the equivalent noise input to the loop is very much broader than β and relatively flat in the low-frequency region. If we denote the power density of this input noise spectrum by D in angular units squared per rad/sec, the variance of the tracking noise will be given by

$$\text{Mean square tracking noise} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{DK^2 d\omega}{\omega^2 + K^2} = \frac{DK}{2} = \left(\frac{S/N}{1 + S/N} \right) \left(\frac{D\beta}{2} \right). \quad (5-91)$$

The next problem is to determine the magnitude D of the input-power density spectrum.

As already noted, the error is assumed zero when the spectrum of the video noise is determined in order to simplify the calculations. This corresponds to the case already considered in Paragraph 5-7. The spectrum of the video noise is thus pictured in Fig. 5-12, and its autocorrelation function is given by Equation 5-81. Dividing the noise power in the square of the envelope as determined from these sources by the square of the d-c level, to account for the effect of the AGC, gives the effective video noise power during a pulse:

$$\text{Video noise power (with AGC)} = \frac{[1 + 2(S/N)]}{(1 + S/N)^2} \quad (5-92)$$

With a pulse width normally on the order of a microsecond, the width of the IF pass band, W cps in Fig. 5-12, must be approximately 1 Mc/sec or greater. The spectrum of the video noise will also be approximately of this width with a correlation time on the order of a microsecond. The repetition rate on the other hand will normally lie in the range from a few hundred to a thousand cps. Pulses will thus be separated by at least a millisecond, and the pulse-to-pulse fluctuations due to internal noise should be very nearly independent.

The effect of the pulse stretching operation is considered next. In Paragraph 5-6 the spectrum of the output of a pulse stretcher was developed from an input of independent noise pulses. This is exactly the situation being considered in this example. Thus the spectrum of the stretched signal plus noise should have the form given by Equation 5-66 which was illustrated in Fig. 5-9. If we denote the repetition period by T , the power spectrum of the input to the demodulator will be of the following form:

$$\begin{aligned} \text{Noise spectrum of demodulator input} &= \frac{[1 + 2(S/N)]T}{(1 + S/N)^2} \left[\frac{\sin \omega T/2}{\omega T/2} \right]^2 \\ &= \frac{[1 + 2(S/N)]T}{(1 + S/N)^2} N(\omega). \end{aligned} \quad (5-93)$$

The effect of the demodulator on its input spectrum was established in Paragraph 5-6 (Equation 5-63). The demodulator input spectrum will be shifted back and forth by the demodulating frequency and multiplied by the factor $(1/4 k^2)$:

$$\begin{aligned} \text{Noise spectrum at demodulator output} &= \frac{[1 + 2(S/N)]T}{(1 + S/N)^2 k^2} [N(\omega + \omega_s) \\ &\quad + N(\omega - \omega_s)]. \end{aligned} \quad (5-94)$$

The width of each component of this spectrum is approximately $1/T$ cps, which normally might be on the order of a few hundred to a thousand cps. Since the bandwidth of the tracking loop will normally be only a few cps, only the power density in the neighborhood of zero frequency is significant;

that is, the noise spectrum may be assumed to be uniform without appreciable error:

$$\text{Power density of demodulator output noise} = D = \frac{[1 + 2(S/N)]T}{(1 + S/N)^2 k^2} N(\omega_s). \quad (5-95)$$

A further simplification can often be made when the ratio of the scanning to the repetition frequencies is small. In this case, the factor $N(\omega_s)$ is approximately unity. For example, when the ratio of these frequencies is 1 : 10, the value of $N(\omega_s)$ is 0.97.

Substituting the power density D given in Equation 5-95 into the relation already derived for the mean square tracking noise (Equation 5-91) and assuming that $N(\omega_s)$ is unity gives the following expression for the tracking noise variance:

$$\text{Mean square tracking noise} = \frac{(S/N)[1 + 2(S/N)]}{(1 + S/N)^3} \left(\frac{T\beta}{4k^2} \right). \quad (5-96)$$

This expression represents the end product of our analysis of the effect of internally generated noise on the performance of a conically scanned angle tracking loop. It is interesting that the tracking noise from this source has a maximum at a signal-to-noise ratio of 1.35 db. The decrease in tracking noise at small signal-to-noise ratios is due to the loss in loop gain and consequent narrowing of the loop bandwidth. When this begins to happen in a practical system, dynamic tracking lags usually cause an early loss of the target. We also note that the rms tracking noise is directly proportional to the square root of the repetition period and inversely proportional to the modulation constant of the antenna. This constant, expressed in per cent modulation per unit error, is itself inversely proportional to the antenna beamwidth.

The analysis in this example was intended to illustrate the sort of considerations which are appropriate to a study of noise in a radar tracking loop which incorporates a variety of components — some of them nonlinear or time dependent. Similar analyses can be made of other types of tracking systems such as monopulse tracking loops, range tracking loops, and frequency tracking loops. The effects of externally generated random disturbances such as glint or amplitude noise will also be handled in a similar fashion.

5-9 AN APPLICATION TO THE ANALYSIS OF AN MTI SYSTEM

In this paragraph we shall make some observations on the performance of a radar system which provides moving target indication (MTI). This analysis will supply another example to illustrate the use of the mathematical techniques which have been developed. The MTI system which we

shall consider is a noncoherent delay-line cancellation system. In such a system, both ground clutter and target reflections are received simultaneously. The RF carrier frequency of the ground clutter is denoted by ω_c , while that from a target moving relative to the ground will possess a doppler shift ω_d and is denoted by $\omega_c + \omega_d$. When the sum signal is detected, a beat is produced at the doppler frequency ω_d . If there is no target present, there is no doppler beat, and the spectrum of the detected video is concentrated at d-c and in the neighborhoods of harmonics of the pulse repetition frequency. The doppler signal can be separated from the clutter background by means of a delay-line cancellation unit. This unit provides the difference of successive returns as an output, that is, returns separated by the repetition period T . Fig. 5-14a shows a block diagram of such a cancellation unit, while Figure 5-14b illustrates its operation. This

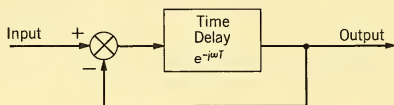


FIG. 5-14a Delay-Line Cancellation Unit.

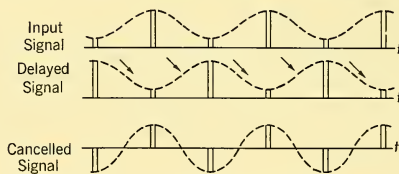


FIG. 5-14b Cancellation of Clutter Echoes.

sort of unit will attenuate the d-c component and all harmonics of the repetition frequency and in this manner cancel most of the clutter. When the doppler frequency lies between these harmonics, it will be transmitted through the cancellation unit. If by chance the doppler frequency coincides with one of the repetition rate harmonics, it will be canceled along with the clutter and produce a blind region or range of doppler frequencies to which the system is insensitive. Blind regions represent one of the most serious limitations of this type of system.

Proceeding with the analysis, the clutter echo at a given range is represented before detection as a narrow band noise process:

$$\text{Clutter echo} = x(t) \cos \omega_c t + y(t) \sin \omega_c t. \quad (5-97)$$

The modulating functions x and y are independent Gaussian noise processes with identical spectra. The clutter spectrum is determined by the motion

of ground objects, the scanning of the antenna, and the motion of the platform on which the antenna is mounted. The clutter power is denoted by $C = \overline{x^2} = \overline{y^2}$, and the autocorrelation function of the x and y processes is denoted by $C\rho = \overline{x_1x_2} = \overline{y_1y_2}$.

During a pulse, the echo received from a moving target is assumed to be a sinusoidal signal with a doppler shift:

$$\text{Target echo} = a \cos(\omega_c + \omega_d)t. \quad (5-98)$$

The peak signal power is denoted by $S = a^2/2$.

We shall assume that the signal plus clutter is rectified by a square-law second detector to give the following video signal during a pulse:

$$\text{Video} = v = [a \cos(\omega_c + \omega_d)t + x \cos \omega_c t + y \sin \omega_c t]^2. \quad (5-99)$$

The video frequencies are, of course, limited by the video bandwidth. Squaring this expression and retaining only the low-frequency components which will be passed by the video amplifier gives

$$\text{Video} = v = \frac{1}{2}(a^2 + x^2 + y^2 + 2ax \cos \omega_d t - 2ay \sin \omega_d t). \quad (5-100)$$

The cancellation unit acts to generate the difference of video signals separated by a repetition period. Denoting the residue from the cancellation unit by $r(t)$, we have

$$\begin{aligned} \text{Residue signal} = r(t) &= v(t) - v(t - T) = v_1 - v_2 \\ &= \frac{1}{2}(x_1^2 - x_2^2 + y_1^2 - y_2^2 + 2ax_1 \cos \omega_d t_1 \\ &\quad - 2ax_2 \cos \omega_d t_2 - 2ay_1 \sin \omega_d t_1 + 2ay_2 \sin \omega_d t_2). \end{aligned} \quad (5-101)$$

In order to evaluate the effect of the cancellation unit in reducing the clutter, it is convenient to define a *video signal-to-clutter ratio*. This ratio is defined as the difference between the video power with a signal v_{s+c}^2 and the video power with clutter only v_c^2 divided by this latter quantity.

$$\text{Video signal-to-clutter ratio} = (S/C)_v = (\overline{v_{s+c}^2} - \overline{v_c^2})/\overline{v_c^2}. \quad (5-102)$$

Similarly, a *signal-to-clutter ratio* is defined for the residue signal output of the cancellation unit:

$$\text{Residue signal-to-clutter ratio} = (S/N)_r = (\overline{r_{s+c}^2} - \overline{r_c^2})/\overline{r_c^2}. \quad (5-103)$$

With these definitions, a gain factor may be determined as the quotient of these two ratios:

$$\text{System gain factor} = G = \frac{(S/C)_r}{(S/C)_v} \quad (5-104)$$

In order to evaluate this gain factor in terms of the system parameters, the average values of the squares of the video and residue signals must be calculated. This is somewhat complicated because of the large number of

terms resulting from the squaring of Equations 5-100 and 5-101, and the details will not be given here. The following average values originally determined in Paragraph 5-6 in connection with a discussion of a square-law device are used in these calculations:

$$\begin{aligned}\bar{x} &= \bar{y} = \bar{x^3} = \bar{y^3} = \overline{xy^2} = \overline{x^2y} = 0 \\ \overline{x^2} &= \overline{y^2} = C, \overline{x_1x_2} = \overline{y_1y_2} = C\rho(T) \\ \overline{x^4} &= \overline{y^4} = 3C^2 = \overline{x_1^2x_2^2} = \overline{y_1^2y_2^2} = C^2[1 + 2\rho^2(T)].\end{aligned}\tag{5-105}$$

The following results were determined for the video and residue signal to clutter ratios:

$$\begin{aligned}(S/C)_v &= 2(S/C) [1 + \frac{1}{4}(S/C)] \\ (S/C)_r &= 2(S/C) \left[\frac{1 - \rho(T) \cos \omega_d T}{1 - \rho^2(T)} \right].\end{aligned}\tag{5-107}$$

The gain in signal to clutter ratio will simply be

$$\text{System gain factor} = G = \frac{1}{1 + (\frac{1}{4})(S/C)} \frac{1 - \rho(T) \cos \omega_d T}{1 - \rho^2(T)}.\tag{5-108}$$

This expression essentially summarizes the ability of a noncoherent MTI system to reduce clutter. Various interesting observations might be made from a study of this factor. For instance, the depth of the blind speed nulls at harmonics of the repetition frequency can be determined as a function of the normalized autocorrelation function of the clutter at the repetition period. The average gain over all doppler frequencies can also be found as a function of the same parameters. These details will not be explored here. The primary purpose of the example has been served by the derivation of Equation 5-108, which showed how a performance equation could be arrived at by a straightforward application of the techniques for signal and noise analysis previously developed.

5-10 AN APPLICATION TO THE ANALYSIS OF A MATCHED FILTER RADAR

In this paragraph, we shall consider how the mathematical techniques which have been developed can be applied to the derivation of optimum radar systems. Besides providing a good illustration of the application of these techniques, this example will also provide an insight into the important basic factors which affect system performance and set theoretical performance boundaries which a practical system may approach but not surpass.

We shall be primarily concerned with the detection performance of radar systems. A fundamental problem in detecting a radar target is to distin-

guish the target echo from random noise which tends to obscure it and render detection a matter of chance. This is the problem that we shall discuss in this paragraph. We shall determine the characteristics of an optimum receiver which will provide the most reliable detection of target echoes obscured by random noise.

There are several possible approaches to this problem, depending upon the generality desired, the definition of most reliable detection adopted, and various assumptions made about the signal. We shall adopt the simplest possible approach, although the receiver design criterion which will be derived is operationally equivalent to the results of more sophisticated analyses in most cases.

We suppose that in the general radar situation a signal is received as an echo from the target. During the process of reception, noise is added to the signal. The question we consider is, "What function must the receiver perform in order that the most reliable detection of the signal may be obtained?" We shall limit our study to receivers which are *linear*. That is, the effect of the receiver on the signal and noise is that of a linear filter. The output of the receiver-filter will consist of a filtered signal and filtered noise. Thus a ratio of the output signal and noise powers can be formed. We shall choose the optimum receiver-filter as that which maximizes this signal-to-noise ratio. We shall subsequently indicate how a maximum signal-to-noise ratio gives a maximum probability of detection for a fixed false-alarm rate and thus provides the most reliable detection in this sense. It will turn out, interestingly enough, that the receiver-filter which is optimum in the sense described above has a transfer function which is the conjugate of the target echo spectrum,⁸ and for this reason such a radar is often called a *matched filter* system. That is, the filter transfer function is matched to the target echo spectrum. We shall also demonstrate that such a system is equivalent to a cross correlation of the signal plus noise with an image of the signal waveform which is the origin of the term *correlation radar* sometimes used in reference to such systems.

We adopt the following notation for this analysis:

$s(t)$ = signal input to receiver-filter

$S(\omega)$ = spectrum of $s(t)$

$s_o(t)$ = signal output of receiver-filter

$S_o(\omega)$ = spectrum of $s_o(t)$

⁸This result is sometimes called the *Fourier transform* criterion and is attributed to a number of authors: namely, D. O. North, W. W. Hansen, N. Wiener, J. H. Van Vleck, and D. Middleton. See particularly Van Vleck and Middleton, "A Theoretical Comparison of Visual, Aural, and Meter Reception of Pulsed Signals in the Presence of Noise," *J. Appl. Phy.* **17**, 940-971 (1946).

$Y(\omega)$ = transfer function of receiver-filter

$n(t)$ = noise input to receiver-filter

D = power density of noise input to receiver-filter

$n_o(t)$ = noise output of receiver-filter

σ^2 = noise power in output of receiver-filter

z^2 = peak signal-to-noise power ratio in output of receiver-filter

t_o = observation time

The target echo is represented by a signal input to the receiver-filter denoted by $s(t)$ with a spectrum $S(\omega)$. The signal output of the filter and its spectrum are denoted by $s_o(t)$ and $S_o(\omega)$. The transfer function of the filter is represented by $Y(\omega)$, and the output signal spectrum is equal to the product of this transfer function and the input signal spectrum:

$$S_o(\omega) = Y(\omega) S(\omega). \quad (5-109)$$

The output waveform will, of course, be simply the inverse Fourier transform of $S_o(\omega)$:

$$\text{Output signal} = s_o(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} Y(\omega) S(\omega) e^{j\omega t} d\omega. \quad (5-110)$$

We choose to make our observation of the output at the time t_o . It is supposed that t_o is selected so that the whole of the input signal is available to the filter. The signal power in the output of the filter at the observation time will be $s_o^2(t_o)$, while the noise power in the filter output is denoted by σ^2 . The input noise is assumed to be Gaussian with a uniform or "white" spectrum with power density D . The output noise power will thus be

$$\text{Output noise power} = \sigma^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} D |Y(\omega)|^2 d\omega. \quad (5-111)$$

The output signal-to-noise ratio at the time t_o is denoted by z^2 :

$$\text{Output signal-to-noise ratio} = z^2 = s_o^2(t_o) / \sigma^2 = \frac{\left| \frac{1}{2\pi} \int_{-\infty}^{\infty} Y(\omega) S(\omega) e^{j\omega t_o} d\omega \right|^2}{\frac{1}{2\pi} \int_{-\infty}^{\infty} D |Y(\omega)|^2 d\omega}. \quad (5-112)$$

The minimum value of this ratio can be determined by means of Schwarz's inequality. This can be derived in the following fashion. Suppose that the functions $f(x)$ and $g(x)$ and the parameter μ are real. Then the

following quadratic function of μ will always be greater than or equal to zero:

$$\int_a^b [\mu f(x) + g(x)]^2 dx = \mu^2 \int_a^b f^2(x) dx + 2\mu \int_a^b f(x)g(x) dx + \int_a^b g^2(x) dx \geq 0. \quad (5-113)$$

This expression is represented by

$$A\mu^2 + 2B\mu + C \geq 0. \quad (5-114)$$

Because this polynomial is always greater than zero, the equation

$$A\mu^2 + 2B\mu + C = 0. \quad (5-115)$$

cannot have distinct real roots, and its discriminant must be less than or equal to zero:

$$B^2 - AC \leq 0. \quad (5-116)$$

Substituting for A , B , and C gives the real form of Schwarz's inequality:

$$\left(\int_a^b f(x)g(x) dx \right)^2 \leq \int_a^b f^2(x) dx \int_a^b g^2(x) dx. \quad (5-117)$$

The absolute value of the product of two complex numbers is always less than or equal to the product of their absolute values. Further, the square of the absolute value of the integral of a complex function is always less than or equal to the integral of the square of the absolute value of the integrand. Combining these ideas, we note that when $f(x)$ and $g(x)$ are complex,

$$\left| \int_a^b f(x)g(x) dx \right|^2 \leq \int_a^b |f(x)g(x)|^2 dx \leq \int_a^b |f(x)|^2 |g(x)|^2 dx. \quad (5-118)$$

This immediately leads to the more general form which we need. Putting $|f(x)|$ and $|g(x)|$ in place of $f(x)$ and $g(x)$ in Equation 5-118:

$$\left| \int_a^b f(x)g(x) dx \right|^2 \leq \int_a^b |f(x)|^2 dx \int_a^b |g(x)|^2 dx. \quad (5-119)$$

Substituting the right-hand side of this inequality for the numerator in Equation 5-112.

$$z^2 \leq \frac{\frac{1}{2\pi} \int_{-\infty}^{\infty} |S(\omega)|^2 d\omega \frac{1}{2\pi} \int_{-\infty}^{\infty} |Y(\omega)|^2 d\omega}{\frac{1}{2\pi} \int_{-\infty}^{\infty} D|Y(\omega)|^2 d\omega} \quad (5-120)$$

The integrals involving the filter transfer function can be canceled:

$$z^2 \leq \left(\frac{1}{D} \right) \frac{1}{2\pi} \int_{-\infty}^{\infty} |S(\omega)|^2 d\omega. \quad (5-121)$$

Thus the right-hand side of this inequality is independent of the filter. Since the signal-to-noise ratio is never greater than the right-hand side of the expression above, and this expression does not contain $Y(\omega)$ at all, it must give the maximum value for z^2 for the optimum choice of $Y(\omega)$. Referring to Equation 5-112, it is apparent that the denominator in that equation will be canceled and the maximum value of z^2 achieved if the filter transfer function is made the complex conjugate of the signal spectrum:

$$Y(\omega) = S^*(\omega)e^{-j\omega t_0}. \quad (5-122)$$

A receiver-filter which is designed on the basis of this principle, where the receiver transfer function is matched to the signal waveform, is often referred to as a matched filter system. Another general term which is also used in reference to such systems is correlation radar. This terminology originates in the observation that the ideal filtering operation is equivalent to a cross correlation of the signal plus noise with an image of the signal waveform. In order to see this, the impulse response of the matched filter is found by taking the inverse Fourier transform of $Y(\omega)$:

$$\begin{aligned} \text{Impulse response of matched filter} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} S^*(\omega)e^{-j\omega t_0 + j\omega t} d\omega \\ &= s(t_0 - t). \end{aligned} \quad (5-123)$$

Denoting the input noise process by $n(t)$ and the output noise process by $n_o(t)$ and using Equation 5-16 to relate the time histories of the input and output signal plus noise gives for the filter output:

$$s_o(t) + n_o(t) = \int_{-\infty}^{\infty} [s(\tau) + n(\tau)]s(t_0 - t + \tau)d\tau. \quad (5-124)$$

In particular, the output at the observation time t_0 is simply

$$s_o(t_0) + n_o(t_0) = \int_{-\infty}^{\infty} [s(\tau) + n(\tau)]s(\tau)d\tau. \quad (5-125)$$

Thus, from this relation it is clear that the optimum receiver could consist of taking the cross correlation of the received signal plus noise and the pure signal waveform and that a matched filter receiver and a cross correlation receiver are equivalent.

Going back to Equation 5-121 for a moment, we might note an interesting basic feature of radar systems which are theoretically optimum in the sense of this paragraph. *The maximum signal-to-noise ratio is equal to the ratio of the received signal energy to the power density of the noise.* That is, the maximum signal-to-noise ratio does not depend upon the waveform of the signal. This is not to say that the waveform is not important. Resolution, tracking accuracy, and many other system characteristics are closely related

to and depend upon the wave shape of the signal. For the detection problem, though, it is the received energy that counts.

As a concrete illustration of a matched filter, suppose that the signal waveform consists of a series of n identical pulses separated by a repetition period T . Such a signal is of common occurrence in radar problems. Denoting an individual pulse by $p(t)$, the signal is defined by

$$\text{Pulse train} = s(t) = \sum_0^{n-1} p(t - kT). \quad (5-126)$$

This signal is depicted in Fig. 5-15a.

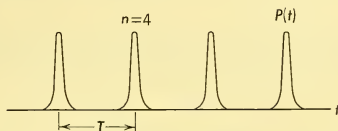


FIG. 5-15a Pulse Train Signal.

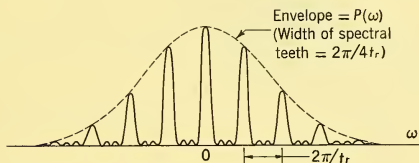


FIG. 5-15b Spectrum of Pulse Train Signal.

From Equation 5-125, the signal component of the filter output at the observation time will be

$$\begin{aligned} \text{Filter output (signal)} &= s_o(t_o) = \int_{-\infty}^{\infty} s^2(\tau) d\tau \quad (5-127) \\ &= \int_{-\infty}^{\infty} \sum_0^{n-1} \sum_0^{n-1} p(\tau - kT) p(\tau - mT) d\tau \\ &= \int_{-\infty}^{\infty} \sum_0^{n-1} p^2(\tau - kT) d\tau \\ &= n \int_{-\infty}^{\infty} p^2(\tau) d\tau. \end{aligned}$$

The effect of the correlation (or filtering) operation has been to select out all the available signal pulses and add them together. A device which will perform this addition is most often referred to as a *pulse integrator*, and

almost all radar receivers whose inputs consist of a series of pulses incorporate such a device in one form or another.

The effect of the pulse integrator on the noise can also be determined from Equation 5-125. The noise output will be

$$\text{Filter output (noise)} = n_o(t_o) = \int_{-\infty}^{\infty} n(\tau) \sum_0^{n-1} p(\tau - kT) d\tau. \quad (5-128)$$

The noise power is determined by squaring $n_o(t_o)$ and finding its average value:

$$\text{Noise power} = \sigma^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \overline{n(\tau_1)n(\tau_2)} \sum_0^{n-1} \sum_0^{n-1} p(\tau_1 - kT)p(\tau_2 - mT) d\tau_1 d\tau_2. \quad (5-129)$$

Since $n(t)$ was assumed to have a uniform spectrum with density D , the average value of the product $n(\tau_1)n(\tau_2)$ is an impulse function with weight D :

$$\begin{aligned} \text{Noise power} = \sigma^2 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} D\delta(\tau_1 - \tau_2) \sum_0^{n-1} \sum_0^{n-1} p(\tau_1 - kT)p(\tau_2 - mT) d\tau_1 d\tau_2 \\ &= D \int_{-\infty}^{\infty} \sum_0^{n-1} \sum_0^{n-1} p(\tau_2 - kT)p(\tau_2 - mT) d\tau_2 \quad (5-130) \\ &= D \int_{-\infty}^{\infty} \sum_0^{n-1} p^2(\tau_2 - kT) d\tau_2 \\ &= nD \int_{-\infty}^{\infty} p^2(\tau_2) d\tau_2. \end{aligned}$$

In evaluating the integral of the double sum, we made use of the fact that when the pulse functions in the integrand do not coincide ($k \neq m$), their product is zero:

$$z^2 = n \left(\frac{1}{D} \right) \int_{-\infty}^{\infty} p^2(\tau) d\tau. \quad (5-131)$$

It is apparent that the effect of the pulse integrator is to increase the signal-to-noise ratio for a single pulse by the factor n . This could, of course, be inferred at the outset from Equation 5-121, since the signal power is directly proportional to n .

The shape of the matched filter response in this case is of some interest. Denoting the spectrum of an individual pulse by $P(\omega)$, the spectrum of the pulse train will be

$$\begin{aligned} \text{Pulse train spectrum} = S(\omega) &= P(\omega) \sum_0^{n-1} e^{-jk\omega T} \quad (5-132) \\ &= P(\omega) \left(\frac{\sin n\omega T/2}{\sin \omega T/2} \right) e^{-j(n-1)\omega T/2}. \end{aligned}$$

The energy spectrum of a typical train of short pulses is shown in Fig. 5-15b. We note that the filter primarily acts to accentuate the harmonics of the repetition frequency. Because of its distinctive appearance, such a device is often called a *comb filter*.

Some explanation on the mechanism of the detection process itself is in order since the previous discussion related only to maximizing the signal-to-noise ratio. The output of the matched filter characterizes a signal-plus-noise situation by a single number $s_o(t_o) + n_o(t_o)$. This number is a random variable with a normal distribution and mean $s_o(t_o)$. The detection process will consist of a decision as to whether the observed number comes from a distribution with mean $s_o(t_o)$ or the distribution of noise alone with a zero mean. This decision can be made by selecting a critical value or threshold and deciding for or against the existence of the signal depending upon whether or not the observed number exceeds the threshold. Fig. 5-16

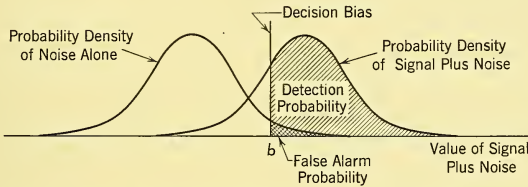


FIG. 5-16 The Use of a Decision Bias for Determining Whether Noise Alone or Signal Plus Noise Is Present.

shows the probability densities of the filtered signal plus noise and noise alone and a decision bias b for distinguishing the two cases. Because the two probability densities overlap, mistakes will be made. On some occasion a target will be thought present when there is none, while at other times the signal plus noise will be thought to be noise alone. The probability of making an error of the first kind is equal to the crosshatched area under the curve of noise alone and to the right of b in the diagram. This probability is normally called the false-alarm probability by radar system designers. The shaded area under the probability density curve of signal plus noise and to the right of b is the probability of detection. The difference between this probability and unity is, of course, the probability of making an error of the second kind or not seeing a target that is actually present. When human operators make a detection, the situation is not nearly so clear-cut, but some similar mechanism must take place. The decision bias might be visualized as diffuse, and it will vary with operators, time, and other conditions.

A basic problem is the choice of the false-alarm probability at which the system is to operate. Most often this operating parameter is chosen

on subjective grounds because the data upon which to base a rational choice are not available. Factors which can be used to determine an optimum false-alarm probability are the cost of a false alarm in time and subsequent commitments, the gain associated with a correct decision, and prior probability of a target's existence. If quantitative estimates of these factors are available, the false-alarm probability can be chosen to minimize the total cost of the detection operation. Or even when the prior probability is not known, it is possible to operate the system at false-alarm rates so as to minimize the cost for the most adverse value of the prior probability. As noted above, however, data on detection costs and prior detection probabilities are known only subjectively in the majority of cases, and most often a rather arbitrary estimate of a desirable value of the false-alarm rate is made after a thorough but subjective study of the problem.

At the beginning of our discussion of the optimum receiving system, it was assumed that the waveform of the signal was known exactly, and the only issue was its existence. In a practical detection situation, however, the signal waveform may depend upon a number of unknown parameters. Three such parameters which are of particular importance are the signal amplitude, the time of arrival of the signal, and the radio frequency of the carrier. The signal amplitude will vary with the range, aspect, and size of the target, while the time of arrival is, of course, directly proportional to range in a radar system. The RF carrier will vary because of the doppler shifts proportional to the relative target velocity. An optimum receiver in this case will consist of a parallel combination of optimum receivers for all the possible waveforms. Luckily, this does not require a duplication of equipment to cover the possibilities of amplitude and time-of-arrival variations. If the signal amplitude is changed by some factor, then the average value of the filter output is changed by the same factor. The same filter will produce the maximum value of z^2 for all possible signal amplitudes. A similar situation applies to variations in time-of-arrival. The optimum receiver produces its maximum output at a time T after a signal is received. Continuous monitoring of the receiver output, then, will provide an observation of the filtered signal over a continuous range of possibilities for the time of arrival. In order to account for variations in the radio frequency, however, it will in general be necessary to have separate receiving systems for the possible radio frequencies which may occur. This situation will be recognized in the design of many doppler systems where a bank of narrow band filters, each connected to its own threshold, is used to cover the possible spectrum of doppler signals.

The situation is complicated further by the fact that some of the signal parameters are random variables in their own right. For example, the amplitudes of echoes reflected by aircraft fluctuate owing to their motions,

and the radio frequency of a magnetron oscillator normally varies randomly from pulse to pulse by a small amount. The statistics of signal parameter distributions would have to be considered in a more realistic optimum receiver, and the result would be somewhat different from that derived in this paragraph.

One should not make the mistake of thinking that great gains over current practice can be attained through some complicated optimizing scheme. Actually, most radar systems are tuned up in this respect about as far as they can be when consideration is taken of limitations in the state of the art and fluctuations in the parameters of the input signals with which the systems must contend. For instance, a pulse radar employing a self-excited magnetron oscillator is not coherent because it is not normally feasible to control the frequency of the power oscillator to a sufficient degree. Because of pulse-to-pulse frequency fluctuations, the receiver must operate upon the envelope of the signal, and it will normally employ a non-linear device to generate it. In this case, the best that can be done is to match the low-pass equivalent of the IF amplifier to the pulse envelope, and this is quite normally done as a matter of course. When there are a number of pulses available in an echo, some provision is usually made to integrate them. Most commonly, this is accomplished on the display where the decay time of the phosphor may be matched to the signal duration. The point is that insofar as is possible receivers are normally matched to the signal waveform, and most radar systems can be quite legitimately regarded as correlation or matched filter radars, although possibly somewhat degraded from the optimum type.

5-11 APPLICATION TO THE DETERMINATION OF A SIGNAL'S TIME OF ARRIVAL

An important function of radar systems is the measurement of a target's parameters such as its range, velocity, size, and location. In this paragraph we shall develop some characteristics of a receiver which provides optimum target tracking in a manner similar to that used in the preceding paragraph to develop the properties of matched filter receivers for optimum target detection. We shall restrict our analysis to the problem of measuring the time of arrival of the signal. Since both angle and range are measured by comparing the return signal with angle and range reference signals which are generated as functions of time, the following discussion can be applied to both types of tracking. As in Paragraph 5-10, the receiver will be supposed to be a linear filter, and where applicable, the notation introduced in that paragraph will be adopted.

Various operational definitions might be used to fix the arrival time of a signal. The mean, the median, or the mode of the distribution of the signal

in time are all quite applicable. We shall find it most convenient to adopt the last of these, the mode or the maximum value of the signal, as the primary indicator of the signal's location. This definition gives a straightforward development which parallels that of Paragraph 5-10 and which to a first approximation leads to results in accord with more elaborate analyses. Even so, we must recognize that since the choice of a definition for the signal location is purely arbitrary, we are optimizing the tracking process only relative to that definition and not in an absolute sense.

In order to use the peak value of the filtered signal plus noise as an unambiguous estimate of the signal location, we shall make several assumptions about the form of the signal and signal plus noise. First, we assume that the signal itself either has a single maximum or that the greatest maximum is sufficiently larger than minor maxima to allow it to be unambiguously distinguished. Second, we assume that the primary maximum of the filtered signal has a finite second derivative, since we intend to locate it by setting the first derivative of the signal plus noise equal to zero. Third, the filtered signal is assumed to be enough greater than the noise that there are no ambiguous noise maxima in the neighborhood of the primary maximum and the shift in this maximum due to the presence of the noise is small enough to be approximated by the first few terms in a series expansion. Suppositions of this kind are not unusual in parameter estimation problems, and equivalent assumptions and approximations almost always must be adopted when a specific example is worked out.

Fig. 5-17 shows a typical example of signal plus noise in the neighborhood of the signal maximum and illustrates how the addition of noise acts to

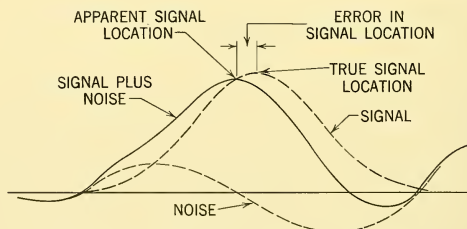


FIG. 5-17 Generation of Signal Location Error.

shift the maximum slightly from its former value. The magnitude of the shift can be determined approximately by differentiating the signal plus noise and setting it equal to zero. The resulting expression will be in the form of a quotient very similar to that given in Equation 5-118 for the signal-to-noise power ratio. Schwarz's inequality can also be applied to

this expression, and we can determine the optimum filter for tracking which gives the minimum tracking error. This error will be interpreted as a simple relation between the signal bandwidth and the signal-to-noise ratio.

The filtered signal and noise are denoted by $s_o(t)$ and $n_o(t)$ as in the preceding paragraph. We suppose that the maximum value of the signal occurs at the observation time t_o . We suppose further that the output signal at the time $t_o + \Delta t$ can be represented by a series expansion about the time t_o for small values of the interval Δt .

$$s_o(t_o + \Delta t) = s_o(t_o) + s_o'(t_o)\Delta t/2 + \dots \quad (5-133)$$

The first derivative of $s_o(t)$ at t_o is zero, of course, because it has a maximum at that time. We assume that the shift in the maximum value of the signal plus noise is small enough that all terms beyond the second in the expansion above can be neglected. The derivative of signal plus noise in the neighborhood of t_o is thus given approximately by

$$\frac{d}{dt}[s_o(t) + n_o(t)]_{t=t_o+\Delta t} \approx s_o''(t_o)\Delta t + n_o'(t_o + \Delta t). \quad (5-134)$$

Setting this expression equal to zero and solving for Δt gives an approximate value for the apparent shift in signal location due to noise:

$$\Delta t \approx \frac{-n_o'(t_o + \Delta t)}{s_o''(t_o)}. \quad (5-135)$$

The variance or mean square value of the signal location error is thus given approximately by the average of the square of this expression:

$$\overline{\Delta t^2} \approx \frac{\overline{[n_o'(t)]^2}}{[s_o''(t_o)]^2}. \quad (5-136)$$

If we denote the transfer function of the filter-receiver by $Y(\omega)$ and the signal spectrum by $S(\omega)$ as in Paragraph (5-10), the following representation of $s_o''(t_o)$ can be obtained by differentiating Equation 5-116 twice:

$$s_o''(t_o) = \frac{-1}{2\pi} \int_{-\infty}^{\infty} \omega^2 Y(\omega) S(\omega) e^{i\omega t_o} d\omega. \quad (5-137)$$

Also assuming, as in Paragraph 5-10, that the input noise has a uniform spectrum of density D , the power spectrum of the output noise is $D|Y(\omega)|^2$ while the power spectrum of the derivative of the output noise is $D\omega^2|Y(\omega)|^2$. The integral of this last spectrum gives the mean square value or variance of the derivative of the output noise:

$$\overline{[n_o'(t)]^2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} D\omega^2 |Y(\omega)|^2 d\omega. \quad (5-138)$$

The quotient in Equation 5-136 giving the variance of the signal location error thus has the following form:

$$\overline{\Delta t^2} \approx \frac{\frac{1}{2\pi} \int_{-\infty}^{\infty} D\omega^2 |Y(\omega)|^2 d\omega}{\left(\frac{1}{2\pi} \int_{-\infty}^{\infty} \omega^2 Y(\omega) S(\omega) e^{j\omega t_0} d\omega \right)^2} \quad (5-139)$$

For convenience, we denote the quotient on the right-hand side by \mathcal{Q} . The denominator of this quotient is in a form to which Schwarz's inequality, given by Equation 5-125, can be applied. Using this relation to split the denominator into two separate integrals leads to the following inequality:

$$\mathcal{Q} \geq \frac{\frac{1}{2\pi} \int_{-\infty}^{\infty} D\omega^2 |Y(\omega)|^2 d\omega}{\frac{1}{2\pi} \int_{-\infty}^{\infty} \omega^2 |Y(\omega)|^2 d\omega \frac{1}{2\pi} \int_{-\infty}^{\infty} \omega^2 |S(\omega)|^2 d\omega} \geq \frac{D}{\frac{1}{2\pi} \int_{-\infty}^{\infty} \omega^2 |S(\omega)|^2 d\omega} \quad (5-140)$$

The integrals involving the transfer function of the filter simply cancel as they did in Paragraph 5-10, where the maximum value of the signal-to-noise ratio was determined. Since the quotient \mathcal{Q} is never less than the expression given on the right-hand side above, which does not contain $Y(\omega)$ at all, this expression must give the minimum value of \mathcal{Q} for the most judicious choice of $Y(\omega)$. Referring to Equation 5-139 above, it is apparent that this minimum value of \mathcal{Q} will actually be achieved if the filter transfer function is chosen to be the conjugate of the signal spectrum. In this case, the numerator in Equation 5-139 cancels one of the factors in the denominator, and we have

$$\overline{\Delta t^2} \approx \mathcal{Q}_{\min} = \frac{D}{\frac{1}{2\pi} \int_{-\infty}^{\infty} \omega^2 |S(\omega)|^2 d\omega} \quad (5-141)$$

The optimum filter transfer function giving this result is

$$Y(\omega) = S^*(\omega) e^{-j\omega t_0} \quad (5-142)$$

This is exactly the transfer function determined in Paragraph 5-10 (Equation 5-128) to give the maximum signal-to-noise ratio. Thus *to a first approximation the matched filter giving the maximum signal-to-noise ratio also provides the minimum error in locating the signal in time.*

The relation given by Equation 5-141 above for the minimum variance of the error in locating the signal can be given an interesting and rather useful physical interpretation. We note that the denominator has the form of the moment of inertia of the energy spectrum of the signal. If this denominator is divided by the total signal energy, we obtain the square of the radius of gyration of the energy spectrum. Now the radius of gyration of a function

is a measure of the width of that function. Thus the radius of gyration of the energy spectrum is a measure of the bandwidth of that spectrum. We shall adopt this definition for the bandwidth of the signal spectrum:

$$\text{Signal bandwidth (rad/sec)} = B = \frac{\left(\frac{1}{2\pi} \int_{-\infty}^{\infty} \omega^2 |S(\omega)|^2 d\omega \right)^{1/2}}{\left(\frac{1}{2\pi} \int_{-\infty}^{\infty} |S(\omega)|^2 d\omega \right)^{1/2}} \quad (5-143)$$

A question may arise in connection with the application of this definition to spectra which are not centered at zero frequency. It is clear that the radius of gyration of an energy spectrum which is concentrated at low frequencies is a good measure of the bandwidth of the spectrum. The radius of gyration of a spectrum whose center is displaced to some high frequency, though, will be very large, and it does not correspond to the conventional idea of bandwidth. Such a signal can be represented as a function, denoted by $f(t)$, whose spectrum is concentrated at low frequencies and which is modulated by a high-frequency carrier:

$$\text{High frequency signal} = f(t) \cos(\omega_c t + \phi) \quad (5-144)$$

This signal contains information about its time location with an accuracy on the order of $1/\omega_c$. But when the ratio of the carrier frequency to the bandwidth of $f(t)$ is large, this information is useless because it is ambiguous. This is illustrated in Fig. 5-18 where a relatively smooth low-frequency

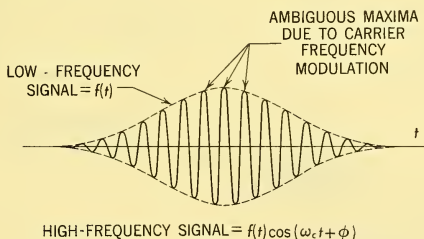


FIG. 5-18 Typical High-Frequency Signal.

signal $f(t)$ is modulated by many carrier cycles. It can be seen in this figure that the carrier frequency modulation produces a number of signal maxima in the neighborhood of the maximum of $f(t)$. Because these maxima are all of about the same magnitude, there is no way of distinguishing one from another, and signal location information provided by the carrier-frequency modulation is ambiguous. In such a case, the low-frequency signal $f(t)$ is normally regenerated by a demodulating operation, and the

location of the signal is determined on the basis of $f(t)$ alone. The effective bandwidth in such a case, then, is that of the spectrum of $f(t)$, and it would be determined relative to the carrier frequency rather than zero frequency as is indicated in Equation 5-143. When the definition of bandwidth given by Equation 5-143 is combined with Equation 5-141, the mean square time error is found to be approximately equal to the ratio of D , the input noise power density, to the product of the square of the signal bandwidth and the signal energy. The ratio of the signal energy to D , however, was established in Equation 5-129 as the greatest possible signal-to-noise ratio and was denoted by z^2 . Thus we can assert that *the minimum rms error in measuring the time of arrival of a signal is approximately equal to the reciprocal of the product of the signal bandwidth and the voltage signal-to-noise ratio:*

$$(\overline{\Delta t^2})^{1/2} \approx 1/Bz \quad (5-145)$$

rms time error $\approx 1/(\text{signal bandwidth})(\text{voltage signal-to-noise ratio})$.

As an example of the application of these ideas, let us consider a pulse radar with a narrow antenna pattern which is scanned over the target at a constant angular rate. Such a system is similar to the AEW example discussed in Paragraphs 2-10 to 2-20 and the results that we shall develop are applicable to the design considerations in that example.

The video signal generated by such a system would have a form similar to that shown in Fig. 5-19. The time at which the envelope of the pulses

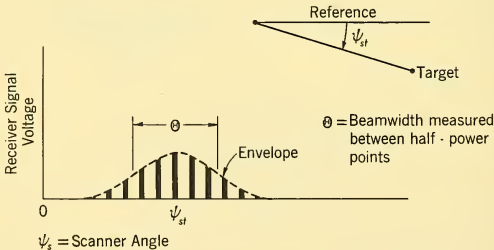


FIG. 5-19 Receiver Voltage Pulse Train Return from a Point Target.

reaches its peak value will be correlated with the angular position of the target so that the problem of locating the target in angle is essentially that of determining the arrival time of the signal, and the ideas and developments of this paragraph are applicable.

The basic functions performed by the system are indicated in the block diagram in Fig. 5-20. The received signal is amplified and filtered by an IF amplifier which is matched to the envelope of an individual pulse. Noise

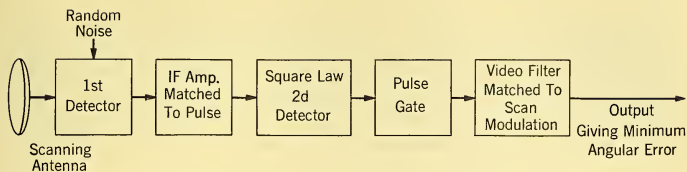


FIG. 5-20 Block Diagram of Receiver of Scanning Radar.

with a uniform power density is introduced at the input to this amplifier. A square-law second detector is used to generate the video envelope of the signal plus noise. The video signal-plus-noise pulses are gated into a video filter which is matched to the scan modulation. That is, this filter is matched to the fundamental component of the gated video signal plus noise. All signals, information, and noise at the repetition frequency and its higher harmonics are filtered out. We assume that the number of pulses per beamwidth is sufficiently large that the signal spectrum about the first harmonic of the repetition frequency does not overlap the fundamental component of the signal spectrum. We also assume that the video noise is sufficiently uniform over the bandwidth of the scan modulation signal for the assumption of a constant noise spectrum under which we derived Equation 5-145 to be valid. Other system configurations are possible. A more practical design might stretch the gated signal-plus-noise pulses before smoothing by the scan modulation filter. Such a system, however, would give slightly greater angular errors than the one we have chosen to study.

The following special notation is adopted for this example.

a = voltage amplitude of received signal

T = repetition period

d = duty ratio

ψ = antenna angle

$\dot{\psi}$ = antenna angular rate

θ = antenna beamwidth (half-power points — one-way)

δ = pulse width

$n = \theta / \dot{\psi} T$ = number of pulses per beamwidth

The signal received from the target will have the following form:

$$\text{Received signal} = a (\text{scan modulation})(\text{pulse modulation}) \cos \omega_c t. \quad (5-146)$$

The peak pulse power will be $a^2/2$. We assume that the individual pulses are rectangular and of width δ . The total energy in a pulse, then, is $a^2\delta/2$. Assuming noise with a uniform power spectrum of density D , the maximum signal-to-noise ratio which can be obtained with a filter matched to a pulse is given by the ratio of the pulse energy to the power density of the noise as was derived in Equation 5-127:

$$\text{Peak signal to noise ratio} = S/N = a^2\delta/2D. \quad (5-147)$$

The noise at the output of the IF amplifier corresponds to a narrow band noise process similar to those discussed in Paragraph 5-8. Since the IF amplifier is matched to the envelope of the pulse signal, the autocorrelation function and power spectrum of the low-frequency in-phase and quadrature components of the noise, $x(t)$ and $y(t)$ in Equation 5-73, will be of the same form as that of a rectangular pulse. This autocorrelation is a triangular function of width 2δ and height equal to the noise power.

The properties of the video signal and noise after a square law detector can be determined from Equation 5-87 which gives the autocorrelation function of the output of a square-law second detector:

$$\overline{r_1^2 r_2^2} = (2\sigma^2)^2(1 + S/N)^2 + (2\sigma^2)^2[\rho^2 + 2(S/N)\rho]. \quad (5-87)$$

The first term in this expression corresponds to the video d-c level during a pulse while the term involving ρ and ρ^2 corresponds to the video noise. In order to exclude the possibility of ambiguities incident to a noise maximum in the neighborhood of the signal maximum, we assumed that the signal-to-noise ratio was large in the development of Equation 5-145. It will simplify the present analysis if we approximate Equation 5-87 for large S/N by considering only the dominant d-c and noise terms in that equation:

$$\overline{r_1^2 r_2^2} \approx (2\sigma^2)^2(S/N)^2 + (2\sigma^2)^2 2(S/N)\rho, \quad S/N \gg 1. \quad (5-148)$$

The normalized autocorrelation function ρ in this expression corresponds to the triangular function noted above of width $2Td$ but of unit height.

The shape of the video spectrum will not be exactly uniform as is required for the developments of this paragraph to be rigorous. The total width of the spectrum, though, will normally be greater than the spectrum of the scan modulation by a factor on the order of 10^4 . Variation of the noise spectrum over the scan modulation bandwidth, then, will be quite small, and we are justified in approximating the noise spectrum by a spectrum with a uniform power density. The power density of the noise spectrum at zero frequency, which we shall assume to be extended to all frequencies, is found by integrating the autocorrelation of the noise. The integral of the triangular function $\rho(\tau)$ is δ . When this value is substituted for ρ in Equation 5-148, the term involving this factor gives the power density of the noise at zero frequency during a pulse or with a c-w signal.

Because the video signal plus noise is gated, the noise power and thus the noise power density will be smaller than the value during a pulse by the duty ratio d . Taking these factors and considerations into account, the effective power density is determined from Equation 5-87 to be

Power density at zero frequency of gated video noise

$$\approx (2\sigma^2)^2 2(S/N)\delta d, S/N \gg 1. \quad (5-149)$$

The d-c component of the video corresponds to the signal which is used to locate the target. From Equation 5-148 it will be noted that this component is proportional to the received signal power. Thus the video signal is modulated by a scan modulation function which indicates how the received power fluctuates during a scan. The first term in Equation 5-148 gives the square of the d-c level during a pulse for large signal-to-noise ratios. To obtain the d-c level of our gated signal, we must multiply the signal level during a pulse by the duty ratio d . The resulting video signal has the following approximate form:

$$\text{Gated video signal} \approx (2\sigma^2)(S/N)^d(\text{scan power modulation}), \quad S/N \gg 1. \quad (5-150)$$

We assume that the antenna pattern of the system has a Gaussian shape and that the same antenna is used for both transmission and reception. The beamwidth of the pattern θ is defined as the angle between the half-power points for one-way transmission. The gain of the two-way power pattern would thus be down by a factor of 4 at these points. The antenna angle is denoted by $\dot{\psi}$. We assume that the scan modulation is generated by a constant velocity scan at the rate $\dot{\psi}$. Supposing that the signal maximum occurs at the time $t = 0$, the scan modulation has the form

$$\begin{aligned} \text{Scan power modulation} &= \exp(-\dot{\psi}^2/0.18\theta^2) \\ &= \exp(-\dot{\psi}^2 t^2/0.18\theta^2) \end{aligned} \quad (5-151)$$

Because we have assumed a square-law second detector, the scan modulation of the video voltage will be proportional to this scan power modulation.

We are now in a position to apply the result of Equation 5-145, giving the rms error in time of arrival, which in turn yields the rms angular error after multiplication by the scan rate. We first compute the signal-to-noise ratio in the output of the video filter matched to the scan modulation. The energy in the signal is given by the integral of its square:

Signal energy at low frequencies

$$\begin{aligned} &= (2\sigma^2)^2(S/N)^2 d^2 \int_{-\infty}^{\infty} \exp(-\dot{\psi}^2 t^2/0.09\theta^2) dt \\ &= 0.53(2\sigma^2)^2(S/N)^2 d^2(\theta/\dot{\psi}). \end{aligned} \quad (5-152)$$

The power density of the video noise is given in Equation 5-149. It was determined in Paragraph 5-10 that the signal-to-noise ratio at the output of a matched filter is equal to the ratio of the signal energy to the noise power density, so we have

$$\begin{aligned} \text{Signal-to-noise ratio at} \\ \text{output of video filter} = z^2 &= \frac{0.53(2\sigma^2)^2(S/N)^2d^2(\Theta/\dot{\psi})}{(2\sigma^2)^22(S/N)d^2T} \\ &= 0.265(S/N)(\Theta/\dot{\psi}T) \\ &= 0.265(S/N)n. \end{aligned} \quad (5-153)$$

The number of pulses per beamwidth given by the ratio $\Theta/\dot{\psi}T$ has been denoted by n in this equation.

The bandwidth of the signal can be determined from the spectrum of the scan modulation. The Fourier transform of the modulating function in Equation 5-151 is as follows.

$$\text{Spectrum of scan modulation} = \left(\sqrt{0.18\pi} \frac{\Theta}{\dot{\psi}} \right) \exp\left(\frac{-\omega^2 0.09\Theta^2}{2\dot{\psi}^2} \right) \quad (5-154)$$

The bandwidth, as defined by Equation 5-143, is easily computed:

$$\begin{aligned} \text{Scan modulation bandwidth} = B &= \frac{\left[\int_{-\infty}^{\infty} \omega^2 \exp\left(\frac{\omega^2 0.18\Theta^2}{2\dot{\psi}^2} \right) d\omega \right]^{1/2}}{\left[\int_{-\infty}^{\infty} \exp\left(\frac{-\omega^2 0.18\Theta^2}{2\dot{\psi}^2} \right) d\omega \right]^{1/2}} \\ &= 2.35\dot{\psi}/\Theta \end{aligned} \quad (5-155)$$

The rms error in measuring the target angle can now be estimated from Equation 5-145 as the scan rate divided by the product of the scan modulation bandwidth and the voltage signal-to-noise ratio in the video filter output:

$$\begin{aligned} \text{rms angle error} &= (\overline{\Delta\psi^2})^{1/2} = (\overline{\Delta t^2})^{1/2} = \dot{\psi}/Bz \quad (5-156) \\ &= \dot{\psi}\Theta/2.35\dot{\psi}\sqrt{0.265(S/N)n} \\ &= 0.825\Theta/\sqrt{(S/N)n}. \end{aligned}$$

The angular error of a scanning radar has been studied in the technical literature⁹ for conditions very similar to the assumptions of this example. In that study the minimum possible rms angular error was found to be approximately proportional to an expression of the form of Equation 5-156

⁹P. Swerling, "Maximum Angular Accuracy of a Pulsed Search Radar," *Proc. IRE* **44**, 1140-1155 (1956).

but only about half as great. Actually, the estimate in Equation 5-156 is optimistic. The mechanism by which the maximum value is chosen can introduce additional errors; nor was any consideration given to the effect of target fluctuations which would act to increase the error. As a typical case, if the signal-to-noise ratio out of the matched video filter as given by Equation 5-153 is 6 db, the rms angle error from Equation 5-156 is

$$\begin{aligned} \text{rms angle error} &= \theta/2.35\sqrt{4}, \quad z^2 = 4 && (5-157) \\ &= 0.213\theta. \end{aligned}$$

In Paragraph 2-15 it was assumed that the target angle could be determined in a scan to 1/4 of a beamwidth. From the relation above, this is not an unreasonable assumption.

CHAPTER 6

GENERIC TYPES OF RADAR SYSTEMS AND TECHNIQUES*

6-1 INTRODUCTION

The radar systems engineer is often asked to solve the following problem:

“Given a set of performance specifications based on the tactical problem requirements, derive a radar system that will meet the specifications.”

For a variety of reasons, it is seldom possible to solve this problem in a straightforward fashion. Probably the most important reason is this: The performance specification — if properly derived — will seldom specify a task which simply cannot be performed by radar techniques (for example, the performance specification could not logically ask the radar to distinguish between red and blue aircraft); however, the performance specification will usually require the radar to perform a *group* of tasks which are not logically consistent with any one radar system mechanization. Even when mechanization limitations are excluded from consideration, there is no such thing as an “ideal” radar system which can perform any group of functions in an optimum manner. Every conceivable type of radar system possesses a combination of good and bad characteristics and both must be accepted and rationalized in a given application.

The usual approach is to assume a generic type of radar system which experience and judgment deem reasonable. The assumed system then is measured analytically against the overall system requirements to determine whether it has the inherent potential for providing an acceptable problem solution. This process is repeated until the best match is found between the performance specification and the basic laws of nature governing what can be done by a given radar system.

*Paragraphs 6-1, 6-2, 6-6, and 6-7 are by D. J. Povejsil. Paragraph 6-3 is by R. M. Page. Paragraph 6-4 is by S. F. George. Paragraph 6-5 is by L. Hopkins. Paragraph 6-8 is by H. Yates.

Unfortunately, there is a tremendous variety of possible choices. In terms of generally recognized system types and subtypes, there are *pulse*, *continuous-wave (CW)*, *pulsed doppler*, *monopulse*, *correlation*, *high-resolution*, and *moving target indication (MTI)* radars. Some of these types represent genuinely different approaches; some of them represent merely alternative means for performing the same job; and some of them are derivatives of particular system types. In each case, however, the selection of one of these types commits the radar system designer to a problem approach that is confined within uncomfortably narrow limits. The radar system designer must therefore have a good general knowledge of the basic system types and the general laws that govern their performance characteristics. Toward this end, this chapter will attempt to accomplish two things.

- (1) It will summarize basic radar laws in a rule-of-thumb fashion to provide a means for understanding the operation of any radar system.
- (2) It will describe the performance characteristics and limitations of generally recognized radar system types and will indicate their general areas of application.

6-2 BASIC PRINCIPLES

The operation of almost any radar system may be visualized and understood by asking and answering the following basic questions:

- (1) Is the system *active*, *semiactive*, or *passive* (see Paragraph 1-4)?
- (2) What information is contained in the signal return from the assumed target complex?
- (3) What are the system sampling frequencies?
- (4) How are the radar data detected and processed in the receiving system?
- (5) Where does the processed information go?

Each of these questions may now be considered in greater detail.

Type of Radar System. The most basic division of radar system types is a classification based on the origin of the target signal information. An *active* system generates the signals which are ultimately scattered back to the point of signal origin. A *passive* system is simply a receiving system which utilizes target-generated radiations as its signal source. A semiactive system employs separate transmission and receiving systems which may be at some distance from each other. Depending upon the degree of coupling between the transmitted and received signals, a semiactive system may resemble either an active or a passive system insofar as its basic operation is concerned.

Type of Information. In any radar system problem there are two basic kinds of target information:

- (1) The *desired* radar-derived target information
- (2) The radar-derived target information that is actually obtained using a given system

The latter will be exemplified by answering the last three questions posed at the beginning of this paragraph.

Most commonly, the desired radar-derived target information takes the form of a four-dimensional information matrix as shown in Fig. 6-1. The

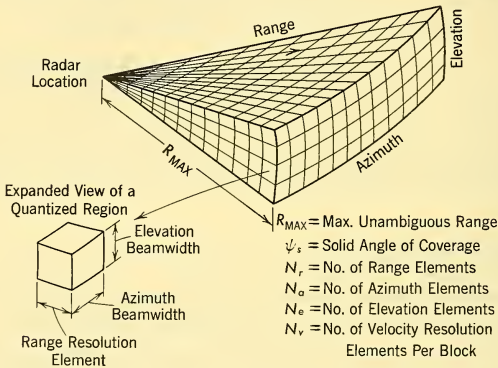


FIG. 6-1 Radar System Information Matrix.

radar is expected to detect and classify targets according to their range, their angular orientation (two dimensions), and their relative radial velocity. Depending upon the tactical problem, each dimension may be characterized by (1) a maximum and minimum value, and (2) a minimum resolution element. Thus, the total amount of information which the radar may gather is

$$N = N_r \times N_e \times N_a \times N_v \text{ elements.} \tag{6-1}$$

Generally, the tactical problem sets some limit on the time taken to gather this information. If we define this as the *total scanning time*, t_{sc} , the required interrogation rate of the radar system is

$$\dot{N} = N/t_{sc} \text{ elements/sec.} \tag{6-2}$$

Often, it is quite informative to translate a system requirement into the form of Equations 6-1 and 6-2. For example if we consider a system which requires 150 n.mi. range with 0.1 n.mi. resolution; a 1° fan beam with 360°

of angular coverage in 2 seconds; and an ability to separate targets with radial velocities of from 0 to 2000 knots in 40-knot increments, we find:

$$\dot{N} = \frac{150}{0.1} \times \frac{360}{1 \times 2} \times \frac{2000}{40} = 13.5 \times 10^2 \text{ elements/sec.}$$

which is a very large number even though the radar is providing only three dimensions — range, one angle, and velocity. This answer implies a system bandwidth requirement of at least 13.5 Mc even if it were possible (which as the reader shall soon see, it is not) to apportion this bandwidth between range, angle, and velocity coordinates in the desired manner.

In the case of a passive system the information matrix is usually only two-dimensional (angle only) since passive systems do not ordinarily have range and velocity measuring capabilities.¹

System Sampling Frequencies. The system sampling frequencies govern the minimum resolution element size and the total unambiguous measurement interval of each coordinate — range, angles, and velocities. In general, there are three basic sampling frequencies which are important in determining the character of the signal entering the receiver: (1) transmitted bandwidth, (2) transmission periodicity, and (3) angular scanning frequency.

1. *The transmitted bandwidth* Δf_t determines the rate at which the radar system can collect pieces of range information. It represents, in effect, the rate at which successive range elements of space can be interrogated. This principle is easily seen for the case of a pulse radar. In this case the transmitted bandwidth is the reciprocal of pulse length ($\Delta f_t = 1/\tau$). At any instant of time following transmission, the received pulse information originates from a range interval ΔR which has the width

$$\Delta R = \frac{c\tau}{2} = \frac{c}{2\Delta f_t}. \quad (6-3)$$

Thus every τ seconds, information is received from a new range interval.

The same principles hold whether the transmitted bandwidth is created by pulsing or by other means such as "FM-ing." That is, the minimum range resolution element is defined by Equation 6-3 and the rate \dot{N}_r at which the system collects pieces of range information is:

$$\dot{N}_r = \Delta f_t. \quad (6-4)$$

If the transmitted bandwidth is large relative to the maximum doppler shift and all the other sampling frequencies, then the transmitted bandwidth also defines the maximum rate at which the radar may collect all

¹A passive system designed to collect and classify radiation sources according to their frequency, bandwidth, polarization, and angular location can encounter bandwidth problems similar to those of an active three- or four-dimensional radar system.

kinds of information, regardless of how it might be split up between range, angular, and velocity coordinates:

$$\text{(Total interrogation rate)} \dot{N} \doteq \dot{N}_r = \Delta f_t. \quad (6-5)$$

2. *The transmission periodicity* f_r is defined as the fundamental repetition frequency of the radar signal. (In a pulse radar, for example, f_r would be equal to the pulse-repetition frequency or PRF.) This quantity governs the radar's ability to split up information between range, angle, and velocity coordinates.

The transmission periodicity defines the maximum unambiguous range interval as follows:

$$R_{\max} = \frac{c}{2f_r}. \quad (6-6)$$

Target returns from greater ranges will be ambiguous because they will enter the receiver after the transmitter has begun another transmission cycle.

The total number of separate range intervals covered by the radar is then

$$N_r = \frac{R_{\max}}{\Delta R} = \frac{\Delta f_t}{f_r}. \quad (6-7)$$

The transmission periodicity also defines the maximum relative velocity interval that may be measured without ambiguity. This may be derived as

$$(\Delta V)r_{\max} = \frac{f_r \lambda}{2} \quad (6-8)$$

where $(\Delta V)r_{\max}$ = maximum relative velocity interval (cm/sec)

λ = wavelength (cm).

For higher velocities, the total doppler spread becomes higher than the sampling frequency f_r . Thus the sampling process will create measurement ambiguities regardless of how the total doppler spread is split between opening and closing velocities.

3. *The angular scanning frequency* is defined as the rate \dot{N}_a at which signal information is collected from separate portions of the angular space volume. This quantity defines the minimum possible bandwidth of any received signal and it may be expressed:

$$\dot{N}_a = f_s = \frac{\psi_s}{\psi_a t_{sc}} \quad (6-9)$$

where ψ_s = total solid angle scanned

ψ_a = solid angle subtended by antenna pattern or instantaneous field of view.

The transmission periodicity f_r places an upper limit on the scanning frequency. If, for example, $f_s > f_r$, then radar returns from the far end of the unambiguous range interval will not be received because the radar antenna will have moved to a new angular position by the time the signal has returned. For this case

$$R_{\max} = \frac{c}{2f_r}, \quad f_s > f_r. \quad (6-10)$$

Since the angular scanning frequency defines the minimum signal bandwidth, it also defines the minimum possible velocity resolution element. Thus

$$\Delta V \geq \frac{f_s \lambda}{2}. \quad (6-11)$$

where ΔV = velocity resolution element, and the total number of separate unambiguous velocity elements is, from Equations 6-8 and 6-11,

$$N_v = \frac{(\Delta V)r_{\max}}{\Delta V} \leq \frac{f_r}{f_s}. \quad (6-12)$$

For passive radar and IR systems, the scanning frequency is the basic system sampling frequency. The interrogation rate of such systems is roughly equal to the scanning bandwidth.

Detection and Data Processing. The detection and data-processing system of any radar may be characterized by several basic properties: (1) *type of detection process*, i.e., coherent or noncoherent; (2) *number of channels*; (3) *filtering techniques*; (4) *signal-to-noise ratio* as a function of target size, range, etc. A knowledge of these properties can provide a ready means for estimating the performance potential of any system. Each of these properties will be discussed briefly; subsequent sections of this chapter will provide illustrations of the various possibilities for the generic systems.

In a *noncoherent* detection process, no attempt is made to correlate the phases of the transmitted and received signals. Thus the signal returns from each target must be detected separately and added together after detection (postdetection integration) as shown in Fig. 6-2. As was shown in Paragraph 3-3, this process improves the S/N ratio of the target return by a factor which may be expressed as

$$\Delta(S/N) \doteq n^k \quad (6-13)$$

where n = number of samples integrated; $n \doteq f_r/f_s$ if all the samples in one scan period are integrated

$K = 0.5 - 1.0$ (0.5 for low S/N and 1.0 for high S/N).

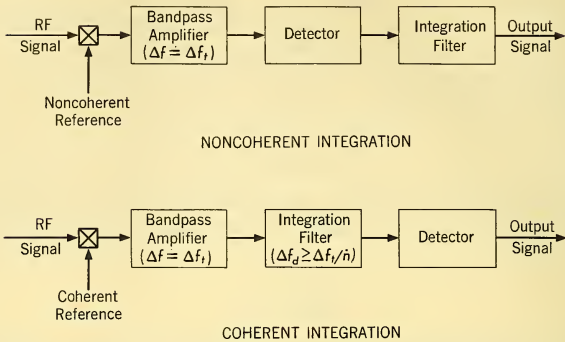


FIG. 6-2 Noncoherent and Coherent Integration Processes.

Thus, noncoherent detection does not make optimum use of system information redundancy; in fact, in the region of low S/N ratio, the S/N ratio improves approximately as the square root of the number of samples integrated.

In a *coherent* detection process, the phase relations between the transmitted and received signals are maintained. This makes it possible to add successive samples before detection to obtain a direct enhancement of S/N ratio (predetection or coherent integration). In this case improvement in S/N ratio may be expressed

$$\Delta(S/N) \doteq n. \quad (6-14)$$

Thus, in a coherent system, the S/N ratio can increase linearly with information redundancy. This means that all other things being equal (average power, frequency, antenna aperture, etc.) a coherent detection system can obtain longer detection ranges than a noncoherent system, the difference between the two being particularly noticeable for high information redundancy. In addition, as will be shown in Paragraph 6-4, a coherent system can employ a more efficient detection law than a noncoherent system, thereby enhancing the relative detection capability of coherent systems even for short observation times. These characteristics coupled with the doppler frequency measurement ability of coherent systems (see Paragraphs 6-4, 6-5, and 6-6) has resulted in a significant shift of development emphasis to coherent systems in recent years.

The *number of channels* required in a radar system depends upon the detection bandwidth and the scanning time. The basic relationships may be ascertained by considering a radar which is designed to measure range,

angles, and velocity. For such a system the information rate may be expressed (from Equations 6-1, 6-2, 6-9, and 6-12):

$$\dot{N} = N_r \times N_c \times N_a \times N_v = \frac{\Delta f_t}{f_r} \times f_s \times N_v. \quad (6-15)$$

The minimum detection bandwidth f_d that could be employed with such a system is of the order of the bandwidth induced by scanning f_d as previously mentioned. Thus the number of parallel channels needed to process all the radar data in minimum time is

$$\dot{N}/f_d \doteq \dot{N}/f_s = \frac{\Delta f_t}{f_r} \times N_v = N_r \times N_v \quad \text{number of channels.} \quad (6-16)$$

This development shows that such a radar would require a separate channel for each range resolution element for a total of N_r range channels; each range channel would possess, in turn, N_v velocity channels. A representation of such a system is shown in Fig. 6-11.

The only means for reducing the number of channels required is to increase the detection bandwidth or to increase the total scanning time and employ time-sharing of the receiving channels. A noncoherent pulse radar is a good example of the first approach: in this case the predetection bandwidth is made equal to (or greater than) Δf_t and only one channel is needed.

A CW radar with a sweeping velocity gate is a good example of the second approach; in this case, the various velocity intervals are examined sequentially. This permits single-channel operation at the cost of increasing the total interrogation time by a factor equal to the number of velocity intervals, as will be explained in Paragraph 6-5.

A number of means — other than the brute force approach indicated — exist for creating parallel information channels. Principal among these are the storage techniques described in Paragraph 6-6 and the delay-line filtering techniques described in Chapter 5.

The *filtering techniques* commonly employed in radar receivers may be listed as follows: (1) mixing, (2) bandpass filtering, (3) gating, (4) demodulation, (5) clamping, (6) cross-correlation error detection, (7) comb filtering, and (8) video integration. Chapter 5 developed the basic mathematical theory of these techniques with illustrations taken from the example of a pulse radar system employing conical scan angle tracking. The basic principles developed for each of these operations do not change; thus the material developed in Chapter 5 provides a means for tracing and analyzing the flow of signal plus noise through any radar receiver. The generic systems discussed in subsequent paragraphs will provide examples of the various filtering and receiver sampling techniques as they are used in other types of systems.

The *signal-to-noise ratio* of the target information is derived from various modifications of the basic radar range equation (Equation 3-1). As the examples in the following paragraphs will show, considerable care must be taken in the derivation of an approximate expression for S/N ratio to allow for system losses and vagaries of the receiving system such as sweeping gates, filter sampling times, and postdetection filters. Three factors are basic in determining S/N ratio and these provide a convenient basis for comparing S/N performance of different systems in the same situation (i.e. same operating frequency, search volume, antenna size, and scan speed). They are (1) average power, (2) type of integration (coherent or non-coherent), and (3) effective integration time.

Information Utilization. The end use of the radar information in a given application constitutes the reference — knowledge of which must be compiled to understand the operation of any given system. The end use requirements for a given application are derived by analyses such as those shown in Chapter 2. Those examples demonstrated a number of different end-use possibilities such as (1) display of radar information for interpretation by an operator, (2) coding and transmission to a remote location, (3) weapon direction computation. Other possibilities include (1) storage by photographic techniques, (2) correlation with information from other sensors such as infrared (IR) and photographic, (3) navigation computations.

The operation of any radar system can be judged only in terms of its compatibility with a set of end-use requirements. This fact is often forgotten by people who like to categorize radar systems on an absolute basis. Such people originate statements such as “Pulse radars have no low-altitude capability” and “Doppler radars have excellent low-altitude capability.” At best, statements such as these are partial truths; at worst, they are quite wrong in certain applications. The systems designer is well advised to avoid generalizations of this sort and analyze radar systems with respect to their applicability to specific problems.

6-3 MONOPULSE ANGLE TRACKING TECHNIQUES

Angle tracking requires measurement of two quantities in a manner that is effectively continuous. These quantities are *magnitude* and *sense* of angle tracking error. As shown in Chapter 5, Fig. 5-13, this is accomplished in conventional conically scanning tracking radar by purposely generating instantaneous tracking errors, but alternating the sign of the error, and averaging to zero. The method is simple and effective, but suffers errors when the signal fluctuates in amplitude in such a manner as to increase apparent

errors of one sign while decreasing apparent errors of the opposite sign (see Paragraph 4-8). Such a phenomenon is possible because plus errors and minus errors are generated *alternately*, not simultaneously.

One obvious method of eliminating this source of error is to generate both positive and negative errors *simultaneously*. A straightforward technique for accomplishing this is to amplify the two signals from two overlapping antenna patterns separately and compare the two amplifier outputs. This technique is operable, but places severe stability requirements on the amplifiers, since relative drifts in amplifier gain produce changes in indicated correct tracking angle.

An analogous method makes use of two spaced antennas in an interferometer arrangement. Signals from the two antennas are amplified separately, with a common local oscillator for the two receivers, and relative phase is measured at intermediate frequency. In this case the phase stability requirement on the amplifiers is severe, since relative phase shifts in the two channels similarly produce changes in indicated correct tracking angle.

Instability in indicated correct tracking angle may be overcome in either the amplitude or the phase comparison approach by connecting the two antennas in phase opposition before amplification, thus requiring only one receiving channel. Direction of arrival of signals is determined as the direction in which the amplifier output is near or equal to zero when increase of signal is produced by misaligning the antenna pattern in either direction from this so-called *null point*. This technique suffers from two objectionable characteristics. When there is no tracking error, there also is no signal to indicate presence of a target; and when there is an error signal, the sense of the error is not indicated.

Monopulse radar, as its name implies, is a tracking radar that derives all its tracking error information from a single pulse and generates new and independent error information with each new pulse. In a broad sense, the simultaneous amplitude or phase comparison systems described above may be called monopulse systems. The name monopulse, however, has become restricted by common usage to still another method for generating both positive and negative errors simultaneously which overcomes the principal objections of the other systems. The method consists in so connecting the RF circuits of two antennas that both sum and difference signals are obtained simultaneously. The patterns of the two antennas overlap in the conventional way for generating tracking error information, as shown in Fig. 6-3. The sum signals from the two antennas merge the two patterns into a single lobe pattern as shown in Fig. 6-4. The difference signals produce the familiar null pattern with the sharp zero at the center, as shown in Fig. 6-5. The sum and difference signals are then amplified separately and recombined in a product detector after amplification.

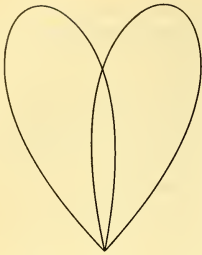


FIG. 6-3 Overlapping Individual Antenna Patterns of a Monopulse Radar.



FIG. 6-4 Sum of Overlapping Patterns in a Monopulse Radar.

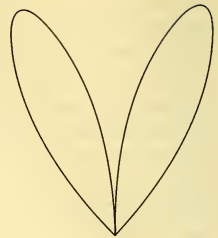


FIG. 6-5 Difference of Overlapping Antenna Patterns in a Monopulse Radar.

The process of generating sum and difference signals results from in-phase connection of the two lobes for the sum pattern, and antiphase connection for the difference pattern. Consequently the sum and difference signals are mutually in phase for directions of arrival on one side of the difference pattern null, and in antiphase for directions of arrival on the other side of the null. Thus the difference signal contains within itself only angle error signal magnitude, while the sum signal contains the phase reference by which angle error sense is determined. The output of the product detector as a function of the direction of arrival of signal energy relative to the

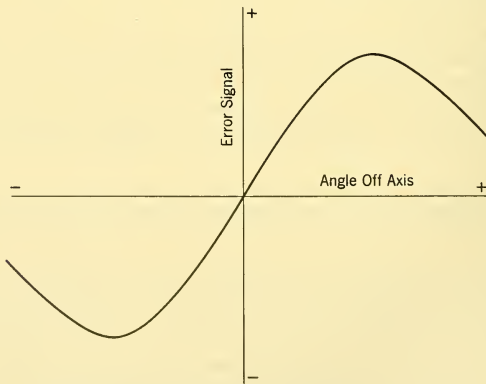


FIG. 6-6 Monopulse Error Signal Curve.

antenna difference pattern null is therefore the familiar error signal curve of Fig. 6-6, with zero signal on target, and polarity of error signals indicating error sense. The output of the sum amplifier provides indication of the presence of a target, an indication which is maximum when on target. The *balance point* representing zero angle error is not significantly affected by relative shifts in gain or phase between the two amplifiers. The *sensitivity* to angle error, represented by the slope of the error curve as it passes through zero, is influenced by relative phase shift between the two amplifiers, becoming zero at 90° relative shift. Since it is a cosine function, however, it is insensitive to phase shift near correct phase, a relative shift at 25° producing a decrease in angle error sensitivity of only 1 db.

A system diagram illustrating the monopulse principle for angle error indication is shown in Fig. 6-7. A conventional hybrid ring (see Paragraph 10-15) is used for deriving sum and difference signals from the two antennas. The transmitter is connected to both antennas by suitable TR circuitry in the sum channel, so that the transmitter radiation pattern corresponds to Fig. 6-4. The output of the sum amplifier is rectified and applied in a conventional manner as a video signal to a radar A scope, giving indication of presence of target and target range. Also the outputs of the two amplifiers are multiplied in the phase-sensitive or product detector to give an error signal whose sign corresponds to error sense. This error signal, which is a video signal, is added to the time base signal

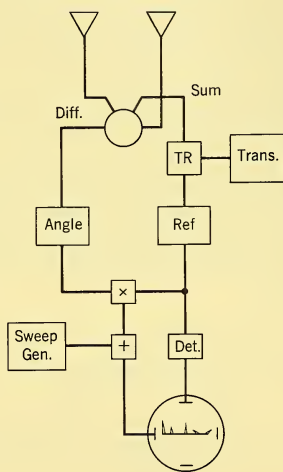


FIG. 6-7 Single-Coordinate Monopulse System.

and the combination applied to the indicator deflection system orthogonally to the output of the sum amplifier. The resulting indication presents targets as pips perpendicular to the time base line for targets which are in alignment with the antenna difference pattern null and which "lean" forward or backward from the perpendicular as angle of arrival deviates to one side or the other from the pattern null. It is apparent that the direction in which the signal pip points is related to the direction of arrival of signal energy relative to the antenna pattern null and is independent of the amplitude of the signal. The length of the signal pip indicates signal amplitude. Sensitivity of the indicator to angle of arrival is a function of relative attenuation

in the orthogonal video deflection circuits. Such an indicator has been called a "Pisa indicator" after the famous leaning tower.

To accomplish the operation described it is convenient to generate the overlapping antenna patterns with a single aperture. This may be accomplished with a single parabolic reflector illuminated by two primary radiators symmetrically displaced laterally from the focus.

The principle has been described for a single angle coordinate. Extension to two angle coordinates is not ordinarily accomplished by duplicating the system, but the one-coordinate system may be modified to operate as a two-coordinate system. It is first necessary to generate four lobes in the antenna pattern. This is accomplished by using a cluster of four primary radiators symmetrically disposed about the focus, with two up, two down, two right, and two left. Sum and difference signals are obtained separately from two pairs. The two difference signals are then added to generate error signals in one coordinate. Sum and difference signals are then obtained from the two first sums. The resulting difference signal is used to generate error signals in the other coordinate. The second or final sum signal, which is the sum of all four lobes, carries target amplitude and range information and provides a common phase reference for both coordinate error signals. The angle information may be utilized in a wide variety of configurations. Shown in Fig. 6-8 is one of the most common: an automatic angle tracking system such as might be employed in an AI radar or guided missile terminal seeker.

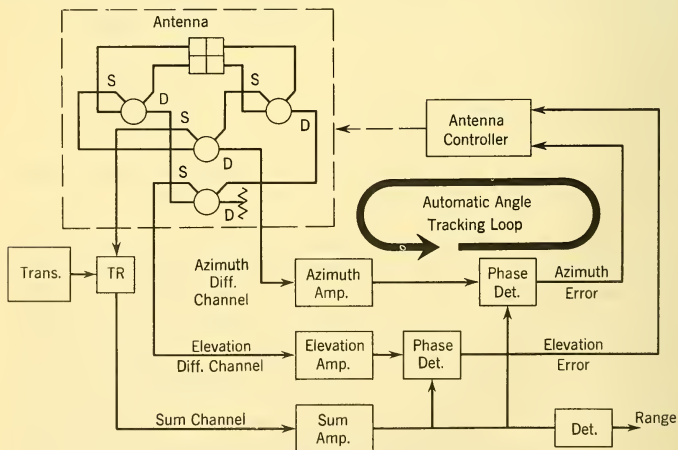


FIG. 6-8 Two-Coordinate Monopulse System.

Monopulse techniques are particularly useful for applications where pulse-to-pulse amplitude fluctuations due to target variations or interference signals can degrade conical or sequential scanning tracking techniques.

6-4 CORRELATION AND STORAGE RADAR TECHNIQUES

Signal storage has played a most significant role in the success of radar. From the earliest use of the cathode ray tube in echo ranging with A-scope presentation to modern sophisticated and complex magnetic storage devices for predetection integration, the use of storage has become increasingly important. Today the lack of high-capacity memory, high-speed operation, and wide dynamic range storage are perhaps major contributing factors impeding the development of more effective long-range radar. The increased emphasis on integration by storage has been brought about in part by the growing popularity of correlation and information theory methods for signal enhancement. The idea of correlation in itself is not new to radar — the World War II SCR-584 used a limited form of cross-correlation detection to separate the bearing and elevation errors. Here the correlation was not of the statistical nature currently in favor for signal enhancement. For this latter purpose, the cross-correlation device requires some form of storage and integration in order to fulfill its mission. Since storage can be considered a part of the correlation process, we will discuss the more general subject of correlation first.

Correlation Processes. Two correlation techniques have appeared in radar during the last two decades: (1) *autocorrelation*, defined mathematically as

$$\varphi_{11}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T f_1(t) f_1(t - \tau) dt \quad (6-17)$$

where τ is a time displacement (delay in the case of a radar echo), and (2) *cross correlation*, defined as

$$\varphi_{12}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T f_1(t) f_2(t - \tau) dt. \quad (6-18)$$

Both of these techniques are defined in the time domain and exist theoretically only in the limit as the total observation time becomes infinite. In practice, of course, infinite time is not available, and it becomes necessary to reinterpret the functions using finite limits. Let us define an incomplete autocorrelation function as

$$\varphi_{11}(\tau, T) = \frac{1}{2T} \int_{-T}^T f_1(t) f_1(t - \tau) dt \quad (6-19)$$

and an incomplete cross-correlation function as

$$\varphi_{12}(\tau, T) = \frac{1}{2T} \int_{-T}^T f_1(t) f_2(t - \tau) dt \quad (6-20)$$

where $2T$ is now a finite *observation or integration time*.

Autocorrelation of a limited nature, specifically for $\tau = 0$, has been used in conventional radar systems almost since the invention of radar. As can be seen from a study of Equation 6-19, the incomplete autocorrelation function as applied to radar for $\tau = 0$ consists of (1) obtaining the instantaneous echo power $f_1^2(t)$, (2) integrating or summing for a finite time $2T$, and finally (3) dividing by the period $2T$, thus forming an average power. These three steps can be seen to be essentially equivalent to the conventional frequency-domain radar processes wherein a square-law second detector converts the echo into instantaneous power and some type of storage provides the required averaging. In early radar sets for echo ranging the averaging was performed aurally by the operator or visually using A-scope presentation. Later the plan position indicator (PPI) used cathode ray tube persistence plus the operator for storage. Finally the use of more sophisticated video integration was adopted.

The relative merits of autocorrelation ($\tau = 0$) and square-law detection versus cross-correlation detection have been studied² with the results shown in Fig. 6-9. The output-versus-input mean power signal-to-noise ratios are plotted for a bandwidth reduction of 2 : 1 (in going from IF to video, for example), a practical value for echo ranging where the pulses must be retained. These curves apply to a single pulse where there is no integration. Signal enhancement resulting from the integration of pulses is discussed subsequently. It is interesting to note that autocorrelation can be thought of as comparable to postdetection bandwidth reduction, whereas cross correlation is comparable to predetection bandwidth reduction. In Fig. 6-9 there is an apparent threshold in the autocorrelation and square-law detection curve starting in the neighborhood of unity signal-to-noise ratio. This threshold is noted by the change from a linear to a square-law relationship between output and input sensitivity. Such a threshold does not exist in cross correlation, where a noise-free reference is used.

Cross-Correlation Radar. As soon as the signal enhancement capability of statistical cross correlation was recognized, applications to radar were considered. In order to obtain the maximum advantage from the process, one of the functions in Equation 6-20 must be noise free. A study of $\varphi_{12}(\tau, T)$ reveals that if $f_1(t)$ is the delayed target echo which contains desired information as well as unwanted noise, then $f_2(t - \tau)$ should be a noise-free reference signal possessing characteristics identical

²Samuel F. George, *Time Domain Correlation Detectors vs Conventional Frequency Domain Detectors*, NRL Report 4332, May 3, 1954.

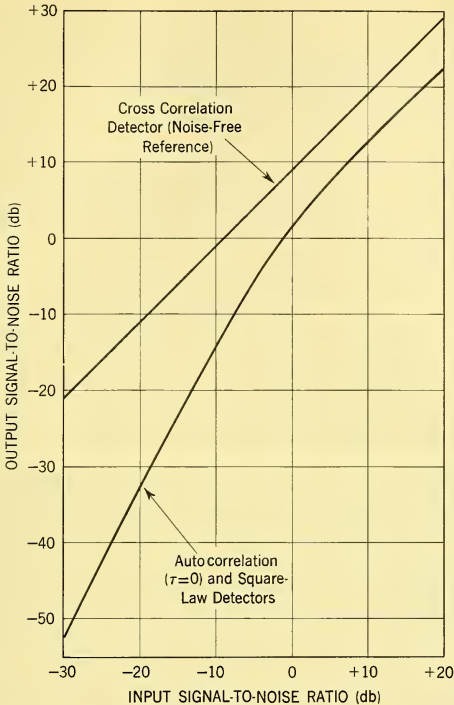


FIG. 6-9 Comparison of Autocorrelation and Cross Correlation.

with those of the signal component in $f_1(t)$ and permitting a variable delay τ to match the echo delay, thereby indicating range. This would be a specific application to the radar problem of echo ranging or the measurement of delay. A simplified block diagram of a cross-correlation detector is shown in Fig. 6-10. The output $\varphi_{12}(\tau, T)$ could be used in the same manner as the output from the second detector of a conventional radar.

Echo ranging as a major application of the cross-correlation principle to radar has been studied comprehensively by Woodward³ in the light of information theory. Woodward shows that in order to extract the most information about the exact target range from a received radar echo in additive Gaussian noise, the optimum receiver is one which forms the

³P. M. Woodward, *Probability and Information Theory, with Applications to Radar*, McGraw-Hill Book Co., Inc., New York, 1953.

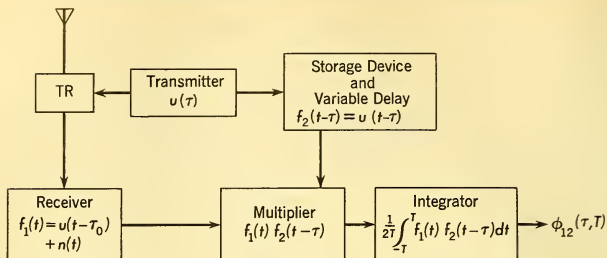


FIG. 6-10 Simplified Cross-Correlation Detector.

incomplete cross-correlation function. The problem of signal detectability has been very exhaustively studied⁴ and reported in 1954 at the MIT Symposium on Information Theory⁵. One conclusion is that for the case of a known signal operating through white Gaussian noise, the cross-correlation receiver is optimum. This result is based upon the *likelihood ratio* criterion.

Cross correlation has become very useful in extracting the doppler frequency shift or range-rate information for moving targets, thus adding a new method to aid in target detectability as well as in more accurate tracking and multiple target resolution. In order to extract the doppler, the incoming echoes must be processed so as to permit coherent integration⁶. This is predetection integration, which in the case of a pulse-doppler system means coherent video or IF integration. The cross-correlation principle is embodied in all of the systems proposed for using range-rate information. First, a coherent or stored noise-free reference must be available; then some storage medium is required to permit integration; and finally some form of very narrow-band doppler filtering must be employed. Fig. 6-11 shows a block diagram of a straightforward or brute-force pulse-doppler system. Here there are n range gates with m doppler filters per gate. It is readily seen that a tremendous duplication of equipment is called for unless some storage device can be placed in the system.

⁴J. Neyman and E. S. Pearson, "On the Problems of the Most Efficient Tests of Statistical Hypotheses," *Phil. Trans. Roy. Soc. London* **A231**, 289 (1933).

I. L. Davis, "On Determining the Presence of Signals in Noise," *Proc. Inst. Elec. Engrs. London* **99** (III), 45-51 (1952).

E. Reich and P. Swerling, "The Detection of a Sine Wave in Gaussian Noise," *J. Appl. Phys.* **24**, 289 (1953).

R. C. Davis, "On the Detection of Sure Signals in Noise," *J. Appl. Phys.* **25**, 76-82 (1954)

W. W. Peterson and T. G. Birdsall, *The Theory of Signal Detectability*, Electronic Defense Group, University of Michigan, Technical Report No. 13, July 1953.

⁵*Transactions of the IRE*, PGIT-4, September 1954.

⁶Bernard D. Steinberg, *Coherent Integration of Doppler Echoes in Pulse Radar*, Report #182-112-1, General Atronics Corp, April 1957.

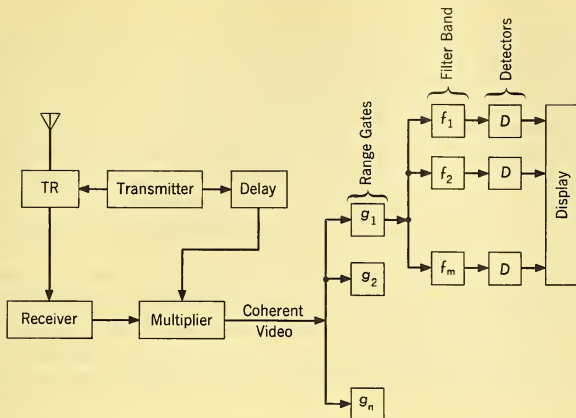


FIG. 6-11 Pulsed-Doppler System.

Storage Radar. We have noted that one of the first and perhaps simplest of all storage mechanisms consisted of a visual observer using an A scope. Next, in the PPI, screen persistence performs the storage, and scan-to-scan integration is attained. Finally, the last form of postdetection integration to become significant was of the video type. Here the storage element could vary from a video delay line to some form of electrostatic storage. A simple delay-line video integrator is illustrated in Fig. 6-12.

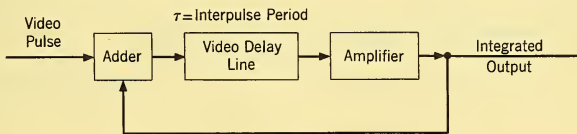


FIG. 6-12 Simplified Delay-Line Video Integrator.

The number of pulses which can be effectively integrated to improve the signal-to-noise ratio depends upon the delay line loss, feedback circuit stability, distortion, and the length of time the target remains essentially at a fixed range. Video integration of pulses embedded in additive Gaussian noise at the radar input improves the signal-to-noise power ratio by the number of pulses added; the total improvement possible after detection, then, is limited by the observation or integration time permissible. From Fig. 6-9 it is seen that there is a detector loss for any detector input below threshold ($S/N \cong 1$) and that the ultimate radar sensitivity is obtained by predetection rather than video integration.

In order to effect an improvement in signal-to-noise ratio by predetection integration, some technique must be used either to obtain signal coherence at IF or video, or to ensure transmitted signal recognition for cross-correlation purposes. The earliest methods employed coherent sources for transmission or coherent local oscillators in reception. As the transmitted signals became more complex and sophisticated, it was considered necessary to use storage to retain an exact replica for cross-correlation purposes. Both electrostatic storage and magnetic-tape storage have been developed and used successfully. The limited dynamic range of electrostatic storage combined with relatively short storage times, and the relatively slow accessibility of the data on magnetic tape have created an interest in magnetic-drum storage. As advanced technical progress provides increased dynamic range and higher frequency operation, the tremendous data handling capacity combined with high-speed record and readout and long-duration storage make magnetic-drum storage appear very desirable for radar use.

Fig. 6-13 shows a system using transmitted-waveform storage to provide a reference for the cross correlation and magnetic-drum storage to reduce

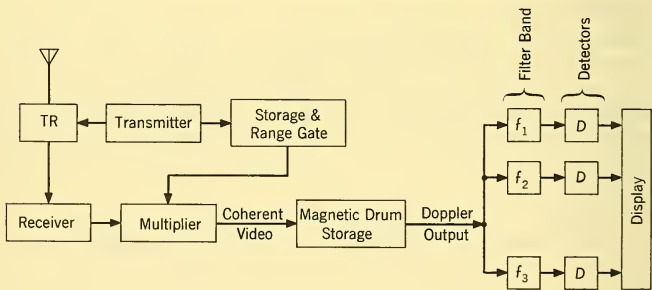


FIG. 6-13 Storage Radar.

the equipment multiplication required in Fig. 6-11. The transmitted-waveform storage unit could be envisioned to consist of multiple delays corresponding to the n range gates of Fig. 6-11. There could be a separate track on the magnetic drum corresponding to each value of delay — i.e., to each range gate. For a given channel or track on the magnetic drum, the return echo pulses could be clipped and converted essentially to a binary code, which could then be painted sequentially so as to form the doppler signal as a modulation on the code.⁷ The doppler frequency could be explored as before by a doppler filter bank as shown. Only an elementary

⁷*Ibid.*

system has been illustrated here, and numerous ramifications become evident.

The ultimate goal in search radar using correlation and storage techniques will be achieved when the range accuracy and resolution are limited only by the transmitted bandwidth, and the range-rate accuracy and resolution are limited only by the total observation time during which the target doppler remains coherent.

6-5 FM/CW RADAR SYSTEMS

Previous discussions have dealt primarily with *pulsed* radar systems — i.e., systems where transmission and reception occur at different times. Another important class of radar systems is composed of systems that employ *continuous* transmission (CW systems). In these systems, the transmitting and receiving systems operate simultaneously rather than on a time-shared basis.

For certain applications — such as semiactive missile guidance systems — continuous-wave (CW) systems can offer important advantages, particularly with respect to high clutter rejection for moving targets, and relative simplicity.

Basic Principles of Operation. In a CW system, the transmitted and received signals are separated on a frequency basis rather than on a time basis as for a pulsed system. This is accomplished by maintaining phase coherence between transmission and reception — a process which permits the measurement of the doppler shift caused by the continual rate of change of phase in the radar reflection from a relatively moving object.

The principles of operation of a simple CW system are illustrated in Fig. 6-14. A signal at frequency f_t is transmitted. The return echo from a target moving with relative velocity V_r is shifted by the doppler effect to a new frequency $f_t + f_d$, where

$$f_d = 2f_t V_r/c. \quad (1-20)$$

Closing geometries shift the received frequency higher than the transmitted; opening geometries lower the frequency.

The transmitted and received signal frequencies are mixed to recover the doppler frequency, which is then passed through a filter whose bandpass is designed to accept the doppler frequency signals from moving targets and suppress the return signals from fixed targets such as ground clutter. (For the example in the diagram, the clutter is shown at zero frequency: a condition that would exist for a ground-based radar.) For a moving airborne radar, the clutter would also possess a doppler shift relative to the transmitted frequency. This complicates the design of the filter; however, the basic principle remains the same.

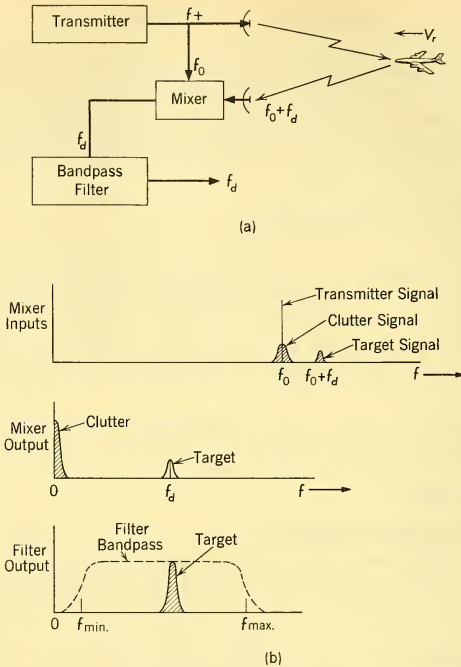


FIG. 6-14 Simple Stationary CW Doppler System: (a) Block Diagram. (b) Signal Frequency Relations.

Very often, the filter is not a single filter as shown in Fig. 6-14 but rather a series of overlapping narrow band filters — each equipped with its own detector — which covers the total desired doppler frequency band as shown in Fig. 6-15. This permits very narrow bandwidth detection, measurement of target velocity, and resolution between targets with different relative velocities within the same antenna beamwidth.

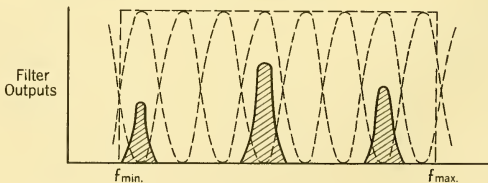
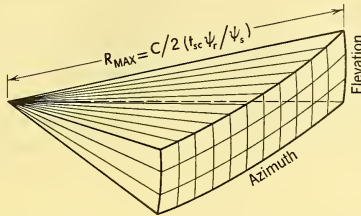


FIG. 6-15 Filter-Bank Detection Showing Contiguous Filters.

The information matrix of this system may be visualized in Fig. 6-16. The maximum possible range is limited by scanning speed; that is to say, beyond a given range, targets cannot be seen because the dwell time on the



$$N_r = 1$$

$$N_a \times N_c = \psi_r / \psi_s$$

$$N_v = \frac{\text{Total Doppler Acceptance Band}}{\text{Individual Doppler Filter Bandwidth}}$$

$$\left(\frac{N_r \times N_a \times N_c \times N_v}{t_{sc}} \right)_{\max} = \text{Total Doppler Acceptance Band}$$

FIG. 6-16 CW Radar Information Matrix.

target is less than the time required for a round trip of the radar energy. Since a simple CW system has no range-measuring capability, the maximum range is also the size of the minimum range resolution element.

Range Measurement in CW Systems. To measure target range, frequency modulation (FM) of the transmitted energy is generally used in CW systems. The maximum deviation of the transmitted signal determines the range resolution obtainable, while the frequency of the FM-ing determines the maximum unambiguous range.⁸

Consider typically a simple sine wave FM as expressed by the transmitted FM/CW waveform of Equation 6-21:

$$e_t = E_T \sin \left(\omega t + \frac{\Delta F}{f_m} \sin \omega_m t \right). \quad (6-21)$$

The received doppler signal e_r as detected by a coherent FM/CW radar is

$$e_r = E_r \cos \left[\omega_d t + \frac{2\Delta F}{f_m} \sin \pi f_m \tau \cos (\omega_m t + \pi f_m \tau) \right]. \quad (6-22)$$

where E_r is the peak received voltage, ω_d is the doppler frequency, ΔF is the peak transmitted deviation, f_m is the deviation rate, and τ is the round trip

⁸For a rough comparison with pulse radar the reader may consider the reciprocal of the bandwidth of the FM transmission to be analogous to an "effective" pulse length, while the FM rate is analogous to the PRF of a pulse radar.

transit time. To determine distance, transit time can be measured by resolving the magnitude of the returned deviation ($2\Delta F \sin \pi f_m \tau$) or by the relative phase lag of the returned modulation ($\pi f_m \tau$). In general, the greater the transmitted deviation ΔF , the greater is the resolvability of range. However, this range resolution is usually purchased at the price of increasing the minimum bandwidth of the doppler filters.

If linear FM is used such as is provided by a sawtooth or triangular waveform, the Doppler frequency is merely shifted by the amount the transmitter has deviated during the round trip transit time. The principle is illustrated in Fig. 6-17. Depending upon range, modulation, and doppler

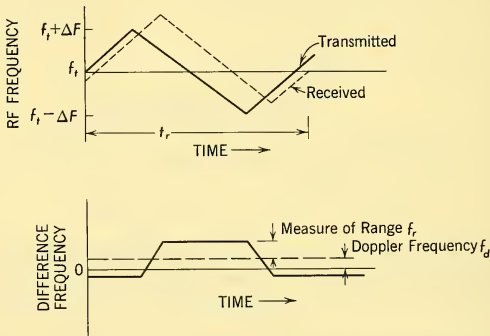


FIG. 6-17 FM Range Measuring Principle.

shift, the instantaneous received frequency will differ from the transmitted reference frequency. When the two are heterodyned, a frequency modulation is superimposed on the detected doppler signal. The magnitude of the FM deviation from the doppler is a measure of range. Care must be exercised in the selection of deviation values or the doppler frequency and range frequency may be difficult to resolve uniquely.

In some applications such as altimetry the range frequency greatly exceeds the doppler values. For example, in doppler altimetry the electromagnetic energy is radiated normal to the direction of flight so that no doppler signal results. By using triangular modulation, as in Fig. 6-17, the magnitude of the resultant detected difference frequency f_r is a measure of altitude; i.e.,

$$f_r = \frac{2\tau_r \Delta F}{c} h \tag{6-23}$$

where τ_r = modulation period in seconds

ΔF = peak transmitted deviation in cps

h = altitude in feet

c = velocity of propagation (984×10^6 ft/sec).

For more conventional radar applications using sinusoidal FM for ranging, the peak frequency deviation f_r on the doppler signal is a measure of range (see Equation 6-22); i.e.,

$$f_r = 2\Delta F \sin \pi f_m \tau \quad (6-24)$$

or

$$\approx 2\Delta F \pi f_m \left(\frac{2R}{c} \right)$$

where R = range in feet

ΔF = peak transmitted deviation in cps

f_m = deviation rate in cps.

Use can be sometimes made of the fact that the doppler frequency is precisely proportional to range rate. Integration of range rate can produce an actual measure of range provided the constant of integration can be determined.

Stability Requirements. The principal theoretical advantage of CW radar systems derives from their ability to distinguish moving targets in the presence of clutter. The maintenance of this advantage in a practical design places strict requirements on the short-term frequency stability of the transmitted energy; i.e., the coherence that exists between the transmitted and received signals. Long-term frequency drift is generally not a problem provided that short-term coherency exists.

Any modulation that tends to broaden the spectrum of a wanted signal may destroy the signal-to-noise ratio in the received gate. Any modulation of interference signals may produce sidebands which interfere with the resolution of desired signals. When interference signals, such as feedthrough or clutter, are very large, short-term stabilities in excess of 10^{-10} may be required.

The process of coherent detection of a time-delayed signal alters the angular modulation index as has been indicated by Equation 6-22.

In a *static situation* the RF reference phase can be adjusted to minimize sensitivity to angular modulations; this is impossible for a gimbaled or moving radar. Alternatively the radar detection reference could be time-delayed an amount equal to the round trip to target delay. However, this is usually impractical because of the continually changing delays of one or more targets.

From Equation 6-22, it is possible to evaluate the extent to which the angular noise sidebands of a large signal will interfere with the detection of another signal. However, it is difficult to apply Equation 6-22 in a general manner since the effective modulation index is a periodic function of the two variables f_m and τ where the maximum values of f_m are at $n/2\tau$ (where $n = 1, 3, 5, \dots$) and the zeros are at $f_m = n/\tau$ (where $n = 0, 1, 2, 3, \dots$). Of more usefulness is a consideration of the two extremes of the relationship as divided by a cutoff frequency f_c at which the returned deviation, by definition, equals the transmitted deviation. Below f_c , where the time delay is short compared to a modulation cycle, the angular indices of the detected and transmitted signals are related simply by

$$M_{fr} = M_{ft}(2\pi\tau f_m). \quad (6-25)$$

Far above f_c , where the time delay is long compared with a modulation cycle, the two indices are, on the average, equal.

For small indices of modulation such as are descriptive of useful transmitting tubes it is more convenient to discuss potential interference in terms of sideband power ratios relative to the carrier. Therefore, the ratio of the power in a single sideband P_{sb} , relative to the carrier power P_c for a detected signal, is determined (for small indices only) by:

$$\frac{P_{sb}}{P_c} = \left[\frac{M_{fr}}{2} \right]^2. \quad (6-26)$$

Typically, for a 1200-cps power supply ripple component producing 240 cps of frequency deviation, a carrier to single sideband power ratio of 100 would exist.

Random noise modulation may be considered to be composed of distributed components having a mean angular excursion per cycle of Δf . The composite mean deviation ΔF associated with a band of frequencies B cps wide is then

$$\Delta F = \Delta f \sqrt{B}. \quad (6-27)$$

A not uncommon composite noise deviation in a 100-kc band B for practical CW radars is 100 cps rms (ΔF), indicating about a 1/3 cps rms/cps density (Δf).

Fig. 6-18 indicates typical signal power levels which might be present in a hypothetical FM/CW radar.

FM/CW Airborne Radar Systems Applications. FM/CW doppler systems are most commonly employed for applications requiring high clutter rejection and a relatively low range information rate. AI radars, missile seekers, and altimeters are good examples of such applications.

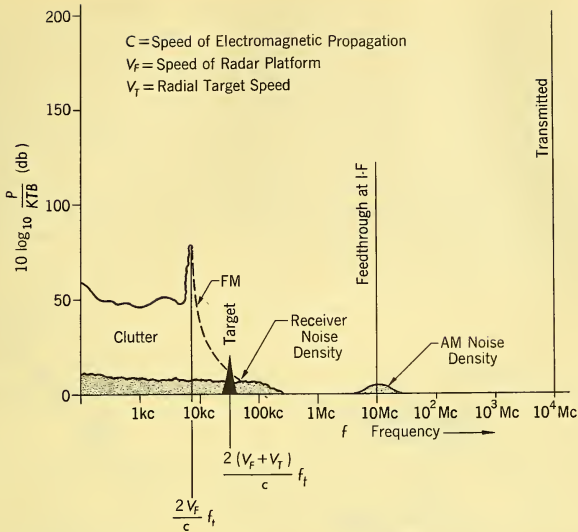


FIG. 6-18 Typical Signal Levels in an Airborne FM/CW Radar.

A principal advantage of CW doppler systems is their simplicity, when compared with other means for obtaining high clutter rejection such as pulsed doppler and coherent AMTI systems. Provided that some of the limitations to be discussed below do not seriously limit tactical utility, an FM/CW system offers a lightweight and potentially reliable answer to many airborne radar system problems.

The use of doppler techniques places several constraints upon the tactical usage of an airborne weapons system. (1) There are approach aspects where the target doppler frequency can be zero, or where the target doppler frequency falls within the clutter spectrum. These conditions lead to "blind regions" and regions of poor signal-to-noise ratio which must receive careful consideration and analysis in the overall system design. For example, approaches in the rear hemisphere of a target can be degraded by these considerations. Fortunately, the most interesting approach region from many tactical standpoints — forward hemisphere or head-on — is a region where doppler sensing devices are most effective in detecting and tracking targets in heavy ground clutter.

(2) *Duplexed-active operation* (transmission and reception through a common antenna) is generally impractical in FM/CW doppler equipment

because of excessive *feedthrough* of the transmitter energy directly into the receiver. By physically spacing the transmitting from the receiving antenna on a common radar vehicle (*spaced-active operation*), the isolation problem becomes resolvable at the cost, however, of degradation in the vehicle's aerodynamic profile. Semiactive operation involves transmission and reception on separate radar platforms, which also minimizes the feed-through problem. The main advantages of semiactive operation in homing missilery are that the transmitting hardware is deleted as an internal missile requirement and a greater on-target illumination power density is practical from the large parent radar. Both a spaced-active and a semiactive system are illustrated in Fig. 6-19.

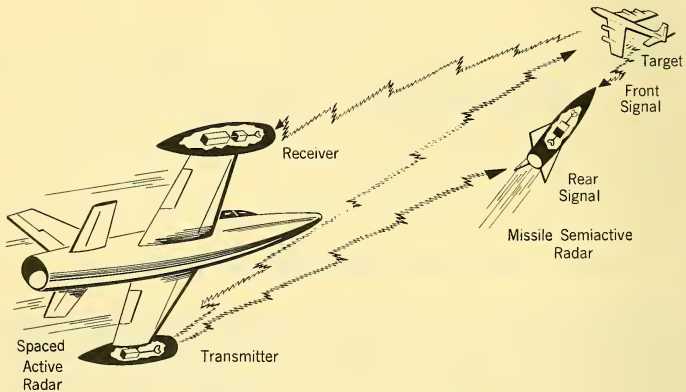


FIG. 6-19 FM/CW Airborne Radar Systems.

(3) The ranging accuracy obtainable with an FM/CW set designed for high ground clutter rejection may be relatively coarse — perhaps of the order of 1–2 n. mi. This is not usually a serious drawback for guided missile applications, although it is an important limitation for fire-control systems employing unguided weapons.

FM/CW Radar Performance. The detectability of the target is determined, for a given false-alarm rate, by the signal-to-noise ratio after final detection. Neglecting the effects of clutter and transmitter modulations, the signal-to-noise ratio may be calculated by approximate modification of the previously derived radar range formula (see Equation 3-9):

$$S/N = \frac{PG_{tr}G_{rec}\lambda^2\sigma}{(4\pi)^3R^4FKTB} \quad (6-28)$$

where P = transmitted average CW power

B = doppler filter bandwidth.

The choice of a detection bandwidth B is governed by a number of considerations derived from the tactical problem and from the realities of radar design practice.

The spectral composition of a modern-type aircraft radar reflection is seldom more than a few cycles wide when it is caused by target characteristics alone. A CW radar, transmitting a truly unmodulated wave, produces the most elementary moving target spectrum. The possibility exists, therefore, of detecting a radar signal as much as 190 db below a watt in about a second, using simple, very narrow band filters matched to the signal waveform. However, a number of practical and tactical considerations usually limit the full exploitation of this potential.

To search the frequency spectrum of expected dopplers requires a series of adjacent filters or one or more individual filters scanning the spectrum. The minimum allowable time on target is that time required for energy to build up in one of a fixed bank of filters; scanning filters increases this minimum by the ratio of scanned to actual bandwidths. Prior information as to target velocity or bearing can reduce the spectrum or area to be searched and thus increase the probability of detection in a given situation.

One practical constraint on exploiting very narrow bandwidths is the shift in the doppler spectrum caused by target maneuvers. This effect can cause the signal to be greatly attenuated if the target doppler transits the filter bandpass range before the filter has time to build up. To avoid this situation, the filter bandpass must satisfy the relation

$$B > \frac{2a}{\lambda}^{1/2} \quad (6-29)$$

where a = acceleration in ft/sec²

λ = RF wavelength in ft.

Another usual requirement in a radar system is to yield additional radar target parameters. Information theory advises that to process more data, more bandwidth and/or time is required. To resolve target range by FM on CW or to resolve bearing by AM will in general increase the minimum radar bandwidth requirements. For example, angular scanning will broaden the return spectrum bandwidth by the scanning frequency modulation, which may be of the order of 50 cycles for a typical airborne radar. Alternatively, the use of very low information rates demands an amount of time that may be unacceptable.

An economically attractive technique is to employ one sweeping filter of a bandwidth such as to satisfy all system requirements. Conventional AFC

methods can maintain the target signal in the velocity gate after lock-on. By allowing adequate time for the energy to build up in the gate,⁹ the maximum allowable sweep rate is proportional to the bandwidth squared. It is this "squared" sweep constraint that makes sweeping *very narrow band* filters very time consuming. As the filter bandwidth approaches the scanning frequency (or the reciprocal of the time on target), fixed filters become mandatory (see Fig. 6-15).

From a performance standpoint, the behavior of the system in the presence of multiple targets is most important. Multiple radar targets include terrain and weather returns as well as reflections from man-made objects; for an airborne radar, each radar target will have a finite radial velocity with respect to the radar platform. Because of their physical size, man-made objects are usually point source targets, whereas clutter is most always angularly expansive.

In the presence of a hypothetical hemisphere of homogeneously reflective clutter, the clutter doppler amplitude versus frequency spectrum would be similar to the integrated transmit-receive antenna radiation pattern versus angle as viewed at the receiver. A typical CW radar doppler clutter spectrum is shown in Fig. 6-18. Note the complete absence of interference at all frequencies above own-speed dopplers; only nose-aspect closing targets appear in this uncontaminated spectrum. In practice the clutter spectrum may be smeared somewhat by noise modulation on the transmitted energy. Reasonable prediction of specific clutter is possible, but there are a myriad of possible situations. Clutter can assume staggering proportions; a terrain return 100 db above the minimum detectable signal power is not inconceivable. Ultrahigh linearity in a receiving system is obviously required to avoid generating the additional interference of distortion products. Most radar systems are advisedly employed in a manner avoiding the most adverse clutter conditions.

With some performance compromises, a practical degree of automaticity can be achieved in sorting multiple-target data in an FM/CW airborne radar. Doppler filtering can provide the necessary resolution to distinguish numbers of targets, but the decision as to which target is wanted may require the application of considerable intelligence. With experience a human being can reduce visual or aural doppler data satisfactorily for some applications.

6-6 PULSED-DOPPLER RADAR SYSTEMS

Pulsed-doppler radar systems represent an effort to combine the clutter rejection capabilities of doppler sensing radars with the range measurement

⁹Commonly, values of from 2 to 17 time constants of the filter are used in practical applications.

and time-duplexing properties of a pulse radar. For applications requiring heavy ground clutter rejection, a common transmitting and receiving antenna, and accurate range measurement, a pulsed-doppler type of system represents the best known technical approach to the problem. However, the pulsed-doppler type of system also has certain drawbacks: principal among these are limited target-handling capacity (when compared with a pulse radar) and a high order of electronic system complexity.

Basic Principles of Operation. A simple pulsed-doppler system is shown in Fig. 6-20. It differs from the CW system of Fig. 6-14 only by the

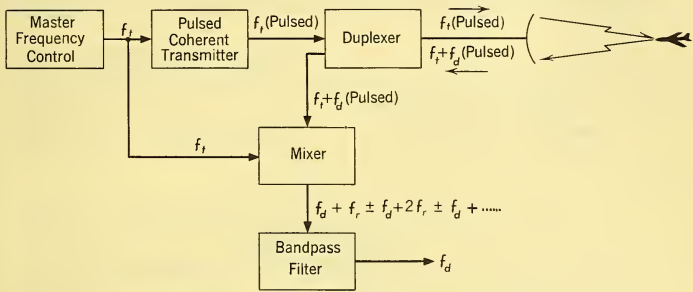


FIG. 6-20 Simple Pulsed-Doppler System.

introduction of a pulsed coherent transmitter in place of a CW coherent transmitter; and a duplexer which turns off the receiver during a pulse period and isolates the receiver from the transmitter during the interval between pulses. A master frequency control is utilized to control the carrier

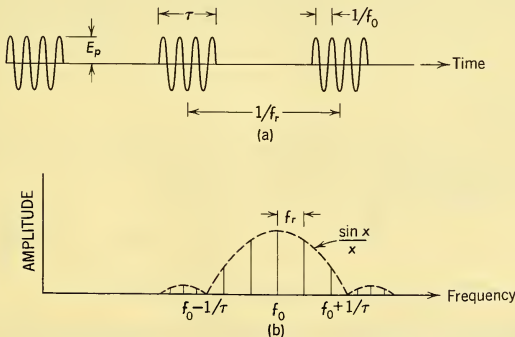


FIG. 6-21 Transmitted Pulsed-Doppler Signals.

and pulse repetition frequencies and to provide coherent references for the receiver mixing processes. The transmitted signal thus consists of an RF pulse train as shown in Fig. 6-21(a) which has the frequency spectrum shown in Fig. 6-21(b). The width of the frequency spectrum is a function of pulse length; the separation between adjacent spectral lines is equal to the pulse repetition frequency f_r .

The operation of a pulsed-doppler system can best be visualized by examining the character of the return spectra from targets and clutter at various points in the receiving system. The target and clutter spectra for a *single* spectral-line transmission from a moving platform is shown in Fig. 6-22. Because of sidelobes, the frequencies of the fixed clutter returns can

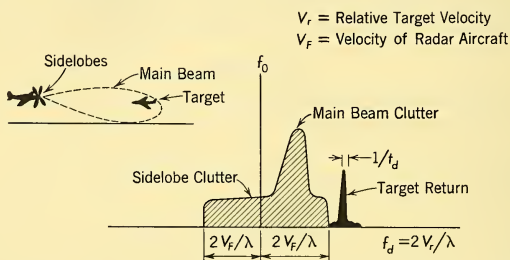


FIG. 6-22 Target and Clutter Spectra for a Single Spectral Line Transmission.

vary $\pm 2V_F/\lambda$ from the transmitted frequency. The clutter possesses a high peak resulting from clutter return in the main beam. The position of this clutter peak depends upon the angle between the antenna pointing axis and the aircraft velocity vector. Quite obviously, if the antenna is scanning, the frequencies of the clutter returns will change as functions of time; in no case, however, can the returns from fixed clutter be doppler shifted by more than $2V_F/\lambda$. For closing targets ($V_r > V_F$), the target returns will be shifted by $2V_r/\lambda$ and will therefore appear in a clutter-free portion of the frequency spectrum.

The effects of scanning and target-induced modulations (see Paragraph 4-8) cause a broadening of the target spectrum. Generally, the latter effects are small compared with the modulation induced by scanning, so that the width of the target spectrum may be expressed

$$\Delta f_{\text{target}} = \frac{1}{t_d} \quad (6-30)$$

where t_d = dwell time of the main beam on the target.

By analogy, then, the frequency spectrum of the signal and fixed clutter returns for a *pulsed* transmission consisting of an assembly of spectral lines

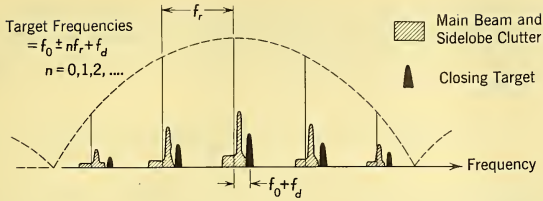


FIG. 6-23 Received Signal Frequency Spectrum for an Airborne Pulsed-Doppler System, Showing Clutter Spectrum and Return from a Closing Target.

can be represented as shown in Fig. 6-23. This diagram illustrates one of the basic limitations of a pulsed-doppler radar system: in order to maintain a clutter-free region for all closing targets of tactical interest, the spacing between spectral lines must be

$$f_r \geq 2(V_F + V_{r,\max})/\lambda \quad (6-31)$$

where $V_{r,\max}$ = maximum closing velocity dictated by tactics.

If this spacing is not maintained, some of the closing targets will be buried in the clutter from adjacent spectral lines. This consideration leads to the use of very high PRF's in pulsed-doppler systems. For example, an X-band (3.2-cm) system designed for operation in a 2000-fps aircraft against 2000-fps targets will require a minimum PRF of 112 kc — a value that is several orders of magnitude larger than the PWF's commonly used in pulsed radars.

When the return is mixed with the coherent reference signal as shown in Fig. 6-20, the output spectrum shown in Fig. 6-24 is produced. This heterodyning operation causes an effect known as *folding*; i.e. each of the

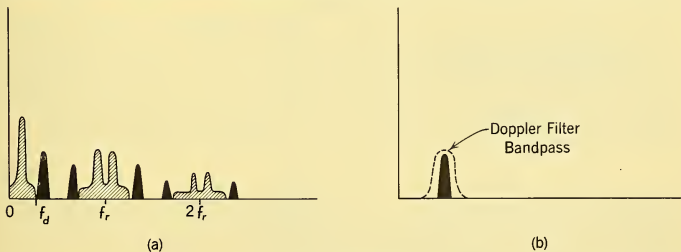


FIG. 6-24 Received Signal Spectrum After Heterodyning to Zero Frequency, Showing Effects of Spectrum Folding: (a) Video Spectrum. (b) Filtered Spectrum.

target and clutter signals produces *two* symmetrical sidebands around each PRF line. The folding effect places a further limitation on the minimum allowable PRF: considering the effects of adjacent spectral-line interference due to folding, the minimum PRF is

$$f_r \geq \frac{4V_{r,\max}}{\lambda} \quad (\text{when signal "folding" occurs}) \quad (6-32)$$

The resulting video signal then is passed through a bandpass filter to remove all the zero-frequency clutter components and all of the higher-frequency sidebands of both signal and clutter. The resulting output, then, is simply the doppler return associated with the central line of the transmitted spectrum.

The process of "folding" also doubles the thermal noise which competes with the doppler return associated with the central line of the transmission. For this reason, some form of single-sideband detection process is usually employed in pulse doppler and CW doppler systems in preference to the simpler system described here.

The doppler filter may be a single filter as shown, a bank of contiguous narrow-band filters as was shown in Fig. 6-15, or a single narrow band filter which sweeps over the total range between $f_{d,\min}$ and $f_{d,\max}$. The comments in the preceding paragraph concerning the various filter types for CW systems also apply to pulsed-doppler systems.

The simple pulsed-doppler system considered does not possess a range-measurement capability. Thus its information matrix and information rate are the same as shown for the simple CW system in Fig. 6-16. Actually the only reason for using a simple system of the type described is to eliminate the duplexing problem of a CW system and permit the use of a single antenna. In most other respects, this simple system is inferior to a simple CW system. Specifically, it possesses as deficiencies (1) greater complexity, and (2) less efficient use of power, since only the target power associated with the central spectral line is detected.

The second point deserves further amplification. In a CW radar of transmitted power P , the peak and average powers are equal because the duty cycle is unity. All of this power is effective for detection of the target. However, for a *pulsed* radar, the peak and average powers are related by the "duty cycle" — i.e., by the ratio d of "on" time to total time. Thus

$$P_d d = P_{ave}. \quad (6-33)$$

For pulsed doppler detection — as previously described — only the power in the central spectral line is used for detecting the target. The ratio of this power to the total power may be expressed

$$\text{Power effective for target doppler detection} = (P_{ave})d. \quad (6-34)$$

Thus, to achieve the same useful power return from the target as for a CW radar of average power P , the peak power of a simple pulsed-doppler system must be

$$P_t = P/d^2. \quad (6-35)$$

and the average power must be

$$P_{ave} = P/d. \quad (6-36)$$

The average power governs the weight of the power source and the peak power dictates the voltage breakdown requirements of the transmitting system components; thus these factors must be weighed against the duplexing difficulties of a CW system.

One of the most common applications of the simple pulsed-doppler system is for doppler navigation radar systems. This will be described in Chapter 14.

Range-Gated Pulsed-Doppler Systems. The full benefits of a pulsed-doppler system can be realized by range-gating the receiver. This technique permits range measurement with the resolution inherent in the radar pulse width and it can also improve the signal-to-noise ratio and signal-to-clutter ratio by the reciprocal of the range-gating duty cycle.

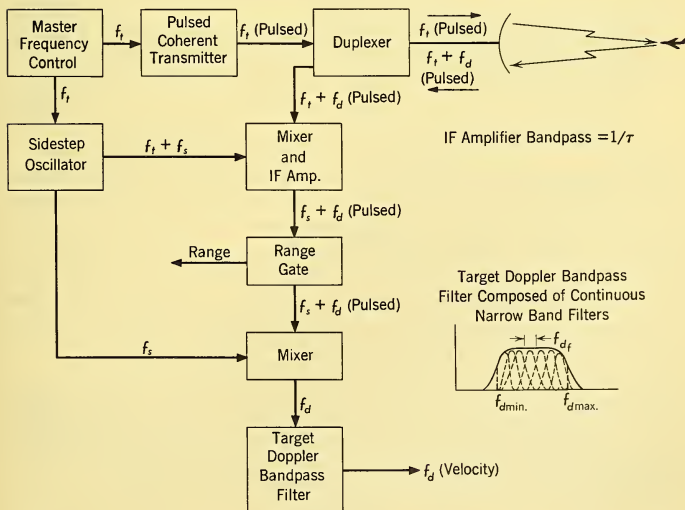


FIG. 6-25 Gated Pulsed-Doppler System with Means for Range Measurement.

A generic range-gated pulsed-doppler system is shown in Fig. 6-25. In this system, the return signal is converted to IF and passed through an amplifier with a bandwidth approximately equal to the reciprocal of the pulse length. The IF amplifier output then is "gated" before the final mixing and doppler detection takes place. The width of the gate usually is made approximately equal to the pulse length.

The operation of the gate is best understood by considering first a fixed gate which opens up the receiver at a time t_1 following transmission and closes the receiver τ_3 seconds later at $t_1 + \tau_3$. This action accomplishes the following:

1. The only returns going into the final detection stage are those from ranges falling between the values

$$\begin{aligned} Rn &= c/2(t_1 + n/f_r) \quad \text{and} \\ Rn + \Delta R &= c/2(t_1 + n/f_r + \tau) \quad n = 0, 1, 2, 3, \dots \end{aligned} \quad (6-37)$$

Clutter originates — in the main — from area extensive targets, whereas the desired signal originates from point targets. Thus the gating will improve the signal-to-clutter ratio of a target in the gate by a factor which is, on the average, equal to the duty cycle of the gate d_g , where

$$d_g = \tau_a f_r. \quad (6-38)$$

2. Noise enters the receiver only during the gating interval. Thus the average noise power is reduced by the duty cycle of the gate.
3. Since the position of the gate is known with respect to the transmitted pulse, any target doppler detected must come from a target in one of the range intervals indicated by Equation 6-35.

The improvement in signal-to-clutter ratio represents an improvement over and above what can be done with a CW radar system. Thus a range-gated pulsed-doppler system can provide greater clutter rejection than any other generic radar system type. The reduced receiver noise incident to gating tends to restore the S/N ratio to the same value as would exist for a CW system of the same average power and bandwidth. In fact, if the gate width equals the pulse length, a target in the middle of the gate would possess the same S/N as the comparable CW system.

The range measurement made by a gated pulsed-doppler system is not exactly the same as a range measurement of a pulse radar. The high PRF that must be employed in a pulsed-doppler system causes the unambiguous range interval to be relatively short compared with the maximum detection and tracking ranges. Since the maximum unambiguous range is

$$R_{\max} = \frac{80,600}{PRF} \text{ n.mi.} \quad (6-39)$$

A PRF of 112 kc, as derived in the previous example, would yield an unambiguous range interval of only 0.74 n.mi. Values of this order of magnitude are typical for airborne pulsed-doppler systems which are constrained by antenna considerations to operate in the general range of S to X band (10 cm to 3 cm). As a result, additional techniques — to be described below — must be employed to measure true range in a gated pulsed-doppler system.

Range gating also levies a cost on the system; a price must be paid in terms of system complexity and/or information rate. The previous discussion considered a single fixed gate. To cover the complete interpulse period, this gate would have to be swept. A sweeping range gate will increase the total required dwell time on the target t_{df} by the reciprocal of the gating duty factor d_g :

$$t_{dt} \approx t_{df}/d_g \quad (6-40)$$

where t_{df} = buildup time for the doppler filter.

An alternative solution is to employ contiguous fixed range gates covering the entire interpulse period (see Fig. 6-11). This "brute force" solution requires a separate doppler filtering system for each range gate interval; however, in combination with fixed contiguous doppler filters it does permit the maximum information rate to be extracted from the system because the separate doppler components of each range interval are examined simultaneously. Paragraph 6-4, Correlation and Storage Radar Techniques, suggested still another means for processing pulsed-doppler radar information.

Range Performance. The idealized range of a pulsed-doppler system may be calculated by the following modification of the basic radar range equation (3-1):

$$R_0 = \left[\frac{P_T d_s^2 G^2 \lambda^2 \sigma}{(4\pi)^3 F k T B d_g} \right]^{1/4} \quad (6-41)$$

where d_s = signal duty cycle

d_g = gating duty cycle [equal to $(1-d_s)$ for an ungated system]

B = doppler detection filter bandwidth¹⁰.

When "folding" occurs in the detection process, an additional factor of 2 is required in the denominator of the one-fourth power expression.

¹⁰In some cases, postdetection filtering will be employed to improve the final signal-to-noise ratio without increasing the number of doppler filters required. In such cases the effective detection bandwidth is $B_{EFF} \doteq \sqrt{2} B \cdot B_{PD}$ as derived in Paragraph 3-5 (Equation 3-62). In these cases, the dwell time should be matched to the bandwidth of the postdetection filter.

Eclipsing. The relatively high duty cycle of a pulsed-doppler system — typical values vary from 0.5 to 0.02 — introduces a strong possibility that part or all of the received target pulse may arrive during a transmission period. Since the receiver is turned off during transmission, target information will be lost or “eclipsed.”

The basic problem is shown for a 0.33 duty cycle pulsed doppler radar in Fig. 6-26. Eclipsing causes an effective change in the duty cycle for returns

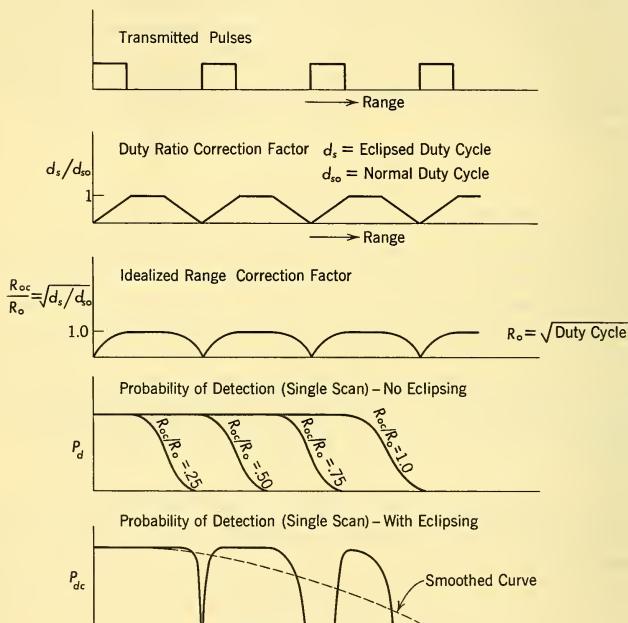


FIG. 6-26 Effect of Eclipsing on Pulsed-Doppler Blip-Scan Ratio.

which overlap transmission periods. Since average power that registers in the doppler filters is proportional to the *square* of the duty cycle, the idealized range will vary as the *square root* of the duty cycle. The blip-scan ratio is a function of idealized range as shown in the fourth figure. When the blip-scan ratio with no eclipsing is corrected for the eclipsing effect, the last curve in Fig. 6-26 results. As can be seen, the effect of eclipsing is to cause “holes” in the blip scan curve in the regions of pulse overlap. In a practical pulsed-doppler system the ratio of PRF to the useful range

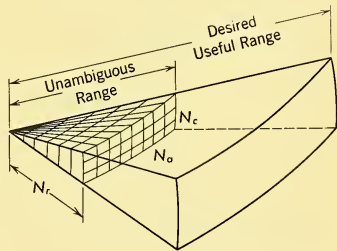
interval would be much higher; thus, there would be many more "holes" than shown in this example. For purposes of calculating the cumulative probability of detection, it is often convenient to approximate the notched blip-scan curve with a "smoothed" curve.

If the pulsed-doppler system is operated with a fixed PRF, there will be certain closing velocities which could result in the target's appearing in a detection notch on each successive scan. For example, the interpulse period of the numerical example was 0.74 n.mi. If the first detection of the target occurred at a range corresponding to a "hole," and if the target moved a multiple of 0.74 n.mi. between scans, then detection would never occur. Eventualities such as this may be largely eliminated by a slow variation of the PRF which would have the effect of producing a smoothed — but nevertheless, degraded — blip-scan curve.

Range Measurements in Pulsed-Doppler Systems.

As previously mentioned, the range gate position measurement produces an ambiguous range indication because of the high repetition frequency that must be used in a practical pulsed-doppler system. The high repetition frequency reduces the total number of separate unambiguous range intervals ($N_r = 1/\tau f_r$) and gives the pulsed-doppler radar an information matrix such as is shown in Fig. 6-27. In almost all practical cases, it is desired to operate the radar against targets at ranges far exceeding the unambiguous range interval. Thus a means must be employed to circumvent the range ambiguity problem in a range-measuring pulsed-doppler system.

There are several means for measuring true range: all are inconvenient and all degrade radar performance in terms of information rate and/or signal-to-noise ratio. One means for accomplishing range measurement is to employ the FM method used for the CW radar. The operating characteristics of this method are essentially the same as for an FM/CW radar; particularly, if the duty cycle of the pulsed-doppler system is relatively high. The range accuracy of this method is relatively poor if a narrow detection and tracking bandwidth is maintained. The range resolution is also poor because the pulse shape information is never utilized.



$$\text{Pieces of Information} = N_r \times N_c \times N_o \times N_v$$

$$N_r = 1/\tau f_r$$

$$N_c \doteq f_r/b$$

$$B = \text{Doppler Filter}$$

$$\text{Detection Bandwidth}$$

FIG. 6-27 Pulsed-Doppler Information Matrix.

$$\begin{aligned} \Delta &= \text{Transmitted Pulses} \\ | &= \text{Received Pulses} \\ \text{True Range} &= \frac{Ct_{tr}}{2} \\ t_{\max} &= \frac{2 \times \text{Desired Max. Range}}{C} \end{aligned}$$

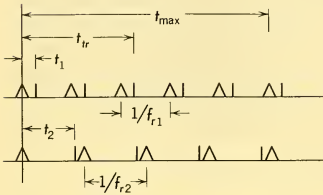


FIG. 6-28 Two-PRF Range Measurement.

A second means is to employ various modulations of the pulse repetition frequency. As an example, step switching of the PRF between several values can provide a ranging capability. The basic principle is shown in Fig. 6-28. Transmission occurs on two PRF's which are multiples of relatively prime numbers (in the example in the figure the numbers are 5 and 4). Because of ambiguities, a target at a true range corresponding to t_{tr} will appear as a target return with time delays t_1 and t_2 relative to the nearest transmitted pulse on each PRF, respectively. Thus we may write

$$t_{tr} = t_1 + \frac{n_1}{f_{r1}} = t_2 + \frac{n_2}{f_{r2}} \quad (6-42)$$

where $n_1, n_2 =$ number of unambiguous range intervals in each PRF (in the figure, $n_1 = 2$ and $n_2 = 1$).

In the example shown there are *two* possible relationships between n_1 and n_2 :

$$n_1 = n_2$$

$$n_1 - 1 = n_2$$

Substituting these relationships in Equation 6-42 we obtain, as expressions for the time delay corresponding to true range,

$$t_{tr} = \frac{t_1 f_{r1} - t_2 f_{r2}}{f_{r1} - f_{r2}} \quad (6-43)$$

or

$$t_{tr} = \frac{t_1 f_{r1} - t_2 f_{r2} + 1}{f_{r1} - f_{r2}}$$

If the first expression is negative, the second must be employed. Thus, the use of two PRF's can provide unambiguous range over a maximum desired ranging interval corresponding to the time delay

$$t_{\max} = \frac{1}{f_{r1} - f_{r2}} \quad (6-44)$$

Three or more PRF's can be used to extend the unambiguous ranging interval further. In these cases, the data processing becomes more complicated for two PRF's; however, methods similar to those used for Equations 6-42 to 44 may be used to derive the required relationships. The multiple PRF system of ranging is severely limited if more than one target return at the same doppler frequency is present. In a two-PRF system, two targets would yield four possible range values: two correct ranges and two "ghosts." Eclipsing also can cause difficulty, since it is quite likely that the target return for one of the PRF's will be eclipsed. The accuracy of this method of ranging is comparable to that of a pulse radar employing the same pulse length.

If the "looks" at the target are taken by sequentially switching the PRF from one value to another, the required dwell time on the target for the same system bandwidth is increased by the number of PRF's employed. Alternative procedures such as simultaneous transmission of the PRF's or wider bandwidth reception could be used to keep dwell time constant. However, these methods will decrease the available S/N ratio for a given amount of total average transmitter power. In addition, simultaneous transmission greatly increases systems complexity in both the transmitter and receiver and gives rise to serious eclipsing problems because of the higher effective duty ratio.

Pulsed-Doppler System Design Problems. Pulsed-doppler systems have the same basic problems of transmitter stability as CW systems. These problems are, in fact, common to any coherent system.

Because of its high duty cycle, the duplexing problem is particularly difficult in a pulsed-doppler system. To cut eclipsing losses to a minimum, the transition from transmit to receive must be made as quickly as possible. Ordinary transmit-receive (TR) tubes are not satisfactory for this application; however, ferrite circulators (see Paragraph 10-16) and ferrite switches have found considerable application because of their low insertion losses (0.5 db) and their very rapid recovery time.

In the receiving system, particular care must be taken to provide sufficient dynamic range to accommodate the maximum clutter amplitudes.¹¹ Linearity must be maintained over this range to avoid intermodulation products which spread the signal spectrum and cause loss of signal-to-noise ratio at the doppler filters.

The design of the doppler filtering system — particularly, the bandpass characteristic and the maintenance of proper frequency spacing between

¹¹In many designs, the clutter from the main beam and the altitude line is eliminated prior to amplification and doppler filtering. This greatly reduces the dynamic range requirements of subsequent stages of the receiver; however, it also makes the system completely "blind" at these frequencies.

filters — is a vital design consideration. A fixed range gating, fixed filter bank pulsed-doppler system may have hundreds or even thousands of these narrow band filters; thus the trade-off between filter performance and size and weight is a vital consideration.

Angle tracking poses certain special problems in a pulsed-doppler radar. The doppler frequency as well as the range must be tracked prior to angle lock-on. The bandwidth of the velocity loop corresponds to the width of the doppler filter. If conical scanning is employed, this filter must be wide enough to transmit the scan modulation sidebands. Actually, the doppler filter width should be about three times the scan rate in order to minimize phase and amplitude variations of the error signal. For example, a 40-cps scanning frequency would require a doppler filter band width of at least 120 cps.

The use of monopulse angle tracking (see Paragraph 6-3) poses a most difficult problem in a pulsed-doppler system. The sum and the difference signals must be handled in completely separate receiver channels — each with its own mixer, amplifiers, range gates, and doppler filters. In addition to the obvious disadvantages of weight and size, the problem of maintaining the proper alignment of these channels relative to each other represents a prodigious design problem.

Pulsed-Doppler Systems Applications. As previously mentioned, pulsed-doppler systems are best employed in systems requiring substantial ground clutter rejection, a common transmitting and receiving antenna, and accurate range and/or velocity measurement.

One other characteristic of a doppler system — either CW or pulsed-doppler — also has great tactical utility. This is the automaticity potential of such systems. Detection in such systems is inherently automatic since the signal is detected by the comparison of a filter output with a preset bias. While the same thing can be done in a pulse radar system, the problem of setting a bias level is enormously more difficult because of false alarms caused by clutter. This necessitates the use of bias levels considerably higher than would be dictated by thermal noise considerations. Thus the detection performance of an automatic pulse radar system is appreciably poorer than can be obtained when a human being is used as the detection element, since the human operator can discriminate between true targets and random clutter peaks so long as the clutter does not completely obscure the target. However, a doppler radar separates closing targets from clutter; thus the bias level may be set on thermal noise considerations alone. For this reason, as well as the others mentioned, pulsed-doppler systems are particularly suited for application as AI radars and guided missile active seekers which must find and lock on to a target buried in clutter in a high closing-rate tactical situation.

Another application of pulsed-doppler systems — doppler navigation — is covered in Chapter 14. In this application, precise velocity measurement coupled with freedom from CW radar duplexing problems make the pulsed-doppler system most attractive.

6-7 HIGH RESOLUTION RADAR SYSTEMS

Certain radar applications such as fuzing and ground mapping often require very fine resolution; i.e. effective radar pulse lengths of from 0.002 to 0.2 μ sec (which correspond to range resolution elements of from 1 to 100 feet, respectively) and/or angular resolutions of the order of 0.1–10 mils. High resolution is also tactically useful for counting the number of separate targets in a given space volume. The AEW radar example of Chapter 2 discussed this basic problem. In this case, high resolution in one dimension — for example range — can provide the requisite capability. Finally, high resolution provides a means for improving signal-to-clutter ratio when the clutter originates from area extensive targets. This is shown by Equation 4-60, where the instantaneous illuminated area of ground is a direct function of pulse length and antenna beamwidth.

There are a number of means for obtaining high resolution in a radar system. Basically, all of them are variations of the following approaches to the problem:

- | | |
|----------------------------|------------------------|
| 1. Angular Resolution | 2. Range Resolution |
| (a) Large antenna aperture | (a) Short pulse length |
| (b) High frequency | (b) Wide bandwidth |
| (c) Beam sharpening | |
| (d) Doppler sensing | |

Angular Resolution. This resolution problem has already been discussed in some detail in Paragraph 3-6. There it was shown that the angular resolution element of a radar system was approximately equal to a beamwidth where the antenna beamwidth θ can be expressed

$$\theta = \frac{\lambda}{d} \text{ radians}^{12} \quad (6-45)$$

where λ and d are the wavelength and aperture size respectively in consistent units.

Increases in the antenna aperture d or the operating frequency ($f \approx 1/\lambda$) will directly increase the angular resolution capability. One limitation on the benefits of increasing antenna aperture size is worthy of mention at this point. For purposes of resolution, the pattern of an antenna has the shape

¹²For a practical antenna, a value of $\theta = 1.2\lambda/d$ radians generally is a closer approximation when the effects of nonuniform illumination of the aperture are considered.

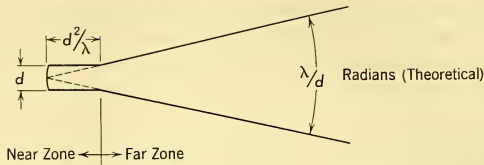


FIG. 6-29 Antenna Beamwidth Pattern.

shown in Fig. 6-29. As can be seen, the concept of angular beamwidth holds only for the so-called far zone (Fraunhofer zone) where the range R is greater than d^2/λ .¹³ At closer ranges, the effective pattern width is variable, but as a general rule it can be considered equal to the antenna dimension. Thus the resolution obtainable with an antenna aperture of d feet cannot be better than d feet regardless of what is implied by the angular beamwidth expression.

For a given antenna aperture and operating frequency, certain techniques such as monopulse and sidelobe cancellation are useful for "sharpening" the beam and thereby obtaining better definition (that is to say the transition of the signal return as the beam crosses an isolated target will be sharper). Improved resolution has been claimed from the use of these techniques. Such claims rest upon relatively shaky theoretical grounds and are based more upon the *appearance* of better resolution resulting from sharper definition than upon a rational repudiation of the basic laws governing the formation of interference patterns.

In certain cases where prior knowledge of the target characteristics exists, *velocity resolution* may be employed to give the appearance of better angular resolution than one would predict from the beamwidth. Such a case is shown in Fig. 6-30 where an antenna points straight down from an airborne

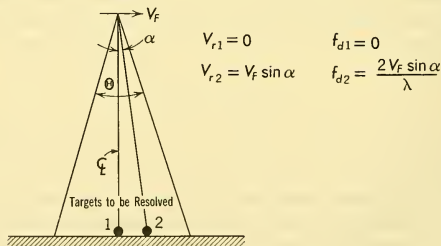


FIG. 6-30 Improvement of Apparent Angular Resolution by Doppler Filtering.

¹³See S. Silver, *Microwave Antenna Theory and Design*, Chap. 6, McGraw-Hill Book Co. Inc., 1949.

platform moving with a horizontal velocity V_F . The beam illuminates two closely spaced fixed targets, 1 and 2; however, because of the angular relation, and the velocity of the radar platform, the returns from these targets differ slightly in frequency. Thus, narrow band filtering may be employed in the receiver to distinguish between the two targets.

If a single narrow band filter of width Δf_d centered about the carrier frequency is used in the receiver, the effective beamwidth may be expressed

$$\Theta_{eff} \doteq \frac{\lambda \Delta f_d}{2V_F} \text{ radians.} \quad (6-46)$$

However, the filter bandwidth is limited by dwell time requirements to a value

$$\Delta f_d \doteq 1/t_d = V_F/h\Theta_{eff} \text{ cps.} \quad (6-47)$$

Substituting, and solving for the minimum value of Θ_{eff} , we obtain

$$\Theta_{eff} \doteq \sqrt{\lambda/2h} \text{ radians.} \quad (6-48)$$

If multiple receiver channels are used in conjunction with appropriate signal storage and correlation techniques, the return from each target can be integrated over the entire dwell time of the actual beamwidth. In this case, the minimum possible effective beamwidth becomes

$$\Theta_{eff} \doteq \lambda/2h\theta \doteq d/2h \text{ radians.} \quad (6-49)$$

and the number of channels required is

$$n_c = 2h\lambda/d^2 \text{ channels.} \quad (6-50)$$

Thus, in theory at least, the resolution performance of a very long antenna (possibly much longer than the aircraft itself) can be obtained by coherently combining signals transmitted and received from various positions along the flight path. Quite obviously, this principle could have application to ground mapping radar systems. It is of some interest to note that the effective angular resolution of the multiple channel correlation system actually improves as the actual antenna beamwidth becomes larger. The reader should also note, however, that achievement of effective beamwidths approaching the minimum possible requires a radar system of enormous complexity. For example, a 3.2-cm system with a 4-ft antenna operating at 10,000 ft altitude has a theoretical angular resolution limit of 0.1 mil. However, such a system would require the equivalent of 267 separate coherent receiver channels to realize this potential.

Short-Pulse Systems. The most obvious means for obtaining high range resolution is to employ a short pulse length in a conventional pulse radar system. However, such a system has a number of important design problems which severely limit the usefulness of this approach.

First of all, the generation of a short, high-power pulse is a difficult problem in itself. The design of the transmitting tube, the modulator, and the TR switching all are complicated by the short-pulse operation.

Short-pulse operation also limits radar performance. The required receiver bandwidth is inversely proportional to pulse length. Thus, the S/N ratio for a given value of peak transmitted power is directly proportional to the pulse length:

$$S/N \sim P_T/B \doteq P_T \tau. \quad (6-51)$$

Usually, peak power cannot be increased to compensate for this effect because of voltage breakdown limitations in the transmitter, antenna, and waveguide. Thus, for a given state of the art in RF components, the S/N performance will decrease with decreasing pulse width. Actually, because of the previously mentioned transmitter design problems, this decrease proceeds at greater than a linear rate. For these reasons, short-pulse systems are limited to relatively short-range operation (such as fuzes) or operation against targets of large cross section (ground mapping) where it is feasible to sacrifice S/N ratio for improved resolution. Short pulse lengths can also complicate certain other problems. For example, if delay line AMTI is employed, the tolerances on the pulse repetition frequency control and the delay line calibration must be held within proportionally closer limits. In addition, the bandwidth requirements of the delay-line elements are increased proportionately.

As a result of limitations such as these, there are certain tactical applications where no physically realizable noncoherent pulse radar system can provide the requisite resolution and range capabilities. To fill this gap, a family of radar systems has grown up during recent years which — for lack of any more suitable name — are called “wide bandwidth coherent systems.”

Wide Bandwidth Coherent Systems. From an information theory standpoint, the fine range resolution capability of a short-pulse system derives from the *wide bandwidth* of such a system. In fact the range resolution capability is a direct function of the bandwidth of the transmitted spectrum.

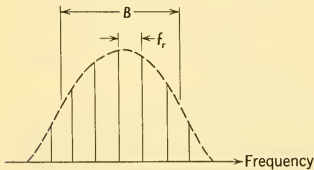


FIG. 6-31 Pulse Radar Spectrum.

This suggests that any system which employs a wide bandwidth has the inherent capability for fine range resolution. Several other observations — useful for inventing new radar systems — may also be made from an examination of the transmitted spectrum of a pulse radar as shown in Fig. 6-31.

First of all, the number of spectral lines contained in the transmitted bandwidth determines the number of individual unambiguous pieces of range information that the radar can collect from an angular volume determined by the antenna beamwidth. Thus, as previously noted,

$$N_r = \frac{B}{f_r} = \frac{1}{f_r \tau} \quad (6-52)$$

Since each piece of range information represents a range interval of $c\tau/2$, the total unambiguous range interval is simply

$$R_{\max} = N_r \times \frac{c\tau}{2} = \frac{c}{2f_r} \quad (6-53)$$

At this point it is worthwhile to recall the development of the matched filter principle presented in Paragraph 5-10 and used as the basis for the storage and correlation radar principles outlined in Paragraph 6-4. This principle stated that the optimum S/N ratio is obtained when the detection filter transfer function is the complex conjugate of the received signal spectrum. Thus for the pulse spectrum shown in Fig. 6-31, the optimum filter would have the comblike appearance of Fig. 6-32, where the width of each tooth of the comb is sufficient to pass the modulations produced by targets and scanning. The total effective detection bandwidth of such a filter may then be expressed

$$B_{\text{eff}} = N_r B_i = B \frac{B_i}{f_r} \quad (6-54)$$

Thus, obtainment of high range resolution with a *narrow band* receiving system is theoretically possible; in fact, for a non-scanning radar operating against nonfluctuating targets, the total required received bandpass approaches zero. The application of the matched filter concept requires a coherent system.

These principles make it possible to conceive a wide variety of high-resolution systems which circumvent the peak power and S/N ratio limitations of noncoherent pulsed radar systems. In general, these systems have the following common characteristics:

1. A wide transmitted bandwidth ($B_t = c/2R_{r,\min}$, where $R_{r,\min}$ = minimum resolution element).

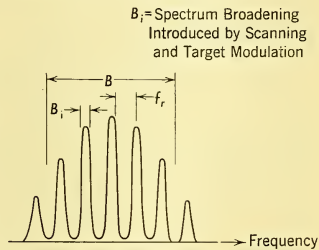


FIG. 6-32 Optimum Filter for Pulse Radar Spectrum.

2. A modulating frequency f_r sufficient to create the desired unambiguous range interval.
3. Some form of storage and cross correlation which attempts to provide an optimum match between the received signal and the effective receiver bandpass filtering characteristic.
4. Usually, great complication in comparison with a noncoherent pulse radar.

The following systems are indicated as possibilities for high resolution wide bandwidth coherent systems.

1. Wide bandwidth FM/CW.
2. Long-pulse, low-PRF systems where the transmitted frequency is FM'd during a pulse transmission period to produce the wide transmission bandwidth. This type of system is often called a *matched filter radar*.¹⁴

6-8 INFRARED SYSTEMS

A book entitled *Airborne Radar* may seem a strange place to find a discussion of infrared techniques, but it must be remembered that the applications and design principles of airborne radars and of infrared detectors and weapon control systems are quite similar. Actually the only real differences between a passive¹⁵ radar and an infrared system are (1) the method of detection¹⁶ and (2) the fact that the infrared radiation emanates from the target itself rather than from its radars or communications equipment. The tuned circuits used in the detection of radio and radar radiations cannot presently be extended to the frequencies (3×10^8 to 1.5×10^7 megacycles, or 1 to 20 microns wavelength) of that portion of the infrared having practical significance here. Therefore optical detectors must be used and these impose their own restrictions.

An advantage of the short wavelength of infrared radiation is that interference patterns have correspondingly small angular relationships and are usually not significant in instrument design or operation. For example, the diameter of an infrared collector mirror depends only on the amount of radiation to be collected and not on the required angular accuracy, as in the case of a radar dish. Also interference patterns such as result in radar from ground reflections and cause target confusion are not encountered with infrared systems. Because the target is itself a source, the signal in a passive infrared system diminishes with range more slowly than in an active system.

¹⁴Note that in Paragraph 5-10, the term *matched* has been used to describe a more general class of radar systems.

¹⁵Only passive infrared is considered here. Active systems do exist and are used, but for long-range detection and tracking no sufficiently intense sources of infrared radiation are available.

¹⁶By detection we mean the manifestation of the presence of electromagnetic radiation.

Conversely, infrared does not have the all-weather capabilities of radar, its ability to penetrate haze, fog, and clouds being only slightly better than that of visible light. Background clutter considerations are also more serious, since everything in a typical tactical environment is to some degree a source of infrared radiation, i.e., a potential source of interference. Passive infrared systems — like passive radar systems — also do not possess the capability for measuring range in the direct and convenient manner of active radar systems.

The use of infrared for detection and tracking is now new, having been vigorously exploited by Germany during World War II. In this country, where reliance was placed more heavily on the development of radar — with obvious beneficial results during the war — serious consideration of infrared systems has been more recent and stems from four facts: (1) modern targets are better sources of infrared radiation than their predecessors and in many cases poorer radar targets; (2) many important targets are encountered at high altitudes where attenuation and absorption of IR energy are minimized; (3) infrared is more difficult to countermeasure than radar, or at least the art is not so advanced; and (4) infrared technology has made significant advances since World War II.

This will be a short discussion of the application of infrared to airborne surveillance and tracking systems. The fundamentals of infrared science are ably covered in an earlier book of this series (*Guidance*, Chapter 5, "Emission, Transmission, and Detection of The Infrared") and a knowledge gained by reading that discussion will be assumed.

Basic Principles. Airborne infrared systems generally use mirrors rather than lenses. Lenses are possible but usually not practical because of limitations imposed by the properties of available materials (see Paragraph 5-7 in *Guidance*). The infrared system is composed, then, of a mirror which collects radiation from the target and focuses it on the detector, a means of modulating the radiation striking the detector in order to produce an a-c signal, and a means of discriminating against spurious targets and background radiation. Frequently, modulation and discrimination are accomplished in the same process.

Consider, as an illustration, the simplified system shown in Fig. 6-33.¹⁷ Radiation from the target, background, and intervening air enters through a dome of transparent material (Irdome) and is focused on the detector after reflection from the two folding mirrors. The instantaneous field of view is determined by the size of the detector and the focal length of the main collecting mirror. Scanning is accomplished by tilting the two folding

¹⁷This arrangement is chosen only to illustrate the significance of the system parameters and not for its desirability or efficacy. It is not an example of a system in actual use, since most such systems are classified and cannot be discussed here.

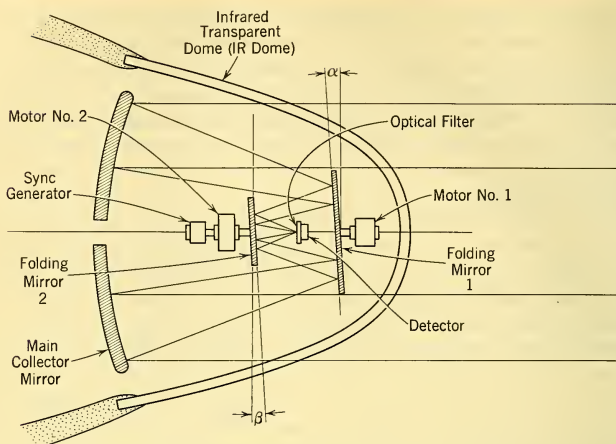


FIG. 6-33 Typical IR System.

mirrors by the angles α and β away from the perpendicular to the optic axis and then rotating them about the optic axis in opposite directions. If the mirrors turn at equal speed, the effect is to move the instantaneous field of view along a straight line, the length of which is determined by the two angles α and β and the intermirror distances. The velocity with which the line is scanned varies sinusoidally; it is most rapid in the center of the line and is slowest at the ends where scan reversal occurs. If one mirror turns a little more slowly than the other, the line scanned in space rotates slowly about its center, resulting in the rosette scan pattern shown in Fig. 6-34. With this pattern the surveillance capability is greatest in the center of the field, which is crossed on each spoke of the rosette, and diminishes toward the edges, a property which may or may not be desirable.

In considering the appearance on the scope of a scene scanned by an infrared device, it should be remembered that while our eyes see almost everything by reflected light (i.e., a "semiactive" process similar to some types of radar), the infrared scanner sees mainly thermal radiation emitted by the observed objects themselves. This is particularly true if the radiometer is filtered to be sensitive only to radiations of wavelength greater than 3 microns, since reflected or scattered sunlight beyond 3 microns is generally negligible compared to emitted radiation. Therefore, a "hot" object such as a city, or one with a high emissivity such as a cloud, will appear bright. The clear sky, bright blue to the eye, will appear black, since the air molecules do not scatter infrared as they do visible light.

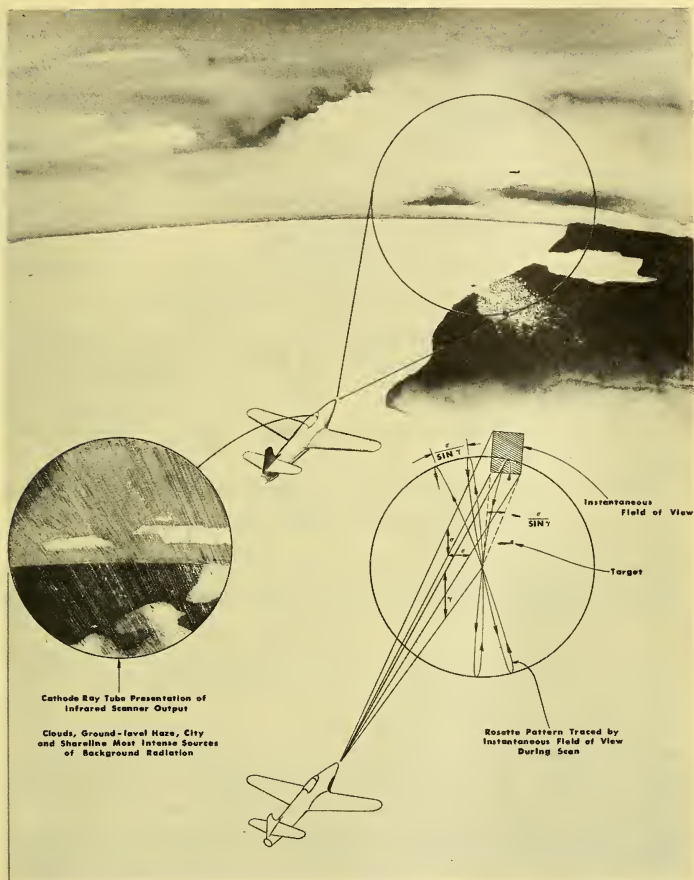


FIG. 6-34 Example of an IR Scan Pattern.

The detector receives the radiation collected by the optical system and converts this to electrical energy, which is amplified and appropriately displayed — in this case on a cathode-ray screen — or used as an error signal for tracking. There is radiation received from the target, from the clouds, haze, land, water, and sky within the scanned field, and from the atmosphere lying between the detector and the target. In some instances the

optical elements themselves, for example an imperfectly transparent dome heated in flight, contribute significant and undesirable radiation. The usable signal results from the difference between radiation collected from the target and all other sources. This "contrast" signal is proportional to:

$$V_s \sim AE \int_{\lambda_1}^{\lambda_2} \left[\frac{\mathcal{F}_{\lambda T}}{R^2} - \mathcal{F}_{\lambda B} \sigma^2 \right] S_\lambda T_\lambda d\lambda \quad (6-55)$$

where V_s = signal output (volts) from the detector

A = area (cm²) of main collector mirror

σ = field of view of radiometer (radians) = width of square detector divided by the focal length of collector mirror

E = efficiency (%) of windows, filters, mirrors, etc. in the optical system (dimensionless). This is here assumed to be independent of the wavelength

$\mathcal{F}_{\lambda T}$ = spectral distribution of target radiation (watts micron⁻¹ steradian⁻¹)

$\mathcal{F}_{\lambda B}$ = spectral distribution of background and other unwanted radiation (watts cm⁻² steradian⁻¹ micron⁻¹)

S_λ = spectral response (volts/watt) of detector

T_λ = spectral transmission (%) of the atmosphere between the target and detector (dimensionless)

λ_1, λ_2 = wavelength limits (microns) of system sensitivity defined by the optical filter or sensitivity limits of the detector.

R = target range (cm).

The detector will have a noise output V_n which will be a function of the type of detector, the bandwidth Δf of the amplification system, and the radiation environment of the detector. Specifying the minimum signal-to-noise ratio V_s/V_n required for reliable detection of a target, the system noise defines the minimum required V_s . Since $\mathcal{F}_{\lambda T}$, $\mathcal{F}_{\lambda B}$, and T_λ are beyond our control and E is always optimized anyhow, the remaining parameters are chosen to give the required V_s at the desired target range. Actually λ_1 and λ_2 are generally determined by consideration of $\mathcal{F}_{\lambda T}$, $\mathcal{F}_{\lambda B}$, and T_λ . Fig. 6-35 shows $\mathcal{F}_{\lambda T}$, $\mathcal{F}_{\lambda B}$, and T_λ plotted as a function of wavelength for a specific application: the detection of a 600°K blackbody viewed against a background of clouds or heavy haze through 10 miles of moderately clear sea level air. It is plain that the best choice of wavelength limits are λ_1 and $\lambda_2 = 3.3$ and 4.1 microns, respectively. The properties of available detectors — sensitivity, time constant, ruggedness, reliability, etc. — may, of

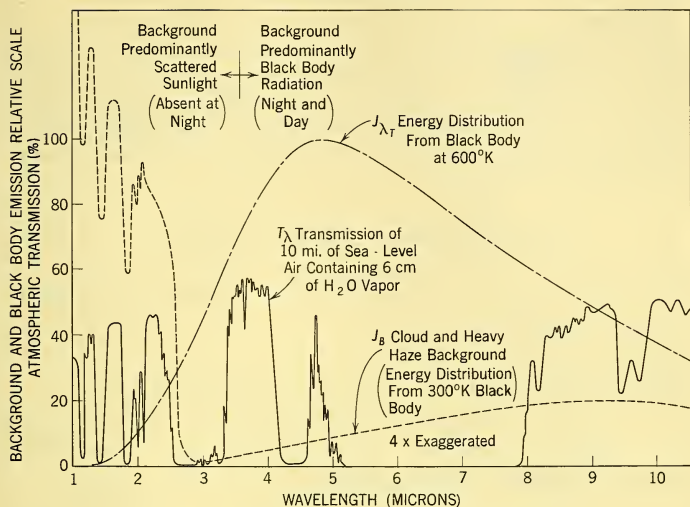


FIG. 6-35 The Useful IR-Frequency Spectrum.

course, force a different choice. The only really flexible parameters for the designer, then, are the areas of the collector mirror and of the cell. These are in turn influenced by the requirements of the scanning system.

Scanning System Characteristics. The choice of a scanning system generally represents a compromise between the requirements of the system and the mechanization advantages of rotary optics (particularly for high-speed scanning) and fixed detector elements (which simplify cooling problems and maintenance of cell sensitivity). As an example of the type of analysis which must be performed to assess a given scanning technique, the rosette scanning pattern previously discussed will be analyzed to determine the interrelations among scan time, resolution, coverage, and detection element characteristics.

The important parameters of the scanning system are: the instantaneous field of view σ (radians) square; the whole field of view which is here (see Fig. 6-34) a function of the half-scan angle γ ; and the time T required for the whole field to be covered. In order to completely cover the field the number of spokes in the rosette pattern is equal to the number of instantaneous fields required to cover the periphery of the whole field, or

$$\frac{2\pi}{\sigma / \sin \gamma} \text{ spokes.} \quad (6-56)$$

The speed of scan is controlled by the time constant τ of the detector. To define a practical upper limit we shall use the time required for the instantaneous field to move its own width σ , equal to τ . To go faster would result in considerable smearing of the display pattern, with resultant loss of resolution. In this sinusoidal scan the velocity is not constant, but for simplicity we use the average value. The time required to scan one complete spoke (2γ radians) is

$$\frac{T}{\text{Total no. of spokes}} = \frac{T\sigma}{2\pi\sin\gamma} \text{ seconds/spoke.} \quad (6-57)$$

Motor 1 then turns clockwise at $\frac{1}{T}\left(\frac{2\pi\sin\gamma}{\sigma}\right)$ rps and motor 2 turns counterclockwise at $\frac{1}{T}\left(\frac{2\pi\sin\gamma}{\sigma} - 1\right)$ rps.

Since a complete spoke is 2γ radians, the average scan rate is

$$\frac{T\sigma}{4\pi\gamma\sin\gamma} \text{ seconds/radian.} \quad (6-58)$$

and the time required to scan σ radians is

$$\frac{T\sigma^2}{4\pi\gamma\sin\gamma} = \tau \text{ seconds.} \quad (6-59)$$

which we accordingly equate to the time constant τ of the detector.

Consider an actual case in which we want:

$$\sigma = \frac{1}{3}^\circ = 0.0058 \text{ rad}$$

$$\gamma = 20^\circ = 0.35 \text{ rad}$$

$$\tau = 10^{-4} \text{ sec}$$

Then the time required to scan a complete field is, from Equation 6-57:

$$T = \frac{4\pi\gamma\sin\gamma\tau}{\sigma^2} = \frac{4\pi \times 0.35 \times 0.342 \times 10^{-4}}{(0.0058)^2} = 4.5 \text{ sec.} \quad (6-60)$$

For many practical cases, this is obviously too long. A modern aircraft will have changed course considerably in this time. Therefore the instantaneous field must be enlarged, the total field reduced, or a faster detector sought — possibly all three. Perhaps the scan mode would have to be abandoned in favor of a more economical one without the multiple retrace encountered in the center of this field — say a raster scan similar to that used in television.

Target Tracking. If the system is required also to track a target, this could be accomplished by orienting the aircraft so that the target image falls in the center of the screen, setting angle $\alpha = 0$, and reducing angle β

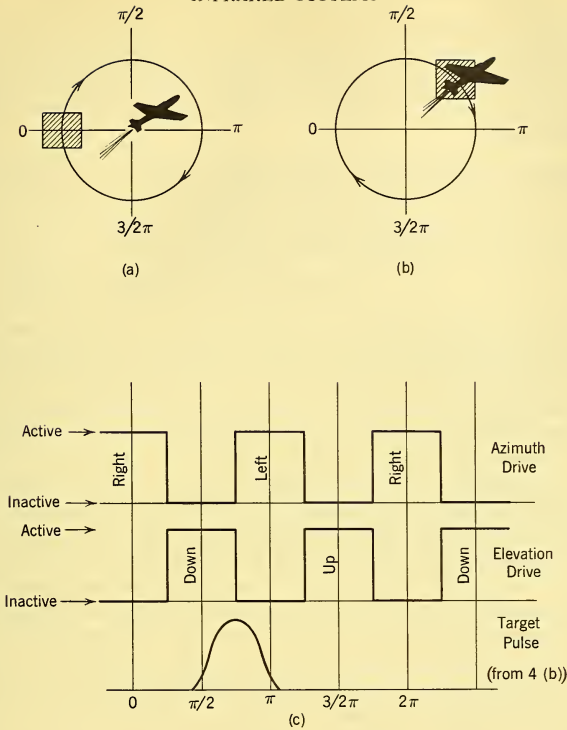


FIG. 6-36 Target Tracking.

to such a value that a circular scan of, say, $\frac{1}{2}^\circ$ results (see Fig. 6-36). Then as long as the image in space of the detector rotates around the target without touching it (Fig. 6-36a) no output (error signal) will result. When the line of sight moves and the detector then encounters the target (Fig. 6-36b) an output pulse will result; by comparing the phase of this pulse with a synchronizing signal generated on the shaft of motor 2, an error signal is generated. By having the entire optical detection device movable in the aircraft and motor controlled, these error signals can be used to keep the device pointed at the target.

In Fig. 6-36c a simplified ON-OFF control is illustrated. The synchronizing pulse alternately activates and deactivates the control motors. If the target pulse occurs during an active period, the motor moves the optical system. In Fig. 6-36c the pulse occurs where both the DOWN and LEFT

controls are activated (as in Fig. 6-36b) to return the target image to the center of the circular scan. Smoother tracking and less tendency to hunt result from a system in which the error signal varies in magnitude with the off-course position of the target, and this is usually a feature of actual tracking systems.

Detection Performance. The actual capabilities achievable with present-day infrared systems can be estimated from target intensities, background intensities, and detector sensitivities. As an example, consider a radiometer having a filter limiting the sensitivity of the radiometer to the region 1.7 to 2.7 microns and a lead sulfide detector at the focus of a collector mirror 1 ft in diameter. With a readily available detector (say an Eastman Kodak Ektron cell) of practical size (say 1 mm square), an output signal just equal to the rms noise from the detector can be achieved under tactical conditions when about 10^{-12} watt/cm² falls on the collector mirror and is focused on the detector. This would represent a signal-to-noise ratio of 1, here arbitrarily construed as a necessary criterion for detectability. If the target is the exhaust port of a typical jet engine, the irradiance (watts/cm² in the 1.7 to 2.7-micron region) at the collector mirror will be about $400/R^2$, where R is the target range. This results from assuming the exhaust port to be a 24-inch-diameter blackbody of emissivity unity and to have a temperature of 600°C. Through a completely clear atmosphere, then, and with no background interference this jet exhaust port could be seen from a distance of 2×10^7 cm or 125 miles, at which distance it would irradiate the collector mirror with the necessary 10^{-12} watt/cm². Atmospheric attenuation, which is severe in the lower atmosphere, and background interference may under average conditions degrade this range to less than a third of this number.

Further, the target we are considering, a single-engine jet aircraft, will be a much fainter target at any other than tail aspect where the exhaust port is visible. In side aspect the radiation emanates from the hot exhaust gases which, while of extended size and quite hot, emit only the wavelengths of the characteristic infrared bands of the gases. If the fuel is a hydrocarbon these are the bands of water vapor and carbon dioxide. Atmospheric attenuation is most severe in this case, since the cold water vapor and carbon dioxide in the intervening air path absorb most of what is emitted. In side aspect a jet will be less than one-tenth as intense a target as in tail aspect and will therefore be detectable at less than one-third the range realizable when looking at the exhaust port. In nose aspect, it will be considerably worse than this, since here most of the exhaust gases are hidden by the aircraft and the hot parts of the engine are not visible.

CHAPTER 7

THE RADAR RECEIVER

7-1 GENERAL DESIGN PRINCIPLES

The airborne radar receiver amplifies and filters the signals received by the radar antenna for the purpose of providing useful signals to display and automatic tracking devices. The receiver accepts all of the signals appearing at the antenna terminals and must filter them so as to provide maximum discrimination against signals which do not originate by surface reflection of the transmitted radar signal from certain desired targets.

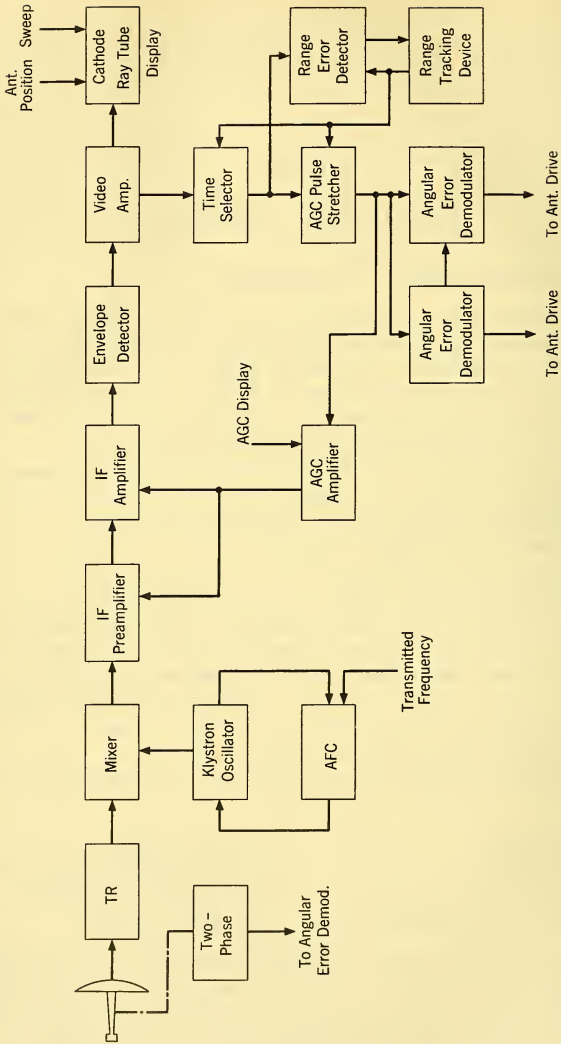
Modulation characteristics of the desired signals must be preserved in the filtering process. The modulation characteristics provide information on the number of targets, their angular position with respect to a given frame of reference, the distance between the radar set and the targets, and the velocity of the targets.

The majority of airborne radar receivers are of the superheterodyne type. Ordinary pulse radar sets are usually of the single frequency conversion type. Doppler radar sets employ single sideband reception. The receivers are more complicated than in the ordinary pulse radar set and generally employ multiple frequency conversion in order to realize the required frequency selectivity. Fig. 7-1 shows a functional block diagram of an elementary receiver of each type.

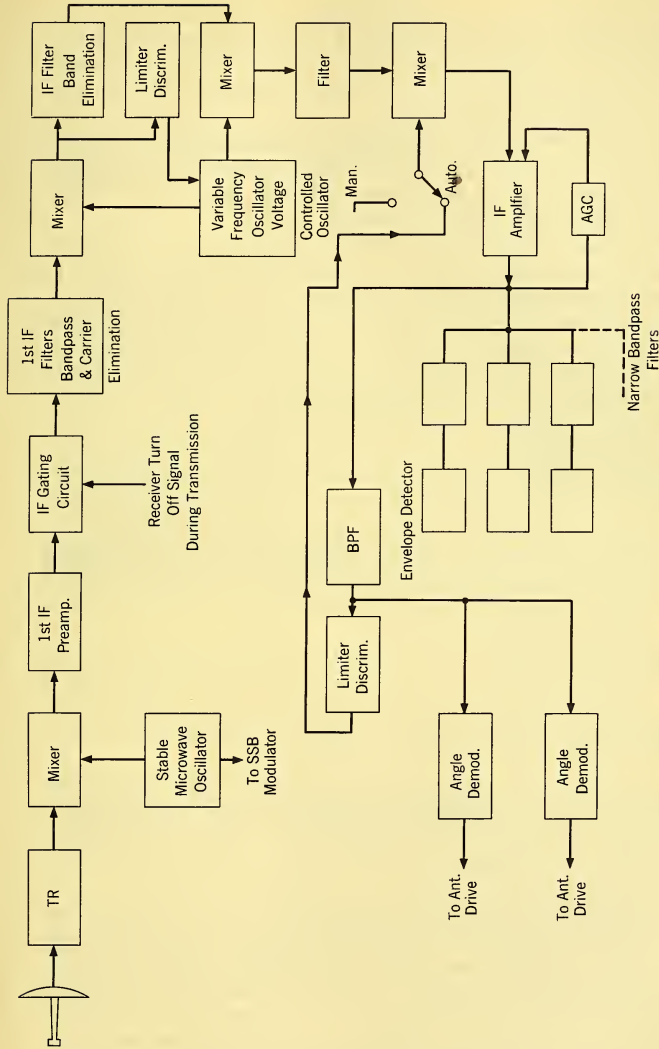
Performance of a radar receiver is described by the following characteristics:

- | | |
|---------------------------------|-------------------------------------|
| 1. Noise figure | 5. Dynamic range |
| 2. Sensitivity | 6. Cross-modulation characteristics |
| 3. Selectivity characteristics | 7. Tuning characteristics |
| 4. Gain control characteristics | 8. Spurious response |

Specification of each of these characteristics depends upon the particular radar application. Analysis of the radar system defines the input signal environment, the required output signal-to-noise ratio, and required fidelity of modulation. Such requirements are then interpreted in terms of the above characteristics. Each of these characteristics is discussed later in this chapter and in Chapter 8.



TYPICAL PULSE RADAR RECEIVER



TYPICAL PULSE DOPPLER RECEIVER WITHOUT RANGE TRACKING PROVISIONS

Fig. 7-1 Typical Receiver Configurations.

It is desirable that the *noise figure* be minimized and the *sensitivity* be maximized. This is not always feasible, as will be indicated in Paragraph 7-2. *Selectivity* is provided in both frequency and time. It is desirable to provide the required selectivity at the low-level signal stages and prior to envelope detection of the signal.

Gain control characteristics are dictated by requirements to provide error signals to range, speed, and angle tracking feedback mechanisms for a specified range of input signal power. Automatic gain control (AGC) systems are discussed in detail in Chapter 8.

Dynamic range is the range of signal levels above the thermal noise level for which a receiver will provide a normal usable output signal. To reproduce faithfully the amplitude modulation on a received signal, the *incremental gain* of a receiver whose output is controlled by the average level of the received signal must be constant for a dynamic range on the order of 12 db above the average signal level. The incremental gain is the slope of the output/input voltage characteristic of the receiver. When an undesired signal appears at the receiver input which is coincident in time with the desired signal and nearly coincident in frequency, a much greater linear dynamic range or range for which incremental gain remains constant is required. In receivers which separate signals by frequency filtering, it is not unusual to require a linear dynamic range on the order of 80 db up to the point in the receiver at which the frequency separation of the desired and undesired signals occurs. On the other hand, short-pulse, low-PRF radar sets which separate signals by time filtering may require only a linear dynamic range on the order of 15 db.

Undesired signals which occur at a different time or frequency than the desired signal may impart their modulation to the desired signal. This is called *cross modulation*. Such a phenomenon arises from nonlinearities in the receiver and is undesirable, since it degrades the output signal-to-noise ratio.

A proper radar system analysis defines the signal environment and allowable degradation of the receiver output signal; thus the principal factors governing the selection of the dynamic-range and cross-modulation characteristics are specified.

Tuning characteristics are dictated by the radar transmitter. The receiver is designed so that it can always be tuned to the transmitter frequency. The design objective is to make the receiver tuning as accurate as the state of the art permits. Both short-term and long-term frequency stability is important. The effect of short-term frequency instability is to introduce modulation on the signal in the receiver. Such modulation degrades the output signal-to-noise ratio. In a noncoherent pulse radar set, the tuning accuracy that can be achieved is on the order of 1 part in 10^5 . Much greater stability is required in coherent radar sets. Automatic

TABLE 7-1

Receiver Characteristic	Desirable Effect	Effects on Other Receiver Characteristics
Direct detection of the RF signal	Simplicity	Poor noise figure; poor rejection of spurious signals. Nonlinear transfer characteristic
Preselection (band pass filter between antenna and receiver)	Greatly reduces spurious signal response	Increased noise figure due to insertion loss of the filter. Imposes long-term frequency stability requirements on the transmitter and preselection filter that otherwise would not be encountered
Desensitization during transmitting time (TR switch)	Reduces degradation of receiver performance with time by limiting the signal power applied to the microwave mixer	Degrades performance at short ranges owing to the deionization properties of the gas switches that are employed. In high-PRF coherent radars, degrades performance in each ambiguous range interval immediately following a transmitted pulse
High IF frequency	Minimizes spurious signal response; simplifies some tuning problems	Results in higher IF noise figure. The receiver noise figure depends on the amount of noise-noise intermodulation due to the local oscillator. If this is negligible, the receiver noise figure will generally be higher with the higher IF
IF bandwidth on the order of the reciprocal of the transmitted pulse length	Maximizes the peak signal to rms thermal noise, thereby providing best detection in thermal noise	Limits signal resolution; imposes strict requirements on the tuning accuracy of the receiver
IF bandwidth characteristic which enhances the signal sidebands greatly removed from the transmitted carrier frequency and attenuates the sidebands near the carrier	Improves the detection capability in clutter	Degrades the detection of signals in thermal noise. Imposes strict requirements on tuning accuracy

TABLE 7-1 (cont'd.)

Receiver Characteristic	Desirable Effect	Effects on Other Receiver Characteristics
Prevention of saturation in a linear receiver	Maximum signal to noise ratio	Requires an AGC loop. When signal contains pulse amplitude modulation that must be recovered, instantaneous AGC is usually not acceptable. A slow AGC is demanded. The information rate is decreased since only a selected signal may operate the AGC. Any other signals that may be examined are modulated by the fluctuations of the signal controlling the AGC. Therefore multiple receiver channels are needed to increase the information rate
Extremely narrow IF bandwidth	Provides maximum signal to thermal noise ratio	Reduces the information rate attainable. Results in a loss of signal resolution unless preceded by a cross correlation operation. Wasteful of power

frequency control (AFC) systems are required to obtain the required tuning accuracy. These devices are discussed in detail in Chapter 8.

Spurious responses are outputs caused by signals at frequencies to which the receiver is not normally tuned. These responses are the result of inadequate selectivity and nonlinear elements in the receiver, e.g. mixers. The receiver is designed to minimize spurious responses by properly selecting intermediate frequencies and mixer circuits and by providing the necessary selectivity.

7-2 THE INTERDEPENDENCE OF RECEIVER COMPONENTS

A particular receiver characteristic may be designed to give optimum receiver performance when the contribution of this one characteristic of overall performance is considered. However, many of the receiver characteristics are interdependent and therefore compromises must be made in the design. The compromises for a specific design are determined by the

performance requirements imposed on the radar system by the tactical requirements.

Some examples of the effects that the choice of a given receiver characteristic has on the overall receiver performance are indicated in Table 7-1.

7-3 RECEIVER NOISE FIGURE

The ultimate sensitivity of a receiver is dependent upon the inherent noise generated in the receiver circuits. A useful measure of this noise is the *receiver noise figure* which is defined as the ratio of the actual noise power output of a linear receiver to the noise power output of a noiseless receiver of otherwise identical characteristics.

Noise in a receiver is made up of *thermal* noise, which results from thermal agitation of charge carriers in conductors, and *shot* noise, which results from random electron motions in vacuum tubes. These noises are characterized by a Gaussian amplitude distribution with time. Such noise sources are independent and uncorrelated. The average power from the various sources is additive, and it is therefore convenient to employ ratios involving power in determining noise figure.

Consider a signal generator described by a short-circuit signal current source I_s and an internal conductance g_s which is at an absolute temperature T_s . Let the generator be connected to a load g_L which is at an absolute temperature T_L . Both g_s and g_L will generate fluctuation currents which are given by¹

$$\overline{i_{ns}^2} = 4kT_s g_s df \quad (7-1)$$

and

$$\overline{i_{nL}^2} = 4kT_L g_L df \quad (7-2)$$

where i_{ns} and i_{nL} are the rms noise currents in a frequency bandwidth element df , k is Boltzmann's constant = 1.37×10^{-23} joule/ K° .

The available signal power from the generator is $I_s^2/4g_s$ and the available thermal noise power is $\overline{i_{ns}^2}/4g_s = kT_s df$.

The available signal power from the circuit composed of the signal generator and the load g_L is

$$\frac{I_s^2}{4(g_s + g_L)}$$

If t is defined as T_L/T_s , the available noise power is

$$\frac{\overline{i_{ns}^2} + \overline{i_{nL}^2}}{4(g_s + g_L)} = \frac{kT_s df (g_s + t g_L)}{g_s + g_L} \quad (7-3)$$

¹J. B. Johnson, "Thermal Agitation of Electricity in Conductors," Phys. Rev. **32** (1928).

The noise figure is defined as

$$F = \frac{S_i/N_i}{S_o/N_o} = \frac{N_o}{N_i G} \quad (7-4)$$

where F is the noise figure (a power ratio)

S_i/N_i is the available signal to noise ratio at the input

S_o/N_o is the available signal to noise ratio at the output

G is the available power gain.

For the case of the generator connected to a load, the noise figure of the combination is then

$$F = 1 + t \frac{g_L}{g_s} \quad (7-5)$$

If both the generator and load are at the same temperature, then the noise figure is merely the attenuation of the signal resulting from the termination.

In a radar receiver it is convenient to associate a noise figure with various elements and then determine the receiver noise figure resulting from their combination in cascade.

Consider that a number of elements characterized by a noise figure F_j and an available power gain G_j are interconnected as in Fig. 7-2. It is

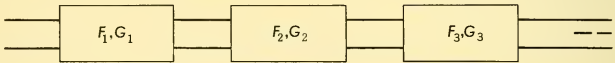


FIG. 7-2 Noise-Equivalent Radar Receiver.

assumed that all the noise sources are at a temperature T (a difference in temperature may be included as a temperature ratio). Assume further that all elements are linear, and that the effective noise bandwidth of each element is B_n .

The input noise is $N_i = kTB_n$

The overall gain G is $G_1 G_2 G_3 \dots$

Output noise originating from the source is $GkTB_n$.

The additional noise at the output contributed by the first box is $GkTB_n(F_1 - 1)$. The additional noise at the output contributed by the second box is $\frac{GkTB_n(F_2 - 1)}{G_1}$. The sum of all noise contributions add up

to $FGkTB_n$, where F is the overall noise figure expressed as a power ratio.

$$FGkTB_n = GkTB_n + GkTB_n(F_1 - 1) + \frac{GkTB_n(F_2 - 1)}{G_1} + \dots \quad (7-6)$$

Then

$$F = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \dots + \frac{F_n - 1}{\prod_{j=1}^{n-1} G_j} \quad (7-7)$$

In the common airborne radar set, RF amplification is not employed. Instead, the signal is heterodyned to some intermediate frequency and then amplified. Microwave crystal mixers are passive nonlinear devices. Their noisiness is characterized by the amount of noise produced by the mixer compared with the noise from a resistance at the same external temperature. The noise is thus expressed as a temperature ratio t_m . A mixer acts as a switch, and in terms of available power exhibits a loss at the conversion frequency. This loss is designated as a power ratio L_x . Following the previous notation the noise figure of a mixer is then $t_m L_x$.

The noise figure of a superheterodyne radar receiver is then

$$F_{rec} = 1 + \frac{T}{T_a}(L_t - 1) + \frac{T}{T_a}L_t[(t_m - 1) + (F_{IF} - 1)] \quad (7-8)$$

where F_{rec} is the receiver noise figure expressed as a power ratio

$t_m - 1$ is the excess noise of the mixer

$F_{IF} - 1$ is the excess IF noise figure expressed as a power ratio

L_t is the product of the conversion loss of the mixer and loss in the microwave transmission circuitry between the antenna and mixer expressed as a power ratio

t_m is the effective noise temperature ratio of the mixer

T is the noise temperature of the receiver

T_a is the temperature of the antenna.

The noise figure is usually defined with respect to room temperature (about 291° K).² A radar receiver is, however, connected to a directional antenna which can be represented as an equivalent generator at a temperature less than room temperature when the antenna is directed toward space. In fact the equivalent antenna temperature under this condition may be about 4°K.

When referred to the antenna temperature under this condition, the noise figure of the best airborne radar sets is on the order of 30 db.

²Under such a definition T/T_a is unity.

7-4 LOW-NOISE FIGURE DEVICES FOR RF AMPLIFICATION

The crystal mixer is a rather fragile element, and its electrical characteristics deteriorate when large-signal inputs are applied to it. Because of the conversion loss of the crystal mixer, the receiver noise figure is highly dependent on a low IF noise figure. From a consideration of the noise figure of cascaded networks it is seen that a large available power gain in the first network minimizes the noise contributions of the later networks.

A low-noise RF amplifier preceding the crystal mixer can provide the power gain required to offset the loss of the mixer. Two types of RF amplifiers that might be employed in an airborne radar receiver are the traveling wave tube (TWT) amplifier and the variable parameter amplifier. These devices are discussed in more detail in Chapters 10 and 11.

Noise in the traveling wave tube results from noise in the electron beam. Theoretically the noise can be reduced to about three times kTB . At present such tubes are not available for airborne radar receivers. Tubes are available, however, that are nearly competitive in noise figure with the microwave crystal mixer in the frequency range employed by the airborne radar set.

A TWT will produce a saturated output under strong signal conditions at maximum gain, alleviating many TR switching difficulties. Since the tube provides gain, the noise figure of the elements which follow is not nearly as important as in the conventional airborne radar receiver. Therefore, much higher intermediate frequencies are feasible without degrading the noise figure. A higher IF results in fewer spurious signal outputs from the receiver. The tube can be gain-controlled by changing the beam current so that it can produce an attenuation equal to the cold loss of the tube if required. This is an advantage when attempting to amplify strong signals with minimum distortion.

One disadvantage in the traveling (forward) wave tube results from the wide bandwidth. The noise spectrum is very wide and this results in more noise at the mixer than desired. An RF filter between the traveling wave tube and the mixer can, however, eliminate this condition if necessary. Another disadvantage is that a number of spurious signals can be generated in the tube, and are likely to be encountered in practice due to the wide RF acceptance bandwidth of the tube. The backward wave amplifier has a narrower bandwidth than the forward wave amplifier and may prove to be the most desirable type of traveling wave tube for use as an RF preamplifier in an airborne radar set. Traveling wave tubes may also be constructed with two slow wave structures to provide mixing.

Variable parameter amplifiers — also called *parametric amplifiers* — are much simpler than the TWT amplifiers. The transmission type of amplifier

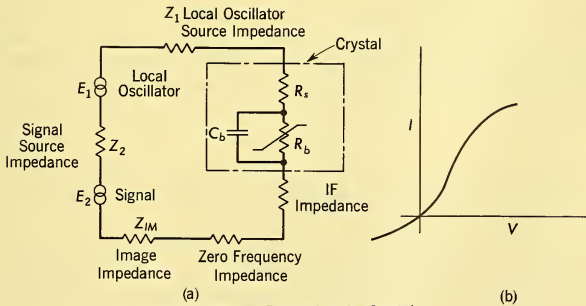
appears to be the best suited for radar receivers, although the noise figure is somewhat higher than the reflection type. Many practical problems associated with these amplifiers, such as stabilization of the loaded Q 's of the resonant circuits and regulation of pump power level, must be solved before these amplifiers find large use in airborne radar sets. However, these negative-resistance amplifiers appear to be a final step in attaining receivers whose sensitivity is truly limited by external noise.

7-5 MIXERS

The SHF (super high frequency) mixer in the majority of airborne radar receivers incorporates crystal diodes. Properties of the crystal mixer which are important to radar system operation are:

1. The effective noise temperature
2. The conversion loss
3. The intermodulation components

A crystal mixer can be represented by an equivalent circuit, as is shown in Fig. 7-3a. The nonlinearity of the crystal arises from the variation



Each Impedance Shown External to the Crystal is Zero to All Frequency Components Except the One to Which it Refers

FIG. 7-3 (a) Equivalent Circuit of Crystal Mixer and (b) V - I Characteristics of a Mixer.

in the barrier resistance R_b which is a function of the voltage applied to the crystal. A typical transfer characteristic is shown in Fig. 7-3b. The spreading resistance R_s and barrier capacitance C_b are detrimental parasitic elements. Because of these elements, not all of the heterodyne signals' energy can reach the IF and image termination.

To obtain a low conversion loss, the voltage applied to the diode by the local oscillator signal is very large so that there is negligible conduction during one half of the local oscillator cycle. The signal voltage is much smaller than the local oscillator voltage.

A current flows through the IF impedance which can be described by the multiplication of the signal and local oscillator voltages by the transfer characteristic of the mixer. The transfer characteristic can be expressed as a power series in the applied voltage. Because of the magnitude of local oscillator voltage a large number of terms are required to describe the mixer behavior.

Thus

$$I = \sum_{n=0}^{\infty} a_n E^n \quad (7-9)$$

where E is the input voltage

I is the current flowing in mixer

a_n are the coefficients of the power series describing the mixer; these are dependent on the local oscillator signal level

n is an integer 0, 1, 2, 3,

Normally E consists of the sum of two voltages, the signal voltage and the local oscillator voltage. In general the input may be

$$E = \sum_{r=1}^m A_r \cos \omega_r t \quad (7-10)$$

with the condition that $A_1 \cos \omega_1 t$ be the local oscillator signal and $A_1 \gg A_2, A_3, \dots$.

If the signal is a single frequency ω_2 and the IF center frequency is $(\omega_1 - \omega_2)$, the desired output spectrum from the mixer is the intermodulation term $K \cos (\omega_1 - \omega_2)t$. An expansion of the expression for the current in the IF impedance yields terms of the form

$$I_{IF} = A_2 \left[a_2 A_1 + \frac{3}{2} a_4 A_1^3 + \dots + \frac{(2n)!}{[2^{(2n-1)}(n)!]} a_{2n} A_1^{(2n-1)} + \dots \right] \cos (\omega_1 - \omega_2)t, \quad r \neq 1. \quad (7-11)$$

for the IF outputs incident to mixing of the signal frequencies with the local oscillator frequency. The term in the brackets is a constant for a particular value of local oscillator voltage, and the mixer thus produces an IF output which can be expressed as

$$I_{IF} = K A_r, \quad r \neq 1. \quad (7-12)$$

When more than one signal frequency is applied to the mixer, intermodulation between the signal components occurs. The output current caused by these intermodulation components is of the following form for each possible pairing of m signal components:

$$I = A_j A_k \left(a_2 + 3a_4 A_1^2 + \dots + \frac{(2n)!}{[2^{(2n-1)}(n-1)!]} a_{2n} A_1^{(2n-2)} + \dots \right) \cos(\omega_j - \omega_k)t. \quad (7-13)$$

where $j = 2, \dots, m$

$k = 2, \dots, m.$

Once again, the term in the brackets is constant for a particular local oscillator voltage, so that

$$I = A_j A_k K_1. \quad (7-14)$$

The voltage developed by these mixer currents is prevented from affecting the receiver performance by the frequency selectivity of the IF networks when $\omega_j - \omega_k$ falls outside the IF passband. Those components that would ordinarily fall within the IF passband, could be eliminated by RF preselection and proper IF frequency. However, such preselection is not always feasible. A balanced mixer is therefore used (see Paragraph 10-15). In the balanced mixer two crystals are placed at two of the ports of a microwave junction, and the signal is fed into one port and the local oscillator into the other port. The junction may be a magic-tee, short-slot hybrid or rat-race.

Each individual crystal develops all of the intermodulation components, but the relative phase of the signal-signal beats differs from that of the signal-L.O. beats and therefore can be discriminated against in the IF coupling circuit. Rejection of the undesired intermodulation components on the order of 25 db is realized in practice. Principal factors affecting the rejection of signal-signal beats are the impedance match between the signal source and each crystal, the rectifier dynamic characteristics, and the balance of the IF circuit.

Among the signal-local oscillator products are two which affect the performance of the mixer. These are included in the value K of Equation 7-12 when it is determined experimentally by measuring the $\omega_1 - \omega_r$ component from the mixer. These two products involve the generation of an *image frequency*, i.e., a signal which is separated from the desired signal frequency by a frequency equal to twice the IF frequency, and which is separated from the local oscillator frequency by the IF frequency. The image frequency is caused by second harmonic mixing and by an up conversion resulting from the IF current which flows through the mixer.

The image frequency signal appears across the crystal and propagates down the waveguide toward the local oscillator and the antenna. If the image wave sees a match, such as would exist if it were allowed to enter the local oscillator channel, the energy in this signal is dissipated and energy that could have appeared in useful IF output is lost. Proper reflection of the wave can cause it to enter the mixer and arrive at the crystal in proper phase so that the output IF is increased. Optimum handling of the image can improve the noise figure about 1 or 2 db. In general, however, conventional pulse type airborne radar receivers have broad band mixers. The image conversion is terminated and the lowest possible noise figure is not obtained.

A number of other intermodulation components involving the second harmonic of the local oscillator occur and can be significant when the RF acceptance bandwidth is great.

The crystal diode voltage-current relationship is given by

$$i = K \left(\exp \frac{eV}{kT} - 1 \right) \quad (7-15)$$

where e = electronic charge

V = applied voltage

K = constant depending on crystal

T = temperature of the junction.

Shot noise is exhibited by the crystal; the mean square fluctuation current is $\overline{i^2}df = 2eI_0df$, where I_0 is the d-c current through the crystal. Equation 7-15 indicates that a given conversion loss could be obtained with a lower d-c current by reducing the temperature and therefore producing less shot and granular noise. In addition to the shot noise there is a frequency-dependent noise. All of this noisiness of the crystal is specified by the crystal noise temperature ratio t_x . The mixer noise temperature is t_m and is given by

$$t_m = \frac{2}{L} \left[t_x \left(\frac{L}{2} - 1 \right) + 1 \right] \quad (7-16)$$

for the broadband mixer. L is the conversion loss; t_m is the value specified by crystal manufacturers. The t_m of an actual mixer may be different, depending on the termination of the image conversion which affects L .

Equation 7-8 shows that a large value of F_{IF} causes the conversion loss of the mixer to be the dominant parameter of the mixer contributing to the noise figure. In fact even for a low IF noise figure the conversion loss is more dominant than the noise temperature. The mixer therefore yields lowest noise figure when it is designed for minimum conversion loss.

A frequency-dependent part of t_m has been observed to vary as $(f_1/f_2)n$, where f_2 and f_1 are IF frequencies and n is between 0.5 and 1.³ Since the IF noise figure varies approximately as f_1/f_2 at high IF, the IF frequency at which minimum F_{rec} is realized is not critical.

The local oscillator is usually a klystron with a wide electronic tuning range. Such oscillators exhibit shot noise whose spectra are determined by the Q of their resonators. To minimize intermodulation components between such noise and the local oscillator signal a high IF frequency is desirable. The use of a balanced mixer, however, reduces this noise significantly.

In any mixer design, the objective is to provide minimum t_m and L . The conversion loss L depends on the match between the signal source at both the signal and image frequency, the RF signal frequency, the IF frequency, the crystal biasing, and the local oscillator signal level. The noise temperature t_m is also dependent on L , the IF frequency, and local oscillator signal.

7-6 COUPLING TO THE MIXER

To obtain minimum noise figure, minimum mixer conversion loss must be realized. Conversion loss is defined on an available power basis; therefore the conversion loss does not depend on the actual IF load admittance connected to the mixer. The conversion loss, however, is dependent on the RF signal source admittance. To obtain minimum conversion loss at the principal beating frequency (signal frequency beating against local oscillator frequency) a mismatch is required between the mixer and the source. The input admittance of the mixer, however, depends on the IF conductance seen by the mixer. This in turn depends on the design of the first IF stage of the receiver and the network which connects it to the mixer.

A condition frequently encountered in airborne radar receivers is that the IF admittance is very large incident to the use of a double-tuned transformer between the mixer and first IF tube. The secondary circuit is usually damped only by the coil losses and circuit losses. A very large admittance is therefore coupled into the primary circuit near the resonant frequency of the secondary circuit.

For this type of coupling between mixer and IF amplifier, an optimum mismatch between the signal source and the mixer is given approximately by

$$\rho = \frac{L_o - 1}{L_o + 1} \quad (7-17)$$

where ρ is the VSWR (voltage standing wave ratio) at the signal frequency and L_o is the optimum conversion loss.

³P. D. Strum, "Some Aspects of Crystal Mixer Performance," *Proc. IRE* **41**, 876-889 (1953).

The source admittance can be designed on this basis when the IF coupling circuit is as specified.

When the image frequency signal generated at the mixer is allowed to be radiated by the antenna or dissipated in the local oscillator source admittance, the value of L_o is that which is normally specified by the crystal manufacturer as conversion loss.

In many airborne radar receivers a short-slot hybrid junction is employed in a balanced mixer. When the crystals are matched in such a mixer, all of the image frequency signal generated in the mixer propagates out the local oscillator port. Normally this port is matched, therefore the image signal energy is lost.

In some radar sets a filter may precede the mixer to reduce interference from other radar sets. Such a filter may appear as a susceptance at the image frequency and reflect the image signal originating in the mixer. If the signal arrives in the correct phase at the mixer crystals, the performance is improved. The phase depends on the distance between the mixer and the filter. However, the distance between the mixer and the filter is also dependent on the mixer to IF coupling circuit, since the filter would be situated so as to give the optimum mismatch of the source to the mixer. To obtain lowest receiver noise figure, design of the RF and IF circuits must therefore be considered jointly, not separately. One solution to this problem might be the use of the short-slot hybrid with a filter in both signal and local oscillator paths.

7-7 IF AMPLIFIER DESIGN

The IF amplifier consists of a cascaded arrangement of vacuum tube amplifiers which employ band pass coupling networks. Frequency of operation is a compromise between several factors such as noise figure, circuit stability, spurious responses, and receiver tuning characteristics. Consideration of these factors usually leads to the choice of an IF frequency between 30 and 60 Mc in the ordinary pulse-type airborne radar receiver.

The IF amplifier is a filter amplifier, and its small-signal transfer function is given by

$$G(s) = H \frac{s^{nq}}{s^{mn} + a_1 s^{mn-1} + a_2 s^{mn-2} + \dots} \quad (7-18)$$

In this expression H is a constant depending on the number of tubes, their transconductance, and the capacitance values of the circuits; s is the complex frequency variable $\sigma + j\omega$; n is the number of circuits; q and m are determined by the network complexity. The transfer function vanishes when $s = 0$ because of the numerator term. The function thus has a zero of order nq at the origin. The denominator can be factored into

$$(s - s_1)(s - s_1^*)(s - s_2)(s - s_2^*)(s - s_3)(s - s_3^*) \dots \quad (7-19)$$

The values of s for which the denominator vanishes ($s = s_1; s = s_1^* \dots$ etc.) are the zeroes of the denominator. At these values of s , the transfer function becomes infinite, so they are called the *poles of the network function*. The synthesis of an IF amplifier is facilitated by use of a potential analogy.⁴ By considering each pole to represent the position of positive line charge normal to the complex frequency plane, and each zero to represent the position of a negative line charge normal to the plane, it can be shown that the potential measured along the $j\omega$ axis resulting from the pole-zero array is equivalent to the logarithm of the magnitude of the normalized transfer function.

When the transfer function consists of poles which are very far removed from the origin and near the $j\omega$ axis (as in Fig. 7-4), an arrangement of the

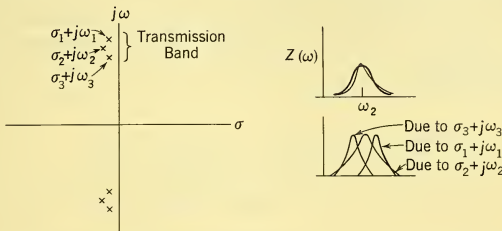


FIG. 7-4 S-Plane and Z-Plane Representations of IF Amplifier Characteristics.

poles at each interval about a semicircle having the $j\omega$ axis as diameter produces an approximate constant potential on the $j\omega$ axis. The networks that are used, however, have zeroes at the origin and conjugate poles in the third quadrant of the plane. When the ratio of bandwidth of the overall receiver to the IF frequency becomes large, the contribution of these zeroes and poles to the transfer function in the passband region becomes significant. A conformal transformation

$$Z = s^{1-\mu} \left(s + \frac{1}{s} \right) \quad (7-20)$$

is used to obtain an exact low-pass transformation, where

$$\mu = \frac{\text{number of zeroes at origin of } s \text{ plane}}{\text{number of poles in upper left } s \text{ plane}}$$

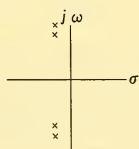
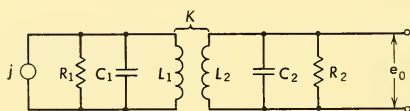
for the individual network elements employed. This transformation moves the zeroes to infinity in the z plane, and results in coincidence of the s plane

⁴W. H. Huggins, *The Potential Analogue in Network Synthesis and Analysis*, Air Force, Cambridge Research Lab. Report, March 1951.

pole pairs in the z plane, so that a single pole cluster is obtained in the z plane.

Placing the poles on a semicircle in the z plane produces a constant potential in the s plane. When the poles in the s plane are placed at the position given by transforming the equally spaced poles of the z plane to the s plane, a maximally flat transfer characteristic is obtained.

The desired transfer function for the amplifier could be realized by a single four-terminal network followed by an extremely wide-band amplifier. (The amplifier, of course, would have a particular pole-zero structure, but the contribution to the selectivity in the frequency band of interest is negligible.) Practically, however, the four-terminal network is limited to a four-pole structure, and most commonly to a one or two-pole structure because of the limitation of realizable unloaded Q 's for the network inductors. The most commonly used network of the IF amplifier is shown in Fig. 7-5. When $k = 1$ and $L_1 = L_2$ the network reduces to a one-pole structure as shown.



$$Z_1 = H \frac{S}{S^4 + \alpha S^3 + \beta S^2 + \gamma S + \delta}$$

$$= H \frac{S}{(S - S_1)(S - S_1^*)(S - S_2)(S - S_2^*)}$$

$$\text{Where } H = \left(\frac{K}{1 - K^2} \right) \left(\frac{\omega_1 \omega_2}{C_1 C_2} \right)$$

$$\alpha = \frac{\omega_1 + \omega_2}{Q_1 + Q_2} \quad \beta = \frac{\omega_1 \omega_2}{Q_1 Q_2} + \frac{\omega_1^2 + \omega_2^2}{1 - K^2}$$

$$\gamma = \frac{\omega_1^2 \omega_2}{Q_2(1 - K^2)} + \frac{\omega_2^2 \omega_1}{Q_1(1 - K^2)} \quad \delta = \frac{\omega_1^2 \omega_2^2}{1 - K^2}$$

$$\omega_1^2 = \frac{1}{L_1 C_1}, \quad \omega_2^2 = \frac{1}{L_2 C_2}$$

$$Q_1 = \omega_1 R_1 C_1, \quad Q_2 = \omega_2 R_2 C_2$$

When $K=1$ and $L_1 = L_2$ the Single-Tuned Circuit With Bifilar Coil is Obtained. In This Case Usually All of the Damping is Provided by R_2

Then:

$$Z_1 = H_1 \frac{S}{S^2 + \gamma S + \lambda}$$

$$= H_1 \frac{S}{(S - S_1)(S - S_1^*)}$$

$$H_1 = \frac{1}{C}$$

$$\gamma = \frac{1}{RC}$$

$$\lambda = \frac{1}{LC}$$



FIG. 7-5 Commonly Used IF Interstage Coupling Network.

A parameter of great importance is the gain bandwidth product. Taking the simplest interstage circuit as a reference network, the *gain bandwidth product* is

$$GB = g_m |Z_t|_{\omega_0} \frac{\omega_0}{2\pi Q} = \frac{g_m}{2\pi C} \quad (7-21)$$

This equation shows that the quantities which determine the gain and bandwidth are so related that high gain can be obtained only at the expense of reduced bandwidth. GB is determined by the tube and the circuit physical layout since it affects C . If two identical circuits are cascaded, then the 3-db bandwidth of the cascaded circuits is the 1.5-db bandwidth of a single circuit. For a given overall bandwidth, the individual stage bandwidths must be increased. With flat staggering of the circuits, this bandwidth shrinkage does not occur, and the GB product can be used to determine overall gain of the amplifier.

The normalized attenuation characteristic of the IF coupling circuits employed in conventional pulse radar receivers can be expressed by the equation

$$\text{Attenuation (db)} = 10 \log_{10} [1 + x^{2n}] \quad (7-22)$$

where $x = f_0 \left(\frac{f/f_0 - f_0/f}{B_1} \right) \cong \frac{2(f - f_0)}{B_1}$

n = number of poles in the circuit (low pass equivalent)

B_1 = 3-db bandwidth

f = frequency at which attenuation is to be evaluated.

If m groups of such circuits are cascaded, then the overall amplifier selectivity is

$$\text{Attenuation (db)} = 10m \log_{10} [1 + x^{2n}]. \quad (7-23)$$

The overall bandwidth of the amplifier is therefore

$$B = (2^{1/m} - 1)^{1/2n} B_1. \quad (7-24)$$

For a given overall IF bandwidth it is desirable that the principal selectivity occur in low-level stages. The input stage selectivity, however, is governed by noise figure considerations. The input stages are therefore designed first and then the remaining amplification and selectivity introduced.

The selectivity of the IF amplifier is provided by one-pole or two-pole networks between the stages in order to realize the maximum dynamic range. With n -pole configurations in which the poles are distributed through the amplifier, the dynamic range is usually not the same at all frequencies within the pass band; such designs are therefore avoided.

Groups of two-poles (staggered pairs) are frequently used. Although the restriction of dynamic range is not too severe, such designs are nevertheless inferior to synchronous stages.

When an amplifier is built with single-pole coupling circuits, the overall frequency response exhibits geometric symmetry with respect to the IF center frequency. With two-pole coupling provided by the magnetically coupled double-tuned transformer, the response is more nearly arithmetically symmetrical. With two-poles having Q ratios of about 3 to 1 the usual amplifier requirements (bandwidth between 1.0 and 10.0 Mc) can be realized with adequate stability margin and fewer components than with single poles.

One of the main difficulties encountered in the design of radar IF amplifiers is accurate control of the stability margin. Pole shifting (regeneration) can occur under strong signal conditions and results in poor transient characteristics or modulation distortion on desired signals.

The principal source of feedback in an IF amplifier is the grid-to-plate capacitance of the tube and circuit. Other feedback paths are:

1. Coupling between input and output leads
2. Coupling due to the chassis acting as a waveguide beyond cutoff frequency
3. Grid-to-cathode feedback
4. Inadequate decoupling circuits resulting from self-inductance of bypass capacitors and their connecting leads
5. Coupling between heaters
6. Coupling between input and output caused by ground currents. (The impedance of the chassis is not negligible. It is necessary that the output and input currents not flow through the same part of the chassis to ensure stable operation.)

The grid-to-plate feedback can be partly compensated by proper circuit design. It is advisable that a common bypass capacitor be employed for the screen grid and plate return of the amplifier. Appropriate choice of this component then enhances the amplifier stability. In high-frequency stages which are gain-controlled it is also desirable that feedback be introduced in the cathode lead to stabilize the input susceptance of the tube. These circuits are shown in Fig. 7-6. It is further desirable that such stages employ vacuum tubes with separate suppressor grid terminals to minimize the feedback from plate to cathode. This feedback path leads to instability when the input susceptance of tubes having internal suppressor grid-to-cathode connections is to be stabilized.

The most economical distribution of gain and selectivity in the IF amplifier occurs when the stages are made identical. However, this condition does not always provide the most stable operation. The latter IF

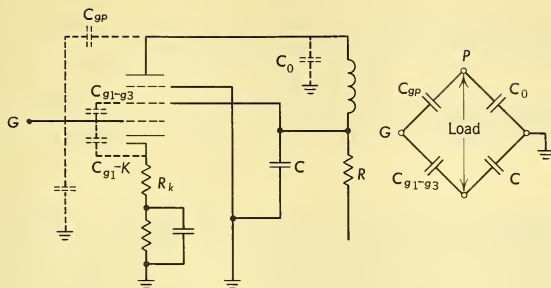


FIG. 7-6 Typical IF Amplifier Configuration Showing How Proper Choice of C Can Self-Neutralize the Stage. (Lead Inductance Has Been Neglected.) $R \gg X_C$.

stages must provide adequate dynamic range and thus are operated in a manner which allows large peak transconductances to occur. The permissible gain in these stages should therefore be less than in the early stages when a tube of the same type is employed in both places. The gain can be controlled by lowering the impedance of the coupling network or by allowing a greater bandwidth in the stages. The maximum allowable gain is sometimes limited to

$$\text{Max gain} = 0.2 \sqrt{\frac{g_m}{\pi f C_{gp}}} \quad (7-25)$$

where g_m is the peak transconductance and C_{gp} is the grid-plate capacitance.

The effective noise bandwidth B_n is the parameter involved in radar performance computation. For any practical amplifier this is very nearly the 3-db bandwidth

$$B_n = \frac{1}{G(\omega_0)} \int_0^{\infty} G(\omega) d\omega \quad (7-26)$$

where $G(\omega)$ is the power spectrum of the receiver. The normalized power spectrum of the radar receiver usually can be given by

$$G(\omega) = \left\{ 1 + \left[\frac{\omega_0}{B} \left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \right) \right]^{2n} \right\}^m \quad (7-27)$$

where ω_0 is the midband frequency

B is the 3-db bandwidth

n is the number of poles in a flat circuit

m is the number of groups of circuits.

The 3-db bandwidth is

$$(2^{1/m} - 1)^{1/2n} B_1, \quad (7-28)$$

where B_1 is 3-db bandwidth of one network consisting of n poles. Then as an example for three staggered pairs of one-poles

$$\frac{\text{Noise bandwidth}}{\text{3-db bandwidth}} = \frac{\int_0^{\infty} \frac{dx}{(1+x^4)^3}}{(2^{1/3}-1)^{1/4}} = \frac{1}{0.714} \frac{\Gamma(\frac{1}{4})\Gamma(3-\frac{1}{4})}{\Gamma(4)\Gamma(3)} = 1.02. \quad (7-29)$$

7-8 CONSIDERATIONS OF IF PREAMPLIFIER DESIGN

The IF amplifier is frequently divided into two units, an IF preamplifier, and the main IF amplifier (postamplifier). This arrangement allows the input stages of the IF amplifier to be physically located near the mixer. When long cables are used between the mixer and IF amplifier, bandwidth and noise figure must usually be compromised. Principal considerations in the design of the IF preamplifier are the noise figure, signal-handling capability, selectivity, and gain.

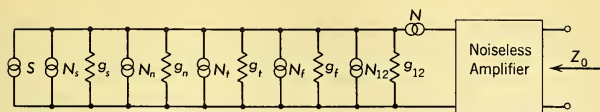
Triode tubes are almost always required for the first two IF amplifiers. They are used because they exhibit less shot noise⁵ than pentodes. (Shot noise is the noise resulting from fluctuations of the currents in a vacuum tube.) The input IF amplifier may be used in a grounded-cathode or grounded-grid arrangement. For the ordinary AI radar with broad band mixer, the grounded-cathode amplifier is usually employed for the first tube. To minimize input admittance variation caused by feedback from grid to plate this amplifier stage is neutralized.

The equivalent circuit representing the sources of noise associated with an IF amplifier stage is shown in Fig. 7-7. From this arrangement of noise generators and signal generator the noise figure as defined by Equation 7-4 becomes

$$\begin{aligned} F &= 1 + \frac{g_n + \beta g_t + \gamma g_f + \delta g_{12}}{g_s} + \frac{R_n}{g_s} \left| g_s + g_n + g_t + g_f + g_{12} + j^b \right|^2 \\ &= 1 + \frac{\psi g_1}{g_s} + \frac{R_n}{g_s} \left| Y_t \right|^2 \end{aligned} \quad (7-30)$$

where b is the total susceptance appearing at the input terminals. The additional parameters involved are defined in Fig. 7-7. This expression yields the single-stage *spot noise figure* of an amplifier which is defined as the noise figure at a specific point of input frequency. When the noise figure of a radar receiver is measured, a noise source is generally employed. If the spot noise figure is constant over the bandpass of the overall receiver, then the noise figure that is measured will be independent of bandwidth

⁵B. J. Thompson, "Fluctuation Noise in Space Charge Limited Currents at Moderate High Frequencies," *RCA Rev.* 4, 269 (1940).



$S, N_s, N_n, N_t, N_f, N_{12}$ = current generators

N = a voltage generator

S = signal current associated with g_s

N_s = source noise, $N_s = \sqrt{4KT_sBg_s}$

(In evaluating IF noise figure T_s is usually taken as reference temperature T for which F is defined, or 290°K . In a radar receiver T_s is larger than T because of noise-noise intermodulation at mixer from local oscillator signal.)

N_n = noise current associated with coupling network losses, $N_n = \sqrt{4KT_nBg_n}$

N_t = noise current associated with grid noise of tube, $N_t = \sqrt{4KT_tBg_t}$

β = usually 4 or 5

g_t = grid damping due to finite transit time only

N_f = noise current associated with feedback such as that due to cathode lead inductance: $N_f = \sqrt{4K\gamma Tg_f}$; γ may be between 0 and 1

N_{12} = noise current associated with output to input feedback conductance: $N_{12} = \sqrt{4K\delta Tg_{12}}$; δ usually 1

N = voltage generator representing shot noise in the tube, $N = \sqrt{4KTBR_n}$

where R_n is the resistance that must be connected between grid and ground to produce the same fluctuation current in the plate circuit of the hypothetical noiseless tube as exists in the actual tube due to shot effect. For the triode, R_n varies between

$$R_n = \frac{2.5}{g_m} \quad \text{and} \quad R_n = \frac{3.0}{g_m}$$

where g_m = operating transconductance.

FIG. 7-7 Equivalent Circuit of IF Amplifier Noise Sources.

as the bandwidth is reduced. The last term of Equation 7-30, however, may cause the spot noise figure to increase for frequencies removed from the carrier frequency and the average value of F is then increased. To minimize this effect in wide-band applications, care is taken to minimize the increase in this term either by virtue of a small value of b or, in some cases, by introducing feedback to increase g_f . In all cases, IF preamplifier tubes are selected which have a high transconductance-to-input capacitance ratio and small transit time.

An important consideration is the mixer-IF coupling network. When g_f and g_{12} are zero, minimum F is obtained for

$$g_s = \sqrt{\frac{g_n + \beta g_t}{R_n} + (g_n + g_t)^2} \quad (7-31)$$

as can be proven by differentiating Equation 7-30. Application of this equation is difficult because βg_t cannot always be easily determined. In

practice, careful measurement of the actual input admittance of the tube under operating conditions and with feedback effects removed by neutralization gives an input conductance which, when employed as βg_l in the amplifier design, results in measured noise figures in close agreement with calculated values.

Fig. 7-8 shows a simplified equivalent circuit for a grounded-grid stage. The noise figure is given by

$$F \cong 1 + \frac{\psi g_1}{g_s} + \frac{g_l}{g_s} + \frac{R_n}{g_s} \left(\frac{\mu}{\mu + 1} \right)^2 |Y_t|^2 \quad (7-32)$$

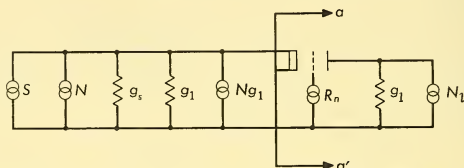


FIG. 7-8 Simplified Equivalent Circuit for a Grounded Grid Stage. N_{gl} includes noise due to grid loading and network loss. N_{gl} is therefore taken as a noise current generator $\sqrt{4KT\psi g_l}$, where g_l is total conductance between cathode and grid and ψ is an effective temperature ratio. The admittance seen to the right of aa' , is

$$\frac{g_l(1 + \mu)}{1 + g_l r_p}$$

where μ is the amplification factor of the tube. In determining the overall noise figure of the IF amplifier, the available power gain of the amplifier stage must be considered. When losses in the interstage coupling network are not included, the available power gain for the grounded-cathode and grounded-grid stages is given by

$$W \cong \frac{g_m^2 g_s r_p}{Y_t^2} \quad (\text{grounded cathode}) \quad (7-33)$$

$$W \cong \frac{g'_m{}^2 g_s r_p}{(g_1 + g_s)(g_1 + g_s + g'_m)} \quad (\text{grounded grid}) \quad (7-34)$$

where $g'_m = \frac{1 + \mu}{r_p}$.

The most frequently used IF preamplifier circuit in airborne radar receivers is the *cascode* circuit. This circuit consists of a grounded-cathode triode stage followed by a grounded-grid stage. Such a configuration results in a lower noise figure than can be obtained with a cascade connection of two neutralized grounded-cathode triode stages. This is because the interstage bandwidth is obtained by virtue of the loading incident to the

large input conductance of the grounded-grid stage. Since such a feedback conductance is not a noise source, less noise exists than in the case when such network damping is obtained by a physical resistor.

Wide-band neutralization of the input tube is employed to stabilize the admittance appearing at the input grid. Such stabilization allows the widest bandwidth for which a low spot noise figure is obtained by minimizing the variation of the last term of Equation 7-30 with frequency.

To minimize the variation of the grid admittance with frequency, a double-tuned (two-pole) mixer-IF coupling network is employed with the cascode input circuit. When this circuit is designed for a flat (Butterworth) response, the bandwidth is given by

$$B = \frac{g_s}{\pi\sqrt{2}C} \quad (7-35)$$

where g_s is the value of source conductance required for minimum F

B is the 3 db bandwidth

C is the total capacitance appearing at the input of the first tube.

The signal transmission bandwidth is slightly wider than this value because of the loading of the source caused by coil losses and the input conductance of the tube.

In high-PRF radars (such as the pulsed-doppler systems described in Chapter 6) and where very short range with high accuracy is required, the double-tuned mixer-IF coupling network is found to introduce objectionable transients following the transmitter signal. These transients result from the nonlinear loading on the network by the mixer crystals. In such cases a grounded-grid input stage is employed. The transmission bandwidth of the mixer-IF coupling is very wide because of the heavy damping caused by the input conductance of the grounded-grid stage. The heavy damping by the tube minimizes transients resulting from the crystal mixer IF admittance variation when the transmitter signal is present at the mixer.

A typical example of IF preamplifier performance is given by the following:

CIRCUIT AND TUBE PARAMETERS

$$g_m = 20 \times 10^{-3} \text{ mho}$$

$$\mu = 44$$

$$\beta g_t = 5.0 \times 10^{-5} \text{ mho}$$

$$g_1 = 10^{-5} \text{ mho}$$

$$R_n = 150 \text{ ohms}$$

$$C_{\text{input tube}} = 15 \mu\mu f$$

$$C_{gp} = 1.5 \mu\mu f$$

$$\Delta C = 1.5 \mu\mu f \text{ due to space charge effect}$$

$$C_{\text{out}} = 3.0 \mu\mu f$$

$$C_{\text{in of third tube}} = 7.0 \mu\mu f$$

$$g_n \text{ of input coil} = 10^{-5} \text{ mho}$$

Problem Solution. Requirements are that the overall bandwidth between the mixer and third tube in which the attenuation is less than 1 db shall be 3 Mc, the noise figure of the amplifier shall be as low as possible, and a cascode circuit employing two of the tubes having the specified characteristics is to be used.

For minimum F , the source conductance from Equation 7-31,

$$g_s \cong \sqrt{\frac{10^{-5} + 5(10^{-5})}{150}} = 6.3 \times 10^{-4} \text{ mho is required.}$$

A source bandwidth (see Equation 7-35) of

$$B = \frac{6.3 \times 10^{-4}}{\pi\sqrt{2} \cdot 20.5 \times 10^{-12}} = 6.95 \text{ Mc}$$

is obtained, assuming $1.0 \mu\mu f$ stray capacitance and a wide-band neutralization of C_{gp} .

At 3-Mc bandwidth the input coupling network produces an attenuation (see Equation 7-22)

$$10 \log_{10} \left(1 + \left[\frac{3}{6.95} \right]^4 \right) = 0.15 \text{ db.}$$

The network between the second and third tube can therefore exhibit 0.85 db attenuation at a bandwidth of 3 Mc. A single-tuned circuit of 6.45-Mc bandwidth is adequate:

$$10 \log_{10} \left(1 + \left[\frac{3}{6.45} \right]^2 \right) = 0.85 \text{ db.}$$

Total capacitance between the second and third tube is $3.0 + 7.0 = 10 \mu\mu f$.

If a step-down transformer is employed, the gain bandwidth product could be improved and the conductance presented to the second tube output is lowered.

The step-down must be $n = \sqrt{\frac{C_1}{C_2}} = \sqrt{\frac{7.0}{3.0}}$

Total capacitance at output of the second tube is then $3.0 + \frac{7.0}{n^2} = 6.0 \mu\mu f$

and the conductance is

$$g_t = 2\pi(6.45)(10^6)(6.0)(10^{-12}) = 2.43 \times 10^{-4}.$$

From Equation 7-30 the noise figure of the first tube is

$$F_1 = 1 + \frac{10^{-5}}{6.3 \times 10^{-4}} + \frac{5.0 \times 10^{-5}}{6.3 \times 10^{-4}} + \frac{150}{6.3 \times 10^{-4}} (6.5 \times 10^{-4})^2.$$

$$1.196 = 0.78 \text{ db}$$

The noise figure of the second tube is (from Equation 7-32)

$$F_2 = 1 + \frac{10^{-5}}{4.55 \times 10^{-4}} + \frac{5.0 \times 10^{-5}}{4.55 \times 10^{-4}} + \frac{2.43 \times 10^{-4}}{4.55 \times 10^{-4}} + \frac{150(0.95)}{4.55 \times 10^{-4}} (4.65 \times 10^{-4})^2$$

$$\left(4.55 \times 10^{-4} = \frac{1}{r_p} = g_m/\mu \right)$$

$$= 1.74 = 2.4 \text{ db.}$$

From Equation 7-33, the available power gain of the first tube is

$$G = \frac{(20 \times 10^{-3})^2 (6.3 \times 10^{-4}) \left(\frac{44}{20 \times 10^{-3}} \right)}{(6.5 \times 10^{-4})^2} = 131$$

while the available power gain of second tube is approximately

$$\frac{4.55 \times 10^{-4}}{2.43 \times 10^{-4}} = 1.87.$$

Assuming 10 db for the third tube F , the preamplifier noise figure is given by Equation 7-7:

$$1.196 + \frac{1.74 - 1}{131} + \frac{10 - 1}{(131)(1.87)} = 1.237$$

$$\equiv 0.92 \text{ db.}$$

7-9 OVERALL AMPLIFIER GAIN

It is necessary that signal amplitudes corresponding to the thermal noise level at the input of the receiver be amplified to a suitable level for detection. The level required at the detectors depends on the use of the signal. For example, for signal detection on an intensity-modulated display providing range and azimuth coordinates, signal voltages on the order of 50 volts are usually required. The total amplification that is required depends on the signal bandwidth and the receiver noise figure. The equiv-

alent noise voltage that must be amplified may be determined by computing the total noise voltage at the input of each stage caused solely by the total grid conductance, the shot noise in the tube, and total grid admittance. This voltage is then referred to the input by dividing by the total gain from the input to the noise considered. Even though consideration is given to the design of a low-noise IF preamplifier, the process of referring all of the noise sources in the receiver to the input is extremely important, especially in multiple-conversion receivers such as are employed in some forms of doppler radar receivers. This method sometimes reveals factors such as noise on beating oscillator signals or noise caused by a method of selectivity distribution that would degrade the IF signal-to-noise ratio and result in a sensitivity poorer than would be estimated from consideration of the IF preamplifier noise figure and overall receiver selectivity only.

A typical equivalent input noise for an ordinary radar receiver having an overall bandwidth of 5.0 Mc is about 3.0 μ volts rms. For signal detection alone, a voltage between 1 and 2 volts rms at the input to the IF envelope detector is satisfactory. Thus the required overall amplifier gain is on the order of 105 to 115 db.

To obtain the voltages for the cathode ray tube an additional gain on the order of 40 db is then required. (Included in this figure is a loss of 6 to 10 db that is usually produced in wide bandwidth second detectors.)

Where the envelope of the signal must be accurately demodulated, higher voltages may be applied to the envelope detector to recover larger negative peak modulation with less distortion. However, dynamic range of the amplifier must be exchanged for the higher operating level. In tracking receivers, use of a range-gated amplifier ahead of the envelope detector allows such an exchange to be made. The detector average output is usually regulated to a relatively fixed level, and noise modulation positive peaks have very small probability of exceeding a level more than 12 db above the regulated level. In a typical case of a receiver having an IF bandwidth of 5 Mc, incremental gain can be maintained for a range of IF signal from zero to about 12 volts rms at the input to the IF envelope detector. Thermal noise can therefore be amplified to a level of about 3 volts rms at the detector input. It is obvious that this does not appreciably alter the IF gain requirements. In the case of very narrow band receivers such as are employed for detection and tracking of targets by means of their difference in doppler frequency, receiver bandwidths are on the order of several hundred cycles per second. Considerably more gain is therefore required over that encountered in conventional radar sets. For example, with a bandwidth of 500 cps, the required gain to the envelope detector would be $10 \log_{10} \frac{5,000,000}{500} = 40$ db more than in the case of the conventional radar receiver of 5 Mc IF bandwidth.

7-10 GAIN VARIATION AND GAIN SETTING

Gain of an amplifier stage which does not incorporate feedback is equal to the product of the transconductance of the vacuum tube and the transfer impedance of the network which the tube drives. Instability of these parameters results in gain variation. The effective signal transconductance of a tube is proportional to the d-c current through the tube. Gain can therefore be stabilized by operating the tubes so that the d-c plate current is stabilized. This can be accomplished by means of large cathode resistors or by operating a number of stages in series d-c connection. However, the first method introduces transient recovery problems and the second method reduces dynamic range. Application of conventional feedback stabilizing techniques may be employed, but in high-frequency IF amplifiers it is usually limited to a small amount of signal-current feedback which is employed to compensate for input admittance variations of the tube.

Network transfer impedance variations are on the order of ± 0.5 db; and when a small amount of d-c current stabilization is employed with the vacuum tube, the variation in signal transconductance is on the order of ± 1.0 db.

Stage gains are limited by bandwidth requirements in the case of wide-band stages, and stability requirements in the case of narrow-band stages. In addition, restrictions are usually encountered in gain distribution through the receiver as a result of dynamic range requirements. Typical average stage gains in a receiver are between 6 and 20 db incident to these limitations. An amplifier providing 100 db gain might therefore require about 10 tubes. Since the variation in gain of each stage is on the order of ± 1.5 db, 15 db reserve gain is required in the design, and provisions for controlling the maximum gain of the amplifier over a 30-db range is required. These gain-control variations do not include the gain control that is required to accommodate target signal variations. The gain setting may take the form of a noise AGC loop which controls the current of several tubes or a manual adjustment which is periodically set.

7-11 BANDWIDTH AND DYNAMIC RESPONSE

A criterion sometimes employed for best signal-to-noise ratio is that signal plus noise should be filtered by a network which maximizes the peak signal-to-rms noise power. The network which will accomplish this result was determined in Paragraph 5-10 to be simply the conjugate of the signal spectrum; that is, the receiver filter should be "matched" to the signal. In the case of the noncoherent pulse radar, each pulse must be considered as a separate entity; therefore the optimum predetection filter is a bandpass filter shaped like the RF pulse spectrum envelope. The IF characteristics usually employed for maximum detection in thermal noise are reasonable

approximations to this value when the 3-db bandwidth is approximately $1.2/(\text{pulse length})$.⁶

Additional filtering can be applied to the postdetection or video signal when there is more than one pulse. A series of periodic pulses will have a spectrum consisting of a number of harmonics. The filter which is matched to such a signal will be tuned to these harmonics so as to amplify them and attenuate the intervening noise. Because of the shape of the frequency response of such a filter, it is sometimes called a *comb filter*. It is often more convenient to obtain the effect of a matched filter by operating in the time domain. The comb filter, which is appropriate for a series of pulses, can be simply represented by adding the pulses after they are delayed by appropriate multiples of the repetition period. This operation is normally called *pulse integration* and, for search radars, is often performed by the phosphor of a B or PPI scope display. When the more elaborate technique of time domain filtering is utilized, it is sometimes referred to as *signal correlation*. A more detailed discussion of matched filters is given in Paragraph 5-10.

In selecting a bandwidth characteristic for the receiver, three considerations must be made over and above signal to thermal noise:

1. Adjacent channel (frequency) attenuation and discrimination against clutter
2. Compatibility of transient response with required resolution
3. Large signal operation

The usual response characteristics that might be encountered were indicated in Paragraph 7-7. The transient response of these networks governs the resolution and large-signal behavior. The rectified envelope of this response corresponds to the video signal.

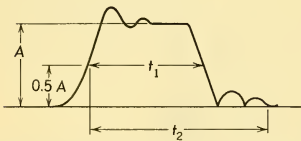


FIG. 7-9 Typical Output Transients as They Appear on the Rectified Envelope of the IF Response to an IF Pulse Input.

A typical transient response would appear as shown in Fig. 7-9. At the receiver output, a loss in sensitivity may occur for the time t_2 shown in Fig. 7-9 if the signal becomes sufficiently large that amplifier stages are driven into saturation. In pulse doppler radar receivers this is a more serious problem than in conventional radar receivers.

When all of the networks have identical transfer impedance of the form

$$K \frac{s}{(s - s_1)(s - s_1^*)}$$

⁶See J. I. Lawson and G. F. Uhlenbeck, *Threshold Signals*, Vol. 24, Sec. 8-6 (Radiation Laboratory Series), McGraw-Hill Book Co., Inc., 1950.

there is no overshoot when the pole frequency and the carrier frequency are coincident. The envelope response of such a network is given by

$$e_o(t) = K \left[\frac{1 + \left(\exp \frac{\omega_o t}{\mathcal{Q}} \right) - 2 \left(\exp \frac{-\omega_o t}{2\mathcal{Q}} \right) \cos (\omega - \omega_o)t}{1 + \mathcal{Q} \left(\frac{\omega}{\omega_o} - \frac{\omega_o}{\omega} \right)^2} \right]^{1/2} \quad (7-36)$$

where \mathcal{Q} is the effective circuit \mathcal{Q}

ω_o is the pole frequency

ω is the carrier frequency of the step sinusoidal input.

Note that oscillatory terms are involved when the carrier is detuned from bandcenter. These terms are relatively insignificant, however, for the amount of detuning that would normally be tolerated. When the circuits are not identical but are stagger-tuned, then the response given by Equation 7-36 becomes important. If the oscillatory signal is sufficiently large, the output of the following stage may be blocked for a period of time in excess of twice the duration of the input signal. To minimize these effects, inter-stage bandpass networks are usually employed which are symmetric about the IF center frequency.

7-12 SNEAK CIRCUITS

When considering the dynamic response of the receiver, it is not sufficient to consider only the performance as a bandpass filter with saturation effects under large-signal input. The transmission characteristics of the amplifier in the low-frequency region of the spectrum must also be considered.

To realize practical high-gain bandpass amplifiers the power supplied to the stages must not be derived from a common-source impedance, since instability will result. Fig. 7-10 shows a typical arrangement of IF stages. The power leads are brought into the amplifiers near the output. Decoupling filter elements $C_1 C_2 C_3 R_1 R_2 L_1 L_2$ are employed. The decoupling is designed so that a single stage will exhibit adequate gain and phase margin over the entire frequency spectrum when the stage is examined as a feedback amplifier. In particular the stability margin must be realized when the tubes operate at the peak transconductance values that would be produced by a saturating signal.

Time domain effects must also be considered. Saturating signals cause the d-c currents to the various tube elements to vary. The cathode circuits will attempt to degenerate the effects of a saturating signal during the time that the signal exists. When the signal input ceases the cathode capacitor is charged to the value which has reduced the gain during the signal on time.

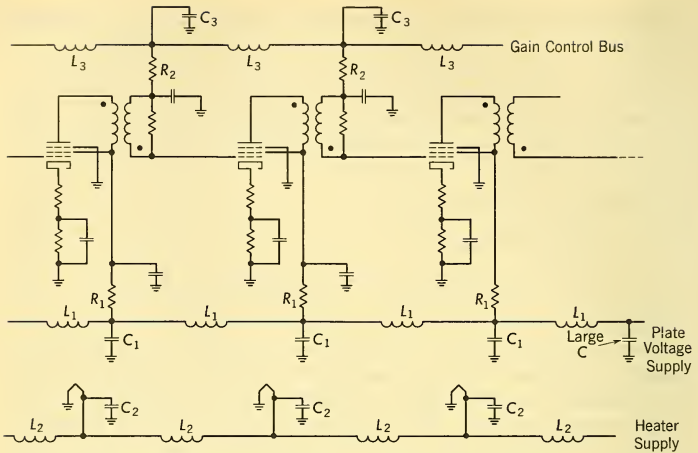


FIG. 7-10 Typical Arrangement of IF Stages Showing Arrangement of Decoupling Circuits and the Feedback Paths Thereby Introduced.

To obtain maximum receiver sensitivity the charge must be removed. This removal occurs with a nonlinear time constant

$$\frac{C_k}{\frac{1}{R_k} + g_m(t)}$$

Short time constants must be used to avoid gain modulation of desired signals when there are large undesired signals such as clutter appearing in the receiver. Grid current may also produce a similar situation, and the time constants must be kept short while at the same time providing sufficient decoupling at low frequencies and at bandpass frequencies.

The plate circuit decoupling is perhaps more critical than the other circuits. With a ladder decoupling chain, the d-c path must be kept low in resistance so that the plate voltage is not dropped excessively. Inductors are therefore used as the series elements. The elements nearest the power input connection have the currents of several tubes flowing through them. When several stages are driven into saturation, each of the stages will send a transient input into the decoupling chain. This transient propagates along the chain and may result in a very complicated transient at the last stage which can gain modulate that stage, causing undesirable transient gain variations following strong pulse signal inputs. To avoid this phenomenon, adequate filtering is provided between the ladder tapping point and the tube.

7-13 CONSIDERATIONS RELATING TO AGC DESIGN

The AGC of the radar may be of two types: (1) a fast AGC which prevents saturation of the receiver or (2) a slow AGC associated with a single target echo. In the radar receiver employed for tracking, AGC circuits of the second type are required. The IF amplifier is one of the limiting factors in the design of a high performance AGC. This subject will be discussed at greater length in Paragraph 8-21.

In designing the IF amplifier great care must be taken to examine signal distribution in the amplifier as a function of the AGC voltage. The AGC voltage must be applied to the amplifier in a manner that will result in minimum signal distortion and limited degradation of the output signal-to-noise ratio of the receiver. For example when the input signal-to-noise ratio is +90 db, it is necessary to reduce the gain in early stages to minimize distortion, and as a result noise from latter stages becomes significant. A typical design might allow the output signal-to-noise ratio to be +30 db minimum for +90 db input signal-to-noise ratio.

For minimum distortion of the modulation on the signal as the gain of an amplifier stage is varied by AGC, it is desirable that the transfer characteristic be a square-law when signal and gain control are applied to the control grid. When sharp cutoff tubes are employed for gain control, considerable distortion is sometimes experienced when gain control is provided for large-signal inputs. Restriction of the gain control to about 10 db per stage in these cases usually results in acceptable signal envelope reproduction. Output stages of the amplifier should operate with linear plate transfer characteristics. This allows the IF signal voltages applied to the last few gain-controlled stages to be small, thereby resulting in less distortion. In addition, wider bandwidths can then be employed in these stages, since filtering of the undesired spectral components of the modulated signal, which result from passing the signal through the nonlinear plate transfer characteristic required for constant incremental gain as a function of AGC voltage, is not required. In a typical case the gains in the IF may be 10 db per stage. Requiring 2 volts rms at the IF envelope detector, the minimum signal voltage at the third from the last stage of the amplifier would be 0.2 volt rms if gain control is not applied to the last two stages. The maximum signal on the controlled stage then depends on the gain reduction allowed. By controlling a number of stages the maximum gain reduction required in any one stage can be limited to something on the order of 10 to 20 db. It is necessary to examine the signal transmission through each stage for the maximum signal allowed at the input of the stage as a result of the distribution of the AGC control voltage. The AGC and transfer impedance of the stages are then arranged to provide a specified allowable distortion of the modulation on the signal appearing at the amplifier output.

In the early stages of the receiver, care must be exercised in applying AGC. When a cascode type input amplifier is employed, relatively large voltages may appear at the input grid and also the third tube grid in receivers which must provide target tracking at very short ranges. An AGC voltage is therefore applied to the first tube in these cases. However, in order that the output signal-to-noise ratio of the receiver shall not be seriously degraded, this AGC is usually not applied at the same input level as the AGC on the other gain-controlled stages but is delayed until the input signal-to-noise ratio is about 20 db. The AGC voltage delivered to the cascode is selected so as to minimize the third-order coefficients of the tube transfer characteristic. The effective cascode transfer characteristic is somewhat superior to that of a single tube because of the d-c series connection which allows control of the current of both tubes. Controlling the current of two tubes in a cascode arrangement has the advantage that the stability is not impaired at low gain. When only one tube is controlled, the grounded grid section may become unstable because of the reduced source conductance which drives it. A disadvantage of controlling the current of two tubes exists; not only does the conductance of the output of the first tube decrease, but the input conductance of the grounded grid section also decreases, thus narrowing the intercascode coupling bandwidth.

Plate and screen grid control for AGC is attractive but reduces the dynamic range of the amplifier stage for large signal input. The operating point can be maintained at a value which minimizes the third-order coefficient, but signal suppression occurs when the signal peaks drive the control grid into cutoff and into grid current.

Suppressor grid control is very attractive, since the third-order curvature can be minimized without sacrificing dynamic range. One difficulty is that the power dissipation of the screen grid is usually exceeded under strong signal conditions.

For a high-performance system the AGC voltage will be staggered, i.e., the amount of AGC voltage applied to the various controlled tubes of the amplifier will be different. This is required to obtain minimum envelope distortion. The AGC decoupling circuits must be designed with the precautions noted in Paragraph 7-12. In particular, the transmission of the IF amplifier at low frequencies must not be significant — i.e., it must operate only as a carrier amplifier.

7-14 PROBLEMS AT HIGH-INPUT POWER LEVELS

In an airborne radar set strong signals are obtained from short-range targets, clutter, and other radar signals. Two situations occur in the receiver. In one case the receiver may be operating at maximum gain and be required to furnish output from signals having an input power of the

order of magnitude of the receiver thermal noise. In the other case the receiver is required to furnish an output from a single signal which has been time-selected.

In the first case, cross modulation caused by the strong signals can deteriorate the weak-signal performance; the extent to which this occurs is a function of the detailed receiver design. If the receiver is linear, the dynamic range for any particular gain setting will usually be between 10 and 20 db. Signals more than 20 db greater than the thermal noise level can be expected to cause saturation in the receiver. The result of the saturation is a paralysis of the receiver for a certain time following the removal of the large signal. To minimize this effect it is necessary that attention be given to the circuits mentioned in Paragraph 7-12, so that a suitable transient characteristic is obtained from the IF amplifier. The transient should exhibit small overshoot and short delay time. Loss of weak signals occurs only when they are time-coincident with the strong signals if adequate IF filtering is provided. In cases where signal information is required and when the interference and signal occur at the same time (range), saturation must be prevented and the two signals separated on the basis of their difference in frequency spectra caused by the doppler shift. In the non-coherent pulse radar this is accomplished by heterodyning the weak signal against the strong signal at the IF second detector.

The second case occurs when a signal is being tracked. The desired signal is gated and may provide range and direction signals from sidebands associated with each of the pulse signal sidebands. The effect of strong signals is to add additional sidebands at the receiver output and thereby cause errors in the range and direction signal. In a well-designed receiver, negligible intermodulation occurs when a strong signal is present which is not time coincident with the desired signal.

In some instances the desired signal power level may approach the order of magnitude of the local oscillator signal power. Fig. 7-11 shows the transfer characteristic of a typical microwave mixer at large-signal levels. The nonlinearities of this characteristic will cause signal distortion. Intermodulation components appear incident to the beating of the various signal components. These components are not highly significant except with some propeller-driven targets in which terms of the order $2\omega_1 + \omega_2$ may introduce more fluctuation in the final bandwidth of the system. The reduction in modulation percentage of the pulse signal at the fundamental modulating frequency results in deterioration of tracking performance, since it corresponds to a change in tracking loop gain. In many cases the signal at the antenna terminals is greatly distorted before it reaches the signal mixer because of the time varying attenuation of a gas discharge TR tube. A controlled TR characteristic is therefore sometimes used to advantage to minimize the deterioration in tracking loop performance.

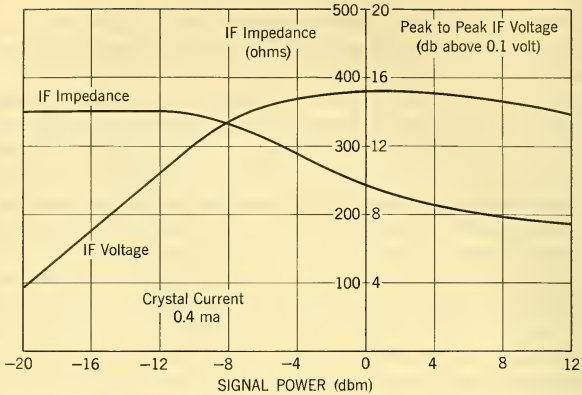


FIG. 7-11 Transfer Characteristic of a Microwave Mixer at Large-Signal Input Levels (1N23C Crystal).

7-15 THE SECOND DETECTOR (ENVELOPE DETECTOR)

An *envelope detector* is employed to produce an output voltage which corresponds to the envelope of the IF signal. The envelope detector is actually a mixer in which the sidebands of the signal are heterodyned against the signal carrier thereby producing as one output the modulation that existed on the IF signal. In the ordinary noncoherent pulse radar set, a diode detector is frequently employed. A typical circuit is shown in Fig. 7-12, together with the current-voltage relations that exist under large-signal conditions. A pulse of IF voltage is indicated as being applied to the detector. A large diode current pulse flows for a short time following the application of the signal. Capacitor C_0 is a relatively low impedance at the IF frequency compared to R_0 , and RFC is a high impedance to these frequency components; therefore negligible voltage appears across R_0 due to the IF frequency components and their harmonics which appear in the diode current. The average value of the current pulse, however, does produce a voltage across R_0 . This voltage builds up at a rate dependent on the capacitance $C_0 + C_1 + C_2$ and the diode resistance, and reaches an average value E_{dc} as shown in Fig. 7-12. The diode only conducts when the instantaneous voltage applied to the diode exceeds E_1 . As shown, conduction during time ab occurs and the capacitance $C_0 + C_1$ is charged at a rate dependent on the diode resistance and this capacitance. When the IF pulse ceases, the diode is back-biased and returns to the unbiased condition with a time constant $R_0(C_0 + C_1 + C_2)$. [The effect of the inductance of the RFC on this transient is usually negligible when the product of pulse length

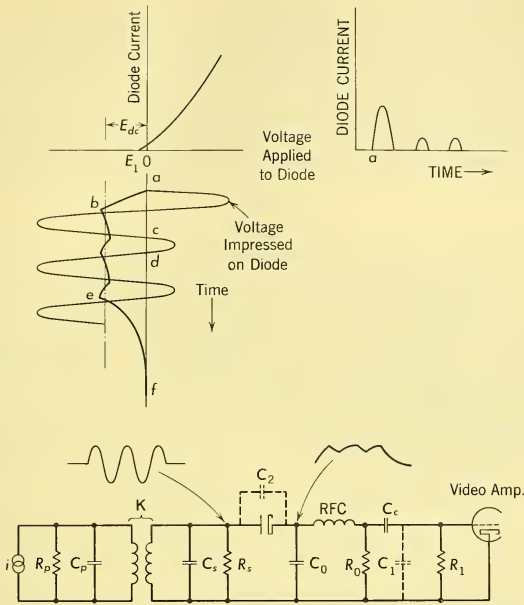


FIG. 7-12 Typical Second Detector Circuit.

times IF frequency is greater than 50. As a result of this operation it is necessary that $R_0(C_0 + C_1 + C_2)$ be considered as a low-pass filter establishing the video bandwidth.]

The efficiency of the diode detector is the ratio of the d-c voltage (E_{dc}) to the peak carrier voltage applied to the circuit. The efficiency depends on the ratio of the diode resistance plus source resistance of the IF network as seen by the diode to the load resistor R_0 . Efficiency, however, also depends on the ratio of the load resistance to the reactance of $C_0 + C_1$ at the IF frequency. In practice C_0 is usually on the order of 10 to 20 μmf . Smaller values of C_0 result in less voltage impressed on the diode because of the division of voltage between C_0 and C_d . R_0 is then selected on the basis of video bandwidth requirements. A typical example is a requirement that the video bandwidth be 10 Mc with a network impedance as seen by the detector of approximately 500 ohms and a capacitance C_1 of 10 μmf . The value of R_0 is then fixed by C_1 and the smallest value of C_0 that can be employed. Assuming C_0 to be 10 μmf , R_0 is required to be 796 ohms. An R_0

of 750 ohms would be used. The efficiency of the detector would be 0.21, assuming a diode resistance of 200 ohms⁷ and a 60-Mc IF frequency. A gain loss of 13.6 db is thus exhibited by the detector. This is a typical loss; the loss usually ranges between 6 and 15 db, depending on the video bandwidth and IF frequency involved.

An important design consideration is the loading on the IF network produced by the detector. An approximation of this loading is given by

$$R = \frac{R_0}{2\eta} \quad (7-37)$$

where R is the IF network loading and η is the efficiency of rectification. Efficiency of rectification depends on the diode resistance R_d plus the IF signal source resistance. Since the value of R_d depends on the voltage applied to the diode, the detector is nonlinear at low levels. A typical second-detector characteristic is shown in Fig. 7-13. Reproduction of the modulation on a PAM (pulse amplitude modulated) signal depends there-

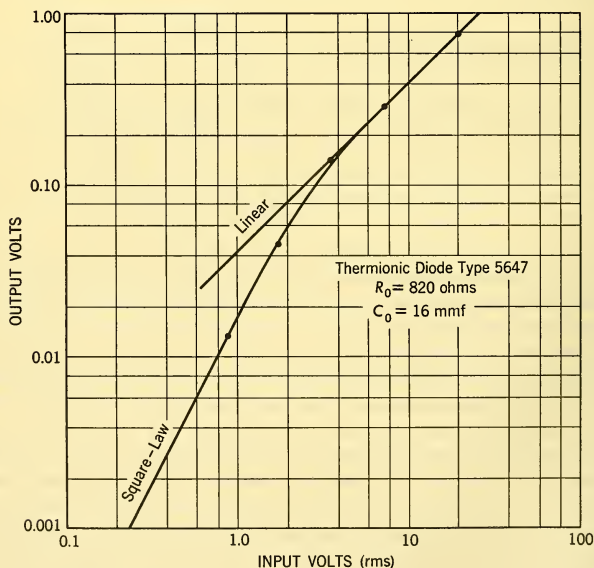


FIG. 7-13 Transfer Characteristic of a Typical Wide-Band Envelope Detector.

⁷Determination of efficiency and input impedance is relatively complicated. Methods for determining these quantities may be found in K. S. Sturley, *Radio Receiver Design*, Vol. 1.

fore on the carrier level of the IF signal applied to the detector. With high percentage of modulation, the negative peak modulation is distorted incident to the nonlinearity of the detector at low levels.

In receivers which provide considerable pre-detection integration (IF bandwidths of a few kilocycles per second) it is feasible to obtain high detection efficiency by use of large R_0 and C_0 . When amplitude modulation on the signal must be recovered in such receivers, it is required that R_1 and R_0 satisfy the relationship

$$\frac{R_1}{R_0 + R_1} > m \quad (7-38)$$

where m is the highest modulation percentage that must be recovered without distortion. Failure to satisfy this condition results in clipping of the negative peaks of the modulation.

When the signal-to-noise ratio of the IF signal is very small and the video bandwidth is less than the IF bandwidth, signal suppression occurs in the second detector.⁸ This is the result of noise-noise intermodulation at the detector. An approximate expression for signal suppression is

$$\text{db suppression} \cong -7 + 20 \log_{10} (S/N)_{IF}. \quad (7-39)$$

It is desirable to provide as much filtering as possible prior to envelope detection to minimize sensitivity loss caused by this signal suppression. However, predetection selectivity is limited by the stability of the IF filters and the tuning accuracy of the receiver. Some receivers, e.g. logarithmic receivers, do not employ a diode envelope detector but obtain the envelope by infinite impedance detection or plate detection in each of the IF stages.

In monopulse receivers the IF detector which is employed to obtain angular error signals is usually a balanced modulator. This may take the form of either a phase detector or a synchronous detector. Such detectors ideally produce an output only when both signals are applied. The output is primarily dependent on one of the two signals present at the input (provided one signal is much larger than the other). If one of the signals, such as the sum signal in a monopulse receiver, is heavily filtered before applying it to the demodulator, significant improvement in detected S/N can be realized for low S/N referred to the difference signal IF bandwidth. Such filtering, however, requires time selection of the sum signal before it is applied to the detector. Such a scheme is, in effect, a carrier reconditioning and exaltation method of detection and, of course, reduces the information rate of the radar.

⁸S. O. Rice, "Mathematical Analysis of Random Noise," *Bell System Tech. J.* **23**, 282-236 (1944), **24**, 46-156 (1945); "Response of a Linear Rectifier to Signal and Noise," *J. Acoust. Soc. Am.* **15**, 164 (1944).

7-16 GATING CIRCUITS

Gating circuits are employed to improve the signal-to-noise and signal-to-clutter ratios at the output of the receiver. A gating circuit consists of a modulator to which the signal and the gating signal are applied. In most applications the only output desired is the intermodulation between gating signal and desired signal. To accomplish this, balanced modulators are required. At video frequencies, such circuits are difficult to realize, the dynamic range usually being small. At IF frequencies such circuits are more easily provided, and dynamic ranges greater than 50 db are common. The choice between the IF and video gating depends on the nature of the signals to be encountered by the radar receiver. Typical gating circuits for video and IF applications are shown in Fig. 7-14. Gating circuits are

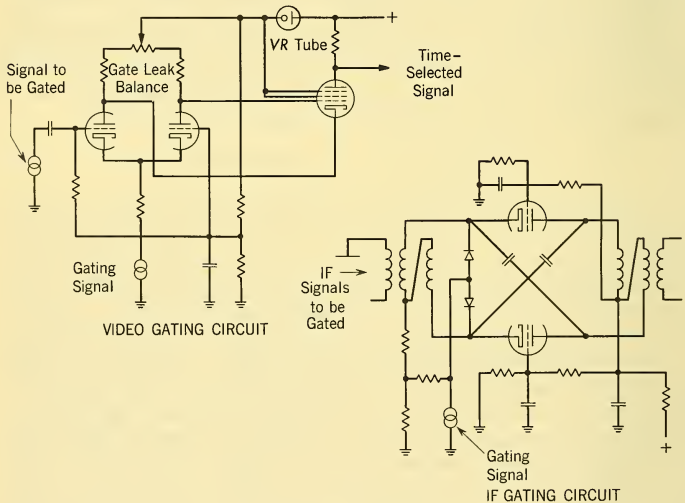


FIG. 7-14 Typical Gating Circuits.

employed having gate lengths equal to the range displayed on an indicator and also with lengths equal to or somewhat less than the transmitted pulse.

When a dynamic range greater than 50 db is required from a gating circuit, component selection is required. This is a result of uncontrolled cutoff characteristics of vacuum tubes that must be utilized. When gating occurs in the IF amplifier, spurious signals are always encountered. These spurious signals occur because it is difficult to suppress completely all of the modulating signal (gate pulse) at the output of the gater. The gating pulse

is not usually coherent with the IF signal. The higher frequency components of the gating signal are the signal components which cannot be adequately filtered. Transients caused by the modulating signal will generally produce outputs from the IF filter when rectangular gate pulses are employed. Noise modulation of the desired signal results from these transients.

In typical designs the noise modulation caused by the transients is at least 40 db below the signal. In addition, the seriousness depends on the signal processing following the gating. Appropriate sampling of the gated signal prior to integration reduces the noise to a negligible value. In video gating circuits the modulating signal is coherent with the detected signal. Thus the noise mentioned does not occur.

7-17 PULSE STRETCHING

In tracking radars it is required that the modulation signal associated with a pulse-amplitude-modulated signal be recovered. The modulated pulse signal is

$$f_1(t) = [1 + m \cos(\omega_m t + \phi)] \sum_{n=-\infty}^{\infty} c_n e^{j2\pi n t / T} \quad (7-40)$$

for periodic pulses of shape $f(t)$

$$\text{where} \quad c_n = \frac{1}{T} \int_0^T f(t) \exp\left(\frac{-j2\pi n t}{T}\right) dt. \quad (7-41)$$

If the pulses are passed through a low-pass filter having a cutoff frequency below the first harmonic of the pulse, the modulation is recovered and will have an amplitude $m(t/T) \cos[\omega_m t + \phi]$. Since t/T typically may be on the order of $\frac{1}{1000}$ this is a very inefficient process. *Pulse-stretching circuits* are therefore used to lengthen a series of pulses without changing the relative pulse amplitudes in order to obtain more gain in the process of recovering the amplitude-modulating signal. For most efficient demodulation the pulse is lengthened for a full period. In either case — whether a pulse is simply filtered or is lengthened and then filtered, time selection of the pulse is required prior to the lengthening to prevent cross modulation by undesired pulses.

A pulse lengthener converts the modulation function $1 + m \cos(\omega_m t + \phi)$ into a new function $F(t)$. Two types of lengtheners are used. In one, $F(t)$, is set at a fixed reference level prior to a signal pulse input; in the other the output is changed from the value measured to the new value. Typical circuits of these lengtheners are shown in Fig. 7-15. The lengthened pulse on which the desired signal is modulated is an exponential pulse. The decrement is small and approaches zero in many practical cases.

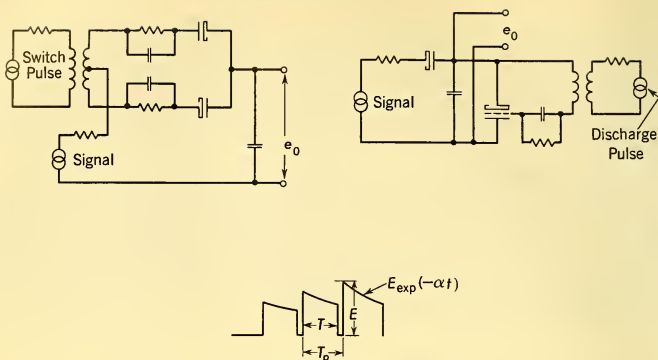


FIG. 7-15 Pulse Lengtheners.

The output spectrum of the lengthener for the case where $\alpha = 0$, and $T = T_p$ is given by

$$\begin{aligned}
 |W_f| = & 1 + 2m \sin \frac{\omega_m T_p}{2} \left[\frac{1}{\omega_m T_p} \cos(\omega_m t + \phi) - \frac{\omega_m T_p}{2} \right] \\
 & + \frac{1}{T_p} \sum_{n=1}^{\infty} \frac{\cos \left[\left(\omega_m - n \frac{2\pi}{T_p} \right) t + \phi + \frac{\omega_m T_p}{2} \right]}{\omega_m - \frac{n 2\pi}{T_p}} \\
 & + \frac{\cos \left[\left(\omega_m + n \frac{2\pi}{T_p} \right) t + \phi + \frac{\omega_m T_p}{2} \right]}{\omega_m + \frac{n 2\pi}{T_p}}.
 \end{aligned} \tag{7-42}$$

When the output from the lengthener is passed through a low-pass filter, the first term becomes the only significant term in the output if the period of the modulating signal is much greater than T_p . If the low-pass filter has a cutoff frequency ω_c , outputs are also obtained for modulation frequencies satisfying

$$\omega_m + n \frac{2\pi}{T_p} \leq \omega_c. \tag{7-43}$$

7-18 CONNECTING THE RECEIVER TO THE RELATED REGULATING AND TRACKING CIRCUITS

The receiver must provide signals to range or speed error detectors, and angular error detectors. It is desired that the outputs of these detectors have a stable characteristic with time and with input power level to the

receiver. If the AGC demodulator is connected to the range error detector, and the angle demodulator connected directly to the AGC demodulator output, both range and angular error characteristics will be determined by the AGC regulation.

It is desirable, however, that the video signals applied to these demodulators be as large as possible to minimize the bias errors resulting from contact potential in the demodulators. Frequently separate filtering of the range and angle video signals may be performed. A single AGC loop operating from the angle channel controls the receiver gain. To obtain a stable range-error detector characteristic, the video amplification between the input to the range detector and the AGC demodulator must then be stabilized by feedback. A typical arrangement is shown in Fig. 7-16.

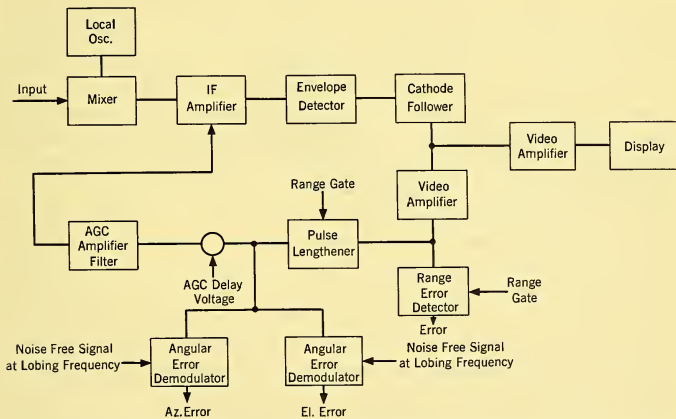


FIG. 7-16 Connection of a Receiver Employed with Sequentially Lobed Antenna to Related Circuits.

7-19 ANGLE DEMODULATION

The antenna tracking error signal can be considered to be proportional to the magnitude of the fractional modulation of the signal resulting from division of the difference signal by the sum signal. In monopulse radar sets the sum and difference signals are separated at the receiver input, whereas in a conical scanning radar the composite signal is passed through the receiver. To obtain a tracking error signal from a monopulse radar, the difference signal is heterodyned with the sum signal, which is effectively a noisy carrier signal. In some cases carrier reconditioning may be performed

and the difference signal heterodyned against a filtered carrier signal. This latter operation, however, is accomplished only with sacrifice of the information rate.

In a conical scanning radar the desired target is selected by range gating so that the other targets, which are also PAM signals, will not be demodulated. The signal is then envelope-rectified and lengthened. Lengthening is employed to minimize additional modulation resulting from PRF variation. The signal at the output of the pulse lengthener still represents the composite signal, i.e. the sum and the difference signal. The low-frequency modulation of the composite signal is caused by the scintillation noise of the target and is independent of the lobing frequency. Both the d-c component of the signal and the low-frequency modulation are fed back to the IF amplifier as a gain-control signal. Modulation at the lobing frequency, however, is not allowed to effect a gain control of the receiver. The signal at the output of the pulse lengthener thus contains primarily the sidebands about the lobing frequency which are caused by the variation in direction of arrival of the signal. To demodulate this signal, and provide control signals for the antenna servo, the signal is multiplied by a noise-free carrier at the lobing frequency. The carrier signal is phase-locked with the antenna lobing. This is usually accomplished by means of an a-c generator mechanically linked to the rotating antenna.

Fig. 7-17 shows three typical demodulator circuits. In all three of these circuits neither the signal nor the carrier frequency appears in the output. The output contains only the beats between the signal and the carrier and certain of their harmonics. Of the three demodulator circuits shown the "ring modulator" is the most desirable because the modulation products are effectively separated in various parts of the circuit. The carrier signal should be as monochromatic as possible for maximum output signal-to-noise ratio.

The process of pulse lengthening merely concentrates all of the noise appearing in the IF in a region less than the PRF. In order that the noise reduction provided by the antenna servo be approximately B_1/PRF , where B_1 is the noise bandwidth of the antenna tracking loop, it is necessary that the demodulator provide a true product demodulation. To approach this performance the ring modulator is employed in conjunction with a bandpass filter which filters the signal applied to the demodulator.

7-20 SOME PROBLEMS IN THE MEASUREMENT OF RECEIVER CHARACTERISTICS

Noise Figure. The most practical method of making noise figure measurements involves the use of a dispersed signal source. An argon-filled gaseous discharge tube will produce a standard noise power output equiva-

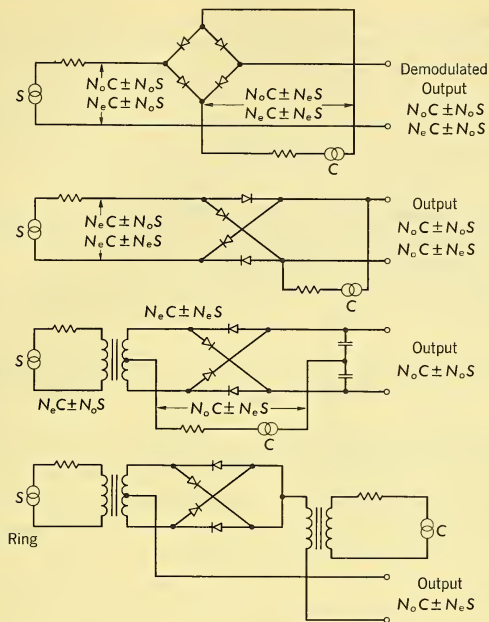


FIG. 7-17 Angular Error Demodulators.

lent to a source temperature of 9775°K . Measurement of noise figure merely involves the measurement of the noise power required to double the output noise power of the receiver under test.

A precision microwave attenuator is used to control the noise power applied to the receiver. The available noise power from the discharge tube is equivalent to a noise figure of 15.28 db referred to a temperature of 290°K . The noise figure is determined by merely subtracting the attenuation required to produce a doubling of the noise power from 15.28 db (corrections for spurious signal response are required).

Several problems arise in this type of measurement. If the noise power output of the receiver is allowed to double, it is necessary that the receiver be linear at the two output conditions and that the response of the detector to noise be known. It is desirable that a 3-db loss be inserted in the receiver rather than let the output noise level change. The 3-db loss must be inserted at a point in the receiver which is preceded by sufficient gain that noise sources following the pad do not contribute to the output. The receiver also

must be linear ahead of the pad. Frequently it is not convenient to provide an accurate 3-db loss in the receiver. An example is the case where the preamplifier and main IF amplifier are contained in a single unit. In such cases an arbitrary attenuation may be introduced by means of the manual gain control. Measurements are made with two arbitrary output levels; it is only necessary that the receiver have a linear transfer characteristic to the noise at the selected levels. The method is as follows.

1. Observe output deflection (d-c voltmeter or milliammeter at the second detector) with no additional noise input. Let the deflection be d_1 . The noise is incident to N_{rec} .
2. Introduce the noise source and adjust the noise power (N_1) applied to the receiver to produce a deflection d_2 . The noise is incident to $N_1 + N_{rec}$.
3. Insert attenuation α by means of the manual gain control so that the noise N_1 produces the deflection d_1 . The noise is incident to $(N_1 + N_{rec})\alpha$.
4. Increase the output from the noise source (N_2) to produce the deflection d_2 . The noise is incident to $(N_2 + N_{rec})\alpha$.

From these observations, the noise figure can be determined from

$$F_{rec} = \frac{T_1^2}{T_a(T_2 - 2T_1)} \quad (7-44)$$

where $T_a = 290^\circ\text{K}$ and T_1 and T_2 are noise temperatures corresponding to N_1 and N_2 .

In making a noise figure measurement with a dispersed signal source, difficulty is experienced with spurious responses of the receiver. In broad band receivers it is usual to add 3 db to the measured result to account for beating at the image frequency. Because of the small available power from the noise source it is necessary to couple directly to the antenna terminals of the receiver rather than through a directional coupler. As a result the noise figure is not usually measured with the transmitter operating in the case of airborne radar sets. The measurements are also correct only if the noise source has the same impedance as the antenna.

Sensitivity. With the transmitter operating, additional noise may appear which will degrade the performance. This is particularly the case with high-PRF doppler radar receiving systems. To determine the performance in detecting and tracking small signals a *sensitivity measurement* is generally made; this is a measure of the least signal input capable of causing an output signal having desired characteristics.

In the case of a radar display it is a simple matter to determine the signal power required to obtain a minimum discernible signal. The signal is

obtained from a standard signal generator which can provide the same modulation characteristics as the radar target. In a noncoherent radar the sensitivity is measured at various ranges. At minimum range the sensitivity is usually reduced owing to the attenuation characteristics of the TR tube. It is sometimes convenient to define the sensitivity of a radar by an A scope measurement. In these cases a *tangential signal* measurement is made. For a tangential signal the signal-to-noise ratio is approximately +4 db.

Measurement of the sensitivity of a tracking receiver requires that the transfer function of the loop be determined at various input signal power levels. The minimum signal power required to produce the full dynamic tracking capability of the loop is determined. Measurement involves the insertion of a fixed power level RF signal having the modulation characteristics of the radar signal, and measurement of the transfer function of the particular tracking loop for this fixed input signal level. More detail on the means for measuring the transfer functions of the regulatory and tracking systems will appear in the following two chapters.

In making sensitivity measurements, accuracy is sometimes limited by signal leakage from the standard signal generator. Frequently it is necessary to put additional shielding around the generator, and connect a second precision attenuator in the line between signal generator and receiver.

When measuring the sensitivity of very narrow band receivers such as are employed for doppler radar applications, it is usual to modulate the STALO (stable local oscillator) signal to obtain a signal source of adequate frequency stability that will remain within the narrow predetection filters. If the long-term stability of the STALO is reasonably good, a standard signal source which is crystal controlled may be used, provided the predetection bandwidth is not less than about 10^{-8} times the RF signal input frequency.

CHAPTER 8

REGULATORY CIRCUITS*

8-1 THE NEED FOR REGULATORY CIRCUITS

To determine the bearing, range, and velocity of a target with high accuracy, three basic conditions must be fulfilled by the radar receiver and its associated data processing system: (1) the desired target intelligence components of the received signal must be faithfully reproduced at the output of the receiver; (2) undesired input signals which tend to reduce the S/N ratio of the desired target intelligence must be suppressed; (3) sources of noise internal to the radar must be minimized.

The desired target intelligence appears as amplitude, phase, and frequency modulations of the received signals. The target information is extracted by taking a cross product between the received signal and a reference signal and filtering the resultant signal to remove extraneous cross products (see Paragraph 1-5). In a practical radar receiver, there are several potentially troublesome sources of degradation in these processes.

An optimum demodulation process depends upon the accuracy with which the receiver can be tuned to the incoming signal. Various environmental and electrical factors will cause receiver tuning to vary or *drift* as a function of time. Receiver tuning control or *automatic frequency control* (AFC) is therefore required to reduce the effects of such variations.

Receiver components must be operated under such conditions that the linear dynamic range of the receiver is very limited. Unless some form of *automatic gain control* (AGC) is utilized, signal distortion will take place in the receiver. For example, saturation effects will tend to erase amplitude modulation on the received signal; this in turn will cause poor tracking or loss-of-track in a conically scanning system.

A large number of vacuum tubes must be employed in the receiver to amplify the noise level to the desired output level. Variations in the tube characteristics occur when the voltages supplied to the tubes vary. The desired output signals are then modulated with the undesired fluctuations of the power supply voltages. Thus *electronic power regulation* is required.

*Paragraphs 8-1 and 8-3 through 8-13, and 8-21 are by D. J. Healey III. Paragraph 8-2 is by D. D. Howard and C. F. White. Paragraphs 8-14 through 8-20 are by C. F. White and R. S. Raven. Paragraphs 8-22 through 8-34 are by G. S. Axelby.

Motions of the aircraft carrying the radar set can modulate the incoming signal and cause loss or degradation of the target signals. *Automatic space stabilization* systems are often required to cope with this problem.

Finally, the measurement problem is complicated by externally and internally generated noise. The origins of such noise and the effects of the noise upon range- and angle-tracking accuracies are described in the next paragraph. This discussion is particularly important to the subsequent discussion of AGC in this chapter and the angle and range tracking as discussed in Chapter 9.

The remainder of this chapter will deal with the basic considerations governing the preliminary design of the AFC, AGC, and space stabilization loops. The problem of electronic power regulation is not discussed in detail since this is largely a matter of good electronic design practice, a topic beyond the scope of this volume.

8-2 EXTERNAL AND INTERNAL NOISE INPUTS TO THE RADAR SYSTEM¹

Paragraph 4-7 presented some of the basic measurements of target noise characteristics. This paragraph will define the noise sources in a form more immediately useful to the closed-loop control designer to illustrate the means for utilizing the measured information for design purposes.

External Noise Inputs. Variations in the external input to the radar system fall into two basic categories, i.e. frequency components associated with motion along the target flight path and other frequency components normally referred to as noise. Noise includes propagation path anomalies and atmospheric noise (sferics) as well as noise caused by the complex nature of the target, random motion, and reflectivity. The emphasis here is on the noise associated with the target motion and reflectivity variations that lead to tracking errors. The various components of external radar noise may be defined as follows:

Range noise, with an rms value of σ_r , is defined as deviation of the range information content in the received echo with respect to some reference point on the target. The reference point may be chosen as the long-time average of the range information. Range noise is independent of the target range since its source is pulse shape distortion caused by variations in the vector summation of energy reflected from target surface elements.

Amplitude noise, with an rms value of σ_{amp} , is defined as the pulse-to-pulse variation in echo amplitude caused by the vector summation of the echoes from the individual elements of the target. Amplitude noise, since it is

¹See J. H. Dunn, D. D. Howard, and A. M. King, "Phenomena of Scintillation Noise in Radar Tracking Systems," *Proc. IRE*, May 1959.

interpreted by the radar as amplitude modulation of the mean signal level, is independent of range if a good automatic gain control (AGC) system is used.

Angle noise, with an rms value of σ_{ang} , is defined as the variation in the apparent angle of arrival of the echo from a target relative to the line-of-sight to the center of reflectivity of the target. Angle noise is a function of the spacing of surface elements producing echoes, and the relative amplitude and phase of these echoes. Since angle noise is a function of the linear dimensions of the target, a variation inversely proportional with range results as long as the target subtended angle is small compared with the beamwidth of the antenna. At times, incident to angle noise, the direction indicated by the apparent angle of arrival of the target echo may fall outside the target extremes.

Bright spot wander noise, with an rms value of σ_{bspw} , is defined as the variations in the center of reflectivity of the target relative to a selected physical reference point on the target. The summation of angle noise plus bright spot wander noise is the variation in the apparent angle of arrival of the echo from a target relative to the selected physical reference point on the target. Bright spot wander noise is a function of the relative spacing of target reflecting elements and the amplitude of echoes from these elements. Like angle noise, bright spot wander noise (in angular units) varies inversely with range. However, the peak excursions of the center of reflectivity of the target cannot extend beyond the target limits.

Internal Noise Inputs. In addition to the primary function of location and tracking of targets in space, radar outputs to computers utilize rates of change of the basic position information. Tracking smoothness and accuracy depend upon the manner in which the external inputs are processed by the radar system. Internal radar noise components may be categorized as follows.

Receiver noise, with an rms value of σ_{rec} , is defined as the variations in the radar tracking arising from thermal noise generated in the receiver and any spurious hum pickup. Receiver noise is inversely proportional to the signal-to-noise ratio in the receiver, and since the signal power varies inversely as the fourth power of the range to the target (excluding propagation anomalies), this effect is directly proportional to the fourth power of range.

Servo noise, with an rms value of σ_{ser} , is defined as the variations in the radar tracking axis caused by backlash and compliance in the gears, shafts, and structures of the antenna. The magnitude of servo noise is essentially independent of the target and is thus independent of the range.

Tracking Noise Definitions. An optimum radar system design can result only from proper consideration of the nature of all the external and internal noise sources. One principal objective of tracking system design

may be taken as minimization of tracking noise, which may be categorized as follows:

Range tracking noise, with an rms value of σ_{rt} , is defined as the closed-loop tracking variations of the measured target range relative to the range to a fixed point on the target. Range tracking noise includes effects of the complex nature of the target and receiver and range servo system noise. Systematic range tracking errors arising from flight-path input information are excluded from σ_{rt} .

Angle tracking noise, with rms value of σ_{at} , is defined as the closed-loop tracking variations of the measured target angular position relative to a fixed point on the target. Angle tracking noise includes effects of the complex nature of the target and receiver and angle servo noise. Systematic angle tracking errors arising from flight-path input information are excluded from σ_{at} .

Range Tracking Noise. The general shape of the dispersion versus range for the various noise factors entering into range tracking is shown in Fig. 8-1. Since the various noise factors are uncorrelated, the total output

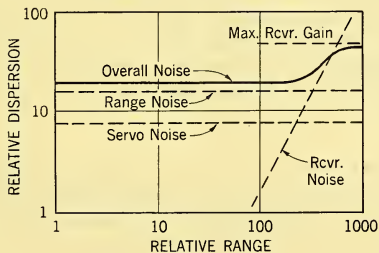


FIG. 8-1 Range Noise Dispersion Factors.

noise amplitude (shown by the heavy line) representing range dispersion in a given tracking system is found by summing the noise components in a root-mean-square manner. To use the diagram of Fig. 8-1 for prediction of system performance, at least one point on each characteristic must be determined by measurements. In the case of external range noise, the following facts are known:

1. Fire-control radar range information contains noise resulting from the finite size of practical targets.

2. The total rms range noise, σ_r (in yards), may be predicted from a knowledge of target size and shape. Measurements made with a split-video error detector on a variety of single and multiple targets show an average

rms value² of $\sigma_r = 0.8$ times the estimated radius of gyration of the reflectivity distribution of the target about its center of reflectivity. These results relate to average noise power. By nature, wide fluctuations from sample to sample may be expected with the actual value dependent upon sample time. Examples of spectral power distributions are shown later.

3. The range noise power spectra for a variety of aircraft targets in normal flight show that the significant power is below 10 cps and, in general, one-half the range noise power lies below 1 cps. The frequency components of range noise are a function of rates of target motion in yaw, pitch, and roll and are influenced by air turbulence, angle of view, maneuvering of the target, and the target type.

4. The influence on the noise values incident to the specific type of range tracking system employed has not been extensively investigated, but the values shown are believed to be typical for fire-control design purposes (assuming good system engineering and performance).

The measured spectral range noise power distributions for an SNB twin-engine aircraft, for two SNB aircraft, and for a PB4Y patrol bomber are shown in Fig. 8-2. The curves represent mean values while the upper and lower maximum excursions from the mean are shown by the arrowed lines. The analysis was based upon 80-sec samples with the indicated mean value for σ_r taken over the number of runs shown. The broad frequency range of the radar range input noise power clearly emphasizes the requirement of range tracking bandwidth minimization consistent with tracking error specifications.

Angle Tracking Noise. The general shape of the dispersion versus range for the various noise factors entering into angle tracking is shown in Fig. 8-3. The various noise components shown are uncorrelated. The rms total output noise for conical scanning or sequential lobing radar is greater^{3,4} than for monopulse⁵ (simultaneous lobe comparison) radars because of the high-frequency amplitude noise at the lobing frequency. For prediction of system performance, at least one point on each characteristic must be determined by measurement. In the case of external angle noise, the following facts have been established.

1. Amplitude noise is an amplitude modulation of the echo caused by the vector summation of echoes from the complex multielement reflecting

²D. D. Howard and B. L. Lewis, *Tracking Radar External Range Noise Measurements and Analysis*, NRL Report 4602, August 31, 1955.

³J. E. Meade, A. E. Hastings, and H. L. Gerwin, *Noise in Tracking Radars*, NRL Report 3759, 15 November 1950.

⁴J. E. Meade, A. E. Hastings, and H. L. Gerwin, *Noise in Tracking Radars, Part II: Distribution Functions and Further Power Spectra*, NRL Report 3929, 16 January 1952.

⁵R. M. Page, "Monopulse Radar," paper presented at the 1957 Institute of Radio Engineers Convention, *IRE Convention Record*, Part 8, Communications and Microwaves, p. 132.

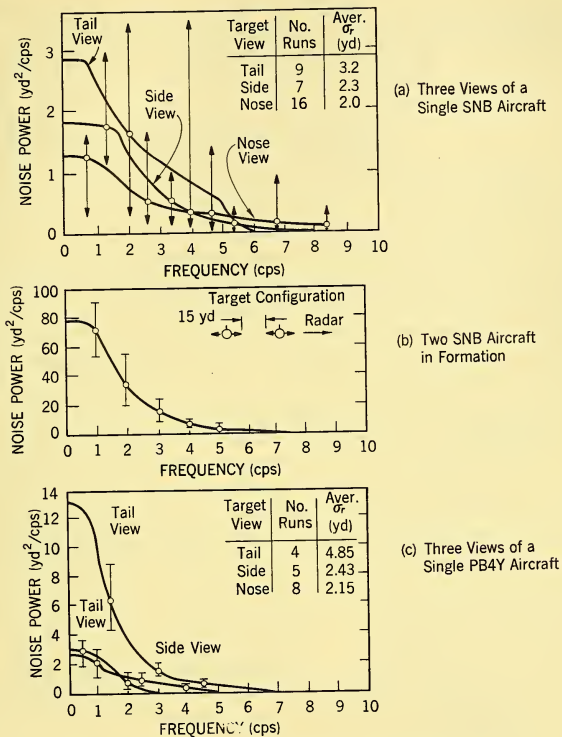


FIG. 8-2 Range Noise Power Spectral Distributions.

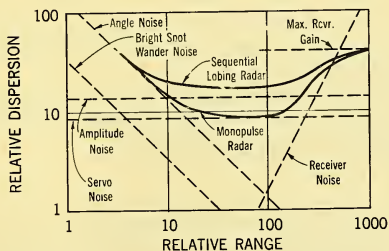


FIG. 8-3 Angle Noise Dispersion Factors.

surfaces of the target. The frequency components of amplitude noise causing angle tracking noise lie in two widely separated bands.

A low-frequency region of noise extending from zero to approximately 10 cps causes a noise modulation within the closed-loop servo-target combination superimposed upon the tracking error caused by flight path input information and associated system tracking errors. The low-frequency band also influences angle noise as explained later. Removal of the effects of low-frequency amplitude noise on angle tracking by suitable AGC design is also discussed later.

A high-frequency region of amplitude noise in the vicinity of the lobing frequency (except in monopulse radars) contributes directly to angle tracking noise. The angle tracking noise power arising from high-frequency amplitude noise is proportional to the square of the beamwidth, the fractional amplitude noise power modulation per cps of bandwidth, and the angle tracking servo bandwidth.⁶ The principal sources of target-generated high-frequency amplitude noise are propeller (power plant) modulation and structural vibrations of the target surface elements.

2. Angle noise is the variation in the apparent angle of arrival of the echo from the target relative to the line of sight to the center of reflectivity of the target. It is caused by variations in the phase front of the reradiated energy from the multielement target. When low-frequency amplitude noise exists incident to narrowband or slow AGC, the angle noise power (in suitable units) equals one-half the square of the radius of gyration of the target reflectivity distribution.⁷ When low-frequency amplitude noise is removed by wideband or fast AGC, the angle noise power is approximately doubled with practical AGC circuitry.

3. Bright spot wander noise results from changes in the center of target reflectivity principally caused by a redistribution of the significant target reflecting surfaces; it does not depend upon the relative phases of the echoes from the individual surface elements. The frequency components of bright spot wander noise lie almost entirely in a low-frequency band since it is associated with major aspect changes of the target. Because bright spot wander noise is an uncorrelated component of target-generated angle tracking noise, a complete elimination of angle noise (as defined above) does not reduce angle tracking noise to zero.

Examples of the spectral energy distribution of amplitude noise were shown in Fig. 4-23.⁸ In the spectra illustrated, the analytical method excluded low-frequency results below 30-40 cps.

⁶*Ibid.*, p. 3.

⁷B. L. Lewis, A. J. Stecca, and D. D. Howard, *The Effect of an Automatic Gain Control on the Tracking Performance of a Monopulse Radar*, NRL Report 4796, 31 July 1956.

⁸Source: D. D. Howard, from measurements made at the Naval Research Laboratory, Washington, D. C.

The effects of the spectral energy distribution of closed-loop angle noise and the contributions of low-frequency amplitude noise modulation of tracking error caused by flight path input information are discussed in Paragraph 8-17.

8-3 AUTOMATIC FREQUENCY CONTROL

Automatic frequency control circuits are employed as a means of overcoming tuning tolerance and stability problems. The operating frequency of the receiver is compared to a reference. An error signal, related to the difference between the operating frequency and the reference, is generated. The error signal is then applied to the system in such a manner as to reduce the difference to an acceptable value.

The general problem of automatic frequency control may be visualized as follows (see Fig. 8-4).

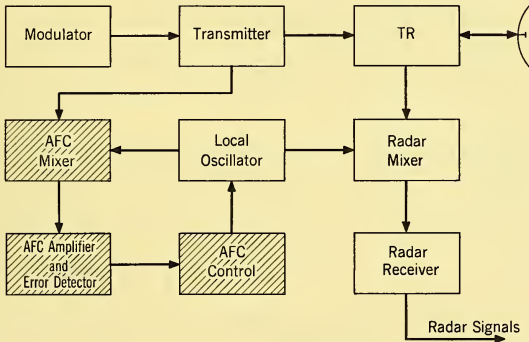


FIG. 8-4 Automatic Frequency Control in a Pulsed Radar System.

In the case of radar employing a pulsed oscillator as the transmitter, it is required that the receiver be tuned to the transmitter frequency. As discussed in Chapter 7, this is done by mixing the incoming signal with a local oscillator signal. The resulting intermediate-frequency (IF) output then is amplified by bandpass amplifiers designed to operate at a fixed intermediate frequency. With such an arrangement, the receiver tuning depends upon the ability of the local oscillator to follow variations in the transmitted frequency and thereby maintain the difference frequency (IF) at the value for which the bandpass amplifiers were designed. The automatic system employed to accomplish the desired regulation of the IF is called an *automatic frequency control* (AFC).

Automatic frequency control is accomplished by applying the generated difference frequency to an error detector whose reference is the desired IF frequency. Such an error detector is a *frequency discriminator*. The frequency discriminator provides an output whose magnitude is proportional to the error and whose polarity indicates whether the IF frequency is above or below the reference. Since a variation in either the transmitter frequency or the local oscillator frequency produces an error in the IF frequency, these variations can be suppressed by suitable control of the local oscillator frequency if they do not exceed the bandwidth limitation imposed on the feedback control loop by the pulsed data.

8-4 VARIATION OF TRANSMITTER FREQUENCY WITH ENVIRONMENTAL CONDITIONS

There are two types of frequency instability which result from the environment in which the transmitter must operate. There are relatively long-term frequency changes which occur incident to the effects of temperature, vibration, deterioration, and the like; there are also short-term frequency changes which are the result of a time-varying load impedance connected to the transmitter, and frequency modulation from the heater-supply and power-supply noise.

Since the reference in an AFC for a conventional radar set is compared with the difference between the transmitter frequency and the local oscillator frequency, corresponding variations in the local oscillator frequency occurring at the same time as the transmitter frequency variations are also important.

Fig. 8-5 shows typical frequency variation of a magnetron and a klystron with ambient temperature.

Some static and slow frequency differences for typical magnetrons and local oscillators are listed in Table 8-1.

TABLE 8-1 FREQUENCY DEVIATIONS OF TYPICAL MAGNETRONS AND KLYSTRON LOCAL OSCILLATORS

<i>Environmental Factor</i>	<i>Maximum Difference Frequency (Mc)</i>
Scatter of magnetron and oscillator frequencies as received from manufacturer	± 50
Warmup of radar set	± 1.5
Temperature	± 15
Pressure (0 to 50,000 ft) altitude	$- 2.5$
Pushing ($\pm 10\%$ line-voltage variation)	± 5.0
Aging	± 10

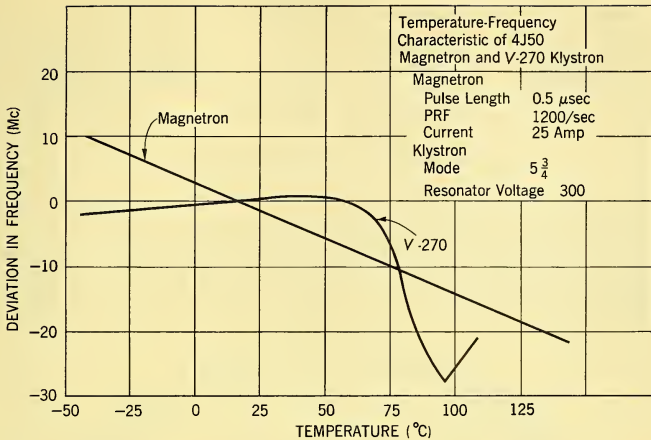


FIG. 8-5 Frequency Stability of a Magnetron and Klystron vs. Ambient Temperature.

The response time of the AFC does not have to be very great to correct for the frequency changes listed in Table 8-1. It is only required that the controlled frequency can be adjusted over the range. To obtain a wide tuning range, control of both the klystron cavity resonator and the reflector potential may be employed. In many cases only reflector control is required if periodic adjustment to accommodate frequency scatter caused by tube replacement and aging is allowed.

Table 8-1 indicates that when a radar set is first energized it is usual for the open-loop frequency error to be rather large. A wide pull in range is therefore required. (*Pull-in* is the process whereby the error in receiver tuning frequency existing at the instant of an off-frequency input signal is reduced by the AFC operation.)

8-5 MAGNETRON PULLING

The single mode equivalent circuit of a magnetron is shown in Fig. 8-6. The magnetron is considered as a conventional self-excited power oscillator with the L-C tank circuit inductively coupled to the output transmission line.

As will be discussed in Paragraph 11-1, loading mismatch can affect both the frequency and power output of the magnetron. Transient variations of the load admittance occur in scanning antenna-radome configurations. In a conical scanning radar, load admittance variations occur with feedhorn

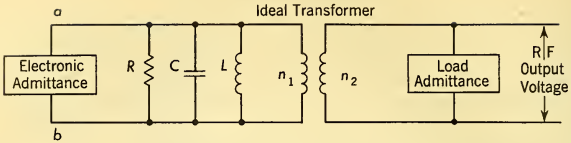


FIG. 8-6 Equivalent Circuit of a Magnetron.

rotation because of imperfect rotary joints. The nature of these transient variations governs the time-response requirements for the AFC.

The *Rieke diagram* is a fundamental performance characteristic of the magnetron which describes the dependence of oscillator power output and frequency on the load. A typical Rieke diagram is shown in Fig. 8-7.

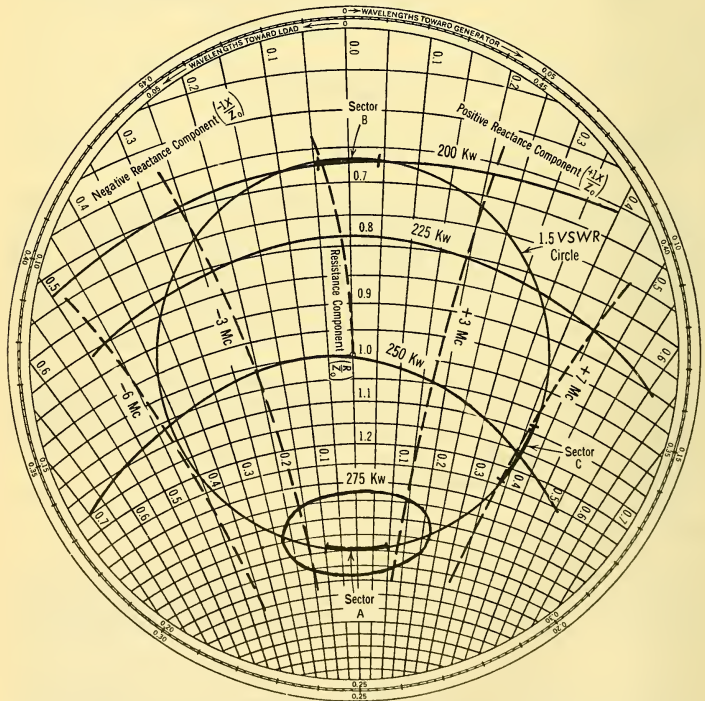


FIG. 8-7 Possible Operating Condition of a 4J50 Magnetron in a Typical Airborne Radar Set (Rieke Diagram).

Although the Rieke diagram specifies the frequency and power output for any load, the *pulling* figure of the magnetron is defined by the total frequency variation resulting from a load which produces a VSWR of 1.5 when it is changed through a phase of 360° . The load corresponding to this condition is shown as a circle in Fig. 8-7. The total frequency variation caused by pulling might therefore be 13 Mc. A typical conical scanning antenna produces relatively small phase variation. The measured phase variation of a typical load is plotted as sectors A, B, and C in Fig. 8-7. The most unfavorable position for the phase is at Sector A, for which the frequency may be pulled a total of approximately 4 Mc. Pulling can also result from a wide-angle scanning antenna looking through discontinuities in the radome as well as the discontinuities in phase of a rotating feedhorn in a conical scanning radar. Sector C is most favorable for elimination of transient frequency pulling caused by phase changes; however, power output variations are relatively large. The transmitted signal will be amplitude-modulated by this effect, and the resultant amplitude modulation on the received signals introduces errors in antenna pointing. Accordingly, Sector B represents the most favorable alternative from the standpoint of low-frequency pulling and minimum amplitude modulation of the transmitter. It will be observed, however, that these advantages are purchased at the price of lower-than-rated power output.

In a well-designed antenna-radome combination, rapid phase changes with the position of the antenna are not usually severe. In a conical scanning radar the greatest pulling effect results from the rotation of the feedhorn. The phase may change quite rapidly with feedhorn position, and the frequency of pulling is therefore high. A typical system has been observed to generate two phase rotations in one revolution of the feedhorn with some abrupt changes. The frequency variation is thus predominantly at frequencies greater than twice the lobing frequency of the antenna.

Table 8-2 gives some typical pulling characteristics of a conical scanning system.

TABLE 8-2 PULLING CHARACTERISTICS OF A
TYPICAL CONICAL SCANNING SYSTEM

<i>Frequency of FM</i>	<i>Peak Deviation (Mc)</i>
f_1 (lobing frequency)	0.5
$2f_1$	1.7
$3f_1$	0.25
$4f_1$	0.25

8-6 STATIC AND DYNAMIC ACCURACY REQUIREMENTS

Tuning errors in the radar receiver degrade the output signal-to-noise ratio of the radar. Maximum range performance of the radar is thus a

function of the tuning accuracy of the receiver. To obtain the maximum performance from a radar system the IF bandwidth must be matched to both the received pulse and the tuning error of the receiver. Fig. 8-8 shows the loss in signal-to-noise ratio (sensitivity) as a function of a static tuning error in the receiver. The video bandwidth is considered to be very much larger than the IF bandwidth.

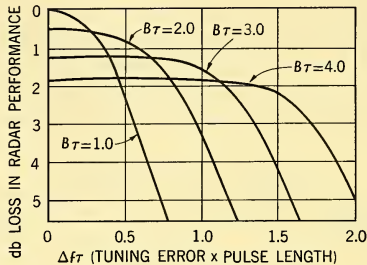


FIG. 8-8 Loss in Sensitivity with Tuning Error. After R. P. Scott, Proc. IRE (Feb. 1948) p. 185.

To obtain maximum system performance a compromise between AFC performance and bandwidth must be made to provide maximum S/N . Maximum performance is obtained with $B\tau \doteq 1.0$ when the tuning error is negligible. In a MOPA (master oscillator power amplifier) system it is feasible to realize this condition. In the case of a separate-pulsed transmitting oscillator and a continuous-wave klystron local oscillator, some allowance for the static error must be made. The reference is a frequency discriminator and associated envelope rectifiers. The reference must be tuned to the center frequency of the receiver IF and must be stable within environmental conditions encountered. With an IF frequency of 30 Mc the reference can be made to be accurate within ± 50 kc of the IF. When a master oscillator, operating continuously, is employed in the radar to obtain the transmitted signal, much greater accuracy can be obtained by the use of crystal control of the IF frequency. The static accuracy that can be realized in the usual radar employing two oscillators is dependent on the accuracy of the reference and the amount of zero frequency gain it is practical to employ in the AFC feedback loop.

The dynamic error characteristic must also be considered in choosing the bandwidth. Since the AFC of a pulsed radar set is a sampled-data feedback control device, the error reduction that can be achieved is dependent on the sampling rate. In practice, the effective bandwidth of a continuous proportional error AFC is limited to about one-twentieth the PRF. Tuning errors

caused by variations in frequency greater than this effective bandwidth are not significantly reduced by the AFC, and their effect on the signal energy at the output of the receiver must be considered. The bandwidth of the receiver IF is then selected so as to maximize the output S/N in the presence of the error resulting from those tuning errors which cannot be removed by the AFC.

In some applications, transmitter pulling is often the determining factor in the accuracy of receiver tuning. In these cases it is not essential that the static error be extremely small, and a simplification of the AFC can be realized by the use of limit-activated correction (see Paragraph 8-8).

In tracking applications a further requirement exists that tuning error in the receiver must not produce an error in tracking exceeding a specified amount. Two types of errors arise from the tuning error. The frequency error is converted to amplitude modulation by the IF characteristic. The additional amplitude modulation from this source produces errors in the direction signal of a conical scanning radar. Distortion of the pulse shape also occurs and may produce errors in the measurement of range.

8-7 CONTINUOUS-CORRECTION AFC

Continuous-correction AFC constitutes a type of closed-loop operation in which the error continuously tends to be minimized. The residual error is a function of the loop gain.

A block diagram of a continuous AFC is shown in Fig. 8-9. The input to the AFC is the frequency difference between the transmitting oscillator and

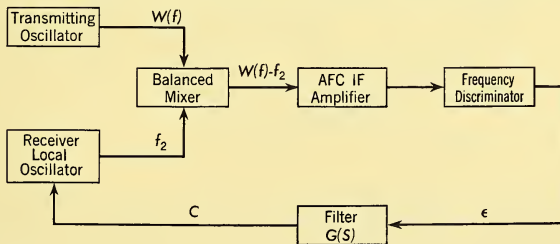


FIG. 8-9 Continuous AFC.

the receiver local oscillator. This frequency is measured by the frequency discriminator which is the error detector. A voltage ϵ proportional to the difference between the input frequency and the crossover frequency of the discriminator is applied to the local oscillator through a filter $G(s)$. The output of the filter is a control C which adjusts the receiver local oscillator to minimize ϵ . C may be a mechanical or an electrical output or a com-

bination of the two, depending on the nature of the control mechanism of the oscillator.

The characteristics of an AFC in stabilizing the receiver tuning when the transmitter frequency or local-oscillator frequency changes can be expressed as

$$\mathcal{L}f_{e1} = \frac{\mathcal{L}f_{d1}}{1 + KG(s)} \quad (8-1)$$

where f_{e1} is the frequency error in the receiver IF frequency

f_{d1} is the frequency error that would result without the AFC

K is the d-c or zero frequency gain of the AFC. ($K = \text{discriminator sensitivity} \times \text{d-c gain of the filter} \times \text{modulation sensitivity of the oscillator}$)

$G(s)$ is the normalized transfer function of the filter.

Fig. 8-10 shows the control characteristics of a typical klystron oscillator. Referring to Table 8-1, the largest tuning error that might exist in a typical

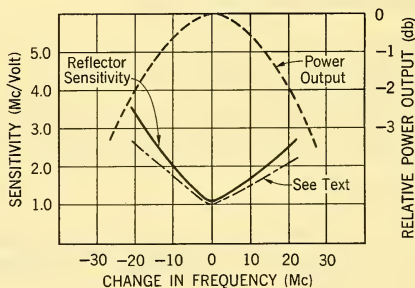


FIG. 8-10 Local Oscillator Characteristics.

system is ± 15 Mc incident to temperature environment, provided that an initial adjustment is made on the AFC whenever a tube is changed. Fig. 8-10 shows that the power output of the local oscillator will vary about 1.5 db and the modulation sensitivity will change by a factor of about 2 (or 6 db) for such a frequency variation. Variations in the power output of the oscillator affect both receiver sensitivity and signal-handling capability. However, the effect of a 1.5-db change is negligible. In cases where such a change cannot be tolerated or in which the frequency variation between the uncontrolled oscillators exceeds ± 15 Mc, an additional feedback loop is sometimes employed. This auxiliary loop measures the oscillator power output and adjusts the klystron cavity for maximum output.

Variation in modulation sensitivity limits the bandwidth that can be employed in the AFC loop. The loop must be designed to have adequate stability with both the highest and lowest gain values. To minimize variations in oscillator sensitivity with tuning, the amplitude of the local oscillator signal is usually made smaller than the amplitude of the sample of the transmitter frequency at the AFC mixer. The IF voltage applied to the discriminator is then proportional to the amplitude of the local oscillator output. The discriminator sensitivity therefore is reduced as the oscillator modulation sensitivity increases. Since loop gain involves the product of the discriminator sensitivity and the oscillator modulation sensitivity, the variation in loop gain that would exist if the IF voltage were maintained constant is reduced. This is shown as an effective modulation sensitivity reduction by the broken curve of Fig. 8-10.

Pull-in and *hold-in* performance of the AFC are determined by superimposing the oscillator control curves on the discriminator characteristic. Fig. 8-11 shows such a curve. *Pull-in* has been defined previously. *Hold-in*

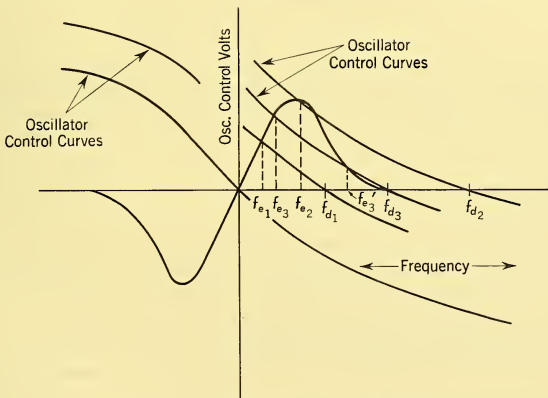


FIG. 8-11 Discriminator Characteristics.

is the maximum frequency interval over which AFC control is effective. Referring to Fig. 8-11, an error f_{e1} in receiver tuning will occur for a frequency deviation f_{d1} of the transmitter frequency. The frequency deviation f_{d2} corresponds to the hold-in range. An error f_{e2} results from an oscillator deviation f_{d2} . The pull-in range corresponds to f_{d3} . The tuning error can be f_{e3} or f_{e3}' ; only the point f_{e3} is stable, however. For all deviations less than f_{d3} the tuning error is a stable condition corresponding to the intersection of the discriminator curve and the oscillator control curve.

The discriminator curve of Fig. 8-11 is the characteristic appearing at the output of $G(s)$ in Fig. 8-9. For high-performance systems $G(s)$ is designed with a large zero frequency or d-c gain. Fig. 8-12 shows the resultant discrim-

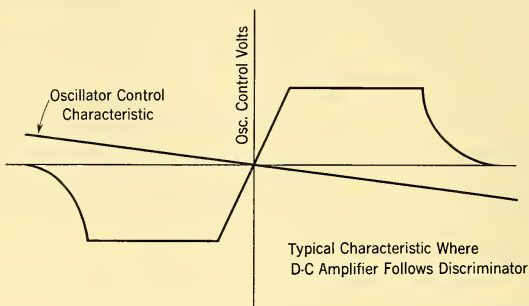


FIG. 8-12 Discriminator Characteristics.

inator curve measured at the oscillator. By increasing the d-c gain for a given IF characteristic the pull-in range is increased. In a number of systems, however, a limited d-c gain follows the IF detectors. To realize a large pull-in range a frequency search sweep is applied to the oscillator. The presence of an IF signal in the AFCIF is employed to remove the sweep. A relatively narrow-band IF discriminator can exhibit a large pull-in range by this technique.

In a typical system the control required on the oscillator may be ± 50 volts, but this voltage is usually at some bias level, e.g. -150 volts. To obtain maximum performance from a given loop, a d-c voltage is added to the output of the IF discriminator so that the control voltage is at -150 considerable energy in modulation sidebands at the IF frequency.

The AFC mixer of a pulse radar set is operated as a balanced mixer to minimize frequency tuning error caused by discriminator outputs resulting from the modulation spectrum of the transmitted signal. With the usual IF frequencies employed, narrow pulse-length transmitted signals have considerable energy in modulation sidebands at the IF frequency.

A typical IF discriminator design can provide an output of 2 to 3 volts per megacycle with a peak-to-peak separation of the discriminator of 4 to 5 Mc. The output from the discriminator is in the form of video pulses. If these pulses are fed directly into the filter the zero frequency gain required from the filter is

$$K_2 = \frac{E_o T}{K f_c \tau} \quad (8-2)$$

where E_o is the maximum output voltage required

T is the interpulse period

f_e is the static error due to finite gain of the loop

τ is the pulse length

K is the discriminator gain.

For a static error of 50 kc, a pulse length of 1.0 μsec , and an interpulse period of 1000 μsec a typical gain required is 250,000. The required gain can be reduced by the use of a pulse lengthener following the frequency discriminator. With a circuit like that described in Paragraph 7-17 the filter gain is reduced to about 250 if $T = T_p$ and $\alpha = 0$. A gain of 250 can be provided with an operational amplifier. This d-c gain can be reduced if the discriminator is designed with higher output or if a video amplifier is employed between the pulse lengthener and the discriminator. Although the loop can provide a static accuracy of 50 kc, the tuning accuracy also includes the inherent accuracy of the reference. With thermionic diodes in the discriminator circuit and with stable capacitances and inductors a reference accuracy on the order of 50 kc can be achieved for the assumed discriminator characteristic. The static accuracy is then expected to be 70 kc, provided that means for adjusting the reference initially to the IF frequency of the receiver exist.

The dynamic accuracy will depend on hold-in and pull-in requirements. A single lag network is usually employed; i.e., the operational amplifier is made an integrator. An error only appears as a signal sampled at the pulse repetition frequency; therefore the integrator time constant that can be employed depends on the allowable overshoot. At a given time t_1 the error in frequency might be f_1 . This error is applied as a voltage to the integrator. The integrator output will change at a rate determined by the RC and the input voltage corresponding to f_1 .

An overshoot of about 50 per cent of the initial error is a reasonable compromise to obtain good dynamic response. The introduction of a step frequency error then results in an output frequency which is 50 per cent of the initial error but of opposite sign. The output of the AFC thus oscillates about the desired frequency with diminishing error. Inputs incident to pulling of the transmitter may occur at the lobe frequency or multiples thereof. The error reduction that can be accomplished by the AFC thus depends on the ratio of the PRF to the pulling frequency. In typical cases there might be 10 to 20 samples during a cycle of the pulling frequency, and error reductions on the order of 10 to 1 are attained when 50 per cent overshoot is allowed.

8-8 LIMIT-ACTIVATED AFC

Limit-activated AFC constitutes a type which has a switching action whenever the error exceeds a predetermined value. The static error in such a system is never zero, but oscillates about the correct value with a constant peak error. The static error is determined by the limit level for the error, the interpulse period, and the integrator characteristic.

Essentially the system is the same as that shown in Fig. 8-9 with $G(s) = K/s$. The difference between this AFC and the continuous-correction AFC is that the input to $G(s)$ does not produce a change in polarity of the oscillator control signal until the frequency error exceeds a certain limit L . The output from the frequency discriminator is applied to an integrator whose output controls the oscillator. A frequency error f_{e1} results in a change of the oscillator frequency at a rate Kf_{e1} . An error correction signal can only be obtained in a shortest time T after a signal occurs which causes the oscillator frequency to change. The quantity T is the interpulse period. The output frequency therefore oscillates at a frequency which is some fractional integer of the pulse repetition frequency. Fig. 8-13 shows the

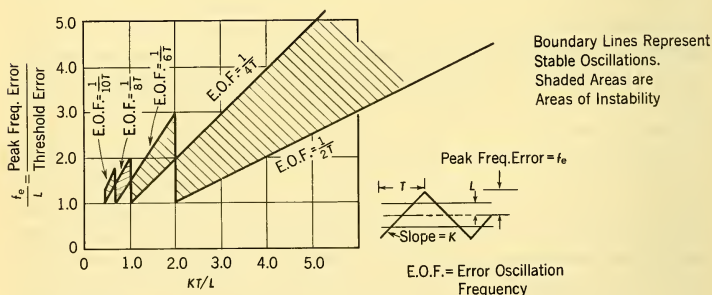


FIG. 8-13 Characteristics of a Switched (Limit-Activated) Type Control Loop for Constant Input Conditions.

relationship between the parameters K , T , and L . It will be observed that multiple modes of oscillation can occur. These result from the fact that the data sample may occur when the frequency error is less than L . The limit level in a practical system is about 50 kc minimum.

With the continuous AFC employing a 100 per cent pulse lengthener the static accuracy is independent of the PRF, provided that the response time of the discriminator is the same for all pulse lengths. (Normally a pulse-length change would be associated with a PRF change.) In the limit-activated correction type of AFC, the static accuracy increases as the PRF increases.

The dynamic performance of this type of AFC is determined by K and the peak rate of frequency change occurring in the input signal. When the peak rate of the input frequency change is less than K , an error reduction is obtained and the peak errors are of the same order of magnitude as the static error. When the peak rate of the input exceeds K , the error is not reduced. When the input to the AFC is predominantly a single frequency, the required value of K may be determined by

$$K = \omega f \quad (8-3)$$

where ω is the angular frequency at which the signal frequency is changing, and f is the peak deviation of the signal frequency.

8-9 THE INFLUENCES OF LOCAL OSCILLATOR CHARACTERISTICS

A typical electronic tuning characteristic of the local oscillator is shown in Fig. 8-10. Note that both the power output and the frequency-modulation sensitivity vary with the tuning. A variation in modulation sensitivity means that there will be a variation in the loop gain of the AFC. In the case of a continuous AFC the change in power can be employed to compensate (to some extent) for the variation in modulation sensitivity. However, perfect compensation is not attained. As a result, if a continuous AFC is designed with only electronic tuning capability and it must accommodate dynamic inputs, there is a degradation of the static and dynamic error characteristics at the extremes of the tuning range if the overshoot is selected to be 50 per cent at the middle of the tuning range. As shown in Fig. 8-10, making the local oscillator signal smaller than the transmitter signal reduces the variation in modulation sensitivity, but the variation is still more than 2 to 1 over the tuning range. The static error therefore oscillates when the receiver is tuned at the extremes of its tuning range. A smaller overshoot and somewhat poorer dynamic performance must be accepted if the tuning range is required to exceed 30 Mc with a fixed static tuning accuracy.

In the limit-activated correction AFC there is no benefit from making the local oscillator signal smaller than the transmitter signal since the rate of correction is independent of the magnitude of the error. (A constant rate of control occurs whenever the error is greater than L .) The static error oscillates at all times and will vary over the tuning range of the receiver in accordance with Fig. 8-13 when the change in modulation sensitivity is inserted into the value for K .

When the system is subjected to severe dynamic error requirements, a double loop is sometimes employed using a low-frequency feedback to the resonator of the klystron and a high-frequency feedback to the reflector.

A smaller variation in electronic modulation sensitivity is thereby obtained, allowing a design of 50 per cent or greater overshoot in the step response of the electronic tuning loop and also providing a much greater receiver tuning range than can be obtained with electronic tuning alone.

8-10 RELATION TO RECEIVER IF CHARACTERISTICS

The dynamic error in the receiver tuning and the received signal IF characteristic are related. The signal at the output of the IF amplifier will contain pulse amplitude modulation arising from the tuning error of the receiver. When the IF response is perfectly symmetrical about the center frequency and the static error is negligibly small, the modulation at the output will be double the frequency of the frequency modulation of the tuning error.

In conical scanning radars the dominant output incident to such effects is thus at two or four times the lobing frequency. The response of the angle demodulators to these frequencies is greatly attenuated by the use of balanced ring demodulators, and additional noise on the direction signal caused by AFC characteristics is then negligible provided that saturation is not present. If the modulation is large there is of course a loss in signal-to-noise ratio which can be determined from the rms error and Fig. 8-8. To minimize the conversion of the tuning error to amplitude modulation the nose of the receiver selectivity is made as flat as possible consistent with the considerations discussed in Chapter 7.

8-11 DISCRIMINATOR DESIGN

There are two types of discriminators employed in pulse AFC — the phase discriminator and the stagger-tuned discriminator. The choice between the two depends on the details of the control circuitry. A slightly higher effective transfer impedance can be realized with the stagger-tuned circuit, but symmetry is difficult to maintain. If a video amplifier is employed after the discriminator but prior to the integration, then the phase discriminator is the more attractive choice. Fig. 8-14 shows a typical phase discriminator circuit, and the form of the transfer impedance.

In designing the discriminator the network elements can be selected so that $s_1 = s_5$, $s_2 = s_7$, $s_3 = s_6$, and $s_4 = s_7$. The discriminator response is then of the same form as the difference in the envelope response of two stagger-tuned one-pole networks. To obtain the maximum sensitivity from the discriminator the poles are located so that the two equivalent response curves cross at their point of inflection. H is inversely proportional to C_1 and C_2 and these quantities are minimized to obtain maximum Z_t .

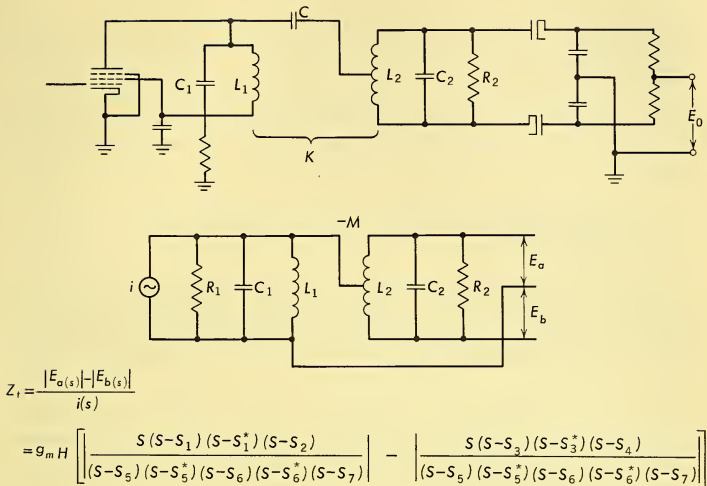


FIG. 8-14 Frequency Discriminator (Phase Type).

8-12 INSTANTANEOUS AFC

When the dynamic inputs are so severe that continuous AFC cannot satisfactorily reduce the tuning error, the possibility exists of using an Instantaneous AFC (IAFC). IAFC is a type in which the error correction is completed before the pulse has ended. Extremely wide bandwidths are required in the AFC-IF amplifier and discriminator in order that negligible time delay may be obtained in these elements. The AFC must include a bidirectional pulse lengthener which is required to have a negligible decrement; the output of the lengthener is the controlled value during the pulse. This type of AFC can potentially provide the best dynamic performance in a sampling-type AFC; however, there are some practical limitations. The discriminator measures the instantaneous frequency and there is negligible lag in the loop, so that if the instantaneous frequency is constant during the pulse the tuning error can be reduced to a value dependent on the gain of the loop. In most cases there are, however, intrapulse frequency changes. In the continuous AFC the average frequency is controlled; in the IAFC the controlled frequency depends on the characteristics of the discriminator and pulse lengthener. The controlled frequency is different from the average when intrapulse frequency variations are large. As a result the static error of an IAFC can be larger than that of a continuous AFC. The wide bandwidth required in the discriminator limits its output

so that the reference accuracy is poorer than in a continuous AFC. Another IAFC problem is associated with holding the voltage precisely during the interpulse period.

The IAFC circuit is not commonly employed because of its limitations and because the dynamic inputs can usually be reduced by proper design of the transmitter and associated circuits so that the continuous AFC is adequate.

8-13 PROBLEMS OF FREQUENCY SEARCH AND ACQUISITION

Table 8-1 indicates that in a typical case the transmitter frequency can vary over a greater range than can the electronic tuning of a klystron as shown by Fig. 8-10. It has been noted that it is also not always feasible to utilize a discriminator-IF characteristic which will provide such a wide pull-in range. To cope with this situation it is necessary that a mode of operation be provided which will allow the AFC to search for the transmitter frequency, acquire it, and track it. With magnetrons which are tunable it is necessary that two loops be provided for the AFC. A slow-response loop which controls the cavity resonator by thermal or mechanical means is sometimes employed. With fixed-frequency magnetrons, however, the electronic tuning range is usually adequate. In these cases, periodic adjustments can be made to the klystron resonator to accommodate aging or replacement of magnetrons. The slow frequency variations that are then encountered are usually well within the electronic tuning capability of the local oscillator. It is sometimes more economical to provide frequency capture within the electronic tuning range by means of a search sweep of the local oscillator than to provide an IF discriminator which has a pull-in range equal to the maximum frequency difference between the transmitter and local oscillator at the time that the radar set may be energized.

A typical frequency range over which the local oscillator must search for the transmitter frequency is 40 Mc. The speed at which this search can occur depends on the bandwidth of the IF discriminator, the interpulse period, and the total search range. A typical discriminator might have a pull-in range of 10 Mc and 10 pulses required for acquisition. The maximum search speed is then equal to $1/T$ Mc, where T is the interpulse period. Circuits are provided so that the search sweep signal is automatically removed when the transmitter frequency has been captured.

8-14 AUTOMATIC GAIN CONTROL

Radar targets act to modulate the amplitude of the reflected signals in several ways. First of all, range variations can produce variations in the received power of more than 100 db. Secondly, amplitude fluctuations

caused by target motion which were discussed in Paragraph 8-2 can also be large. These variations in signal strength can seriously interfere with tracking of the target unless steps are taken to protect the receiving system from their effects. This is particularly important in angle tracking where the signal amplitudes in two offset antenna lobes are compared (either sequentially or simultaneously) to generate an angle error. The angle error will normally be directly proportional to differences in the received amplitudes in the two lobes, so that signal strength variations common to the two lobes must be removed if a usable error signal is to be obtained.

Regulation of the received signal level is normally accomplished by an *automatic gain control (AGC) circuit*. This is a feedback loop which adjusts the receiver gain to maintain the average receiver output at a constant reference value. A simplified block diagram of an AGC loop is shown in Fig. 8-15. In operation, the gain of the IF amplifier presented to its input

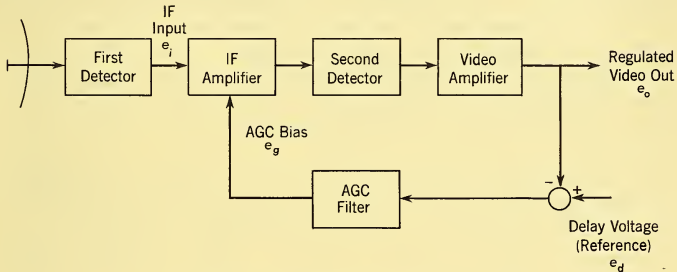


FIG. 8-15 Radar Receiver with AGC Loop.

voltage e_i is automatically adjusted by an AGC bias e_g . This bias is developed as the difference of the video output voltage e_o and a reference voltage e_d referred to as the *AGC delay*. The system basically acts to maintain the output equal to the delay. The degree to which this is done when the input fluctuates is determined by the AGC filter in the feedback path.

An AGC loop is a nonlinear servo in that the feedback signal e_o is not linearly combined with the input, but acts to modify the gain with which the input is amplified. In a sense, an AGC loop in combination with an angle tracking loop is an example of an adaptive servo system in that the system gain is automatically adjusted to compensate for externally generated variations in the received signal strength. Although basically nonlinear, an approximate linear analysis of the system operation for small deviations of the input from an average operating level is very useful and forms the basis for system design.

In designing an AGC loop, particular attention should be paid to the following four areas of performance which are of primary importance:

1. The steady-state or d-c regulation of the video voltage
2. Attenuation of amplitude fluctuations
3. Fidelity with which intelligence is transmitted
4. AGC loop stability

The steady-state regulation determines the degree to which the AGC loop compensates for slow variations in the average signal level caused by changes in target aspect and range. As noted above, such variations can be as great as 100 db. The AGC loop is often required to reduce slow variations of the output level to only a few decibels. For instance, in an angle tracking loop, the loop gain is proportional to the output signal level so that variations in this output produce corresponding variations in loop gain. When the average output varies by 2 to 1, or 6 db, the angle track loop gain will also vary by the same factor, and this may have a serious effect on the overall angle track loop stability and performance.

Fluctuations in the strength of radar echoes reflected from aircraft targets have been discussed in Paragraph 4-8, and typical spectra of this amplitude noise for two types of aircraft are shown in Figs. 4-23 and 4-24. In Fig. 4-23 the amplitude noise spectrum for a propeller-driven aircraft illuminated by X-band radiation is shown. Very predominant propeller modulation peaks at harmonics of about 60 cps persist to over 300 cps. In Fig. 4-24, the amplitude noise spectra generated by a B-45 jet bomber illuminated by several wavelengths are shown. With no propeller modulation, the spectra all fall off within a few cps. In general, it is desirable for the AGC loop to remove amplitude noise whose frequency components fall within the pass band of the angle tracking loop. Otherwise, modulation of an angular lag error by amplitude fluctuations in the receiver output can produce excessive angle noise.

Besides removing noise modulation from the receiver signal, the AGC loop must also transmit intelligence modulation without appreciable distortion. This is a critical problem in systems which employ sequential or conical scan lobing to generate angle error signals. For instance, in a conically scanned system, the angle error is contained in the amplitude and phase of a sinusoidal error signal at the scan frequency which may vary because of poor scan rate generator regulation. Generally, the AGC loop must be able to transmit this signal with negligible phase shift or change in amplitude.

Since the AGC circuit is a feedback loop, stability questions are important and servomechanism design techniques are applicable. These techniques are applied to a linear small-signal approximation to the nonlinear loop which will be derived in the following paragraph. Adequate

stability and dynamic response of the AGC loop is often difficult to achieve in combination with other requirements on d-c regulation (proportional to AGC loop gain) and fidelity of intelligence modulation. For pulsed radars where fast AGC action is desired (common in monopulse systems), methods for analyzing sampled-data servos must be used and the AGC loop bandwidth is limited to about half the repetition rate by stability considerations.

8-15 LINEAR ANALYSIS OF AGC LOOPS

Design of AGC loops is based upon a first order or linear approximation to the nonlinear action of the IF amplifier for small deviations from average operating points.⁹ This approximation is illustrated in Fig. 8-16a. The

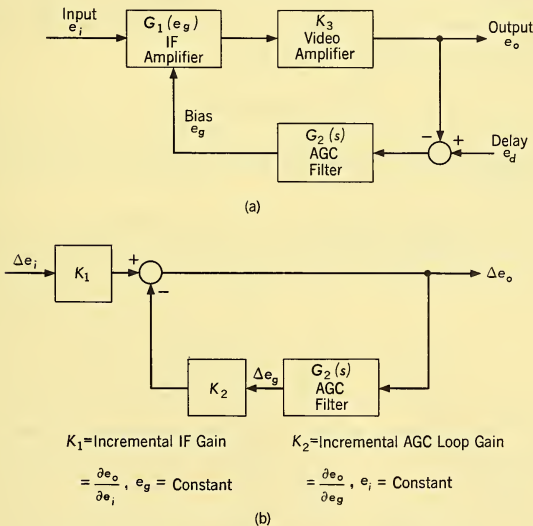


FIG. 8-16 Linear Approximation to AGC Loop.

upper block diagram in this figure shows the essential components of an AGC loop. The nonlinear relation of the IF amplifier gain to the AGC bias is indicated by $G_1(e_g)$. The system equations have the following forms:

$$e_o = e_i K_3 G_1(e_g) \tag{8-4}$$

$$e_g = (e_d - e_o) G_2(s). \tag{8-5}$$

⁹B. M. Oliver, "Automatic Volume Control as a Feedback Problem," *Proc. IRE*, 36, 466-473 (1948).

Small deviations in the output can be related to small deviations in e_i and e_o in the following manner:

$$\begin{aligned}\Delta e_o &\approx \frac{\partial e_o}{\partial e_i} \Delta e_i + \frac{\partial e_o}{\partial e_o} \Delta e_o \\ &\approx K_1 \Delta e_i + K_2 \Delta e_o.\end{aligned}\quad (8-6)$$

The gain factors K_1 and K_2 represent the incremental gain of the IF amplifier to the input and the incremental AGC loop gain, respectively. Deviations of the bias will be simply related to output deviations through the AGC filter:

$$\Delta e_o = -G_2(s) \Delta e_o. \quad (8-7)$$

The approximate linear feedback loop corresponding to Equations 8-6 and 8-7 is pictured in Fig. 8-16b. The output-input ratio for this linear regulating loop will have the following form:

$$\frac{\Delta e_o}{K_1 \Delta e_i} = \frac{1}{1 + K_2 G_2(s)}. \quad (8-8)$$

8-16 STATIC REGULATION REQUIREMENTS OF AGC LOOPS

The transfer function represented by Equation 8-8 gives the small-signal modulation transmission characteristics of the loop. If the zero frequency gain of the AGC filter is assumed unity [$G_2(0) = 1$], the static gain around the loop is K_2 as is indicated in Fig. 8-16b. The static regulation performance is directly related to the loop gain K_2 . In order to display this relation, though, the gain control characteristic of the IF amplifier indicated in Equation 8-4 must be examined in detail. Typically, the logarithm of the IF amplifier gain is approximately a linear function of the AGC bias voltage. That is, the slope of the gain-bias curve in decibels per volt is a constant. Fig. 8-17 shows a typical IF amplifier gain control characteristic. Such a linear relation can be generally expressed in the following form:

$$20 \log_{10} G_1 = A + B e_o. \quad (8-9)$$

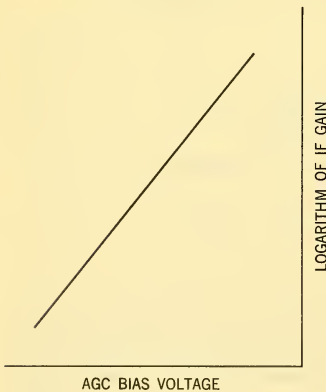


FIG. 8-17 Typical IF Amplifier Gain Control Characteristics.

The constant B in this equation gives the slope of the gain control characteristic in decibels per AGC bias volt. Differentiating with respect to e_g :

$$(20 \log_{10} e) \frac{1}{G_1} \frac{\partial G_1}{\partial e_g} = B \quad (8-10)$$

$$\frac{1}{G_1} \frac{\partial G_1}{\partial e_g} = 0.115B.$$

Multiplying and dividing the LHS by $e_i K_3$ allows us to express it simply as the ratio of the loop gain K_2 and the video output e_o by utilizing Equation 8-4.

$$\frac{1}{e_i K_3 G_1} \times e_i K_3 \frac{\partial G_1}{\partial e_g} = 0.115B \quad (8-11)$$

$$\frac{1}{e_o} \times K_2 = 0.115B.$$

For the d-c or static case with $G_2(0)$ equal to unity, changes in the bias are directly proportional to changes in the output. Thus the slope B can be expressed as the ratio of the total change in gain to the change in output voltage:

$$B = \frac{\text{gain change (db)}}{e_{g,\max} - e_{g,\min}} = \frac{\text{gain change (db)}}{e_{o,\max} - e_{o,\min}} \quad (8-12)$$

Substituting Equation 8-12 into Equation 8-11 yields the following expression for the AGC loop gain:

$$K_2 = \frac{0.115 e_o}{e_{o,\max} - e_{o,\min}} \times [\text{gain change (db)}] \quad (8-13)$$

It is apparent from this expression that with the linear gain control characteristic shown in Fig. 8-17, the loop gain will vary somewhat with the video output e_o . Normally, the video output will be well enough controlled that its variation can be neglected and an average value used in Equation 8-13. It is possible to compensate for this variation in the loop gain K_2 by introducing a slight curvature in the gain control characteristic. Generally, though, uncontrolled departures from linearity with accompanying uncontrolled variations in the loop gain are a much more important design factor to consider.

To illustrate the use of Equation 8-13, suppose that static input variations of 100 db must be regulated by the AGC loop to output variations of only ± 1 db or between $0.89\bar{e}_o$ and $1.122\bar{e}_o$. Substituting these numbers into Equation 8-13 yields

$$K_2 = \frac{(0.115)(100)}{1.122 - 0.89} = 49.5 = 33.4 \text{ db.} \quad (8-14)$$

Thus, with an AGC loop gain of about 50 or 34 db, input variations of 100 db can be reduced to output variations of only ± 1 db.

8-17 DYNAMIC REGULATION REQUIREMENTS OF AGC LOOPS

It was previously noted that amplitude noise fluctuations in the receiver output will modulate steady-state lag errors in the angle tracking loop output and can thus produce excessive angle tracking noise. For this reason and also to minimize the possibility of saturation, the AGC loop should be designed to remove most of the input amplitude fluctuations, particularly those within the pass band of the angle tracking loop. Actually, if there were no systematic errors, some slight improvement in the glint noise or deviations in angle of arrival could be achieved with no AGC. The reason for this is that there is a correlation between large deviations of the apparent center of reflection of an aircraft target and deep amplitude fades, since both effects are produced by destructive interference of the reflected signals. An effective AGC will increase the receiver gain to compensate for fades and thus increase the magnitude of the glint deviations. In a practical case, this effect is more than balanced by the benefits of removing spurious modulation from the error signal.

Typical results from a simulator study of this problem are shown in Fig. 8-18.¹⁰ In this case, the target noise spectrum (amplitude and angle) had a width of 1 cps while the tracking servo had a similar bandwidth. The

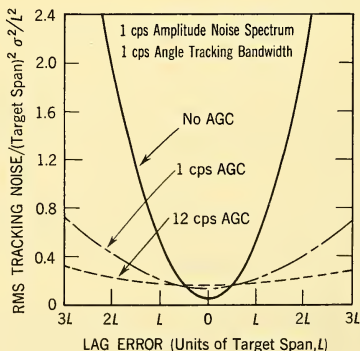


FIG. 8-18 Effect of AGC on Angle Tracking Noise as a Function of Servo Lag Error.

¹⁰J. H. Dunn and D. D. Howard, "The Effects of Automatic Gain Control Performance on the Tracking Accuracy of Monopulse Radar Systems," *Proc. IRE* **47**, 430-435 (1959).

mean square tracking error is plotted versus the lag error for no AGC, a 1-cps AGC, and a relatively fast AGC of 12 cps. With small lag errors, less noise is produced with no AGC because of the correlation between amplitude and glint noise noted above. For typical tracking conditions, though, in which appreciable lag errors exist, the tracking noise and no or slow AGC greatly exceeds that associated with fast-AGC designs.

Other factors to be considered in the design of the low-frequency response of the AGC loop are the transient recovery time of the receiver from deep fades (as great as 60 db) which should be such that the angle error is not blanked for longer than the angle tracking loop response time. This can be achieved by providing a high enough velocity constant for the AGC loop and allowing sufficient dynamic range in the output.

The amplitude noise spectrum from most aircraft targets falls off with frequency approximately as if it were filtered by a single section, low pass, RC filter (see Paragraph 4-8). In order that no particular noise frequencies be emphasized in the output, it is desirable, although not absolutely necessary, that the AGC filter match this spectrum; that is, it should fall off with frequency with a -1 slope in the frequency region covering the angle tracking pass band.

The significant factor in determining the quantitative effects of the AGC on received modulations is the transfer characteristic

$$\text{AGC transfer characteristic} = \frac{1}{1 + K_2 G_2(s)}. \quad (8-15)$$

The required behavior of this function and the open-loop characteristic $K_2 G_2(s)$ will be examined in more detail.

8-18 AGC TRANSFER CHARACTERISTIC DESIGN CONSIDERATIONS

From the discussion in Paragraphs 8-14 through 8-17 of factors significant to the design of a radar receiver automatic gain control, several basic specifications emerge as AGC transfer function desiderata in sequential lobe comparison radars, namely:

1. High gain at low frequencies to provide adequate static regulation. Some system specifications contemplate allowing only ± 1 db variation in the modulated envelope output for a range of input signal levels of approximately 100 db.

2. An initial transfer function slope of zero from dc to as high a frequency (approaching the angle tracking bandwidth) as possible.

3. Gain drop-off with a -1 slope on a db-versus-log frequency plot to ensure, in view of the established nature of radar noise, that the receiver output shows no noise emphasis at any particular frequency.

4. Adequate AGC bandwidth (fast AGC) to ensure isolation from amplitude fluctuation in the received echo and minimization of closed-loop angle tracking noise in the practical employment of the radar system (particularly including recognition of situations where angle tracking errors will exist). Estimates of the half-power frequency of amplitude noise for some target tracking problems are as high as $\omega = 10$ (a gain of +20 db or $|K_2G_2| = 10$ at $\omega = 20$ may be taken as a practical design objective).

5. Adequate AGC bandwidth to ensure suitable transient response. This requirement is another aspect of system demands consistent with item (4) above.

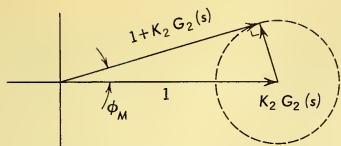
6. The phase of the quantity $[1/(1 + K_2G_2(s))]$ should not vary excessively over the range of angle tracking modulation frequencies surrounding the scan frequency to limit crosstalk effects between the azimuth and elevation angle tracking axes. Phase shifts of up to 5° or 10° can be allowed in most systems although some applications may require this phase shift to be maintained less than a few degrees. As will be shown in Paragraph 9-9, with phase shifts of more than 10° , the antenna has a tendency to spiral or circle, and with even larger shifts, it will become unstable. This phase shift must often be maintained in the face of uncontrolled variations in the incremental loop gain K_2 of as much as 10 db and uncontrolled variations in scanning frequency of up to ± 5 per cent.

7. In order to ensure adequate loop stability and transient response, a minimum gain margin of 6 db should be maintained for all possible variations of the incremental loop gain. Similarly, a phase margin of from 40° to 50° should be maintained.

8. In pulsed systems, it is necessary to provide a minimum gain margin of at least 6 db at one-half the repetition rate to ensure stable operation. This is particularly important in monopulse systems where rapid AGC action is desired.

8-19 THE MODULATION TRANSMISSION REQUIREMENT

When large fluctuations in the AGC loop gain are possible or very small scanning frequency phase shifts are required, special design techniques must be used to maintain the phase of the intelligence signal being transmitted through the system. Two such techniques are available. In the first, additional high-frequency lag and lead terms are incorporated into the AGC filter to provide an open-loop phase shift of 180° at and near the lobing frequency. The phase shift of the closed loop is then zero at the lobing frequency and insensitive to variations in loop gain and lobing rate. A second approach is to attenuate $K_2G_2(s)$ with a null over the required frequency band so that the maximum closed-loop phase ϕ_m will be limited to a small value regardless of the phase of $K_2G_2(s)$.

FIG. 8-19 The Vector $[1 + K_2G_2(j\omega)]$.

Using the latter approach and referring to Fig. 8-19,

$$\tan \phi_m = \frac{K_2G_2}{1 + K_2G_2} \approx K_2G_2. \quad (8-16)$$

If, for example, it is required that ϕ_m be maintained less than 1.5° , then the gain at the lobing frequency

must comply with

$$K_2G_2 < \tan 1.5^\circ = 0.0262 = -31.9 \text{ db.} \quad (8-17)$$

If it is assumed that the lobing frequency is 50 cps, that its regulation is ± 5 per cent or ± 2.5 cps, and that the angle tracking loop bandwidth is 1 cps, the open-loop attenuation should be at least 31.9 db between 46.5 cps = 292 rad/sec and 53.5 cps = 336 rad/sec in order to maintain the phase shift less than 1.5° over the anticipated range of operating conditions.

8-20 DESIGN OF AN AGC TRANSFER FUNCTION

As a trial design, a single time constant RC filter is selected for the AGC filter. This will provide a -1 slope, which was previously noted as desirable. The maximum gain is chosen on the basis of the static regulation requirement. In the example of Paragraph 8-16, a static gain of 34 db was required to regulate input variations of 100 db to ± 1 db in the output. This requirement is adopted as the loop gain in this example.

In Paragraph 8-18, it was mentioned that a practical AGC loop design for a system tracking aircraft targets should have a loop gain of 20 db

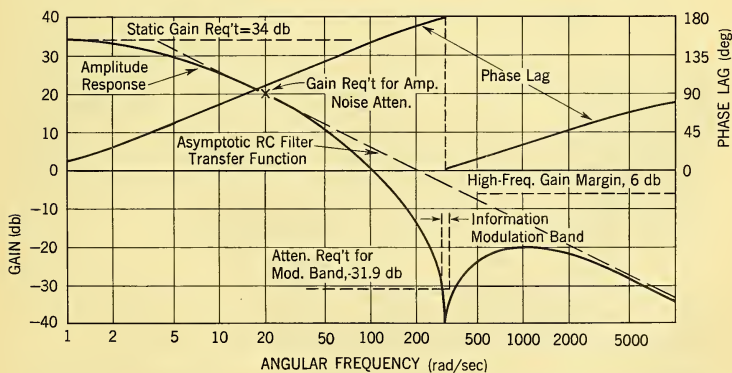


FIG. 8-20 Trial Design of AGC Open-Loop Transfer Function.

at $\omega = 20$ in order to provide sufficient attenuation of amplitude noise fluctuations. This requirement, in connection with the static gain requirement, fixes the location of the low-frequency corner of the AGC filter at 4 rad/sec. The AGC loop transfer characteristic thus developed is shown in Fig. 8-20.

In order to maintain the fidelity of transmitted modulation, a network will be introduced into the transfer function to provide a null at the lobing frequency as described in Paragraph 8-19. It will be supposed that the phase shift must be kept less than 1.5° and the lobing frequency, its regulation, and the angle tracking bandwidth have the values assumed in the example in Paragraph 8-18.

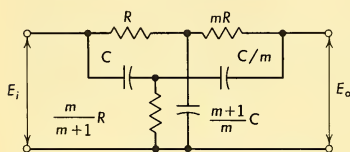


FIG. 8-21 Parallel-T Null Network with Symmetry Pattern m .

Fig. 8-21 shows a parallel-T network which can be used to provide the required null.¹¹ The transfer function of this network is

Voltage transfer function of network

$$= \frac{u^2 + 1}{u^2 + u \left(2 + \frac{2}{m} \right) + 1} \quad (8-18)$$

where $u = j\omega RC = j\omega/\omega_c$

ω = angular frequency, rad/sec

ω_c = null location, rad/sec

m = symmetry parameter determining null sharpness.

The effect of this network with a null of 50 cps = 314 rad/sec and a value of $m = 0.54$ is shown in Fig. 8-20 in combination with the single time constant RC filter. The value of m (Fig. 8-21) was selected to provide a total attenuation of 31.9 db (as required in Paragraph 8-19) at 292 rad/sec, the lowest possible modulation frequency. The high-frequency gain margin should be checked, particularly at one-half the PRF. An inspection of Fig. 8-20 shows that a minimum gain margin of 20 db, in comparison with the 6-db requirement, is maintained at all high frequencies.

The closed-loop response of the AGC loop indicates most directly the dynamic AGC action in attenuating low-frequency amplitude noise and transmitting modulation frequencies. Fig. 8-22 shows the closed-loop response corresponding to the trial design illustrated in Fig. 8-20.

¹¹C. F. White, *Transfer Characteristics of a Bridged Parallel-T Network*, NRI Report R-3167, 27 August 1947.

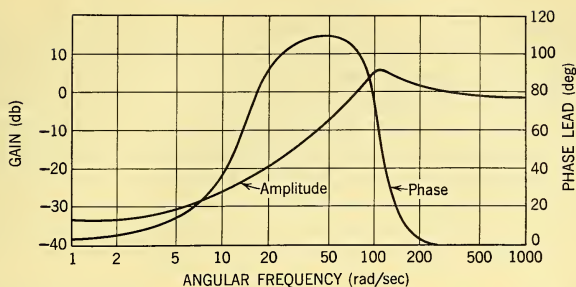


FIG. 8-22 Closed-Loop Response of AGC, Trial Design.

8-21 THE IF AMPLIFIER CONTROL CHARACTERISTIC

The AGC loop must maintain the receiver output constant so that the loop gain of the various tracking loops is negligibly affected by the large variations in input signal power that are encountered. The analysis of Paragraph 8-18 is based on maintaining the average value of the receiver output constant. In receivers which must recover modulation from a PAM signal to obtain an error signal, additional consideration must be given to the distortion of the IF signal as it passes through the amplifier. Ideally it is desirable that gain control be applied to IF tubes which exhibit square-law transfer characteristics. Under these conditions no distortion of the IF signal will be apparent at the demodulated output. Usually the radar receiver must incorporate tubes having good gain bandwidth products. Such tubes are invariably of the sharp cutoff variety and are likely to produce distortion of the IF signal with accompanying excessive variations in the AGC loop gain if proper precautions are not taken.

To determine the actual distortion through an IF stage, an accurate description of the transfer characteristic is required. In general any characteristic may be expressed as a power series in e_g , the grid-cathode voltage. If the signal input to the IF stage is a modulated signal $e_i = A \sin \omega_c t (1 + m \cos \omega_m t)$ and if an AGC voltage E_1 is applied to the number one grid of the tube, then a convenient measure of the distortion is the change in the effective modulation of the signal at the output of the stage. Thus

$$\frac{m'}{m} = 1 + \frac{\frac{3}{2}a_3A^2(1 - \frac{3}{8}m^2)}{a_1 + E_1(2a_2 + 3a_3E_1) + \frac{3}{4}a_3A^2(1 + \frac{3}{2}m^2)} \quad (8-19)$$

where m' is the fractional modulation at the output and a_1 , a_2 , and a_3 are the first three coefficients of the power series expansion for the transfer characteristic. The distortion of the signal is seen to be a function of the

third derivative of the transconductance. In designing an amplifier it is not convenient to obtain an infinite series expressing the transfer characteristic. The published transconductance curves can, however, be employed to obtain a reasonable estimate of the distortion. The transconductance corresponding to an infinitesimal signal applied to the tube at a grid voltage E_1 is determined, and then the transconductance at the positive and negative peak values of the IF signal is determined. From these three values the distortion is computed from

$$\frac{m'}{m} = \frac{1}{2} \left[1 + \frac{g_{m,max} + g_{m,min}}{2g_{m0}} \right] \quad (8-20)$$

When A and E_1 are so large that the negative excursion of the signal extends beyond cutoff, Equation 8-20 is not sufficiently accurate. However, it serves as an estimate of the distortion provided that the tube is cut off. Sharp cutoff tubes do not always cut off at the voltages indicated by the tube characteristics. The tubes are only required to exhibit less than a specified maximum value of plate current at cutoff bias. As a result of inadequately controlled cutoff characteristics $g_{m,min}$ does not go to zero for large E_1 and A , and the distortion in an actual amplifier is sometimes observed to be much greater than estimated by Equation 8-20. In the design of a gain-controlled amplifier employing sharp cutoff tubes the AGC voltage applied to the stages is therefore restricted so that the peak negative voltage $E + A$ does not exceed cutoff. The number of stages in the IF amplifier must then greatly exceed the minimum number determined by gain, bandwidth, and stability requirements.

A more suitable arrangement in the radar receiver involves the use of two or three remote cutoff pentodes in the early stages of the amplifier. Gain reductions of 35 per stage with negligible distortion can be obtained with some of the available semiremote cutoff tubes having reasonably good gain bandwidth products. Very little or no AGC is then applied to the remaining stages of the IF amplifier.

It is desirable to limit AGC loop gain variations with the input signal level in order to maintain loop stability and the required dynamic performance. AGC loop gain variations of 2 : 1 or 6 db represent a practical goal employed in the design of the IF amplifier.

As was noted in Paragraph 8-16, the AGC loop gain is proportional to the derivative of the logarithm of the IF amplifier gain with respect to the AGC bias. The contribution of an individual stage is thus proportional to the derivative of the logarithm of the transconductance curve. Unfortunately, this quantity, like the third derivative of the transconductance whose importance was noted above, is not normally specified or controlled in tube manufacture, and large variations can occur in the cutoff region. The use of a greater number of remote or semiremote cutoff tubes, limiting

the controlled gain variation to the order of 20 db per stage, will minimize these gain variations. When a gain control range in excess of 50 or 60 db must be provided, it is frequently necessary to control the first stages of the radar receiver. When this is done, control voltage must be provided in a manner that causes the least deterioration of signal-to-noise ratio.

In addition to grid-1 control of the amplifier stages, grid-3 or plate and screen control is sometimes employed. Grid-3 control provides minimum third-order distortion, but the screen dissipation is generally excessive when the tubes are operated with reasonable gain bandwidth factor. As a result the best compromise is grid-1 control of remote or semiremote cutoff tubes.

8-22 THE ANGLE MEASUREMENT STABILIZATION PROBLEM

In airborne radars the measurement of target angular position is complicated by the angular motion of the airborne platform. This paragraph will discuss the general features of this problem and the approaches that are employed to solve it. Subsequent paragraphs will show how a specific stabilization problem — the AI radar search and track stabilization problem — might be approached. The techniques and lines-of-reasoning used in this example are typical of those which must be employed for the solution of any airborne radar stabilization problem.

The essential features of the problem are illustrated by the simple one-dimensional representation of Fig. 8-23. The space pointing direction of the antenna A_{TL} is made up of two components: (1) the angle A_L of the antenna with respect to the aircraft and (2) the space orientation angle of the aircraft A_A . Thus changes in the orientation of the aircraft — due either to maneuvering or disturbances from wind gusts, etc. — will cause corresponding changes in the space pointing direction of the antenna.

From a tactical standpoint, this situation is undesirable. The line of sight from the radar to the target is relatively independent of radar aircraft orientation (neglecting long-term kinematic effects, it is completely independent). Thus, the effect of aircraft platform motion is to degrade the radar's ability to measure the target's position in space.

The term *angle stabilization* refers to the family of techniques employed to isolate the radar measurements from the degrading influences of aircraft motions. These techniques fall into two general classes: (1) data stabilization and (2) antenna stabilization.

Data Stabilization. With this technique, no changes are made to the basic control loops illustrated in Fig. 8-23. The effects of aircraft motion are compensated in the data-processing system by correcting the antenna angle measurements by appropriate functions of the measured platform motion.

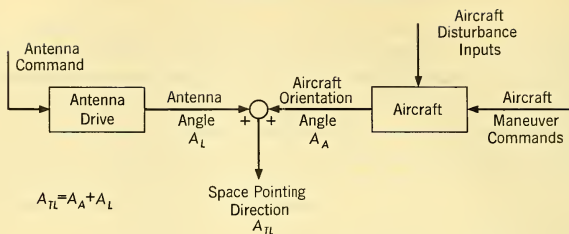


FIG. 8-23 Basic Relationships in the Airborne Antenna Drive System.

This technique is generally applicable to fan-beam AEW systems and other similar applications where platform motions cause measurement errors but do not cause loss of the target. Data stabilization finds particular favor where the antenna structure is so bulky as to preclude any other approach.

Antenna Stabilization. For the vast majority of airborne radar applications — missile seekers, AI radars, side-looking radars, infrared systems — stabilization of the antenna itself usually is required.

The basic objectives of such a stabilization system may be derived as follows.

From Fig. 8-23, the space pointing direction of the antenna may be expressed:

$$A_{TL} = G_a \times (\text{antenna command}) + G_d \times (\text{aircraft disturbance inputs}) + G_m \times (\text{maneuver commands}) \quad (8-21)$$

where G_a , G_d , and G_m are the transfer functions of the antenna drive and the aircraft.

From a tactical standpoint, the desired relationship is

$$A_{TL} = G_a \times (\text{antenna command}) = A_{TL} \text{ desired.} \quad (8-22)$$

Thus, the stabilization system must have two primary objectives:

1. It must provide control loops which reduce the effective couplings (G_d and G_m) between aircraft and antenna motion. (The required amount of reduction is a function of the expected tactical use requirements).
2. It must provide control means for driving the antenna to the desired space pointing direction.

Antenna Stabilization During Search. During the search phase of radar operation, the problem is to maintain surveillance of a predetermined volume of space despite the perturbations caused by platform motion. During this phase, target data are not used for control of the antenna; rather the antenna is driven by open-loop command data to sweep out the

desired space volume. The antenna is commanded to move in a direction opposite to that of the aircraft. The general means for solving the stabilization problem in this phase are shown in Fig. 8-24. The antenna is driven by

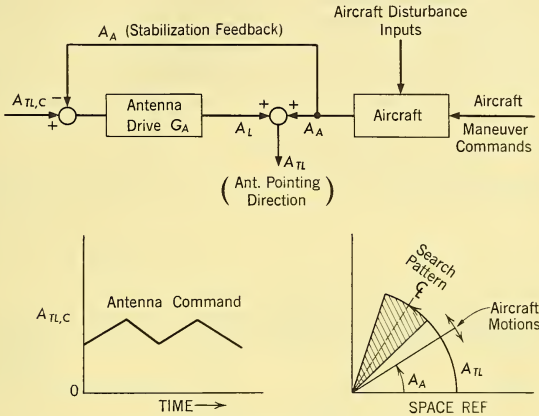


FIG. 8-24 Basic Search Stabilization System: Single Axis.

a generated command function $A_{TL,c}$ as shown. A feedback signal A_A provides stabilization by subtracting the aircraft orientation angle from the command angle. Thus, we may write

$$A_{TL} = G_a(A_{TL,c} - A_A) + A_A \quad (8-23)$$

$$A_{TL} = G_a A_{TL,c} + (1 - G_a) A_A.$$

If G_a is essentially unity over the frequency range of $A_{TL,c}$ and A_A

$$A_A \doteq A_{TL,c} \quad (8-24)$$

which is the desired result.

The critical elements of such a system are seen to be:

1. The accuracy of the angular reference which provides the feedback signal.

2. The closed-loop gain and frequency response of the antenna drive which must be sufficient to follow the input commands $A_{TL,c}$ and the stabilization feedback signals. Generally, the dynamic response requirements imposed by the command function are the most severe.

Additional complications are introduced by the more practical problem of stabilization in two or three axes. While the basic principles remain the same, the problem geometry will involve somewhat complex angular

transformations. These will be discussed later in the example of a detailed search stabilization design for an AI radar.

Tracking Stabilization. During tracking, information from the target can be employed to position the antenna. As described in Chapter 6, various techniques such as conical scan or monopulse can be used to create an error signal which indicates the amount of error between the line-of-sight and the antenna pointing direction. Generally, however, the stabilization provided by the radar angle tracking control is not sufficient: a faster, tighter inner stabilization-control loop must be used to provide the necessary isolation from aircraft motions.

A basic tracking radar stabilization system is shown in Fig. 8-25.¹² The outer control loop represents the generation of the automatic angle tracking

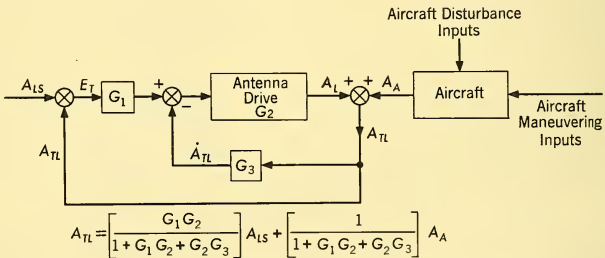


FIG. 8-25 Basic Tracking Radar Stabilization System. A_{LS} = sight target line; E_T = angular error signal; A_{TL} = stabilization feedback; G_3 = rate-measuring device (gyro).

error signal. The design considerations for such control loops are covered in Chapter 9. Stabilization is provided by measuring the *space angular rate*, \dot{A}_{TL} , of the antenna and feeding this signal into the antenna drive as shown in Fig. 8-25. This arrangement yields the interrelations among antenna pointing direction line-of-sight inputs and aircraft motions shown.

If the gain of the control loops, the products of $G_1 G_2$ and $G_2 G_3$, are much greater than unity for frequencies greater than those inherent in the line-of-sight angle A_{LS} and the aircraft space angle A_A , the equation in Fig. 8-25,

$$A_{TL} = \left[\frac{G_1 G_2}{1 + G_1 G_2 + G_2 G_3} \right] A_{LS} + \left[\frac{1}{1 + G_1 G_2 + G_2 G_3} \right] A_A \quad (8-25)$$

becomes

$$A_{TL} = \left(\frac{G_1}{G_1 + G_3} \right) A_{LS} + \left[\frac{1}{G_2(G_1 + G_3)} \right] A_A. \quad (8-26)$$

¹²Variations of this system are discussed in Paragraph 8-31.

Over the low-frequency range of A_{LS} , $|G_3| \ll |G_1|$, and the equation becomes

$$A_{TL} \approx A_{LS} + \frac{A_A}{G_1 G_2}. \quad (8-27)$$

Since $G_1 G_2 \gg 1$ in the frequency region of interest, $A_{TL} \approx A_{LS}$ as desired.

Of course, these relationships hold only if the loop gains are high and the control loops are stable. These are conflicting requirements which are discussed in the following paragraphs where the design of the stabilization loops is covered in detail.

8-23 AI RADAR ANGLE STABILIZATION

The primary function of the AI radar control system is to detect a target and to provide tracking information to the interceptor fire-control system about the target's relative position and motion.

The degree of space stabilization that must be provided depends on (a) the magnitude of the interceptor space motions during an attack and (b) the accuracy with which the target position and rates must be known. These topics are considered in more detail in subsequent sections.

8-24 AIRCRAFT MOTIONS

The first step in the design of the stabilization system is to obtain a description of the aircraft angular¹³ motions that will occur in the detection and tracking phases of interceptor operation. The basic angular motions are the roll, pitch, and yaw of the aircraft. The origins of these motions may be outlined as follows. First of all, the aircraft must maneuver in accordance with the vectoring commands or the fire-control system commands. Superimposed on these desired *maneuvering motions* are the *oscillatory motions* resulting from lightly damped aircraft motions which are excited by the control actions and the tendency of the human pilot (or autopilot) to overcorrect an error. Finally angular motions will be excited by disturbances such as wind gusts and release of armament or other stores (*interference motions*).

For purposes of preliminary design of the stabilization loops, these motions may be described in several ways:

1. By the maximum expected roll, pitch, and yaw angles and angular rates and derivatives. These data can be estimated from attack-course studies and knowledge of aircraft operation.

2. By the time-response characteristics of the aircraft in yaw, pitch, and roll incident to impulse inputs. This information can be derived from equations which describe aircraft dynamics.

¹³Linear motion is not considered here since stabilization control loops are principally concerned with angular motion. Linear motion is considered in Paragraph 9-18 in systematic errors.

3. By the frequency-response characteristics of the aircraft in yaw, pitch, and roll. This information can also be derived from the equations which describe aircraft dynamics.

4. By an actual time response made of angular aircraft motions as an ideal attack course is flown. This assumes that an actual aircraft is available or that it can be simulated on an analogue computer and "flown" realistically with an autopilot or human pilot. An ideal, simplified radar and antenna tracking system may be assumed, but noise and approximate error filtering should be included in the simulation. The time responses may be used as follows:

(a) A Fourier analysis of the time responses may be made. This may be made by conventional methods, but it usually is not as useful in design synthesis as the other techniques are.

(b) Segments of the time response may be represented by sinusoids or parts of sinusoids of various amplitudes and frequencies. These data are particularly useful in the synthesis of antenna stabilization control loop frequency responses.

5. By a statistical description of the aircraft motion. This is usually not available, and the effort required to obtain the power density spectrum is considerable. However, motions due to gust disturbances are better described statistically, as discussed in following sections.

The most useful descriptions of aircraft motion for preliminary stabilizations considerations are given by methods (1), (3), and (4). Typical numerical values of modern interceptor aircraft motion described by these methods are given in the following pages.

Maximum Disturbances Incident to Aircraft Maneuvers. The controlled interceptor motions, as limited by the aerodynamic characteristics of the interceptor, which affect the angle tracking system design, are roll and roll rate, yaw and yaw rate, and pitch and pitch rate. Estimates of a typical interceptor's capabilities are:

(a) ROLL

Roll angle: $+180^\circ, -180^\circ$
Roll rate: $80^\circ/\text{sec}$ to $90^\circ/\text{sec}$

(b) YAW

Oscillation frequency: 3 to 6 rad/sec
Sideslip angle: $1-2^\circ$

(c) PITCH

Pitch angle: $+180^\circ, -30^\circ$
Pitch rate: $20^\circ/\text{sec}$ up to $40^\circ/\text{sec}$
Pitch frequency: 2-13 rad/sec
Pitch oscillation angle: Pitch rate/pitch frequency = 1.5° to 10°

The values are typical of maxima that may be encountered. Actually, the *kinematics* of most attack courses do not require maneuvers of this magnitude; for example, the lead collision type of attack described in Chapter 2 theoretically requires no maneuvering at all once the initial error has been corrected. Despite this fact, however, the aircraft will experience relatively large angular rates during an attack because of lightly damped oscillatory modes in the aircraft response and the marginal stability which characterizes the pilot-aircraft steering loop.

The data of Fig. 8-26, which were obtained from a typical simulation program, illustrate the principle. In this simulation, a human pilot attempted

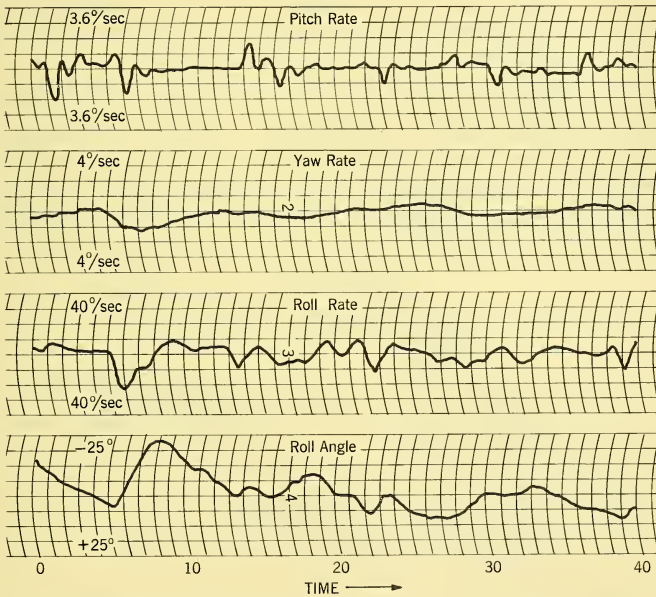


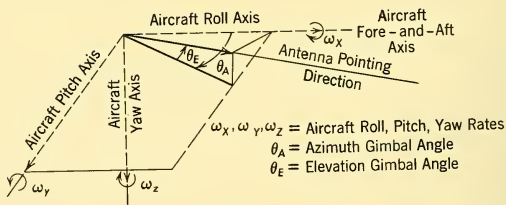
FIG. 8-26 Typical Simulation Results, Showing Aircraft Motions During an Attack.

to fly an attack course using information from a display which presented steering error and aircraft roll and pitch angles. The steering error signal was contaminated by radar tracking noise. Both the steering error and the noise were passed through a 0.5-sec filter prior to display.

Under these conditions a major portion of the aircraft motions took place at several relatively well defined frequencies. Rolling motions predominated; these took place at frequencies between 0.5 and 3.0 rad/sec, with maximum rolling rate amplitudes in the range of 0 to 20 deg/sec. Also evident is a yawing oscillation at a frequency of 0.5 rad/sec, and a maximum yawing rate amplitude of 1.2°/sec, and a small pitching oscillation at a frequency of 6 rad/sec and a maximum pitching amplitude of 2°/sec.

Sinusoidal Representation of Disturbances. This method is more useful in determining the stabilization control loop specifications. Specifically, this information may be obtained from actual time responses of a simulated aircraft on an analog computer as was shown in the preceding discussion. Portions of the time responses may be approximated by sine waves, and the amplitudes and frequencies of the sine waves can be recorded for various aircraft motions from several different courses.

To study the effect of aircraft motion on tracking-line stabilization, the aircraft motions are converted into motion with respect to the axes of the antenna gimbals. Usually, the antenna has two gimbals.¹⁴ The azimuth gimbal allows the antenna to rotate about an axis parallel to the aircraft's vertical axis; the elevation gimbal permits the antenna to nod up or down. The basic angle and angular rate relationships for such a two-gimbal system are shown in Fig. 8-27.



Transformations

Antenna Rates Due to Aircraft Angular Rates:

$$\text{Azimuth } \omega_A = \omega_x \cos \theta_A \sin \theta_E + \omega_y \sin \theta_A \sin \theta_E + \omega_z \cos \theta_E$$

$$\text{Elevation } \omega_E = -\omega_x \sin \theta_A + \omega_y \cos \theta_A$$

FIG. 8-27 Angle and Angular Rate Relationships for a Two-Gimbal Radar Antenna.

When the antenna tracking lead angle is large, the aircraft rolling motions appear as azimuth and elevation disturbances as is demonstrated by the ω_x terms in the transformation relationships in Fig. 8-27. This fact

¹⁴In some missile applications it is necessary to provide a third gimbal to space-stabilize the antenna in roll. This is not considered here.

is important because the rolling rates can be quite large relative to the yawing and pitching rates (see Fig. 8-26).

When the sinusoids representing aircraft motions are transformed into equivalent motions in antenna coordinates, results like those shown in Table 8-3 are obtained.¹⁵ These results are typical of what might be

TABLE 8-3 ANTENNA DISTURBANCE FROM
AIRCRAFT MOTION OF FREQUENCY ω_D

\dot{A}_A ($^\circ$ /sec) Peak to Peak	A_A ($^\circ$) Peak to Peak	ω_D (rad/sec)
35	33.6	1.04
30	9.55	3.14
20	6.04	3.31
20	4.77	4.19
22	4.48	4.92
5	0.497	10.5
7.5	0.715	10.5

obtained for lead angles of 45 deg from a large sample of the type of simulation data displayed in Fig. 8-26. It is assumed that the magnitude of this motion could disturb either channel of the tracking control loops at any time without further attenuation.

It should be emphasized that in a missile system or in an autopilot-controlled aircraft, these motions can be calculated by considering the equations of aircraft motion in three dimensions as it follows a prescribed course, assuming that the aircraft and autopilot design are known well enough to be described mathematically. The resulting data can also be approximated by calculating the aircraft frequency response from its design equations; this is often done. Although the computations may be simplified through the use of matrix notation and block diagram representation, the cross-coupling between the control loops is complex and nonlinear because of trigonometric functions involved, and it is difficult to interpret except for simple cases which are discussed in the next section. The task becomes more difficult, if not impossible, when a human pilot is involved because the human transfer function is not defined to a usable degree of accuracy or with sufficient reliability necessary for realistic results. Therefore, when available, analogue simulation is the most propitious method of obtaining information about aircraft motion in space with or without autopilot control.

¹⁵In practice many more values should be used because the nature of these disturbances is essentially random and a large sample should be made to obtain representative data. Note, for example, that a single frequency may have different amplitudes at different times.

Aircraft Transfer Functions. The aircraft cannot maneuver with large amplitudes at high frequencies. This can be shown from simplified transfer functions of the aircraft relating control surface motion to aircraft motion. The transfer functions can be obtained from transforms of the differential equations which describe aircraft motion. When simplified, to eliminate the short term yawing oscillation term, the transfer function relating aircraft heading to control surface position reduces to the form $K/S(1 + T_s)$.¹⁶ K and T depend upon the particular aircraft characteristics, and a frequency plot of this function should resemble the plotted data described in the preceding paragraph.

The general form of the space isolation required by the radar antenna should be the reciprocal of the aircraft response transfer function. The equivalent gain factor of the isolation transfer function ultimately depends upon the amount of isolation needed, the equivalent gain of the aircraft, pilot or autopilot, course computer, and the error presentation as discussed in Paragraph 8-32.

Gust Disturbances. As an aircraft flies an attack course, it is subjected to winds and turbulence or velocity fluctuations in the surrounding air. Turbulence disturbs the aircraft in a random manner, and its general effect is referred to as a *gust disturbance*.

Because of their random nature, gust disturbances are best determined by measurement and then described by power density spectra. The data of typical measurements and the associated normalized power density spectra are presented in the following documents:

- (a) *An Investigation of the Power Spectral Density of Atmospheric Turbulence* by G. C. Clementson, Report No. 6445-T-31, Instrumentation Laboratory, M.I.T., May 1951.
- (b) *A Statistical Description of Large-Scale Atmospheric Turbulence* by R. A. Summers, Report No. T-55, Instrumentation Laboratory, M.I.T., May 1954.

The normalized power density spectra may be applied to a specific aircraft by scaling both abscissa and ordinate. The effect on the tracking loop antenna position and rate may be then found by multiplying the gust power density spectrum by the square of the transfer function magnitude relating the disturbance to the antenna position rate in the channel corresponding to the direction that the gust disturbances were measured. The square root of the integral of this product is the rms value of the antenna motion or rate. For most tactical situations the effect of gust disturbances is negligible when compared with other factors and it will not be considered further in this text. For high-speed, low-altitude flights, however, gusts

¹⁶Actually, the transfer function varies in roll, pitch, and yaw. The most useful transfer functions are those which transform aircraft motion to antenna motion.

can sometimes be quite severe and in such cases their effects should be studied as part of the systems design.

8-25 STABILIZATION REQUIREMENTS

The stabilization requirements vary for the search and track modes. In search, a lack of space stabilization would allow the search pattern to move with the interceptor from its preassigned space sector, and the desired target might not be found.

Deviations from true space stabilization in search are produced mainly by

- (a) Aircraft angular motions¹⁷
- (b) Dynamic antenna control loop errors
- (c) Incorrect commands to the control loops
- (d) Inaccuracies in the vertical reference

The total deviation from true space stabilization that can be allowed is related to the loss in target detection probability that it produces.¹⁸ Usually 0.25° to 1.5° deviation from the ideal space stabilized pattern can be permitted. However, the deviation varies as a function of the search angle with respect to the aircraft.

In the tracking mode, antenna space stabilization is needed for several reasons:

(a) To prevent the course computer from operating on inaccurate antenna motions due to interceptor space motion rather than target sight-line motion in space.¹⁹

(b) To prevent the antenna beam from drifting off the target because of aircraft motion during short periods when the radar signal fades.

(c) To prevent system instability caused by coupling of the interceptor motion with its commands through resulting antenna space motions. This is a form of positive feedback, because as the interceptor moves the antenna in space, the antenna motion produces signals used by the computer to direct the interceptor farther in the same direction. To prevent instability in this positive feedback loop, it is necessary to have the loop gain²⁰ always less than unity. This is most important in systems using an autopilot. The problems involved in providing the necessary isolation in the search and track modes are discussed in the paragraphs that follow.

¹⁷Another source of error may be produced, especially in some missile systems, from control loop disturbance torques created by an unbalanced antenna undergoing large rotational space accelerations. However, a detailed discussion of this is beyond the scope of the text.

¹⁸Detection probability is discussed in Paragraphs 3-9 to 3-14.

¹⁹Another source of error may be produced, especially in some missile systems, from control loop disturbance torques created by an unbalanced antenna undergoing large linear or rotational space accelerations. However, a detailed discussion of this is beyond the scope of the text.

²⁰This includes antenna motion detectors, filters, computer, pilot or autopilot, aircraft transfer functions, and the isolation factor provided by the closed, space stabilized antenna control loops (See Paragraph 8-32).

8-26 SEARCH PATTERN STABILIZATION

The space reference for the search pattern control loops is derived from a vertical gyro. The accuracy of the vertical gyro in maintaining a true vertical need not be extremely good; available vertical gyros provide sufficient accuracy for stabilizing the antenna search pattern during the relatively short times that a particular target is being sought. The vertical gyro used has two degrees of freedom and is provided with position detectors that measure the aircraft pitch and roll angles with respect to the vertical. Except for some missile applications, yaw angles are generally not measured because the aircraft yaw motions are better controlled and the search pattern is usually much wider in a horizontal direction (see Paragraph 2-26).

8-27 SEARCH STABILIZATION EQUATIONS

Because the antenna motion of a two-gimbal antenna does not include roll correction directly, the aircraft roll motion must be converted into the proper azimuth or elevation commands for the two-antenna control loops.

The exact transformation of aircraft motions to antenna commands is rather complex as shown by the following formulae:

$$\sin E_A = \sin A_S \cos E_S \sin \phi - (\cos A_S \cos E_S \sin \theta - \sin E_S \cos \theta) \cos \phi \quad (8-28)$$

$$\sin A_A \cos E_A = \sin A_S \cos E_S \cos \phi + (\cos A_S \cos E_S \sin \theta - \sin E_S \cos \theta) \sin \phi \quad (8-29)$$

where E_A is the elevation antenna angle command to move the antenna with respect to the interceptor

A_A is the azimuth antenna angle command to move the antenna with respect to the interceptor

E_S is the elevation search angle with respect to space

A_S is the azimuth search angle with respect to space

ϕ is the aircraft roll angle in space

θ is the aircraft pitch angle in space.

It is possible to mechanize these equations to within a few minutes of arc²¹ by using several resolvers, as shown in Fig. 8-28; but to simplify the mechanization, the transformation equations are often simplified. This can be done in several different ways. One of the approximations which

²¹This inaccuracy is due to components, principally the resolvers.

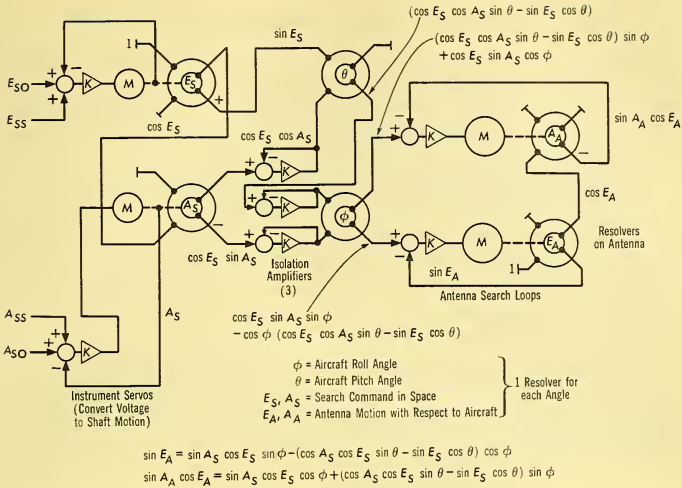


FIG. 8-28 Exact Coordinate Conversion Mechanization for a Two-Gimbal Antenna Search Pattern.

creates very little significant error (in the command signals for search angles within 50 deg) is expressed by the following equations:

$$E_A = A_S \sin \phi - (\theta - E_S) \cos \phi \tag{8-30}$$

$$A_A = A_S \cos \phi - (\theta - E_S) \sin \phi \tag{8-31}$$

These equations are shown mechanized in Fig. 8-29.

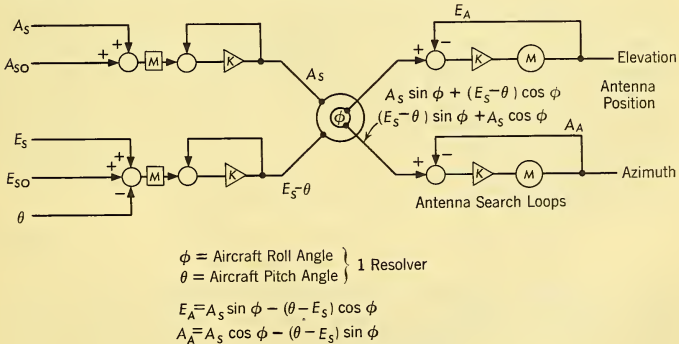


FIG. 8-29 Approximate Coordinate Conversion Mechanization.

8-28 STATIC AND DYNAMIC CONTROL LOOP ERRORS

Perhaps the largest errors in the search-pattern stabilization are engendered in the antenna control loops. These errors are reduced to satisfactory limits by proper design of the control-loop gains and bandwidths or by modification of the pattern command signals. Errors to be considered are:

- (a) Static or steady-state errors due to aircraft motion
- (b) Static or steady-state errors due to search-pattern velocity
- (c) Dynamic errors due to changes in search-pattern command signals

It may be considered that part of the total allowable search-pattern errors may be allotted to the antenna control loops. For example, a 0.35° steady-state error may be assumed and it may be divided equally between aircraft motion and search-pattern velocity signals; i.e., the allowable error contribution of each source is 0.25° .²²

To provide a means for translating the error requirements into a design specification, a generic form must be assumed for the search stabilization and drive system transfer function. For the example to follow, the assumed open-loop transfer function will have the form

$$\frac{K_v(1 + S/\omega_2)}{S(1 + S/\omega_1)(1 + S/\omega_3)} \quad \omega_1 < \omega_2 < \omega_3. \quad (8-32)$$

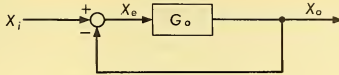
The following analysis will demonstrate how the values of K_v , ω_1 , ω_2 , and ω_3 can be chosen to meet a set of system requirements.

Aircraft Motion Errors. To reduce the steady-state errors caused by aircraft motion to 0.25° , the nature of the aircraft motion must be known. For example, Table 8-3 shows the amount of antenna movement that would take place at large lead angles if the antenna were not stabilized. The search stabilization loop generates position command signals which — if computed correctly — are equal and opposite to the disturbance caused by aircraft motion. However, the control loops that drive the antenna with respect to the aircraft have finite gain and bandpass. Thus, the actual position of the antenna will tend to lag the stabilization commands. As shown in Fig. 8-30, the amount of lag depends upon the frequency and magnitude of the input relative to the open loop gain of the stabilization loop at the input frequency. In order that the lag be kept below 0.25° at all input frequencies, the minimum loop gain must be

$$G(\omega) \geq \frac{x_i(\omega)}{\text{error specification}} = \frac{x_i(\omega)}{0.25}. \quad (8-33)$$

As an example, the input at 1.04 rad/sec is 33.6° peak-to-peak or $x_i = 16.8^\circ$. Thus, the required open-loop gain of $f = 1.04$ is 67.2. Similar

²²Since aircraft motion is independent of the search pattern, the individual errors may be added by taking the square root of the sum of the squares.



Let: X_e = Loop Error
 X_i = Loop Input Command
 X_o = Loop Output

In Search Loop, X_i Contains Signals to Move Antenna and Stabilize it in Space. It is Desired to keep the Loop Error X_e Less Than a Particular Magnitude $|X_e|$

Since

$$|X_e| = \frac{|X_i|}{1+G_o} \approx \frac{|X_i|}{|G_o|} \text{ for } G_o \gg 1$$

$$|G_o| \approx \frac{|X_i|}{|X_e|}$$

$|X_i|$ As a Function of Frequency is Known and $|G_o|$
 As a Function of Frequency is Found as Desired

FIG. 8-30 Derivation of Required Loop Transfer Function.

calculations can be made for other frequencies, resulting in the circled points shown in Fig. 8-31. These points define the *minimum* open-loop

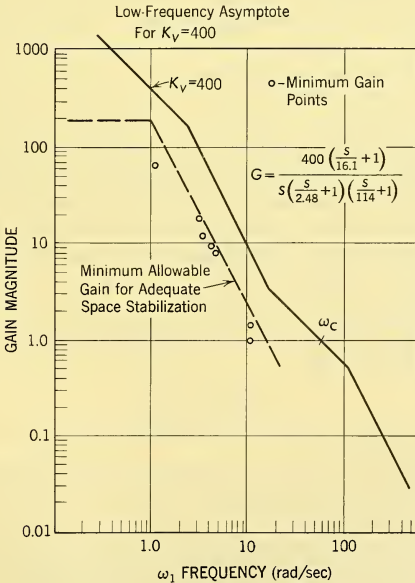


FIG. 8-31 Search Loop: Open-Loop Transfer Function.

gain necessary to provide isolation from the expected aircraft angular motion.²³

Search-Pattern Velocity Errors. In addition to providing isolation from aircraft motion, the steady-state error caused by the antenna scanning motion must also be considered. Usually the antenna motion is uniform, and in the azimuth or horizontal space direction it is a constant angular velocity. The steady-state position error is the velocity divided by the velocity constant.²⁴ The constant antenna sweep velocity is determined as discussed in Paragraph 5-7 and is usually between 75 and 150°/sec.²⁵ Therefore, to keep the error below 0.25° for a 100°/sec search velocity, the velocity constant must be

$$\frac{100^\circ/\text{sec}}{0.25^\circ} = 400 \text{ sec}^{-1} = K_v. \quad (8-34)$$

Dynamic Errors Due to Search-Pattern Velocity Changes. Some distortion is likely to occur at the ends of the search pattern because physical control loops cannot follow the rapid changes in command signal which are used to change the antenna velocity and position at the end of a horizontal sweep to another horizontal sweep.²⁶ The transient at the end of the sweep requires a longer time to reach a small steady-state value than does the vertical-motion transient because the vertical motion is usually a small position step instead of a large velocity reversal. Since it is desired to resume the steady-state error in the shortest time after the sweep direction is changed, both azimuth and elevation control loops may be designed to realize the desired sweep transient. If there were no aircraft roll, each control channel would have different characteristics. But because a large roll angle transfers much of the horizontal motion to the elevation channel, both channels should be designed to have the characteristics of the azimuth channel in horizontal flight.

²³The specification derived in this manner is somewhat pessimistic because the antenna is not always at the large lead angle used to obtain the data in Table 8-3.

²⁴If the open loop has the form

$$\frac{K_v(1 + s/\omega_2)}{s(1 + s/\omega_1)(1 + s/\omega_3)}$$

K_v is the velocity constant.

²⁵The search pattern may have several forms — a horizontal Palmer scan with several horizontal sweeps spaced at the beam width and a diagonal return, a horizontal sweep and return spaced at the beamwidth, a spiral scan, a circular scan, a vertical scan, or combination of these types.

²⁶The command changes can be made gradually to eliminate transients, but this usually involves a more elaborate signal generator. Even then, the ideal pattern will not necessarily be obtained, and more time may be required for the antenna sweep to resume the desired constant velocity.

Fig. 8-32 shows the nature of the error transient that occurs if the direction of the horizontal sweep changes instantaneously. An important

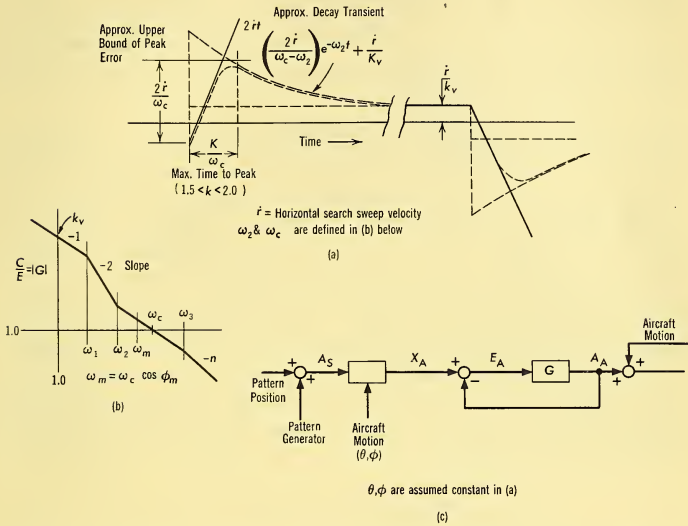


FIG. 8-32 Dynamic Transients. (a) Approximate position error transient for linear search sweep. \dot{r} = horizontal search sweep velocity; ω_2 and ω_c are defined in design equations. (b) Open-loop transfer function. Design equations (approximations usually accurate enough for engineering design purposes): (1) Gain:

$$K_v = \frac{\omega_c (\omega_2)^2}{\omega_2 \omega_1} \omega_1$$

$$(2) \text{ Phase: } \frac{-180^\circ + \phi_m}{57.3} = \beta_m = \frac{\pi}{2} - \left(\frac{\pi}{2} - \frac{\omega_1}{\omega_m}\right) + \left(\frac{\pi}{2} - \frac{\omega_2}{\omega_m}\right) - \frac{\omega_m}{\omega_3}(n-1)$$

$$(3) \text{ Minimum Phase at } \omega_m: \frac{2\beta_m}{2\omega_m} = 0$$

NOTE: Substitute $\left(\frac{\omega_m}{\omega_3}\right)$ in Equation 2 when solving for ω_1 or ω_2 .

(c) The search loop (azimuth channel).

point is that the peak error of the transient, frequently neglected, can be much larger than the steady-state error, and a relatively large time may be required for it to settle if ω_2 in Fig. 8-32 is small. Actually, there are several ways in which the dynamic error can be specified. For example, the time

required for the peak error to decay to within a certain percentage of the final value could be specified, a maximum time for the peak error to occur could be given, or the percentage distortion of the overall pattern dimensions which can be tolerated could be specified. Fig. 8-32 indicates approximate but useful relationships between the dynamic transient, the open-loop transfer function, and the steady-state error that help determine the search control-loop design characteristics.

For example, if the allowable distortion incident to peak dynamic error is to be within 10 per cent of an overall pattern sweep of 30° , the peak error would be 3° . From Fig. 8-32, the upper bound is about $2\dot{r}/\omega_c$ above the negative error, and the approximate peak error above zero is $2\dot{r}/\omega_c - \dot{r}/K_v = 3$ deg; and if $K_v = 400 \text{ sec}^{-1}$ and \dot{r} is $100^\circ/\text{sec}$, $100(2/\omega_c - 1/400) = 3$ and $\omega_c = 61.6 \text{ rad/sec}$. This is the search-loop bandwidth. If greater accuracy is required, a more sophisticated pattern command would be necessary with special accelerating and decelerating controls—perhaps nonlinear control for maximum effort. This is not usually necessary, however, to obtain a relatively constant sweep velocity. Other characteristics of the open-loop transfer function are found from stability considerations, and an optimum system can be determined directly from the three following equations relating the corner frequencies, peak phase margin, and the loop gain shown in Fig. 8-32.²⁷

SEARCH LOOP SYNTHESIS

1. Loop gain equation:

$$\frac{\omega_c \left(\frac{\omega_2}{\omega_1} \right)^2}{\omega_2 \left(\frac{\omega_2}{\omega_1} \right)} \omega_1 = K_v = 400 \quad (8-35)$$

$$\frac{\omega_2}{\omega_1} = \frac{K_v}{\omega_c} = \frac{400}{61.6} = 6.49 \quad (8-36)$$

2. Phase equation (frequency response peak = 1.3,²⁸ peak phase margin

$\phi_m = 50.3^\circ$ at a frequency $\omega_m = \omega_c \cos \phi_m = 0.64\omega_c$):

$$\beta_m = \frac{(-180^\circ + 50.3^\circ)}{57.3} = \frac{-\pi}{2} - \left(\frac{\pi}{2} - \frac{\omega_1}{\omega_m} \right) + \left(\frac{\pi}{2} - \frac{\omega_2}{\omega_m} \right) \quad (8-37)$$

$$-\frac{\omega_m}{\omega_3} + 0.69 = \frac{(\omega_2 - \omega_1)}{\omega_m} + \frac{\omega_m}{\omega_3}$$

where $\beta_m =$ phase angle of G at ω_m .

²⁷The derivations of these equations are discussed in the paper, "Synthesis of Feedback Control Systems with a Minimum Lead for a Specified Performance," by George S. Axelby in *IRE Transactions in Automatic Control*, PGAC-1, May 1956.

²⁸The closed-loop frequency response peak M_p occurs at a frequency $\omega_m = \omega_c \cos \phi_m$ and sin $\phi_m = 1/M_p$ in the optimum, minimum lead system.

3. Minimum phase equation:

$$\frac{\omega_2 - \omega_1}{\omega_m^2} = \frac{1}{\omega_3} \quad (\text{differentiate Equation 8-37}) \quad (8-38)$$

Combining Equations 8-37 and 8-38

$$0.69 = \frac{2(\omega_2 - \omega_1)}{\omega_m} = \frac{2(\omega_2 - \omega_1)}{0.64\omega_c} \quad (8-39)$$

$$\text{and} \quad 0.69 = \frac{2\omega_1(\omega_2/\omega_1 - 1)}{0.64\omega_c} = \frac{2\omega_1(5.49)}{(0.64)(61.6)} \quad (8-40)$$

from which $\omega_1 = 2.48$ and $\omega_2 = 16.1$

$$\omega_m = 39.4$$

$$\omega_c = 61.6$$

$$\omega_3 = 114.$$

These are the search loop corner frequencies, bandwidth, and maximum phase frequency in radians per second. A log magnitude phase diagram or

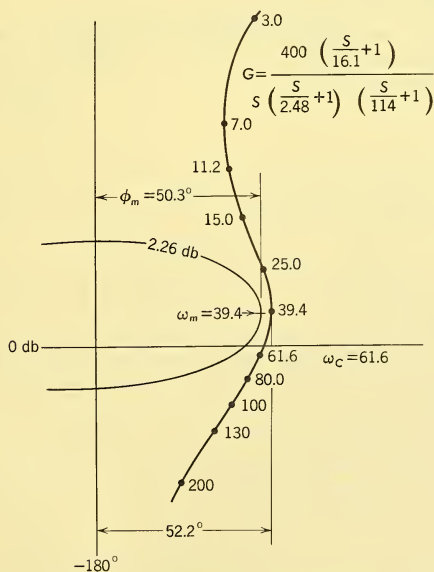


FIG. 8-33 Nichols Chart of Search Loop.

Nichols chart is shown in Fig. 8-33.²⁹ Note that if the corner frequency in Fig. 8-32 had been a double corner, ω_3 would be doubled; if it had been a triple-corner frequency or equivalent, ω_3 would be tripled, etc. However, Fig. 8-33 would be essentially the same in the crossover frequency region. This example illustrates one method of synthesizing the search control loop directly with a minimum of the usual cut and try effort. The procedure is similar, even if other criteria are used to specify the search loop performance.

8-29 SEARCH LOOP MECHANIZATION

Actual circuit details of search loop mechanization cannot be discussed in general terms because the control loop components vary with specifications, with the nature of the power available, and with the antenna size. However, a few general considerations can be discussed.

The basic components needed to mechanize each of the two search loops are shown in Fig. 8-34. The coordinate converter needed to correct the

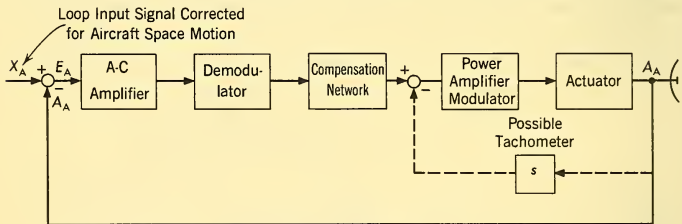
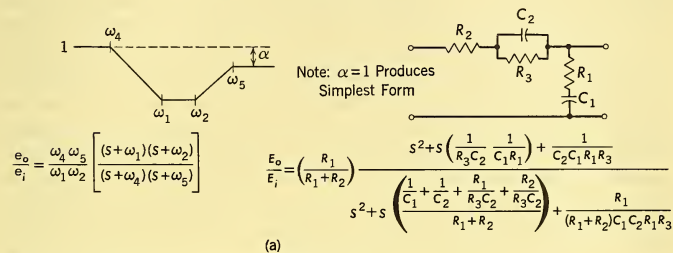


FIG. 8-34 General Block Diagram of Search Loop (One Channel).

input signal for aircraft motion is not shown because it was discussed in Paragraph 8-27 and illustrated in Fig. 8-28 or 8-29. The practical problems involved in its construction are those of making proper resolver connections with correct phasing of a-c signals. For the exact transformation, the signals between the six resolvers must pass through isolating amplifiers or phase-shifting devices. The loop-actuating signal, or error signal, is usually an a-c voltage proportional to the difference between the input signal X_A and the controlled antenna position A_A as shown in Fig. 8-34. It is obtained in the exact coordinate transformation from the windings of resolvers, which are mounted on the antenna. In the approximate transformation, the a-c error signal is obtained from the sum of voltages from the vertical gyro roll

²⁹Note that the calculated maximum phase margin is about 2° greater than the desired design value. This discrepancy, conservative but negligible in an actual system, occurs because the locus was calculated exactly from the transfer function which was determined from approximate equations.

resolver and a potentiometer on the antenna. The latter is not shown, but it is implied in the summing symbol in Figs. 8-29 and 8-34. Wire-wound potentiometers may be used in each channel, but induction potentiometers are preferred (especially if the loop gain is high) to prevent oscillations between potentiometer wires.

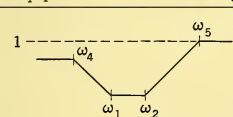


(a)

Design Formulas

$$C_1 = \text{arbitrary} \quad R_2 = R_1 \left(\frac{1 - \alpha}{\alpha} \right) \quad C_2 = \frac{C_1}{C_1 K - 1} \quad K = (\omega_4 + \omega_5 - \omega_2)(R_1 + R_2)$$

$$R_1 = \frac{1}{\omega_1 C_1} \quad R_3 = \frac{1}{\omega_2 C_2} \quad \alpha = \frac{\omega_4 \omega_5}{\omega_2 \omega_2} \quad K = (\omega_4 + \omega_5 - \omega_2) \left(\frac{R_1}{\alpha} \right)$$



$$\frac{e_o}{e_i} = \frac{(s + \omega_1)(s + \omega_2)}{(s + \omega_4)(s + \omega_5)}$$

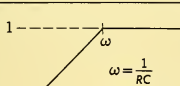
$$\frac{E_o}{E_i} = \frac{s^2 + s \left(\frac{1}{C_1 R_1} + \frac{1}{C_3 R_4} \right) + \frac{1}{C_3 C_1 R_1 R_4}}{s^2 + s \left(\frac{1}{C_3 R_1} + \frac{1}{C_3 R_2} + \frac{1}{C_3 R_4} + \frac{1}{C_1 R_1} \right) + \frac{R_4 + R_2}{C_1 C_3 R_1 R_2 R_4}}$$

(b)

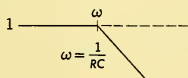
Design Formulas

$$C_1 = \text{arbitrary} \quad R_2 = \frac{1}{N C_1} - R_1 \quad R_4 = \frac{1}{\omega_2 \omega_3}$$

$$R_1 = \frac{1}{\omega_1 C_1} \quad C_3 = \frac{\omega_1}{R_2(\omega_4 \omega_3 - \omega_1 \omega_2)} \quad N = \frac{(\omega_4 \omega_5 - \omega_1 \omega_2)}{(\omega_4 + \omega_5)(\omega_1 + \omega_2)} \approx \omega_4$$



(c)



(d)

FIG. 8-35 RC Compensation Networks.

As in all feedback control systems, the error signal is used to position the antenna in a direction which will decrease the error. Generally, however, the control system characteristics shown in Fig. 8-31 do not exist naturally, and it is necessary to provide some form of compensation. On small, low-power systems this compensation can be provided mechanically by adding extra inertia to the antenna inertia with fluid coupling; but with large antennas where space is limited and the power is relatively high, it is more practical to provide electrical compensation in the form of RC networks as shown in Fig. 8-35. Design formulae are given to illustrate how the network parameters are related to the corner frequencies. Generally, the corner frequencies are chosen so that, in combination with the corner frequencies of other equipment in the loop, the desired overall characteristic shown in Fig. 8-31 is obtained. This is illustrated in Fig. 8-36. It should be noted that networks A and especially B of Fig. 8-35 would be used in series with the low-power circuits in the forward path of the search loop, and that

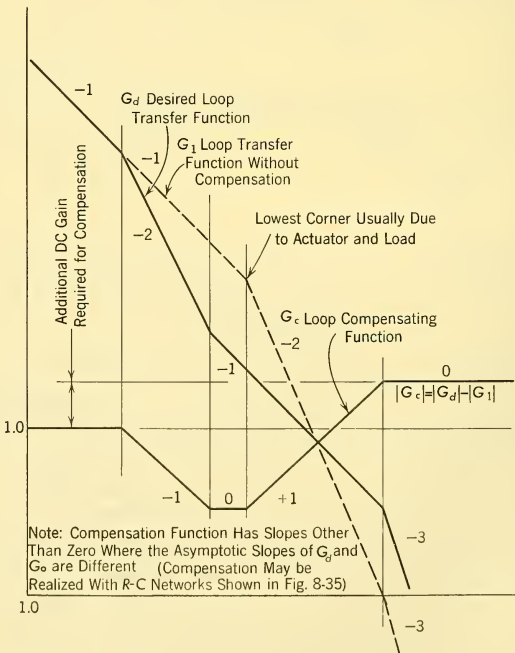


FIG. 8-36 Method of Determining Control-Loop Compensating Function.

network C, possibly in combination with D, might be used in a tachometer feedback path around the antenna, actuator, and power amplifier.

To use the RC compensating networks, it is necessary to convert the a-c error signal to a d-c signal with a demodulator.³⁰ A peak demodulator is often used in the search loop because it has (1) less high-frequency noise and (2) a smaller time constant than an averaging demodulator. After the error signal is demodulated and passed through d-c networks and a power amplifier, it is applied to the power actuator which moves the antenna. Generally, the actuator is a two-phase a-c electric motor or a hydraulic actuator controlled with an electrically operated valve. If an a-c motor is used, the d-c actuating signal from the compensating network must be modulated before it is applied to the motor through a power amplifier. This may be done electronically with vacuum tubes or with magnetic amplifiers which provide amplification, modulation, and power in one reliable unit. However, if a hydraulic actuator with a control valve is used, the d-c actuating signal may be used to control the valve. Some signal amplification may be provided with tubes or with transistors.

There are advantages and disadvantages in both types of antenna drives. The electric motor is cheaper and lighter than the hydraulic actuator; it does not need oil lines or rotary joints with oil seals; but it does not run as smoothly at low speeds, it is much less efficient, and it cannot produce as much steady-state torque or velocity as a hydraulic actuator of the same size. In addition, gear trains with troublesome backlash and friction are needed with electric drives, whereas they are not used with hydraulic actuators.³¹

Regardless of its type, an actuator must be selected which will provide the required search-pattern velocities and accelerations. The output power of the actuator must be greater than the power required to move the antenna inertia along the search-pattern paths in space in the presence of antenna friction and unbalance as well as aircraft pitch and roll motion which may be directly opposed to the desired antenna space motion. In fact, aircraft motion adds considerably to the required actuator torque, velocity, and power because the actuators move with respect to the airframe to generate a pattern in space rather than with respect to the airframe; thus the expected aircraft motion must be combined with the required search-pattern velocities and accelerations to determine the actuator performance characteristics. Note that the antenna inertia and unbalance in elevation may be less than in azimuth and the actuator may be correspondingly

³⁰Simple RC networks can be directly approximated with "notch" networks or resonant filters for a-c signals; but for airborne systems, this is usually not practical because of the accuracy required and because the carrier frequency varies.

³¹It is assumed that the hydraulic actuator discussed here consists of a shaft with a vane enclosed in a housing through which oil may be ported to either side of the vane to produce a shaft rotation. Of course, 360° rotation is not possible with this type of actuator.

smaller. In practice, however, the actuators often have the same size for production economy. Finally, it should be emphasized that the control loop and actuators should not be designed to have a performance much greater than that required, not only because increased size and weight would be involved but also because physical limitations inherent in the gimbals and the antenna structures of a given size place an upper practical limit on linear design,³² and as this limit of performance is approached, the cost and complexity of control equipment increases rapidly. Specifically, noise is always present in the search control loops, although not to the extent that it is in other control loops associated with fire-control systems, and its detrimental effects become more of a problem as an attempt is made to increase loop performance. In addition, structural resonant frequencies in the antenna make it nearly impossible to construct a stable loop with a crossover frequency near the resonant frequencies. Thus, there is a practical upper limit for the control loop bandwidth which is governed by the antenna characteristics. Since the performance of the loop is primarily a function of the bandwidth, the performance itself is limited.

8-30 STABILIZATION DURING TRACK

As was discussed in Paragraph 8-25, the tracking antenna must be stabilized in space to prevent:

- (a) system errors caused by the course computer operating on information from coupling between the interceptor and antenna motions.
- (b) the antenna radar beam from drifting away from the target during brief periods of radar signal fading
- (c) instability incident to coupling between the interceptor and antenna motions.

A portion of the required space stabilization is provided by the automatic tracking loops discussed in Chapter 9, except during periods when the radar signal fades. However, during normal operation, the typical tracking loop cannot provide effective isolation above frequencies equal to about one-half the track loop bandwidth, or about 3.0 rad/sec for a typical system. On the other hand, the interceptor may have appreciable motion at higher frequencies as indicated in Table 8-3. To provide the necessary space stabilization, a special automatic control loop is designed to move the antenna relative to the aircraft in a direction opposite to that of the aircraft motion in space.

To provide space stabilization, the control loop must obtain antenna space motion information. This is obtained and converted into useful electrical signals with gyroscopes mounted on the antenna or on the aircraft,

³²It is possible to devise nonlinear control loops which will provide increased performance in special cases.

and these signals are used to move the antenna in a direction opposite to undesired space motions. Of course, the correction signals are not exact; the gyros sense space velocity only, not the space position actually desired. However, a control loop with sufficient accuracy and speed of response can reduce antenna space motion to magnitudes much lower than those of the interceptor, and the residual space position errors are further reduced by the track loop which uses the target sight line as a reference. Details of designing these stabilization loops are discussed in the subsequent paragraphs.

8-31 POSSIBLE SYSTEM CONFIGURATIONS

Several physical configurations are used to mechanize the stabilization loops. Some of them are shown in Fig. 8-37.³³ Theoretically, all of them

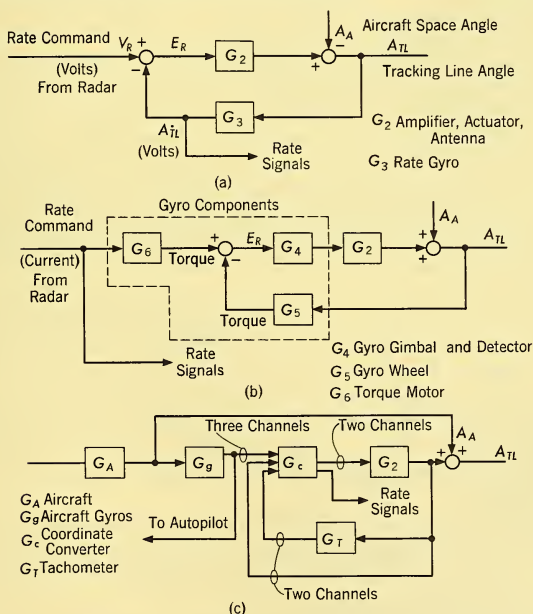


FIG. 8-37 Three Stabilization Loop Configurations. (a) Rate Gyro. (b) Integrating Gyro. (c) Aircraft Gyro.

³³Another form is not illustrated. It consists of a gimballed antenna dish which is rotated at high speeds to become an effective gyroscope. It is used in some missile tracking systems to provide stabilization.

are equivalent in a mathematical sense, but practically the arrangement of the physical equipment is entirely different, and, depending on the application, the mechanization can create discrepancies between the theoretical and realized loop performance. In fact, in many control loops the mechanization characteristics, which are incidental to the primary loop function, may prevent the realization of a workable system. Therefore, it is necessary to know as much as possible about equipment characteristics before a design is finalized. Unfortunately, component characteristics vary widely depending on the application, and it is impossible to discuss them in detail. However, the general principles of operation are outlined below for three different systems.

Rate Gyro Stabilization Loop. The rate gyroscope³⁴ is a self-contained unit which produces an electrical signal, usually a voltage, proportional to a space rate about a particular axis. In the stabilization loop design shown in Fig. 8-37a two rate gyros are used, one to measure space rates about the elevation axis and the other to measure space rates about the azimuth axis of the antenna. Both gyros are mounted on the antenna dish³⁵ where the antenna space rates are measured directly. As shown in the figure, the *space rate* (\dot{A}_{TL}) of the antenna tracking line angle, A_{TL} , is measured by gyro and converted to a voltage proportional to \dot{A}_{TL} . This voltage is compared with an antenna rate command voltage V_R and the voltage difference, the rate error E_R , is used to control the antenna actuators through appropriate amplifiers in a direction which will reduce the rate error.

Although only one channel is shown, two control loops are needed — one for each antenna motion — and these loops are interconnected, not only through the common space platform and antenna structure but through the gyroscopes as well, because space accelerations in one channel influence the gyro output voltage in the other channel to a degree which is determined by the gyro characteristics and the gyro orientation. However, unless the stabilization loop is of very large bandwidth, this effect can be made negligible by selecting the proper gyro and by choosing the optimum gyro axis mounting.³⁶

As shown in the figure, motion of the aircraft angle, A_A , is also detected by the gyro and converted to a voltage, essentially a rate error, which is used to move the antenna at a rate in space opposite to the aircraft space

³⁴See Locke, *Guidance*, Chap. 9, pp. 350-353.

³⁵Note that the effective azimuth gyro gain in volts per unit of velocity is proportional to the cosine of the elevation angle and that it will change during normal operation. For computational purposes this is often desired, but it may affect the stability of the azimuth stabilization loop if the elevation angle becomes large.

³⁶A complete discussion of this problem is given in a paper, "Analysis of Gyro Orientation," by Arthur Mayer in the *Transactions of the Professional Group on Automatic Control*, PGAC-6, December 1958, p. 93.

motion. Of course, this principle of operation is common to any stabilization loop.

Integrating Gyro Stabilization. The integrating rate gyroscope is a self-contained unit which produces an electrical signal, usually a voltage, proportional to the *time integral* of the torques applied around the gyro gimbal axis. A schematic representation of the integrating gyro and the simplified equations expressing its behavior are shown in Fig. 8-38. Such

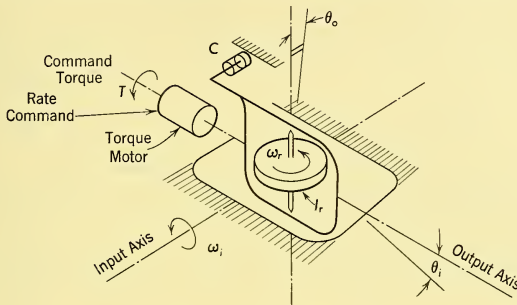


FIG. 8-38 Integrating Rate Gyro Relationships.

θ_o = gimbal angular displacement about output axis relative to gyro case

θ_i = input angular displacement about input axis

I_g = moment of inertia of gimbal assembly about the output (gimbal) axis

C = viscous damping constant

$T = K\omega_c$ = rate command torque

$H = I_r\omega_r$ = rotor angular momentum

I_r = rotor inertia

ω_r = rotor angular rate

DIFFERENTIAL EQUATION:

$$\frac{d^2\theta_o}{dt^2} + \frac{C}{I_g} \frac{d\theta_o}{dt} = \frac{H}{I_g} \omega_i - \frac{K}{I_g} \omega_c$$

$$\text{or for } K = H$$

$$\theta_o = K_o \frac{(\omega_i - \omega_c)}{S(1 + T_s)} \doteq K_o \frac{\omega_i - \omega_c}{S} \quad \text{for } T \ll 1$$

a device may be used as an integral part of the antenna control loop as shown in Fig. 8-37b. The rate command, in the form of variable current, is used to produce a torque on the gimbal holding the gyro wheel, with an electromagnet or torque coil in the gyro. This torque is opposed by the gyroscopic torque of the spinning gyro wheel which is induced primarily by antenna space motion about a particular axis. If these torques are not

equal, the gyro gimbal moves at a rate determined by the amount of viscous damping. The motion is detected with a sensitive transducer, often a microsyn, and transformed into a proportional voltage. This voltage is amplified and applied to the antenna actuator in a direction to reduce the rate error. Aircraft rates are also compensated in the same way.

A principal difference between the integrating gyro and the rate gyro control loop is that the steady-state error in the integrating gyro loop is zero³⁷ for a constant rate command while the steady-state error in the rate gyro loop has a finite value proportional to the rate command and inversely proportional to the d-c loop gain. However, this steady-state error in the rate loop can be reduced to a negligibly small value without great difficulty. The integrating gyro must have compensating networks to make it stable, whereas the loop with the rate gyro may not need compensating networks, depending on the degree of performance required. Practically, some compensation is usual in both forms of the stabilization loop.

Another difference in performance between stabilization loops using the integrating gyro and those using the rate gyro is incident to the saturating characteristics of the gyro. Large steady-state rates do not usually saturate the integrating gyro, because the gyro gimbal is in the forward path of the loop and its motion is proportional to the rate error, which is small. Even if it should saturate, it does not change the loop performance appreciably, because changes in the forward gain of a feedback control loop do not affect the overall loop characteristics or its performance significantly. On the other hand, the rate gyro saturation does occur for large, steady rates and its measuring and performance characteristics effectively change. This is not desirable, nor even allowable in most cases, because the rate gyro must measure antenna space rates accurately for the fire-control computer, it must provide a stabilizing signal proportional to the antenna rate at all times to prevent drift, and it must have the proper transfer function to provide track loop stability. This may be particularly serious for systems using guns because the large random space rates induced in the antenna during periods of gun fire cause rate gyro saturation. Consequently, it is necessary to provide rate gyros with linear rate measuring ranges far in excess of those needed to measure the aircraft space rates for the fire-control computer. Unfortunately, a large, linear measuring range reduces the accuracy of the gyro in the important low rate region. However, with missiles for armament and jet aircraft as an antenna platform, this problem is not serious.

In recent years, HIG gyros,³⁸ fluid damped and hermetically sealed, have been commonly used in antenna stabilizing loops. They are extremely

³⁷This neglects the effects of actuator stiction which can make the loop nonlinear and produce a small error.

³⁸Hermetic Integrating Gyros developed at the Massachusetts Institute of Technology.

accurate, with low drift rates, but they are expensive and require temperature-compensating circuits to maintain the preset value of damping torque. Generally, this accuracy is not needed to provide antenna stabilization, but is needed for the fire-control computer which uses the gyro signals to provide accurate information about the line-of-sight motion.

Aircraft Gyro Stabilization Loops. Another of the many methods of providing antenna stabilization is shown in Fig. 8-37c. Because the aircraft with a fire control system and tracking antenna frequently has autopilot gyroscopes, it is possible to use them to provide the antennas with stabilization signals. This reduces the total number of rate gyros needed in the system, but other components must be added as shown in the simplified figure.³⁹

The other components are tachometer and position indicators on the azimuth and elevation antenna actuators and a converter to change the space rate signals from three aircraft coordinates to two antenna coordinates. Although the converter is not complex, the conversions must be made accurately, and the three aircraft gyros should be mounted near the antenna base to prevent discrepancies from occurring in the rate measurements which are different in various parts of the aircraft owing to structural flexing.

From this discussion it is evident that several methods of providing antenna stabilization are available. Although the integrating gyro may have slight advantages over the other systems, the ultimate choice of a system configuration will depend upon the required accuracy, the allowable expense, and the permitted complexity of the application.

8-32 ACCURACY REQUIREMENTS ON THE ANGLE TRACK STABILIZATION LOOP

The stabilization accuracy requirements for the angle track stabilization loop are determined from the basic functions of the stabilization loops (Paragraphs 8-25 and 8-30) and from different criteria depending on whether a pilot or autopilot is used, a lead collision course or pursuit course is being flown, or a ballistic or target-seeking missile is being fired. The significance of these factors is discussed in the following paragraphs.

Stabilization Accuracy Required to Reduce Aircraft Rate Signal Errors. The magnitude of the error in the antenna rate signal caused by own-ship's motion must be reduced to an acceptable value. This value depends on the nature of the system. If an autopilot is used, it must be

³⁹Note that there are three aircraft gyros which do not measure the desired antenna rates in azimuth and elevation. Therefore the information in three channels must be converted to two channels using the relative rates and angles of the antenna with respect to aircraft in two dimensions. For simplicity the figure depicts a one-channel conversion.

defined from a specified system error, the allowable miss distance, or hit probability density.⁴⁰ Actually, for an autopilot system, the criterion involving the magnitude of the rate error due to aircraft motion may be less important than a criterion concerning the amount of attenuation needed from the stabilization loop to prevent system instability through the coupling between the rate commands, autopilot, aircraft, and antenna.

In the case of manual control of the aircraft flight path, the specification of the allowable stabilization error is governed by the following basic observations.

(1) The stabilization error may be considered as a random error in the measurement of angular rate. Thus the low-frequency component of the stabilization error must be compatible with the specification for angular rate measurement accuracy to avoid inaccuracies in the fire-control computation.

(2) The high-frequency components of the stabilization error (greater than 1–2 rad/sec) do not affect the fire-control problem directly because the aircraft heading cannot change this rapidly. However, they do increase the apparent amount of noise on the steering indication, and this does degrade the pilot's ability to fly an accurate course (see Paragraph 12-7).⁴¹ This degradation is proportional to the rms contribution of stabilization error to the total apparent noise appearing on the pilot's indicator.

An example using the basic AI radar problem presented in Chapter 2 is informative in illustrating how these principles might be applied. The applicable specification for the azimuth and elevation channels are (see Paragraph 2-27):

Allowable random error in rate measurement. 0.11°/sec rms
 Allowable magnitude of indicator noise 1° rms
 Allowable filtering. 0.5 second.

For purposes of analysis, we will assume that the contributions of the stabilization error should be limited to the following:

$$\left[\begin{array}{l} \text{Low frequency rate measurement} \\ \text{error due to stabilization errors} \end{array} \right] K_L \leq 0.03^\circ/\text{sec rms}$$

$$\left[\begin{array}{l} \text{Indicator noise due to high} \\ \text{frequency stabilization errors} \end{array} \right] K_L \leq 1.25^\circ/\text{sec rms}$$

⁴⁰For a detailed discussion of this problem see "Control System Optimization to Achieve Maximum Hit or Accuracy Probability Density" by G. S. Axelby, Wescon Record of the *IRE*, 1957.

⁴¹The fact that the stabilization errors are actually correlated with the pilot-induced motions does not seem to be important (as it is for autopilot applications where the correlation results in degradation of system stability). Thus for analysis of manually flown systems, the rms stabilization error must be combined with the rms noise errors from other sources to produce an equivalent noise error.

The problem now is to compute the stabilization loop attenuation K_S as a function of frequency needed to achieve these levels of performance. Table 8-3 will be used to provide aircraft input data.

For the low-frequency (less than 2 rad) inputs

$$K_S \leq \frac{K_L}{0.7\dot{A}_{TL}(\omega)} \quad (8-41)$$

where $0.7\dot{A}_{TL}$ is the rms value of the sinusoidal inputs from Table 8-3, for example, at $\omega = 1.04$, $\dot{A}_{TL} = 35/2 = 17.5^\circ/\text{sec}$ and $K_S \leq 0.00245$. However, for the high-frequency inputs, the effects of filtering and error sensitivity as a function of angular rate must be taken into account to ascertain the amount of stabilization attenuation needed. The basic expression may be written

$$K_{S(\omega)} \leq \frac{K_L}{0.7\dot{A}_{TL}(\omega)(\partial\epsilon_{Hc}/\partial\dot{A}_{TL})G_f} \quad (8-42)$$

where $\partial\epsilon_{Hc}/\partial\dot{A}_{TL}$ = sensitivity of computed error signal to angular rate inputs

G_f = rate noise filter characteristic.

From Table 2-3 the value of the angular rate sensitivity factor at the time of firing on a head-on course is 14.2.

If a 0.5-second filter is used, all sinusoidal signals above 3 rad/sec passing through it are attenuated by a factor ω_f/ω_s , where ω_s is any signal frequency and ω_f is the filter corner frequency equal to 2 rad/sec.

Therefore, if K_L is assumed to be equal to 1.25° rms for sinusoidal frequencies and G_f equal to ω_f/ω_D for disturbance frequencies greater than 3 rad/sec,

$$K_S = \frac{1.25}{(0.7)(\dot{A}_{TL})(14.2)(\omega_f/\omega_D)} \leq \frac{0.125}{(\omega_f/\omega_D)\dot{A}_{TL}} \quad (8-43)$$

However, for sinusoidal motion $A_{TL} = \dot{A}_{TL}\omega_D$ and

$$K_S \leq \frac{0.125}{\omega_f A_{TL}} = \frac{0.0625}{A_{TL}} \quad (8-44)$$

Using the values for A_{TL} in Table 8-3, the desired K_S is given in the following table as a function of frequency.

A plot of K_S is shown in Fig. 8-39. This is the minimum attenuation that must be provided by the stabilization loop if rate signal errors caused by aircraft motion are to be maintained below acceptable levels on the pilot's indicator. The isolation required for the low-frequency inputs is also indicated. The design of the stabilization loop characteristics to achieve this attenuation is described in Paragraph 8-33.

TABLE 8-4 K_S AS A FUNCTION OF FREQUENCY

A_{TL}	ω_D	K_S
9.55/2	3.14	0.013
6.04/2	3.31	0.021
4.77/2	4.19	0.026
4.48/2	4.92	0.028
0.497/2	10.5	0.25
0.715/2	10.5	0.174

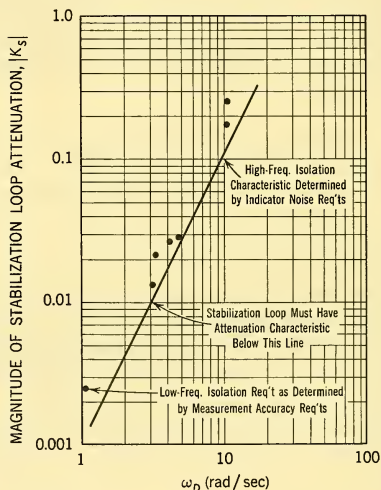


FIG. 8-39 Desired Stabilization Loop Attenuation.

The magnitude of the error in the antenna rate signal may be specified in other ways. For example, it may be determined as the allowable error in an input to the course computer from a study of overall system error specification. In this case the total rate error may be specified where the total rate error includes tracking loop inaccuracy as well as unwanted aircraft rates due to aircraft motion. The allowable total error depends on the overall accuracy of the system and it may vary from about 1 mil/sec to 4 mils/sec. In Chapter 9, where tracking loop inaccuracies are considered (Fig. 9-5) it is shown that for a total allowable random error of $0.15^\circ/\text{sec}$ (2.62 mils/sec/channel), the rms error component due to the aircraft motion could be large enough to yield the same stabilization attenuation characteristic that is shown in Fig. 8-39.

Stabilization Accuracy Required to Reduce Position Errors Due to Aircraft Motion. The radar antenna must not deviate from its desired position within a specified angle incident to aircraft motion even if the radar signal should not be present for a short time because of temporary loss of target return. In discussing the accuracy required to accomplish this, it will be necessary to discuss the position accuracy required and the temporary loss of target return which is often referred to as *signal fading*.

The position accuracy needed again depends on the type of fire control being considered. In some missile systems where a proportional navigation course is flown, only antenna rates of the missile seeker are needed to compute the proper course; angular information is not used. Therefore, the antenna angle need not be maintained very accurately. It is only necessary to keep the antenna positioned toward the target within the antenna beamwidth, which may vary from 1° to 15° depending on the antenna size and radiating frequency. Actually, the stabilization accuracy needed to maintain antenna position is an order of magnitude less than that needed to provide the necessary attenuation of antenna motion in the rate signal; and therefore it is usually not used as a criterion for stabilization loop design for this type of application.

In many fire-control systems, however, the antenna position is used to compute the course of the aircraft. From a system analysis, analytical or simulated, the random and bias errors in position may be specified. These values may range between 2 mils and 10 mils as described in Paragraph 2-27 for a typical lead collision course. The total random error should be less than 2.62 mils (0.15°) rms and the bias error should be less than 9.15 rms (0.525°).

Tracking loop errors contribute to the total position error, but as will be shown in Chapter 9, where the tracking loop errors are considered, the position error component due to platform motion does not require the attenuation needed to reduce the corresponding rate errors to the specified values. Usually, the allowed error in antenna position may be greater than that allowed in the rate signal because the rate signal, and its errors, are multiplied by a large multiplying constant, the error signal to angular rate sensitivity factor discussed in Chapter 2. Therefore, the rate error specification is usually more severe, and the required stabilization accuracy may be determined from it.

At long ranges the target may be detected and tracking may begin, but the radar return signal may become weak shortly thereafter and even fade away completely as the target changes its aspect angle with respect to the interceptor.⁴² After two or three seconds, if the target signal does not reappear, the antenna search mode is reactivated. However, during the

⁴²The change in magnitude of the reflected radar signal is discussed in Paragraph 4-3.

signal fading period, it is necessary to prevent the antenna from being moved in space away from the target direction by own-ship's motion. Here the stabilization loop plays a major role in attenuating antenna space motion caused by aircraft motion. Usually, the attenuation needed is not extremely large, because when the radar error signal (which is also the antenna rate command) fades, it is replaced by the thermal noise of the radar receiver. As is shown in Fig. 8-40, this noise causes the antenna in a

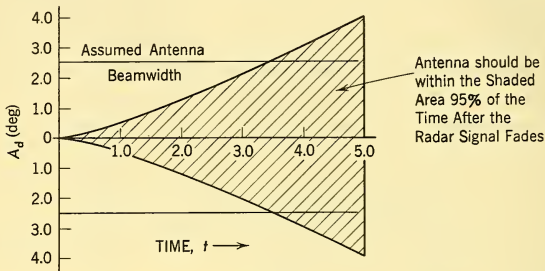


FIG. 8-40 Antenna Drift During Radar Signal Fade.

$$A_d \approx \frac{0.214}{K} \sqrt{\tau} \omega_c^2 t^{3/2} \left[1 - \frac{3}{32} \frac{\omega_c}{I} t \right]$$

where:

K is per cent modulation/degree tracking error
 τ is time between radar pulses
 ω_c is track loop bandwidth, rad/sec
 I is amount of integration in track loop
 t is time in seconds
 A_d is antenna drift angle in degrees

Typical Values
 50%/deg
 1×10^{-3} sec
 6 rad/sec
 $> 10 : 1$
 $0 < t < 5$
 $0 < 15^\circ$

typical system to drift away from the desired position by as much as 2.0° in 3 seconds. Thus, even if the stabilization loop operated perfectly, the noise commands would probably drive the antenna off target.

Actually, the positional accuracy of the antenna cannot be maintained in space without a radar signal to provide a space reference. Sufficient accuracy for fire-control computation cannot be obtained, and it is only necessary to prevent the antenna from being displaced from the target sight line by more than the antenna beamwidth, which may be between 1° and 15° , to ensure that tracking will be resumed when the target signal strength increases again. Therefore, for the system of Fig. 8-40 it would be well to limit the duration of the fading period to a time less than 3 seconds. If the system is not returned to search at this time, the antenna would probably drift from the target sight line and prevent resumption of tracking. An

alternative to returning to the search mode would be to replace the average level of the stabilization loop command, ordinarily the radar error signal, with a fixed d-c voltage to eliminate the noise. Unfortunately, this requires extra mechanization which may not be worth the additional space and complexity. However, during the fading period, it is necessary to provide sufficient antenna space stabilization, with the stabilization control loop, to keep the positional deviation incident to aircraft disturbance motion an order of magnitude lower than the maximum allowable deviation. Therefore, for the illustrative case depicted by Fig. 8-40, the peak positional error incident to aircraft motion should be less than 0.125° .⁴³ Although this is not as much attenuation as is required to reduce the rate signal errors, it may not be so in all cases. Usually, it is necessary to check the amount of stabilization needed for each type of error to determine which will require the greatest amount of attenuation.

Stabilization Accuracy Required to Prevent System Instability.

As mentioned in Paragraphs 8-25 and 8-32, antenna stabilization is needed, especially in a system with autopilot control, to prevent system instability incident to coupling between the aircraft space commands from the antenna gyros and the resulting antenna motion from the commands. Unlike the human pilot, the autopilot responds to all signals. Ordinarily, it does not learn to reject the signals generated by the coupling between the aircraft and antenna motions. Since this is a form of positive feedback (an undesired or parasitic loop) it is necessary to keep the magnitude of this loop gain less than one-half at all frequencies to ensure that instability will not occur.⁴⁴

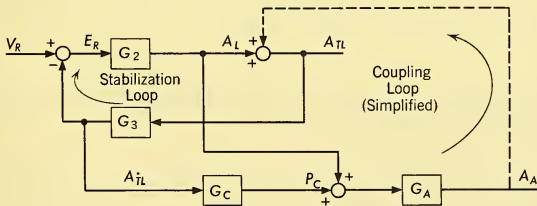


FIG. 8-41 Coupling Between the Air Frame and the Antenna.

G_2 = amplifier, actuator, antenna; G_3 = rate gyro; G_C = fire-control computer; G_A = autopilot and aircraft; $G_{CL} \approx G_3 G_C G_A$; P_C = computed projectile angle with respect to aircraft; V_R = rate command; A_{TL} = tracking line angle; A_{TL} = tracking line rate; A_A = aircraft space angle; θ_{TL} = antenna angle with respect to aircraft.

⁴³This is an arbitrary figure. It is one-twentieth of the antenna beamwidth to make it insignificant with respect to the fading error.

⁴⁴Actually, the degree of stability of any feedback loop depends on the magnitude and phase of the loop characteristic at all frequencies, but if the gain is less than one-half, a stable loop is assured regardless of the phase variation with frequency.

A simplified block diagram of this positive feedback loop is shown in Fig. 8-41. Since the gains of the autopilot, aircraft, and computer are determined by the performance specifications, only one gain element in the loop may be readily adjusted. This gain, or attenuation, from A_{GL} to A_{TL} in Fig. 8-41 is the required stabilization loop attenuation K_S . If the gain of the coupling loop is represented by $K_S G_{CL}$ (where G_{CL} is the overall characteristic of the computer, autopilot, and aircraft) as indicated in Fig. 8-41, the stabilization loop attenuation becomes

$$|K_S| \leq \frac{0.5}{|G_{CL}|} \quad (8-45)$$

Since $|G_{CL}|$ is a function of frequency, $|K_S|$ may be determined as a frequency function.

Unfortunately, however, G_{CL} is not easily determined analytically.⁴⁵ It varies as a function of antenna elevation and azimuth angles as well as the aircraft roll, pitch, and yaw angles in space because it involves the conversion of three aircraft coordinates of motion into two antenna coordinates including the dynamics of coordinated aircraft turns. Although the magnitude of $|G_{CL}|$ may be estimated or approximated for various operating conditions, the maximum value may be most easily and accurately found using analog computer simulation. Of course, the magnitude of $|K_S|$ should provide adequate attenuation for the largest magnitudes of G_{CL} that are likely to be encountered under normal operating conditions. In high-performance autopilot systems, especially missiles, the necessary stabilization loop accuracy or attenuation may be greater than that derived by the criteria for manually controlled aircraft.

8-33 DYNAMIC STABILITY REQUIREMENTS ON ANGLE TRACK STABILIZATION

In order to provide the necessary antenna isolation from aircraft motion, the stabilization loop must have an attenuation characteristic similar to that shown in Fig. 8-39. It must be stable, and it must provide a smooth, constant antenna rate for a constant voltage input. Although it might be desirable to have a transient response with negligible overshoot similar to that produced by a critically damped quadratic system, this is not always necessary or even possible in some cases. It is more important to achieve the required attenuation characteristic even with some sacrifice in stability margin.

⁴⁵This function is not as complex in a cruciform missile configuration as it is in a conventional aircraft which must roll to turn.

Consequently, the peak overshoot to a step input may be as high as 40 per cent.⁴⁶ Actually, this overshoot never appears in the antenna signals during actual operation because the input signals are never step functions and the output signals are heavily filtered. Therefore, the primary objective is to design a stabilization loop with the necessary isolation characteristics and then to obtain the minimum bandwidth possible with a stability margin that is not seriously affected by normal component tolerances, gain changes, or environmental effects. Essentially, reasonable stability margin in a minimum lead system for a stabilization loop would be: phase margin greater than 40°, gain margin greater than a factor of 4 (or 12 db). However, if precision components are used which are not greatly affected by environmental conditions, these margins may be reduced to: phase margin, 35°; gain margin, 6 db.

The attenuation characteristic K_S of the stabilization loop with respect to aircraft disturbances A_{GL} may be derived from Fig. 8-37a as

$$K_S = \frac{\dot{A}_{TL}}{\dot{A}_A} = \frac{1}{1 + G_2G_3} \quad (8-46)$$

$$|K_S| \approx \frac{1}{|G_2G_3|} \quad \text{if } G_2G_3 \gg 1.0^{47} \quad (8-47)$$

The magnitude of K_S is determined as a function of frequency from one of the criteria in Paragraph 8-32 and plotted as shown in Fig. 8-39. The magnitude of $1/G_2G_3$ must be below these points as shown, and consequently, the magnitude of G_2G_3 must be greater than the reciprocal values of K_S . The magnitude of G_2G_3 corresponding to the values of K_S in Fig. 8-39 but increased by 175 per cent is shown in Fig. 8-42.⁴⁸ This characteristic need not be extended to frequency regions below 3 rad/sec as discussed at the beginning of Paragraph 8-30. As shown in Fig. 8-42, the magnitude of G_2G_3 is 250 at 3 rad/sec to make the actual gain equal that desired. This becomes the d-c gain in the rate gyro loop shown in Fig. 8-37a. In an integrating gyro loop, shown in Fig. 8-37b, the d-c gain is infinite, but the velocity constant would be 750 sec⁻¹ as indicated in Fig. 8-42.

As in the search loop design, the other characteristics of the open-loop transfer function, for a minimum bandwidth, are determined directly from stability considerations, using the following equations relating the loop

⁴⁶This corresponds to a maximum phase margin of about 40° or a frequency response peak of about 1.6 in a minimum lead system such as that discussed in Paragraph 8-28.

⁴⁷Actually this approximation holds very well for $|G_2G_3| \geq 3$ in most cases.

⁴⁸The 175 per cent increase in gain is made to counteract changes in the stabilization loop gain. One source of gain change is the normal variation in production tolerances, and ± 20 per cent may be allowed for this. The other large gain change is in the gyro (azimuth only) because the azimuth gyro gain changes in proportion to the cosine of the antenna elevation angle when it is mounted on the antenna dish. Since the antenna may have elevation angles of 50°, this causes a 55 per cent gain change, a total of 175 per cent.

REGULATORY CIRCUITS

A_{TL}	ω_D	K_S	$ G_2 G_3 _{\ominus}$	$ G_2 G_3 _{\Delta}$
9.55/2	3.14	0.013	134	129
6.04/2	3.31	0.021	83.5	81.9
4.77/2	4.19	0.026	67.5	64.6
4.48/2	4.92	0.028	62.5	60.8
0.497/2	10.5	0.25	7.00	6.74
0.715/2	10.5	0.174	10.1	9.67

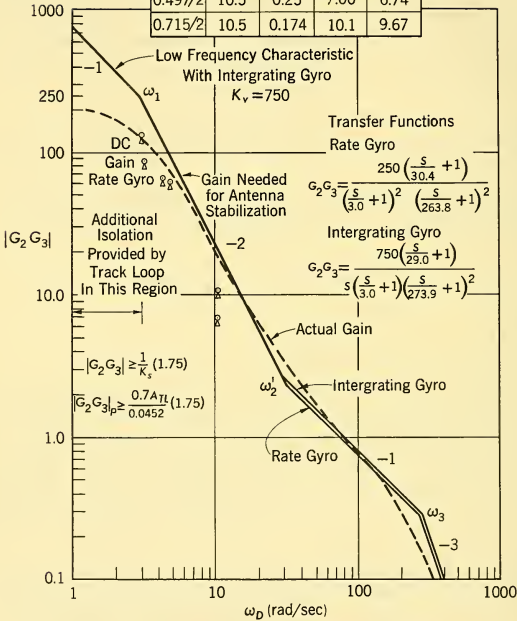


FIG. 8-42 Stabilization Loop Transfer Ratio Designs.

corner frequencies of Fig. 8-42 to the maximum phase margin and the gain of the loop.

RATE GYRO LOOP

$$\text{Gain: } K = 250 = \frac{\omega_c \omega_2}{(\omega_1)^2}$$

$$\text{Phase: } \beta_m = -\frac{180 + \phi_m}{57.3} = -\frac{\pi}{2} + \frac{2\omega_1}{\omega_m} - \frac{\omega_2}{\omega_m} - \frac{2\omega_m}{\omega_3}$$

$$\text{Maximum Phase Equations: } \frac{\partial \beta_m}{\partial \omega_m} = 0; \frac{\omega_2 - 2\omega_1}{\omega_m^2} = \frac{2}{\omega_3} \quad \frac{\omega_2 - \omega_1}{\omega_m^2} = \frac{2}{\omega_3}$$

INTEGRATING GYRO LOOP

$$K_V = 750 = \frac{\omega_c \omega_2}{\omega_1}$$

$$\beta_m = -\frac{\pi}{2} + \frac{\omega_1}{\omega_m} - \frac{\omega_2}{\omega_m} - \frac{2\omega_m}{\omega_3}$$

For each system: $\omega_m = \omega_c \cos \phi_m$, $\phi_m = 40^\circ$, $\omega_1 = 3$ rad/sec.

Solving the equations:

$\omega_2 = 30.4$ rad/sec	29.0 rad/sec
$\omega_m = 56.6$ rad/sec	59.6 rad/sec
$\omega_c = 74.0$ rad/sec (11.8 cps)	77.8 rad/sec (12.4 cps)
$\omega_3 = 263.8$ rad/sec	273.9 rad/sec

The derived corner frequencies are plotted in Fig. 8-42. This establishes the required stabilization loop transfer function. As shown in the figure, there is little practical difference in bandwidth between the stabilization stability margin and the same attenuation characteristic.

The gain margin was not used to establish the stabilization loop transfer function because it is more than adequate unless resonant frequencies in the antenna structure are near the crossover frequency of the loop. In fact, if a resonant frequency is close to ω_3 with a damping factor less than 0.1, it will reduce the gain margin below 6 db without changing the phase margin appreciably. Since the lower frequency characteristics of the loop are specified as described in previous sections, it is necessary to specify that the antenna resonant frequencies with low damping should be above ω_3 by 50 per cent or more. Thus, the lowest resonant frequency should be greater than 400 rad/sec or 64 cps.⁴⁹ If resonant frequencies below this figure should exist in the stabilization loop, the loop stability or the isolation specifications must be reduced or precision components must be used with closer stability margins.

In practice, ω_3 need not be a double corner as shown in Fig. 8-42; but it should be an equivalent where ω_3 would be the geometric mean of two corner frequencies. Actually ω_3 should be slightly higher to allow for the inevitable but unknown higher corner frequencies which are characteristic of all physical equipment and which often reduce the phase margin at the crossover frequency. However, if these frequencies are known, or if they can be estimated, they can be included easily in the equations which are used to calculate the frequencies ω_c , ω_m , ω_2 , and ω_3 . It is not always possible to do this with more rigorous methods. In fact, it should be emphasized that the approximate equations give excellent results, but only if all the corner frequencies, up to 20 times the loop bandwidth, of the actual equipment are included or estimated and used in the calculation. Actually,

⁴⁹Theoretically, resonant frequencies can be canceled by electrical networks, and this may be done in practice to increase stability margins. However, in practice, the cancellation cannot be made perfectly under all conditions, and the resulting improvement is not very large. For more details about resonant frequency effects on control loop performance see "Some Loading Effects on Servomechanism Performance," by G. S. Axelby, *Aeronautical Electronic Digest*, 1955, pp. 226-241, National Conference on Aeronautical Electronics, Dayton, Ohio, May 1955.

any control loop design, based on a mathematical analysis, should be verified with actual equipment tests, and the approximate equations may be used to give an extremely useful preliminary design even if the exact mathematical description of the actual equipment is not known.

8-34 STABILIZATION LOOP MECHANIZATION

The type and size of the antenna actuators are chosen largely from the search loop requirements; and to save space and weight, the same antenna and actuators are used in the stabilization loop. However, the remaining components in the loop are not identical; they are selected according to type of configuration desired in Fig. 8-37. Actually each of the three designs has different mechanization problems and a complete discussion of the details is beyond the scope of this text. However, a few important considerations, common to the designs, will be given.

In Fig. 8-37a two rate gyros provide electrical signals proportional to antenna space rates about the azimuth and elevation gimbals. The signal from each gyro is compared with an azimuth and elevation reference voltage from a track loop amplifier, and the resulting difference voltage (passed through appropriate compensating networks) is then used to move the antenna in its channel through a power amplifier and actuator. Usually the same power amplifiers serve the search loop and the stabilization loop, but the compensating networks are different and relays convert the search loops to stabilization loops. To prevent transients during this brief transition period, it is often necessary to short-circuit the capacitors in the compensating networks. The rate gyro is a self-contained unit obtainable in various sizes and with various degrees of accuracy. It is important that the rate gyros have low threshold voltages and a dynamic range large enough to measure the highest space rates due to vibration, antenna unbalance, and aircraft motion expected in normal operation. The gyros are mounted securely on the antenna after being carefully aligned; but although they are on the antenna during the search mode, they are not usually used to provide space rate stabilization in the search mode. As discussed in Paragraph 8-27, search stabilization must be derived from a space position reference. Therefore, the dynamic range of the gyros need not be so large as to measure the higher velocities that occur during the search loop transients. On the other hand, the gyros must be sturdy enough to withstand the extra forces that occur incident to rapid changes in direction of motion. Integrating rate gyros such as the HIG gyros with gimbals floating in highly viscous fluid provide considerable damping, and the problem is not as serious as it is with gyros for which damping is provided only by electrical feedback loops. This is a necessary consideration also in configuration (b); it is not important in configuration (c)

because the gyros are not on the antenna. In some cases it is necessary to cage the gyro gimbals during the search mode to prevent damage and to keep the gimbals near neutral position in readiness for the tracking mode of operation when it is initiated.

The location of the gyro mounting involves a problem that is often overlooked. Usually the gyros cannot be mounted directly on the actuator output shafts because of space limitations. In fact it is desirable to have the gyros mounted on the antenna dish, and the necessarily light, compact structural members between the actuator drive shaft and the gyro base have compliance, mass, and resonant frequencies which limit stabilization loop performance. Usually there is more than one resonant frequency which can affect the performance of the control loop; although electrical networks may be used to partially cancel the detrimental effects of these frequencies, they are extremely complex if more than one resonant frequency is to be attenuated. The most economical method of avoiding this difficulty is to design the antenna with the resonant frequencies ten times the loop bandwidth or to limit the bandwidth and performance of the stabilization loop.

In Fig. 8-37b an integrating gyro is used in each stabilization loop as a device for comparing the antenna rate with the commanded rate and for providing an electrical signal proportional to the difference between these rates. This signal is amplified and passed through appropriate compensating networks to drive the antenna through the power amplifiers and actuators designed for the search loops. Theoretically, any gyro could be used as an integrating gyro; practically, it is difficult to provide sufficient stability, loop gain, and damping to prevent the gyro gimbal from moving into its mechanical stops during antenna transient motions unless a large amount of viscous damping is provided at the gimbals. Therefore, HIG gyros are commonly used as integrating gyros although they may be used as rate gyros also.

Note that in the integrating gyro configuration, antenna rates are not measured directly. They are assumed to be proportional to the rate loop command which is the track loop radar error signal. As is discussed in Chapter 9, the track loop has a low bandwidth and removes high-frequency signals incident to antenna vibration from the measured rate signal. This is not a great advantage, because the rate signal is actually measured through a low pass filter in any configuration in order to remove system noise.

In Fig. 8-37a and b the amplifier is assumed to have compensating networks to convert the measured or calculated component characteristics to the derived loop transfer ratio shown in Fig. 8-42. The transfer function of the compensating elements can be derived as indicated in Fig. 8-36.

The configuration of Fig. 8-37c has been described in Paragraph 8-31 and the design details are beyond the scope of this text. However, it should be

noted that the gyros are not mounted on the antenna and that the antenna resonant frequencies do not influence the stabilization loop design because relative rate motion between the antenna and aircraft is measured with tachometers and converted to space motion in a coordinate converter. Thus, the difficulty of designing an accurate stabilization loop is transferred to the converter. The three gyros must be more accurate than those usually used for autopilots, the converter must be carefully made, and the various input signals must be phased and combined with precision resolvers to provide the necessary accuracy for the fire-control computer. As indicated in the figure, this method of providing space stabilization does not use the automatic regulation function of a feedback loop. It is primarily a balancing arrangement in which the antenna actuators are moved as directed by a computed signal. In this type of system great accuracy is required of all elements not directly enclosed by a feedback loop.

By the methods outlined in this chapter, antenna stabilization loops may be designed. Primarily regulating loops, they are not capable of directing the antenna toward a target. Automatically controlled tracking loops are used for this purpose and these loops are discussed in the next chapter.

CHAPTER 9

AUTOMATIC TRACKING CIRCUITS*

9-1 GENERAL PROBLEMS OF AUTOMATIC TRACKING

The general tracking problem has already been mentioned as one of prediction. It is desired to know not only where a target is but also where it will be in the near future. If this can be determined, the target can be intercepted and destroyed. Mathematically, this is not a particularly difficult problem. If the past and present motion of the target is known, its course can be projected into future time. Theoretically, the equations of target motion may be formulated rather easily, and they can be solved rapidly with an analogue or digital computer to determine the target position at any time. Paragraphs 1-4 and 2-7 illustrated some of the basic principles.

Practically, however, the information needed to formulate the equations must be obtained from measurements made with physical equipment at a considerable distance from the target. Unfortunately, there are no perfect measuring devices and the true signal information cannot be found exactly. Actually, it is obtained as an electrical signal veiled with noise and shrouded with the inaccuracies of physical equipment. The true signal — the desired information about the target — can be completely extracted from the noise with appropriate filtering only if infinite time is used to obtain it. Of course, this would be of no use in predicting a future target position because, in actual time, the target would have passed over the course and the real time errors would be infinite.

Therefore there must be a compromise between the length of time that the signal can be filtered and the accuracy of the real time prediction that can be obtained. As shown by Wiener,¹ there is an optimum filter for any specified signal, noise, and prediction time which yields the required signal with an error that is the minimum theoretically obtainable with perfect physical equipment.

This noise problem is much more severe if the target is an enemy target because, through countermeasure methods, the target will generate noise

*Secs. 9-1 through 9-9 are by G. S. Axelby. Secs. 9-10 through 9-14 are by C. F. White.

¹N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Devices*, John Wiley & Sons, Inc., New York, 1949. (Published in a classified report in 1942.)

and submit as much false information as possible to the measuring devices. The accuracy of the measured data can be improved over the theoretical minimum only by using more than one type of measuring device to obtain signals from the same target. By correlating the different signals, the signal strength can be enhanced without increasing noise in the same proportion. The resulting improvement in the signal-to-noise ratio permits a more accurate prediction of a target's position. Practically, such improvements are limited by the complexity, reliability, weight, and cost of the added equipment needed to obtain independent signals.

The correlation techniques may have many different forms. Various frequencies of radiation may be used to illuminate the target, not only in the microwave range but in the infrared spectrum as well. Among the many systems using a type of correlation technique are those with frequency diversity, micropulses, "quickenings," matched "frequency stretching" filters, phase and amplitude comparators, and angle, range, and velocity error comparators in pulsed doppler systems.

The specific type of information needed for tracking a target depends on several factors — the type of course being used in an attack on the target, the relative position and motion between the target and the tracking equipment, the amount of noise received, the weapons to be used against the target, the type of course being flown by the target, and the type of equipment used to obtain tracking information. Most of these factors are different each time a target is being tracked; and theoretically, different tracking information for computation and different tracking equipment should be used for every attack condition in order to obtain the most accurate information about the target. Of course it is impossible to have an infinite variety of tracking equipment; but because the range of variation in the optimization factors is not extremely large, it is possible to fabricate a system with one or two modes of operation which will be nearly optimum for the expected range of operating conditions.

The basic information required for tracking is the position of the target with respect to a space reference system as a function of time. In actuality the target is usually some distance from the tracking device, moving with respect to it, and consequently the position of the target may be described by a vector which represents the range between the tracking device and the target and is oriented along the space line of sight. Target information is obtained by measuring characteristics of this vector. Frequently, it is only the angular position of the vector that need be obtained, and in some cases, only the angular rate of the vector in space need be known. Of course, radar is a convenient means of obtaining this information because an angular deviation between the radar beam and the target sight line, the angular space error, can be obtained directly as an electrical signal; if this signal is small, the measured antenna direction is almost identical with that

of the target space vector. If the radar error signal from the target is maintained near zero while the target moves, the antenna motion is that of the target vector, and the required angular rate information about the target is obtained from measurements of the antenna motion. Although contaminated with noise, the radar error signal is a measure of the true tracking error and it is held to a minimum value by automatic angular tracking circuits. In a similar manner, the length of the vector between the target and antenna is measured with automatic range tracking circuits to an accuracy consistent with the physical equipment available, the existent noise, and the rapidity with which the target position changes.

As there are several types of radar systems that may be used to obtain space error information, there are many types of tracking systems. Fundamentally, however, the radar must furnish not only the magnitude of tracking error but its space direction as well. This may be obtained by several different methods (conical scanning, monopulse, etc.), but basically, the space direction must be used to direct mechanical actuators that move the antenna about supporting pivots or gimbals in a space direction which will reduce the error continuously and thereby direct the antenna toward the target. The number of gimbals needed to move the antenna in space is kept to a minimum to save space, to avoid excessive weight, to prevent excessive structural flexure, and to minimize the number of rotary joints in the radar waveguides.

In theory the radar error signal could be separated into appropriate components, modified to obtain a desired driving function, and applied directly to the antenna actuators. In practice the antennas must be driven by devices which amplify power, and generally power amplifiers have transfer characteristics which vary considerably with environment and operating conditions. To obtain constant antenna drive characteristics, a feedback loop is therefore designed around the antenna actuators. This may be a conventional tachometer feedback loop; or the stabilization loop discussed in Chapter 8 may provide this function and thus serve a double purpose.

The angle tracking problem is further complicated in an airborne system by the fact that the radar antenna is on a platform moving with respect to the space reference system as discussed in Chapter 8. Fortunately, the angular motion of the target vector does not have significant components as high in frequency as the platform motion does, and the automatic tracking loop may therefore have a much lower bandwidth than the stabilization loop. If this were not the case, it would be difficult to design a stabilization loop as discussed in Chapter 8 that would result in a suitable system when operating with the tracking loop. As implied earlier in this discussion and as shown later in this chapter, the tracking loop bandwidth is limited to prevent excessive transmission of noise and to provide the

optimum relationship between the tracking bias errors caused by the moving target and the random errors caused by inherent fluctuations in the radar error signal. A bandwidth too narrow would produce large bias errors in real time with small random errors, and a bandwidth too wide would produce small bias error with large random error components. The basic problem of tracking loop design is to find the optimum compromise between these conflicting considerations.

The automatic range and angle tracking loops must be endowed with memory during short intervals when the radar signal fades and target information is not available. This is usually done with relatively simple networks which provide an approximation to an integral.

Another significant problem which arises in tracking control systems is that of cross coupling between control channels, particularly in the angle track loops. Cross coupling exists in basic space geometry when vector rates are translated into various reference frames in the antenna, in the platform, or in the weapon coordinate systems. There are other sources of coupling between channels — improper separation of the radar error signal into the proper antenna gimbal coordinate, incorrect choice and orientation of rate-measuring gyros, poor mechanical design in which motion in one channel is induced by motion in another, and high impedance in common power supplies, which may be electrical or hydraulic.

Details of these space-vector tracking problems are discussed in the following paragraphs.

9-2 AUTOMATIC ANGLE TRACKING

The primary function of the angle tracking system is to derive the direction and angular rate of the line-of-sight with an accuracy needed for an adequate solution of the fire-control problem. As discussed in Paragraph 8-23, space stabilization must be provided during the tracking operation.

The angular direction to the target is determined basically by an automatic control system which directs the radar beam toward the target. The error between the antenna boresight and the target sight line is detected as a modulated signal. This signal, proportional to the angular space error, is amplified and applied to actuators which drive the antenna to reduce the angular space error. The direction and magnitude of the motion that the servo motor must provide to correct the error is usually obtained as a d-c signal by demodulating the tracking error signal derived by the radar.

As described in Paragraphs 8-23 and 8-29, angular space stabilization is provided by gyros, usually mounted on the antenna, and by the radar tracking loop itself.

9-3 EXTERNAL INPUTS: UNDESIRE AND DESIRE

Before an automatic control system can be designed, some information about the system inputs and the allowable tracking accuracy must be known. This knowledge may be used to determine what the control loop must do, and it dictates what characteristics are necessary to accomplish the desired function. The track loop inputs are discussed in this paragraph and the allowable errors are described in the next paragraph.

The undesirable inputs to the tracking loop may, in general, be classified as noise or as disturbances. Ultimately it is desired to eliminate the effects of the undesirable inputs from the useful output signals. One form of disturbance, aircraft motion, was described in Paragraph 8-24. Another form of undesirable input, noise, is described in Paragraph 4-4. If radar is used as a detector, there will be considerable noise in the system, and the system output noise increases with increasing bandwidth. Therefore it is necessary to keep the system bandwidth as low as possible without creating significant time delays in the measured information about the target's motion in space.

The desirable inputs are those which the angle tracking system must detect and follow. Essentially, the desirable inputs are the time functions which describe the space motion of the sight line between the radar antenna and the target. The expected sight line motion may be determined from a study of the various attack courses that the interceptor is to fly.

Possible target maneuvers should also be considered in the general study of course dynamics, although for preliminary considerations the target is often assumed to have a constant velocity.

Course Dynamics. Course dynamics, with respect to the angle track system, concern the angular motion of the sight line in space corresponding to all tactically feasible attacks. The maximum angular rates $\dot{\omega}_{\max}$ and angular acceleration $\ddot{\omega}_{\max}$ of the total line-of-sight motion resulting from these attack courses may be calculated from knowledge of the equations describing the course. Typical sight-line rates and accelerations with time during a lead collision course have been shown in Fig. 2-45.

Another possible mode of attack for a high-speed fighter is the lead-pursuit course wherein the aircraft is flown in a trajectory such that the correct lead angle for firing is maintained throughout the course, rather than at a single point as for the lead collision course. Typical data for lead-pursuit courses against maneuvering targets are shown in Fig. 9-1. The maximum values of angular rates and angular accelerations of the lead pursuit course sight lines are summarized in Table 9-1 along with the lead-collision data. Higher derivatives have been found to be relatively small and unimportant in establishing system specifications, and they are not included in the table. However, $\dot{\omega}_{\max}$ and $\ddot{\omega}_{\max}$ are the principal charac-

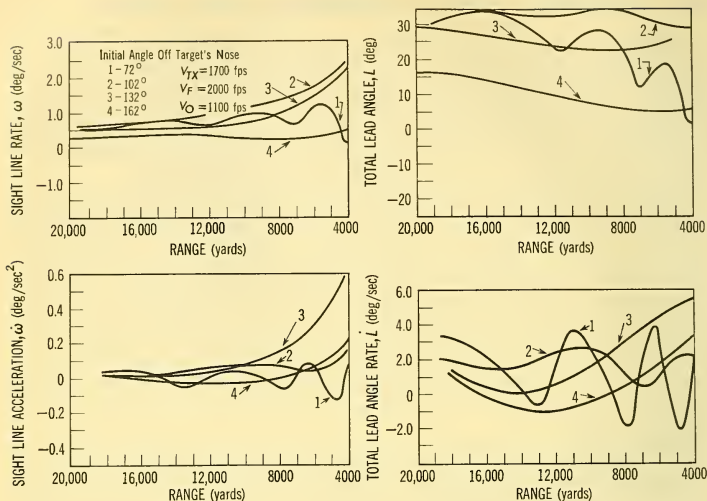


FIG. 9-1 Lead-Pursuit Course Inputs.

teristics of the motion which are used to establish angle tracking loop specifications as described in Paragraph 9-6.

TABLE 9-1 TYPICAL MAXIMUM ANGULAR RATES AND ACCELERATIONS OF THE SIGHT LINE

	ω mils/sec	$\dot{\omega}$ mils/sec ²
Lead Pursuit (Horizontal Target Maneuvers)	46	11.4
Lead Collision Course ^a	18.3	2.27

^aAt time of firing, 10 seconds to go. (Much more severe at short ranges; see Fig. 2-45.)

Fig. 9-1 also shows plots of the lead angle and lead angle rate to illustrate how the antenna must be moved with respect to the antenna platform. Space platform disturbances are not included here, but curves such as these help define the necessary antenna gimbal motion and actuator capabilities.

Figs. 9-1 and 2-45 are typical of the plots needed to define the angle track inputs, and in practice many others would be made during studies of various attack conditions to determine the most severe operating requirements.

Signal Fades. At long ranges, signal returns from a target are of extremely low power and may be detected only periodically. This variation, of random nature, is referred to as *fading*. Fading is not a desirable input,

but it is an inherent characteristic of the desired target input. At shorter ranges where the signal strength is greater, fading is less likely to occur.² Although fading is not a predictable phenomenon and there is no accurate mathematical description of it, it may persist for periods greater than 1 sec and possibly as great as 3 sec at long ranges.³ The tracking system must be capable of maintaining the antenna position approximately on the target during a fading period with an angular accuracy which will permit restoration of tracking when the target return signal becomes strong enough to be detected after the fade.

Target Information Signal. Information about the angular space error between the sight line and antenna tracking line is presented to the antenna drive circuits as sidebands on a suppressed carrier signal. In a conical scan radar, as shown in Fig. 5-13 for example, the carrier frequency is obtained from a rotary motion of the antenna beam in space. This motion is often provided by mechanically driving the antenna feedhorn with a small motor. The spinning feedhorn provides a modulated pulse return from the target when it is not on the tracking line. An unmodulated carrier frequency is also obtained from a generator on the feedhorn to provide a reference for extracting space error information from the modulated carrier. The suppressed carrier signal actually contains error information for two antenna control channels simultaneously, and two phase reference signals from the feedhorn generator are needed in a tracking circuit demodulator to separate the information required for each channel. The tolerances which apply to the carrier frequency amplitude and phase are now discussed.

The selection of the carrier (lobing) frequency is dictated by the following considerations:

(1) Natural modulating frequencies of the target (see Paragraph 4-8) should be avoided. For example, a propeller-driven aircraft might produce a 70-cps modulation on the signal return; if the lobing frequency was also 70 cps (or a harmonic of 70 cps) poor tracking performance would result.

(2) The lobing frequency cannot be higher⁴ than $\frac{1}{10}$ of the basic radar sampling frequency; otherwise stability problems will arise incident to the effective delays introduced by the sampling process. For example, if the PRF of a pulse radar is 1000 pps, the lobing frequency should not be higher than 100 cps.

²Except in a pulsed doppler radar with high pulse repetition where fading is definite and periodic because of eclipsing.

³Because of flight and attack time limitations, longer fades would probably dictate a return to the search mode of operation discussed in Chapter 8 with reference to the AGC design problem.

⁴To achieve almost complete protection from the effects of using sampled data in the usual case, a 20 : 1 relationship should be used.

(3) Mechanical stresses increase with the square of the lobing rate. Thus, to avoid excessive wear and vibration the lowest possible lobing frequency should be employed.

(4) The lower limit on lobing frequency is dictated by stability considerations.

For these reasons, the lobing frequency is usually between 30 and 100 cps; it may even be variable to provide a defense against countermeasures. As previously discussed, the design of the AGC loop, as well as the track loop, is influenced by the choice of the lobing frequency.

The carrier frequency and a band of frequencies 10 cps above and below it must be transmitted by the receiver over the full range of input power, or possible lobing frequency variation, without an attenuation greater than 1 db or with a phase shift greater than 10° . This is necessary because the target information needed to provide accurate tracking line motion may have significant frequencies up to 1 cps, and adequate stability and accuracy in each control channel can be obtained in practice only if frequencies up to 10 times this significant frequency are provided within the specified tolerances. This is discussed in more detail in Paragraph 9-8.

The receiver transmits the space error information in terms of a voltage with respect to milliradians of target angle from the boresight line. This scale factor should be controlled within ± 1 db as stated above to prevent amplitude modulation and false error information, although its nominal design value is not critical. For practical realization of the angle track amplifiers, however, it would be desirable to have a scale factor compatible with the practical threshold levels and amplification limits in the control system amplifiers. In the actual mechanization, it is necessary to distribute the tracking loop gain appropriately between a-c amplifiers, d-c amplifiers, and intervening compensating networks. Proper location of the compensating networks will do much to lessen saturation problems during transient operation.

Glint Noise Inputs. The magnitude of the glint noise at the antenna drive angle track input will be essentially the same as that appearing at the receiver input. It is unmodified by the receiver because glint noise is actually angular noise which affects the magnitude and phase angle of the carrier frequency. Since the carrier frequency contains the true angular error information as well as glint noise, it must be transmitted through the receiver unmodified. Only after the angular information is extracted from the carrier frequency can the glint noise be modified, to a limited extent, without seriously affecting the target information. In Chapter 4, the anticipated target glint noise is described. It is assumed here to have a spectral density of the form

$$\sigma_o = \sigma_i \sqrt{\frac{2}{\omega_o} \left(1 + \frac{\omega^2}{\omega_o^2} \right)^{-1/2}} \quad (9-1)$$

where $\sigma_i \approx 0.15 \frac{L}{R}$; L is target length subtended to the radar; R is the range to the target; ω_o is the noise "bandwidth"; $1 \leq \omega_o \leq 15$.

Amplitude Noise Inputs. The magnitude of the pulse-to-pulse amplitude noise appearing at the antenna track input will be reduced considerably by the AGC in the receiver if it has a bandwidth 10 cps or more.⁵ At relatively short ranges, it is much smaller in magnitude than the glint noise. However, the amplitude noise may still appear as an angle track input signal after it passes through the AGC circuits and receiver filter. In the angle track loop it is further attenuated by the angle track demodulators which transmit narrow frequency bands of noise only at odd harmonics of the carrier frequency.⁶ However, there is a possibility that noise could cause saturation of the track loop input amplifier. Therefore, the dynamic range of the amplifier must be made large enough to prevent saturation by noise most of the time even with steady state bias errors.

9-4 REQUIREMENTS IN ANGLE TRACKING ACCURACY

Generally, the antenna angular position and antenna space velocity must be measured for use in a weapons control computer. Position is measured with resolvers or potentiometers and rates are usually obtained from gyros on the antennas. As discussed in Paragraph 2-27 the errors in measurement have two essential components — bias errors and random errors. To provide maximum system accuracy, all units including the tracking loop must be designed to provide a proper relationship between the bias and random errors that are generated in the physical equipment.

Maximum Dynamic Lag or Bias Errors. The output of the tracking system, the antenna tracking line, must follow the sight line in space with a maximum permissible space error of 0.525° total as specified in Paragraph 2-27, Table 2-4, assuming that a particular attack, a lead collision course, is being flown. The errors may be different for other types of attacks. This is referred to as a bias error since it is predictable from a given target flight path, attack course, and tracking loop configuration. This space error of 0.525° is equivalent to 0.372° per channel. In a similar manner, the line-of-sight bias rate error is specified as less than $0.2^\circ/\text{sec}$ total or $0.14^\circ/\text{sec}/\text{deg}/\text{sec}/\text{channel}$. These particular errors and the random errors below will be used in an example to illustrate the relationship between the errors and the design of angle track loops.

⁵See Paragraph 8-18. Reference: J. H. Dunn and D. D. Howard, "The Effects of Automatic Gain Performance on the Tracking Accuracy of Monopulse Radar Systems," *Proc. IRE*, **47**, 430-435 (1959).

⁶Assuming that an averaging type demodulator rather than a keyed demodulator is used, because of the noisy signal.

Maximum Random Tracking Errors. The random errors in the tracking loop should be less than 0.15° rms total or 0.105° rms per channel in the position measurements, and $0.15/\text{sec}$ rms or $0.105^\circ/\text{sec}$ rms in the rate measurements (see Table 2-4).

The random errors are produced not only from radar noise but also from inaccuracies in the measuring devices and from random platform space motion. These errors and the bias errors are described in more detail in Paragraph 9-6.

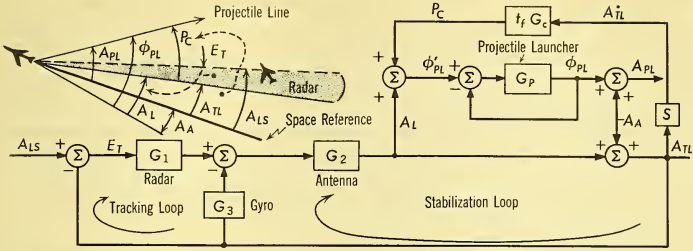
9-5 ANGLE TRACKING SYSTEM ORGANIZATION

For an adequate solution of the fire-control problem, measurements must be maintained within limits discussed in Paragraph 9-4. As implied, there are many factors affecting the accuracy of the system: (a) target signal fading, (b) interceptor-controlled maneuvers, (c) interceptor gust disturbances, (d) target noise inputs, (e) interceptor induced vibrations (due to buffeting, structural resonances, etc.), and (f) radar receiver noise.

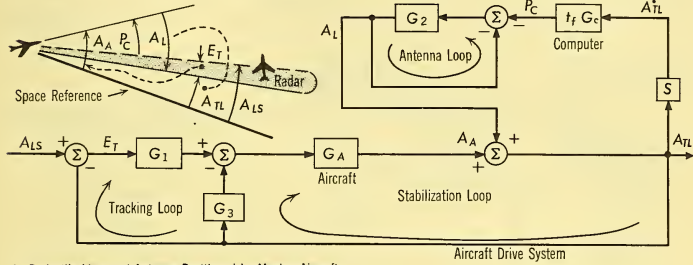
To minimize the effects of these disturbances, some form of feedback control must be used. There are a variety of basic control systems that can be devised. A few of these are indicated in Fig. 9-2. The figure shows representative block diagrams of three types of systems, and it has corresponding diagrams which define the angular variables that are being measured or commanded in space.

The first part of the figure, (a), represents a system where the projectile launching mechanism is on a gimbal system, separate from that of the antenna in the aircraft, and it moves the projectiles into proper launching position according to commands computed from the tracking antenna position and rate information. A tail turret is often mechanized in this manner. In this case the projectile launching mechanism is a cannon and, as shown in the figure, it is commanded through a position feedback loop to reduce the effect of actuator and amplifier variations. The turret and the antenna are subjected to the same aircraft motion disturbances, and in addition there is usually a structural coupling between the tracking antenna and the turret which can actually produce system instability as well as induce disturbances into the antenna tracking loop.

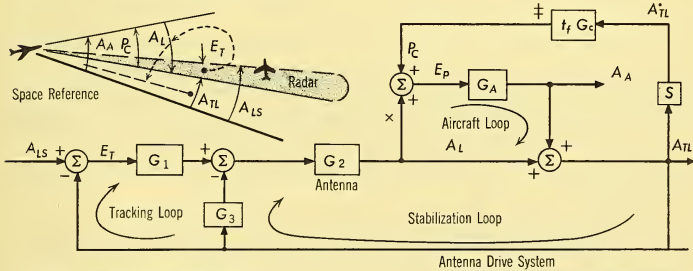
The second part of Fig. 9-2, (b), illustrates a system in which the aircraft is controlled directly from the tracking radar error signal. In this way the antenna is moved directly by the aircraft and also by a position command, the lead angle with respect to the aircraft, computed from measured antenna space rates. However, it is the aircraft which is continuously directed by the radar tracking error signal toward the future position of the target. The antenna lead angle is similar to a calculated bias signal in the loop which prevents the aircraft from flying directly toward the present target position as the tracking error is reduced.



(a) Projectile Launcher With Separate Drive (Turret)



(b) Projectile Line and Antenna Positioned by Moving Aircraft
(Fixed Guns or Missiles $A_A = A_{PL}$)



(c) Projectile Line Positioned by Moving Aircraft
(Fixed Guns or Missiles)

Antenna Drive System

× = Eliminated in Missile Seeker Systems

‡ = Modified for Missile Seeker Systems

FIG. 9-2 Forms of Antenna Drive Systems.

t_f = projectile flight time
 E_T = tracking error
 E_p = pilot error
 A_A = angle aircraft in space
 P_c = computed projectile angle

A_{TL} = antenna tracking line
 A_{LS} = angle of line of sight
 A_{PL} = angle of projectile line
 θ_{PL} = angle of projectile line from aircraft
 A_L = angle of tracking line from aircraft (lead angle)

The last part of Fig. 9-2, (c), represents the system discussed most often in this text. In this system, the antenna is automatically directed toward the target by the radar error signal, and the aircraft is flown toward the future target position with commands computed from the measured antenna position and velocity.

The design of each is developed from primary system error requirements and knowledge of the inputs such as those discussed in Paragraph 9-3.

To illustrate the design considerations, a system similar to Fig. 9-2(c) with a conical scan radar and rate gyro feedback will be discussed.⁷

Control System Diagram. It is necessary to obtain the required functions of antenna angular position and sight line space rates in two axes since the antenna motion is to be controlled mechanically in two orthogonal axes. This necessitates two separate control channels. Fig. 9-3 is a block

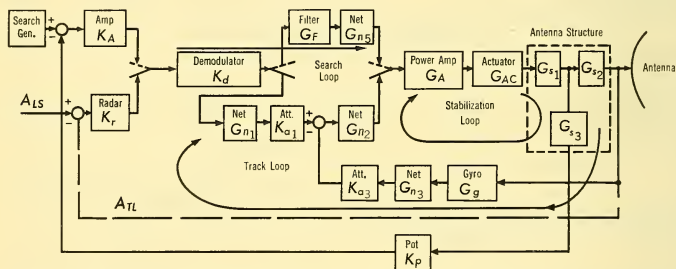


FIG. 9-3 General Block Diagram of Angle Tracking Configuration, One-Channel.

diagram of a typical control system for one channel. The two channels have duplicate components with the possible exceptions of K_r and K_d , the radar and demodulator⁸, which are usually common to the two channels. Because of the similarity between the two channels, only one channel will be described.

As discussed previously, the radar signal contains space error information for both channels; it may be obtained from a rotating antenna feedhorn. It appears as suppressed carrier modulation at the output of K_r , the radar receiver. The rest of the diagram can be subdivided into three loops: (a) search loop, (b) stabilization loop with a rate gyro, and (c) track loop.

The search loop has been described in Paragraph 8-26, but is included here to show that some of the track circuits not in use during the search

⁷Actually, many of the design principles could be applied to any of the configurations shown in Fig. 9-2.

⁸In many systems a different demodulator is used in the search loop. A common demodulator is shown in this figure.

mode may be used as search loop components. The stabilization, or inner loop, primarily isolates the antenna from aircraft motion as described in Paragraph 8-30; it provides measurements of the tracking line rates in the presence of external space disturbances. The angle track loop includes the stabilization loop as well as the radar equipment; it provides the magnitude and direction of the space error between the sight line and tracking line E_T in Fig. 9-2.

Simplified Angle Track Loop. As noted above, the two channels are nearly identical.⁹ Therefore, for purposes of analysis, all of the sight line motion may be referred to only one channel, rather than be distributed between the two, to establish the necessary loop specifications. The form

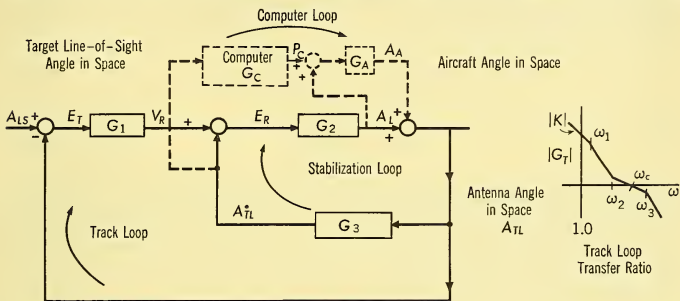


FIG. 9-4 Simplified Block Diagram of One Channel of Angle Tracking and Stabilization Loops.

G_1 = radar receiver, amplifier, bandwidth shaping; G_2 = power amplifier, actuator, antenna structure; G_3 = rate gyro; E_T = radar angle error; E_R = rate loop error; $G_C G_A$ = computer, pilot, aircraft (or missile and autopilot only) (not easily described mathematically); V_R = rate command

TRANSFER FUNCTIONS

$$G_1 = \frac{K_1 \left(1 + \frac{S}{\omega_2}\right)}{\left(1 + \frac{S}{\omega_3}\right) \left(1 + \frac{S}{\omega_3}\right)} \quad \text{NOTE:} \quad \frac{G_2}{1 + G_2 G_3} \approx \frac{1}{K_3 S}$$

$$G_T = \frac{A_{TL}}{E_T} = \frac{K_1 \left(1 + \frac{S}{\omega_2}\right)}{\left(1 + \frac{S}{\omega_1}\right) \left(1 + \frac{S}{\omega_3}\right)} \times \frac{1}{K_3 S} = \frac{K \left(1 + \frac{S}{\omega_2}\right)}{S \left(1 + \frac{S}{\omega_1}\right) \left(1 + \frac{S}{\omega_3}\right)}, \quad K = K_1$$

⁹Although both channels may not be required to have the same performance or the same component design for a given tracking problem, it is desirable to consider that the same performance may be required of either channel because of the wide variety of attack conditions that may occur.

of the stabilization and angle track loops in Fig. 9-3 can also be simplified as an aid to the analysis and understanding of the tracking function. This simplified diagram is shown in Fig. 9-4.

The block G_1 comprises the radar receiver (including the automatic range tracking circuit discussed in Paragraph 9-11 and the common track demodulator, voltage amplifiers, and bandwidth shaping circuits). The block G_2 contains a power amplifier, actuating motor, and antenna structure with associated mechanical resonances. The block G_3 represents the rate gyro and possibly a network and an attenuator as shown in Fig. 9-3.

The transfer function of the open stabilization loop (assuming the angle track loop is not closed) will be equal to the product of G_2G_3 . This was described in Paragraph 8-23. The transfer ratio G_T for the angle tracking loop will be equal to the product of G_1 and the closed-loop transfer function of the stabilization loop,

$$\frac{A_{TL}}{V_R} = \frac{G_2}{1 + G_2G_3}. \quad (9-2)$$

Thus,
$$G_T = \frac{G_1G_2}{1 + G_2G_3}. \quad (9-3)$$

In Fig. 9-4 the transfer functions for the simplified blocks G_1 , G_2 , and G_3 are shown in Laplace transform notation. Note that numerical values are dependent upon the design specifications which are themselves determined from the target and interceptor information available, and that the mechanized system transfer functions may differ in numerical value in different systems. Actually, the functions become much more complex if high-frequency terms are included. However, the functions will have the basic form shown in the figure, and this may be readily synthesized, especially at low frequencies, by networks which compensate for undesirable corner frequencies inherent in the physical components used in the mechanization.

Consideration in Specifying Control Loops. The control system specifications may be determined from the output required, the type of inputs that can be anticipated, and the necessary accuracy. Two characteristics generally used to specify control-loop performance are frequency responses and time responses. To obtain satisfactory results, both will be used to obtain the control system characteristics in the next paragraph.

In general, the gain at specific frequencies and the bandwidth (unity crossover) of the control-loop response (unity gain of the loop transfer ratio) will be used. By specifying the gain at various frequencies, the errors in the system will be maintained within the desired limits; by specifying the bandwidth and the gain constants or error constants, the necessary speed of response to aperiodic inputs will be provided, and the magnitude

of the peak transient errors will be maintained within acceptable levels. The bandwidth specification will also affect the magnitude of the noise in the output of the control loops.

9-6 TRACKING LOOP DESIGN

Solutions to most engineering problems are obtained as a result of compromising between ideal but conflicting characteristics. The angle track loop design problem is no exception. It would be desirable to have a fast, accurate loop to keep the antenna pointed at the target during any target or antenna platform motion in space. On the other hand, it is necessary to have a relatively slow tracking system which will not respond to the noise inherent in the radar receiver. A perfect solution to the problem cannot be obtained; compromises are necessary: the fast loop response needed to reduce antenna platform motion is obtained with a stabilization loop, as discussed in Chapter 8, and a slow tracking loop is designed to provide noise filtering without causing large dynamic errors while tracking a moving target.

The general configuration of the track loop was discussed and shown in Fig. 9-4. The bandwidth may be expressed in several ways,¹⁰ but in this paragraph, it is the frequency ω_c in Fig. 9-4, where the open loop gain magnitude is unity. Usually this is the same as the asymptotic gain plot at unity gain. The velocity constant K_v reduces steady-state errors caused by constant velocity inputs. Theoretically, it may be as large as desired, but practically, it is usually between 100 and 500 sec^{-1} because higher gains cause amplifier saturation during transients, and the resulting time response is not satisfactory unless nonlinear networks are used.

The corner frequency ω_1 is used to reduce the high gain rapidly at higher frequencies to realize a relatively low bandwidth. This is sometimes a double corner to reduce the gain more rapidly as the frequency increases, but the additional complexity is not often warranted.

The lead corner frequency ω_2 is needed to provide stability or damping for the tracking loop. The frequency is usually one-half to one-fifth of ω_c . If it is too low, the bandwidth will be excessive for the same low-frequency gain; and although there will be little overshoot in the response to a step input, the loop will not reach its steady-state operating levels very rapidly. On the other hand, as ω_2 approaches ω_c , the overshoot to a step input will become undesirably large. An optimum choice for ω_2 is discussed in the next paragraph.

¹⁰Other ways of expressing bandwidth: (a) The frequency where the phase angle of the closed loop is 60° , (b) the frequency where the maximum value of the closed-loop response occurs, (c) the frequency where the closed-loop response returns to unity, and (d) the frequency where the closed-loop response is -3 db.

The corner frequency ω_3 is used to attenuate high frequency noise and resonances. It is usually about three times greater than ω_c .¹¹

It is the choice and physical realization of K_v , ω , ω_1 , ω_2 , and ω_3 that constitute the tracking loop design and the consequent compromises. In particular, K_v , ω_1 , and ω_2 determine the system tracking errors, and as mentioned earlier, the errors which result from the design compromises may be divided into two categories: random and bias errors. The relationship between these errors and the track loop design parameters are discussed in the next paragraphs.

9-7 ANGLE TRACKING LOOP RATE ERRORS

Usually the most important signals to be measured in the track loop are the antenna rate signals, and the errors in these signals will be considered first.

Random Errors. The components of the random errors in the rate signal and the methods of computing them are shown in Fig. 9-5.

The principal random errors are caused (a) by gyro inaccuracies e_g which are assumed to be proportional to the rate being measured, (b) by the radar noise e_n transmitted through the tracking loop to the gyro output, and (c) by the antenna rates e_s caused by the platform motion which is attenuated by the stabilization loop as discussed in Chapter 8.

The gyro error is a function of the gyro construction. For a particular gyro, this error is fixed; it cannot be influenced by the tracking loop design. The noise in the filtered rate signal e_n is calculated as shown in Fig. 9-5. It is a function of the track loop parameters ω_2 and ω_c ,¹² the rate filter corner ω_F ,¹³ and the noise power density parameters ω_i and ω_o .

The random rate error caused by platform motion e_s was discussed in Chapter 8. Because it is the noise component most easily attenuated and controlled by loop design, its allowable design value is derived from the specified total rate error and the sums of the squares of the other two principal errors as shown in the figure.

Although a tracking loop bandwidth much lower than the 6.28 rad/sec (1 cps) shown could be tolerated without making the steady-state bias unduly large, it is usually not made much less because the system errors caused by sudden input changes would be excessive for relatively long periods of time.

¹¹A definite ratio can be determined by using approximate design equations relating gain, phase, and frequency as shown in Chapter 8 where the search and stabilization loop characteristics are derived.

¹² K_v is not an important parameter in these calculations if it is much larger than ω_c .

¹³A simple lag corner is assumed here, but more complex filters may be used.

FIG. 9-5 Antenna Rate Error Calculations.

RATE ERROR COMPONENTS

Error	Symbol	Approx. Formula	Comments
Gyro (random)	e_g	$K\dot{\theta}$	$0.002 < K < 0.02$
Rate noise (random)	e_n	$\sigma_i \omega_c \left[\frac{1(K^2 + Kr + r)}{K(K + Kr + r^2)} \right]^{1/2}$ $\left[\left(1 + \frac{\omega_c}{\omega_F} \right) \left(1 + \frac{\omega_c}{\omega_F} \right)^{1/2} \right]$	ω_F = rate filter corner frequency $\sigma_i \approx 0.15 \frac{L(180)}{R(\pi)}$; deg rms L = target length (yd) R = target range (yd) ω_0 = effective noise bandwidth of noise power density ϕ_{nn} $\Phi_{nn} \approx \frac{2\omega_0 \sigma_i^2}{\omega_0^2 - S^2}$ $r = \frac{\omega_c}{\omega_0}$
Stabilization (random)	e_s	$\frac{0.7 A_T L}{G_2 G_3} (\)$ rms	Described in Chapter 8

Random error components are independent. Therefore assume total random error; $\sigma_T^2 = e_g^2 + e_s^2 + e_n^2$ Specified $\sigma_T \leq 0.105^\circ/\text{sec}/\text{channel}$

Track loop (bias)	e_b	$\frac{\ddot{\theta}}{K_v} + \frac{\ddot{\theta}^a}{K_a}$	Specified $e_b \leq 0.14^\circ/\text{sec}$ $\omega_c I = K_v$, track loop velocity constant ω_c = track loop bandwidth I = track loop integration $\frac{\omega_c^2}{K} = K_a$, track loop acceleration constant $k = \frac{\omega_c}{\omega_2}$ ω_2 = track loop lead corner frequency
-------------------	-------	---	---

CALCULATIONS

Assumed Values

- | | | |
|--|---|---|
| $I = 20$ | $\dot{\theta} = 1.05^\circ/\text{sec}$ | } course conditions of FIG. 2-44.
Time to go: 10 sec |
| $\omega_c = 6.28 \text{ rad/sec}$ | $\ddot{\theta} = 0.13^\circ/\text{sec}^2$ | |
| $k = 3.0$ | $\ddot{\theta} = 0.0246^\circ/\text{sec}^3$ | |
| $r = 1.0$ | $R = 5500 \text{ yd}$ | |
| $L = 30 \text{ yd}$ | $K = 0.01$ | |
| $\omega_F = 2 \text{ rad/sec}$ | $e_s = \text{unknown, to be determined}$ | |
| $\sigma_T \leq 0.105^\circ/\text{sec}$ | | |

$$\begin{aligned} \sigma_i &= 0.15 \left(\frac{30}{5500} \right) \left(\frac{180}{\pi} \right) = 0.047^\circ \text{ rms} \\ e_n &\approx (0.047)(6.28) \left[\frac{1}{3} \frac{(9 + 3 + 1)}{(3 + 3 + 1)} \right]^{1/2} (1 + 3)^{-1} \\ e_n &\approx 0.0578^\circ / \text{sec rms} \\ e_b &= \frac{0.13}{(20)(6.28)} + \frac{(0.0246)3}{(6.28)^2} = 0.00292^\circ / \text{sec} \\ e_g &\approx (0.01)(1.05) = 0.0105^\circ / \text{sec} \\ e_s^2 &= \sigma_T^2 - (e_g^2 + e_n^2) = (0.105)^2 - [(0.0100)^2 + (0.0578)^2] \\ &= 0.011025 - 0.00343 (\text{sec}^2) \\ e_s &= 0.0870^\circ / \text{sec rms} \end{aligned}$$

^aThis is actually part of an infinite series. Higher-order terms are neglected because the higher derivatives of θ are negligible in most cases. If the higher derivatives have large magnitudes, they must be included with more terms in the series or the series must be terminated with simple transient terms as described in a paper "A Simple Method for Calculating the Time Response of a System to an Arbitrary Input" by G. A. Biernson, MIT Servo Lab Report #7138-R-3, Jan. 1954.

The parameter k (the ratio of ω_c to ω_2) was chosen as 3.0. A larger ratio would produce long transient settling times and larger steady-state errors due to accelerations. A smaller ratio would cause a substantial increase in rate noise. This is shown in Fig. 9-6. For the simplified track loop shown,

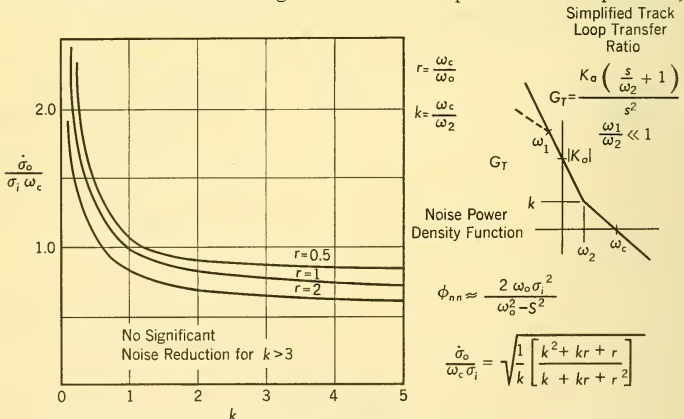


FIG. 9-6 Variation of Rate Noise with Tracking Loop Transfer Function.

the rate noise formula in the figure has been derived without simplification. From the curves, it can be seen that the choice of $k = 3$ is a rather arbitrary compromise. Actually $k = 2$ could have been selected, but, although not indicated in the figure, it would have produced a less stable system.

Bias Errors. As shown in Fig. 9-5, the bias error is obtained from a segment of an infinite series — an error series, the sum of line-of-sight angular derivatives¹⁴ attenuated by the velocity constant K_v and acceleration constant of the loop. Although it is not shown, the higher-order loop constants of the assumed track loop become larger and the input derivatives become smaller. Therefore, sufficient accuracy is obtained by using a series with only two terms. For inputs with larger, higher-order derivatives, more terms would have to be used as discussed in the footnote of Fig. 9-5. As shown, the bias rate error e_b is much less than the value specified. If it were not for transient conditions, omitted from the equations, the bandwidth and error constants could have much smaller values without causing the bias error to become excessive.

9-8 ANGLE TRACKING LOOP POSITION ERRORS

As discussed in Chapter 8, the position errors are generally not as significant as the errors in rate measurements. However, this is not always true and the position errors should be calculated. Again the errors are divided into random and bias categories.

Random Errors. The components of the random errors in the position signal and the methods of computing them are shown in Fig. 9-7.

The most significant random errors in position are caused by (a) mechanical components and misalignments e_c — including boresight mechanical transducers, servo noise, etc. — (b) antenna platform motion e_{sp} not completely counteracted by the stabilization loop, and (c) radar noise e_{np} transmitted through the control loop. Note that the formula for radar noise in the position signal is different from that used to calculate the noise in the rate signal.

For purposes of comparison with Fig. 9-5 the allowable stabilization errors in position are calculated. Although the magnitude of this error appears to be one-half of the allowable rate error, the loop gain required of the stabilization loop is only slightly lower.¹⁵ Of course, if more accurate components and alignments could be obtained, the stabilization loop gain could be less. On the other hand, if the stabilization loop gain could be

¹⁴The input derivatives, $\dot{\theta}$, $\ddot{\theta}$, and $\dddot{\theta}$ are the first, second, and third derivatives, respectively, of A_{LS} in Fig. 9-4. They are components in one channel only.

$${}^{15} e_{sp} = \frac{0.7A_{TL}}{G_2G_3} \leq 0.0452 \quad (\text{from Fig. 9-7})$$

$$|G_2G_3|_p \geq \frac{0.7A_{TL}}{0.0452} \quad \text{at } \omega = 3.14, A_{TL} = 4.76^\circ \text{ peak; } |G_2G_3|_p \geq 73.8$$

Increasing this by 75 per cent to allow for loop gain variations, $|G_2G_3|_p = 129$. The $|G_2G_3|_p$ for other frequencies are plotted in Fig. 8-42.

made larger, somewhat less accurate components might be used or the track loop bandwidth could be reduced. This illustrates the compromises that have to be made between the interrelated tracking loop design parameters.

Bias Errors. As shown in Fig. 9-7, the bias error e_{bp} is calculated approximately, again using a truncated error series. A third term is

FIG. 9-7 Antenna Position Error Calculations.

POSITION ERROR COMPONENTS

Random Error	Symbol	Approx. Formula	Comments
Component and Alignment	e_c	Constant	$0.5 < e_c < 3$ milliradians or $0.0288 < e_c < 0.173^\circ$
Position noise	e_{np}	$\sigma_i \sqrt{\frac{r(K+r+1)}{K+Kr+1}}$	$\sigma_i \approx 0.15 \frac{L}{R} \left(\frac{180}{\pi} \right)$; deg rms L = target length (yd) R = target range (yd) ω_0 = effective noise bandwidth of noise power density Φ_{nn} $\Phi_{nn} = \frac{2\omega_0\sigma_i^2}{\omega_0^2 - S^2}$ $r = \frac{\omega_c}{\omega_0}$
Stabilization	e_{sp}	$\frac{0.7 A_{TL}}{ G_2 G_3 _p}$	Discussed in Chapter 8
Total		$e_c^2 + e_{np}^2 + e_{sp}^2$	Specified $\leq 0.105^\circ/\text{channel}$
Bias Error			
Track loop	e_{bp}	$\frac{\dot{\theta}}{K_v} + \frac{\ddot{\theta}}{K_a} + \frac{\ddot{\theta}^a}{K_a'}$	Specified $\leq 0.372^\circ/\text{channel}$ K_v = velocity constant track loop K_a = acceleration constant track loop K_a' = acceleration derivative track loop constant $K_v = \omega_c I$; $K_a \approx \frac{\omega_c^2}{K}$; $K_a' \approx \frac{\omega_c^3}{K^2}$

CALCULATIONS
Assumed Values

$I = 20$
 $\omega_c = 6.28 \text{ rad/sec}$
 $K = 3.0$
 $r = 1.0$
 $L = 30 \text{ yd}$
 $\sigma_{TP} \leq 0.105^\circ/\text{channel}$

$\dot{\theta} = 1.05^\circ/\text{sec}$
 $\ddot{\theta} = 0.13^\circ/\text{sec}^2$
 $\ddot{\theta}^a = 0.0246^\circ/\text{sec}^3$
 $R = 5500 \text{ yd}$
 $e_c = 1.5 \text{ mils or } 0.086^\circ$
 $e_{sp} = \text{to be found}$

} Course conditions
of Fig. 2-44.
Time to go: 10 sec

$$\sigma_i = 0.15 \left(\frac{30}{5500} \right) \frac{180}{\pi} = 0.047^\circ \text{ rms}$$

$$e_{np} = 0.047 \sqrt{\frac{1(3+1+1)}{3+3+1}} = 0.0397^\circ \text{ rms}$$

$$e_{bp} = \frac{1.05}{125} = \frac{3(0.13)}{(6.28)^2} + \frac{9(0.0246)}{(6.28)^3} = 0.0193$$

$$e_{sp}^2 = \sigma_{TP}^2 - (e_c^2 + e_{np}^2) = (0.105)^2 - [(0.086)^2 + (0.0397)^2] = 0.002053$$

$$e_{sp} = 0.0453^\circ$$

^aSee footnote to Fig. 9-5.

included to illustrate how the higher-order terms become less important in the series for the particular conditions used. Again the bias error is considerably less than that allowed. However, as before, the input did not include the sudden changes that often occur in some attack courses, and under these circumstances, the bias error could be much larger. For example, if the 1.05° rate input had been applied suddenly, the peak error would be about 0.16° for a short time.¹⁶

From the preceding discussions, it appears that the track loop transfer function shown in Fig. 9-8 would meet the tracking loop error requirements. It should be noted that the loop phase margin is greater than 50° . This is necessary to provide adequate damping during the tracking operation.

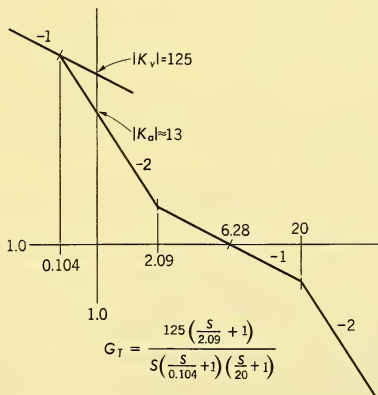


FIG. 9-8 Tracking Loop Transfer Function.

¹⁶See Chapter 8 where transient errors due to ramp inputs are discussed.

9-9 ANGLE TRACKING LOOP MECHANIZATION

In the preceding paragraphs, desirable transfer functions for the simplified stabilization and tracking loops, shown in Fig. 9-4, have been derived from typical operating conditions and accuracies. Before mechanizing transfer functions such as these, however, it would be desirable to check the theoretical frequency responses and transmit responses for the combined track and stabilization loops. Actually, with the antenna, actuator, amplifier, and networks, the stabilization loop is not the simple function shown in Fig. 9-4. It is much more complex; and although all the closed-loop corner frequencies of the stabilization loop are at considerably larger frequencies than those in the tracking loop, they may cause appreciable phase shift in the region near the tracking bandpass frequency. These calculations should be made as early as possible, perhaps without great precision, because it is not possible to measure many system characteristics accurately and there are usually many transfer characteristics which cannot be defined in a rigorous fashion. Often, the mathematical models are refined to resemble actual system characteristics after preliminary measurements and trials have been made. After the transfer functions have been combined and checked satisfactorily, the mechanization design may be finalized.

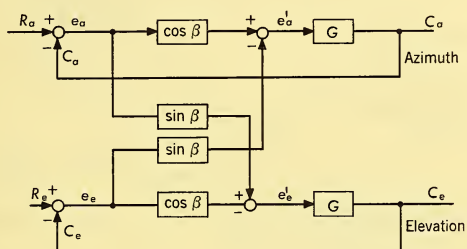
The significant elements in the track loop are depicted in Fig. 9-3. The major component, the radar, has been thoroughly described in earlier chapters. Although it is exceedingly complex, it performs a relatively simple function in the tracking loop: it converts angular space errors into a modulated voltage. If properly designed, the radar is essentially a gain with no appreciable time delays.

Another major part of the track loop is the closed stabilization loop described in Chapter 8. Although it has a complex transfer function, it provides the track loop with a pole at zero frequency — i.e. a simple integration. This makes the track loop a type I system¹⁷ which reduces steady-state position errors to zero.

Demodulator and Cross-Coupling Effects. The angle track demodulator K_d in Fig. 9-3 converts the modulated radar output voltage, which represents the total space error, into two voltage components, azimuth and elevation. Mechanically, K_d consists of two ordinary demodulators. One has a phase reference which produces an output voltage proportional to the azimuth space error and the other is phased to produce a voltage proportional to the elevation space error. Any of several types of demodulators could be used; but because of the large amount of noise inherent in the radar signal, averaging demodulators are usual. These demodulators have filters on the output to attenuate carrier frequency

¹⁷Harold Chestnut and Robert Mayer, *Servomechanism and Regulating System Design*, Vol. 1, p. 194, John Wiley & Sons., Inc., 1951.

harmonics. Ordinarily the phase shift produced by a low-pass noise filter would be detrimental, but because the track loop bandwidth is low the filter has little effect on the track loop stability. The demodulator selected, however, should have a low d-c drift. To provide the proper allocation of the total space error components into the correct channels, the demodulator reference voltage must have the correct phase with respect to the incoming voltage. If this phase is not adjusted properly, part of the azimuth error will be directed into the elevation channel, and part of the elevation error will be introduced into the azimuth channel. The cross coupling of the azimuth and elevation channels due to this carrier frequency phase shift is shown graphically in Fig. 9-9 for a typical two-channel control loop. As



$$C_a = \left[\frac{G(\cos \beta + G)}{1 + 2G \cos \beta + G^2} \right] R_a - \left[\frac{G \sin \beta}{1 + 2G \cos \beta + G^2} \right] R_e$$

$$C_e = \left[\frac{G(\cos \beta + G)}{1 + 2G \cos \beta + G^2} \right] R_e + \left[\frac{G \sin \beta}{1 + 2G \cos \beta + G^2} \right] R_a$$

FIG. 9-9 Cross-Coupling Due to Carrier Frequency Phase Shift β : Block Diagram and Equations.

indicated in the equations, normal individual control of each loop is obtained only when the phase shift error β is equal to zero. However, when β is other than zero, the input of each loop affects each output. This cross coupling effectively modifies the normal loop transfer ratio G by the function G' which is described in Fig. 9-10. As indicated, G' introduces phase lag proportional to $\sin^2 \beta$ at the normal crossover frequency ω_c of the control loop. Actually, when β is equal to the phase margin of the loop, $(\pi - \phi_c)$ at ω_c , the additional phase shift due to the cross coupling is also equal to $(\pi - \phi_c)$. This reduces the normal phase margin to zero and the loop becomes unstable.

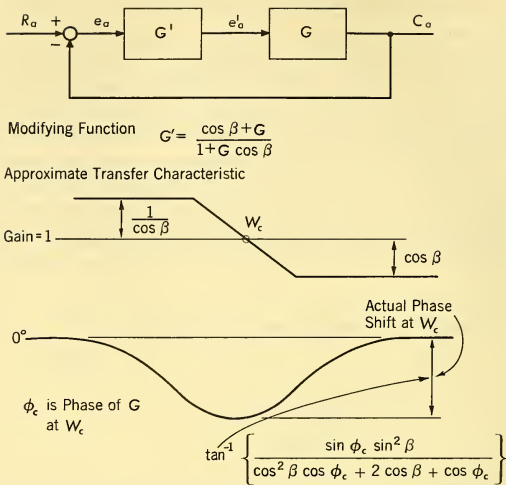


FIG. 9-10 Cross-Coupling Transfer Function Channel Output for Same Channel Input.

Therefore, as the phase of the reference voltage is shifted, the stability of each track loop channel changes. The overall effect on performance is that the settling time of each loop is extended, and the antenna moves in a spiral toward a preset position rather than in a straight line. The nature of the spiral for various amounts of reference phase shift is pictured in Fig. 9-11. The loop for which these illustrations were made had a large phase margin at the crossover frequency equal to 71° . The data indicate that the reference phase shift variation should be under 20° for loops with this large normal phase margin, and that the phase shift should be under 10° for loops with a small normal phase margin, because a noticeable deviation from an ideal straight-line response may be noted for even a 5° reference phase shift in a system having a large phase margin.

A similar form of cross coupling between channels can come from the gyros used in the space stabilization loop. The gyros are not the perfect instruments that have been represented in the block diagrams. They may produce voltage outputs, not only from space rates about their input axes but also from space accelerations about axes orthogonal to the input axis. Thus, accelerations in the elevation axis can appear as a voltage or an equivalent space rate in the azimuth axis; and, conversely, accelerations in the azimuth axis can create false rate signals in the elevation axis. The

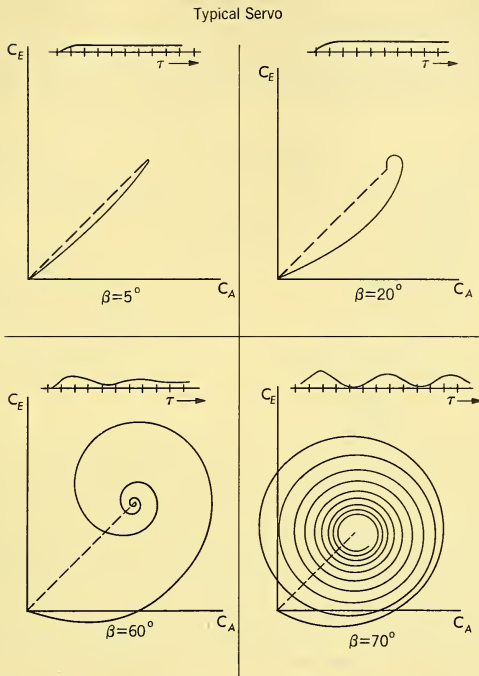


FIG. 9-11 Two-Dimensional Responses to Step Input.

amount of coupling between the channels from each gyro is proportional to \mathcal{J}/H , where \mathcal{J} is the moment of inertia of the gyro gimbal about its output axis and H is the angular momentum of the gyro wheel. Therefore it is important to select and orient the gyros so that the \mathcal{J}/H ratio is a minimum. If the \mathcal{J}/H ratio becomes too great, the cross coupling will become large enough to cause instability in the two loops. This is particularly true if the stabilization loop bandwidth is greater than 15 cps, because the \mathcal{J}/H coupling becomes more pronounced at higher frequencies and it is the normal attenuation of the higher frequencies by the stabilization loops that prevents instability in most cases.¹⁸

¹⁸A complete discussion of this problem is given in a paper "Analysis of Gyro Orientation," by Arthur Mayer in the *IRE Transactions on Automatic Control*, PGAC-6, December 1958, p. 93.

Tracking Loop Network and Amplifier. The frequency characteristics and the gain of the track loop shown in Fig. 9-8 are obtained with a shaping network and amplifier indicated by G_1 in Fig. 9-3. To obtain this function, an amplifier may be cascaded with a simple, passive RC network such as that shown in Fig. 9-12. For reference purposes the frequency

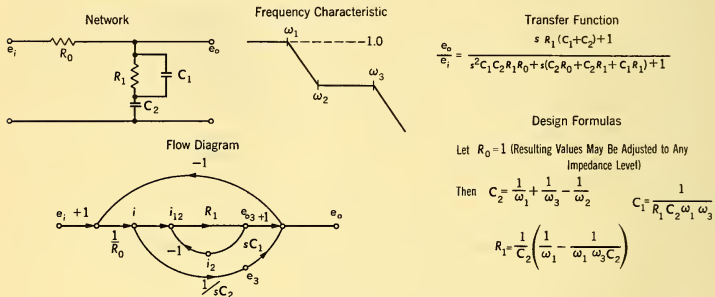


FIG. 9-12 Passive RC Tracking Network.

characteristic, a flow diagram, the transfer function, and useful design formulas are given. Note that the design formulas are expressions for deriving the parameter values directly in terms of the corner frequencies desired in the transfer function. This eliminates a cut-and-try procedure in the design synthesis. Although this is a simple network, it has some disadvantages: (a) the parameter values, particularly the capacitor C_2 may become quite large for loops with narrow bandwidths, and (b) a change in amplifier gain will create a corresponding change in track loop bandwidth.

These objections can be largely eliminated by using an active network, the operational amplifier shown in Fig. 9-13. It should be noted that the transfer function is essentially the same as that of the passive network except that the input resistor value is modified by the amplifier gain. Consequently, the lower lag corner is also a function of amplifier gain; and if the gain of the amplifier changes, the low-frequency corner will change inversely. Thus the bandwidth, stability, and noise transmission of the tracking loop will not be altered. It should be noted that the low-frequency corner is achieved by using feedback which permits the use of smaller parameters, and the output impedance of the active network is much less than that of the passive network. This reduces the problem of loading, but the output impedance still will not be zero and its value must be considered if the impedance of a device following the network is low.

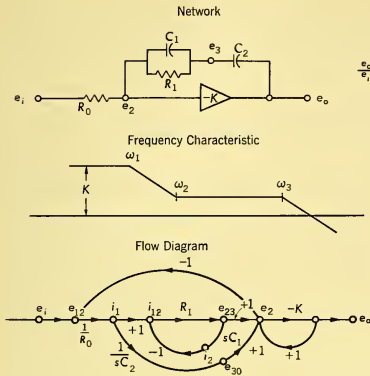


FIG. 9-13 Active RC Tracking Network.

Transfer Function

$$\frac{e_o}{e_i} = \frac{sR_1(C_1 + C_2) + 1}{s^2 C_1 C_2 R_1 R_0' + s(C_2 R_0' + C_2 R_1 + C_1 R_1) + 1} \quad \text{where } R_0' = R_0(1+K)$$

Design Formulas

Let $R_0' = 1$ (Resulting Values May be Adjusted to Any Impedance Level)

$$\text{Then: } C_2 = \frac{1}{\omega_1} + \frac{1}{\omega_3} - \frac{1}{\omega_2}$$

$$R_1 = \frac{1}{C_2} \left(\frac{1}{\omega_2} - \frac{1}{\omega_1 \omega_3 C_2} \right)$$

$$C_1 = \frac{1}{R_1 C_2 \omega_1 \omega_3}$$

General Design Considerations. In general, considerable care must be used in designing the track loop circuits. Proper grounding and shielding must be utilized to minimize ground loops and induced noise. Attenuation and transfer function modifications due to impedance loading must be considered, especially in the summing networks where the gyro rate signal is added to the track loop. Adequate dynamic range must be provided in the amplifier to minimize the possibility of producing grid current and saturation due to large signal and noise inputs, especially during transient operation. It may be necessary to place Zener diodes across large capacitors to prevent voltages from accumulating at times when large step inputs are applied. At the stabilization summing point, a voltage sufficient to balance the large gyro voltage resulting from maximum antenna rates must be available from the angle track circuits.

Finally, the tracking loop must be constructed and evaluated. This cannot be done satisfactorily until the actual antenna is constructed and used in the loop; ideally the measured antenna characteristics should be available in the initial design stages of the stabilization loop. After construction, the antenna performance, essentially its tracking accuracy, must be measured. Usually, as the result of preliminary measurements on the component parts of the system, the design may have to be modified to obtain the transfer functions desired. In airborne equipment it is important to measure the track loop performance over wide ranges of environment — pressure, temperature, and vibration.

The physical design problems associated with each high-performance antenna tracking loop, including its stabilization loop, vary with different

antennas and in different systems. No specific design or design method can be given concisely, but the basic design principles discussed in this chapter must be considered when determining the angle track characteristics for any system.

9-10 INTRODUCTION TO RANGE AND VELOCITY TRACKING

The subject of radar range tracking is first mentioned in this book in Paragraph 1-4 preceding the radar range equation, Equation 1-2. The concepts of indirect determination of range rate by differentiation of radar range information and of direct measurement by doppler frequency shift, Equations 1-19 and 1-20, are introduced in subsequent paragraphs. Doppler frequency is again discussed in Paragraph 4-4, in connection with sea return in Paragraph 4-11, and in connection with cross-correlation radar in Paragraph 6-4. Paragraph 6-6 is entirely devoted to concepts related to pulsed doppler radar systems. This last paragraph discussed a method of range tracking wherein the entire interpulse period was covered by contiguous fixed-range gates. Each range gate interval was divided into contiguous velocity (doppler frequency) intervals (see Fig. 6-11). In such systems range and velocity tracking is provided by a radar data processing system (or computer) which records the channel location of each signal return. The AEW system discussed in Chapter 2 is an example of a system that might employ such tracking areas.

In single target systems such as the AI radar of Chapter 2 or a missile target seeker, the range and velocity tracking problem is somewhat different. Generally these systems have only a single channel for the flow of target information. Range and velocity tracking is accomplished by time and frequency gates which are controlled to bracket the received pulse and doppler frequency of the target return (See Paragraph 6-5 for an example of a moving velocity gate).

The following paragraphs will discuss the basic principles of range and velocity tracking as they are commonly applied to single target systems. The principal emphasis will be given to the dynamic problems of automatic systems designed for such purposes, since the electronic problems of gating circuits and time and frequency discriminators have been adequately covered in the general literature.

9-11 AUTOMATIC RANGE TRACKING

The principal functions of the range tracking unit may be listed as follows:

1. To provide means for automatic measurement of target range

2. To provide means for discriminating against returns which originate at ranges different from those of the desired target
3. To provide means for effecting the transition from range search to automatic range tracking.

The first function involves measuring the round-trip transit time of the radar energy; the second involves positioning a range gate such that only signals in time coincidence with the target pulse are passed through the receiving system; the third involves the dynamic problems of stopping a sweeping range gate.

Range gating — as illustrated in Fig. 5-13 and described in Paragraph 5-8 — is necessary to the satisfactory operation of the angle tracking system. In fact the target pulse train must be time selected (i.e. range lock-on must occur) before angle tracking can be started.¹⁹ This sequence of operations ensures that returns from other targets on the ground or clouds — at ranges other than the range of the desired target — do not compete with the desired target signal.

Range measurement generally is employed to provide inputs to the threat-evaluation and fire-control systems. Accuracy requirements vary widely; errors of no more than a few yards may be allowed for unguided weapons, while guided missile systems can often tolerate rather large range errors, as was shown in the example of Chapter 2 (Table 2-4).

Typical means for providing automatic range tracking and gating are shown in Fig. 9-14. Also shown (in dotted lines) are the parts of the radar

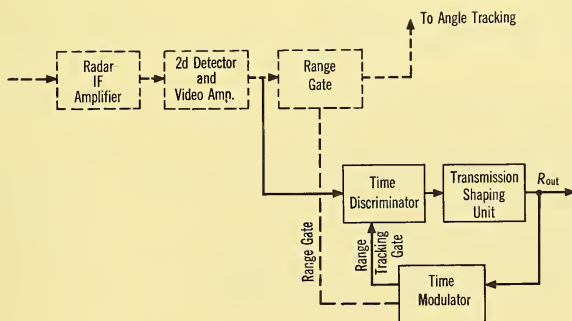


FIG. 9-14 Automatic Range Tracking System.

¹⁹In doppler radars which do not employ range gating, a similar function is served by the velocity gate — in this case frequency selection replaces time selection. In some pulsed doppler systems both time and frequency selection are utilized preparatory to angle lock-on (see Paragraph 6-6).

system with which the range tracking system must be integrated. The operation of the range tracking system may be described in general terms by considering the functions of each element — time discriminator, time modulator, and transmission shaping unit — and the essential features of their closed-loop operation.

Time Modulator and Time Discriminator.²⁰ These units provide a means for generating a range error signal. They accomplish this by dividing the video signal to be tracked into an early part and a late part and comparing the energies in each part to develop a voltage proportional to range position error. Two methods commonly employed for these

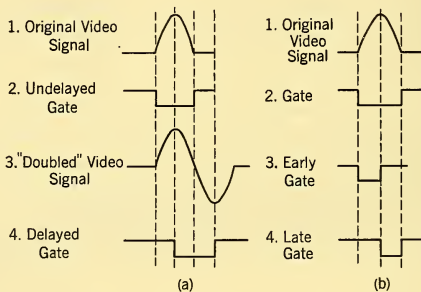


FIG. 9-15 Wave forms in a Time Discriminator: (a) Double Signal Type. (b) Split Gate Type.

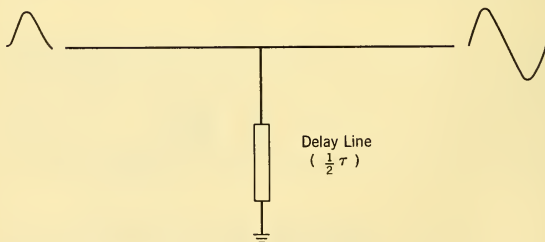


FIG. 9-16 Double Signal Time Discriminator.

²⁰See *Guidance*, Par. 10-31, which forms the basis for this discussion.

purposes are shown in Fig. 9-15. In the *double signal* type, the original video signal is transformed into a bipolar signal by adding to the video signal its reflection from a shorted delay line. Since the delay line is shorted, the reflection is the negative of the original video. By making the delay equal to one-half the pulse width, the total delay of this negative signal equals the pulse width. A schematic representation of this process is shown in Fig. 9-16.

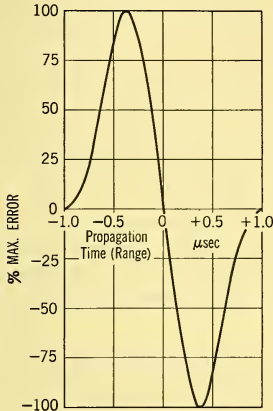


FIG. 9-17 Time Discriminator Characteristic.

The time modulator generates two gates: one is the range gate used to time-select the video information which goes to the angle tracking system; the other is a delayed range gate used to integrate the energy in the bipolar pulse to produce a range error signal. The typical output of the time discriminator as a function of gate position with respect to the echo center is shown in Fig. 9-17. A second means for tracking is the so-called split gate system of Fig. 9-15b. In this scheme two adjacent gates are employed. The pulse energy falling in one gate is subtracted from that in the other to form an error signal. The time discriminator characteristic produced thereby is the same as that of the

other system.

A variation of these methods is the so-called *jittered range gate* method. With this technique, the position of a range gate is modulated at some low frequency — say 30 cps. The product of the jittered range gate and the received pulse will possess a modulation component at the jitter frequency whose amplitude and phase signify the misalignment between the pulse and the average gate position. This a-c error can be demodulated to provide an error voltage for controlling the gate. One disadvantage of this particular scheme is that there is an extra modulation frequency upon which extraneous noise and jamming can enter the system.

Transmission Shaping Unit. This is a filter whose characteristics determine the bandpass and lag errors of the range tracking system. Subsequent discussion in this chapter will deal with the methods for selecting the parameters of this unit.

9-12 SERVO SYSTEM TRANSFER FUNCTION RELATIONSHIP TO INPUT TIME FUNCTION FOR A RANGE TRACKING SYSTEM²¹

Target and interceptor flight-path analysis provides a time-function specification for automatic range and angle tracking radar servo-system inputs. Whereas the tactical problem is entirely a time-domain phenomenon, servo-system design procedures depend upon both time- and frequency-domain quantities. The salient points in the practical realization of an appropriate servo-system transfer characteristic in the preliminary design stage are outlined in the following paragraphs.

In radar systems, the received signal contains the desired information together with uncorrelated noise components. To proceed with the servo-system design on a rational basis, the nature of the undesired noise should be determined. The noise spectral density, its mean value relative to the desired information, and the maximum excursions should be known.

Tactical Problem Inputs. The initial considerations common to automatic tracking servo-system preliminary design studies include:

1. From the space geometry involved, derive analytical expressions for the input position as a function of time. Take derivatives to find the velocity and acceleration.

2. Study the tactical situation and choose as typical problems those that appear to place the most stringent demands upon the servo system under study, i.e., the highest values of input velocity and acceleration.

3. Select one set of parameter values to define the *design problem*. Prepare accurate graphs of position, velocity, and acceleration as functions of time.

4. Make a critical appraisal of the desired performance to establish the maximum allowable tracking error.

Transformation to Frequency Domain. The intelligence to be tracked is given by the input time function; however, the equivalent frequency-domain transformation is desired. "Initial conditions" at the start and end of the tracking period affect the resulting frequency-domain equivalent. For some geometries a direct analytic transformation may be made, but in general a graphical procedure is needed.

Any arbitrary function of time may be transformed to the frequency domain by Guillemin's^{22,23} impulse method, in which the integrand of the

²¹Discussions with members of the Radar Division, U.S. Naval Research Laboratory — especially with J. P. Dougherty and S. F. George — materially aided the author in rounding out some of the concepts presented.

²²E. A. Guillemin, "Computational Techniques which Simplify the Correlations Between Steady-State and Transient Response of Filters and Other Networks," *Proc. Nat. Electronics Conf.* 1953, 9, 513-532 (1954).

²³J. G. Truxal, *Automatic Feedback Control System Synthesis*, pp. 379-382, New York: McGraw-Hill Book Co., Inc., 1955.

Laplace integral is converted to a set of impulses in order that the integral may be evaluated without numerical methods. Two operations are involved. The time function, or a derivative of it, is first approximated with a series of broken lines. A broken-line approximation of the first derivative is equivalent to fitting the time function itself with a series of parabolic segments. Broken-line approximation of the acceleration curve is equivalent to fitting the original time function with a series of cubic segments. When the original function is based upon other than an analytic function, the errors resulting from graphical differentiations tend to mitigate against any increased accuracy resulting from fitting with segments of higher-order curves, so that a point of diminishing return is reached. Once the broken-line approximation has been made, the second step in the process is to differentiate the broken-line approximation twice to obtain a set of impulses. The impulses are an approximation representing the information within the time-function interval and are based upon the tacit assumption of a smooth function at the start and end of the interval. That is, immediately before and immediately after the starting and ending points, the position, velocity, acceleration, and all higher order derivatives of motion have the same value. When other initial and final conditions are to be imposed, additional impulses (doublets, etc.) appear at these two points.

Using the Transformed Input Information. A log-magnitude log-frequency plot is prepared from the mathematical expression equivalent to the transformation to the frequency domain of the set of impulses. Straight-line segments are drawn (having integral values of slope) tangent to the maximum values on the log-log plot to form the asymptotic equivalent. The servo-system excitation function as represented by the asymptotic segments is then compared with the spectral density of the extraneous noise. This comparison may lead to careful shaping of a regulation frequency response in the case of an AGC design or to merely emphasizing the necessity for minimizing the servo-system transmission bandwidth. The excitation function is then studied in relationship to possible servo open-loop transfer characteristics with a view to providing, in general, gain variation with frequency matching the excitation function. Closed-loop transient response requirements preclude steeper slopes through unity gain than perhaps $-1\frac{1}{2}$ (-2 results in instability).

Choice of Gain Level. Graham's²⁴ design procedure makes use of a steady-state error series having the general form

$$\epsilon_{ss} = \frac{R_i}{1 + K_p} + \frac{\dot{R}_i}{K_v} + \frac{\ddot{R}_i}{K_a} + \dots \quad (9-4)$$

²⁴R. E. Graham, "Linear Servo Theory," *Bell System Tech. J.* **25**, 616-651 (1946).

where ϵ_{ss} = steady state error

R_i = input position

\dot{R}_i = input rate

\ddot{R}_i = input acceleration

K_p = position error constant

K_v = velocity error constant

K_a = acceleration error constant.

The values for K_p , K_v , and K_a are readily derived for a given system²⁵ by writing the expression

$$\frac{\epsilon_{ss}}{R_i} = \frac{1}{1 + \mu} = \frac{N(s)}{D(s)} \quad (9-5)$$

with powers of the complex frequency $s = j\omega$ in increasing order. A direct division of $D(s)$ into $N(s)$ gives the desired constants. Variations in design predicated upon higher derivatives than the second (acceleration) should be avoided.²⁶

Steady-state error as a function of time may be found using the input position, velocity, and acceleration found above. The contemplated system is satisfactory if the parameters (corner frequencies) have been selected to cause the maximum error to equal the maximum allowable.

The best system becomes the one that simultaneously meets the tracking accuracy specifications and requires the minimum bandwidth. Such a system is more readily realizable physically and excludes the maximum amount of extraneous noise.

Redesign. In the foregoing paragraphs a design procedure leading to an initial trial design of a linear servo system has been given. Before such a system would be recommended, the transient response, the possibilities of improvement using nonlinear techniques, and the compatibility with respect to the system environment should be studied. Methods of design permitting prescribed steady-state and transient response have been developed. Other developments indicate that the entire design may be accomplished in the time domain. The foregoing notes refer to the older steady-state design procedure.

²⁵A. S. Locke, *Guidance*, pp. 230-263.

²⁶J. G. Truxal, *op. cit.*, pp. 236-245.

9-13 RANGE TRACKING DESIGN EXAMPLE

As an example of going from the tactical to the practical in preliminary servo-system design²⁷ consider a bombing aircraft on a constant-velocity straight and level flight that is radar range tracking a fixed surface target as depicted in Fig. 9-18 for which

t = time in seconds ($t = 0$ at minimum range or crossover)

V = aircraft horizontal speed in knots

H = aircraft altitude

R_i = range to target in yards

R_0 = slant range to target at crossover

O = fixed ground surface target.

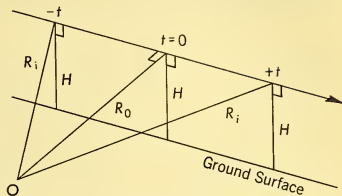


FIG. 9-18 Pass-Course Range-Tracking Problem.

The aircraft radar range tracking time function becomes

$$R_i = [R_0^2 + (KVt)^2]^{1/2} \quad (9-6)$$

where $K = 0.5626 \frac{\text{yd hr}}{\text{n.mi. sec}}$

Equation 9-6 may be normalized to the form

$$\frac{R_i}{R_0} = \left[1 + \left(\frac{t}{\tau} \right)^2 \right]^{1/2} \quad (9-7)$$

where $\tau = R_0/KV$.

As was suggested earlier, an asymptotic-segment representation of the upper bound of the excitation function transformed to the frequency domain that is reasonably accurately located provides the amount of detail useful in the engineering design of a servo system transfer function. Since our objective is to determine the upper bound of the magnitude on a log-magnitude log-frequency plot, with any errors resulting from approximation methods such that a servo-system transfer function based upon the analysis will be conservatively designed, the tedious details associated with close

²⁷Pass-course angle tracking is treated in a comparable manner in C. F. White, *Tactical to Practical in Preliminary Servo-System Design*, NRL Report 4879, February 1957.

approximations to the input time function may be avoided. We proceed by taking derivatives of Equation 9-7. Thus

$$\frac{\dot{R}_i}{R_0} \tau = \frac{\left(\frac{t}{\tau}\right)}{\left[1 + \left(\frac{t}{\tau}\right)^2\right]^{1/2}} \quad (9-8)$$

$$\frac{\ddot{R}_i}{R_0} \tau^2 = \frac{1}{\left[1 + \left(\frac{t}{\tau}\right)^2\right]^{3/2}} \quad (9-9)$$

Plots of Equations 9-7, 9-8, and 9-9 are shown in Fig. 9-19. A three-segment broken-line approximation of range rate is shown in Fig. 9-20a.

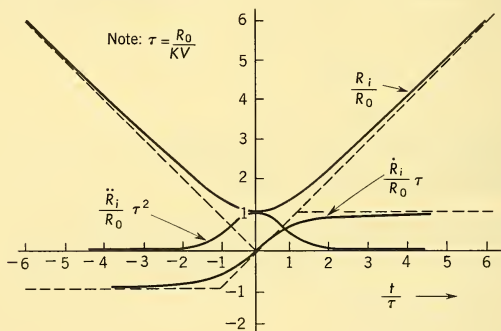


FIG. 9-19 Normalized Position, Velocity, and Acceleration Components of Motion for Pass-Course Range Tracking.

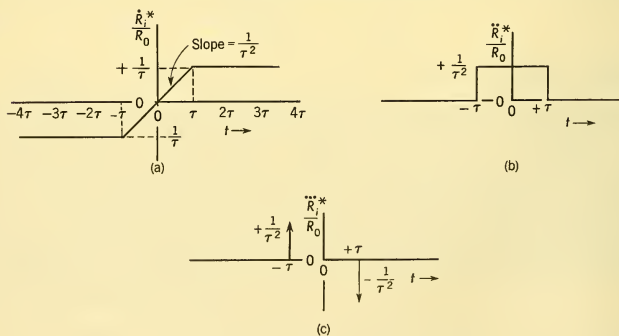


FIG. 9-20 Approximations for Components of Motion Leading to Impulses.

Two additional derivatives lead to the impulses of Fig. 9-20c. The mathematical expression for the steady-state frequency-domain equivalent of these impulses becomes

$$\frac{R_i^*}{R_0}(s) \xrightarrow{s \rightarrow j\omega} \frac{1}{(j\omega)^3} \frac{1}{\tau^2} (-e^{-j\omega\tau} + e^{+j\omega\tau}) \quad (9-10)$$

$$\left| \frac{R_i^*}{R_0}(s) \right| \xrightarrow{s \rightarrow j\omega} \left| \frac{2 \sin \omega\tau}{\omega^3 \tau^2} \right| \quad (9-11)$$

The low-frequency upper bound is found by using the small angle $\sin x \rightarrow x$ approximation. Thus,

$$\frac{\left| \frac{R_i^*}{R_0}(s) \right|}{\tau} \xrightarrow[\omega \rightarrow 0]{s \rightarrow j\omega} \left[\frac{\sqrt{2}}{\omega\tau} \right]^2. \quad (9-12)$$

The high-frequency upper bound is found by using the maximum value (+1) for the sine function. Thus,

$$\frac{\left| \frac{R_i^*}{R_0}(s) \right|}{\tau} \xrightarrow[\omega \rightarrow \infty]{s \rightarrow j\omega} \left[\frac{\sqrt[3]{2}}{\omega\tau} \right]^3. \quad (9-13)$$

Figure 9-21 shows the results given by Equations 9-12 and 9-13 on a log-magnitude, log-frequency plot. The low-frequency upper bound for

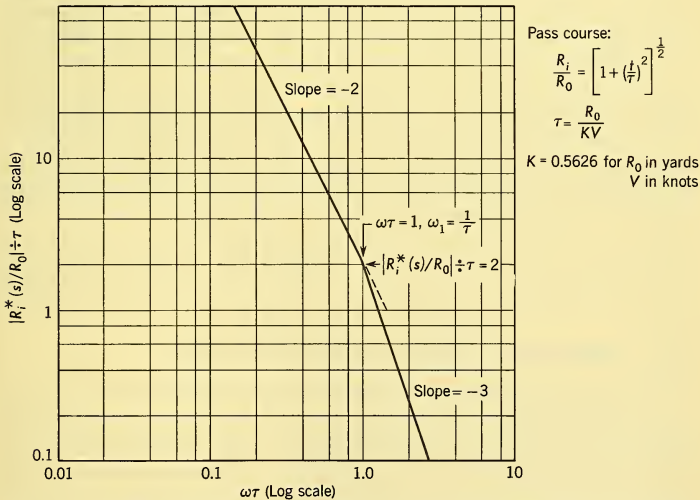


FIG. 9-21 Asymptotic Upper Bound Frequency Domain Representation of Pass-Course Range Tracking.

$|R_i^*(s)/R_0| \div \tau$ is a straight line with a -2 or double-integrator slope which, when projected to a log unity or zero value, crosses the abscissa at $\omega\tau = \sqrt{2} \cong 1.414$. The high-frequency upper bound has a -3 slope through $\omega\tau = \sqrt[3]{2} \cong 1.2599$. These asymptotic segments intersect at $\omega\tau = 1$ at a value $|R_i^*(s)/R_0| \div \tau = 2$.

The asymptotic upper bound of the amplitude distribution shown in Fig. 9-21 may be expressed by the equation

$$R_i^*(s) = 2R_0 \left(\frac{R_0}{KV} \right) \left[\frac{\left(\frac{KV}{R_0} \right)}{s} \right]^2 \left[\frac{\left(\frac{KV}{R_0} \right)}{s + \left(\frac{KV}{R_0} \right)} \right] \quad (9-14)$$

in units of yard seconds. If the definition of a *corner frequency*

$$\omega_1 = \frac{1}{\tau} = \frac{KV}{R_0} \quad (9-15)$$

is introduced, Equation 9-14 may be written in the form

$$R_i^*(s) = \frac{2R_0}{\omega_1} \cdot \left(\frac{\omega_1}{s} \right)^2 \left(\frac{\omega_1}{s + \omega_1} \right). \quad (9-16)$$

Gain level Frequency characteristic

The pass course qualifies as an excellent test signal for investigation of the performance of automatic range tracking servo systems. At times that are widely removed from crossover, the input approaches a constant range rate. In the vicinity of crossover there is a transition from the approximation of a constant rate to a zero range rate at the instant of crossover. In the sense that all circles (or parabolas, etc.) have a fixed shape, all pass courses may be characterized by one general function. This fact is emphasized by Fig. 9-19 which specifies, in the time domain, the nature of all pass courses in a fully normalized manner. One needs only to evaluate the time constant $\tau = (R_0/KV)$ for a particular case. Fig. 9-21 specifies, in the frequency domain, the upper bound for the pass-course range tracking input information important in design of the automatic tracking servo system.

9-14 PRACTICAL DESIGN CONSIDERATIONS

For range tracking, the low-frequency asymptotic slope should be that of a double integrator. To this basic fact one adds the practical considerations that airborne radar systems employ electronic range tracking systems that have a definite gain limit. This means that the -2 slope cannot be extended indefinitely, and in practice a zero slope at extremely low frequencies results. An additional practical consideration is the ever-present requirement that the feedback system be stable. This requirement means, in

general, the presence of an asymptotic segment with a -1 slope in the vicinity of unity gain.

The foregoing material, then, provides the reasons for the nature of the "typical" airborne radar range transfer function shown in Fig. 9-22.

In the detailed consideration in Chapter 2 of an airborne radar system, the dynamic variation of lead collision fire-control parameters was depicted in Fig. 2-45. Range and range-rate plots as functions of time-to-go were given there; when compared with the pass-course range function considered here, these are seen to be remarkably comparable in all aspects important to choice of range tracking servo system transfer function. Thus the pass course is seen to qualify equally well for air-function. Thus the pass course is seen to qualify equally well for air-to-air interceptor-target range tracking system considerations.

If we make a straight-line approximation of the range rate curve of Fig. 2-45 by a zero slope at the -1266 -ft/sec ordinate value to 3.3-sec time-to-go followed by a straight line with a slope of 80 ft/sec² through the -1000 -ft/sec ordinate value at zero time-to-go (and assume a comparable positive-time function to avoid transients extraneous to the real range tracking problem), the equation comparable to Equation 9-10 becomes

$$R_i^*(s) \xrightarrow{s \rightarrow j\omega} \frac{1}{(j\omega)^3} 80(-e^{-j\omega 3.3} + e^{+j\omega 3.3}). \quad (9-17)$$

Then,

$$\left| R_i^*(s) \right|_{s \rightarrow j\omega} \rightarrow \left| \frac{160 \sin 3.3 \omega}{\omega^3} \right| \quad (9-18)$$

$$\xrightarrow{\omega \rightarrow 0} \frac{52.8}{\omega^2} \quad (9-19)$$

$$\xrightarrow{\omega \rightarrow \infty} \frac{160}{\omega^3}. \quad (9-20)$$

The -2 , -3 corner (see Fig. 9-21) becomes

$$\frac{52.8}{\omega^2} = \frac{160}{\omega^3}; \quad \omega_1 = \frac{160}{52.8} = 3 \text{ rad/sec.} \quad (9-21)$$

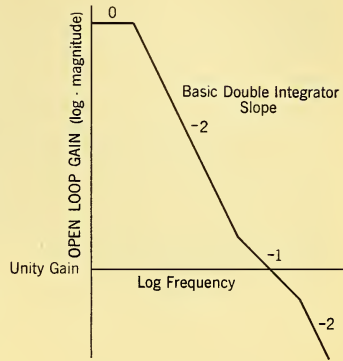


FIG. 9-22 General Shape of Typical Airborne Radar Range Tracking Servo System Transfer Function.

With the information contained in the range tracking input time function lying below 3 rad/sec, every effort should be made to match the automatic range tracking system bandwidth to this input spectrum. Factors precluding close match are the requirements on tracking error and stability of the range tracking loop as discussed in connection with Fig. 9-22. Completion of all details of design for a range tracking servo system transfer function follows the basic procedures presented earlier in this chapter in connection with angle tracking systems. In all precision tracking systems the servo bandwidth greatly exceeds the information bandwidth. However, by making a preliminary study of this nature, the servo-system designer can proceed with confidence in that the low-frequency portion (where the largest amount of energy lies) has been matched and no high-frequency information has been discarded.

CHAPTER 10

ANTENNAS AND RF COMPONENTS*

SYMBOLS USED IN CHAPTER 10

A_e	effective antenna area or aperture		of angular error
A_p	antenna physical cross section	S_{\min}	minimum detectable signal
c	velocity of light in free space	v_g	group velocity
D	antenna directivity	v_p	phase velocity
E_T	target error normalized to antenna beamwidth	α	attenuation constant
f	frequency	β	phase constant
G	gain	γ	propagation constant
j	$\sqrt{-1}$	ϵ_r	relative dielectric constant or permittivity
L_c	crystal diode conversion loss (power ratio)	θ	angular displacement from the parabolic centerline
n	number of slots in an array	Θ	beamwidth
N_c	crystal diode noise ratio (power ratio)	Λ	lobing angle (squint angle)
N_{if}	IF amplifier noise figure (power ratio)	λ	free-space wavelength
N_r	receiver noise figure (power ratio)	λ_c	cutoff wavelength
r	polar radius of arithmetic or geometric growth curves	λ_p	wavelength in transmission line
S	per cent modulation per unit	μ_r	relative permeability
		σ	target back-scattering cross section
		ω	angular frequency ($2\pi f$)

There is at the front end of a radar receiver an almost one-to-one analogy with components of a conventional communications receiver. The only reason for the difference in appearance of this section of the radar as compared with the communications receiver is the difference in frequency handled at the antenna and at the input to the receiver. Both receivers are usually heterodyne receivers and for much the same reason. As such, a local oscillator and mixer are involved. In the radar receiver, the local

*Paragraphs 10-1 through 10-9 are by M. S. Wheeler. Paragraphs 10-10 through 10-20 are by P. J. Allen. Paragraph 10-21 is by Dr. P. M. Pan.

oscillator is likely to be a klystron. The mixer is usually a pair of semiconductor diodes. If the radar frequency is low enough for a traveling wave tube to be available with a competitively low noise figure, such a tube may be used for an RF amplifier.

Radar frequency techniques are in a transition region between the methods of ray optics and lumped circuits. Here a host of specialized antennas and components have been developed which have in common an outer metallic covering to prevent loss of energy by radiation. Here, also, special techniques of measurement and calculation are used which have in common their dependence on field theory. In some ways, this is a difficult part of the frequency spectrum in which to work. The components are expensive, large, and closely toleranced. The measurements are restricted, and analysis by field methods is cumbersome. On the other hand, there are certain functions that can be performed here more easily than at other frequencies, especially in the field of nonreciprocal devices and of the so-called hybrid couplers.

This chapter will inspect a few of the antennas and components in this frequency band and will show the engineering approach to problems related to these devices. A method of looking at antennas is developed which is particularly useful for microwaves, and several antenna types are described, such as the parabolic reflector and the linear array. These are typical of important classes of radar antennas. Quite a number of waveguide components are presented in enough detail to give the reader a familiarity with their problems and applications, but not enough detail for design without further reference. For this reason, references are included to which the reader can refer for design.

10-1 ANTENNAS: INTRODUCTION TO RADAR ANTENNAS

Radar information concerning a specific target consists of range and direction quantities and their rates of change. The range information is concerned with time measurement and as such has little to do with the antenna design. Direction information, on the other hand, comes directly from the antenna through its ability to concentrate radiated power in one direction relative to another. It is to this end that the radar antenna is planned.

In general, there are two distinctly different types of direction information required by the radar of the antenna. First, information may be necessary to a comparatively rough degree of accuracy concerning the many targets that may exist in a sector of space of tactical interest around the antenna. This function is referred to as *search* in fire-control radar or as *ground surveillance* in mapping radar. Second, information involving both direction and rates of change of direction may be required to a high

degree of accuracy on one or more specific targets. This function is referred to as *tracking*. Often, to save weight in a fire-control radar, these two functions are required from the same antenna. If this is the case the search and track functions are usually not required simultaneously. The AI radar discussed in Chapter 2 is an example of such a system. In other cases, such as the AEW system of Chapter 2, the search and tracking functions are carried on simultaneously in what is called a *track-while-scan mode* of operation.

In view of these things, the most salient feature of the radar antenna with which we are concerned is directivity. It is related to the radar angular resolution and range. Explicitly, directivity is 4π times the ratio of the maximum radiation power angular density to the total power radiated by the antenna.¹ That is to say, directivity is a measure of the antenna's ability to concentrate power in a particular direction. This antenna characteristic and others such as beamwidth and gain are related through the antenna effective aperture.²

10-2 SOME FUNDAMENTAL CONCEPTS USEFUL IN THE DEVELOPMENT OF RADAR ANTENNA REQUIREMENTS

The effective antenna area or aperture A_e may be defined for a *lossless, matched* antenna as

$$A_e = \frac{P_{\text{load}}}{P_{\text{density}}}, \quad (10-1)$$

where P_{load} is the power delivered to its load and P_{density} is the power density in an incident plane wave. It should be said at this point that the antennas of which we speak are assumed to be linear and reciprocal. Antenna characteristics such as aperture, directivity, etc. are unchanged whether the antenna is used in transmission or reception. Sometimes it is more convenient to give a physical picture of the phenomena in terms of a receiving antenna. At other times the physical picture as a transmitter is more easily visualized as in the previous definition of directivity. This selection is only a matter of convenience and in no way affects antenna performance, whether the antenna is being used in reception or in transmission.

In either case the aperture concept is especially useful for many radar antennas because (as in the definition) losses are relatively low, the antenna is usually well matched to the load, and the effective aperture is within

¹*Standards on Antennas: Definitions of Terms*, The Institute of Radio Engineers, 1948.

²John D. Kraus, *Antennas*, Secs. 3-2 and 3-3, McGraw-Hill Book Co., Inc., New York, 1950.

50 per cent of the physical cross section normal to the transmission direction. Antenna directivity D , as previously defined, is simply related to the effective aperture³ A_e , at a wavelength λ by

$$D = \frac{4\pi}{\lambda^2} A_e. \quad (10-2)$$

This expression shows that the directivity of an antenna increases in direct proportion to its effective area expressed in square wavelengths.

Thinking of the antenna, now, as a transmitter, if the actual physical aperture is fed in such a manner as to contain a wave of constant phase and amplitude, then the physical aperture is equal to the effective aperture in the above relation for directivity. Generally, however, a variation from a plane wave in phase and amplitude may exist across such an aperture. For example, an amplitude distribution which is much lower at the edges of the aperture is usually required in practice to improve some antenna characteristics. Thus the effective aperture may be smaller than the physical cross section A_p , by a factor which is called *illumination effectiveness*.

$$A_e = k_{\substack{\text{illumination} \\ \text{effectiveness}}} A_p. \quad (10-3)$$

Antenna gain is closely related to directivity. It is usually defined, however, for an actual antenna which is not perfectly matched to its load and which is not dissipationless. Gain is less than directivity because of these two losses:

$$G = k_{\substack{\text{dissipation} \\ \text{efficiency}}} \times k_{\substack{\text{matching} \\ \text{efficiency}}} D. \quad (10-4)$$

In many practical radar antennas, $k_{\substack{\text{matching} \\ \text{efficiency}}}$ will be essentially unity,

$k_{\substack{\text{dissipation} \\ \text{efficiency}}}$ will be greater than 0.9, and $k_{\substack{\text{illumination} \\ \text{effectiveness}}}$ will be greater than 0.5.

Thus, gain is nearly equal to directivity and it is within 50 per cent of $4\pi/\lambda^2$ times the antenna physical cross-section area.

Finally, the antenna beam shape is described by certain terms such as beamwidth, sidelobe level, etc. The sidelobes, such as those in Fig. 10-1 are an inherent but generally undesirable characteristic of the antenna. They are described by their magnitude and direction relative to the main lobe. The pattern in Fig. 10-1 has 0.05 or -13 db sidelobes at ± 6 deg azimuth.

³H. T. Fris and W. D. Lewis, "Radar Antennas," *Bell System Tech. J.*, p. 229 (April, 1947).

Some antennas have *cylindrical symmetry* resulting in equal beamwidths in azimuth and elevation. Such a pattern is called a *pencil beam* and is used for tracking and search functions. Other antennas have a fan-shaped beam with greatly different beamwidths and sidelobes in azimuth and elevation. These antennas are used for search and ground surveillance, for example. Still others have essentially non-directional patterns and are called *omnidirectional antennas*.

The concept of antenna beamwidth is useful in narrow-beam radar antennas. *Beamwidth* is defined as the angular displacement between points at which the received signal power is one-half of its value at the center of the beam. The pattern in Fig. 10-1 has a 4° beamwidth. From the preceding discussion on directivity, beamwidth is in some manner inverse to directivity. The exact relation is not simple, however. For example, beamwidth becomes undefined for very low directivity where the received signal power may never drop below one-half its peak intensity. An approximate relation between directivity and beamwidth is given in a later discussion on pattern approximations.

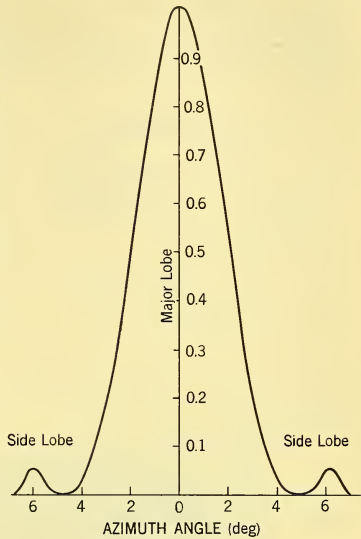


FIG. 10-1 Typical 4° -Beamwidth Antenna Pattern Showing the Major Lobe and Two Minor or Sidelobes.

10-3 THE PARABOLOIDAL REFLECTOR AS A RADAR TRACKING ANTENNA

There are several different fire-control systems in common use requiring some difference in antenna design. Common to them all, at least in airborne radar, is a paraboloidal reflector. The chief difference between antennas for a conically scanned, a sequentially lobed, a monopulse, etc. radar is in the means by which energy is coupled to the reflector in order to obtain an error signal for automatic tracking. For brevity, it will be necessary to select one system for discussion such as the conical scan. The other systems have identical aperture (that is, the reflector) problems and different but analogous feed problems. Extension of this discussion and these principles to other antennas can be made.

The Aperture. With a paraboloidal reflector, it is convenient to project energy from the end of a waveguide, which is small compared with the final aperture size required. The feed, being small, is effectively a point source of radiation which is placed near the focus of the parabola (dish) and with its peak radiation aimed toward the center of the dish. The final transmitting aperture can be considered as the plane containing the outer diameter of the reflector. The use of such a reflector is based on two independent and distinct properties of the parabola. First, all rays striking the dish surface from the focus reflect from the dish parallel to its central axis. Second, all ray paths from the focus to the dish to the aperture plane are of the same length. These are shown in Fig. 10-2, where lengths abc for all paths are equal and bc is parallel to the central axis.

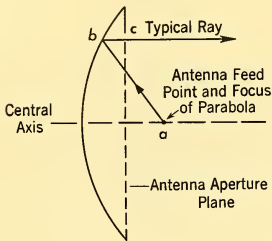


FIG. 10-2 Properties of a Parabolic Reflector, Showing a Typical Ray with All Path Lengths abc Equal and the Path bc Parallel to the Central Axis.

Thus, a wave front on all ray paths from a point source, at the final crossing of the aperture, are automatically of equal phase.

Properties of a Parabolic Reflector. The relative phase and power density of the wave at various points in the aperture plane is called the *aperture illumination*. The constant-phase characteristic provided by the parabola is quite universally desirable. With constant phase, it can be shown that maximum directivity is achieved by constant amplitude illumination.⁴ The amplitude characteristic of the aperture, however, is usually controlled in a prescribed manner by the beamwidth of the feed. If the illumination were so adjusted that constant illumination were achieved, the directivity would be maximized and the antenna physical aperture would be equal to the effective receiving aperture. Unfortunately, as has been mentioned, this gives rather high sidelobes; further, considerable power is lost over the edge of the dish and is not returned to the aperture. Thus the actual gain is maximized and sidelobes are reduced by lowering the power density from the feed about 10 db at the edge of the dish relative to that at the center.⁵

The Tracking Mechanism. Now consider the tracking function required of the antenna. In conical scan radar, the feed is not placed

⁴Samuel Silver, ed., *Microwave Antenna Theory and Design*, Radiation Laboratory Series (MIT), Volume 12, p. 178, McGraw-Hill Book Co., Inc., New York, 1949.

⁵*Ibid.*, pp. 423-433.

exactly at the focal point of the paraboloid but rather is slightly displaced laterally and rotated in a circle of constant radius. This provides a motion of the radiated pattern conically about the antenna centerline—hence, the name, conical scan. The results of this are twofold. First, the signal, upon passing through the antenna, is modulated by the spinning feed for all targets except those on the centerline of the antenna. The phase of this modulation gives the direction in which the antenna must move to bring the target onto the centerline. Second, angular resolution is increased so that a target can be located to a small fraction of the actual beamwidth.

Adversely, however, the greater the feed offset, the greater the beam asymmetry and sidelobes that accompany the beam offset. For a typical antenna, however, with the beam offset so that extreme beam positions cross one another about $1\frac{1}{2}$ db below the peak, these difficulties are not excessive. A typical antenna of this type is shown in Fig. 10-3. This has a 4.0° beam from a 20-inch dish with a $1\frac{1}{2}$ db crossover.

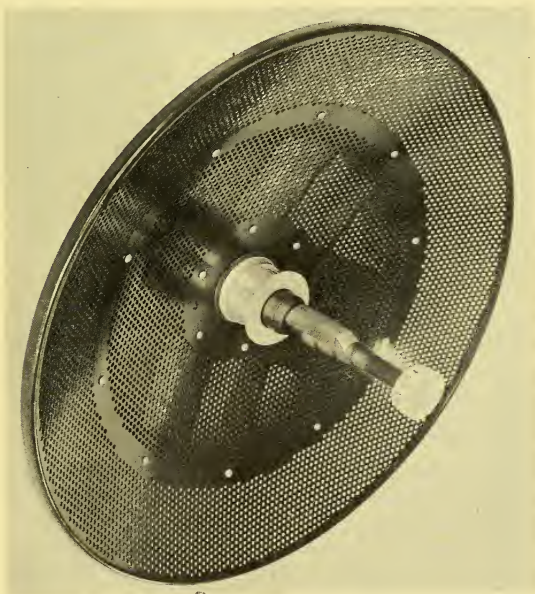


FIG. 10-3 A Typical Center-Fed Parabolic Tracking Antenna with a 4° Beam and a $1\frac{1}{2}$ -db Crossover. (Courtesy, Westinghouse Electric Corporation)

Quite analogous to the conical scan antenna are the sequentially lobed and the simultaneously lobed tracking antennas. Where the conical scan beam makes a complete circle in deriving error signal, these antennas look only up-down and left-right. With the sequentially lobed tracker, this is accomplished in a programmed sequence (as the name implies), while in the simultaneously lobed tracker, multiple receiving channels measure up-down and left-right tracking errors simultaneously. Both of these antennas have four stationary feeds suitably offset from the focal point of the reflector and derive error information by signal comparison from the several feed points. Monopulse systems have already been discussed in Paragraph 6-3.

10-4 SYSTEM REQUIREMENTS FOR RADAR ANTENNAS

Up to this point, antenna characteristics such as gain, directivity, etc. have been discussed as referred principally to a parabolic reflector. These are directly measurable antenna characteristics that describe not only radar antennas but many others as well, regardless of the way in which they are used. This is the vocabulary of the antenna trade. From the point of view of system requirements as derived in Chapter 2, however, it is necessary to start from different characteristics — maximum range, resolution, tracking accuracy, etc. The required antenna parameters are then derived.

In the search function, the relation between the antenna and radar is straightforward. The antenna gain may be used directly in the radar range equation, (Paragraph 3-2)

$$R^4 = \frac{P_t \sigma \lambda^2 G^2}{S_{\min} (4\pi)^3} \quad (10-5)$$

where R is the maximum useful range to a target

P_t is transmitted power

S_{\min} is the minimum detectable signal

G is antenna gain

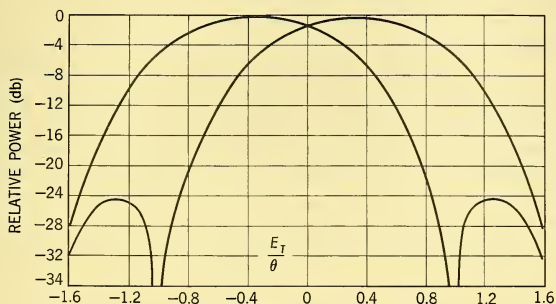
λ is wavelength

σ is target back-scattering cross section.

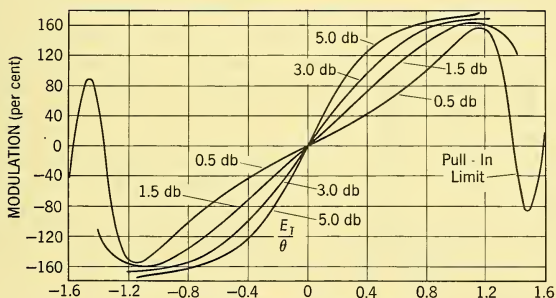
This equation shows the importance of antenna gain in determining the useful range of a radar as compared with transmitted power, for example.

A second characteristic of the search function of radar is angular resolution. This is defined as the minimum angular separation of two point targets at the same range. As shown in Paragraph 3-6, resolution is a complicated function of beam shape as well as the way the radar circuitry handles the information; but at least to a rough degree, angular resolution is about one beamwidth.

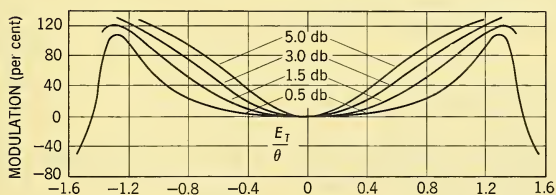
The tracking function stands in need of more analysis, however. For a conical scanning radar, the tracking error signal is an amplitude modulation



(a) TYPICAL CROSSOVER PATTERN



(b) FUNDAMENTAL MODULATION



(c) 2d HARMONIC MODULATION

FIG. 10-4 Tracking Error Signal as Derived from a Conical Scan Radar Antenna: (a) Typical Pattern with the Beam Offset to a $1\frac{1}{2}$ -db Crossover. (b) The Per Cent Fundamental Modulation or Error Signal in Conical Scan as a Function of Target Error and Beam Offset. (c) The Per Cent Second Harmonic Signal Which Is Not Useful as Tracking Information but Which Can Cause System Problems.

on the signal returned from the antenna after a round trip between the radar and target. The modulation is almost sinusoidal near zero error, but becomes high in harmonics at large errors. It is found that the per cent fundamental modulation of the received signal is the important quantity through which the tracking servo is linearly closed.⁶ That is, the amplitude of the per cent fundamental modulation is proportional to target displacement from the antenna centerline. The phase of this signal is the target position around the antenna in a roll coordinate. Thus the per cent fundamental modulation goes to zero with no target error and reverses sign (phase) when the error goes through zero. This error curve is shown in Figure 10-4(b) for several values of beam offset. Here (E_T) is target error normalized to beamwidth (θ). For a high angular resolution, the function should have a high slope and should be linear, at least in the region of zero error.

The transition of radar function from search to track is called *lock-on*. During this period, the radar antenna is positioned manually as close as possible to the correct aiming direction and then a switch is closed to change the radar function. As the initial aiming error may be relatively large, it is important that the error curve have at least the correct sign (even if it is no longer linear) out to this initial probable error. The error signal sign reversal is the limit of target lock-on; beyond this point the antenna will be driven in the reverse direction by the antenna position servo and the target will be lost. Finally, a false lock-on may occur at a target error other than zero where the per cent fundamental modulation again goes through zero with the same polarity of slope as it does on the antenna centerline. Such a possibility is shown in Figure 10-4(b) for $\frac{1}{2}$ -db crossover with a target error of 1.57 beamwidths. See Paragraph 10-6 for more discussion of false lock-on.

It can be seen that fundamental modulation is important to the tracking function. It can be measured by harmonic analyzer methods on an actual antenna or it can be calculated from a known beam shape. In this manner the system requirements can be related to the usual antenna parameters such as beamwidth. For this calculation and others, an approximation to the radiation pattern by a simple function is required.

10-5 PATTERN SIMULATION AS A LINK BETWEEN SYSTEM REQUIREMENTS AND ANTENNA CHARACTERISTICS

A mathematical description of the antenna pattern is a link between antenna terminology and system requirements. For problems involving

⁶Aeronautical Electronics Digest, 1955, IRE National Conference on Aeronautical Electronics, Page 13. (Also see Paragraph 5-8, above.)

the main lobe of a cylindrical beam, it is found convenient to use

$$p = e^{-(a\theta)^2} \text{ with } a = \frac{2\sqrt{\log_e 2}}{\Theta} \quad (10-6)$$

where p is the normalized one-way power pattern, θ is angular displacement from the parabola centerline, and Θ is the beamwidth.

When a more complete description of a pattern is needed having side-lobes, one of the Bessel functions is found useful.⁷ For example, the second-order function has first sidelobes about 25 db below the main lobe. This might be considered a good design practice for a narrow-beam tracking system.

$$p = 64 \left[\frac{J_2(b\theta)}{(b\theta)^2} \right]^2 \text{ with } b = \frac{1.27\pi}{\Theta} \quad (10-7)$$

Both of these functions are shown in Fig. 10-5 as compared with measurements on a typical pencil beam. The measured pattern is for an offset beam tracking antenna. This accounts for its lack of symmetry about the vertical axis, particularly in the side lobes

For any calculations within 10 db of the peak of the beam such as error signal sensitivity, the radiation pattern may be fitted very nicely by the exponential of Equation 10-6. With this assumption, the modulation sensitivity for zero target error is given by⁸

$$S = \frac{4 \log_e 2}{\Theta^2} \Lambda \times 100 \text{ per cent.} \quad (10-8)$$

Here, Λ is the lobing angle between extreme beam positions during the cycle of conical scan (squint angle), Θ is beamwidth, and S is the per cent modulation per unit of angular error.

Another useful result derivable from this pattern approximation is the *db difference* expression that has been

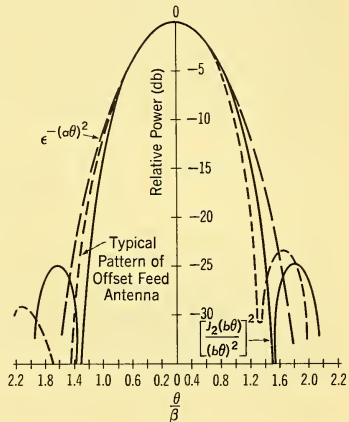


FIG. 10-5 The Radiation Pattern of a Typical Offset Feed Conically Scanning Antenna as Compared with Two Pattern Shape Approximations That Are Often Used to Represent the Actual Antenna Pattern.

⁷Silver, *op. cit.*, p. 194.

⁸J. B. Damonte and D. J. Stoddard, "An Analysis of Conical Scan Antennas for Tracking," 1956, *IRE Convention Record*, Part I, Page 42.

used in the past to express system performance. Explicitly, db difference is the difference between the oneway antenna relative power curves in decibels at extreme positions of the conical scan cycle. While this term is less frequently used now (per cent fundamental modulation more accurately describes the tracking performance) the term *db difference* is often seen in specifications. For the exponential beam approximation per cent modulation equals

$$\text{Per cent modulation} = \frac{\log_e 10}{20} \times (\text{db difference}) \times 100 \text{ per cent.} \quad (10-9)$$

The above relation is derived for the nonsinusoidal waveform as the difference between maximum and minimum amplitudes divided by the average value. As the modulation is nearly sinusoidal for small target error, the db difference is very nearly so related to the per cent fundamental modulation in this region.

To illustrate the utility of the Bessel approximation to the narrow-beam pattern, the relation between directivity and beamwidth may be calculated. It is necessary to restrict this to narrow beams to make the integration practical, and as has been mentioned, because the beamwidth becomes undefined for wide beams. From the definition of directivity and for narrow beams

$$D = \frac{K \left[\frac{J_2(a\theta)}{(a\theta)^2} \right]^2 \left[\frac{J_2(b\theta)}{(b\theta)^2} \right]^2}{\frac{1}{4\pi}} \quad (10-10)$$

where $\frac{1}{4\pi}$ is the power density at unit distance from a unit power source radiating omnidirectionally. The proportionality constant K may be determined by integrating⁹ the pattern over all space with a narrow-beam approximation, giving

$$D = \frac{25,500}{\Theta_a \Theta_b} \quad (10-11)$$

where Θ_a and Θ_b are beamwidths in degrees along the major and minor axes of the beam.

Further use of the Bessel representation is made in calculating tracking performance out to the sidelobe region. The fundamental and second harmonic modulation calculations are made as outlined in the reference of footnote 6. The results may be shown as in Fig. 10-4. Here (E_T) is target error, which is normalized to antenna beamwidth (Θ).

⁹Higher Transcendental Functions, Bateman Manuscript Project, California Institute of Technology, Volume 2, p. 51.

These data are typical of measurements on conical scan systems out to one beamwidth; in this region there is little difference between one antenna and another. Beyond this point, however, the characteristics depend upon the detail of the sidelobe structure. Fig. 10-4 is carried to this extreme to show that angular pull-in (as would be expected intuitively) is maintained to approximately the point where the two extreme lobe positions cross for the second time. Compare Figs. 10-4(a) and 10-4(b). From this kind of information, the optimum antenna crossover may be selected for a given application. Note that a $1\frac{1}{2}$ -db crossover gives optimum linearity. As the crossover is deepened, angular sensitivity increases and the angular pull-in increases slightly. Adversely, however, per cent harmonic distortion increases. Typical second-harmonic distortion produced at various crossover values are shown in Fig. 10-4(c). Higher harmonics also are generated, but to a smaller degree. The presence of these harmonic distortions requires special consideration in the design of the receiver and angle tracking system. Another factor to consider is tracking range, which may decrease somewhat at deeper crossover because of the lower power incident on the target. A $1\frac{1}{2}$ -db crossover is often considered optimum.

Procedures very similar and analogous to this may be repeated for other problems such as monopulse radar, beam-riding missiles, etc.

10-6 SEVERAL ANOMALOUS EFFECTS IN ANTENNAS FOR TRACKING SYSTEMS

Radar return from targets or ground return may dictate the wave polarization to be used for a particular design. At other times, the choice between vertical, horizontal, and circular polarization is less important; or possibly, it is a weighted consideration between conflicting requirements. The effects of polarization upon signal return are discussed in another chapter of this book. One other factor in the polarization choice needs mentioning, however.

In addition to the ground return normally returned from the sidelobes, there is inevitable scattering from the antenna and aircraft in a vertical direction, giving a strong return from the ground directly beneath the aircraft. This has been termed *altitude line* because of its appearance on a B scope as a line at a range equal to the airplane's altitude. Sometimes this return is so strong that tracking of a target at a range equal to the altitude line becomes difficult. As the ground return directly beneath the antenna is polarization insensitive, and as a parabolic antenna often has one direction of this right angle scattering which is substantially less than the other, the antenna may be polarized to favor the direction of low ground return. This factor should be taken into account when the radar polarization is chosen.

Another antenna problem affecting system performance has been termed *false lock-on*. In this event, a radar tracking lock is achieved with an exceptionally strong target which is at some angle other than zero with the paraboloid centerline. This is seen in all types of radar systems, but practice has found conical scanning systems less susceptible to the problem than others. With beams having cylindrical symmetry in their patterns, including the sidelobes, a false lock-on would not be possible. The beam asymmetry, however, caused by offset feed which is necessary to produce the required error signal gives sidelobes which are not cylindrically symmetric. A three-dimensional picture of this phenomenon would show the major lobe as a large symmetrical central peak, but the sidelobes as a rather irregular range of low mountains arranged around the foot of the major lobe. In this case, a sidelobe peak may follow a conical motion similar to the main lobe, and some form of lock-on can be achieved at this crossover with a sufficiently strong target.

It is difficult to exclude this possibility from antenna measurements alone, and as a result, it is usually best to test the completed antenna for false lock-on with the actual system with which it is to be used.

A final effect to be mentioned with narrow-beam antennas is *harmonic generation* by large targets at close ranges. When the target subtends a large portion of or exceeds a beamwidth, the target structure modulates the return as the beam sweeps past a wing or tail surface. This modulation gives a false error signal causing angular steering errors at short ranges. The phenomenon is not related to target glint noise, although they both become effective at short range. The only relief for this problem as far as the antenna is concerned is to widen the antenna beam.

10-7 THE LINEAR ARRAY AS A FAN BEAM ANTENNA FOR SURVEILLANCE

An antenna array is a combination of two or more elementary radiators so connected to give pattern — and thus gain — reinforcement in certain directions and pattern cancellation in other directions. One class of antennas used often in radar is a long linear array of elementary radiators so driven as to produce field reinforcement broadside to the array. These antennas find application in ground surveillance and in AEW radar.

The Slotted Waveguide as a Fan Beam Antenna for Ground Surveillance. System requirements for these search antennas have little in common with the fire-control antennas previously discussed. The pattern is broadbeam in elevation of a modified *cosecant squared* ($\text{csc}^2 \theta$) form¹⁰ to give uniform illumination of a considerable segment of the earth's

¹⁰Silver, *op. cit.*, p. 471.

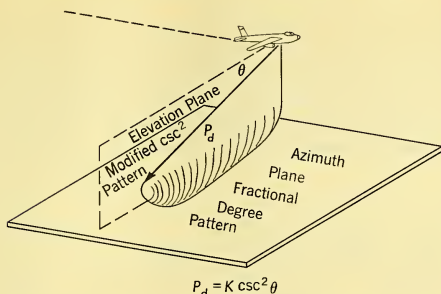


FIG. 10-6 A Perspective View Showing Schematically the Radiation Pattern of a Side-Looking Fan Beam for Ground Surveillance Radar.

surface. This type of radiation pattern is shown in Fig. 10-6 for a side-looking ground surveillance system.

The pattern in azimuth is a fraction of one degree to give the required angular resolution. Targets are resolved in elevation on a time basis. That is, for a given altitude, objects directly beneath the antenna are at minimum range, and those on the horizon are at maximum range. The antenna will be designed for a specific angular coverage in elevation which is less than these two extremes — the actual pattern being so shaped in elevation as to give uniform display brightness within the design range. Mathematically speaking, such an antenna has, in elevation, a power density pattern which varies as the square of the cosecant of the angle between the horizontal and the slant range line-of-sight. Some antennas of this type cover as much as 70° in elevation in this manner.

The antenna, then, must be swept in azimuth. One usual system — already described in Chapter 1 — is to mount the antenna rigidly to the airframe for side looking and allow the forward motion of the aircraft to sweep the coordinate as the aircraft flies straight and level over the ground to be studied. In this event, a storage or photographic means will be required to give a coherent picture of the slowly unfolding panorama.

For such a fan-shaped beam as described, it is almost impossible to feed a rectangular parabola as the radiating aperture. In addition, a rectangular aperture does not have the optimum mechanical form for the application. As a result it is usual to approximate a continuous antenna aperture by a discontinuous array of slots in a waveguide wall.

Along the long aperture, illumination is first chosen based upon the gain and sidelobe requirements for a given aperture length. The type of slot used in the guide wall determines the antenna polarization. Longitudinal slots in the top wall of the waveguide result in a polarization in the plane of

the fan, while essentially transverse slots in the side wall of the guide result in a polarization perpendicular to the plane of the fan. If the polarization is not specified by the system requirements, it is preferred to use the side-wall slots for several practical reasons. Side-wall slots are easier to fabricate, and the vertical pattern in the fan is more readily controlled with this polarization. Slots of these two types are shown in Fig. 10-7.

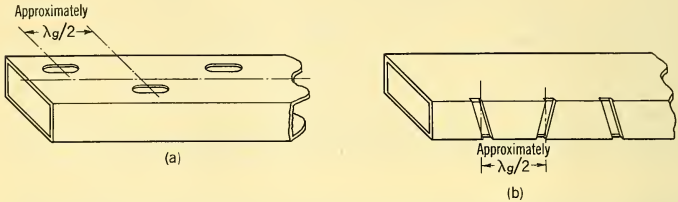


FIG. 10-7 Waveguide Slotted in a Manner Suitable for Antenna Arrays with Reverse Phase Coupling in Adjacent Slots: (a) Top Wall Slots Which Radiate with Electric Polarization in the H Plane of the Guide. (b) Side Wall Slots with Electric Polarization in the H Plane of the Guide.

Consider the slot spacing. There is a minimum slot spacing required to prevent deterioration of the main beam, given by Silver ¹¹

$$S \gtrsim (1 - 1/n)\lambda \quad (10-12)$$

where n is the number of slots, λ is the free space wavelength; and as all slots must be driven in approximately the same phase, it is usual to space the slots by essentially half a guide wavelength and arrange for the resulting required phase reversal at alternate slots. For example, in using side-wall slots, the direction of tilt in adjacent slots is reversed to provide this phase reversal. The result is that E -plane components of the electric field in adjacent slots tend to cancel, whereas H -plane components add. Thus, an array of in-phase radiators has been obtained spaced at half a guide wavelength, giving polarization parallel to the waveguide axis.

It is not usual to space the slots at exactly half a guide wavelength, however. This spacing of many small discontinuities gives a resonant antenna, often with an impractically narrow bandwidth. A small change in the slot spacing allows the many reflections to tend to cancel one another, resulting in a good bandwidth. In this case, the fan beam is tilted a small amount from the normal to the waveguide. For this reason, many arrays of this type will be found with beam centers 1° to 5° from a perpendicular position.

¹¹Silver, *op. cit.*, p. 319.

The azimuth pattern of this type of antenna is well controlled and quite predictable. For a given beam shape, the aperture illumination is well known, and this illumination is forced by the many slots in the waveguide wall. With the vertical slots again, for example, the angle of each slot to the normal is adjusted to give the required coupling, taking into account the fact that the power in the waveguide is becoming progressively less in passing down the array.

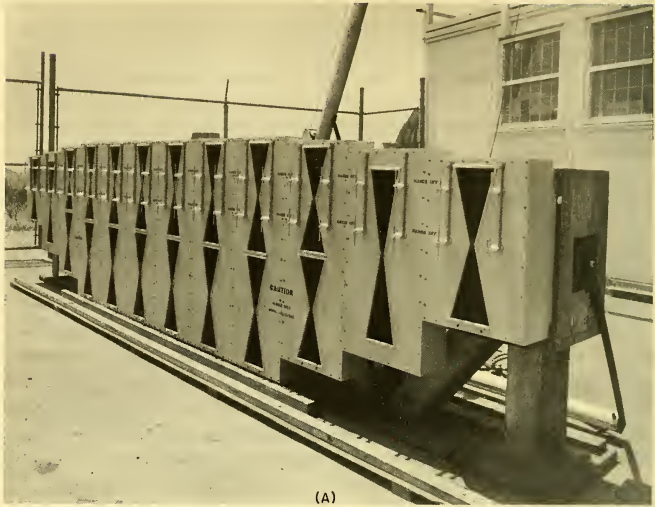
Finally, the beam is shaped in elevation to a modified cosecant-squared pattern by a reflector and/or lens system adjusted to give the most power at the depression angle of maximum range and the least power at the angle of minimum range. For uniform display intensity, Silver states that this pattern determined experimentally is approximated by $\csc^2 \theta \times \cos \theta$ for wavelengths from 1 to 10 cm. Pattern irregularities in this plane need be reduced below the intensity resolution of the system to prevent strips of light and darkness on the recorded picture. A typical system limit between peaks and valleys is 2 db.

The Loaded Resonator Array as a Fan Beam for AEW Radar.

Closely related electrically, but differing physically from the slotted waveguide array, are the low-frequency arrays used for long-range AEW radar. Because the AEW radar usually requires less in angular resolution and more in range than the ground surveillance radar, it is usual to go to lower frequencies where much greater power-handling capacity is available and atmospheric attenuation is negligible.

Here the waveguide array is unrealistic. It is usual to find an array of dipole elements or of cavity elements fed from a branching line which has been termed a *corporate feed structure*. The term "corporate feed" describes the branching nature of the feed, analogous to charts of line authority in corporation management. By suitably designing this branching line, the power distribution between elements in the array is established to control the radiation pattern, beamwidth, and sidelobe structure — consistent with the number of elements allowed within the array. The general nature of the power division is similar to that described under the ground surveillance antenna. The number of elements, however, is substantially smaller, so that it is preferable to analyze the problem from the point of view of an N element array rather than from the point of view of a continuous distribution of fields in an aperture.

The radiating element is frequently a slot, suitably backed up by a cavity to limit radiation to one primary direction. In Fig. 10-8(a) an array of twenty "bow-tie slots" are shown in an AEW antenna developed by Hughes Aircraft. The feed structure is visible in Fig. 10-8(b). The slots are horizontally polarized. The reasons for narrowing the slots in the center are multiple. This is one dimension in controlling the impedance match



(A)

between the slot and its driving coaxial line. Further, this wasping effect capacitively loads the slot at its center, allowing for a physically smaller slot and cavity than would have been possible otherwise. This is important at low frequencies where components are becoming heavy and cumbersome.

The particular antenna shown in Fig. 10-8 is seen to have a secondary antenna consisting of an array of dipole elements at a high frequency and lower power. Because of the frequency difference, they do not couple to one another and the two arrays are conveniently suspended by the same structure.

10-8 TWO-ARM SPIRAL ANTENNAS

A final antenna type to be described, finding application to passive radar direction finding and semiactive missiles, is logarithmic in nature. Typical of antennas in this class are the spiral antennas which are wide band and wide beamed.

Geometry. Flat spiral-wound antennas have been studied extensively by the University of Illinois¹² and others. This antenna in its pure form is

¹²"The Equiangular Spiral Antenna," Technical Report No. 21, Sept. 15, 1957, Electrical Engineering Research Laboratories, University of Illinois, Contract AF33(616)-3220, Project 6(7-4600), Task 40572, W.A.D.C.

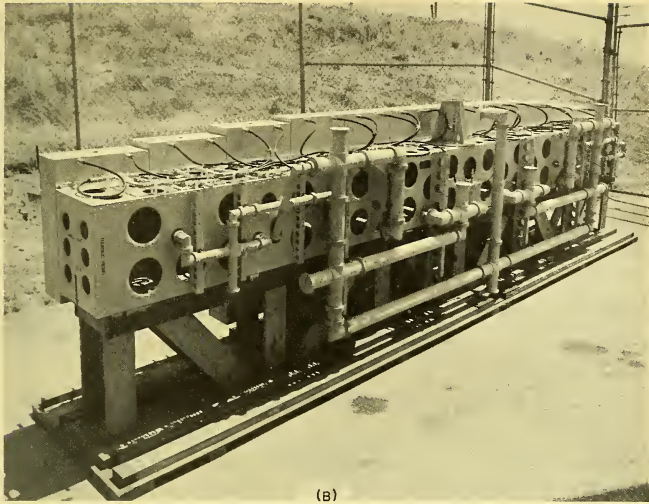


FIG. 10-8 A Hughes Aircraft Co. AEW Antenna, Showing (a) the 20-Element "Bow Tie" Structure and (b) the Corporate Feed Structure of Coaxial Lines at the Rear of the Antenna.

described by angles alone (at least within its region of radiation), and thus it has no lengths to be frequency sensitive. Physically, the antennas consist of dipole-like conductors with boundaries defined by a geometric growth,

$$r = e^{C\theta}, \quad (10-13)$$

and wound in a spiral in a plane perpendicular to its radiating direction (see Fig. 10-9a). When the spiral is so proportioned that the antenna is in its

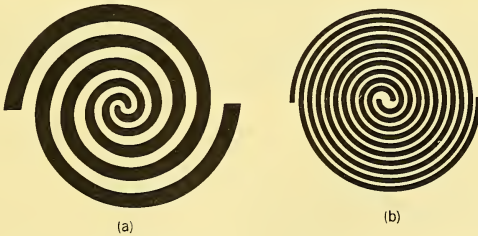


FIG. 10-9 The Basic Two-Arm Spiral Antennas: (a) Geometric Spiral. (b) Arithmetic Spiral.

most useful form as a constant beamwidth, circularly polarized, broadband radiator, an engineering description of the antenna pictures a traveling wave moving outward along the spiral to a diameter where the circumference is 1 wavelength. Around this circumference, the waves on adjacent turns are such as to reinforce one another, producing radiation in this ring; whereas, in general, at other diameters, no reinforcement takes place and radiation is small. Since most of the energy leaves the antenna at a circumference on 1 wavelength, little is left at the next resonant circumference of 2 wavelengths.

A spiral so proportioned, fitting this engineering description and having a circle-like turn in the radiating region, gives conductor sizes impractically fine in the central driving point area. As a result, another type of spiral has been studied at the Massachusetts Institute of Technology¹³ and elsewhere, which is similar in all respects except that it is defined by an arithmetic growth,

$$r = C\theta. \quad (10-14)$$

This spiral, pictured in Fig. 10-9(b), has electrical characteristics similar to the geometric spiral, including wide bandwidth and constant beamwidth. Thus although the conductor width is a specifying linear dimension, the antenna is a good approximation to the principle of description by angles only when the conductor width is made small compared to a wavelength everywhere in its radiating band.

Analysis. With wirelike radiator elements such as have been described, it is more satisfactory to consider the problem from antenna array theory than from aperture consideration, because in small apertures of the order of wavelength squared, the effective aperture cannot be determined by a mere inspection of the physical aperture. The radiating ring of the spiral can be reduced to an array in the following manner.

The current in the radiating ring has been described as a traveling wave,

$$I = \cos(\omega t - \beta Z) \quad (10-15)$$

where ωt produces the uniform time variation of the current and βZ produces the linear space variation of the wave measured around the circumference of the ring. To make the ring circumference 1 wavelength and from definitions,

$$\lambda = 2\pi r, \quad \beta = \frac{2\pi}{\lambda}, \quad \text{and} \quad Z = r\theta. \quad (10-16)$$

The current may now be rewritten

$$I = \cos \omega t \cos \theta + \sin \omega t \sin \theta.$$

¹³B. H. Burdine and R. M. McElvery, *The Spiral Antenna*, Massachusetts Institute of Technology, Research Laboratory of Electronics, Report No. 1, March 15, 1955.

This represents two current standing waves in space and time quadrature. That is to say, the first has two current maximums on opposite sides of the radiating circle, and the second has current maximums where the first wave is zero. Further, the first standing wave achieves its maximum value in time when the second is at zero. The four current maximums at each quadrature of the circle may now be considered an array of four elementary dipole radiators. This convenient mathematical model of the spiral antenna radiates in a manner similar to the antenna itself, and thus its use has been justified experimentally. Using the four-element array set in a ground plane, the normalized radiation pattern will be found to be¹⁴

$$p = \frac{1}{4} \cos^2 \theta [1 + \cos(\sin \theta)]^2 \quad (10-17)$$

to a circularly polarized detector. For a linearly polarized detector, however,

$$p = \cos^2 \theta [\cos(\sin \theta)]^2 \quad (10-18)$$

when the pattern is measured perpendicular to the plane of the polarization of the detector and

$$p = \cos^2 \theta \quad (10-19)$$

when the pattern is measured in the plane of the detector polarization. These patterns give beamwidths of 77°, 68°, and 90° respectively. A somewhat improved model of the spiral may be had by replacing the short dipole radiator elements by half-wave dipole radiators. The effect is to reduce the 90° beamwidth of Equation 10-19 to 78°, which is closer to the actual measurements on the spirals.

Application. Electrical characteristics such as these find application to antennas for direction finding or possibly for semiactive missile guidance. As a specific example, a cluster of four nonmoving geometric spiral antennas could be used in the nose of a small semiactive, homing missile giving steering information in up-down, left-right coordinates. With a typical beamwidth of 70° and with rigidly mounted antennas, low-accuracy steering information is available from the antennas over plus or minus 70° of azimuth and elevation for proportional navigation, etc. Information from the antennas may be collected by standard monopulse or sequential lobing techniques.

10-9 RADOMES

One of the most limiting components in radars for high-speed aircraft has been the radio transparent window or *radome*. The system may suffer a loss in useful range because of a poor radome; equally important, the

¹⁴M. S. Wheeler, "An Electrically Scanned Spiral Antenna," *Fifth East Coast Conference on Aeronautical and Navigational Electronics*, p. 55.

radome may result in a loss of accuracy. The range loss is simply calculated from the percent loss in power that occurs when a radome is inserted between an antenna and a detector.¹⁵ Losses are typically between 10 and 30 per cent and are more or less independent of the angle of look through the radome. Let us glance first at the system problem.

The System Approach to Radome Design. The aforementioned radome losses degrade the range capability of the radar by 2 to 8 per cent. The loss in accuracy, on the other hand, arises primarily from refractions of the radar beam as it transits the radome at other than normal incidence. This causes the apparent line-of-sight to change, thereby introducing a static measurement error. Further, the amount and direction of refraction is not constant over the entire radome; thus when the radar is tracking a target, spurious angular rate signals will be produced by relative movement between the radome and the line of sight.

Other contributions to radar inaccuracy come from radome reflections affecting the transmitter frequency and from general deterioration of the antenna sidelobes. Paragraph 8-5 considered the effect of the transmitter frequency pulling problem on the design of the AFC. The deterioration of the sidelobes can similarly have a serious effect upon tracking and the altitude line. The system designer must remember this fact when he evaluates antenna patterns measured without the radome.

While it is possible in principle to compensate for known radome errors by suitable system design, this line has not been followed. Rather, the effort has been to reduce the errors below some tolerable figure. Depending upon the use to which the system is put, the tolerable errors may vary for different portions of the radome. For example, let us consider one facet of the radome design problem for the AI radar discussed in Chapter 2. During the tracking phase, the AI radar measures range, azimuth, and elevation lead angles, and the azimuth and elevation space angular rates of the radar line-of-sight to the target. The magnitude of the lead angle during the final phases of tracking may vary from near zero for head-on or tail-on attacks to 40-50° for attacks upon the target's beam. During steering, the pilot may roll the aircraft to bank angles exceeding $\pm 60^\circ$ at rolling rates up to 60°/sec. This rolling action causes the position of the radome relative to the antenna beam to change with a speed that is a function of rolling rate and lead angle. If the boresight error caused by the radome is variable, this variation, coupled with the rolling action, will cause spurious angular rate signals which degrade system accuracy and can also contribute to instability of the steering loop.

From this discussion, it can be seen that boresight error changes are likely to be more significant for large lead angles. Thus, to provide the

¹⁵Radar Design Criteria for Precision Guidance Radar, p. 77, Contract AF33(038)-12283.

radome designer with sufficient flexibility, the specification on boresight error shift should not be the same for the entire radome; rather, it should be matched to the tactical problem. The generation of radome requirements on bases such as these is a tedious and time-consuming system study task. However, it is easier than trying to design a complete radome to meet a set of rigid specifications that actually are necessary for only a small portion of the radome's area. It is becoming more and more necessary to approach the radome problem in this manner because aerodynamic considerations require radome shapes of increasingly poor geometrical properties so far as the transmission of electromagnetic energy is concerned. In fact, historically the radome problem has advanced somewhat as follows.

History of Development. The earliest radomes were dielectric material made of organic composition in a single half-wavelength layer. The form was hemispherical. This shape made possible good accuracy, since by reason of symmetry a hemisphere will not affect the antenna boresight. The half-wavelength thickness minimizes reflective losses at least one frequency, and a strong low-loss dielectric material need only be found.

Typically, radome material has a dielectric constant between 4 and 7 with a loss tangent less than 0.02. The principal radome problem is that strong dielectric materials inevitably have a high dielectric constant. Thus, the transition reflections in and out of the dielectric are large and the bandwidth is narrow. To increase bandwidth, various sandwich arrangements have been used with materials of differing dielectric constants to arrange for reflection cancellation at more than one frequency.¹⁶

At the same time, aircraft speeds increased, requiring a special high-density, protective coating on the outside surface of the radome to protect against rain erosion. Further, the hemisphere is no longer satisfactory aerodynamically, and progressively sharper structures were required. A typical radome for a transonic aircraft is shown in Figure 10-10-a.

Still it was found possible to maintain good system accuracy through radomes of conical symmetry by carefully and experimentally compensating the radome by large dielectric rings suitably fastened to the inside of the radome.¹⁷

At this point in the history comes the transition to missiles, which changes the requirements considerably. For aerodynamic reasons, the required shape is definitely ogival. Further, aerodynamic heating precludes the use of most organic plastics. In Fig. 10-10-b, the temperature expected at the nose of a missile is shown as a function of speed. Some help has come

¹⁶*Radome Design Criteria for Precision Guidance Radar*, p. 41, Final Report, Contract AF33(038)-12283. N. M. Weiderborn and A. F. Kay, McMillan Laboratory Report No. 1276.

¹⁷H. A. Schetne, "The Use of Dielectric Rings in Reducing Radome Error, *Proceedings of the OSU-WADC Radome Symposium*, Vol. II, June 1956.



FIG. 10-10-a A Typical Transonic Radome Resting on the Concrete Apron in Front of the F4D Aircraft, Showing the Fire-Control Radar and Antenna Exposed in the Nose Section.

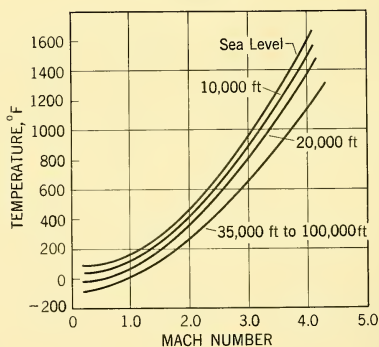


FIG. 10-10-b Temperatures Expected at the Tip of a Missile Radome as a Function of Velocity and Altitude.

however, in an easing of the electrical requirements. With a proportional navigation system such as is often used in missile guidance, the actual bore-sight error is not of consequence compared with the rate of change of apparent error with respect to angle.¹⁸ Practical limits for this rate of change of radome error are nearly $\pm 0.05^\circ$ per degree.

¹⁸*The Radome Problem in Guided Missiles*, Final Report, N60NR27023. March 1, 1952, Princeton University.

The high temperature problem for supersonic missile radomes has hastened the introduction of a new class of structural ceramics called Pyroceram by Corning Glass. This material is unlike glass in that it is crystalline and about twice as elastic as glass. Its hardness and softening temperatures correspond more to those of stainless steel, whereas its density is closer to that of aluminum. Its flexural strength is two-thirds that of steel. Electrically, its loss factor is good and its dielectric constant is 6. This dielectric constant is about midway between those of most of the organic plastic radome materials and of some of the ceramics that have been considered. As this is not one particular material, but rather a class of materials, it is to be expected that as time goes on, compositions suitable for a wide range of radome applications will be found.

10-10 INTRODUCTION TO TRANSMISSION LINES AND MODES OF PROPAGATION

Any physical separation of the RF units of a radar, such as that between the antenna and the transmitter, requires the use of an RF *transmission line* to convey the electromagnetic energy in an essentially lossless and radiation-free manner. Often more than one type of transmission line is employed in a single radar to achieve most effective design. Selection of

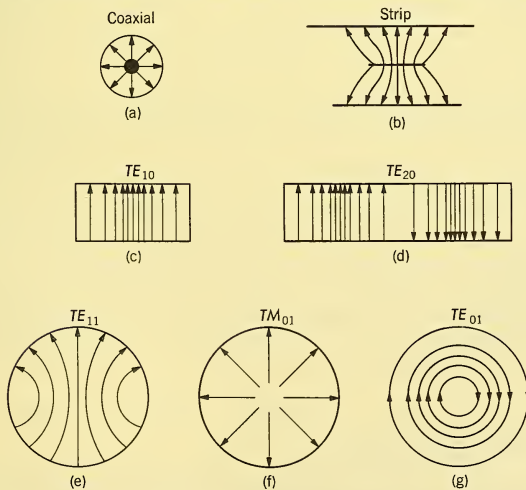


FIG. 10-11 Cross Sections of Various Transmission Lines, Showing Electric Field Configurations for (a) Coaxial, (b) Strip, (c) TE₁₀, (d) TE₂₀, (e) TE₁₁, (f) TM₀₁ Modes, and (g) TE₀₁ Modes.

the proper transmission lines requires familiarity with the system specifications, environmental requirements, and the characteristics of the various line types.

There are many types of RF transmission lines, differing basically in geometry and in mode of propagation. Fig. 10-11 shows cross sections of several types of transmission lines and the electric field configurations of representative modes.¹⁹ Each transmission line mode represents a different and unique distribution of the electric and magnetic fields of the propagating wave on or within the transmission line. There is no limit to the number of modes which can propagate in a transmission line if no upper frequency limit is specified. Transmission lines are generally operated in the *lowest-order mode*, however, and an upper frequency limit is specified to prevent propagation of energy in the *higher-order* modes. Higher-order mode and multimode transmission lines are often used to advantage in special applications.

10-11 TYPES OF TRANSMISSION LINES AND MODES OF PROPAGATION

Transmission line modes are of *four* general classes:²⁰ (1) the TE or *transverse-electric modes* (sometimes referred to as H waves) in which the *electric* field is everywhere perpendicular to the direction of propagation; (2) the TM or *transverse-magnetic modes* (sometimes referred to as E waves) in which the *magnetic* field is everywhere perpendicular to the direction of propagation; (3) the TEM or *transverse-electromagnetic modes* in which *both* the electric and magnetic fields are perpendicular to the direction of propagation of the wave; and (4) the HEM or *hybrid electromagnetic modes* having components of *both* electric and magnetic fields *in* the direction of propagation.

Electromagnetic energy always propagates at *group* velocity, which never exceeds c , the velocity of light. The wavelength λ_g in a transmission line may be greater than, equal to, or less than the free-space wavelength, λ , depending upon the type of transmission line, the mode of propagation, and whether the line is air or dielectric filled. The *phase velocity* v_p of a wave in a transmission line can be determined from a measurement of the wavelength in the line and the relation

$$v_p = \frac{\lambda_g}{\lambda} c \quad (10-20)$$

where c is the velocity of light.

¹⁹For a very descriptive discussion of waveguide modes see George C. Southworth, *Principles and Applications of Waveguide Transmission*, Chapt. 5, D. Van Nostrand Co., Inc., Princeton, N. J., 1950.

²⁰"IRE Standards on Antennas and Waveguides: Definition of Terms, 1953," (53IRE2.S1), *Proc. IRE*, **40**, No. 12, (December 1953).

In air or in air-filled transmission lines, the product of phase velocity and group velocity is constant and is equal to c^2 , i.e.

$$v_p v_g = c^2. \quad (10-21)$$

In a refractive medium or in a transmission line filled with a refractive medium this is modified so that

$$v_p v_g = \left(\frac{c}{\sqrt{\mu_r \epsilon_r}} \right)^2 \quad (10-22)$$

where μ_r is the relative *permeability* of the medium,

ϵ_r is the relative permittivity or *dielectric constant* of the medium,
and

$\sqrt{\mu_r \epsilon_r}$ is the *refractive index* of the medium.

The *propagation constant* of a transmission line is

$$\gamma = \alpha + j\beta \quad (10-23)$$

where α , the *attenuation constant*, expresses the loss per unit length of transmission line, and β , the *phase constant*, expresses the phase change per unit length of line. The phase constant β is related to the *phase velocity* by the equation

$$\beta = \frac{\omega}{v_p} \quad (10-24)$$

where ω is the angular frequency, $2\pi f$. In general, each of the foregoing parameters will have a different value for each mode and for each type of transmission line.

Coaxial Lines. Operating in the principal (TEM) mode (Fig. 10-11a), a coaxial transmission line is *nondispersive* and will transmit a wave of any frequency. Higher-order modes can propagate, however, when the mean internal circumference of the coaxial line is greater than a wavelength. As one approaches microwave frequencies, coaxial lines become less practical because of increased losses and reduced power-handling capability incident to reduced dimensions. Consequently, coaxial transmission lines find greatest application below 3000 Mc.

Rectangular Waveguides. The most important line for microwave use, the rectangular waveguide operating in the TE_{10} mode (Fig. 10-11c), is a *high-pass* transmission medium having a low-frequency cutoff characteristic which is governed by its greater width a , as illustrated in Fig. 10-12. The height b of the waveguide has no influence on the cutoff wavelength of the dominant mode, but does affect impedance, breakdown power, and transmission loss. For each possible mode of propagation in a waveguide

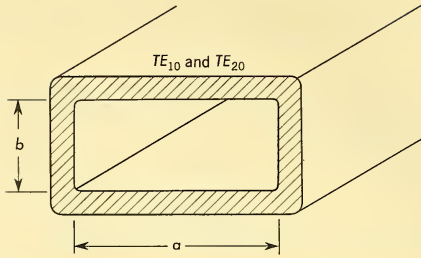


FIG. 10-12 Cross Section of Rectangular Waveguide, Showing Width a Which Governs Cutoff Wavelength.

there is a definite *cutoff wavelength*. At wavelengths greater than cutoff for a particular mode, the transmission line will not propagate energy in that mode.

The lowest-order mode or *dominant mode* is that waveguide mode having the longest cutoff wavelength (or lowest cutoff frequency). In rectangular waveguide this is the TE_{10} mode (Fig. 10-11c), which has a cut-off wavelength λ_c of $2a$ (Fig. 10-12). The TE_{20} mode (Fig. 10-11d) is the next

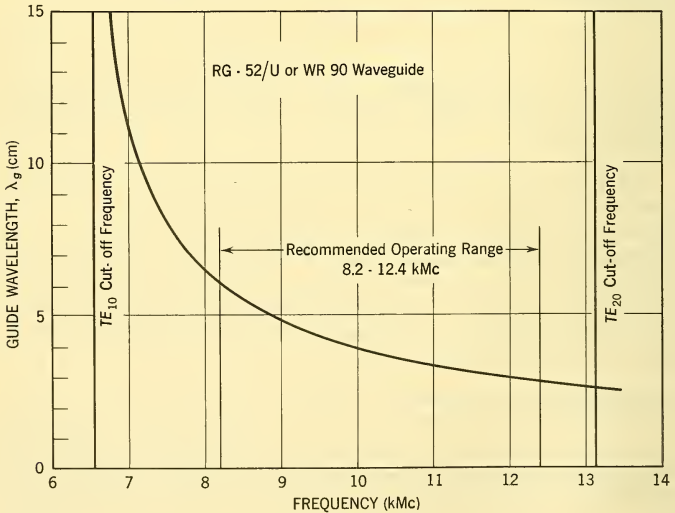


FIG. 10-13 Plot of Guide Wavelength vs. Frequency for RG-52/U or WR90 X-Band Waveguide, Showing Cutoff Frequencies for TE_{10} and TE_{20} Modes.

higher-order mode and has a cutoff wavelength one-half that of the TE_{10} mode. Thus there is a one-octave frequency range between cutoff points for the two modes. The recommended operating range is appreciably less than this, however, as illustrated by Fig. 10-13, which shows cutoff frequencies and guide wavelengths for representative X-band rectangular waveguides. Typical recommended bandwidth for rectangular waveguide operating in the dominant mode is about 40 per cent, and over this range loss *decreases* as frequency increases.

Ridge Waveguides. For very wide band or multiband operation, *ridge waveguide* has particular utility. Cross sections of two types of ridge waveguide are shown in Fig. 10-14. High-power ridge waveguides are

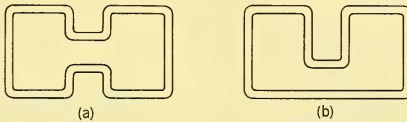


FIG. 10-14 Cross Sections of (a) Double-Ridge and (b) Single-Ridge Wideband Waveguides.

available covering a 2 or 3-to-1 frequency range, although frequency ranges of 4 to 1 or more are possible in ridge lines having reduced power-handling capacity. A representative commercial ridge waveguide covers the frequency range from 4700 to 11,000 Mc, a range which could be covered with a minimum of three standard rectangular waveguides.

Circular and Square Waveguides. Being axially symmetrical, circular waveguides are able to play certain unique roles. There are three circular waveguide modes of particular practical importance. These are the TE_{11} "linearly" polarized dominant mode, the TM_{01} circular magnetic mode, and the TE_{01} circular electric mode (Fig. 10-11e, f, g). The TM_{01} mode finds its principal application in rotary joints. The TE_{01} mode is unique in that for a given waveguide size, loss decreases without limit as frequency is increased, and an oversize guide can provide very low loss over appreciable transmission distances.

Circular waveguide finds its main application in the TE_{11} mode. The polarization of a circularly-symmetrical antenna feed can be controlled simply by orienting the TE_{11} mode in the connecting circular waveguide, or it can be excited with either sense of circular polarization. A similar feed with its radiating end deflected slightly from the antenna axis can be rotated about the antenna axis to generate a nutated conical scanning action with unvarying yet controllable polarization. Mode multiplexing is feasible using two TE_{11} modes which are orthogonally polarized in circular waveguide.

In many applications, square waveguide, employing orthogonal TE_{10} modes, can be used instead of circular waveguide, with the advantage of greater bandwidth and reduced problems in mode coupling and ellipticity. It should be appreciated that circularly polarized waves *can* propagate in square waveguide. Faraday rotation-type microwave ferrite devices, for example, which depend on circularly polarized waves for their operation, are often designed in either circular or square waveguide. These devices are described in more detail later in this chapter.

Strip Lines. Strip-type transmission lines offer many possibilities for reduction in size, weight, and cost of microwave assemblies. Two types of strip lines are illustrated in Fig. 10-15. In other types the strip conductor

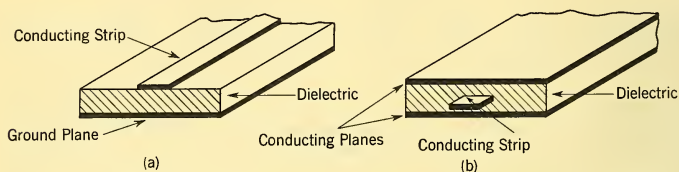


FIG. 10-15 Two Types of Strip Transmission Lines.

may be principally air supported, or supported by a thin dielectric sheet in the plane of the strip conductor. Certain types are especially adaptable to printed-circuit production techniques like photoetching.

In general, strip lines have higher loss and lower power-handling capabilities than conventional waveguides and generally require a shielding enclosure to prevent leakage and crosstalk. Current applications are to microwave filters, receiver heads, printed antenna arrays, and in miniaturized microwave assemblies. Strip lines are undergoing intensive research,²¹ while present applications require custom development of the product.

10-12 STANDING WAVES AND IMPEDANCE MATCHING

The problem of standing waves universally arises in any dealing with practical RF transmission line systems. This is particularly true in the complex antenna and plumbing assemblies of modern radar systems, where numerous circuit elements are involved. The presence of standing waves in a transmission line system is generally undesirable and often harmful to the proper operation of the system. Standing waves, caused by reflections from impedance discontinuities somewhere along the transmission line, are a simple example of interference phenomena between two coherent waves

²¹"Symposium on Microwave Strip Circuits," *IRE Trans. MTT-4*, No. 2 (March 1956).

traveling in opposite directions. It is the periodic constructive and destructive interference between the incident and reflected waves which creates the standing wave pattern.

Minima occur every half wavelength where the incident and reflected waves are out of phase and subtract ($E_i - E_r$), while maxima occur where the two waves are in phase and add ($E_i + E_r$). The voltage standing wave ratio (VSWR) is the ratio of these two quantities:

$$\text{VSWR} = \frac{E_i + E_r}{E_i - E_r} = \frac{E_{\text{max}}}{E_{\text{min}}} \quad (10-25)$$

where E_i and E_r are the voltage amplitudes of the incident and reflected waves respectively. The relative position of a voltage minimum is a function of the phase angle of the reflected wave, which is determined by the complex impedance of the mismatch. When the transmission line is perfectly matched and terminated, there is *no* reflected wave and hence no standing wave, and the VSWR reduces to unity. For the case of *total reflection*, $E_r = E_i$, and VSWR is infinite, meaning zero voltage at the standing wave minima, while the voltage maximum is $2E_i$. This condition is approximated in a short-circuited transmission line.

It is often more convenient to measure and express the condition of mismatch in terms of a *reflection coefficient*. The magnitude of the voltage reflection coefficient is simply the ratio of the reflected voltage to the incident voltage, or $\rho = \frac{E_r}{E_i}$. The voltage standing wave ratio and the voltage reflection coefficient are thus related by the following equations:

$$\text{VSWR} = \frac{1 + \rho}{1 - \rho}, \quad \rho = \frac{\text{VSWR} - 1}{\text{VSWR} + 1}. \quad (10-26)$$

Unless kept to a very low magnitude, standing waves can significantly affect proper operation of a microwave radar. The reflected signal responsible for the standing wave represents RF power flowing back toward the source. If the voltage reflection coefficient is known, the relative power reflected is simply ρ^2 , as the reflected power is proportional to the square of the reflected voltage. When transmitting, this is power which does not reach the antenna. Fig. 10-16 shows the amount of power lost as a function of the voltage reflection coefficient. Good system design practice calls for an overall VSWR not to exceed 2 : 1 ($\rho = 0.33$). Even this can represent an 11 per cent reflected power loss, both on transmitting and receiving. Power handling capacity of a transmission line is reduced under standing wave conditions because of the increase in voltage at the standing wave maxima. The voltage maximum is increased by the factor $(1 + \rho)$.

A more frequently encountered effect of standing waves in microwave systems is their reaction upon the RF generator. Magnetrons in particular

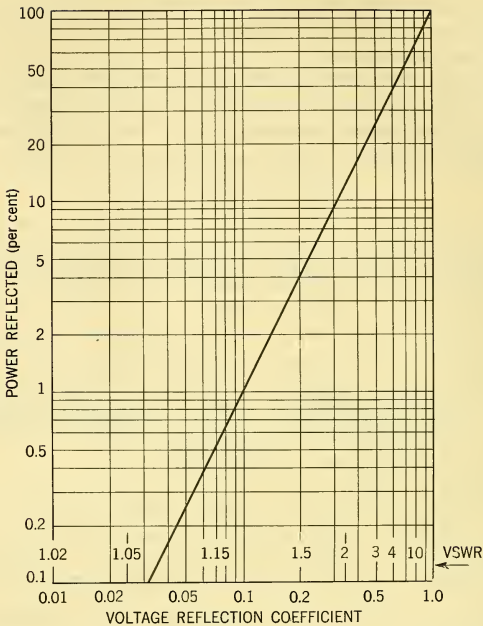


FIG. 10-16 Plot of Power Reflected as a Function of the Voltage Reflection Coefficient of a Mismatched Transmission Line.

are affected in their operation by energy reflected back to them. Frequency of oscillation, power output, and frequency stability are all functions of the phase and amplitude of reflected waves. Long transmission lines and the changing impedance of scanning antennas present particular problems in this respect. The design of Automatic Frequency Controls (AFC) is strongly affected by these problems, as was indicated in Paragraph 8-5.

It is important to realize that standing waves can exist at the input of a transmission line even though it is properly terminated at the far end. Any irregularity or abrupt change in dimension of a transmission line can be the source of undesirable reflected waves. Poorly brazed or soldered waveguide joints, for example, are often the source of objectionable reflections.

A frequently encountered matching problem arises when two transmission lines of different impedances must be connected together. Such a transition is most readily accomplished by using a *quarter-wave transformer*

between the two different transmission lines. The transformer is simply a quarter wavelength of transmission line having a characteristic impedance which is the geometric mean of the characteristic impedances of the two transmission lines to be joined. Thus $Z_t = \sqrt{Z_1 Z_2}$, where Z_1 and Z_2 are the characteristic impedances of the two lines to be joined, and Z_t is the characteristic impedance of the quarter-wavelength section of line used for the transformer. A simple quarter-wave transformer can provide a perfect match at a single frequency, but is not broadband. A series of quarter-wave transforming sections, successively differing in impedance according to formula, are often used to obtain greatly increased bandwidth.²²

In addition to the various stubs and impedance transformers employed in matching waveguide systems, an assortment of obstacles have been adapted to this purpose. Such devices as ball bearings, dents, screws, pins, posts, irises, and windows are being used. While all of these elements are reactive and lossless, the inductive iris illustrated in Fig. 10-17 is generally preferred, since it does not reduce breakdown power of the line. For low-power applications the flexibility of the capacitive tuning screw is convenient.

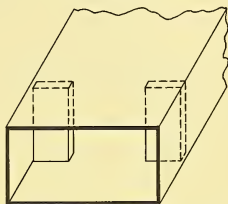


FIG. 10-17 Symmetrical Inductive Iris in Rectangular Waveguide.

Matching elements, discontinuities in themselves, are so designed and so positioned on the transmission line as to set up a reflected wave of such phase and magnitude as to cancel the wave reflected from the offending discontinuity. The result is the elimination of standing waves on the generator side of the matching element. A standing wave will still exist between the matching element and the offending discontinuity, however. Best bandwidth will be realized when the matching element is placed as close as possible to the mismatch.

Matching elements, discontinuities in themselves, are so designed and so positioned on the transmission line as to set up a reflected wave of such phase and magnitude as to cancel the wave reflected from the offending discontinuity. The result is the elimination of standing waves on the generator side of the matching element. A standing wave will still exist between the matching element and the offending discontinuity, however. Best bandwidth will be realized when the matching element is placed as close as possible to the mismatch.

10-13 BROADBAND SYSTEM DESIGN

Increased concern over radar countermeasures, and the need for greater flexibility of airborne radar systems as the number of equipments operating in a given band increases, is forcing a continual extension of the frequency band over which systems must be able to operate. Advances in microwave power generators, both mechanically and electronically tuned (see Chapter 11) are paving the way for improved system bandwidth capability. Fixed-

²²Seymour B. Cohn, "Optimum Design of Stepped Transmission Line Transformers," *IRE Trans.*, **MTT-3**, No. 3, 16-21 (April 1955).

frequency radar systems may become largely a thing of the past for military use.

Special care is required in the design and selection of antennas and components for broadband systems. Each separate element and component of the system must be individually broadband matched to a high degree of excellence to ensure satisfactory broadband operation of the complete microwave assembly. If the individual elements of a transmission system are not in themselves well matched over a broadband, then the prospect of matching the assembled system over a broadband is poor. Several components, individually well matched, when cascaded in a microwave system can create an objectionably high standing wave ratio at certain frequencies if the reflection coefficients of the several elements add in phase.²³ Thus the quality of components for broadband systems, particularly systems employing numerous components, must be exceptionally good to avoid wide fluctuations of reflection coefficient with change in operating frequency. In the worst case, however, the highest reflection coefficient for the assembled system will not exceed the arithmetic sum of the individual voltage reflection coefficients.

Components, to be broadbanded, basically should not involve severe impedance transformation. Where the transformation is large, either a gradual tapered transformation or a well-designed stepped-type transformation is required for broadbanding. Techniques for matching components with irises are well developed,^{24,25} but as a simple first principle, for best broadband performance, the matching element should be located as close as possible to the component to be matched.

There are some important principles to follow in the design of broadband systems: (a) The number of transmission line components used should be kept to a minimum. A proposed layout should be critically reviewed to eliminate superfluous bends, twists, transitions, flange joints, etc. (b) Components used should be individually well matched. (c) Whenever possible, the elements of a transmission circuit should be grouped closely together with a minimum of transmission line between components. This will aid in matching the system as a whole. (d) Subassemblies of groups of elements should be broadband matched before final assembly in the complete system.

²³Richard K. Moore, "The Effects of Reflections from Randomly Spaced Discontinuities in Transmission Lines," *IRE Trans. MTT-5*, No. 2, 121-126 (April 1957); J. A. Mullen and W. L. Pritchard, "The Statistical Prediction of Voltage Standing-Wave Ratio," *IRE Trans. MTT-5*, No. 2, 127-130 (April 1957).

²⁴R. G. Fellars and R. T. Weidner, "Broadband Waveguide Admittance Matching by Use of Irises," *Proc. IRE* **35**, 1080-1085 (1947).

²⁵Peter A. Rizzi, "A Low VSWR Matching Technique," *IRE Trans. MTT-4*, No. 3, 185-186 (July 1956). Supplementary comments in the *IRE Trans. MTT-5*, No. 2, 163 (April 1957).

The consequences of operating a system whose reflection coefficient is a sensitive function of frequency include frequency selectivity of the transmission line causing spectrum distortion of pulse transmission, variation in pulling effect as frequency is changed, variation in transmission loss, and the possibility of resonances which might cause magnetron moding. High VSWR resonances, which may be overlooked in spot frequency measurements of VSWR, are sometimes responsible for unaccountable breakdown in high-power transmission line systems. The testing method specified for broad-banded systems must be designed to avoid this possibility. A swept-frequency reflectometer²⁶ is invaluable for revealing resonances of this nature.

Many broadband radar systems can operate over a 10 per cent band or more, but considerable effort is being directed toward development of very wideband components which will permit design of systems to operate over the full 40 per cent band of a dominant mode rectangular waveguide. Operating bandwidths even greater than this are possible with strip line and ridge waveguide.

10-14 PRESSURIZATION

The microwave plumbing of airborne radar systems is generally pressurized for two important reasons: first, to prevent RF breakdown²⁷ at altitude pressures, and second, to eliminate problems of moisture condensation and corrosion in the transmission line due to breathing. To completely seal the microwave assembly involves capping the antenna feed in some manner with a dielectric cover or window, which then becomes an integral part of the feed itself and is involved in any measurement of the electrical characteristics of the antenna feed. Pressure-sealing gaskets are required at each transmission line coupling, and all branch lines must be sealed with some form of pressurizing window. A TR tube, for example, can serve this function if properly gasket-sealed. Most magnetrons and some of the newer local oscillator klystrons also can be pressure-sealed to the waveguide system. A wide selection of broadband windows are commercially available for pressurizing waveguide systems. Most of these are rated for a differential pressure of 30 psi or more, when mounted with the glass seal on the high-pressure side. Where breakdown power level is not raised sufficiently by pressurization alone, sulfur hexafluoride gas²⁸ (SF₆) has been used to raise significantly the power-handling capability of

²⁶N. L. Pappas, "Measuring Microwave Impedance with the Reflectometer," *Tele-Tech* 15, No. 72 (May 1956).

²⁷Lawrence Gould, *Handbook on Breakdown of Air in Waveguide Systems*, Microwave Associates, Inc., Boston, 1956.

²⁸*Sulfur Hexafluoride*, Technical Bulletin TB-85602, General Chemical Div., Allied Chemical and Dye Corp., Philadelphia, 1955.

a microwave transmission line. Pressurization has become an important factor in the reliability of airborne radar systems, for seal rupture could completely disable a system at high altitude. Dielectric-filled waveguides, proposed as a solution to the problem, are under development.²⁹

10-15 MISCELLANEOUS MICROWAVE COMPONENTS

As the elements which compose microwave assemblies and systems, microwave components encompass many forms and functions. Certain components serve only to change transmission line direction, while others perform some transformation of the electrical description of the propagated wave. A limited number of the more important components will be discussed. The reader is urged to consult the references and bibliography for greater coverage of this extensive topic.



FIG. 10-18 Waveguide "Twist-Turn" Simultaneously Executes 90° Twist and 90° Turn. (Kearfott Company, Inc., Western Division)

Bends, Twists and Rotary Joints. The compactness required of airborne systems may invite generous use of waveguide bends. As pointed out earlier in this chapter, in the interest of broadband performance, their number should be kept to a minimum, as each will introduce a finite reflection. Whether a long, fabricated bend is better than a compact, die-cast bend will depend more on the relative quality of the components than on the radius of the bend. At slight disadvantage reflectionwise, the precision-cast bends permit very direct and abrupt changes in waveguide direction. The novel component shown in Fig. 10-18, called a twist-turn, simultaneously executes a 90° twist and a 90° bend.

Waveguide twists often are fabricated from a portion of the waveguide run by relatively simple shop techniques. This results in a rather long section of waveguide devoted to the twist. A significant reduction in length is possible using *step-twist* techniques.³⁰ Very effective wideband step

²⁹Tore N. Anderson, "Practical Dielectric Filled Waveguide," *IRE Trans. MTT-3*, No. 2, 82-86 (March 1955).

³⁰H. A. Wheeler and H. Schwiebert, "Step Twist Waveguide Components," *IRE Trans. MTT-3*, No. 5, 44-52 (October 1955).

twists are feasible in surprisingly short lengths. A typical 3-face, 90° step twist for X-band is only 1 inch long, yet has a very low VSWR over a 12 per cent band.

The step-twist principle has been applied to an interesting variable twist useful as an oscillating rotary joint having both ports on the axis of rotation. The commercial unit shown in Fig. 10-19 will rotate $\pm 90^\circ$, with a VSWR of under 1.15 over the full 40 per cent waveguide band.

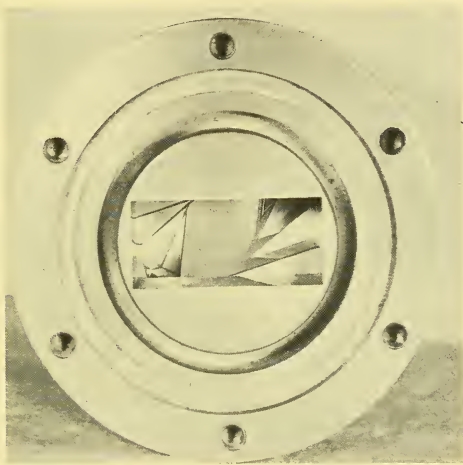


FIG. 10-19 Step-Twist Rotary Joint Permits $\pm 90^\circ$ Rotation of a Waveguide About Its Own Axis. (Courtesy of Lieco Incorporated)

The more conventional waveguide rotary joints, capable of continuous rotation, involve a mode conversion from rectangular waveguide to an axially symmetrical mode and back to rectangular waveguide again. Modes commonly used are the TEM coaxial mode, the TM_{01} circular magnetic mode, and the TE_{01} circular electric mode (Fig. 10-11a, f, g). Such rotary joints commonly have both rectangular waveguide ports extending radially, or one radially and the other axially. Although rotary joints vary considerably in bandwidth and performance, units having a VSWR under 1.15 over a 12 per cent band are now common. Disturbing resonances, occasionally found to occur in rotary joints, are best detected with a swept-frequency reflectometer while slowly rotating the joint through its full range. Spot frequency checks of VSWR seldom reveal such resonances.

Directional Couplers. A directional coupler is a device for coupling signals directionally between two transmission lines, or between a *main guide* and an *auxiliary guide*, with a fixed power division ratio. Referring to Fig. 10-20, power entering the main guide at A divides between the main guide and the auxiliary guide. The ratio between the power input and the power coupled to the auxiliary arm C, expressed in db, is called the *coupling* of the device. Thus a directional coupler in which 1 per cent of the input power to the main guide reaches the auxiliary guide has a coupling of 20 db. The balance of the input power continues down the main guide. The *directivity* of the coupler is the ratio, expressed in decibels, of the auxiliary guide power flowing to C and that reaching D, with input at A. To measure directivity accurately, both B and C must be terminated with reflectionless loads. Arm D of the directional coupler is normally terminated in a well-matched load.

The directional coupler³¹ is often used as a power divider and for injecting signals into the main guide through the padding or isolation provided by the directional coupler. A more frequent use is for sampling or monitoring power flow in the main guide. Thus if a transmitter is attached to A and an antenna at B (Fig. 10-20), the amount of power flowing from A to B can

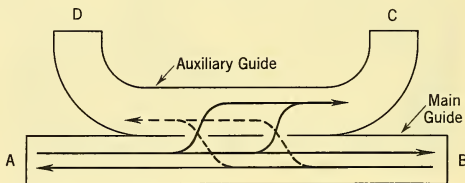


FIG. 10-20 Waveguide Directional Coupler.

be monitored by a detector on C. If the detector is instead attached to arm D, it will sample the power reflected from the antenna on B to give a relative indication of the condition of match of the antenna. Directional couplers in which coupling ranges from 3 db to 30 db or more are most common, with directivity of up to 40 db possible in high-quality couplers.

A *bidirectional coupler* is a dual directional coupler which permits simultaneous sampling of signal flow in both directions in the main guide. A bidirectional coupler is often used in reflectometer setups.

Hybrid Junctions. The term *hybrid junction* implies a multiport transmission line component in which certain ports couple in common to

³¹For a detailed discussion of directional couplers see C. G. Montgomery, *Technique of Microwave Measurements*, Chapt. 14, McGraw-Hill Book Co., Inc., New York, 1947.

other ports, yet are isolated from each other. The four-port hybrid junctions³² are most common and include the *magic-tee*, *ring hybrid* or *ratrace*, *short-slot hybrid*,³³ and the *folded-tee hybrids*. All of these hybrid junctions have the common property of dividing input to one port equally between two other ports, while the fourth port remains isolated from the first. Because of this characteristic of effecting equal power division, they are classed as *3-db hybrids*. As such, they also function as 3-db directional couplers.

The waveguide magic-tee junction shown in Fig. 10-21 will be used to illustrate the action of a four-port hybrid. A signal entering port 1 of this junction will divide *equally* and *in phase* between *collinear* arms 3 and 4, as arm 1 is *in shunt* with these two arms. A signal entering port 2 will also divide equally between arms 3 and 4, but with components which are *out of phase*, as arm 2 is in series with arms 3 and 4. Arms 1 and 2 are isolated from each other because of orthogonality of their respective modes.

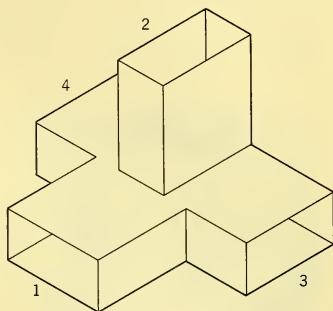


FIG. 10-21 The Waveguide Magic Tee, a 3-db Hybrid Junction.

In a properly designed magic-tee, all ports are matched and arms 3 and 4 are also isolated from each other. Isolation between ports of a hybrid is generally realized, however, only when all other ports are terminated in matched loads. The isolation which one measures between ports of a hybrid junction is, in fact, affected by the terminations attached to the other ports of the junction. In commercial units, isolations range from 20 to 40 db.

The *folded-tee hybrids* are functionally similar to the junction just described except that the "collinear" arms (arms 3 and 4 in Fig. 10-21) are folded in the *H* plane in one type, and in the *E* plane in a second type to obtain a physical arrangement more convenient for many applications. A commercial *H*-plane folded tee is shown in Fig. 10-22.

The *balance* of a hybrid junction, which is a measure of the equality of power division expressed in db, is of importance in many precision applications as in monopulse antenna feeds and in microwave bridges. The quality of balance of various commercial hybrid junctions ranges from ± 0.25 to ± 0.1 db for a 10 to 15 per cent bandwidth.

³²W. A. Tyrrell, "Hybrid Circuits for Microwaves," *Proc. IRE* **35**, 1294-1306 (1947).

³³H. J. Riblet, "The Short-Slot Hybrid Junction," *Proc. IRE* **39**, 180-184 (1952).

The uses of 3-db hybrid junctions are many in radar systems,³⁴ from their use as a 3-db directional coupler and for power splitting, to deriving "sum" and "difference" signals in the microwave bridge circuits of monopulse antenna feeds. The use of 3-db hybrids in balanced mixers and duplexers will be described later.

Many other multiport hybrid junctions have been developed or devised for special application. Among these are two of particular interest for circularly polarized systems: the *turnstile* waveguide junction^{35,36} and the *trimode turnstile* waveguide junction.³⁷ The latter junction, shown schematically in Fig. 10-23, has particular utility in a compact dual balanced mixer for circularly polarized applications.³⁸

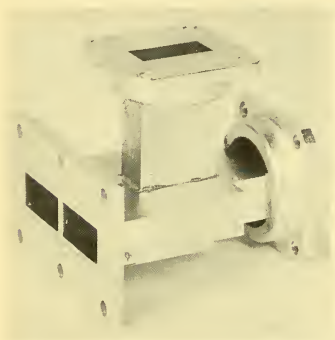


FIG. 10-22 The Hughes H-Plane Folded Hybrid Tee, a 3-db Hybrid Junction. (Microwave Development Laboratories, Inc.)

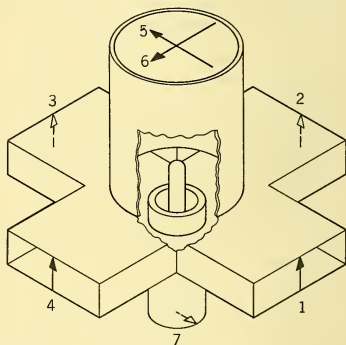


FIG. 10-23 The Trimode Turnstile Junction, a Symmetrical Seven-Port Hybrid Junction.

The Mixer. The purpose of the *mixer*³⁹ in a radar receiver is to convert the incoming microwave signal to an intermediate frequency (IF) signal of similar character which can be amplified using low-frequency techniques. This process is accomplished by mixing the received signal

³⁴J. W. Sutherland, "Waveguide Hybrid Circuits and Their Use in Radar Systems," *Electronic Eng.* **28**, 464-469 (1956).

³⁵M. A. Meyer and H. B. Goldberg, "Applications of the Turnstile Junction," *IRE Trans. MTT-3*, No. 6, 40-45 (December 1955).

³⁶R. S. Potter and A. Sagar, "A New Property of the Turnstile Waveguide Junction," *1957 Proc. Natl. Electronics Conf.* **13**, 452-458 (1958).

³⁷R. S. Potter, "A Trimode Turnstile Waveguide Junction," *IRE Convention Record* **4**, Part 5, 36-43 (1956).

³⁸P. J. Allen and R. D. Tompkins, "An Instantaneous Microwave Polarimeter," *Proc. IRE* **47**, 1231-1237 (1959).

³⁹R. V. Pound, *Microwave Mixers*, McGraw-Hill Book Co., Inc., New York, 1948.

with a local oscillator signal in a crystal diode to obtain the IF or difference frequency signal which then passes to the IF amplifier.

The ultimate sensitivity of a radar receiver is limited not by the gain of the receiver, but by the magnitude of losses and noise power present in the input circuitry. The overall receiver *noise figure* is given by the expression

$$N_r = L_c(N_{if} + N_e - 1) \quad (10-27)$$

where N_r is the receiver noise figure

L_c is the crystal diode conversion loss

N_{if} is the IF amplifier noise figure

and N_e is the crystal diode noise ratio or noise temperature, all expressed as power ratios.

The receiver noise figure is usually expressed in db, and as such is

$$\text{Noise figure (db)} = 10 \log N_r \quad (10-28)$$

This represents the best noise figure which theoretically can be realized with the given values of L_c , N_{if} , and N_e . There are factors not included in this definition of noise figure which can deteriorate the receiver noise figure realized in practice. Thus any loss between the antenna and the mixer will reduce practical receiver sensitivity a corresponding amount; hence, duplexer losses, etc., should be kept to a minimum. A further deterioration of the practical noise figure can be caused by noise sidebands present in the local oscillator (LO) output. In the simple single-diode mixer, this klystron noise produces extraneous noise components in the IF output of the mixer which result in a deterioration of receiver noise figure. Much of this noise could be eliminated by use of a very narrow-band filter between klystron and mixer, although this is not generally practical. The preferred method of eliminating the effect of local oscillator noise is to use a balanced mixer.

Balanced mixers commonly employ some form of 3-db hybrid junction and two crystal mixer diodes in a manner which permits the desired IF output signals to add, while the klystron noise components cancel by subtraction.⁴⁰ Most common are the short-slot hybrid mixers,³³ the folded-tee mixers, and the magic-tee mixers. Fig. 10-24 illustrates a balanced mixer employing an *H*-plane folded hybrid tee. In the mixer shown, the local oscillator signal divides equally in the hybrid and arrives at the crystals *in phase*. The input signal, isolated from the local oscillator, also divides equally, but arrives at the crystals *out of phase*. By using a push-pull IF input and reversed balanced crystals in the mixer, the local oscillator noise components will cancel in the mixer output circuit. A variety of input, output circuit, and crystal arrangements are possible in balanced mixers.

⁴⁰L. D. Strom, "Noise Cancellation in Microwave Mixers," *Tele-Tech* **15**, No. 3 86ff. (March 1956).

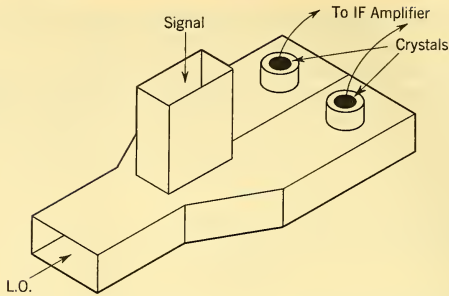


FIG. 10-24 Balanced Mixer Using Folded-Tee Hybrid Junction.

The principal advantage expected from using a balanced mixer is an improvement in receiver noise figure through reduction or elimination of local oscillator noise. The improvement in receiver noise figure which can be realized from a balanced mixer depends, of course, on the klystron noise level and on the degree of noise suppression achieved, which is dependent upon the quality of balance achieved in the mixer, although adequate suppression can be obtained with moderate crystal unbalance. Balanced pairs of crystals should always be used in balanced mixers, and replaced with a balanced pair when change is necessary. Although in a well-balanced mixer the local oscillator noise suppression may be near total, the improvement to be expected in receiver noise figure is of the order of 2 db at X band. Greater advantage can be realized at higher frequencies because local oscillator noise increases with frequency. Certain secondary advantages are realized from use of a balanced mixer, including reduced local oscillator power requirement.

Generally it is desirable in a mixer that reflection of any input signal should be back out the port through which it entered. Among the reasons for this is the prevention of local oscillator radiation through the antenna, and variations of crystal current which would arise with variations in antenna impedance. Also, local oscillator power reflected back into the signal input of the mixer introduces local oscillator noise to the signal channel; hence some of the noise-suppressing advantage of the balanced mixer is lost. The symmetrical hybrid tee mixers have the desired characteristics while the ordinary slot hybrid mixers do not. In these, signal reflections enter the LO channel, while LO reflections travel toward the antenna by the signal channel. In many applications this may be of no consequence, while in others it may be of importance. LO radiation and effects of antenna variations can be significantly reduced, if necessary in such instances, by employing a low-level ferrite isolator (see Paragraph

10-16) in the signal input arm of the mixer. Because of the bandwidth advantage of the slot hybrid it has been proposed that slot hybrid mixers incorporate a broadband 90-deg phase shift section in one crystal arm to obtain the preferred characteristic.

As balanced mixers for the higher frequencies are readily available commercially, and as they are often more convenient to use than the single crystal mixer, they are finding rather general application in new system designs. Depending principally on how well crystal balance is maintained, the noise figure improvement may or may not be realized. Improvements in crystal diodes have resulted in measurable reductions in crystal conversion loss (L_c) and in crystal noise ratio (N_c) such that an overall receiver noise figure of 8 db or less should be realizable at X band. Complete cancellation of local oscillator noise is assumed, however.

Microwave Noise Sources. Unless some means of checking is provided, the operator of a radar may be unaware of a deterioration in system performance, particularly if it is due to loss in receiver sensitivity. One of the simplest ways of providing a quick overall check on receiver performance is with a built-in calibrated *microwave noise source*, which serves as a standard signal with which an absolute measure of receiver sensitivity can be made with reasonable accuracy. Such noise sources, which are special gas discharge tubes in a section of transmission line, are now available in compact form for incorporation as a component of the radar system. These discharge tubes typically generate an excess noise power of 15 to 18 db above random noise power at room temperature, depending on the kind of gas fill in the tube. A simple comparative measurement of receiver noise output level, with and without the noise tube turned on, provides a valuable check of receiver sensitivity. Fig. 10-25 illustrates how one type of gas tube noise source can be inserted directly between the duplexer and the mixer;

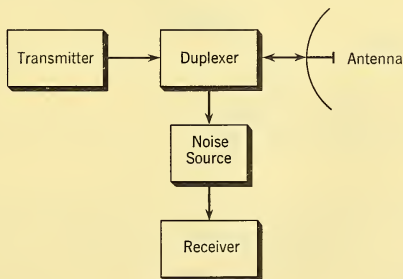


FIG. 10-25 Simplified Block Diagram of Radar Incorporating Double-Ended Noise Source for Monitoring Receiver Noise Figure.

in other instances the noise power is introduced at reduced level through a directional coupler.

Components for Circular Polarization. Circularly polarized waves are used in many applications from microwave ferrite devices to radars which discriminate against rain.⁴¹ There is nothing circular about a circularly polarized wave, however. A circularly polarized wave is one in which the electric polarization rotates at the carrier frequency, and may be either right- or left-handed, depending on direction of rotation of the electric vector. A circularly polarized wave is equivalent to two linearly polarized waves differing in phase by 90° and polarized at right angles to each other, and can be resolved into two such waves. Thus, two identical TE_{11} waves, differing in phase by 90° and orthogonally polarized in a circular waveguide, will result in a circularly polarized wave.

Quarter-wave plates are frequently used to generate circularly polarized waves. These elements, also called $\Delta 90^\circ$ plates, produce a phase delay in one plane of polarization which is 90° greater than in the orthogonal plane of polarization. Microwave quarter-wave plates take several forms. One of the most popular, the dielectric quarter-wave plate,⁴² consists of a longitudinal dielectric slab in a section of dominant mode circular (or square) waveguide. The dielectric slab introduces a phase delay to a wave in its plane which is just 90° greater than the phase delay undergone by a wave polarized at right angles to the first. As shown in Fig. 10-26, a dielectric quarter-wave plate oriented at 45° relative to an incident linearly polarized wave will introduce a 90° differential phase delay between the component in the plane of the dielectric plate and the component at right

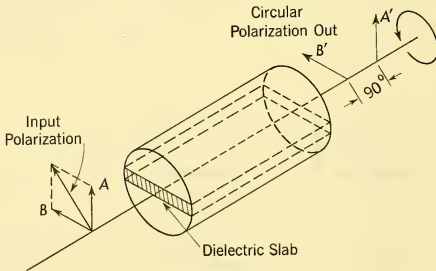


FIG. 10-26 Dielectric Quarter-Wave Plate Used for Converting Between Linear and Circularly Polarized Waves.

⁴¹P. A. Crandell, "A Turnstile Polarizer for Rain Cancellation," *IRE Trans. MTT-3*, No. 1, 10-15 (January 1955).

⁴²R. M. Brown and A. J. Simmons, *Dielectric Quarter-Wave and Half-Wave Plates in Circular Waveguide*, NRL Report 4218, Washington, D.C., November 1953.

angles to the dielectric plate. As a consequence, the emerging wave is circularly polarized. The *sense* of circular polarization will be reversed if the quarter-wave plates is rotated 90° from its indicated position. Other forms of microwave quarter-wave plates employ elliptical waveguide, periodically loaded square waveguide,⁴³ and capacitive pins in circular waveguide.⁴⁴

Half-Wave Plates, or $\Delta 180^\circ$ plates, resemble the quarter-wave plates in construction but introduce a 180° phase delay difference. Two quarter-wave plates having the same orientation are equivalent to a half-wave plate. Because of the 180° differential phase delay introduced by a half-wave plate, it will reverse the sense of rotation of a circularly polarized wave passing through it, and if rotated about its axis will produce a continuous change in phase of the circularly polarized wave, advancing or retarding the phase, depending on direction of rotation. Such a device is often used as a *continuous phase shifter*,⁴⁵ although generally in conjunction with two quarter-wave plates — the first to convert from linear to circular polarization, and the second to convert back to linear polarization. The construction of a commercial continuous phase shifter⁴⁶ of this type is shown in Fig. 10-27.

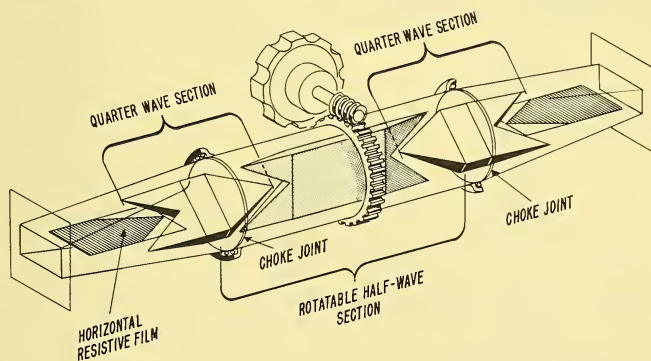


FIG. 10-27 Phantom View of a Commercial Continuous Phase Shifter. (Courtesy of Hewlett-Packard Company. Reprinted from Reference 46 by permission of The Institute of Radio Engineers)

⁴³A. J. Simmons, *A Method of Producing Broadband Circular Polarization in Square Waveguide*, NRL Report 4286, Washington, D.C., January 1954.

⁴⁴A. J. Simmons, "A Compact Broad-Band Microwave Quarter-Wave Plate," *IRE Proc.* **40**, No. 9 1089-1090 (1952).

⁴⁵A. G. Fox, "An Adjustable Wave-Guide Phase Changer," *Proc. IRE* **35**, 1489-1498 (1947).

⁴⁶E. F. Barnett, "A Precision X-Band Phase-Shifter," *IRE Trans. PGI-4*, 150-154 (October 1955).

In such a phase shifter, phase change is a linear function of rotation angle of the half-wave plate, and one revolution of the half-wave plate results in 720° of phase shift. A continuous phase shifter can also be used as a single-sideband generator, the input carrier being shifted an amount equal to twice the angular frequency of rotation of the half-wave plate. Whether the carrier frequency is increased or decreased by this amount depends upon the direction of rotation of the half-wave plate. The polarization of a linearly polarized wave can be rotated by passing it through a half-wave plate; the polarization will rotate in the direction of rotation of the half-wave plate, and at *twice* the angular rate.

Phase Shifters. Many methods have been used for introducing controllable phase shift into microwave systems for one purpose or another. The continuous phase shifter just described is one of the most valuable types because it can be readily provided with a calibration which is independent of frequency. Being a fairly sizable device, it often cannot be used where a phase adjuster is required. Very few practical methods have been developed for providing continuous phase shift. A number of methods for producing limited ranges of phase shift are known, however. For example, a dielectric slab moved from the side of a rectangular waveguide to the

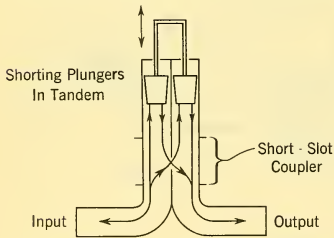


FIG. 10-28 Principle of the Slot Hybrid Phase Shifter.

high-field region at the center of the waveguide will increase the phase delay and is used where only a moderate amount of phase shift is required. A more elaborate device, but one capable of providing appreciable phase variations, is illustrated in Fig. 10-28; this employs a short-slot hybrid junction and a ganged pair of movable short circuits in a very compact, wide-range component. Phase shift is controlled by simultaneously moving both

short circuits. As indicated in Fig. 10-28, the input wave is split by the hybrid and, after reflection from the two short circuits, the two components recombine as shown. This simple device is made possible because of the quadrature phase characteristic of the slot hybrids³³; i.e. the slot hybrid splits the input signal into two equal components which are 90° out of phase. On being reflected back through the hybrid, these components recombine in the output arm. The amount of phase shift is a function of short-circuit motion, 360° of phase shift being obtained for a half-wavelength motion of the short circuits. Methods for producing phase shift electronically are discussed in Paragraph 10-16.

10-16 MICROWAVE FERRITE DEVICES AND THEIR APPLICATION

The solid-state materials are certain to play major roles in future microwave systems. Currently, the new *microwave ferrite devices*⁴⁷ are causing a near revolution in microwave system concepts through application of nonreciprocal properties and rapid electronic control to microwave circuitry. The ferrites, ceramic-like ferromagnetic semiconductor materials, when magnetically biased exhibit unusual microwave permeability characteristics due to mutual interaction of the magnetic fields with certain electrons of the ferrite.

For the special case of circularly polarized waves, the microwave permeability of the ferrite is a scalar quantity and is different for opposite senses of circular polarization. The manner in which the microwave permeabilities vary with magnetization of the ferrite is shown in Fig. 10-29. Note that for the positive sense of circular polarization, the permeability, starting at unity in the unmagnetized state, decreases, passes through zero, and then becomes negative. At a certain value of magnetization, which is propor-

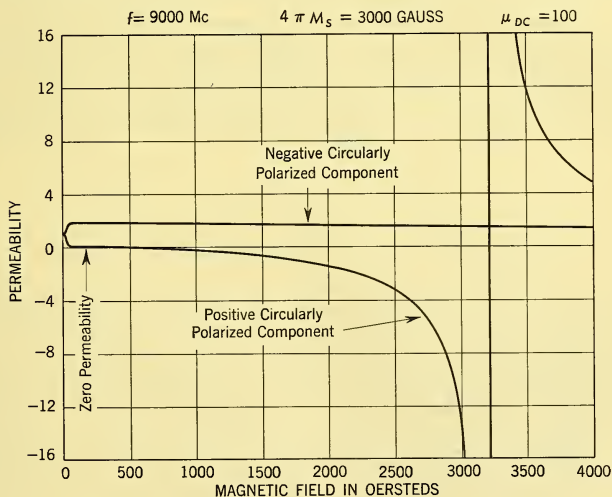


FIG. 10-29 Real Part of the Effective Permeability of a Ferromagnetic Medium Which Is Magnetized Parallel to the Direction of Propagation of an Infinite Plane Wave. (Courtesy of C. L. Hogan. Reprinted from Reference 48 by permission of The Institute of Radio Engineers).

⁴⁷See particularly the Ferrites Issue, *Proc. IRE* 44, No. 10 (October 1956).

tional to frequency, *gyromagnetic resonance* occurs and the permeability for the positive sense of circular polarization reverses abruptly between extreme values. In this same region the positive circularly polarized wave experiences a high *resonance* absorption. Except in the region of resonance, the ferrite is practically lossless for both senses of circularly polarized waves.

Various regions of the unusual and controllable permeability characteristics shown in Fig. 10-29 and the resonance loss characteristic are exploited in many ways to achieve the remarkable reciprocal as well as nonreciprocal properties of the microwave ferrite devices.^{48,49} Most of these new devices depend upon the differential permeability for their operation, while the *resonance absorption isolator*⁵⁰ exploits the high absorption differential between positive and negative circularly polarized waves in the region of gyromagnetic resonance.

Faraday Rotation. A linearly polarized wave incident on a longitudinally-magnetized ferrite element, as shown in Fig. 10-30, will undergo a

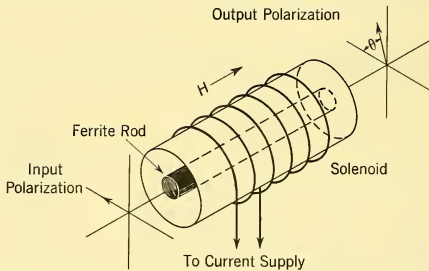


FIG. 10-30 Principle of the Ferrite Microwave Faraday Rotator.

rotation of its plane of polarization. This is because of the different permeabilities seen by the positive and negative circularly polarized components of the linearly polarized wave. (A linearly polarized wave can be resolved into two oppositely rotating circularly polarized waves of equal amplitudes.) Each circularly polarized component manifests a different phase delay in traversing the ferrite section; and if losses are negligible, they combine in a linearly polarized wave rotated by some angle θ from the input wave. The microwave Faraday effect, which permits electronic

⁴⁸C. L. Hogan, "The Elements of Nonreciprocal Microwave Ferrite Devices," *Proc. IRE* **44**, 1345-1368 (1956).

⁴⁹Benjamin Lax, "Frequency and Loss Characteristics of Microwave Ferrite Devices," *op. cit.*, pp. 1368-1386.

⁵⁰Max T. Weiss, "Improved Rectangular Waveguide Resonance Isolators," *IRE Trans. MTT-4*, No. 4, 240-243 (October 1956).

control of polarization, is utilized in a number of different types of ferrite devices.

The Microwave Gyrator. One of the most fundamentally important nonreciprocal ferrite elements is the *microwave gyrator*.⁵¹ The gyrator is a two-port circuit element, ideally lossless, having an electrical length which is greater by 180° in one direction of transmission through it than in the other. It can be represented schematically as in Fig. 10-31. A 90° Faraday rotator has the nonreciprocal property required of a gyrator, while another form of gyrator is based on a direct differential phase shift of 180° obtainable in a rectangular waveguide ferrite component. Several other microwave gyrator elements are known.

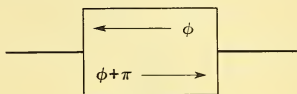


FIG. 10-31 Schematic Representation of the Gyrator.

The Circulator. One of the most important and valuable applications of the gyrator is in a nonreciprocal network called a circulator. This multiport network is characterized by nonreciprocal coupling properties such that a signal entering one port is coupled out a second port, a signal entering the second port is coupled out a succeeding port, etc., while a signal entering the last or n -th port couples to the first port. This coupling property is illustrated schematically in Fig. 10-32. Ideally there is no

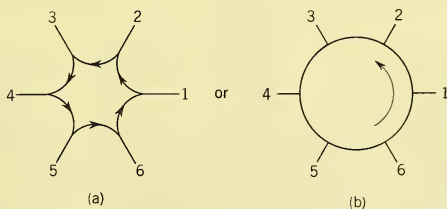


FIG. 10-32 Symbolized Coupling Properties of a Six-Port Circulator.

coupling between other ports of the circulator, but in practice coupling between the several ports is finite though often exceedingly small. A practical circulator has three or more ports; but the number is not limited, and two or more circulators can be cascaded to increase the number of ports available.

⁵¹C. L. Hogan, "The Microwave Gyrator," *Bell System Tech. J.* **31**, 1-31 (1952).

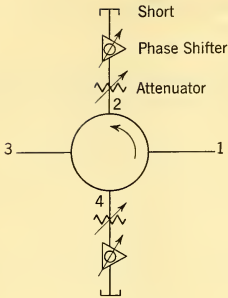


FIG. 10-33 Anisotropic Transmission Line Using a Circulator. Transmission Properties from 1 to 3 and from 3 to 1 Can Be Independently Controlled.

Circulator applications are many and include duplexing, diplexing, isolating, and switching. The circulator is also valuable as a unity-coupled directional coupler in reflectometer applications,⁵² for example. Fig. 10-33 illustrates the application of a four-port circulator to a novel *anisotropic transmission line* having independently adjustable propagation characteristics in opposite directions through the network. An attenuator and phase shifter in series with short-circuited arm 2 can control the attenuation and phase delay from arm 1 to arm 3, while a similar arrangement on arm 4 can

independently control the attenuation and phase delay from arm 3 to arm 1. The differential attenuation and phase delay thus can be made to have any desired values.

Circulators are commercially available having forward coupling insertion losses of generally less than 0.5 db. Isolation between decoupled ports ranges between 20 and 40 db; in general is not the same between various isolated ports, nor is the isolation necessarily the same in opposite directions between any two isolated ports. A broadband *differential phase shift circulator* for X-band, rated at 100 kw peak pulse power, is shown in Fig. 10-34.

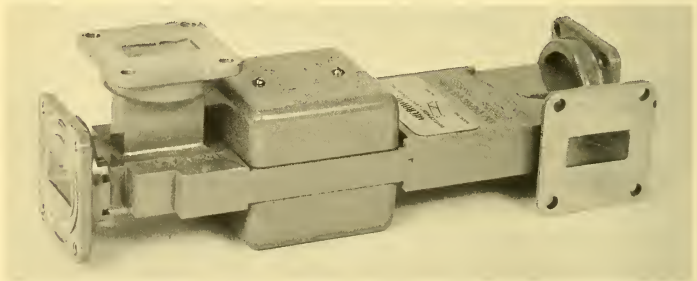


FIG. 10-34 Four-Port Circulator for X Band Rated at 100-kw Peak Power with Less Than 0.2 db Insertion Loss. (Microwave Development Laboratories, Inc.)

⁵²Howard Scharfman, "Measurement of Small Complex Reflection Coefficients," *IRE Convention Record*, Part 8, 82-94 (1955).

Switches. There are numerous ways in which ferrites can be used for microwave switching. For example, reversing the direction of the magnetic biasing field applied to a circulator will reverse the direction of circulation in the device (i.e. the direction of the arrow in Fig. 10-33 reverses with direction of applied field). In this manner a four-port circulator can be used as a reversing switch for such applications as lobe switching.

A solenoid-controlled 90° Faraday rotator can be used to switch an input signal between two orthogonal output ports of a dominant mode circular waveguide. With a slightly different geometry, a reversible 45° rotator can accomplish the same thing. The first switch type is bidirectional, but has the property of a gyrator when the magnetic field is applied. The second type functions as a reversible three-port circulator and as such is not bidirectional. It should be evident that in applying a ferrite component it is important to consider whether or not it is a reciprocal device. Fig. 10-35 shows an experimental four-port reversible ferrite switch which is

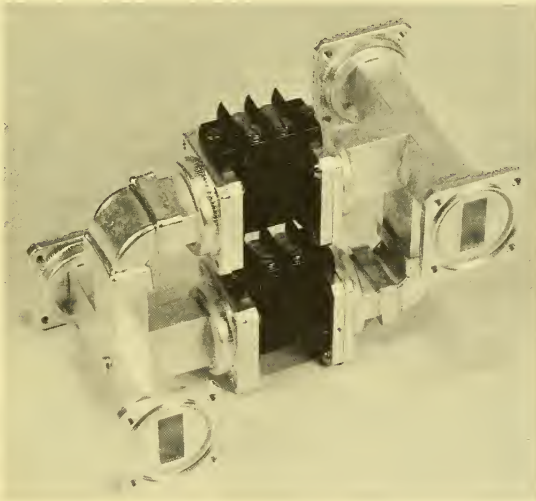


FIG. 10-35 Experimental Model of a Four-Port, *Reciprocal* Ferrite Switch. Two Faraday Rotation Gyrators are Used, and Switching is Accomplished by Reversing the Current in One of Them. (U.S. Naval Research Laboratory.)

reciprocal, and consequently bidirectional. Two 90° Faraday rotators are used with two magic-tee hybrids in a symmetrical bridge circuit. Switching is accomplished by electronically reversing just one of the 90° rotators. The

electrical symmetry of this device permits high switching ratios to be achieved over a broad band.

Off attenuations of 30 db over a 10 per cent band are readily obtained in ferrite switching devices, and 40 db can be achieved with special care in design, construction, and operation of the device. Insertion losses of a fraction of a decibel are often realized. Switching speed of microwave ferrite devices is usually limited not by the ferrite, but by eddy-current induction effects in the waveguide structure. Switching times of a fraction of a microsecond have been achieved by pulsing specially designed Faraday rotation type switches.⁵³ Ferrite switches capable of handling peak powers of 100 kw and more have been made for X-band. New low-loss materials with improved high-level characteristics will make feasible devices of still higher power.

Phase Shifters. The controllable microwave permeability of ferrites (Fig. 10-29) can be utilized to obtain either reciprocal or nonreciprocal phase shift in either circular or rectangular waveguides.⁵⁴ The phase of a circularly polarized wave passing through a longitudinally magnetized ferrite element in circular waveguide can be controlled by varying the applied magnetic field. The fact that there are regions in a rectangular waveguide where circularly polarized components exist^{50,55} permits phase control by magnetically biased ferrites in a TE_{10} mode waveguide as well. Cross-sectional configurations of three different phase shift elements of this type are shown in Fig. 10-36. The nonreciprocal type is known as a *differen-*

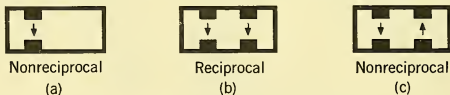


FIG. 10-36 Various Configurations for Ferrite Phase Shifters in Rectangular Waveguide: (a) Nonreciprocal. (b) Reciprocal. (c) Non-Reciprocal. Arrows Indicate Direction of Biasing Magnetic Field.

tial phase shifter, as the phase shift is different in opposite directions through the device. If this phase differential is made 180° the element has the property of a gyrator. Two 90° differential phase shifters, sometimes called *jayrators*, can function as a gyrator as illustrated in Fig. 10-37. This shows schematically the construction of the circulator of Fig. 10-34, which utilizes two 90° differential phase shift elements to perform the gyrator function required in any circulator.

⁵³R. C. LeCraw, "High-Speed Magnetic Pulsing of Ferrites," *J. Appl. Phys.* **25**, 678-679 (1954).

⁵⁴A. G. Fox et al., "Behavior and Applications of Ferrites in the Microwave Region," *Bell System Tech. J.* **34**, 5-103 (1955).

⁵⁵M. L. Kales et al., "A Nonreciprocal Microwave Component," *J. Appl. Phys.* **24**, 816-817 (1953).

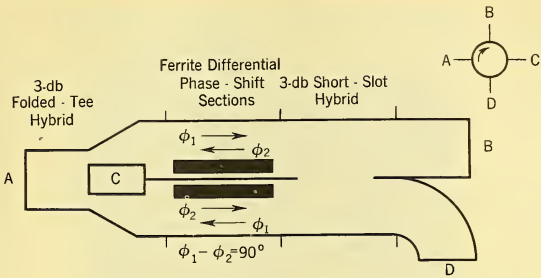


FIG. 10-37 Schematic of a Differential Phase-Shift Circulator.

Electronically controlled quarter-wave and half-wave plates for circularly polarized applications can be made using ferrites in circular or square waveguide. An electronically driven ferrite half-wave plate, in conjunction with two conventional quarter-wave plates, can be employed as an electronically controlled continuous phase shifter or single-sideband generator⁵⁶ for such applications as generating doppler frequencies.

Isolators. The first nonreciprocal component to achieve almost universal adoption in radar systems is the *ferrite isolator*. An isolator is a unidirectional attenuator having low loss in the forward direction and high loss in the reverse direction. Having gained a reputation as a panacea for many magnetron ills, this very effective component commands consideration in any new system design.

The *resonance absorption isolator*,^{50,55} one of the most important types for high-power applications, is an ingenious device which exploits the absorption differential between positive and negative circularly polarized waves in a ferrite which is magnetically biased into the resonance region. The cross-sectional geometry of such an isolator is essentially the same as the differential phase shift section of Fig. 10-36a, but magnetic field strength, generally provided by an external permanent magnet, must be higher in the isolator. The ferrite is intimately bonded to the waveguide to provide for dissipation of the heat generated in the ferrite by the RF power absorbed. A typical commercial resonance absorption isolator is shown in Fig. 10-38. There are many other ferrite devices, both low and high power, which also can function as load isolators.⁵⁷

⁵⁶J. Cacheris, "Microwave Single-sideband Modulator Using Ferrites," *Proc. IRE* **42**, 1242-1247 (1954).

⁵⁷See, for example, Alvin Clavin, "High-Power Ferrite Load Isolators," *IRE Trans. MTT-3* No. 5, 38-43 (October 1955).

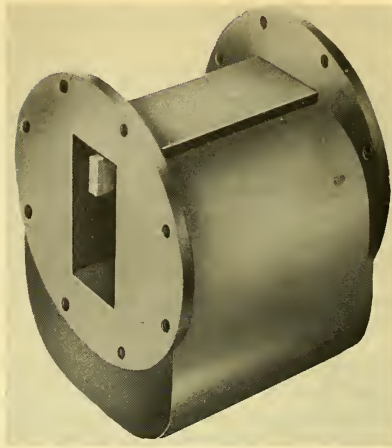


FIG. 10-38 This High-Power Resonance-Absorption Isolator for S Band Handles 400 kw Peak Power. (Courtesy of Airtron, Inc.)

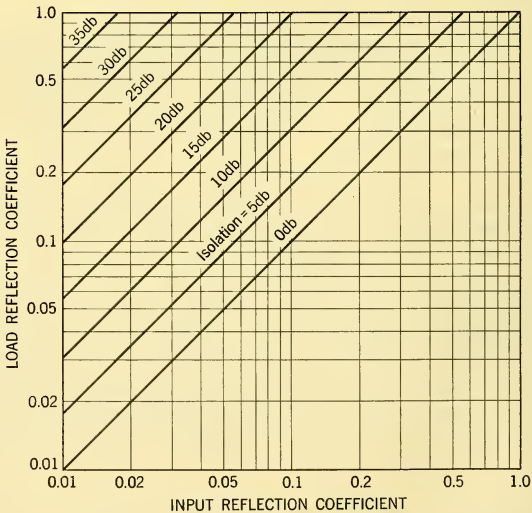


FIG. 10-39 Effectiveness of an Isolator in Reducing Input Reflection Coefficient of a Mismatched Transmission Line.

In a radar, the isolator is inserted near the magnetron to "isolate" the magnetron from any load mismatch by largely absorbing any power reflected back toward the magnetron. The effectiveness of an isolator in reducing load mismatch is shown in Fig. 10-39, where the reflection coefficient presented to the magnetron is plotted against the load reflection coefficient, for various amounts of "isolation." In general, however, the reflection coefficient presented to the magnetron is not likely to be better than the input reflection coefficient of the isolator. Although resonance absorption isolators are generally rated for a maximum load VSWR of 2 : 1, higher VSWR's are permissible with reduced power. Benefits to be derived from use of an isolator include improved spectrum, improved AFC performance, and reduction or elimination of long-line effect⁵⁸ and magnetron pulling, moding, skipping, and sparking.

Many other ferrite devices and applications are now available to the microwave engineer and system designer, who is referred to the numerous current works and references on the subject.

10-17 MICROWAVE DIELECTRIC, MAGNETIC, AND ABSORBENT MATERIALS

The role of dielectric, magnetic, and absorbent materials in microwave applications is expanding at a rapid rate. Dielectric synthesis, or the custom tailoring of dielectric materials for specialized microwave applications, is becoming commonplace. Numerous liquid, solid, and gaseous dielectric materials having favorable microwave characteristics are available for application by the design engineer. Teflon,* for example, noted for its zero moisture absorption and low loss-factor and for its frequency-independent dielectric constant, has exceptional microwave properties. Dielectric foams are finding wide use as filling and supporting structures within transmission line components where low loss and low dielectric constant are important. Relatively low loss, high dielectric constant materials ("high-K" materials) having a wide selection of electrical and physical properties are also available.

The subject of dielectrics has received extensive treatment^{59,60} and the dielectric properties of numerous materials have been studied and tabulated.^{60,61,61a} Among the newer dielectric materials of particular importance

⁵⁸J. F. Hull et al., "How Long-Line Effect Impairs Tunable Radar," *Electronics* **27**, No. 2, 168-173 (February 1954).

*Teflon is a trademark of the DuPont Company, Wilmington, Delaware.

⁵⁹A. R. von Hippel, *Dielectric and Waves*, John Wiley & Sons, Inc., New York, 1954.

⁶⁰A. R. von Hippel et al., *Dielectric Materials and Applications*, John Wiley & Sons, Inc. New York, 1954.

⁶¹A. R. von Hippel et al., "Tables of Dielectric Materials," *Laboratory for Insulation Research, Technical Report 119*, MIT, April 1957, Vol. 5.

^{61a}"Tables of Dielectric Materials," *Laboratory for Insulation Research, Technical Report 126* MIT, June 1958, Vol. 6.

for microwave applications are the *aluminas*, the *ferrites*, and the special glass-derivative crystalline materials of the *Pyroceram*† family. The aluminas and the Pyroceram materials are characterized by unusual electrical, mechanical, and thermal properties. Pyroceram possesses the important advantage that it can be fabricated to shape by conventional glass-forming techniques. Heat treatment after forming converts it to a hard, non-porous, fine-grained crystalline material having a density slightly less than that of aluminum. A most important application of the material is in the manufacture of precision radomes for high-speed airborne systems as was mentioned in Paragraph 10-9.

Low-loss ferrites have found their principal microwave application in the various reciprocal and nonreciprocal microwave ferrite devices which exploit the gyromagnetic resonance properties of the materials. The low dielectric loss and the high dielectric constant (ϵ_r , typically 13) of the microwave ferrites make them potentially valuable as special-purpose dielectrics. In an interesting application to dielectric rod antennas,⁶² both the dielectric and the magnetic properties of ferrite have been exploited in a single element. The high dielectric constant of the ferrite permits the rod antenna to be miniaturized, while the magnetic properties are utilized in an integral Faraday rotator for electronic lobe switching or control of polarization. Ingenious application of dielectric and magnetic materials in microwave systems can lead to simplification, size reduction, and improvement in performance.

Numerous microwave absorbent materials, some still in a classified category, have been developed for a wide range of applications. Lossy plastic materials, readily molded or machined, are used in such applications as dummy loads and attenuators. Thick blocks of impregnated plastic form, or mattress-like slabs of treated hair fiber,⁶³ some with surfaces elaborately carved to minimize surface reflections, are used as camouflage materials and for lining microwave "darkrooms."⁶⁴ Many of these thicker absorbent coverings reflect less than 1 per cent of the incident power, over wide frequency ranges. For narrow-band applications, absorbers can be had in rather thin sheet form.

10-18 THE DUPLEXING PROBLEM

It is common practice in radar systems to use a single antenna for both transmitting and receiving whenever possible. This is done not only to

†Pyroceram is a trademark of the Corning Glass Works, Corning, New York.

⁶²F. Reggia et al., "Ferrod Radiator System," *Proc. IRE* **45**, 344-352 (1957).

⁶³W. H. Emerson et al., "Broadband Absorbing Materials," *Tele-Tech* **14**, No. 11, 74 (November 1955).

⁶⁴A. J. Simmons et al., "Anechoic Chamber for Microwaves," *Tele-Tech* **12**, No. 7, 47 (July 1953).

ensure collimation of transmitting and receiving patterns, but to minimize space and weight requirements of the antenna system, a critical factor particularly in airborne system design. To use a common antenna for both transmitting and receiving poses a severe receiver isolation problem because of the extreme differences in signal level between the strong transmitted signal and the feeble echo return from a distant target. Some form of *duplexer*⁶⁵ is required.

The prime function of a duplexer is to adequately protect the vulnerable crystal mixers, so generally used in radar receivers, from the paralyzing power of the transmitter, and with a minimum insertion loss to both transmitted and received signals. A radar duplexer may be required to isolate a 100 kw pulse from a mixer crystal which would deteriorate if subjected to 100-mw pulses. This is a power ratio of a million-to-one, indicating the need for more than 60 db isolation between transmitter and receiver for the duration of the transmitted pulse.

Where transmission and reception take place sequentially, as in pulsed systems, this high degree of isolation is readily obtained using transmitter-actuated gas tube switches of one form or another. Duplexing in FM and CW systems, on the other hand, cannot be accomplished by switching, for transmission and reception are simultaneous and continuous; so other means must be employed. Pulsed systems basically require a nonlinear duplexing system which is amplitude selective, while continuous systems require a static duplexing scheme which provides the require high degree of isolation at all times between receiver and transmitter. In broadband continuous systems; the duplexing problem becomes most acute.

In pulsed radar systems employing a traveling-wave-tube RF amplifier in the receiver, the duplexing problem is considerably relaxed because of the relatively high input power which the tube can tolerate without damage. There are more important considerations, however, which determine the advisability of using a traveling-wave-tube preamplifier in a particular system. These are discussed in more detail in Chapter 11.

10-19 DUPLEXING SCHEMES

With few exceptions pulse radars have always employed gas tube duplexers in one form or another. Controlled by the transmitted RF pulse, the gas tube duplexer depends upon the nonlinear characteristic of the gas to perform the necessary switching action. Under the influence of the transmitted pulse, the gas tubes become intensely ionized and effect a switching action to connect the antenna to the transmitter. Upon deionization of the gas tubes, the antenna is switched back to the receiver.

⁶⁵For a detailed treatment of the subject see L. D. Smullin and C. G. Montgomery, *Micro-wave Duplexers*, McGraw-Hill Book Co., Inc., New York, 1948.

Ionization time is almost negligible, while deionization or *recovery time* may be several microseconds or more. The recovery characteristic of a duplexer plays an important role in the short-range performance of a radar. Too long a recovery time will prevent detection of small targets at short range, while too short a recovery time may result in receiver overload on short-range targets.

The *branched duplexer* and the *balanced duplexer* are the most widely used gas tube duplexers. The *branched duplexer* employs a TR and ATR tube to effect switching. On being ionized, the TR tube provides the high attenuation between transmitter and receiver required to protect the mixer crystal. On deionization, it is the function of the ATR tube to direct the received signals through the TR tube to the receiver and at the same time prevent them from entering the magnetron branch. Very high-power systems sometimes employ a third tube called a *pre-TR* tube which functions as a second TR tube in cascade with the first to provide the greater isolation needed with the higher power. The branched duplexer has certain disadvantages — notably, rather limited bandwidth. In its un-ionized state the ATR tube presents a serious mismatch to the magnetron, and under certain phasing conditions, troublesome “moding” can occur in the magnetron.⁶⁶ For this reason the magnetron-to-ATR spacing becomes an important factor in the design of a branched duplexer.

A number of advantages over the simple branched duplexer are offered by the nonlinear *balanced duplexer*. Now being used extensively, the nonlinear balanced duplexer provides broadband operation, eliminates the ATR tube and its mismatch problem, provides somewhat better crystal protection, and generally introduces somewhat less loss than the branched duplexer. The balanced duplexer employs two TR tubes suitably mounted between two 3-db hybrid junctions in a bridge arrangement. Balanced duplexers have been made using many types of hybrid junctions, but the *slot hybrid duplexer*³³ has found widest acceptance because of its compactness and unexcelled broadband performance. A dual TR tube having a common gas fill is now commonly used in the slot hybrid duplexer to obtain best bandwidth through closely balanced characteristics. Duplexers which incorporate solenoid-operated shutters in the TR tubes are available to provide crystal protection against stray radiation whenever the radar is turned off.

An integral slot-hybrid balanced duplexer for X-band, incorporating a dual TR tube with shutters, is shown in Fig. 10-40. The signal path in such a slot hybrid duplexer for the transmit condition is shown in Fig. 10-41a. On entering the duplexer, the transmitter power is split by the 3-db hybrid and, on ionizing the TR tubes, is reflected. However, because of the

⁶⁶*Magnetrons and Traveling-Wave Tubes*, Tube Division, Radio Corporation of America, Harrison, N. J., 1956.

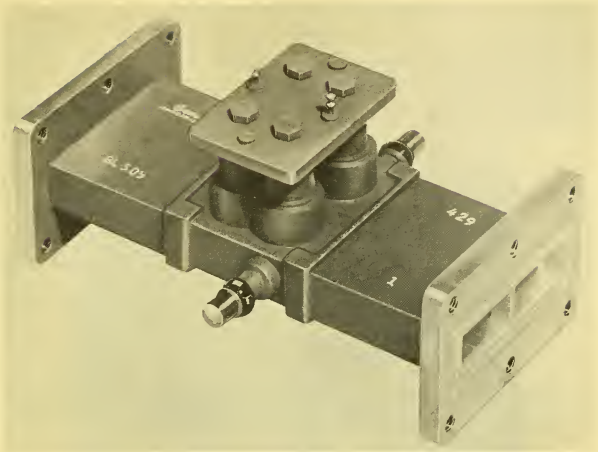


FIG. 10-40 Integral Slot-Hybrid Balanced Duplexer Assembly with Shutter-Type Dual TR Tube. (Courtesy of Bomac Laboratories, Inc.)

quadrature phase characteristic of the slot hybrid,³³ all the power leaves by the adjacent port as shown in Fig. 10-41a. After TR recovery, received

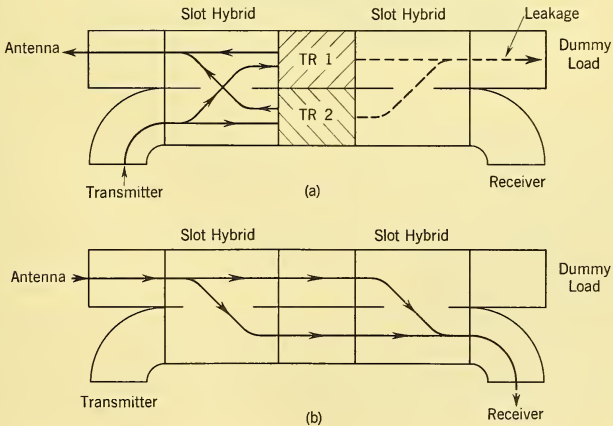


FIG. 10-41 Schematic of a Slot-Hybrid Balanced Duplexer in (a) the Transmitting Condition, and (b) the Receiving Condition.

echo signals follow the paths shown in Fig. 10-41b. The dummy load on the fourth arm of the duplexer serves only to absorb a small amount of leakage power during transmission.

The novel *crossed guide duplexer* pictured in Fig. 10-42 has a number of features which make it attractive for certain airborne radar applications.

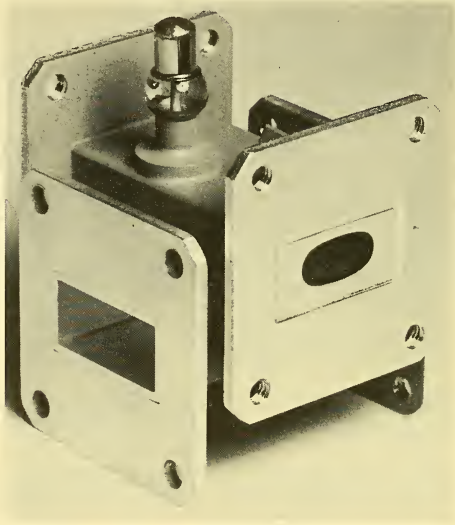


FIG. 10-42 Cross-Guide Duplexer. (Courtesy of Bomac Laboratories, Inc.)

A complete duplexer in itself, the device requires no ATR tube and does not present a large mismatch to the magnetron before it ionizes. Compact, lightweight, and convenient to apply, the device functions as a unity-coupled directional coupler on reception; but on transmission, ionization takes place in the region of the clover-leaf coupling iris and provides the necessary receiver isolation. Its application is limited, largely because of its restricted bandwidth.

The gas tube duplexer, although universally used in pulse radar systems, is by no means the only form of duplexing. FM and CW systems, as described in Chapter 6, in fact require some linear form of duplexing to permit simultaneous reception on a continuous basis. There are a number of linear methods for isolating the receiver from the transmitter in a radar system. The most direct solution of course is to employ separate transmitting and receiving antennas and achieve isolation by physical separa-

tion, taking advantage of the directional characteristics of the antennas. Seldom is this a very practical solution in airborne systems. The method requires two separated antennas located where space and weight are at a premium; it poses problems of interconnection, as well as problems of alignment and parallax between the antennas. A preferred solution is to employ some duplexing device or circuitry which permits use of a common antenna for simultaneous transmission and reception.

There are several *linear* duplexing schemes which might be used in certain applications; however, in general they involve a two-way transmission loss of 6 db. Any of the familiar 3-db hybrid junctions can be employed in a *linear balanced duplexer*⁶⁵ as illustrated schematically in Fig. 10-43. The transmitter is shown connected to port 1 of a hybrid ring

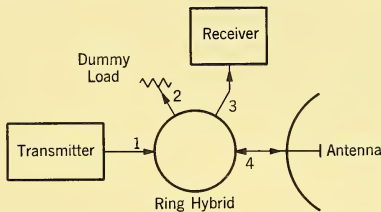


FIG. 10-43 Linear Balanced Duplexer Using Ring Hybrid.

which divides the power equally between arms 2 and 4. The receiver is connected to arm 3 and will be isolated from the transmitter on arm 1 if arms 2 and 4 are properly terminated. As with any linear duplexer, antenna mismatch will cause direct reflection of transmitter power back to the receiver. This reflected component can be canceled out by an adjustable reactance in arms 2 or 4, but a very frequency-sensitive duplexer may result. To achieve a high degree of isolation between transmitter and receiver in a duplexer of this type the hybrid must be well balanced and the antenna and dummy load well matched. The 6-db loss arises in this type of duplexer because half the transmitter power is "dumped" in the dummy load and only half the received power reaches the receiver.

Linear duplexing can also be accomplished using *mode multiplexing* techniques. Thus orthogonal linearly polarized modes are employed, for example, in circular or square waveguide, transmitting on one and receiving on the other, and relying on target depolarization of the transmitted signal to permit reception of an echo.⁶⁷ A variation of this duplexing technique involves transmitting one sense of circular polarization and receiving the other.

⁶⁷See Paragraph 4-3 for a discussion of the effects of the target upon polarization.

The ideal linear duplexer is a matched, lossless three-port device in which the first port couples only to the second and the second couples only to the third. Such an "impossible" duplexing device recently came into being as the *microwave circulator*. A perfect circulator is the ideal linear duplexer. In practice, one must settle for something less than ideal duplexing performance. The practical microwave circulator is, however, the best known solution to the linear duplexing problem and does not suffer the loss which is characteristic of other linear duplexers. Four-port circulators providing isolation in excess of 30 db, and one-way insertion losses of a fraction of a decibel are a practical reality for moderate-powered systems (see Fig. 10-34). As previously mentioned, the isolation achieved between trans-

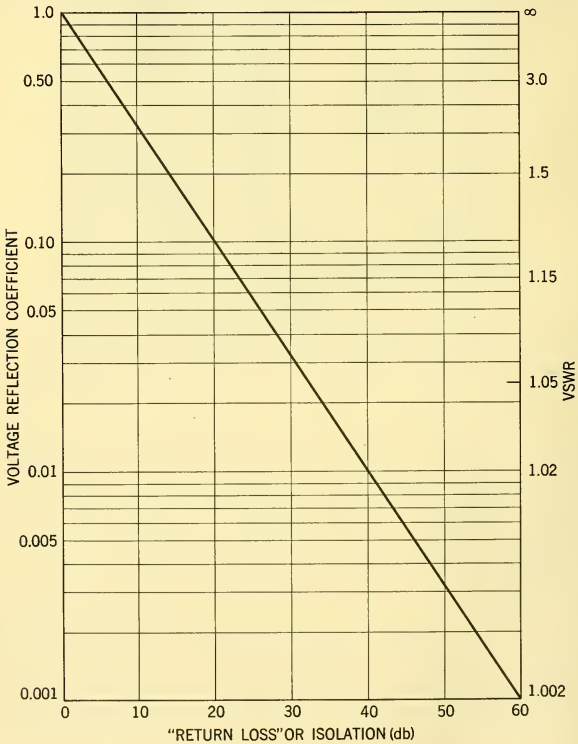


FIG. 10-44 Return Loss or Receiver Isolation as Affected by Load Reflection Coefficient.

mitter and receiver using a linear duplexer is dependent on the quality of antenna match. With a perfect circulator, the maximum transmitter-receiver isolation achievable is determined by the mismatch on the antenna arm. The degree of isolation which could be obtained using a perfect circulator as a duplexer is shown in Fig. 10-44 as a function of the reflection coefficient of the antenna arm. It is plain that to achieve high isolation the antenna arm must be exceptionally well matched. For the same antenna arm reflection coefficient an additional 6-db isolation between transmitter and receiver is obtained by using the linear balanced duplexer, because of the split of both the transmitter power and the reflected power. But of course system performance is down by 6 db in this case.

The circulator is also useful for duplexing in pulsed radar systems where it is possible to use a supplementary TR tube to obtain much greater receiver isolation than is provided by the circulator alone.⁶⁸ The advantage of such a duplexer arrangement is that the circulator greatly reduces the RF power incident upon the TR tube. The result is considerably increased tube life. Reduced ionized intensity of the TR tube effects a reduction of recovery time, thus permitting close-range reception of weak signals.

As shown in Fig. 10-45 the TR tube in the circulator duplexer might be replaced by an electronically controlled ferrite device to provide attenuating or switching action which can be synchronized with the transmitted pulse. The ferrite element can also function as an RF gain control or as a sensitivity-time control (STC) which regulates receiver sensitivity as a function of target range.

Certain other advantages are gained in using a circulator duplexer. The circulator provides load isolation for the transmitter, as antenna reflections will be absorbed in the dummy load (Fig. 10-45) when using a TR tube, or possibly in the ferrite control element when one is used. Local oscillator radiation is also reduced, as it too is absorbed in the dummy load.

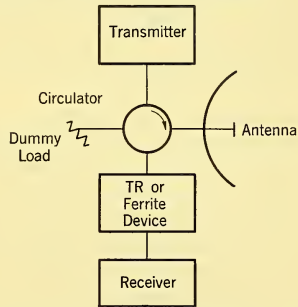


FIG. 10-45 Circulator Duplexer with Auxiliary TR Tube or Ferrite Device to Improve Receiver Isolation.

10-20 SPECIAL PROBLEMS OF COHERENT SYSTEMS

By a coherent system is meant one in which a phase or frequency comparison is made between the transmitted signal and the received echo.

⁶⁸"X-Band Ferrite Duplexer," *Electronic Design*, August 1, 1956, pp. 32-33.

The FM, CW, and pulse doppler systems described in Chapter 6 fall in this category. In addition to duplexing problems already mentioned, special problems of phase, amplitude, and frequency stability are encountered. Extraneous modulation of any of these factors by sources other than the radar target of interest can result in troublesome noise which, if within the passband of interest, will deteriorate system performance.

Mechanical resonances and vibrations in any part of the microwave plumbing may lead to serious system noise. Thus, components which vibrate or flex under operational stress may introduce phase modulation which adds to system noise. Sections of flexible waveguide, for example, are especially prone to phase modulating when flexing, and unless damped or well secured can be a troublesome source of noise.

Not the least of the problems in coherent systems is that of duplexing. In CW and FM systems which must rely on a linear duplexer for receiver isolation, even minute variations in antenna reflection coefficient will seriously affect receiver isolation, as can be inferred from Fig. 10-44. Antenna scanning is virtually impossible in such systems because of inevitable modulation of the reflection coefficient with scanning motion. Separate transmitting and receiving antennas are indicated if scanning is required. Separating transmitting and receiving antennas in such a system to obtain the necessary isolation does not, however, solve all the problems of extraneous modulations.

Mechanical vibrations of the antenna feed or dish also can introduce phase modulation. One especially troublesome area is that surrounding the radome — not only the mechanical vibrations and noise introduced by buffeting of the radome, but also the high-intensity sound waves generated within the radome, which increase seriously as Mach number rises. Added to this are the problems of signal modulation and reflection from the shock wave and turbulence which shrouds a radome traveling at supersonic speed. At hypersonic speeds ionization creates additional modulation problems. Pulsed doppler systems avoid the duplexing problems of other coherent systems, but are faced with many of the same problems of extraneous modulation.

10-21 SOLID STATE AMPLIFIERS

Since the original ammonia gas MASER work of C. H. Townes and others in 1954, the use of molecular systems to provide low noise amplification at high frequencies has received great attention. In an effort to obtain increased bandwidth with little increase in noise output, suggestions were made in 1956 that solids be used in MASER amplifiers. A further development was the realization in 1957 that nonlinear reactance MAVAR amplifiers are also capable of providing very low noise amplification.

Since then, remarkable progress has been made, and these types of devices will certainly be used in many radar and communication systems of the future.

MASERS (Molecular Amplification by Stimulated Emission of Radiation). Amplification is accomplished in a MASER by establishing an excess population density in an excited molecular energy state so that the assembly of molecules emits electromagnetic energy in making transitions to a lower energy state when stimulated by a low-power signal. Coherent amplification of stimulating frequencies corresponding to the energy-level change is obtained in this manner. Solid-state MASERS are constructed using paramagnetic salts as the active material, with the external electromagnetic field interacting with the magnetic moments of bound electrons or their spin levels. For this reason such MASERS are often called *paramagnetic amplifiers*. Normally, a very strong and uniform d-c magnetic field is required to separate the spin levels sufficiently. Also, very low operating temperatures are normally required to provide usable efficiencies and the low noise operation of MASERS. For some applications, it may be possible to obtain the required low temperatures by magnetic cooling with only a moderately cold environment. To date, the auxiliary equipment needed to provide the d-c magnetic field and the cooling has led to rather bulky designs.

Various possibilities exist for the choice of energy levels and methods for producing inverted populations of molecules in which there are more than a normal number in a high-energy level state, and the assembly of molecules is in a condition to emit radiation when stimulated. The simplest type is a two-level device. In the most common design for a two-level MASER a strong auxiliary microwave pulse at the transition frequency inverts the populations at the two levels by causing more upward than downward transitions, since there are normally more molecules in the lower level. Two-level MASERS normally provide only intermittent operation, since amplification is not possible while the inverting pulse is being applied, although some two-level systems have been constructed which operate continuously.

In three-level MASERS, a continuous pumping signal is used to equalize the populations of the lowest and highest levels. An inverted population condition is thus produced between one of these levels and a third intermediate level. This method provides continuous operation since the pumping signal is at a different frequency than that which is emitted.

Another possibility which has been proposed as a means of obtaining an emissive state is to pump with polarized light of a wavelength corresponding to an optical transition of the active material. The light from the pump is absorbed and shortly reradiated as unpolarized light. This results

in a rearrangement of atoms among the various energy levels. In this manner, it is possible to invert the populations of pairs of energy levels connected by microwave transitions. The feasibility of such a device has not yet been demonstrated experimentally.

MAVARS (Modulating Amplification Using Variable Reactance). In a MAVAR amplifier, three resonant circuits are coupled through a common nonlinear reactive element, so that for certain choices of resonant frequencies for the circuits, power is pumped from a generator in the highest frequency circuit into the lower frequency circuits and

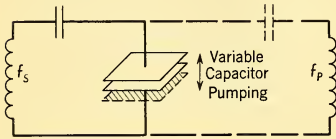


FIG. 10-46 Single Resonant Circuit with Variable Reactance.

amplification of a low power signal at one of these lower frequencies results. Because they depend for their operation upon the variation of a circuit parameter, they are often called *parametric amplifiers*. The principle of operation can be illustrated with the single resonant circuit shown in Fig. 10-46.⁶⁹ Suppose that this circuit is oscillating at its resonant frequency f_s . It is possible to introduce additional energy into this circuit and thus amplify the oscillating signal by separating the capacitor's plates when it is charged and pushing them together again when the voltage across the capacitor is zero. This pumping action is analogous to the manner in which a child pumps a swing. The frequency at which the capacitor is pumped must be exactly twice the resonant frequency of the original circuit f_s , since the capacitor value is increased by pushing together

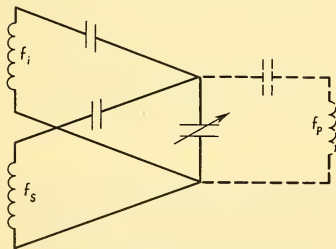


FIG. 10-47 Equivalent Electrical Circuit of Parametric Amplifier with Nonlinear Capacitance Element.

its plates every time the voltage across the variable capacitor is zero, or twice each cycle. The periodic variation of the capacitor would not be obtained in practice by physically moving the plates but rather by applying a large voltage of the proper frequency to a nonlinear capacitor whose incremental capacitance varied with applied voltage.

An external resonant circuit which would supply this pumping frequency is indicated with dotted lines in Fig. 10-47.

⁶⁹Neither this circuit nor the one in Fig. 10-47 is an exact representation of an equivalent circuit of practical parametric devices; these two are used to illustrate the principles involved.

With the circuit in Fig. 10-47, the proper phase must be maintained between the signal which is being amplified and the pumping voltage. In order to eliminate this requirement, another resonant circuit is paralleled across the nonlinear capacitor as shown in Fig. 10-47. This circuit is called the *idler* and its frequency is adjusted so that $f_i + f_s = f_p$. The sum of the signal and idle voltages will have zeros at the frequency $f_i + f_s$ since

$$\cos \omega_s t + \cos \omega_i t = 2 \cos \frac{1}{2} (\omega_s + \omega_i) t \cos \frac{1}{2} (\omega_s - \omega_i) t \quad (10-29)$$

With the proper phase, then, the pumping voltage should introduce energy into the system and increase the signal and idle voltages. It turns out that the phase of the idle voltage automatically adjusts itself to a value allowing pumping action of this kind. The effective result insofar as the signal circuit is concerned is that the idling circuit appears as a negative resistance whose value depends upon the value of the pumping voltage and the characteristics of the nonlinear capacitance. By adjusting the effective negative resistance to be only slightly less than the physical resistance in the signal circuit, considerable very low noise amplification can be obtained. These results are possible at room temperatures; it is not necessary, as it is in the case of MASER amplifiers, to cool such a device to get very low noise figures.

In addition to the negative resistance type of parametric amplifier discussed above, in which the input and output frequencies are equal, it is possible to couple into the idling circuit to provide an amplifying up-converter. If the output resonant frequency is greater than the pumping frequency, the device will be unconditionally stable (unlike the negative-resistance type of device, which could become oscillatory) although the maximum possible gain is limited.

Currently, the most successful and promising types of solid-state parametric amplifiers have been constructed using semiconductor junction diodes as nonlinear capacitive elements. The depletion layer formed at the p-n junction acts like a capacitance. Because the width of this layer is a nonlinear function of the applied voltage, the capacity varies with this voltage in a nonlinear fashion.

Another method of obtaining parametric amplification at microwave frequencies employs ferromagnetic materials to obtain the equivalent of nonlinear inductive elements. Proposals have also been made for obtaining nonlinear circuit elements at microwave frequencies by means of a number of other methods. These include the use of ferroelectric effects, the Hall effect, and cyclotron resonance in semiconductors.

Fig. 10-48 briefly summarizes the present status of solid-state amplifier development. The diagrams in that figure are intended to convey a rudimentary physical concept of typical configurations.

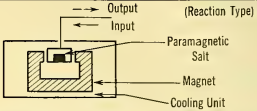
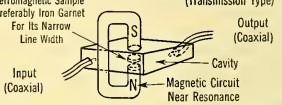
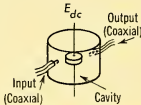
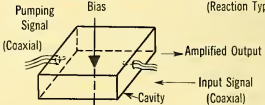
Type	Development Status	Conceptual Diagram
Maser	Paramagnetic	 <p>(Reaction Type)</p> <p>Output</p> <p>Input</p> <p>Paramagnetic Salt</p> <p>Magnet</p> <p>Cooling Unit</p>
	Ferromagnetic	 <p>(Transmission Type)</p> <p>Output (Coaxial)</p> <p>Input (Coaxial)</p> <p>Magnetic Circuit Near Resonance</p> <p>Cavity</p> <p>Ferromagnetic Sample Preferably Iron Garnet For Its Narrow Line Width</p>
Parametric Amplifier	Ferroelectric	 <p>(Transmission Type)</p> <p>Output (Coaxial)</p> <p>Input (Coaxial)</p> <p>Cavity</p> <p>E_{dc}</p> <p>Ferroelectric Sample Such as $BaTiO_3$ or $ZrTiO_3$ For Low Loss and Nonlinear Amplification Properties</p>
	Diffused Junction Diode	 <p>(Reaction Type)</p> <p>Amplified Output</p> <p>Input Signal (Coaxial)</p> <p>Cavity</p> <p>Bias</p> <p>Pumping Signal (Coaxial)</p>

FIG. 10-48 Summary of Present Status of Solid-State Amplifier Development.

The discussion of this paragraph has been very brief and of a general nature. The references given below provide more detailed information⁷⁰⁻⁷⁴.

BIBLIOGRAPHY FOR CHAPTER 10

1. John D. Kraus, *Antennas*, McGraw-Hill Book Co., Inc., New York, 1950.
2. George C. Southworth, *Principles and Applications of Waveguide Transmission*, D. Van Nostrand Co., Inc., Princeton, N.J., 1950.
3. H. J. Reich, P. F. Ording, H. L. Krauss, and J. G. Skalnik, *Microwave Theory and Techniques*, D. Van Nostrand Co., Inc., New York, 1953.
4. L. G. H. Huxley, *Waveguides*, The Macmillan Co., New York, 1947.

⁷⁰J. P. Wittke, "Molecular Amplification and Generation of Microwaves" *Proc. IRE* **45** 291-316 (1957).

⁷¹H. E. D. Seovil, G. Feher, and H. Seidel, "Operation of Solid State Maser" *Phys. Rev.* **105** 762 (1957).

⁷²D. I. Bolef and P. F. Chester, "Some Techniques of Microwave Generation and Amplification Using Electron Spin States in Solids," *IRE PGMTT Trans. MTT-6*, 47 (January 1958).

⁷³R. D. Haun, Jr., "Small Signal Theory of Microwave Parametric Amplifiers and Up-Converters Using High Q Non-Linear Reactances," Westinghouse Research Scientific Paper 8-1047- P. 2, Nov. 3, 1958.

⁷⁴S. Bloom and K. K. N. Chang, "Theory of Parametric Amplification Using Non-Linear Reactances," *RCA Rev.* **18**, 578-593 (1957).

5. Edward L. Ginzton, *Microwave Measurements*, McGraw-Hill Book Co., Inc., New York, 1957.
6. Theodore Moreno, *Microwave Transmission Design Data*, McGraw-Hill Book Co., Inc., New York, 1948.
7. T. N. Anderson, "Rectangular and Ridge Waveguides," *IRE Trans. MTT-4*, 201-209 (October 1956).
8. T. N. Anderson, "Double-Ridge Waveguide for Commercial Airlines Weather Radar Installation," *IRE Trans. MTT-3*, 2-9 (July 1955).
9. George L. Ragan, *Microwave Transmission Circuits*, McGraw-Hill Book Co., Inc., New York, 1948.
10. N. Marcuvitz, *Waveguide Handbook*, McGraw-Hill Book Co., Inc., New York, 1951.
11. *Techniques for Airborne Radome Design*, Wright Air Development Center, WADC Tech. Report 57-67, Sept. 1957, ASTIA Document Number AD142001 (Report Confidential).

CHAPTER 11

THE GENERATION OF MICROWAVE POWER

11-1 THE MAGNETRON

The magnetron has been the source of transmitted microwave power in nearly all of the airborne radar equipment constructed in recent years. It is a self-excited power oscillator that in its more modern versions can be tuned mechanically over moderate frequency ranges, typically 10 per cent.

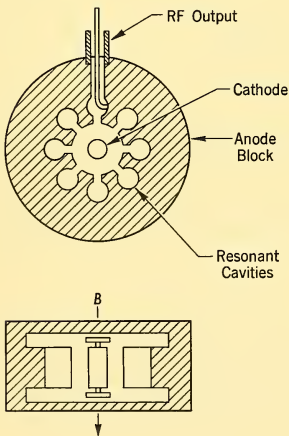


FIG. 11-1 Physical Configuration of Typical Microwave Magnetron.

At the present time other types of tubes that permit more sophisticated systems are being given increasingly serious consideration by designers of modern radar systems. However, the primary virtues of the magnetron that have led to its present dominant role have yet to be excelled by other types of tubes. These virtues are small size, light weight, high efficiency, and low operating voltage. All of these virtues are important to the designer of airborne radar equipment.

The physical configuration of a typical microwave magnetron is shown in Fig. 11-1. The magnetron is basically a cylindrical diode, with a central cathode enclosed in a larger, cylindrical anode. The magnetron is immersed in a strong

magnetic field parallel to the cylindrical axis.

Electron Motion. To understand the operation of a magnetron, it is necessary first to understand the nature of the electron trajectories in a static, nonoscillating magnetron. When an electron leaves the cathode and is accelerated toward the anode by a small anode potential, its trajectory is curved by the action of the magnetic field. The electron follows a quasi-

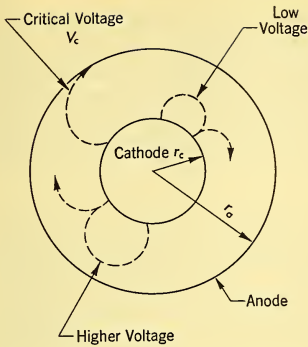


FIG. 11-2 Electron Trajectories in a Static, Nonoscillating Magnetron.

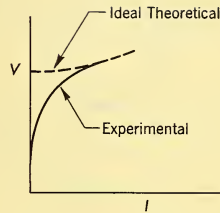


FIG. 11-3 Voltage-Current Relationship in a Nonoscillating Magnetron.

cycloidal path and returns to the cathode, as shown in Fig. 11-2. As the voltage is increased, the maximum distance from the cathode reached by the electron also increases. Finally, as a critical voltage is reached, the electron will strike the anode. The critical voltage is given by

$$\frac{V_c}{B^2} = \frac{er_a^2}{8m} \left[1 - \left(\frac{r_c}{r_a} \right)^2 \right]^2 \quad (11-1)$$

where V_c is the voltage, B the magnetic field, e/m is the charge to mass ratio of the electron, and r_c and r_a the radii of the cathode and anode respectively.

If this ideal theoretical behavior were actually followed, the voltage current relationship would be as shown in Fig. 11-3. Experimentally, the abrupt transition at the critical cutoff voltage is not observed, although a rapid increase of current is observed, as shown. The reasons for this departure from theory are not well understood, although it has been postulated that electronic interaction of some nature takes place in the whirling cloud of electrons with crossing trajectories that surrounds the cathode. The possibility of an electronic interaction in the whirling cloud of space charge is strengthened by another experimentally observed fact. Some of the electrons that return to the cathode strike it with considerable velocity, again in contradiction to the simple theory. These electrons release secondary electrons which contribute to the electron emission of the cathode. Cathode emissivity depends then upon both primary and secondary emission properties.

In normal operation of a magnetron, the heater voltage must be turned on initially to heat the cathode to a temperature sufficient for electron

emission. But when anode voltage is applied, the back bombardment of the cathode by accelerated electrons will further heat the cathode. To prevent overheating of the cathode, the heater voltage is normally reduced after the anode voltage is applied. It is believed that this electron bombardment of the cathode may have a harmful effect upon its life, although the effects of ionic bombardment which depend upon residual gas pressure are probably more severe.

Resonant System. The anode of a microwave magnetron is normally a heavy metal structure, typically copper, in which is machined or formed the RF circuitry of the magnetron. Typical anode structures consist of a number of slots, wedges, or slots and holes uniformly spaced around the anode, as shown in Fig. 11-4. Each of these slots may be regarded as a

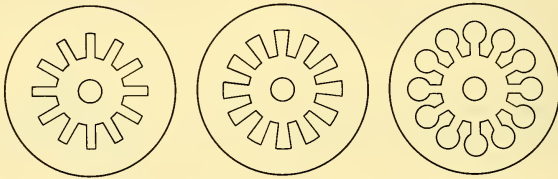


FIG. 11-4 Typical Anode Structures for Microwave Magnetrons.

simple cavity resonator, with a strong electric field developed across the opening of the slot facing the cathode. The fields of each resonator extend into the region between cathode and anode, and also into the end spaces at the ends of the anode block. The individual cavity resonators are therefore coupled together, and the entire anode block, together with the end spaces and other parts that contain the electromagnetic field, form the resonant system of the magnetron.

The complete resonant system presents too complex a problem for accurate, quantitative analysis. Many of its properties can be understood by considering a resonant system with N slots or cavities to be a system of N coupled resonant circuits. Such a system normally has N resonances at different frequencies, although degeneracies exist in the magnetron structure that reduce the independent resonances to $N/2$. Normally, it is desired to operate the tube at a particular one of the resonant frequencies, and the other possible resonant modes are troublesome because they may lead to spurious oscillations at undesired frequencies.

The desired mode of oscillation is usually that in which the phase shift between adjacent segments of the anode is π radians, or 180° (π mode).

For the other modes, the phase shift between adjacent segments of the anode is some value other than 180° , the fields in the different resonators will not be uniformly strong, and less efficient operation is observed.

The undesired modes are most troublesome when their frequencies are close to the desired π mode. To increase the frequency separation between modes, strapping of the anode is frequently resorted to. Strapping may take a variety of forms. One typical form, *double ring strapping*, is shown schematically in Fig. 11-5. The effect of strapping upon increasing the separation of the resonant frequencies of the various modes is shown in Fig. 11-6.

Another type of anode block that is widely used is the *rising-sun anode* shown in Fig. 11-7. In this system, the alternate cavities are of unequal

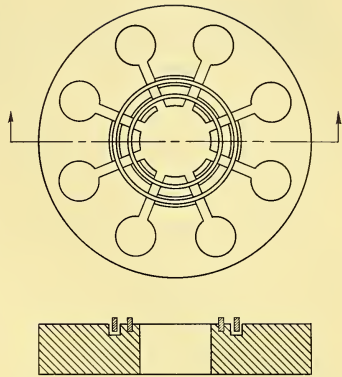


FIG. 11-5 Double-Ring Strapping on a Magnetron Anode Block.

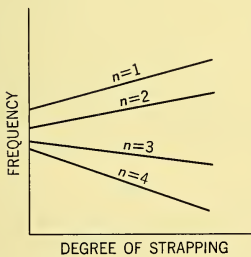


FIG. 11-6 Effect of Strapping upon Mode Separation for an Eight-Resonator System.

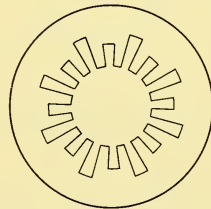


FIG. 11-7 Rising-Sun Resonant System for Magnetron.

dimensions. This results in a wider separation of modes than is found with an unstrapped system of equal resonators, and it is not necessary to strap a rising-sun anode. There is one important difference between the mode spectrum of a strapped system and a rising-sun system. With the strapped system, the desired π mode is the lowest frequency mode, while for the rising-sun system, the desired π mode lies between groups of modes at higher and lower frequencies.

Mechanism of Oscillation. The mechanisms that lead to oscillation in a magnetron may be understood by considering first what happens when a voltage is applied between cathode and anode in the magnetron. Although no current is drawn initially to the anode, a whirling cloud of electrons is formed about the cathode that expands in radius as the voltage is increased. Neglecting the electronic interaction effects, the outer radius of the cloud will reach the anode when the cutoff voltage is reached. For voltages below the cutoff voltage, the electrons near the outer edge of the cloud have considerable angular velocity as they sweep by the slots in the anode.

If an RF voltage exists across the slots such that the electron picks up energy from the field, its path will be distorted in a manner such that the electron will be returned to the cathode. It will strike the cathode and be removed from the cloud, although it may be expected possibly to release some new secondary electrons into the electron cloud. On the other hand, an electron which delivers energy into the RF field and is thereby slowed down will not be returned to the cathode, but will continue to travel in the cloud. There therefore exists a sorting mechanism, whereby the electrons that extract energy from the RF field are removed from the electron cloud, and those that deliver energy to the field remain in the cloud.

A condition of synchronism between the rotating electrons and the electromagnetic field must also exist for efficient operation. For the normal π mode of operation, at any given instant of time the electric field across one of the slots is exactly 180° out of phase with the field across the slots adjacent to it on each side. If the conditions of synchronism are met, the angular velocity of the electrons will be such that in the time it takes an electron to travel from one slot to the next, one half-cycle will have elapsed. The rotating electron will therefore encounter RF fields of identical phase as it passes by successive slots. Equivalent conditions of synchronism exist for the other possible modes of oscillation.

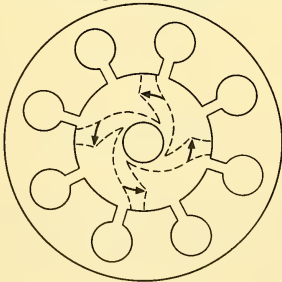


FIG. 11-8 Rotating Space-Charge Cloud in an Oscillating Eight-Resonator Magnetron.

When the conditions of synchronism are met and the voltage is below the cutoff voltage, the sorting mechanism will begin to operate and there will be a net transfer of energy from the electron cloud to the RF field, causing oscillations to build up. The electron cloud will be distorted into the form of a spoked wheel, with $N/2$ spokes if there are N cavities, as shown in Fig. 11-8. These spokes sweep by the slots in the proper phase for the electrons to deliver energy to the WF field, and the

slowed-down electrons will eventually reach the anode. It is not necessary that the conditions for synchronism be met exactly; a phase-focusing mechanism holds the electrons in synchronism long enough for them to be slowed down sufficiently to reach the anode. The conditions for synchronism are given mathematically by

$$V_H = \frac{\pi f}{n} (r_a^2 - r_c^2) \left[B - \frac{2m \pi f}{e n} \right] \quad (11-2)$$

where f = operating frequency

r_a = anode radius

r_c = cathode radius

B = magnetic field strength

m = mass of electron

e = charge of an electron

n = Hartree harmonic of the interaction field.¹

V_H is called the *Hartree voltage*, after the man who developed the theory. As the anode voltage of a magnetron is increased, oscillations can be expected when the voltage reaches the Hartree voltage, provided it is below the cutoff voltage. When the tube breaks into oscillation, the current drawn to the anode will increase sharply. The relationship between the Hartree voltage and the cutoff voltage is shown in Fig. 11-9. In practice these conditions are not met exactly, because of the conditions of space charge that were neglected in the theory but which affect the electron trajectories.

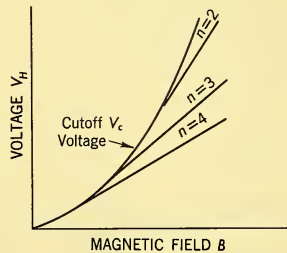


FIG. 11-9 Hartree Diagram, Showing Relationship Between Cutoff Voltage and Hartree Voltage for Oscillation in the Various Modes, for an Eight-Resonator System.

Moding. The experimental current-voltage relationship of the static, nonoscillating magnetron is shown in Fig. 11-10, along with the voltage-current relationship of two of the modes of oscillation. The form of the curve shown is typical of experimentally observed results. In the example

¹For a more detailed definition of n , the reader is referred to H. D. Hagstrum, "The Generation of Centimeter Waves," *Proc. IRE* 35, 556 (1947). (In this article, the symbol k designates the harmonic that we designate by n .)

shown, the Hartree voltages for other possible modes of oscillation are greater than the cutoff voltage. For any one of the given modes of oscillation, the current-voltage relationship will be a function of the external load on the magnetron, as shown in Fig. 11-11.

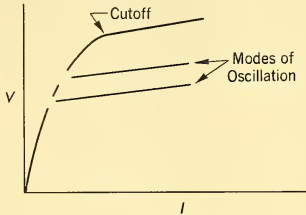


FIG. 11-10 Voltage-Current Relationship for Nonoscillating Magnetron and for Two Modes of Oscillation.

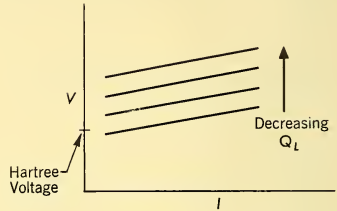


FIG. 11-11 Variation of Voltage-Current Relationship with Changing External Load on the Magnetron.

When the magnetron is operated, there is competition between the possible modes of oscillation, and for proper operation the conditions must be such that oscillation occurs only in the desired mode. The conditions are several and they interact. For example, when a pulse is applied to the magnetron, and the leading edge of the pulse reaches the Hartree voltage, a finite time is required for the oscillations to build up. If the leading edge of the pulse is too steep, the oscillations may not build up fast enough for the increase of current to limit the rise in voltage, and the voltage will continue to rise to the point where oscillations start in the next higher voltage mode of oscillation. To prevent this, it is usually necessary to specify the maximum permissible steepness of the leading edge of the voltage pulse that is applied to the magnetron.

With a rising-sun anode, undesired modes of oscillation can be excited at voltages lower than the voltage of the π mode. To avoid oscillation in these undesired modes, the leading edge of the voltage pulse cannot be too gradual.

The voltage-current characteristic of the modulator is also important. Fig. 11-12 shows the voltage-current characteristic of two different types of modulators. As shown, one curve is typical of hard-tube modulators, and the other of line-type pulsers. The two voltage-current characteristics of two modes of oscillation are also shown. It is apparent that with one modulator, the curve intersects both mode curves, and with the other modulator, only one. Oscillation in both modes is therefore possible with the first modulator, but oscillation in only a single mode is possible with the second.

As was illustrated in Fig. 11-11, the position of the current-voltage relations of a mode depend upon the load impedance. The undesired modes are at frequencies different from the desired mode. The impedance presented to the magnetron at those other frequencies may therefore determine in part whether the magnetron will operate stably in the desired mode.

Changes of modes may occur in two different ways. In one form the shift from one mode to the other occurs during the pulse. This is known as a *mode shift*. In the other form the frequency is stable during an individual pulse, but may shift back and forth between modes in a random fashion from one pulse to the next. This is known as a *mode skip*. The two forms of instability may occur in combination. The mode skip is the more common form. With a mode skip the voltage pulses and current pulses appear as in Fig. 11-13. With a mode shift, the voltage and current are as in Fig. 11-14. Either of these phenomena presents a problem to the magnetron and system designer that requires careful attention to detail to solve.

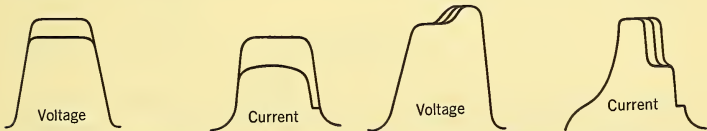


FIG. 11-13 Voltage and Current Pulses with a Mode Skip.

FIG. 11-14 Voltage and Current Pulses with a Mode Shift.

With a pulsed microwave magnetron, the cathode emission is at a high current density during the pulse. The various types of oxide cathodes that are most commonly used are capable of high peak current for only a limited length of time. If the pulse is too long, the cathode emission will be exhausted and sparking will result. The term *sparking* refers to the cathode limitation that is characterized by the physical transfer of incandescent particles of oxide coating from the cathode to the anode. Sparking may also be observed as the magnetron nears the end of its life and the emissivity of the cathode is reduced. Much effort has been devoted to cathode design and to improvement of processing techniques in recent years, and substantial improvements in cathode emissivity have resulted.

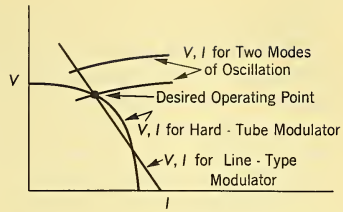


FIG. 11-12 Voltage-Current Characteristic of Two Modes of Oscillation in a Magnetron, and the Characteristic Curves of Two Modulators.

Effect of Load on Magnetron Performance. The operation of a magnetron will depend upon the load impedance that is presented to the magnetron by the output transmission line. The magnetron is normally designed to operate into a matched transmission line, and it is specified that the standing wave ratio in the transmission line may not exceed some value, usually $VSWR = 1.5$ (see Paragraph 10-12). The effects of the load may be established by considering the magnetron as a conventional self-excited power oscillator, with the LC tank circuit inductively coupled to the output transmission line. A mismatch reflected into the tank circuit will affect both output power and frequency of the oscillator. For the magnetron, the voltage-current curve is also shifted by the load impedance, and this shift may under certain conditions lead to moding.

A phenomenon that is usually more serious is the *long-line effect*. This effect is not peculiar to magnetrons, but will be found in any oscillator whose output frequency is affected by the output loading. If the magnetron is connected to a relatively long transmission line terminated in a small mismatch, the electrical length of the line (line length measured in wavelengths) will change rapidly with frequency, and the load impedance presented to the magnetron will also change rapidly with frequency. If the line length is sufficiently long, or the mismatch at the end sufficiently large, the result may be discontinuities in the tuning curve, if the magnetron is tunable, or moding when the discontinuities are approached.

The critical line length beyond which the problems associated with long lines may be encountered, is given by

$$L = \frac{130\lambda/\lambda_g}{\Delta F(\sigma^2 - 1)} \quad (11-3)$$

where L is the line length in feet, λ is the free space wavelength, λ_g is the waveguide wavelength, ΔF is the pulling figure² of the magnetron (measured at $VSWR = 1.5$), and σ is the VSWR of the termination at the far end of the transmission line.

In a practical system, when the inevitable mismatch of antennas, etc. must be located at the end of a long transmission line, the long-line effect may be avoided by using a ferrite isolator at the magnetron output.

Noise. Noise in magnetrons is important in two different forms. The first is *preoscillation noise*, or noise that is developed at voltages below that at which the magnetron breaks into oscillation. It is observed that as the voltage approaches the oscillation threshold, the noise output increases rapidly. The mechanism by which this noise is developed is not understood, but the noise itself is important as it affects the buildup of oscillations.

²The pulling figure is the maximum change in frequency that occurs as a standing wave with a VSWR of 1.5 is presented to the tube and the phase is varied through 360°.

The second form in which noise is important is in the signal-to-noise ratio of the oscillating magnetron. The noise level of the oscillating magnetron is frequently higher than can be explained theoretically, and for CW magnetrons is frequently higher than can be tolerated for CW radar systems. Magnetrons sometimes change in noisiness during life. "Quiet" magnetrons have been built, but there is still much to learn about the design and construction of magnetrons for quiet operation.

Tuning Means. A variety of means have been employed to tune magnetrons mechanically. In one form, pins are lowered into the resonant cavities of the anode block to change the inductance of these cavities. In another form, a cylinder is inserted between the two rings used for strapping, so that the strap capacitance is increased. In still another technique, a high- Q cavity is coupled to the resonant system of the magnetron and tunes the magnetron by pulling. These mechanical schemes give varying amounts of tuning up to 10 per cent or more of the center frequency. The magnetron can also be tuned over a narrow frequency range by varying the output load impedance. A number of all-electronic tuning techniques have been investigated.

Typical Performance of Modern Magnetrons. High-power pulse magnetrons for a variety of frequencies and power levels are currently in production. Typical of X-band magnetrons is the 4J50, which produces a

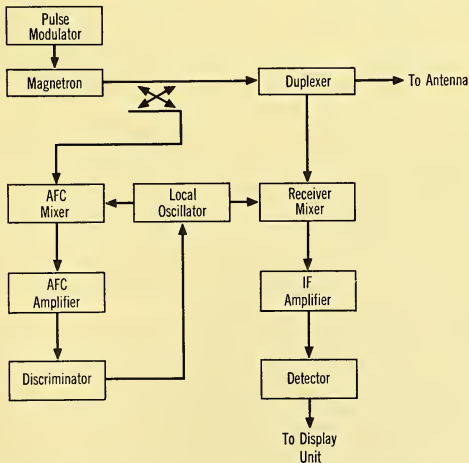


FIG. 11-15 Simplified Block Diagram of a Radar with a Magnetron Oscillator as the Transmitter.

minimum power of 225 kw at 0.001 duty cycle and can be operated at 5 μ sec pulses. At 3000 Mc, magnetrons of 5 Mw output at 0.0007 duty cycle have been developed and are in production. These values probably do not represent performance limits, although problems such as back bombardment of the cathode are sufficiently severe to limit the rate of advancement in peak and average power ratings to a relatively slow pace.

Radar System Employing a Magnetron. A simplified block diagram of a typical radar system employing a magnetron oscillator as a transmitter is shown in Fig. 11-15. The magnetron oscillator feeds through the duplexer to the antenna, and the return signal from the target passes from the duplexer to the receiver. An AFC circuit is required to keep the receiver tuned to the frequency of the magnetron oscillator.

As shown, the system is not capable of distinguishing between fixed and moving targets, although a variety of methods have been devised to refine such a system and give it MTI capabilities (see Chapter 6). Generally, these compare the phase of the received signal from one pulse with the phase of the received signal from the preceding pulse, but phase coherence is not maintained for more than two successive pulses.

11-2 THE KLYSTRON

For many years the role of the klystron in airborne radar has been that of local oscillator in the receiver. The reflex klystron oscillator has been almost universally accepted for this role. More recently the virtues of klystrons as high-power transmitting tubes have become more widely recognized, and they are being given serious consideration for this application by many designers of modern systems.

Perhaps the most important virtue of high-power klystrons is that they can be operated as amplifiers with high, stable gain. As will be discussed in detail below, this makes it possible to design relatively simple yet sophisticated radar systems such as are not possible with transmitting tubes that are self-excited oscillators. Other virtues of klystrons that cause them to be given serious consideration are relatively long operating life, high stability and low noise, and very high peak and average power capabilities.

Principles of Operation. A schematic diagram of a klystron amplifier is shown in Fig. 11-16. The cathode, focus electrode, and anode are designed so that the electrons emitted by the cathode are focused into a beam which passes through an aperture in the anode. For low- and medium-power tubes a grid may be placed across the aperture in the anode to assist in the formation of the electron beam. After passing through the aperture in the anode, the electron beam, traveling at a constant velocity, passes through the gap of a re-entrant cavity resonator. This gap is shown

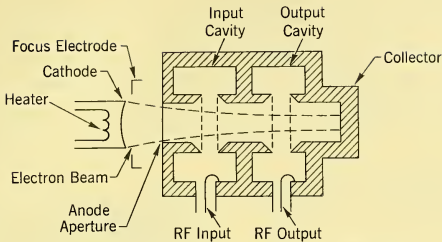


FIG. 11-16 A Two-Cavity Klystron Amplifier.

gridded in Fig. 11-16. If RF power is supplied to this first cavity resonator from an external source of power, an oscillating electromagnetic field will be developed wholly within the cavity resonator. The electromagnetic field developed by the driving signal does not reach the cathode; cathode current is therefore unaffected by the level of drive power.

With drive power supplied to the first resonator, an electromagnetic field is developed within the resonator as shown in Fig. 11-7. The electric field is largely concentrated across the gap. When the electron beam passes through this gap, the electrons are acted upon by the electric field. Electrons that pass through the gap at the instant when the electric field is in a direction to speed them up are accelerated during their passage through the gap. A half-cycle later, the field will have reversed in direction and will decelerate the electrons that pass through at that moment. When the electron beam enters the cavity gap, all electrons are traveling at approximately the same velocity. When the beam leaves the cavity gap, some electrons have been speeded up, and others have been slowed down. This action is known as *velocity modulation*.

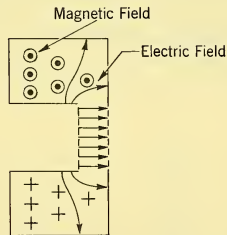


FIG. 11-17 Electromagnetic Field Inside a Re-entrant Cavity Resonator.

After the electron beam leaves the first cavity and proceeds down the drift tube (see Fig. 11-16), the electrons that have been speeded up will tend to catch up with and bunch together with the electrons that passed through the gap a half-cycle earlier but were slowed down by the field in the cavity. Clumps or bunches of electrons will form along the beam. A lumpy or bunched electron beam is called *density-modulated*. A Fourier analysis of a

density-modulated electron beam formed in this way shows that it contains a component of RF current at the frequency of the drive signal supplied to the first cavity resonator. Therefore, when a second cavity resonator similar to the first is placed along the electron beam, this RF current flowing through the impedance of the resonant circuit will develop a voltage across the gap or capacity of the cavity resonator. If an external load is coupled to this second cavity resonator, power may be delivered to the load. This power is normally greater than the drive power supplied to the first cavity resonator, and the klystron therefore acts as an amplifier. For a two-cavity amplifier as shown in Fig. 11-16, a power gain of 10 db is typical.

Klystron Oscillators. If part of the energy developed in the second cavity is fed back into the input cavity, the two-cavity klystron can be operated as an oscillator. The energy may be fed back in any of several ways. The simplest way is to cut an aperture in a common wall separating the first and second cavities so that the electromagnetic energy will couple through the aperture. Oscillations will only occur when the phase of the signal fed back to the first cavity is such as to cause regeneration rather than degeneration. The total phase delay of the feedback loop depends in part upon the transit time of the electrons between the grids of the first cavity and the grids of the second, and this transit time depends upon the voltage between cathode and anode. Modes of oscillation are found as the voltage is changed; this is illustrated in Fig. 11-18a. Fig. 11-18b shows

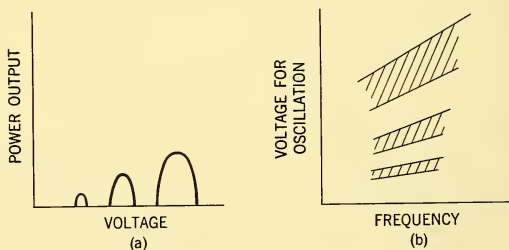


FIG. 11-18 Modes of Oscillation of a Two-Cavity Klystron Oscillator.

how this mode pattern changes with operating frequency. Because the operating voltage and output power change with frequency, and because of the mechanical difficulty of tuning the two cavities properly for oscillation, two-cavity klystron oscillators are almost never used for applications where the frequency must be changed. They have found wide application for low-power continuous-wave radars that are used in aircraft and missiles as ground-speed indicators (see Sec. 6-5). These CW radars require that the

transmitted signal be unusually stable and free from sideband noise of a character that might mask the doppler-shifted frequency that is returned from the ground. The two-cavity klystron oscillator has been operated very successfully in this application, and also in CW radar seekers for active homing missiles, where similar requirements for high stability and low noise are encountered. A typical example of a fixed-frequency two-cavity klystron oscillator is shown in Fig. 11-19.



FIG. 11-19 Two-Cavity Klystron Oscillator.

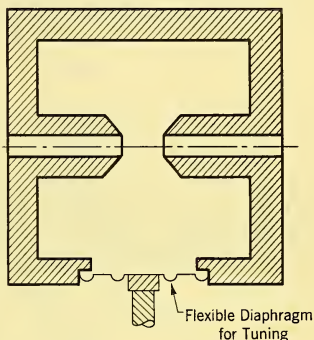


FIG. 11-20 Resonant Cavity for a High-Power Klystron with a Gridless Gap.

High-Power Klystron Characteristics. In the klystron that is shown schematically in Fig. 11-16, the electron beam is focused through grids that form the interaction gaps in each cavity. As the power level is raised and the power density in the electron beam is increased, grids cannot be built that will not be burned out by the beam. It is then necessary to have ungridded apertures at the interaction gaps. If the dimensions of these gaps are not too large, there will still be adequate interaction between the electromagnetic fields in the cavity and the electron beam. A typical cavity for high-power klystrons with a gridless gap is shown in Fig. 11-20. The power level above which it is necessary to go to gridless gaps depends upon the diameter of the electron beam, which in turn depends upon the operating frequency. At 10,000 Mc, the klystron with gridded gaps is limited to roughly 50 watts output; for high powers, gridless gaps must be used.

The two-cavity klystron amplifier of Fig. 11-16 will give roughly 10 db gain. The gain can be greatly increased by spacing additional cavities along the electron beam, as shown in Fig. 11-21. The cavities that are intermediate between input and output cavity are coupled to the electron beam

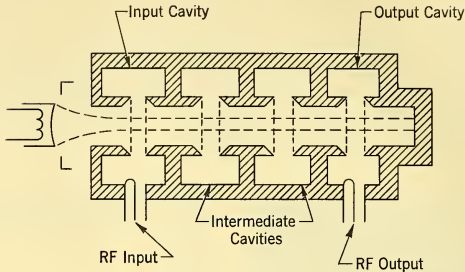


FIG. 11-21 Schematic Diagram of a Multicavity Klystron Amplifier with Higher Gain than a Two-Cavity Amplifier.

at the gaps, but are not coupled to any external sources or sinks of power. Each additional cavity that is added gives effectively another stage of gain. The exact amount of gain that can be obtained depends substantially upon the details of design, but the following data are typical:

NO. OF CAVITIES	GAIN IN DECIBELS
2	10
3	30-35
4	50-60
5	70-80

If possible sources of feedback, such as external leakage or internal secondary electrons, are avoided, the high values of gain shown can be obtained stably and reproducibly. This is often of great value to a system designer, as he can obtain kilowatts of output power with milliwatts of drive power.

For various reasons it may be desirable to limit the gain in the power klystron amplifier to something less than the maximum value. For example, the signal-to-noise ratio of the amplifier becomes poorer as the gain becomes higher. Also, the physical size and weight may become excessive. Another alternative that is available is to use several cavities in the klystron and then stagger-tune them for wide bandwidth. In a manner somewhat analogous to the well-known stagger-tuning techniques of IF amplifiers, it is also possible to stagger-tune the cavities of a klystron amplifier and trade gain for bandwidth.

In a multicavity klystron amplifier, the electron beam is necessarily rather long. The electron-optical system of the cathode and anode are normally designed to form a dense, small-diameter electron beam where the beam enters the aperture in the anode. The forces of mutual repulsion

between the electrons would cause the beam to expand and be lost to the walls of the drift tubes between cavities if additional magnetic focusing were not provided. To constrain and collimate the electron beam, the beam is brought into a strong magnetic field parallel to the beam axis. This magnetic field is continued until the beam has passed through the final cavity, and is then removed. This allows the electron beam to spread because of its own space charge until the power density in the beam is reduced to an acceptably low value, so that the beam can strike the surface of the collector without melting it. A schematic diagram of a high-power multicavity klystron amplifier with magnetic focusing is shown in Fig. 11-22.

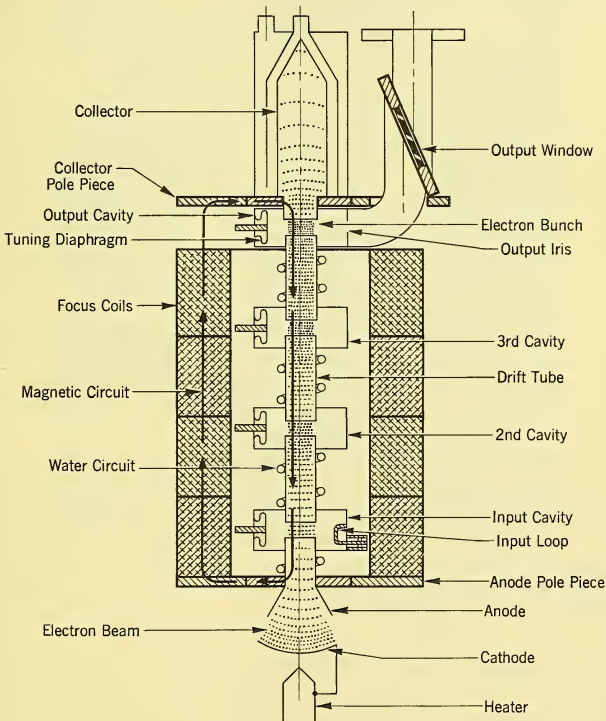


FIG. 11-22 Schematic Diagram of a High-Power, Multicavity Klystron Amplifier, with Magnetic Focusing.

The magnetic field required can be supplied with either an electromagnet or a permanent magnet. If a permanent magnet is used, the magnet weight increases very rapidly with increasing number of cavities and may become impractically large.

Comparison of Klystron and Magnetron. Compared with a magnetron, a klystron of the same power output will require higher voltage, typically by a factor of 2. Associated with the higher voltage may be a problem of X-ray radiation. Also the required modulator may be heavier and bulkier and present a more severe problem of insulation at high altitude. The efficiency of a klystron is also typically lower than that of the magnetron. For modern power klystrons, the efficiency will typically be in the range of 30 to 45 per cent, while the magnetron will fall in the range of 40 to 60 per cent. The klystron will typically also be larger and heavier than the magnetron, especially when a multicavity, high-gain klystron is being compared with a magnetron oscillator.

Power klystrons are capable of producing peak and average powers as great as or greater than the magnetron over most of the microwave frequency range. At the lower frequencies, the klystron is unquestionably superior; at millimeter wavelengths the magnetron has a decided edge. Klystrons have been built that produce peak powers of 30 Mw at frequencies of approximately 3000 Mc and peak powers of 1 Mw at 10,000 Mc. Average powers of several kilowatts have been produced at 10,000 Mc, and greater than 10 kw at 3000 Mc. These numbers compare very favorably with magnetron performance.

The operating life of high-power klystrons appears to be substantially longer than that of magnetrons. A fundamental reason for this is that the cathode is located where it can be operated at a more conservative emission density and is not subject to electron bombardment as is the magnetron cathode.

A principal difference between the klystron and the magnetron is that the klystron is usually operated as an amplifier, while the magnetron is a self-excited oscillator. The klystron is free from long-line effects that trouble the magnetron oscillator, although the advent of ferrite isolators has made it possible to operate magnetrons into long transmission lines.

Because the klystron is an amplifier, it can be used to build systems to higher performance specifications and to design more sophisticated systems. The klystron amplifier offers higher frequency stability and lower noise output. A greater variety of modulation techniques can be used. Phase coherence can be maintained over a chain of pulses, making possible a variety of systems that can separate moving targets from fixed targets. Circuit techniques can be used that eliminate the troublesome AFC circuit that controls the local oscillator when a magnetron is used.

The importance of these advantages depends upon the degree of system sophistication and performance that is desired. For the simpler systems, the magnetron is usually chosen. For more advanced systems, the benefits of using a power amplifier become increasingly important. A photograph and typical performance data of a high-gain klystron amplifier suitable for airborne radar are shown in Fig. 11-23.

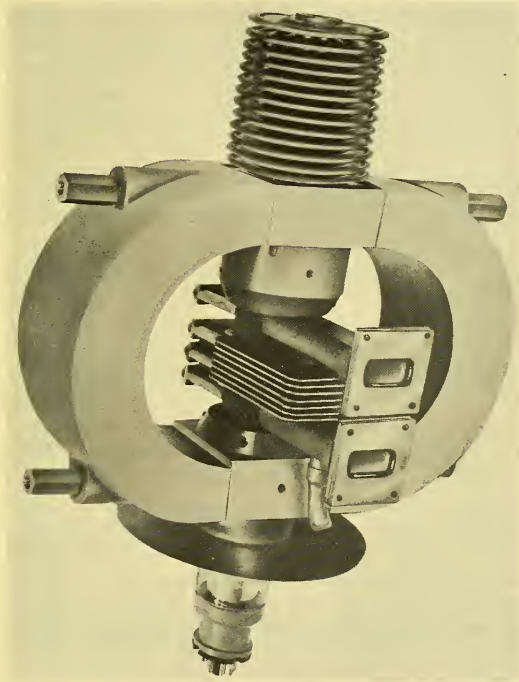


FIG. 11-23 The VA-24B Klystron Amplifier Which Produces 50 kw Power Output at 0.002 Duty Cycle, with 57-db Gain and 25-Mc Bandwidth at X-Band Frequencies.

11-3 TRAVELING WAVE TUBES FOR HIGH POWER

Although traveling wave tubes have been under intensive engineering development for more than a decade, it is only recently that much effort has been devoted to models that might be suitable for airborne radar transmitter applications. None of these high-power traveling wave tubes

has yet reached operational service, but the advantages that they offer will very probably lead to operational use at some future date.

The traveling wave tube is a beam-type amplifier as is the klystron. The high-power versions can be applied in systems in the same manner as klystron amplifiers. The most significant advantage that traveling wave tubes can offer is bandwidth. Whereas high-power pulse klystrons have demonstrated bandwidths as great as 5 per cent of the center frequency and can undoubtedly be designed for even wider bands, the traveling wave tube has greater ultimate capabilities. The wide bandwidth can be of substantial value in systems that employ frequency diversity as a means of combatting countermeasures. With wide bandwidth it is also possible to design advanced types of systems for distinguishing moving targets from ground clutter when the radar is mounted on a moving platform.

Principles of Operation. A schematic diagram of a traveling wave tube is shown in Fig. 11-24. An electron beam is formed by a cathode and

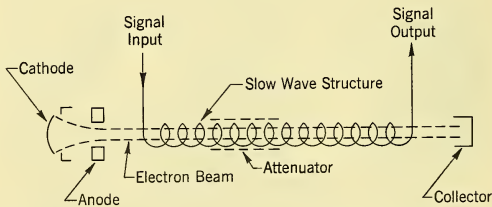


FIG. 11-24 Simplified Diagram Showing Basic Parts of a Traveling-Wave Tube.

anode structure. After passing through an aperture in the anode, the electron beam travels down the axis of a slow wave structure, shown symbolically as a helix. The slow wave structure is so designed that an electromagnetic wave propagates along it at a relatively slow velocity, a velocity nearly synchronous with that of the electron beam. When the velocity of the beam is nearly synchronous with the velocity of the propagating electromagnetic wave, the electromagnetic fields associated with the beam continuously interact with the propagating wave on the slow wave structure. As a result of this interaction, there is transfer of energy from the electron beam to the propagating electromagnetic wave. A wave that is started by a signal injected at the input end is thereby amplified and emerges from the output with greater signal strength.

A wave that is reflected from a mismatch in the output circuit will travel in the opposite direction along the slow wave structure toward the input. The feedback of energy in this reflected wave may well be sufficient to cause oscillations if additional means are not incorporated in the tube to suppress

this reflected wave. An attenuator of some sort must generally be introduced near the midpoint of this slow wave structure to absorb the reflected wave if a high-gain traveling wave tube is to be a stable amplifier. The wave traveling in the forward direction along the slow wave structure is also attenuated by the attenuator, but the signal information is carried forward through the attenuator by the modulated electron beam, which re-induces a traveling wave in the slow wave structure on the output side of the attenuator. Proper design of the attenuator is one of the most difficult problems in designing a high-power traveling wave tube.

A helix, which is shown symbolically in Fig. 11-24 as the slow wave structure, is actually used as the physical slow wave structure in low-power traveling wave tubes. It is not suitable for high-power tubes because its ability to dissipate heat is very limited and because the interaction between the electron beam and the traveling wave on a helix falls off at the very high voltages necessary for very high power. Other types of structures must be used, generally heavy metallic structures that are capable of dissipating large amounts of heat. Many such structures have been suggested, a large number of them being forms of loaded waveguide with the loading designed to reduce the phase velocity of the traveling wave tube in the waveguide structure. Structures of this class generally have useful bandwidths much less than the simple helix, but still large enough to be very interesting to system designers. Very high-power traveling wave tubes with bandwidths of 10 per cent or greater appear capable of attainment in the near future. These tubes are linear amplifiers and have demonstrated reasonably high efficiencies and stable gains of 30 db and more.

11-4 MODULATION TECHNIQUES FOR BEAM-TYPE AMPLIFIERS

One of the disadvantages of beam-type amplifiers such as the klystron and traveling wave tube as compared with the magnetron is that for the same output power, relatively high voltages are required. When cathode modulation is used, design of the modulator is usually more difficult and breakdown problems associated with high operating voltages are aggravated. For some applications it is advantageous to introduce an additional modulating electrode into the tube. The resulting complication in the tube may be more than offset by the simplification of the modulator.

Modulating electrodes can take several forms. One of these, a mesh-type control grid close to the cathode, is illustrated in Fig. 11-25. If such a control grid is curved to conform with one of the equipotentials in the electron beam, it will have little effect upon the beam shape when the potential of the grid is set at the value of that equipotential. This operating potential is normally positive with respect to the cathode potential. When

the control grid is sufficiently negative with respect to the cathode, current from the cathode will be cut off. The μ of a structure such as this is relatively high and can be greater than 100. A relatively high-power beam can therefore be switched on and off by relatively small pulses applied to the control grid. A disadvantage of this type of grid is that the grid is heated by its proximity to the cathode and is further heated by electron bombardment during the time the beam is on. If the grid is too hot as a result, there may be electron emission from the grid, which will form a low-current electron beam at a time when the beam should be completely off. An excessive amount of noise or signal may be produced between pulses as a result of this grid emission.

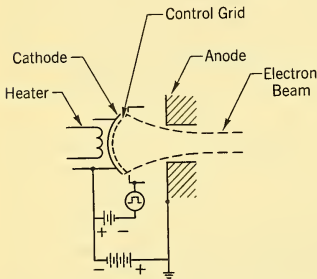


FIG. 11-25 High- μ Mesh-Type Control Grid for an Electron Beam. The Control Grid Is Biased Negative with Respect to the Cathode to Cut Off the Electron Beam, and Is Pulsed Positive with Respect to the Cathode When the Beam is Turned On.

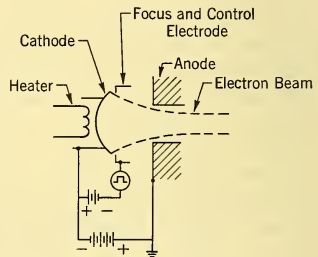


FIG. 11-26 Low- μ Control Electrode for an Electron Beam. The Control Grid Is Biased Negative with Respect to the Cathode to Cut Off the Electron Beam, and Is Pulsed up to Cathode Potential when the Beam is Turned On.

A second type of control electrode, in this instance nonintercepting, is shown in Fig. 11-26. Here the focus electrode adjacent to the cathode, which helps form the electron beam, is insulated from the cathode proper. For proper formation of the electron beam, this control electrode should be operated at cathode potential and will intercept none of the current emitted by the cathode. The beam may be turned off by biasing the control electrode negative with respect to the cathode by a voltage sufficient to shield the cathode from the positive anode voltage. A structure of this type has relatively low μ , in the order of 2, but the problem of grid emission can be avoided.

A third type, sometimes called a modulating anode, is shown in Fig. 11-27. Here the anode of the diode which forms the electron beam is

insulated from the main body of the tube which contains the RF circuitry. If this modulating anode is run at the potential of the main body of the tube, the electron beam will pass through the hole in the modulating anode and continue through the interaction circuitry. When the modulating anode is at cathode potential, it completely shields the cathode from the positive potential of the tube body and no current will be drawn from the cathode. It is not meaningful so use the technically correct definition of μ in this structure. Because the tube body is completely shielded from the cathode, the control μ is technically infinite. It is of more interest to the system designer to note that the control voltage must be relatively large. The pulse voltage will be equal to the voltage required for cathode modulation if the modulating anode is operated at tube body potential during the ON period.

The modulating anode can be so designed that it need not be operated at the tube body potential during the ON period, and the modulating voltage required can thereby be reduced by perhaps a factor of 2. More important is the fact that the modulating anode does not draw much current. Nearly all of the beam current will pass through the aperture in the modulating anode. This permits greater flexibility in modulator design.

The modulating anode can be so designed that it need not be operated at the tube body potential during the ON period, and the modulating voltage required can thereby be reduced by perhaps a factor of 2. More important is the fact that the modulating anode does not draw much current. Nearly all of the beam current will pass through the aperture in the modulating anode. This permits greater flexibility in modulator design.

11-5 A TYPICAL RADAR SYSTEM EMPLOYING A HIGH-GAIN AMPLIFIER

A typical radar system employing a high-gain amplifier is shown schematically in Fig. 11-28. A low-power stable oscillator drives an intermediate amplifier which is synchrodyne to shift the output frequency from the input frequency by an amount equal to the intermediate frequency of the receiver. The intermediate amplifier then drives the high-power, high-gain final amplifier, and the stable oscillator provides the local oscillator signal for the receiver. With this system, no AFC circuit is required to keep the receiver in tune with the transmitted signal. In addition, phase coherence is maintained continuously through a chain of transmitted pulses. A system of this type gives excellent MTI performance when properly designed.

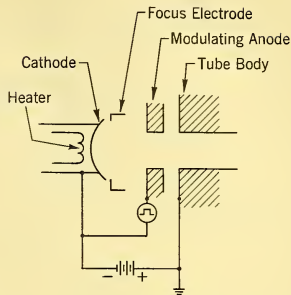


FIG. 11-27 Modulating Anode for Control of an Electron Beam. The Modulating Anode Is at Cathode Potential or Lower When the Beam Is Off, and Is Pulsed Toward the Potential of the Tube Body to Turn the Beam On.

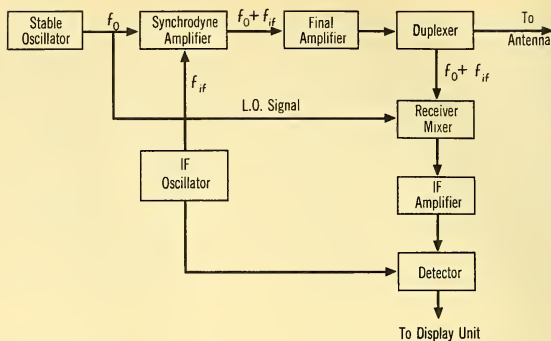


FIG. 11-28 Simplified Block Diagram of a Radar with a High-Gain Amplifier as the Power Output Tube of the Transmitter.

11-6 BACKWARD WAVE OSCILLATORS — CARCINOTRONS

Another type of microwave tube that may ultimately find application in airborne radar is the backward wave oscillator, or Carcinotron. These oscillators employ a slow wave circuit that supports an unusual type of electromagnetic wave in which the phase velocity is in one direction along the structure and the group velocity is in the opposite direction. It can be shown that all periodic, slow wave structures support these backward waves. If an electron beam is directed along the structure, oscillations are possible at a frequency at which the phase velocity of the backward wave is synchronous with the velocity of the electron beam. The phase velocity changes with frequency, and the significant feature of these oscillators is that the frequency of oscillation depends upon the velocity of the electron beam. The oscillator can be tuned over a wide range of frequencies by varying the voltage of the electron beam.

There are two types of backward wave oscillators. The first, the *axial-field type*, resembles a traveling wave tube in that the magnetic field is parallel to the electron beam, which passes along the axis of the slow wave structure. This is illustrated in Fig. 11-29. This type of oscillator is also known as a *Type-O Carcinotron*. It is generally a low-power, low-efficiency tube suitable principally for receiver and test-equipment applications.

The second type of backward wave oscillator is the *crossed-field type*, which resembles a magnetron in that the electric and magnetic fields are perpendicular to each other rather than parallel. It is illustrated in Fig. 11-30. The electron trajectories are also perpendicular to the magnetic field. The crossed-field backward wave oscillator is also called a *Type-M Carcinotron*. The efficiency of this type is relatively high, and this high

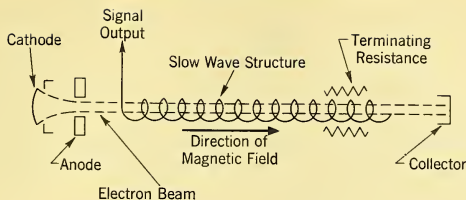


FIG. 11-29 Axial-Field Backward Wave Oscillator — Type-O Carcinotron.

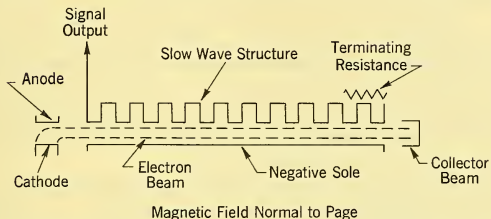


FIG. 11-30 Crossed-Field Backward Wave Oscillator — Type-M Carcinotron.

efficiency combined with wide-range electronic tuning makes these tubes attractive for countermeasure or jamming applications. Backward wave oscillators for high pulse powers have yet to be developed.

11-7 THE PLATINOTRON

The platinotron is a relatively new type of microwave power tube. It bears a strong physical resemblance to the microwave magnetron and the M-Carcinotron, but the differences in performance are very significant. In its most interesting form, known as the *amplitron*, the tube functions as a wideband amplifier and combines the magnetron virtues of high efficiency, light weight, and low operating voltage with the ability to amplify signals over a wide band of frequencies.

The basic structure of the platinotron as compared with the magnetron is shown in Fig. 11-31. The cathode structure is essentially the same in both tubes. Both tubes are immersed in a magnetic field parallel to the axis of the cathode, so that many of the comments that apply to the electron trajectories in the magnetron are equally applicable to the platinotron. The essential difference is in the anode structure. In the magnetron, the slow wave structure, generally consisting of a series of cavity resonators, is re-entrant or closed on itself, so that the RF circuitry forms a resonant system with multiple modes of oscillation. In the platinotron, the slow

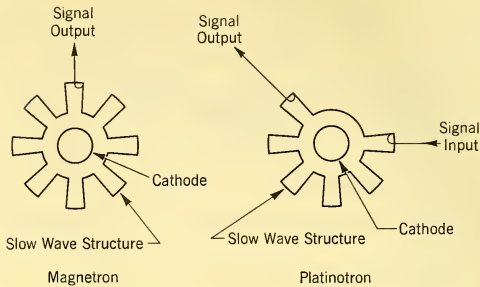


FIG. 11-31 Basic Structure of the Platinotron as Compared with the Magnetron.

wave structure in the anode is not re-entrant. It extends most of the way around the circumference of the anode, but has an input and output and is a nonresonant rather than a resonant structure.

If both input and output are externally terminated in matched terminations, there is no mechanism to determine the frequency of operation of the platinotron. If it is pulsed like a magnetron, the output will consist of noise and random frequencies of oscillation, spread over a considerable band of the spectrum. For useful operation it is necessary to add to the basic platinotron some means of determining its frequency of operation. This can be done by coupling a high- Q resonant cavity to the input end. The frequency of oscillation will then lock to the frequency of the added cavity. To the device operated in this form the term *stabilotron* is applied. The stabilotron is a relatively stable self-excited oscillator, many times more stable than the ordinary self-excited, unstabilized magnetron. Higher circuit efficiency is obtained than with the cavity-stabilized magnetron because the stabilizing cavity placed at the input to the circuit absorbs relatively little power.

The frequency of operation of the platinotron can also be established by injecting a signal into the input. The output signal will be affected only slightly until the strength of the input signal exceeds a threshold level. Above this threshold level, the output signal will be locked in frequency to the input signal, and the device will act as a saturated amplifier. To the device operated in this manner the term *amplitron* is applied. The input signal is also transmitted through the amplitron to the output, so that as the input is increased above the threshold level, the increase in input signal will be added to the output signal. A signal which is reflected from the load under these conditions of operation is transmitted in a reverse direction through the amplitron, and comes out of the input end with little attenuation.

Nearly all of the signal generated within the amplitron comes out of the output end, and very little out of the input end. When the direction of magnetic field is reversed, the functions of input and output terminals are exchanged, but there is no other effect on the performance. It has been determined that the rotating electron cloud interacts with a backward wave on the slow wave structure. However, the device differs from the M-Carcinotron in the cathode structure employed. The electron beam in the stabilotron is re-entrant; in the M-Carcinotron it is not.

The conditions under which the platinotron is operated determine the gain that can be obtained. With low modulator input and magnetic field, relatively large gain can be obtained, in the order of 20 db. The available gain is reduced as the modulator input and magnetic field are increased. Under the most efficient conditions of operation, a relatively large signal is required to lock the output frequency to the input and the available gain is only a few decibels. This characteristic is illustrated in Fig. 11-32.

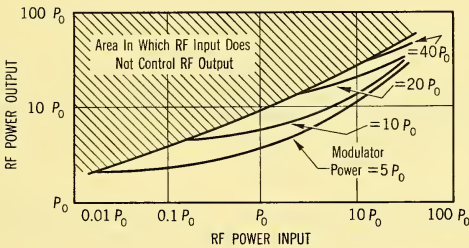


FIG. 11-32 Gain Characteristics of Amplitron.

Amplitrons can be operated in cascade to increase the gain, but this operation is made difficult by the fact that a signal reflected from the high-level output is reflected back through the chain of amplitrons and will appear as a relatively large reflected signal to the input of the chain. This situation can be improved by placing attenuators or isolators between adjacent amplitron stages.

The potential usefulness of the low-gain, high-efficiency mode of operation can be appreciated when it is remembered that the drive signal is transmitted through the amplitron with little attenuation and adds to the amplitron output. In this way the amplitron can be used to boost the power of a conventional radar. A signal reflected from the target will be transmitted back through the amplitron in the reverse direction, so that duplexing can be accomplished at the amplitron input. This can be important when the duplexer limits the power level of the radar.

Little unclassified information is available on performance of specific amplitrons. On one tube, the QK434, it has been reported that the bandwidth is at least 10 per cent centered at a frequency of 1300 Mc, and that a maximum of 10 db gain can be obtained with 1 Mw output and 67 per cent efficiency. At lower gains and higher powers, efficiencies greater than 70 per cent are observed. A block diagram of a typical radar system employing duplexing at the amplitron input is shown in Fig. 11-33.

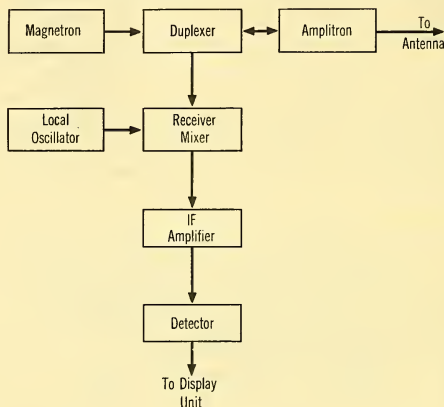


FIG. 11-33 Simplified Block-Diagram of a Typical Radar System Employing Duplexing at the Amplitron Input.

CHAPTER 12

DISPLAY SYSTEM DESIGN PROBLEMS*

12-1 INTRODUCTION

Almost every airborne radar system requires presentation of radar information to a human operator for interpretation, decision, and action.¹ With the growth of problem complexity, the development tendency has been towards greater automaticity which relieves the man of routine functions. Corresponding emphasis has been placed on utilizing man's capacity for judgment and innovation in unexpected or complex tactical situations. In this respect, man cannot be even remotely approached by the most advanced control-system technology.

The growing creative role of man as a weapon system element has increased the problem of man-machine integration to the point where it is the most difficult and perplexing of all system design problems. The display system which translates radar and other information into a form suitable for interpretation and decision is a central element of this problem.

This chapter will outline the basic features of display systems, indicate the types of display systems that are commonly employed, develop the important characteristics of display devices and their interrelations with human operations, and, finally, illustrate how display system requirements might be derived for several specific applications. The general characteristics of special display devices such as storage tubes and solid-state (electroluminescent) devices are covered to provide the reader with an introduction to present and future display system componentry. The circuit design problems of specific display systems are not treated since the purpose of this chapter is to provide principles of design applicable to the display system.

Display system design is complicated by the fact that electrical signals must be converted to light and then interpreted in the human mind. The conversion of information through these transducers is extremely complex

*Paragraphs 12-1 through 12-6 and 12-9 and 12-10 are by M. Taubenslag. Paragraph 12-7 is by Dr. A. Kahn. Paragraph 12-8 is by R. S. Raven.

¹Certain classes of guided missile terminal seekers form the major exceptions to this rule. While human intervention into the operation of these systems is often desirable, the technical problems associated with the establishment of communication links between the missile radar and a human observer can outweigh the operational advantages.

and is not as easily defined as the familiar electrical to electrical conversions in units such as the receiver. Many important facets of the problem therefore are not well understood; many others are understood in different ways by different people. This makes it difficult to compose a set of broad principles which can cover any of the conceivable situations. Thus, while the material in this chapter can provide guidance, the actual design of a specific display system for a specific application usually requires some new approaches which must be invented and proved in the course of the system design. Flight test research facilities and simulation laboratories are invaluable for collecting the necessary empirical data.

One approach to the design problem that has been quite successful is to form a balanced design team of systems engineers, human engineers (or psychologists), and display equipment design personnel. This approach facilitates the solution of the many problems posed by the strong interrelationships between system considerations, human factors, and component limitations.

12-2 USES OF DISPLAY INFORMATION

The general display problem is shown in Fig. 12-1. Information from a variety of sources — in this case an airborne radar — enters the display

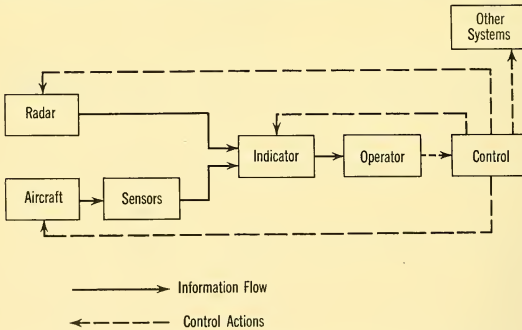


FIG. 12-1 Typical Airborne Weapons System, Showing Functions of Indicator, Operator, and Control Subsystem (Display and Control Subsystem).

system in the form of electrical signals. The display device is a *transducer* which changes the electrical signals into some form of visual indication (and may, at times, code the signals in terms of an established tactical doctrine). The human operator observes this indication, assimilates information contained thereupon, and takes some action such as positioning a control, communicating a message, etc.

Broadly speaking then, the display system may be represented as a communications link as shown in Fig. 12-2. In designing such a system, it is necessary to answer — in order — three questions:

1. What are the output requirements? That is, what is the use to which the transmitted message is to be put?
2. What is the nature of the input information (messages)?
3. How should the information be coded and transmitted to provide the most efficient link?

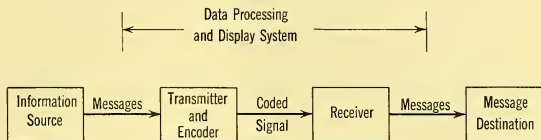


FIG. 12-2 The Data Processing and Display System Viewed as a Communications Link.

Each of these questions will now be treated in a general fashion to provide the reader with background knowledge. Then, the characteristics and problems of specific types of displays and display devices will be treated.

The weapons system application determines the use requirements of the display information. The following is a summary of some of the more important functional uses of display information in airborne weapons systems.

Detection and Identification. In many applications, the human operator serves as the final detection stage of the radar system. (This is particularly true of noncoherent pulse type radar systems.) From a display of the *radar information*, he is required to detect the presence of targets and to distinguish between bona-fide targets and signal indications caused by noise, ground clutter and other sources of interference. The advantages of such an operation lie in the human operator's superior ability to utilize size and shape of the return signal information in addition to S/N ratio. Also, as pointed out in Paragraph 3-3, the display itself can provide signal integration which enhances the S/N ratio. The principal disadvantage is the necessity for an optimum match between the characteristics of the radar, the display, and the human operator. Although detection and identification is one of the oldest uses of radar information, many facets of this problem remain unknown or unsolved.

Tactical Situation Assessment and Monitoring. A display system may combine radar and other information to construct a pictorial or symbolic representation of an overall tactical situation. For example,

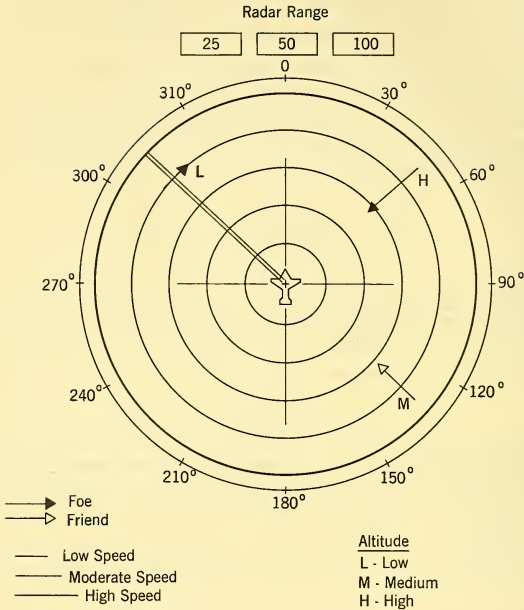


FIG. 12-3 Hypothetical Tactical Situation Display.

as shown in Fig. 12-3, the position, heading, and speed of targets in a given sector might be displayed as arrows of variable length. Friendly aircraft might be designated in the same manner by arrows of a different color. Additional coding of the symbols might be employed to indicate altitude — in this case the symbols H, M, and L are used to designate high, medium, and low altitudes. In theory, at least, such representations can provide the human operator with information necessary to make the optimum analysis of a given tactical situation. However, the generation of such displays poses difficult mechanization problems. Even more important is the problem of matching the display coding to operator interpretation characteristics. An infinite variety of display codes — all designed to accomplish the same end result as Fig. 12-3 — can be devised. For such applications, it is particularly important to define — in advance — what the operator is expected to do with the tactical display and monitoring information and to compare these use requirements with the “bandwidth” and recognition capabilities of a human being (see Paragraph 12-7). Then,

having established a set of tasks within the ken of the human operator, one can begin to devise and evaluate various clever and aesthetically pleasing means for coding the information.

Control. The purpose of many displays is to allow the operator to function as an element in a control loop. For example, the operator may be required to position the antenna or the range gate so that it brackets the target preparatory to initiating automatic lock-on. In such cases a marker (or strobe) is provided on the display which provides an indication of the actual position of the antenna or range gate — or both. The operator is provided with a control handle, button, or wheel and is required to manipulate the control to put the strobe on the target. This sounds very simple; however, experience has shown that operators are so sensitive to the coding of the strobes and the configuration of the control that the mean time to position the range and angle strobes of an AI radar can vary by 5 : 1 or more depending upon the configuration employed. The time required has a direct effect upon either the tactical utility of the weapons system or the detection requirements of the radar system. For example, in Chapter 2 the AI radar detection requirements were based on mean lock-on times of 6 to 12 seconds. At the maximum closing speed of 2000 fps, every additional 3 seconds spent fumbling with the control will increase the detection range requirements by 1 mile.

The AI radar example provides a second important example of the control function: weapon aiming control. As explained in Chapter 2, the radar information is combined with aircraft flight data and stored weapon information to produce an aiming error signal. This signal is presented to the pilot, who then attempts to fly the aircraft in such a manner that the error becomes less than 2° by the time missile launching range is reached. The display also may present an indication of when the missile can be launched and when the pilot should break off the attack to avoid collision with the target.

This operation (data interpretation and subsequent control) requires a great deal from both the operator and the display. First of all, the basic steering task itself is most difficult and normally requires additional information to produce a stable steering loop (see Paragraph 12-7). Second, in addition to the steering and firing information, various tactical situation indications such as range, range rate, and bearing angle must be presented. (With respect to these, the previous remarks on tactical situation displays apply.)

Flight control and navigation functions form another important area for the utilization of radar display information for control purposes. These functions may require a radar map which allows the pilot to navigate in the same manner as if visual contact were maintained.

Data Read-Out. In many applications — notably, airborne early warning systems and semiautomatic bombing systems — the radar operator functions as a data processor and filter. For example, in the AEW system discussed in Chapter 2, the operator(s) might be presented with a display of target azimuth, range, and height. The operator then is required to “read out” these data in a form that can provide inputs for an air-to-ship communications link.

A general diagram of such a system is shown in Fig. 12-4a. The operator “reads” the display, selects the desired information, and codes into a form

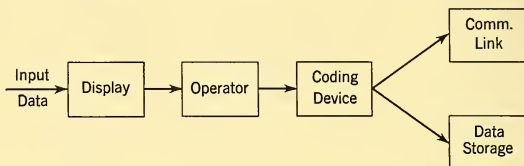


FIG. 12-4a Data Read-Out System.

suitable for communication and/or storage. This type of display problem combines many of the features discussed in the previous applications. First of all, the operator may have to use size, shape, and contrast information to distinguish between desired and undesired data. He may also be required to measure or estimate the coordinates of the target. Second, in applications such as this, the operator usually is presented with a large amount of data. Thus, many of the same considerations discussed for tactical situation assessment and monitoring apply.

Finally, the operator is expected to perform some function in a more or less continuous fashion to keep up with the flow of data. This process involves many of the same problems as the control function previously discussed. The transfer characteristics of the operator — bandwidth, time lag, etc. — are vitally important. These characteristics will, of course, vary with the type of physical action required to accomplish the coding.

An important variation of the data read-out problem is shown in Fig. 12-4b. In this case, the displayed information is recorded photographically. The operator then interprets the photograph rather than the display itself. Many ground-mapping systems utilize display information in this manner. Systems such as this permit time-scale changes between the rate of data

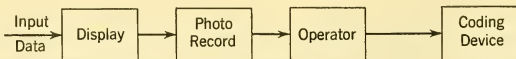


FIG. 12-4b Data Storage and Read-Out System.

collection and data processing. For example, a ground mapping radar may produce 100,000 picture elements per second. As will be shown in Paragraph 12-7, a human operator can interpret information at the rate of about 30 bits a second. Obviously a large time-scale change is required if full use is to be made of the radar information.

12-3 TYPES OF DISPLAYS

The basic types of displays are categorized by (1) the method of signal insertion, (2) the method of deflecting the signal into an understandable pattern on the indicator, and (3) the type of information that is displayed.

The display device itself is some form of electrical-to-light transducer. For most applications, the cathode ray tube — or one of its many variations — is the most readily adaptable element. Thus, this discussion will be confined to the types of display presentations which can be utilized with a cathode ray tube. Aural presentations and other types of optical displays will be discussed at the end of the chapter.

In a cathode ray tube, the intelligence signal can be used either to move or to intensity-modulate the electron beam. *Amplitude-modulated displays* are achieved on the cathode ray tube (CRT) by insertion of the signal in the deflection circuit. The signal amplitude is shown in such a display as a variable along a time base sweep. *Intensity-modulated displays*, on the other hand, are achieved by modulating the cathode ray tube electron beam by inserting the signal on the grid or cathode of the tube. Amplitude of the signal is therefore displayed as a variation in light intensity on the CRT phosphor screen.

Amplitude-Modulated Display. The simplest and oldest radar indicator utilized a *type A presentation* in which only target range is displayed. Fig. 12-5 shows this presentation. The sweep of the A scope starts simultaneously with the transmitted pulse and moves linearly with time from left to right. The video signals (target returns) are impressed on the vertical deflection plates. They give a short vertical deflection (pip) at distances along the horizontal trace corresponding to the two-way travel time of the transmitted pulse.

Between sweeps the cathode ray tube is biased to cutoff by applying a negative pulse to the grid of the tube. The biasing of the tube improves the display by blanking out the illumination that would ordinarily build up (incident to receiver noise) at the start of the sweep. Other signals that may be added to this display include range marks, which are displayed along the range axis, and range notches which can be strobed along the axis to measure the precise position of a target. This position is measured as a function of the d-c voltage required to move the notch to the target position. The range marks and range notches have been displayed in the type A scope

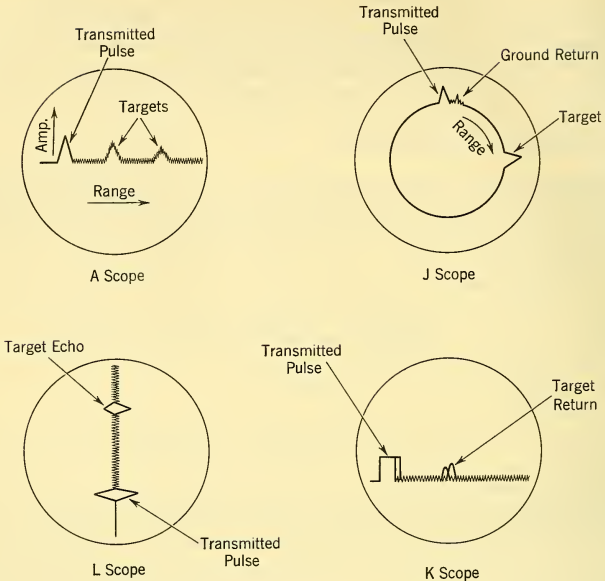


FIG. 12-5 Amplitude-Modulated Displays.

as either intensity-modulated signals (applied as positive signals on the grid of the CRT) or as amplitude-modulated signals (applied to the vertical deflection plates together with the receiver video signals).

The type A display was first used because of its simplicity of mechanization. Its limitation in wide-angle coverage radar systems is apparent, for it is incapable of presenting the target directional coordinates. The advantages of the type A scope are numerous. It is easily interpreted (all oscilloscopes use this type of sweep). The amplitude modulation of the A scope has better bandpass characteristics than intensity-modulated presentations because the signal is traced along a longer path on the phosphor screen than the signal of an intensity-modulated presentation. Therefore, there is not as much signal compression as in other displays. It is the most adaptable scope for observing signals through interference for this reason. Lastly, the dynamic range of signal display is greatest of all the displays because none of the cathode ray tube electrical characteristics except the deflection sensitivity enter into the modulation reproduction. The signal may be displayed along the full scope face if the video amplifier

has enough gain; thus the signal drive limitation is not in the tube. This is not true of intensity-modulated displays.

A modification of the type A scope, the *type J*, increases the length of the time base (Fig. 12-5). All the disadvantages and advantages of the type A presentation exist, with the additional advantage that a longer time base can be displayed for the same diameter of tube. There are also additional disadvantages in that system circuitry complexity is increased and a special cathode ray tube is required.²

Other modifications of the type A are the types K and L displays. These presentations are used to obtain directional information of a target when two antennas or a sequential antenna lobing system is used. The antenna beam patterns are offset so as to make a small angle between the two patterns. The ON TARGET position is obviously that position which is equally illuminated by both antenna beams. This would exist only when the target is along the centerline of the sequentially lobed antenna (or along the centerline between two separate antennas).

As shown in Fig. 12-5, the *type L scope* presents the target information from each beam on the cathode ray tube in a back-to-back manner so that correction can be made to equalize the video signals and thereby align the antenna system to target. In the mechanization of sequential lobing and antenna switching, alternating pulses are used for the right and left antenna beams. The display is so arranged that the signal appearing on each side of the display is associated with the proper antenna beam. The antenna motion necessary to align the antenna and target is then evident from the scope.

The *type K scope* is a variation of the type L scope. Instead of a back-to-back display, a side-by-side display is used. The sequence of antenna lobe switching is similar to the L scope. This display is shown in Fig. 12-5. Neither the type L nor type K scopes are used much in airborne installations, except for programed flights at low levels.

Intensity-Modulated Displays. In order to give additional information on the two-dimensional cathode ray tube, grid or intensity modulation is provided in the types B, C, and PPI scopes; thus the two dimensions of the CRT face are available to display pertinent target information other than signal-return intensity (for example, range, azimuth, elevation, height).

The *type B presentation* is a form of intensity-modulated display which is commonly used in airborne radar search displays. The indicator presents a cartesian plot of antenna azimuth position along the horizontal axis and range on the vertical axis. The video information is used to intensity-

²L. N. Ridenour, *Radar System Engineering*, pp. 166-167, McGraw-Hill Book Co., Inc., New York, 1947.

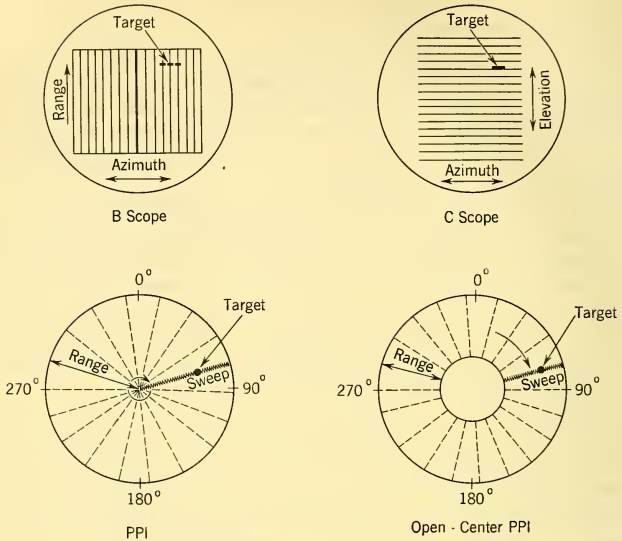


FIG. 12-6 Intensity-Modulated Displays.

modulate the electron beam, as shown in Fig. 12-6. The antenna azimuth motion is measured to provide a horizontal deflection signal.

The *type C presentation* is a plot of elevation angle on the ordinate versus azimuth angle on the abscissa, as in Fig. 12-6. This presentation is commonly employed for the display of angular tracking information such as target angular position and computed steering error signals. It is not normally used for target detection because all of the noise from one range sweep piles up on one spot to compete with the signal return, if any, from a given angular location. Thus, in comparison with the B scope, the signal-to-noise ratio of the C scope is degraded by a factor of the order of

$$S/N \text{ degradation factor for C-scope presentation} = (\tau f_r)^k \quad (12-1)$$

where τ = pulse length

f_r = repetition frequency

$K \doteq (0.5 - 0.6)$ (see Sec. 3-5)

In the *PPI (plan position indicator)*, range and bearing are presented in polar coordinates. The beam of the cathode ray tube starts from the center of the screen when the transmitted pulse is generated and is swept radially

as shown in Fig. 12-6. The radial direction of the sweep corresponds to the true bearing of the antenna. The video is presented as an intensity-modulated signal similar to the B or C scope. The distance from the center of the screen to the intensity-modulated signal indicates the range of the target. The relative bearing of the target is given as the angle between forward heading (direction is such that forward position is at the top of the display) and the position of the target as shown in Fig. 12.6.

The antenna for a PPI is driven in the horizontal plane at a speed of 1 to 30 rpm. Since the direction of radiation changes as the antenna is rotated, the sweep on the scope is rotated so as to be always correlated with the direction of radiation. The PRF is very high compared with the frequency of rotation; therefore the antenna is relatively motionless during a radial range sweep. The displacement between sweeps is also relatively short. If, for example, the antenna is rotated at 20 rpm and the PRF is 800 cps, then the radial changes in sweep position (and change in antenna position) between transmitted pulses is 0.15° . This radial change is directly proportional to the antenna speed and inversely proportional to the PRF.

The high azimuth resolution coupled with the wide-angular coverage makes this type of display particularly adaptable to AEW and mapping or navigational radar systems. One of the disadvantages of the polar coordinates plot is the poor resolution at short ranges. This problem is particularly important for ground-mapping radar displays which employ very narrow antenna beamwidths.

At only one range can the angular sector corresponding to the azimuth beamwidth of the antenna and the spot size be equal. Below this range, the spot size determines the azimuth resolution rather than the radar. Beyond this range, the radar angular resolution is the determining factor. These conditions are illustrated in Fig. 12-7. The degree of poor focus is determined by a number of variables: the spot size, the azimuth beamwidth, the maximum range displayed, the physical length of the sweep, the angle of sector scanned, and the angle of display sector.

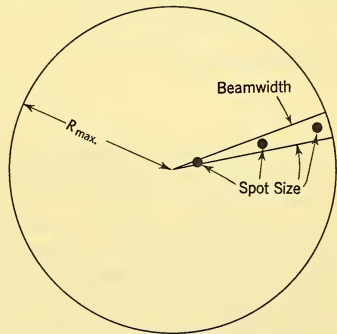


FIG. 12-7 Resolution Limitation of PPI Display.

One method of improving the azimuth resolution at short ranges is by mechanizing an *open center PPI*. This type of display, as shown in Fig. 12-6, overcomes the minimum spot size limitation by changing the zero range origin out beyond the point where the angular resolution of the

antenna is not obscured by the minimum spot size. This type of display distorts the very near information since all points along the circumference of the circle are in reality a common point (the transmitting point). This type of display is used sparingly in conjunction with the conventional PPI as a closing stage in navigational or homing displays.

Other general radar displays have been cataloged³ such as the E scope for height finding, the F scope, G scope, and I scope for error correcting, the M scope and N scope for range reading. These presentations are not common in airborne radar displays.

Several general comments which apply to all types of intensity-modulated displays are in order. A sacrifice in *bandwidth* is made in the intensity-modulated display as compared with the type A scope since the phosphor excitation is more confined. When the spot size is greater than the product of pulse length and sweep speed, then several spots overlap, integrating the video information contained in the overlap period. Thus the display acts like a low-pass filter with a buildup time of

$$T_{bu} = \frac{d}{S} \quad (12-2)$$

where T_{bu} = display build-up time, sec

d = spot size, cm

S = sweep speed, cm/sec

or a bandwidth of

$$B_d = 1/T_{bu} \text{ cps.} \quad (12-3)$$

When the pulse to be reproduced has a length greater than the display filter buildup time (or has a bandwidth less than that of the display), then it is passed (displayed) with little distortion and no adjacent noise pulse integration occurs. However, when the pulse length is shorter than T_{bu} (or the signal bandwidth is greater than that of the display), then the signal is distorted and adjacent signal and/or noise pulses are integrated, degrading detection capability and possibly picture quality.

The *voltage-to-brightness transfer function* of the display device is also important for intensity-modulated displays. A square-law relationship, for example, tends to preserve a linear relationship between signal power and brightness. This factor is discussed in more detail in Paragraph 12-6.

Another important factor is the *total brightness range*. If this is smaller than the range of signal amplitudes, then video compression will occur, with a consequent loss of some of the signal intelligence.

³See Chapter 1 of reference 1 given at the end of this chapter.

Combination Displays. In almost all radar systems the indicator displays tactical and control information to the operator or pilot as well as radar information. For example, in the operational AI radar systems the indicator displays antenna position, IFF information, aircraft guidance in the track mode, breakway time and aircraft attitude as well as the raw radar data. Information coding of these signals given as circles, crosses, and bars.

With the advance of the concept of the contact analogue display (Paragraph 12-10) and complex radar systems where computers are required to sort out even raw radar data, additional dimension is given to the symbolic logic by adding letters, numbers, and color to the method of data presentation. Only special tubes, now in advanced stages of development, are capable of presenting these new symbolic concepts. We shall discuss color tubes, character type presentation tubes, and electroluminescence displays in Paragraph 12-9.

12-4 TYPES OF INPUT INFORMATION

The preceding paragraphs have indicated many of the more important types of inputs to radar displays. This paragraph will discuss these inputs with the object of pointing out some of their salient characteristics which can affect the display design problem. The primary purpose is to emphasize the importance of obtaining a definition of display system inputs as a prelude to system design.

Three basic types of input signals are available from the radar and the associated data processing system:

1. Video signals
2. Angle, range, and velocity reference signals
3. Computed quantities

In addition to the radar signals, the display may be required to handle information from a number of other sources — e.g., aircraft instrument signals, data link information, and equipment test and check-out signals.

Radar Signals. The radar video signals and their associated range, angle, and velocity reference signals together constitute an analogue representation of the radar system's information matrix (see Paragraph 6-2). From the standpoint of the display device, the most important properties of these input signals are:

1. Number of matrix elements
2. Number of matrix dimensions and the number of elements per dimension

3. Bandwidth
4. Number of amplitude levels
5. Linearity, accuracy, and resolution of the reference signals

The total number of matrix elements provides a measure of the total number of "chunks"⁴ of information that the display system may have to handle. This number is instrumental in establishing the required size of the indicator.

The number of matrix dimensions and the number of elements per dimension have a strong influence on the type of display to be used. As was indicated in Paragraph 12-3, the display device is usually two-dimensional. Additional information dimensions are supplied by coding the two-dimensional symbols or by providing additional displays. Regardless of the means employed, the extra dimensions are supplied at a considerable price in complexity and operator interpretability. This price tends to increase with the number of separate elements in each of the extra dimensions. Thus there is a strong premium on using the indicator dimensions for those information dimensions which have the largest number of elements. The extensive use of the B scope for AI radars illustrates this principle. Such a radar may have 100 range elements, 30 azimuth angle elements, and 2-3 elevation angle elements. Customarily, range and azimuth are displayed on the indicator dimensions, while a separate strobe is used to indicate elevation. The bandwidth of the input information is governed by the video bandwidth of the radar receiver (or by the devices placed between the video amplifier and the indicator). The number of amplitude levels is governed by the S/N ratio. For example, a maximum S/N ratio of 20 db implies that there are ten distinguishable voltage levels in the incoming signal.

Closely related to questions of bandwidth and number of levels is the character of the video itself. Three types of video may be used as an input to the display system: (1) raw video, (2) processed video, and (3) artificial video.

Raw video is obtained directly from the radar video amplifier output. This signal contains the target information and internally and externally generated noise. The presence of noise complicates the indicator design problem particularly in the detection phase where operation of very low S/N ratios is required to obtain maximum performance from the radar. This was indicated in Paragraph 12-3 in the discussion of the use of C scopes for detection.

In some cases, the raw video may be processed in some fashion before display. Typical processes are (1) pulse stretch (see Paragraph 5-6),

⁴A *chunk* of information may be defined as a separate message. For example, the range, azimuth, elevation, velocity, and amplitude of a single target is a chunk of information.

(2) pulse length discrimination, (3) integration, (4) amplitude limiting, (5) noise clipping, (6) logarithmic amplifications. Each of these processes may affect the signal-to-noise ratio, the bandwidth, and the size and shape of the video pulse information. The methods outlined in Chapter 5 must be employed to form an analytical definition of the display system input.

Certain types of radar systems — notably doppler systems — may require the generation of artificial video — i.e., video that is generated to represent signals already detected. In these applications, the concept of S/N ratio no longer has meaning since the artificial signals are generated at an arbitrarily high S/N ratio to ensure that only signal information appears on the display. With respect to the requirements imposed on the indicating device, this type of signal poses the same sort of problems as the display of instrument readings and data link information.

Computed Signals. In many important applications the radar video information is processed automatically and used in conjunction with other information to compute various quantities which are then displayed to the operator. The tracking phase of the interceptor fire-control problem represents such a case. The signal sent to the display will be affected by the S/N ratio of the radar information as modified by the processes of computation and smoothing. For example, range and angle tracking noise will result in jitter of the computed steering indication. Filtering can cause a time delay detrimental to steering loop stability.

In addition to the noise and time delay, it is important to define the following other properties of computed signal inputs:

- | | |
|--------------------------|-----------------------------|
| 1. Total amplitude range | 3. Linearity and resolution |
| 2. Bandwidth | 4. Number of dimensions |

For the display of information derived from sources other than radar, similar comments apply.

12-5 THE CATHODE RAY TUBE

The general elements of the display system problem have been discussed and some of the more common types of displays have been illustrated. Now we shall turn our attention to the display device itself to ascertain the characteristics that are most important to the solution of the display problem.

To provide a reference for this discussion, the important general characteristics of display devices will be illustrated by the specific performance characteristics of cathode ray tubes. Despite the development of many more esoteric display devices, the cathode ray tube is still the most generally useful radar display device; moreover, an understanding of the operation

and limitations of the cathode ray tube contributes to a proper understanding of other visual display devices.

As previously mentioned, there will be no discussion of display system circuitry or the construction and operating details of the display devices themselves. A bibliography at the end of this chapter provides the interested reader with appropriate references for such material.

The cathode ray tube (CRT) is basically a two-dimensional indicating device which paints a picture on a phosphorescent screen. All CRT's contain three major sections:

1. A combination of electrodes, called an electron gun, for producing and focusing an electron beam
2. A means for deflecting the electron beam to a selected area of the fluorescent screen
3. A fluorescent screen which is luminous when excited by an electron beam.

These items will now be separately discussed.

The Electron Gun. The electron gun includes a thermionic source of electrons (the cathode). This cathode is usually a small capped cylinder of sheet nickel which is oxide coated to provide an electron-emitting material. The cathode is indirectly heated by a filament. The cathode element is surrounded by a cylinder which acts as a tube shield and increases the cathode thermal efficiency. A grid for controlling the electron volume density of the beam is placed in front of the cathode. This electrode takes the form of a cylindrical can completely surrounding the cathode, having a circular aperture in front of the emitting cap of the cathode.

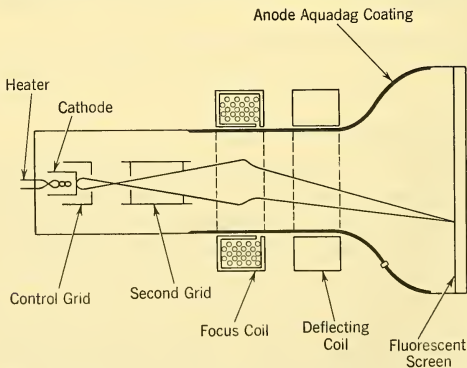


FIG. 12-8 Magnetically Focused CRT.

Focusing of the electron beam is accomplished by either electric or magnetic fields. A magnetically focused CRT is shown in Fig. 12-8. The control grid is followed by a second grid which serves to screen the region between the cathode and control grid from the high accelerating potential of the anode and thereby reduces the control voltage necessary to cut off the beam current. This electrode serves no focusing function. Focusing in a magnetic system is accomplished by means of a longitudinal magnetic field (of circular symmetry) which increases in intensity from the center to the edge of the tube. External coils or permanent magnets are used to produce this field. Electrons moving exactly along the axis of the tube pass through the field with no deflection because they move parallel to the magnetic field at all times. Electrons which are diverging from the axis of the tube as they leave the cathode have a radial velocity component and are acted on by the magnetic field in such a way as to cause rotation of these electrons about axes parallel to the tube. The magnetic forces acting on the diverging electrons produce the inward radial accelerations necessary to reverse the initial diverging trajectory of the electrons to bring the beam to a focus on the fluorescent screen.

When an electromagnet is used to produce the field, the coil is wound with many turns of wire on a soft iron ring provided with an annular gap. The strength of the magnetic field and therefore the focusing is controlled by adjusting the direct current passed through the coil. Instead of a focusing coil, a permanent magnet may also be used. This magnet is made in the form of a two-section annular ring. Focusing is adjusted by changing the air gap between the pole pieces.

The electrostatic focusing scheme utilizes an indirectly heated cathode a control grid, a first or focusing anode, and a second or accelerating anode

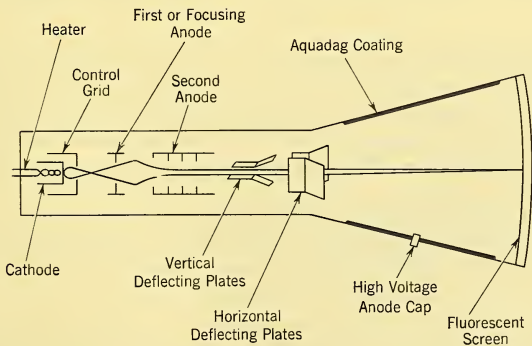


FIG. 12-9 Electrostatic Cathode Ray Tube.

The electrostatic tube is shown in Fig. 12-9. The cathode, control grid, and anodes are cylindrical in shape, and their axes coincide with the axis of the tube. The voltages which are applied to the first and second anodes accelerate the electrons and cause them to be focused into a narrow beam. Focusing is usually controlled by adjusting the first anode voltage; some focusing action can be effected by varying the voltage of the second anode.

Beam Deflection. Deflection of the cathode ray beam can be accomplished by either a magnetic or an electric field. In the magnetic field deflection system the electron beam is deflected by a magnetic field which results from passing current through a set of deflection coils. The coils are multilayer windings on either a square or a circular iron core as shown in Fig. 12-10. Because of the large air gap and the low magnetizing force that

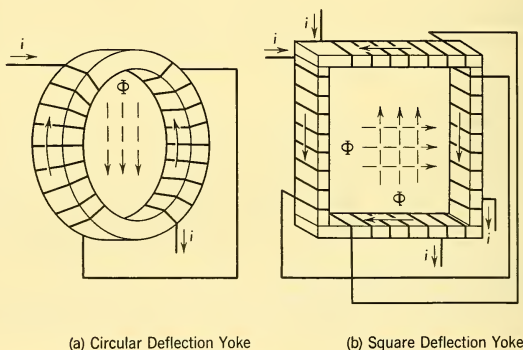


FIG. 12-10 Types of Deflection Yokes.

is usually used, the iron core serves more to shape the magnetic field than to increase the sensitivity. The two coils of Fig. 12-10a are connected so that their developed fluxes are in opposition around the iron core but add along the vertical diameter; thus horizontal deflection occurs. The square yoke of Fig. 12-10b may be used to provide both horizontal and vertical deflection in a single unit. The two horizontal coils produce a horizontal field (and vertical deflection) while the two vertical coils produce a vertical field (and horizontal deflection). The deflection produced by a magnetic field is normal to the direction of the field. The magnitude of deflection is a linear function of the magnetic field strength, which is in turn proportional to the coil current.

Deflection of the electron beam by an electric field is accomplished by developing an electric potential between two pairs of plates located just

beyond the second anode. The two sets of plates are so mounted at right angles that the plane of one set of plates is vertical and the plane of the other set is horizontal. The field is normal to the surfaces of the deflecting electrodes. The lateral movement of the electron beam is a result of the electric field which is perpendicular to the direction of the electron beam. The magnitude of the deflection of the electron beam is directly proportional to the deflecting voltage and inversely proportional to the voltage applied between the second anode and cathode.

Since the brightness of the luminescent screen increases as a function of the electron volts which strike the screen and this energy is a direct function of second anode to cathode potential, it follows that higher brightness is attained at the expense of deflection sensitivity. This difficulty is overcome by adding a postdeflection accelerating voltage (some defocusing is caused by this method of increasing sensitivity while retaining high brightness). The auxiliary high voltage anode placed near the screen consists of an aquadag coating painted on the inside of the tube. *Aquadag* is a colloidal solution of graphite. Although it forms a high-resistance coating, the current density of this coating is small and the voltage drop is negligible.

Some of the basic and important differences between magnetic and electrostatic CRT's may be listed as follows:

- (a) At the excitation levels of intensity-modulated displays, magnetic tubes provide better focus.
- (b) Electrostatic tubes are longer, but the overall equipment has smaller size, weight, and power dissipation than the magnetic tube.
- (c) The grid modulation characteristic (beam current versus grid bias) of most electrostatic tubes follows a square law, while that of the magnetic tube follows a cube law with the reservation quoted above.
- (d) Multiple presentations can be more easily displayed on electrostatic tubes because several guns may be accommodated in the same glass envelope and operated independently without interfering noticeably with each other. This cannot be done with magnetic deflecting tubes because the magnetic yoke encompasses the entire tube and would mutually affect multiple guns.

Because of size limitations in airborne indicators, and because the usual radar display uses multiple presentations of radar information and tactical symbols, the electrostatic tube is most commonly used.

Fluorescent Screens. Much of the ability to display the signal information content is dependent on the phosphor. Persistence, resolution, light level, and contrast are functions of the phosphor quality.

The usefulness of the cathode ray tube is dependent on the fluorescence, phosphorescence, spot resolution, and secondary emission characteristics of the phosphor material. Some terms commonly used in describing the properties of screen phosphors are luminescence, fluorescence, and phosphorescence. *Luminescence* is the emission of visible to near visible radiation in phosphors incident to excitation. Some methods of excitation are: (a) by bombardment of electrons (cathode luminescence), (b) by an electrical field (electroluminescence), and (c) by radiation (photoluminescence). *Fluorescence* is luminescence occurring during excitation. In cathode luminescence — the method of excitation of all cathode ray tubes — fluorescence is the light emitted *during* the period of electron bombardment. *Phosphorescence* is luminescence occurring *after* bombardment has ceased.

Many nonmetallic inorganic crystals exhibit the properties of luminescence when bombarded by an electron beam. However, most do not have the properties of brightness, long phosphorescence, and suitable color necessary in cathode ray tubes. The chemical makeup of most practical phosphors includes a base material and an activator. The base materials are usually the oxides and sulfides of zinc, cadmium, magnesium, and silicon. The activator is mixed in the ratio of 10 to 100 ppm with the base material to give a greatly increased light output. This activator supplies the additional energy levels necessary to make the jump between energy levels (caused by the excited electrons of the beam) small enough to give off light efficiently.

Phosphors are available in almost every color. A few of the phosphors used in cathode ray tube indicators are listed in Table 12-1.

TABLE 12-1 PHOSPHOR CHARACTERISTICS

Type of Phosphor	Fluorescent Color	Phosphorescent Color	Time Persistence ^a (sec)	Decay Law
P_1	Green	Green	Medium 0.05	e^{-t}
P_2	Green-blue	Green	Long 0.5	t^{-n}
P_4	White	White	Medium 0.06	e^{-t}, t^{-n}
P_7	Blue-white	Yellow	Long 3.0	t^{-n}
P_{11}	Blue	Blue	Short 0.005	t^{-n}

^aTo 1 per cent of peak

This table indicates the relative time duration of phosphorescence for these phosphors and the decay laws they follow. For many radar displays, long persistence is required because of the long frame time of one antenna scan. Yellow-green sulfides have the required long persistence if they

are excited by light. To obtain both flash highlight intensity and long persistence, some phosphors such as the P_7 are composed of two materials in cascade, i.e., two layers. The layer nearer the electron gun ($ZnS:Ag$) fluoresces with a high, blue light intensity. This light in turn excites the longer-persistence yellow phosphor, thereby giving long persistence. If the yellow layer ($ZnCdS:Cu$) were bombarded by the electron beam directly, it would give a very bright flash but would have shorter persistence.

12-6 IMPORTANT CHARACTERISTICS OF ELECTRICAL-TO-LIGHT TRANSDUCERS

The most important characteristics of a display device which affect the radar system design are:

- (a) Resolution and display size
- (b) Indicator and background brightness
- (c) Signal integration properties of the tube
- (d) Tube size and form factor
- (e) Susceptibility to physical environment

Resolution and Display Size. Previous chapters have indicated the high price paid for range and angle resolution in terms of system bandwidth and antenna aperture size. It is often important, therefore, that the resolution of the display device be such that it is not the limiting factor in determining overall system resolution. Several simple examples will illustrate the possible nature of the resolution problem.

Example 1. Assume a side-looking ground-mapping system designed to provide a high resolution map of a strip twenty miles wide on each side of the aircraft (see Paragraphs 1-4 and 14-3 for more details concerning this type of radar system). Assume the beamwidth and pulselength of this radar are such as to produce an average resolution element of 100 ft on a side. The instantaneous output of each side of such a radar system may be presented as an intensity-modulated line on the face of a cathode ray tube as shown in Fig. 12-11. The line may be photographed on a moving film strip which is synchronized with aircraft motion to produce a continuous map.

From the data given, it can be seen that the line will contain $(20)(6020) / 100 = 1200$ resolvable range elements. If the line is 6 inches long, there are 200 resolvable range elements per inch, provided the indicator does not limit resolution. To achieve this latter condition, the indicator must have 400 resolvable elements per inch,⁵ i.e. a spot size of 0.0025 inch or 2.5 mils.

⁵The display may be viewed as a sampling device. In accordance with the sampling theorem outlined in Paragraph 3-3, the effective sampling frequency must be twice the input bandwidth. Hence, the size of each sampling element must be half the size of the input elements.

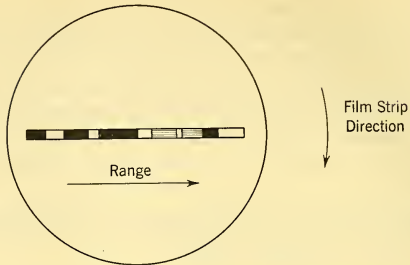


FIG. 12-11 Intensity-Modulated Ground-Mapping Display.

Example 2. As a second example, the AI radar discussed in Chapters 2 and 3 might be considered. This system employed a beamwidth of about 3.6° , a pulse length of 1 microsec, an azimuth coverage of 120° , and a total surveillance range of about 25 miles. Thus, a B-scope search display (see Fig. 12-6) must be capable of handling 33 angular elements and 300 range elements. If the display height is 5 inches, then there are 60 range elements per inch — a resolution capability that may be preserved by an indicator with 120 resolvable elements per inch. In this case, however, it is important to note that radar resolution is not a vital factor for mission accomplishment — the radar antenna size and pulse length were governed by other considerations such as detection range, and the radar resolution capability need only be great enough to resolve targets separated by several miles. Thus, in contrast to the first example, the system performance is relatively insensitive to indicator resolution, and this property of the indicator may be traded for other properties more important to the AI radar detection problem.

Relative to requirements of this type, it is of interest to consider the typical resolution characteristics of cathode ray tubes. These devices provide the highest resolution of all the various types of display devices. The spot size of the CRT is obviously a determining factor in the resolution of the display. The smallest spot size obtainable at this time with careful control of the electron-optics system and at a low beam current is approximately 0.8 mil at the center of the screen. Tubes having such small spot size utilize magnetic focus. However, to reduce misalignment problems and weight, electrostatic focus is used in most airborne cathode ray tubes. At low beam currents and high anode potentials spot sizes of 1.5 mils have been measured at the center of the scope. Usually, though, with the cascaded phosphors which are used in most radar indicators, spot sizes of 10 to 40 mils are measured, depending on the beam current. The spot size is degraded as the spot is deflected away from the center of the cathode ray

tube since the curvature of the phosphor front face does not always coincide with the surface of best focus.

The major effects limiting spot size are:

1. Random direction of the electron emission
2. Space charge limitations
3. Halation
4. Light scattering

Focus is obviously dependent on the electron optics of the CRT. Because of the random manner in which electrons are emitted from the cathode, the electron beam has a finite width to begin with. This width cannot be improved on by the electron-optics system — in fact, it is increased.

The space charge effect on spot size is a result of the mutual repulsion between electrons in the beam, preventing the electrons from coming together at one point. As the beam electrons converge to a smaller diameter, the mutual repulsion increases, causing the converging beam to straighten out and finally to diverge.

Halation is produced by reflection of the light rays back and forth between the front and back surfaces of the front glass panel of the cathode ray tube. When the electron beam strikes the phosphor screen, light rays emanate in random directions from the excited spot. These are transmitted at all angles into the glass panel, where multiple reflections occur, some light escaping to the observer in areas adjacent to the excited one. The appearance of this halation, as it is called, is a bright spot surrounded by a region of lower brightness.

Spot size is also limited by the scattering of light created at one point within the phosphor layer by phosphor particles nearby.

The presentation resolution is also limited by other parameters such as sweep speed and cathode ray tube size. The restricted space in aircraft limits the size of cathode ray tubes to a 5- to 7-inch screen diameter. A radar display on a 5-inch tube is usually limited to 3.5–4 inches on a side. The usable display size by itself does not limit resolution, but the spot size and pulsed information that must be shown in a given size of display do. The sweep speed, which is determined by the number of discrete pulses to be displayed in a given time and tube dimension, is a more important parameter in limiting picture resolution, as was illustrated in the two examples.

Indicator Brightness. Target and background brightnesses on the indicator are important parameters in determining target detectability. Variations in the ambient light level affect target detectability since they alter the contrast and therefore the ability to distinguish target from the background. The ambient light levels encountered in airborne radar

compartments vary from very low levels (under 10 foot lamberts) in enclosed areas or under night operation conditions to extremely high levels (over 10,000 foot lamberts) at high altitudes in open compartments.

Brightness is defined as the intensity per unit of projected area of the surface. It is measured in candles per square foot or more commonly, foot lamberts. The spot brightness of a CRT depends upon several factors: (1) beam current and screen voltage, (2) duty cycle of excitation, (3) dwell time of the electron beam on a particular spot, and (4) type of phosphor. The spot brightness of the phosphor is extremely high at the instant of excitation — many thousands of foot lamberts. The electron beam rests on a particular spot of the screen, however, for a very short time compared with the frame time. The average brightness of the spot, therefore, is but a few foot lamberts.

Fig. 12-12, shows the transfer characteristic of the electrical-to-light transducer. The lower usable light limit is dependent on the noise level of

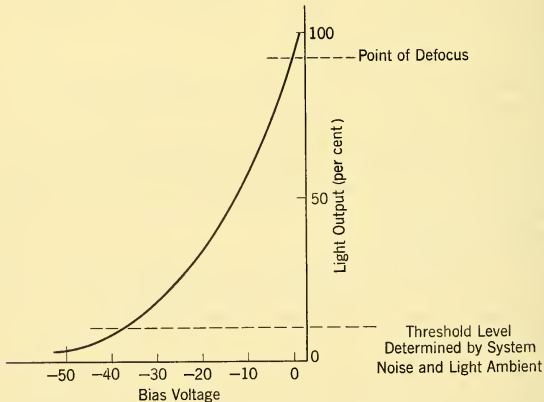


FIG. 12-12 Operating Points of Intensity-Modulated CRT.

the video signal and the ambient light level. In this region near cutoff, the light output varies in most tubes as the $3/2$ power of the input signal. This relationship tends to enhance the contrast between signals that are large relative to the noise level; on the other hand, at low S/N ratios, the signal information tends to be depressed by this characteristic. As the ambient light level is increased, it is necessary to raise the minimum threshold of visibility by driving the tube toward zero bias. The upper limit of grid drive is determined by the application of the indicator. This determination must be made on the basis of whether high resolution or high brightness is

the dominant requirement. It is hard to satisfy both requirements because driving the signal toward zero grid not only increases brightness but also increases the spot size. This is so because of the phosphor characteristics and the mutual repelling characteristic of the electrons in the beam. The electrical-to-light curve also changes as the bias is reduced. More of the beam is intercepted by the apertures of the electron-gun structure and so the electrical-to-light relationship is determined by the particular electron optics of the tube.

The CRT as an instrument has a large light-contrast capability. From complete cutoff to zero bias, the CRT is capable of better than 10 shades of gray (it is assumed that each discernible increase in gray intensity represents a 3-db increase in light output). In a radar system, however, with the defined lower limit specified by the noise level of the video signal and the upper limit by defocus, a reduced contrast ratio can only be obtained in a low light-level surrounding. This contrast represents a 30-130-to-1 range of signal brightness.

Insofar as detectability is concerned, the quantity of importance is *contrast*:

$$\text{Contrast} = \frac{B}{B_N + B_A} \quad (12-4)$$

where B_N = mean brightness of the background noise

B_A = mean ambient brightness level

B = brightness of the signal spot above the background noise level.

Environmental conditions and target brightness levels yielding the same contrast will have equal detectability. Thus when the ambient light level B_A is low compared with the noise light level B_N , little advantage is gained by altering the mean presentation brightness (except as the signal-to-noise ratio may be affected by CRT nonlinearities). On the other hand, when B_A is large compared with B_N , increasing the target brightness will provide increased contrast and hence detectability. Thus, high light output levels are required of the CRT under high ambient light conditions. A limit is quickly reached, however, incident to the upper limit of spot defocusing. The dynamic brightness range is decreased as the CRT brightness is increased.

Signal Display Integration. One of the more important functions of the display device is to serve as an instrument of pulse correlation or integration to improve signal detectability. When pulses are painted on the display screen *alongside* one another, the process of assimilation of this information is frequently called *correlation*. When they are *superimposed* one on top of another, the process is called *integration*. The intensity-

modulated presentation of most fire-control and search radar indicators effects a signal which is a combination of pulse correlation and integration. Target information is presented on the scope not as a dot but as a line made up of a series of dots where each dot represents one pulse return. The scanning motion causes the dots to be displaced one from another. Depending upon the radar parameters — beamwidth, scan speed, PRF, and spot size — overlap of the dots may or may not occur. The ratio of length of the line made up of a series of dots to the picture size is usually proportional to the ratio of the antenna beamwidth to the azimuth scan pattern. Integration is effected by superposition of the signal return along the dash line during the antenna dwell time.

The ideal integrator assumes no information decay as the summing process is made. If the display device were a perfect integrator of a pulse, a signal-to-noise power ratio improvement of as much as \sqrt{N} might be realized (noncoherent integration). The use of a perfect integrator, however, would necessitate the erasure of screen information on the successive frames. If this were not done, noise pulse variations would eventually saturate the integrating device.

Conventional cathode ray tubes are not perfect integrators. The phosphorescent characteristic of all phosphors follows a temperature-independent exponential decay law of $B_0e^{-\alpha t}$ or a temperature-dependent power law decay of B_0t^{-n} . The exponent n varies greatly with temperature, duration of pulse and degree of excitation. Depending on the type, decay phosphor approaches the t^{-n} type as the depth and distance of the activator from the parent atom is increased. Both heating and electrical conduction accompany a decay of the t^{-n} type.

Typical buildup and decay of CRT phosphors are as shown in Fig. 12-13. The average tube bias, the pulse width, the beam current and the PRF

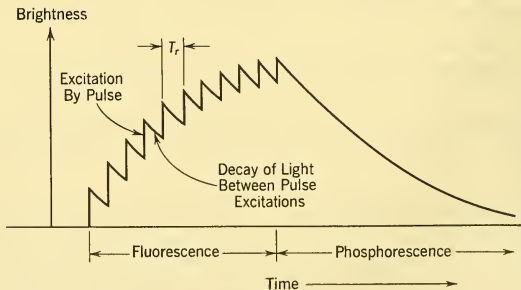


FIG. 12-13 Buildup and Decay of Light for Intermittent Excitation of $e^{-\alpha t}$ Type Phosphor.

determine the number of pulses necessary to build up to a required brightness. The growth of fluorescence is dependent on the rise of brightness with each pulse (representing the radar received pulse), dwell time, the exponential decay of brightness during the beam off time, and the PRF. As is seen from Fig. 12-13 a point is reached where additional pulses do not add further to the brightness (or integration) of the CRT. The number of pulses required to reach this equilibrium condition is dependent on the above-mentioned factors as well as on the type of phosphor. Increase in PRF decreases the number of pulses required to reach equilibrium.

The visual bias level setting on the indicator is dependent on the average noise power which excites the phosphor, on the ambient light level, on the characteristics of a particular phosphor, and on the discernibility of the human operator. These characteristics are all interrelated.

Bandwidth. The bandwidth of an intensity-modulated display is proportional to the writing speed and inversely proportional to the spot size.

$$\text{Bandwidth (in cycles)} = \frac{\text{writing speed (inches/sec)}}{\text{spot size (inches)}}$$

Writing speed is dependent on the system-required time base sweep speed and the tube size. Further if the display pictures the azimuth trace of a radar antenna (such as in a 'B' or 'C' scope) then the bandwidth is also dependent on the speed of antenna travel. Typical numbers would be the display of 20 n.mi. (248 μ sec) along a 3.5-inch vertical trace and the display of antenna motion of 100°/sec for 2 seconds along a 3.5-inch horizontal trace. Spot size, as previously discussed, is dependent on beam current and the characteristics of the phosphor. Spot size can vary from 1 to 2 mils at extremely low light levels to approximately 30 mils at very high light levels. It must be mentioned here, however, that some reward is obtained by impression of higher-frequency signal components than the indicator bandwidth. The high light-flash activation of the phosphor incident to the fast rising leading edge of the impressed video pulse accentuates the presence of a signal.

Tube Size and Form Factor. It is necessary in many radar systems to limit the size of the indicator so as to fit into a confined area of the aircraft. This problem is further complicated by the frequent necessity for installing the indicator in the most desirable location in the aircraft—directly in front of the pilot, for example, in an AI radar. Unfortunately, the form factor of cathode ray tubes is poor. The electron optics needed in the tubes we have been discussing require a large length for a given front face diameter of the tube. The bandwidth and multiple information requirement of most airborne indicators dictate the use of at least a 5-inch-

diameter front face. The length of such a tube using electrostatic deflection is usually greater than 18 inches. Associated with the tube, and usually packaged in the same unit, is the high-voltage supply for the tube. The combination makes for inefficient form factor. Developmental display devices such as electroluminescence and the flat plate tube (these are described in Paragraph 12-9) give hope of reducing the volume.

Special Features. The requirements of certain applications may necessitate certain display device characteristics not normally found in the ordinary CRT — e.g., color, storage, and erase features and three-dimensional representations. As will be described later in this chapter, such features may be provided by special types of display devices. Generally speaking, however, such characteristics are provided at the sacrifice of some of the other important characteristics, such as resolution, contrast, bandwidth, and tube size and complexity. Thus, the designer must scrutinize *all* the important characteristics of a display device to ensure that he has not purchased freedom from one problem at the price of creating several new problems.

Susceptibility to Influence of the Physical Environment. The displays discussed are subject to environmental conditions which have an effect on the presentation. The ambient light level on the open cockpit indicators of fighter aircraft may vary from darkness to several thousand foot-lamberts. In the latter case, the presentation of the CRT tends to be washed and the phosphor exhibits practically no persistence. The tube has to be driven to a maximum beam current; therefore, resolution and gray scale are limited. The tubes are also subject to a wide temperature range as well as vibration and shock and therefore must be ruggedly constructed. For military application the requirements are given in the MIL-E-1 specification.

12-7 IMPORTANT CHARACTERISTICS OF THE HUMAN OPERATOR

When the human operator is an integral part of an airborne radar system, the achievement of optimum performance depends upon how well his characteristics are considered in the design of the system. The characteristics of the human operator that can influence the system can be divided into the following categories: (1) physical, (2) physiological, (3) psychological, and (4) anthropometric. The anthropometric data, or measurements of the parts of the body, will not be considered in this discussion since these data have been adequately presented elsewhere⁶;

⁶C. W. Bessemer, *Missile Engineering Handbook*, pp. 378-380, D. Van Nostrand, Princeton, N. J., 1958.

it is a consideration only in the actual hardware design and not in the development of the system concept and design. The discussion of the physical and physiological characteristics will be limited to a discussion of the eye, since in the more conventional airborne radar systems it is through the eye that the human being receives the information of the system operation. The discussion of psychological characteristics will consider some information handling and extraction problems merely to indicate the nature of the methods that must be used to obtain the requisite data.

Definition of Terms. The following terms are used conventionally in discussing the response of the human visual system.

1. *Intensity* is the ratio of the flux, or light emitted across a surface from a source, to the unit of solid angle formed by the source, when the source is the apex of the angle.

2. *Candle* is the unit of intensity:

$$1 \text{ lumen/steradian} = 1 \text{ candle}$$

3. A *lumen* is the quantity of light radiated into a unit solid angle by a "standard candle."

4. *Brightness* or *luminance* is the intensity per unit of the projected area of the source.

5. A *lambert* is the unit of brightness. It is the brightness of a surface that reflects or radiates 1 lumen per square centimeter. The conventional measure is a *foot lambert*. 1 foot lambert = $1/\pi$ candles/ft² = 1.076 millilamberts.

6. *Visual angle* is the angle subtended by an object at the eye. Fig. 12-14 demonstrates this concept.

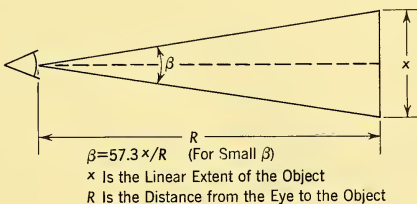


FIG. 12-14 Angle Subtended by an Object at the Eye.

7. *Visual acuity* is the reciprocal of the smallest visual angle that can be resolved in minutes of arc.

Characteristics of the Human Eye. Although one usually speaks of the eye as a camera, this statement is the crudest of approximations and can only mislead the system designer. The eye can be better considered in

two parts, (a) a photosensitive portion called the *retina* and (b) a control system for positioning the eye. The following discussion will be limited to a consideration of the retina and its characteristics that influence system design. The response of the control system will not be considered in this analysis because its characteristics manifest themselves as part of tracking behavior.

The major subdivisions of the retina that can influence design are (1) the periphery and (2) the fovea or center portion. Each portion has important characteristics. The center portion or fovea is color sensitive and has high resolution properties. The periphery is sensitive to very small amounts of light when the eye is dark-adapted.

The basic characteristics of the eye will be considered in terms of their effect upon the performance of a simple task, i.e. the detection and resolution of spots of varying size, color, and brightness under varying conditions of background brightness, exposure time, and eye adaptation time. The basic problem and the sensitive parameters to be considered are shown in Fig. 12-15.

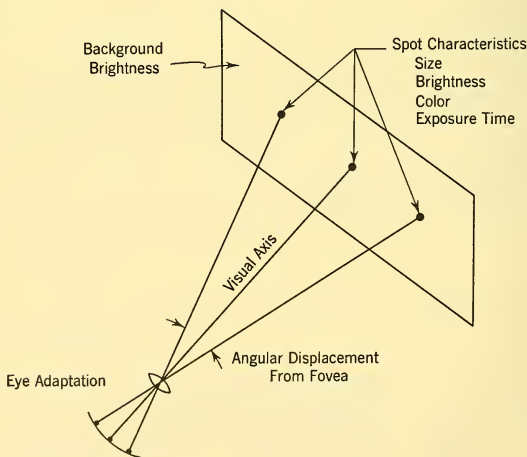


FIG. 12-15 Sensitive Parameters of the Spot Detection and Resolution Problem.

Visual acuity (i.e. the ability to resolve small visual angles) depends upon the angular displacement from the fovea as shown in Fig. 12-16. At 10° from the fovea on the nasal side, the visual acuity is 20 per cent of the value that exists when the object falls on the fovea. (That is to say, the minimum

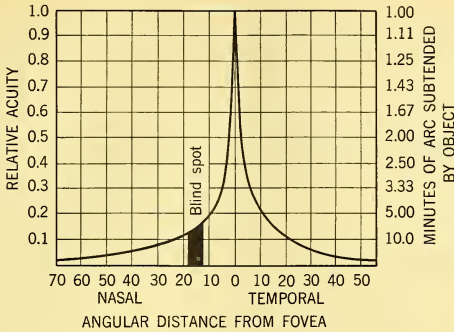


FIG. 12-16 Variation of Acuity with Angular Displacement from the Fovea.

resolvable element is five times as large.) If light falls upon the “blind spot” (entrance of the optic nerve) no visual response will be obtained.

Fig. 12-17 shows the range of sensitivity (the absolute brightness required to elicit a response) of the retina as well as the extreme sensitivity of the periphery. In this figure, the ordinate indicates the sensitivity of the eye after dark-adaptation has taken place. Each curve presents these data for a different subject.

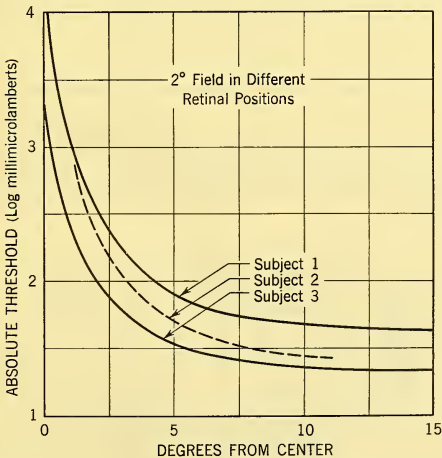


FIG. 12-17 Sensitivity of the Eye as a Function of Angular Displacement from the Fovea.

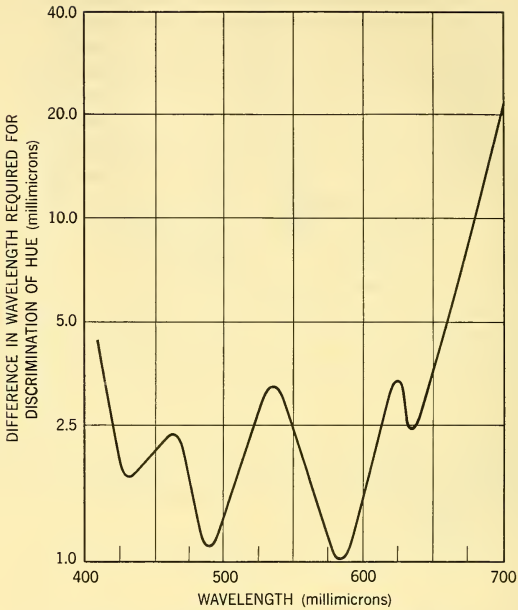


FIG. 12-18 Color Differentiation of the Fovea.

Figure 12-18 shows the discrimination of wavelength that the fovea can perform. This figure indicates that when two colors are placed side by side, the eye has excellent chromatic resolution. For example, if yellow (600 mμ)

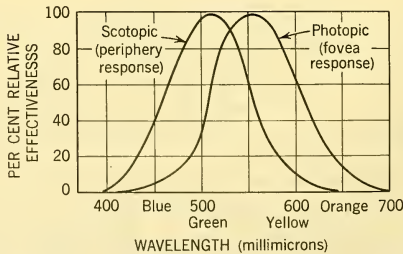


FIG. 12-19 Visibility Curve for the Human Eye.

FIG. 12-20 Table of Approximate Brightness.

DAY VISION (PHOTOPIC)	100,000 mL	— Upper limit of visual tolerance
	10,000 mL	— New snow on clear day or upper surface of clouds at noon
		— Average sky on hazy day at noon
	1,000 mL	— Average sky on clear day
		— Average earth on clear day
	100 mL	— Average sky on cloudy day
		— Average earth on cloudy day
	10 mL	— White page in good reading light
1 mL	— White paper 1 ft from a candle	
0.1 mL	— Average chart brightness at lowest readable brightness level	
0.01 mL	— Snow in full moon	
NIGHT VISION (SCOTOPIC)	0.001 mL	— Average earth brightness in full moon
	0.0001 mL	— Snow in starlight
	0.00001 mL	— Green grass in starlight
	0.000001 mL	— Absolute threshold of seeing

is placed alongside a color whose wavelength is $601.75 \text{ m}\mu$, a difference in hue will be discerned.

The eye responds differentially to radiant energy; i.e., the energy required to elicit an equal brightness response varies with wavelength. To respond differentially does not necessarily imply color discrimination, but merely that the response to different spot colors requires more or less energy. Fig. 12-19 is called the visibility curve and shows the relative

response of the fovea and the periphery to energy. The periphery response is indicated by the curve labeled SCOTOPIC, while the foveal response curve is labeled PHOTOPIC. The ordinate of Fig. 12-19 is a measure of the energy required to obtain an equal brightness response as a function of wavelength. For example, referring to the photopic curve, a greenish-yellow hue (550 $m\mu$) will require approximately 20 per cent of the energy required by an orange hue (650 $m\mu$) to elicit an equal brightness response.

Although the terms periphery (scotopic) and fovea (photopic) have been used to designate the parts and response of the eye, there has been no discussion as yet of the quantitative levels of energy at which these two parts of the eye become effective. Fig. 12-20 gives a crude practical estimate of the brightness at which the fovea and the periphery become effective. The periphery begins to be effective at about 0.1 millilambert, depending upon the length of dark adaptation.

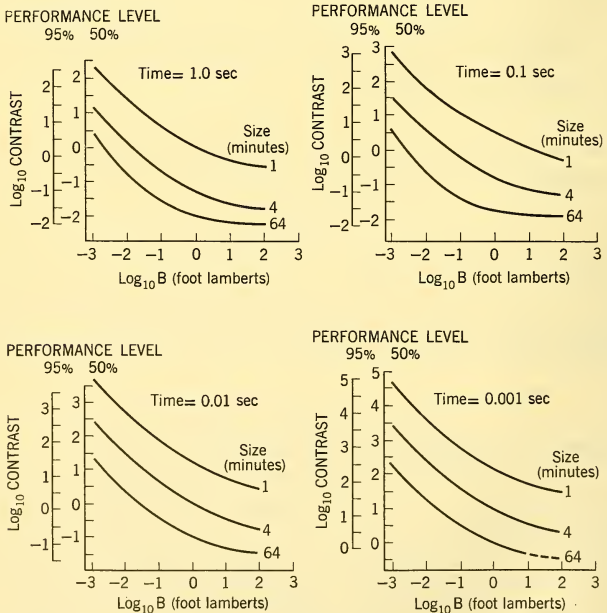


FIG. 12-21 Contrast Value for a Given Detection Probability as a Function of Performance Level, Time of Exposure, Background Level, and Spot Size.

In practical cases the detection of a spot is strongly dependent upon the brightness of the spot relative to the background. Intuitively, contrast is usually defined as the difference in brightness or intensity between the spot and the background. In vision experiments, *contrast* is defined as the ratio of the difference in energy between the spot and the background to the energy of the background. Conventionally the measure of energy that is used in determining contrast is brightness:

$$\text{Contrast} = \frac{B_s - B_B}{B_B} = \frac{\Delta B}{B} \tag{12-5}$$

where B_s = spot brightness

B_B = background brightness.

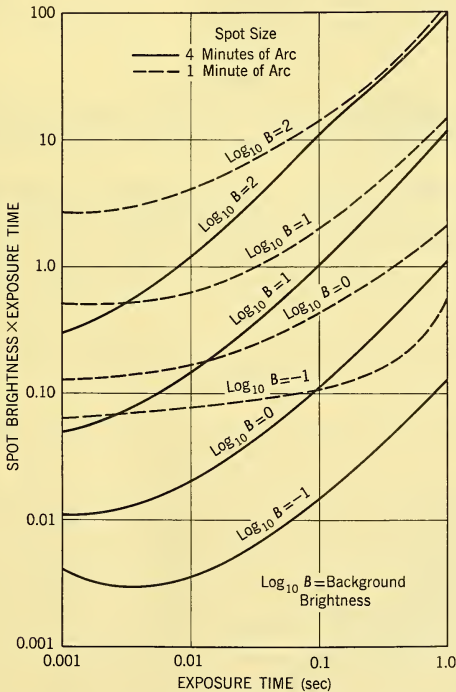


FIG. 12-22 Brightness of Spot × Exposure Time as a Function of Exposure Time, Background Brightness, and Spot Size for a Detection Probability of 50%.

The plotted data on this phenomenon are usually for threshold values, i.e., a given probability of detection. However, the required contrast will change as a function of (a) probability of detection, (b) the background brightness, (c) size of the spot of light, and (d) the time of exposure. The data are shown in Fig. 12-21.

Using the data of Fig. 12-21, Fig. 12-22 has been derived. In this figure, the relationship between the required spot brightness of an assumed 50 per cent threshold value is shown as a function of size, background brightness, and the time of exposure to the spot. At the higher background brightnesses, the effect of the spot size becomes negligible for longer exposures. As the time decreases, the importance of size to achieve threshold becomes more critical.

Exposure to high levels of brightness immediately prior to spot detection significantly affects the time required for spot detection. The pertinent data are shown in Figs. 12-23a, b, and c. These data are based upon

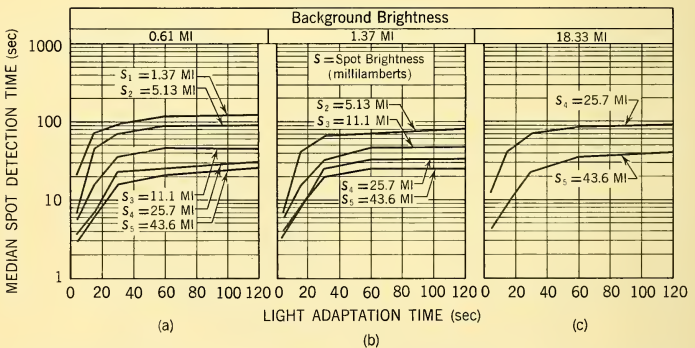


FIG. 12-23 Time for Detection as a Function of Target Brightness and Duration of Light Adaptation.

exposing the subjects to a brightness level of 3100 millilamberts for 5, 15, 30, 60, and 120 seconds prior to attempting spot detection. The three sets of curves differ only in the amount of background brightness around the spot. Each figure shows the detection time as a function of exposure to the high brightness level. These figures suggest that if the exposure lasts as long as 60 seconds, an additional 60 seconds has little effect.

Fig. 12-24 is a plot of the data for an exposure time of 120 seconds and shows the relationships that exist among the variables as a function of contrast. These figures suggest that if the operator is exposed to high

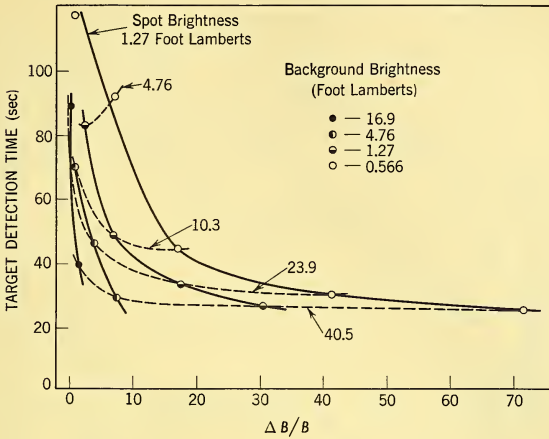


FIG. 12-24 Detection Time as a Function of Contrast, Target Brightness, and Background Brightness (Light Adaptation Duration 120 sec at 3100 ml).

brightnesses as a result of his other duties, time will be lost in target detection even if the target is present on the scope. It is obvious from these figures that the system designer must consider these additional duties in the development of his system.

A study typical of those often required in the development of the display system concept has been performed by Gardner and Carl.⁷ In this study, the experimenters were interested in determining whether there exists an optimum pedestal amplitude voltage as a function of the ambient brightness and the receiver noise. (The *pedestal amplitude voltage* is used to set the trace brightness on intensity-modulated radar scopes. This brightness is independent of the brightness that is incident to the receiver noise on the signal.) Fig. 12-25 shows the relationship between pedestal amplitude voltage, receiver noise, and trace brightness. Fig. 12-26 shows detectability as a function of pedestal amplitude voltage, receiver noise, and ambient light levels. Using the data shown in Figs. 12-26 and 12-21, it is possible to determine the contrast that was obtained at the detectability threshold for each pedestal amplitude voltage and receiver noise level. In such instances the trace brightness can be considered to be equivalent to the background brightness shown in Fig. 12-21. With this information and a measure of the ambient brightness, it is possible to state the requirements

⁷R. E. Gardner and J. M. Carl, *The Effects of Ambient Illumination, CRT Bias, and Noise upon Target Detectability with a B-Display*, NRL Report 5264, Jan. 28, 1959.

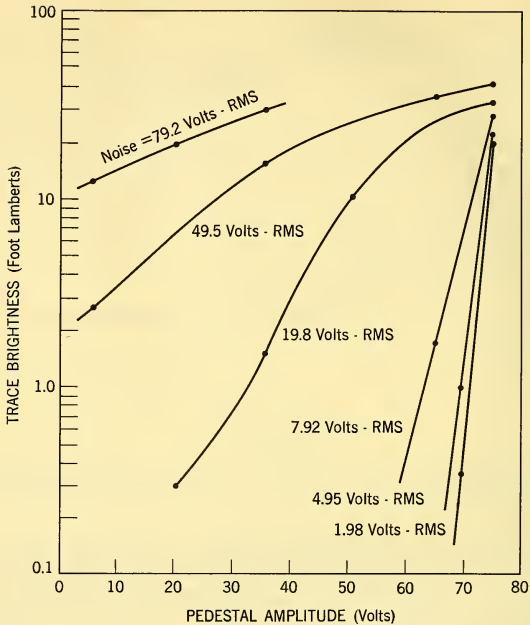


FIG. 12-25 Average Trace Brightness as a Function of Pedestal Amplitude and Receiver Noise.

for the additional brightness that must be generated by the electron beam so that the observer can detect the spot at a given level of probability.

Effects of Operator Environments. The data discussed in the foregoing paragraphs are for normal conditions when the observer is not under environmental stress such as acceleration and low atmospheric pressures. These characteristics will change as acceleration and low atmospheric pressure affect the operator. Fig. 12-27 shows the loss of visual acuity as a function of acceleration and body position. A comparison of the minimum spot size which can be resolved at 1 g (0.91 minutes of arc) and that which can be resolved at 3 g (1.11 minutes of arc) indicates that a 22 per cent increase in spot size is required for equivalent detectability at high acceleration.

The effect of low atmospheric pressure manifests itself as a lack of oxygen. Fig. 12-28 shows the required oxygen at different altitudes to maintain oxygen pressure in the lungs at sea-level conditions.

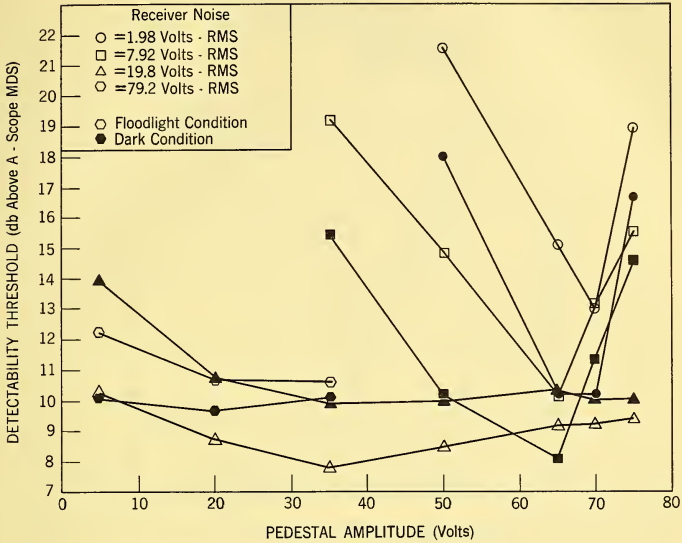


FIG. 12-26 Detectability vs. Pedestal Amplitude.

A lack of oxygen will result in a decrease in the sensitivity of the eye. Fig. 12-29 shows the change in absolute threshold as a function of dark-adaptation time and oxygen tension. A higher threshold is an indication of decreased sensitivity. The effect of lack of oxygen becomes more pronounced as the eye becomes more sensitive. This fact is indicated by the increased separation of the curves as length of time in the dark is increased.

Man-Machine Interrelationships. The foregoing analysis has been primarily concerned with the physical, physiological, and anatomical aspects of the visual system. However, when the human operator is part of the system, his psychological characteristics can influence system design. Unfortunately the necessary data to develop a cohesive system of design specifications relating to human psychological characteristics are not available at this time. Therefore it is necessary to study each problem and gather data that are specific to the particular problem, thus adding to the general body of knowledge which will eventually lead to a systematic procedure of specification. The following paragraphs will present indications of the kind of effort that is required to obtain data that can be used in development of an airborne radar system.

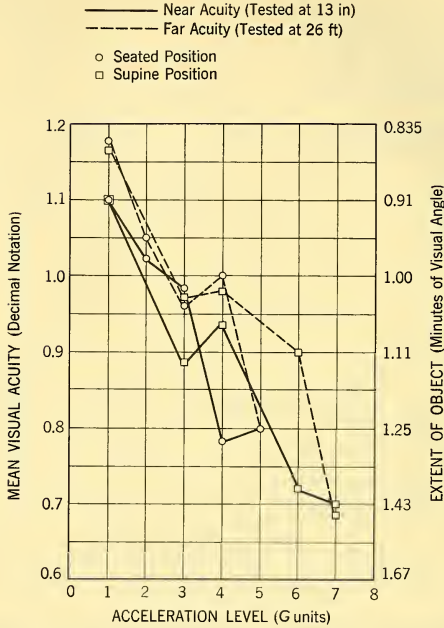


FIG. 12-27 Mean of Binocular Acuity as a Function of Acceleration Level and Body Position.

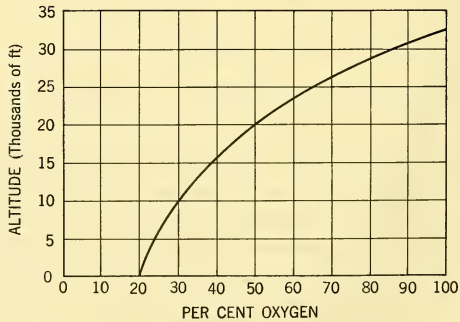


FIG. 12-28 Per Cent Oxygen as a Function of Altitude Required to Maintain Normal Oxygen Tension in the Lungs.

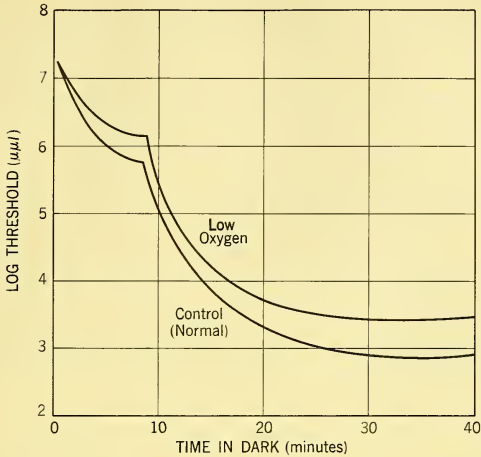


FIG. 12-29 Effect of Low Oxygen Tension on Sensitivity of the Visual System.

With the advent of high-speed systems, the importance of time of response or recognition becomes a critical factor in the design of the system. When the human operator is part of the system, his response characteristics can influence design. Fig. 12-30 shows the result of an experiment in which

- 0.1 Effective Foot Candles
0.5ft Diameter
8.0ft Viewing Distance
- 2.0 Effective Foot Candles
3.0ft Diameter
12.0ft Viewing Distance

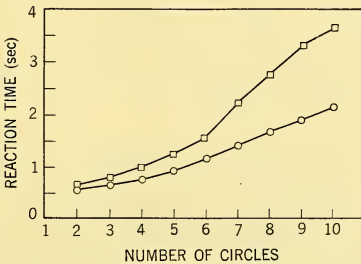


FIG. 12-30 Reaction Time as a Function of Display Variables.

the subjects were required to report the number of concentric circles that were on a large display that subtended an angle of 14° at a viewing distance of 12 ft. An experiment such as this is typical of the type of basic investigations made for tactical display systems to determine the operator's ability to report bearing and range information as a function of display and system parameters. This kind of display problem would be involved in CIC and AEW installations. The data reported were obtained at two levels of brightness and two viewing distances. It appears that there is an interaction between the number of circles, the size of the display, and the illumination. It is also obvious that if the number of circles could be kept below five, the effect of size and brightness can be minimized and the reaction time will be a negligible factor as a human engineering problem.

An important characteristic of the human operator is his ability to receive and act upon information. This characteristic is predicated on the fact that all the characteristics of the signal are well above the threshold of the observer and he has merely to extract the information from the symbol and act upon it. In this class of behavior, the operator is acting as a digital device in that he absorbs information and acts upon it in serial fashion. There is a separate response to each signal even though the signals may be in sequence. This last statement is to differentiate this form of activity from that which exists when the operator is acting as an error detector in a tracking system. This case will be considered in a subsequent paragraph.

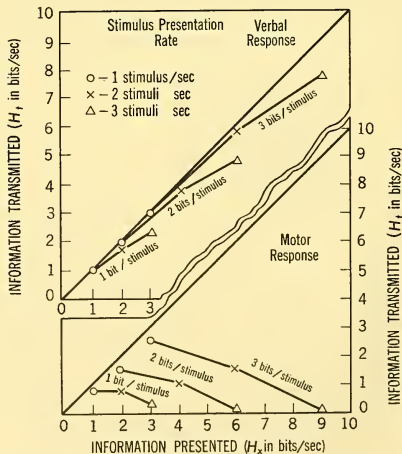


FIG. 12-31 Information Transmitted as a Function of Speed of Presentation and Mode of Response.

Fig. 12-31 shows the results of a study on the amount of information absorbed as a function of the rate of presentation and the mode by which the subjects transmitted the information to a recorder.⁸ In this study the subject was presented with a series of numeral stimuli. These numerals were presented at different rates on a screen in front of the subject. The subject either pressed keys to indicate the correct number or called out the name of the numeral. The bits per stimulus is a measure of the number of possible numerals that can be displayed in one block of trials. The data presented indicate that the complexity of the human data processing and read-out system can be adequately studied to the point that system specifications can be based on objective and reliable data.

This point of view is further emphasized in the results of a study in which the human operator was asked to perform as an error sensor in a tracking system.⁹ Fig. 12-32 is a block diagram of this system. In this study it was necessary to determine the characteristics of the display that would lead to optimum performance. An examination of the engineering and psycho-

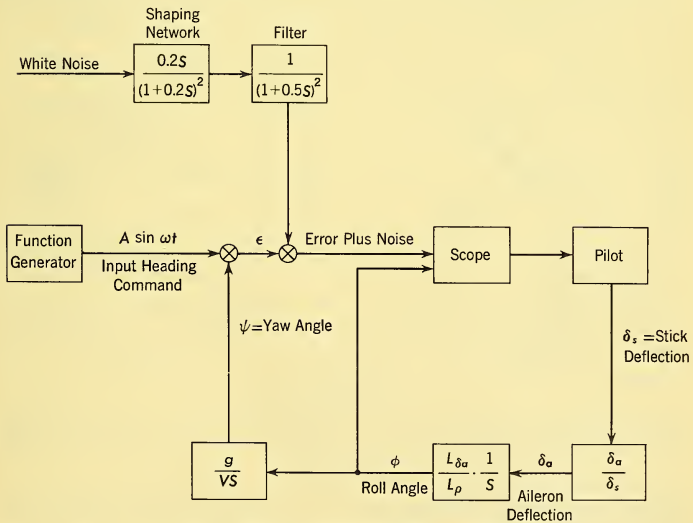


FIG. 12-32 Block Diagram of Experimental Setup.

⁸E. A. Alluisi, P. F. Muller, Jr., and P. M. Fitts, *Rate of Handling Information and the Rate of Information Presentation*, WADC Technical Note 55-745, December 1955.

⁹A. Kahn and M. Mazina, *Human Tracking Performance I. Tracking Ability as a Function of System Noise, Scope Sensitivity and Forcing Function*, Westinghouse Electric Corp., System Research and Development Rpt. 57-103-6-116-A-1, March 1957.

logical factors indicated that there might be an interaction between the radar noise level and the scale factor used on the scope. An experiment was therefore designed in which the amplitude of the radar noise, the scope scale factor and the frequency of data inputs to the system were the variables. The results of this study are summarized in Fig. 12-33. The

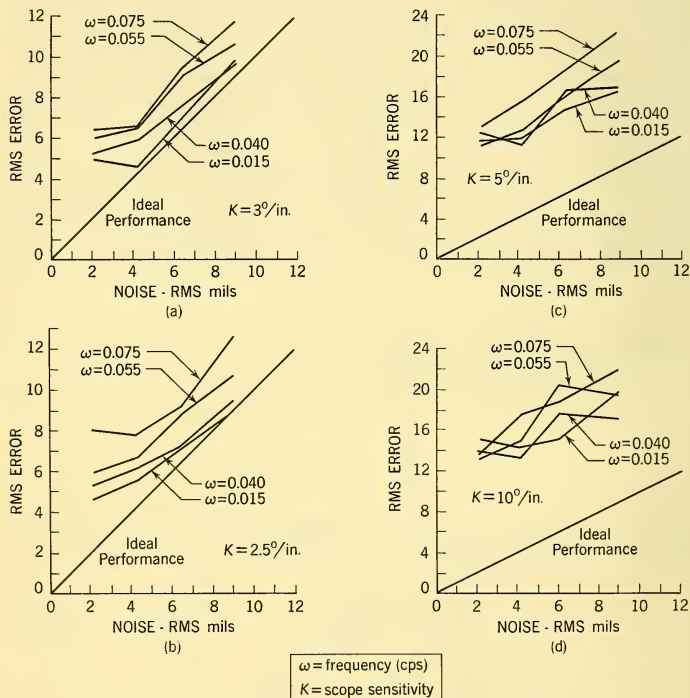


FIG. 12-33 Performance vs. Noise for Variable Forcing Frequency and Scope Sensitivity.

diagonal on the curves labeled IDEAL PERFORMANCE is the performance that would obtain if the human operator did not contribute to the error of the system. Analysis of the data indicated that there is no significant difference in system performance when the $0.3^\circ/\text{inch}$ or the $2.5^\circ/\text{inch}$ scope sensitivity is used. These results suggest that there is no need to have a scale factor more sensitive than $2.5^\circ/\text{inch}$, and that the scale factor is independent of the radar noise level.

Summary. This paragraph has discussed some of the characteristics of the visual system of the human operator and has indicated the kind of analysis, experimental and logical, that is required if the human operator is to be properly integrated into the system design.

12-8 DEVELOPMENT OF REQUIREMENTS FOR A DISPLAY SYSTEM

In this paragraph, the use of the data and considerations relating to display system design will be illustrated by the development of a possible solution to display system problems for the airborne intercept mission originally defined and discussed in Paragraphs 2-22 through 2-30. To recapitulate, the interceptor is vectored towards the target on a collision course, until the target is detected. Upon detection, the pilot locks his radar on the target, converts from an approach to an attack course, guides the aircraft on this attack course, and launches his weapon at the appropriate time. The weapon is assumed to be an air-to-air missile with semiactive homing guidance. During the missile flight, the radar must remain locked-on to provide illumination for homing although the aircraft itself may make moderate turns off the attack course after launching.

Search and Vectoring Display. In Paragraph 2-26 and in Fig. 2-29, the quantities which must be displayed to the pilot during the vectoring phase are specified. These quantities are (1) the target altitude, (2) time to collision, (3) the relative range to the target, and (4) the interceptor heading commands. In addition, of course, the video information itself must be displayed to provide the detection function along with aircraft attitude information and the antenna look angles. It is very difficult to provide all this information in one display, particularly on a single scope face, without generating a great deal of confusion. One possibility is to present some of the auxiliary data which change slowly on either counters or dials rather than directly on the scope face. This can be done with target altitude, the time to collision, and the range to the target. Accordingly, we shall require these three quantities to be displayed numerically on separate counters. The information which must now be combined on one scope are (1) the video signal for target detection, (2) a vectoring heading command, (3) aircraft roll and pitch and (4) the antenna look angles.

The video presentation which has been found most satisfactory for this type of application is an azimuth-range B scope. The AI requirements summarized in Sec. 2-30 specify that the search pattern be 60° in azimuth and 17° in elevation with a search range of 20 n.mi. The total look angle can be $\pm 60^\circ$ in both azimuth and elevation.

A 7-inch scope will be preferred for an AI application of this type where space is very limited. This will provide a 5-inch square for the B display.

We suppose that the AI radar system has parameters similar to those in the example of Sec. 3-2 with a beamwidth of about 4° and a $1\text{-}\mu\text{sec}$ pulse width. If the 60° azimuth scan pattern is presented over the full 5 inches, only $60^\circ/4^\circ = 15$ horizontal resolution elements are required. The vertical resolution, however, must be quite good, since there will be $20 \times 12 = 240$ 1-microsecond range elements within the 20-n.mi. maximum range. The number of vertical elements per inch will be 48, allowing a 21-mil spot size. In order to be on the safe side, we will require a spot size of 10 mils, which is a practical goal.

In order to present information on the azimuth and elevation look angles of the antenna, separate strobes will be required. The azimuth strobe is specified to move horizontally below the B display with the elevation strobe moving vertically at the side of the display. Range strobes can also be conveniently used to bracket the expected target ranges and reduce the uncertainty volume which the pilot must scan during detection.

It is desirable that the display of heading command or error and aircraft attitude not overlay the B pattern, as the flight information can very easily mask a faint target. Unfortunately, there would not be room on a 7-inch scope for both the B display and the flight information separately, and the search presentation will have to be overlaid. In order to minimize any deterioration in the detection function, we shall specify the use of a two-color tube to present the search and flight data in two different colors.

There are several possibilities for the presentation of heading information. One would be to display a plan view of the locations and courses of both the interceptor and target in order to give the pilot a good idea of the general tactical situation. This type of presentation will allow considerable flexibility in the pilot's selection of the most appropriate type of approach. A typical display of this type was illustrated in Fig. 12-3. Because of the small scale required to represent the large distances involved, however, the accuracy with which a desired course will be flown might be marginal. An alternate possibility, where the approach doctrine is prescribed, is simply to present the heading error. Greater vectoring accuracy can be obtained in this manner and possibly a smoother transition to the track mode at the expense of restriction to a single mode of approach. The vectoring doctrine originally adopted in Paragraph 2-11, which applies to this AI example, actually contemplates a single type of vectoring, i.e., a collision approach to minimize bomber penetration.

The type of display of heading commands which has been selected for the present example is simply to present the heading error or deviation from the computed collision course rather than the complete tactical situation. This will be accomplished by horizontal movement of an error circle along a horizontal reference line. The size of the error circle will represent the allowable piloting error during vectoring, which was specified

to be less than 5° rms. The pilot will have to maintain the center of the scope within the error circle for a successful approach.

Aircraft attitude will be indicated by the motion of an inverted tee representing the aircraft. Roll and pitch of the aircraft will correspond to rotation and vertical movement of the tee relative to a horizontal reference line.

The primary features of the display unit which we have just specified are indicated in Fig. 12-34. In operation, the pilot attempts to fly the tee into the error circle, although it is actually the circle which moves horizontally.

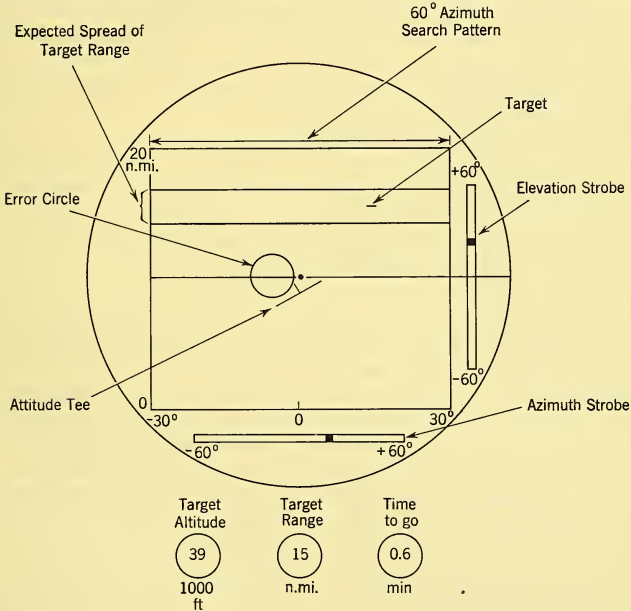


FIG. 12-34 AI Display Specified for Vectoring and Search Phase.

The data in Paragraph 12-7 are applicable to several aspects of the search display design. In particular, the data in Fig. 12-26 indicate some of the effects on detection of high ambient illumination as a function of receiver gain and pedestal voltage in a typical case. It is apparent in that figure that optimum performance will be provided when specific related values of receiver gain and pedestal voltage are used. Means therefore should be provided in the search display for ensuring that optimum settings for these controls are maintained. One possibility is simply not to provide any

brightness or gain control for the pilot. The data in Fig. 12-26 indicate definite optimum settings for the receiver gain and pedestal amplitude. This, of course, applies to a particular display. Similar data should be determined experimentally for the scope selected for the display being discussed.

Attack Display. After detection and lock-on, conversion to the attack mode is effected and the aircraft is precisely flown along a straight-line attack course to the missile launch point. The heading error relative to the correct course must be displayed in such a manner that the tracking error caused by the pilot is maintained at less than 1° rms as specified in Paragraph 2-27. The time to launch must also be displayed as well as the interceptor attitude. After launching, a breakaway symbol must be presented when there is a danger of collision.

The attack display will be specified to be very similar to the heading error and attitude symbols in the search display. There will be a fixed horizontal reference line across the center of the scope. Attitude will be indicated by the motion of an inverted tee relative to this line. The motion of this tee in response to roll and pitch of the aircraft will be identical with that in the search display previously described.

The attack course error will be indicated by a circle which moves relative to the center of the scope. As in the search display, the size of the circle will be proportional to the allowable error. The size of this circle will of course be somewhat smaller than that for the vectoring phase. Time-to-go until missile hit will be indicated by a thermometer type of symbol on one side of the scope. Full-scale reading of the time-to-go thermometer will be 40 seconds. This type of attack display is illustrated in Fig. 12-35.

The data in Fig. 12-33 can be used to determine the best choice of scope sensitivity. These results establish that there is very little improvement in pilot tracking performance with scope sensitivities greater than $2.5^\circ/\text{inch}$. In general, a small scope sensitivity is desirable to give as large as possible a linear dynamic range for the error. With $2.5^\circ/\text{inch}$ on the 7-inch scope, the dynamic range will be $\pm 2.5^\circ \times 3.5 = \pm 8.8^\circ$. This should be sufficient since a relatively smooth transition will be made from the vectoring display where the error sensitivity is considerably less. The diameter of the error circle with a radius corresponding to the 1° allowable tracking error will be 0.8 inch. This should be satisfactory. From Fig. 12-33b, the tracking error with a scope sensitivity of $2.5^\circ/\text{inch}$ is only slightly larger than the actual noise on the scope. Thus, if the noise on the scope can be kept less than 1° rms per channel, the aircraft tracking accuracy will be within the specified bounds. From these considerations, it is indicated that the attack display illustrated in Fig. 12-35 will provide satisfactory performance, and it is adopted as a satisfactory solution to the system requirement.

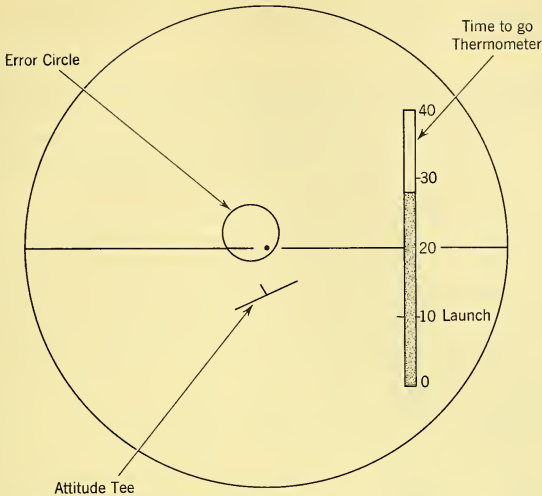


FIG. 12-35 AI Display Specified for Attack Phase.

Breakaway Display. After launching, precise interceptor tracking is no longer required, but the radar must remain locked-on until impact to provide illumination and so only moderate turns are possible. In order to preclude the possibility of collision, the interceptor should take violent evasive maneuvers after impact. Danger of collision will be indicated by the display of a breakaway symbol when there is danger of collision, accompanied by a flashing light at the side of the scope. The breakaway symbol could have a number of forms. We shall require it to be a large arrow pointing in the direction in which the interceptor should break away.

Alternative Solutions. It should be emphasized that other solutions to this display problem could be found. In an actual design study, several solutions might be derived using the general lines of reasoning presented in this example. These then would be evaluated to determine their relative efficiencies as solutions to the display problems. Such evaluations would involve use of data already presented in this chapter and carefully designed experiments which would disclose the specific problems of the application.

12-9 SPECIAL DISPLAY DEVICES

New types of phosphors, electron optics, and display tubes are being developed which are improvements over the conventional cathode ray tube

described above, in that they have higher brightnesses, controlled persistence, or good form factor.

Storage Tubes. The ideal display device is capable of (1) accepting a picture with many bits of information, (2) displaying this for a full frame time as a uniformly illuminated display, and (3) removing the past information for rewrite in a new frame. The newest display systems approach these goals by using storage tube principles. A *storage tube* is a device capable of producing a charge pattern on an insulator surface corresponding to input video signals. The charge pattern is developed by secondary emission techniques. Reading of this information by either visual observation or electrical scanning of the storage information is incorporated in the storage tube. The classification of tubes into electrical-to-electrical or direct viewing is determined by the method of reading. The direct-view storage tube, which is finding application in the fixed fire-control system where high ambient light level is a problem, has the capability of displaying a high light level, uniformly illuminated radar presentation. Whereas the light output of a target integrated over the frame time in a conventional CRT is a few foot lamberts, it can be made higher than 1000 foot lamberts in the direct-view storage tube. A typical direct view storage tube usually has two guns. One gun (for writing) produces a fine beam which deposits a pattern of electric charges on an insulating mesh surface located just under the phosphor screen. The other gun produces a highly divergent beam which uniformly floods the whole tube screen. The purpose of developing a wide collimated beam is to illuminate the phosphor continuously so as to obtain a high level of brightness and a uniformly illuminated picture. The electrons of this beam are, however, able to penetrate the insulating mesh surface only in accordance with the charge pattern produced by the write

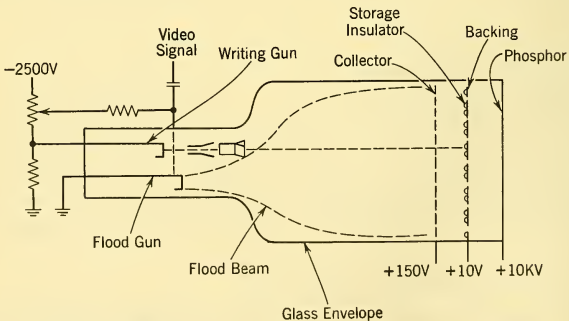


FIG. 12-36 Direct-View Storage Tube.

beam. A replica of this electric charge pattern is thereby produced as a light pattern on the phosphor screen. Fig. 12-36 shows the arrangement.

The storage tube has the capability of retaining information for periods comparable to the radar frame time because of the charge control developed by secondary emission action on the surface of the insulator material. The problem of writing or placing a charge on the insulator surface has been approached in different ways. One solution, which is used in most direct-view tubes, charges a unit element of the insulator off the floating potential equilibrium curve and then shifts the potential toward the equilibrium point by a controlled amount of secondary emission action dependent on the input signal. Equilibrium potential of the insulator is never reached in this method of writing and therefore it is called *nonequilibrium writing*.

The direct-view storage tube makes use of transmission modulation or grid control reading. This is achieved by the established electrostatic fields on the insulator surface. These electrostatic fields modulate a low-velocity reading beam (flood beam) as it passes through the field. The insulator target for this type of reading is usually made of fine wire mesh (250 mesh to the inch) over which a thin layer of insulator material is placed. Erasing of the charge information in the transmission modulation type storage tube is mechanized by discharging the entire surface of the insulator by purposely causing slow-velocity electrons to land on the insulator. This is accomplished by pulsing the backing electrode of the insulator material with a positive pulse such that the entire insulator surface, by capacitive effect, is made positive with respect to the flood gun. With flood electron landing, the potential of the cathode and the insulator is made the same. When the positive pulse is removed, the negative-going waveform develops a negative potential on the insulator with respect to the flood cathode to prevent primary electrons from the flood gun from landing.

The technical disadvantages of the available direct-view storage tubes are limitations in resolution, shades of gray, and duration of viewable storage without degradation. Resolutions of 30 photographic lines to the inch at a brightness level of 1000 foot lamberts are obtainable today. The resolution limitations are caused by the coplanar grid effect and by the electron lens characteristic of the insulator mesh grid holes. The tangential direction imparted to the low-velocity flood beam, the distance to the phosphor screen and the potential of the phosphor screen become important resolution considerations. Shades of gray, where each shade is defined in 3-db brightness steps, are limited by the background brightness of the tube. The direct-view storage tube screen, unlike conventional cathode ray tubes, has a uniformly illuminated brightness area. This illuminated area (which is the flood-gun illuminated area) is the integrated brightness caused by the flood-beam excitation of the phosphor when the erase pulses are present. During the erase pulse, flood electrons are not biased and therefore can

traverse the distance to the phosphor screen. Fortunately, the duty cycle of erase to store can be kept small. The number of gray shades has been measured, for a particular background brightness, as four or five. This implies a maximum brightness contrast ratio of 32 to 1.

For definition purposes the *duration of storage* and the *duration of viewable storage* must be considered separately. The duration of storage is the retention time of information on the insulator material if writing and reading are separated. With presently used materials this storage duration is many hours, if not days. The duration of viewable storage is the retention of the insulator information for continuous display. This implies that the flood beam is continuously illuminating the phosphor. In this sense the storage time in present tubes is not much longer than 10 seconds. The reason for this short retention time is that gas molecules are present in the imperfect vacuum. Collision between the flood electrons and these gas molecules cause positive ion landing on the insulator surface so that stored information is obliterated. With this type of tube, ion bombardment is especially severe, since the insulator is the most negative element, being sandwiched between a positive collector and a positive phosphor, and is therefore most susceptible to positive ion bombardment. Viewable storage time can be increased at the expense of brightness by pulsing the flood beam to lower its duty rate and thereby cause fewer collisions with gas molecules.

Flat Tubes. One of the biggest limitations with all the tubes we have discussed is the form factor of the tube. To obtain a usable deflection sensitivity the tubes must be long. Electrostatic deflecting tubes, in almost all cases, are longer than magnetic deflecting tubes. Tubes using the latter deflecting method have been built with deflection angles up to 110 deg. Major development effort is being conducted to further reduce the length of the tubes. Aiken¹⁰ has placed the electron gun and accelerating and deflection elements in an arrangement other than the in-line method used in the tubes discussed above. As Fig. 12-37 shows, the electron beam is made to move along an edge of the tube. A set of deflection plates arranged along this edge deflects the beam if the voltage is lowered on any of the plates. In the absence of an electrical field the beam would continue in its initial path. By placing the proper sequence of the voltages on the plates, the beam can be made to move up to any point on the tube. A second series of deflection plates which extends across the tube deflects the beam to the phosphor front face. By choosing the proper time sequence on the two sets of deflection plates, the electron beam can be made to strike any point on the phosphor front face.

Focusing is simplified in this tube. The electron beam which initially is circular becomes elliptical as the beam is deflected up from its normal travel

¹⁰W. R. Aiken, "A Thin Cathode Ray Tube," *Proc. IRE* 45, No. 12 (December 1957).

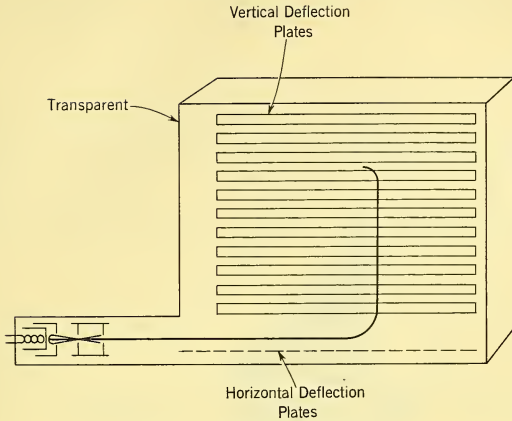


FIG. 12-37 The Aiken Tube.

to the right. This is true since the electrons closer to the deflection plates have a sharper radius of curvature than the electrons farther removed from the plate. The coarsely focused beam can now be considered as a ribbon beam when it travels upward. This process is repeated in the other plane when acted on by the second set of deflection plates as shown in Fig. 12-38. This double focused beam is brought to bear on the phosphor.

One of the limitations of the tube is the need for multiple plates to deflect the beam. Present experimental tubes use ten deflection plates for horizontal deflection and ten more plates for vertical deflection; it is necessary to provide separate and sequential sweep voltages to each plate. Various deflection mechanization schemes have been used, including the use of a separate amplifier for each plate, delay lines, radial switch tubes, and other special devices.

Although this tube still has sealing and stress problems and deflection mechanization problems, the excellent form factor and special features such as transparency, multiple color display, and high resolution make it very attractive for radar system display use.

Transparent Phosphor Films. Research work is being done on the development of thin transparent phosphors to improve the CRT resolution by negating the optical scattering properties of conventional powders used to make the phosphor screen and to overcome the reflectivity problem at high ambient light levels. The transparency of the film limits reflectivity and allows viewing the backside of the screen. The thinness and homoge-

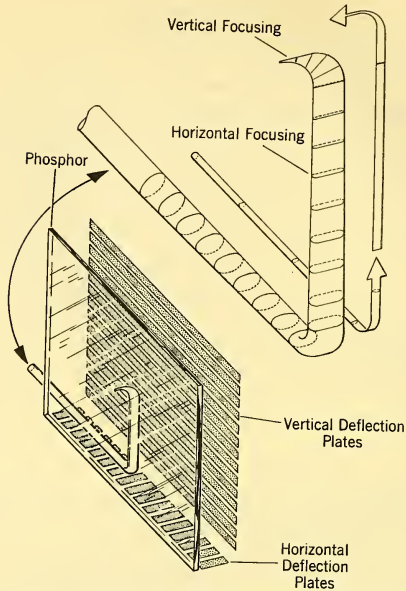


FIG. 12-38 Thin Cathode Ray Tube Focusing Action.

neity of the material precludes much optical scattering. The transparent phosphor film is not as efficient as conventional powders but the losses of the transparent phosphor film at high ambient light levels are not as great as for conventional phosphors.

Various unsuccessful attempts have been made to prepare this phosphor by evaporation. The difficulty in most cases was that either (1) the cathode-luminescent efficiency was too low or (2) the phosphor deteriorated under electron bombardment. However, it has been shown¹¹ that the nonluminescent and discolored phosphor film usually obtained when evaporating and condensing most phosphors in vacuum can be made luminescent and transparent by heating the deposited phosphor to a high temperature in a suitable atmosphere. At present, it is necessary because of the temperatures involved in this process to deposit the film on a substrate such as quartz or vicor instead of the glass front surface used in cathode ray tubes. The film is supported by a metallic clamp near the front face in the tubes presently available.

¹¹Feldman and O'Hara, *J. Opt. Soc. Am.* **47**, 300 (1957).

The advantages and disadvantages of available transparent thin phosphor films are listed here.

DISADVANTAGES

1. No persistent phosphors are presently available.
2. Efficiency is less than conventional powder because of the limited light-generating ability of the phosphor and the internal light-trapping properties of the film.
3. The film cannot be deposited on the front face of the cathode ray tube.

ADVANTAGES

1. Resolution is improved.
2. Contrast is improved by reduction of halation.
3. Visibility increases at high ambient light levels.
4. There is more resistance to deterioration under electron bombardment because of better heat transfer of the substrate.

Color Tubes. Not many radar indicators have utilized color tubes in the past. The information obtained by gradation of hues is not definitive enough to warrant use of color in fixed fire-control, surveillance, or bomber defense radar systems. There are indicators, however, where contrasting colors such as red and green could display zones of danger (profilometer indicator) or distinctive signals (differentiation of fixed and moving targets) and would therefore be of great use.

The color tubes and color phosphors presently available are not ideal for radar applications. A tube which can withstand the rigors of a fighter aircraft environment is still not available. Persistence limitations of single-layer phosphors which also have color purity still provide a deterrent to their use in radar color display tubes. It is anticipated that the development work being conducted in electroluminescence, in thin transparent phosphors, in the Aiken type of deflection tube, and in color storage tubes will produce several useful color indicators.

The Shadow Mask Tube. The tubes available today are the outgrowth of industrial developments aimed at the home television market. The most widely used tube, the shadow mask tube, cannot be used in airborne applications since it is sensitive to the position of the indicator with respect to the earth's magnetic field, but it is worthwhile to consider its construction and operation. The tube utilizes three beams, a shadow mask, and a phosphor dot array as shown in Fig. 12-39. In this dot-array color tube, it is necessary for the three beams to converge exactly at all points in the shadow mask plane, as shown in Fig. 12-40, so that the beams strike the

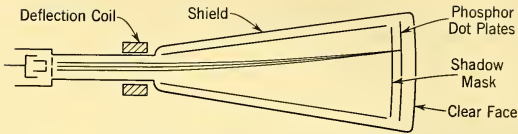


FIG. 12-39 Physical Arrangement of the Shadow Mask Color CRT.

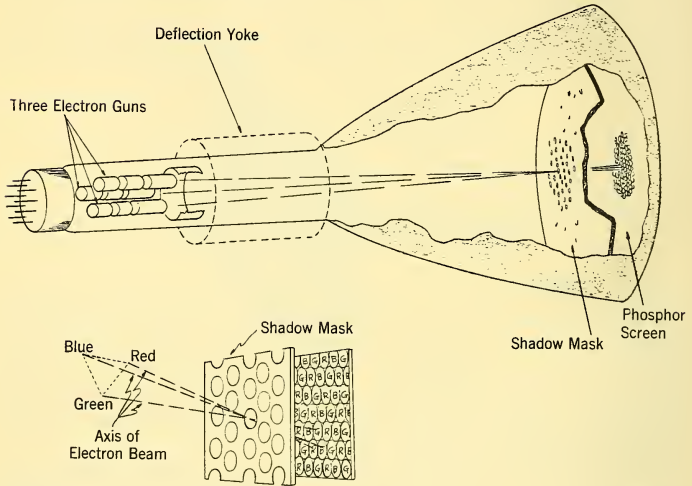


FIG. 12-40 Operation of Shadow Mask Color CRT. [H. B. Law, *Proc. IRE* (Oct. 1951) p. 1187]

proper phosphor dots without affecting adjacent color dots. The shadow mask, a thin perforated diaphragm, blocks any one beam (such as the beam designed to excite only green phosphor dots) from striking dots other than the proper one (green in this case). Separation of colors is obtained by the direction of entry of the three beams through the holes of the shadow mask and by the geometry of the shadow mask and phosphor dot screen. To minimize convergence and focus difficulties throughout the front face of the tube, curvature of the shadow mask and phosphor screen are sometimes used.

In the shadow mask tube different colors are obtained by modulating the three beams to obtain an additive color of any hue, the observer's eyes integrating over the individual dots. All three beams, which are generated by three guns placed symmetrically about the neck of the tube, are deflected

together. The transfer characteristic of each gun must be altered to compensate for the different efficiencies of the three phosphors. If equal radiant energy is not obtained from each color, then the mixed hue is predominantly tinted by one color.

The difficulties associated with the fabrication of a uniform shadow mask and phosphor dot plate are being resolved by using mass production on photoengraving and silk screen techniques. Even with these mass production techniques, the stringent alignment and uniformity requirements make these tubes complex to manufacture and therefore expensive. Continuous operational misalignment and sensitivity to external magnetic fields make this tube unsatisfactory for military display. For the home television system an external magnetic shield of high permeability is used to minimize the effects of external fields. In a moving vehicle such as an aircraft, vibrations are enough to vary the convergence of the beams even with a tube shield.

The Chromatron. A simpler color tube which has been flown in experimental systems has been given the trade name of *Chromatron*. This tube has the advantage of displaying multiple colors with a single beam and the further advantage of not being sensitive to magnetic fields. Although it is not as simple to assimilate into compatible home television systems as the shadow mask tube and is therefore not in general use for this purpose, it will certainly become a useful radar display tool when persistent color phosphors are developed and when the grid vibration problem is solved. The Chromatron tube, as shown in Fig. 12-41a, uses a single beam which is magnetically deflected over a phosphor screen. The three primary-color phosphors are arranged alternating in horizontal strips across the screen and the electron beam made to focus at any point on any one of the color

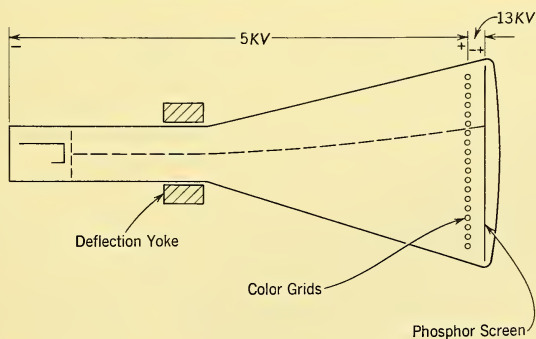


FIG. 12-41a The Chromatron Tube. [J. Gow and R. Dorr, *Proc. IRE* (Jan 1954) p. 309]

strips. Control of the landing of the electron beam on the proper phosphor strip is accomplished by a grid network placed in front of the phosphor. A grid wire is placed in optical alignment with each red and blue phosphor strip. All the grids associated with the red phosphors are interconnected, as are all the grids associated with the blue phosphors. A high accelerating potential is maintained with the blue phosphors. A high accelerating potential is also maintained between the grid array and the phosphor screen. With the proper deflection conditions and no potential difference between the two grids, the electron beam incident on the grid plane will land on the green phosphor. If a potential difference is placed between the wires, however, the beam will be deflected toward the wire which is most positive, producing a blue or red hue as shown in Fig. 12-41b.

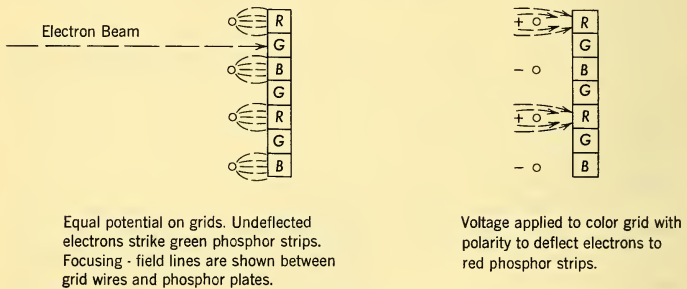


FIG. 12-41b Color Production by Beam Deflection in Chromatron Tubes. [J. Gow and R. Dorr, *Proc. IRE* (Jan. 1954) p. 309]

Switching between colors can be accomplished by a line or frame sequence. Line sequence color switching is accomplished by having the beam oscillate over the three color strips, while synchronously providing the correct color video information. Frame sequence color switching is accomplished by applying one-color video information during one frame and then switching to the other colors on succeeding frames.

In radar systems with long frame times, line sequence color switching is more desirable. Gating circuits for such applications have been developed around sinusoidal switching circuits rather than pulse-type circuits (multivibrators) because of the large capacitance associated between the color control grids. Present tubes have as much as $1400 \mu\text{mf}$ capacitance between the grids. For radar displays where a typical range sweep line might represent 300 microsec (24 n.mi.), a switching frequency of 0.5 to 1 Mc/sec is adequate to present a good color picture. Present mechanizations of gating circuits yield approximately 300 volts rms, utilizing high-Q circuits.

Power loss in the inductors and the grid capacitance requires an oscillator with a power output of approximately 20 watts. Since each color phosphor has a different efficiency, it is necessary to compensate the video information of each color to obtain good hues. Actually, it is easier to accomplish this on the single gun chromatron tube by electronic bias control than on the three-gun shadow mask tube by compensating the three transfer characteristics.

Two-Color Tubes. In many radar displays only two colors are required. For example a two-color presentation might be used to show moving targets among stationary targets. Work in the transparent phosphors, the Aiken deflection system, and electroluminescence indicates that satisfactory progress is being made to realize dual color presentations. With respect to transparent phosphor development, it has been demonstrated that two phosphors can be placed in tandem. One beam excites the first phosphor while a second beam (of much higher energy) penetrates the first phosphor and excites the second phosphor. As can be expected, the color purity is not as good as it is in the tubes discussed above.

Electroluminescence. The control circuitry of this solid-state display is still in the laboratory stage. However, the potentials of this display medium, which offers the multiple advantages of good form factor, storage, no vacuum, no high voltage, high light output, and color capability, are discussed to point out some of the goals of indicator improvement in the future.

Light is generated in electroluminescent cells in basically the same way as it is generated by cathode luminescent methods. An a-c field is applied across a photoluminescent phosphor to change the light level of the screen. The alternating electrical field, although high — of the order of $10^4 - 10^5$ volts/cm — is not enough to excite the phosphor. There are, however, regions of higher electrical stress inside the phosphor particles. These local fields may be as high as 10^7 volts/cm. In these areas of high local fields there are present electron-trapping centers as well as luminescent centers. (These electron-trapping centers lie below the conductive band by only a small amount of energy.) When the intensity of the electric field reaches a high point, electrons in these traps are freed and enter the conductive band. These electrons are accelerated by the high electric fields and thereby attain considerable energy above the bottom of the conductive band. Some of these accelerated electrons collide with filled luminescence centers, liberating electrons in the center. Eventually the avalanche of free electrons will reach regions of lower field strength and will be terminated. When the applied field is reversed (a-c voltage) the electrons will return to the region containing empty luminescence centers and recombine with them and thereby emit light.

The best known materials which exhibit these electroluminescent qualities are ZnS and ZnSe. Other materials which are being investigated are Zn_2SiO_4 , $BaTiO_3$, and some organic compounds.

The electroluminescent "cells" are made by one of three methods:

1. Suspending the powder in a plastic layer which is sandwiched between two conductive layers, the first layer being a glass coated with tin oxide (NESA) and the second layer vaporized aluminum. The a-c voltage is applied between the conductive layers.

2. Embedding the phosphor in a coating of low-melting-point glass or enamel, which is then fired onto a metal or glass base. A second electrode of transparent conductive film completes the sandwich. This type of construction develops a very rugged "cell."

3. Suspending the phosphor in a plastic medium and sandwiching this between a conducting glass cloth or metal mesh and (second electrode) vaporized aluminum to develop flexible electroluminescent cells.

The electrical properties (of the electroluminescent "cells") of time constancy, brightness, and frequency characteristics must be considered. If the phosphor has been completely de-excited (by heat, for example) and an electrical field is placed across it, then it requires approximately 5 cycles to reach 50 per cent of the light output. Although the actual time of excitation can be decreased by increasing the frequency of the alternating field, the number of cycles cannot be decreased. On the contrary, the higher the frequency, the greater the number of cycles that are required to reach the required light level. If the phosphor has not been de-excited (has some of its traps filled from previous electroluminescent excitation) the build up to 50 per cent of maximum output would require one or two cycles.

Electroluminescent phosphors show rapid decay after excitation. A typical decay of 0.5 time constant is 0.5 msec for ZnS:Mn. This figure represents an exponential decay reaching a 10 per cent output level. Temperature does not noticeably affect either the time constant of excitation or decay.

Electroluminescent brightness increases nonlinearly and is dependent on the applied field. It may be expressed as follows:

$$L = Ae^{-(E_0/E)^{1/2}} \quad (12-6)$$

where L = brightness, E = field strength, and A and E_0 are constants characteristic of a particular material at a particular frequency. It is typical of electroluminescent phosphors that the brightness increases linearly with frequency. However, the voltage drop incident to the impedance of the cell electrodes begins to limit the output at high frequencies.

The cells can be operated at any voltage by varying the thickness of the phosphor, since it is the electric field intensity which is of paramount

consideration in generating light. The thickness of the material, which is a few thousandths of an inch, is limited by pinholes and nonuniformities. It is necessary to limit the thickness of the material to prevent dielectric breakdown. Of course, an increase in thickness of material means the application of a higher voltage to obtain the same electrical field intensity. Since there is more phosphor in the thicker material, there will be more light emitted.

Electroluminescent material emits light as a function of magnitude and frequency of an electrical field. The electroluminescent display device therefore must make use of controlling elements to vary the light intensity of each point of the luminescent material. Variation of the light intensity is of interest to us in two ways — first as a light amplifier and second as a display device.

Light amplifiers or image intensifiers are of interest to us because of their ability to improve the brightness of our presentation. They are also a first step toward a highly sophisticated solid-state presentation device.

The first light amplifier was a single-layer type. This single-layer type is excited by the direct action of incident radiation such as ultraviolet or X-rays. Destriau¹² first showed that an a-c field applied across a photoluminescent phosphor screen would change the light level of the screen. If the exciting radiation were ultraviolet rather than visible light then the light output caused by the electrical field dropped by 50 per cent. D. A. Cusano¹³ applied a d-c voltage rather than an a-c voltage and with ultraviolet radiation demonstrated a light enhancement of about 60 times. Fig. 12-42 illustrates the Cusano cell. The Cusano image intensifier has a thin manganese-activated zinc phosphor (10 microns thick) which is vapor deposited. It has extremely high resolution and a capability of reproducing

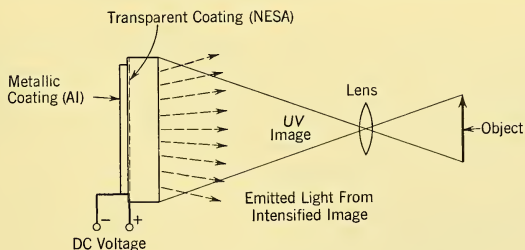


FIG. 12-42 The Cusano Light Amplifier.

¹²G. Destriau and H. F. Ivey, "Electroluminescence and Related Topics," *Proc. IRE* **43** No. 12 (Dec. 1955).

¹³D. A. Cusano, "Field Enhanced Solid-State Luminescence," referenced in *Phys. Rev.* **98** No. 4 (May 1955).

halftones. There are limitations to this device: (1) this type of cell has energy gain only in the ultraviolet part of the spectrum, while our requirement for intensification is in the visible part of the spectrum; (2) the response time for changes in the ultraviolet radiation is a few seconds. This single layer type has been described here simply for its historical importance.

Much work has been done in a double-layer type of intensifier, where the control of the electrical field across the electroluminescence is developed across a photoconductor by incident visible light radiation. A double-layer image intensifier¹⁴ is illustrated in Fig. 12-43. A thin photoconductor, whose

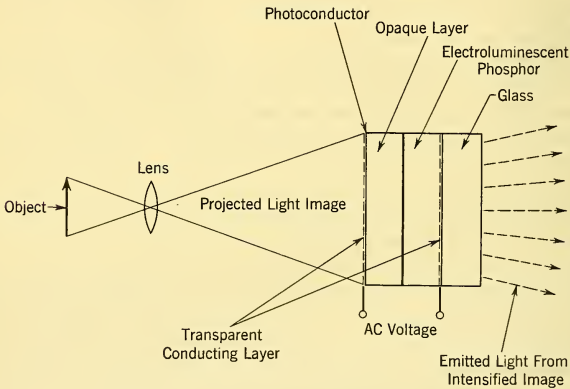


FIG. 12-43 Double Layer Image Intensifier.

impedance per unit area is very great compared with the impedance of the electroluminescent layer per unit area, is placed electrically in series with the light emitter material. An a-c voltage is placed across the sandwich. If incident light is made to fall on the localized area of the photoconductor, thereby changing its resistivity, then the increased a-c voltage (or increased dielectric field) across the electroluminescent cell causes light emission. The ability to obtain gain in this device is dependent on the change in resistivity of the photoconductor and the light efficiency of the electroluminescent phosphor.

The major limitations presently encountered in the development of photoconductors are (1) inability to obtain uniform photoconductor layer materials, (2) difficulty of proper impedance matching to the electroluminescent layer and (3) poor response time. Very few photoconductive

¹⁴F. H. Nicoll and B. Kagan, *J. Opt. Soc. Am.* **45**, 657 (1955).

materials have the proper impedance necessary to make a good electro-luminescent control device. Single crystal CdS has the proper impedance, but techniques for producing large single crystals have not been developed. Vapor deposition of CdS has met with limited success because of the difficulty of obtaining a thick enough layer to get a proper impedance matching.

Consideration has also been given to utilizing the combination of photoconductor and electroluminescent phosphor as direct display devices. Although an adequate resolution of about 50 lines per inch is obtainable and will probably be further improved, the limitation of rise and fall-off is the greatest deterrent to developing this control element into a useful radar display device. If a powder with a short time constant is developed, another limitation is encountered. The short persistence would not allow enough light storage at one unit area (assuming a display with moving information which is made up of many bits) as the radiation source is sequentially moved to write a pattern. Since human intelligence requires at least a 1/30-second frame time, the integrated light level from any particular unit area is extremely small and becomes smaller as the bit rate is increased. What is required, as in all indicators, is methods of (1) placing information on a given unit area in as fast a time as is demanded by the system, (2) retaining this information for a full frame rate until a second bit of information needs to be placed on the same unit area, (3) erasing the first bit information at a fast rate, and (4) writing the new information.

Development of such a solid-state device by E. A. Sack¹⁵ contemplates using ferroelectric capacitors in a matrix array to control the electro-luminescent phosphor. A simple schematic of one element of this array appears in Fig. 12-44. The ferroelectric and electroluminescent capacitors

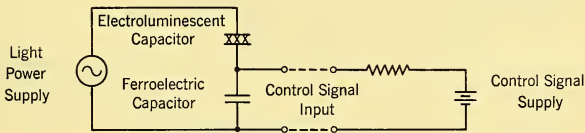


FIG. 12-44 Schematic of ELF Screen.

are connected in series with a source of a-c power. The voltage drop across the electroluminescent phosphor is controlled by the impedance of the ferroelectric capacitor. As shown in Fig. 12-45, a ferroelectric material such as barium strontium titanate has the property of changing its capacitance value as a function of change of potential across the dielectric. The impedance change which is caused by the d-c potential bias and input signal

¹⁵E. A. Sack, "ELF — A New Electroluminescent Display," *Proc. IRE* **46**, 1694-99 (1958).

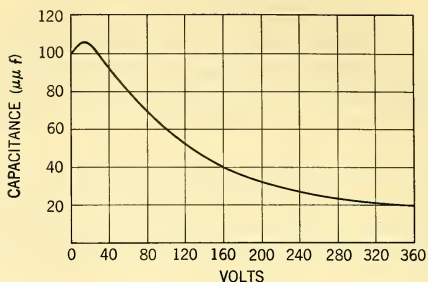


FIG. 12-45 Capacitive Characteristic of Ferroelectric Material.

shown in our schematic varies the voltage drop across the electroluminescent phosphor. The change in voltage across the light-producing element is rapid with change in control potential, since we are basically dealing with a capacitor-divider. The leakage resistance of this ELF (electroluminescence-ferroelectric) circuit is extremely high, thereby retaining the voltage across the light-producing phosphor for time durations of several minutes. We can see in the ELF material the capabilities of fast time constant and storage that are prerequisite to the ideal display device.

Although the ferroelectric material has the advantage of good impedance matching and fast time constant, much development must be done to overcome the disadvantages of nonuniformity, temperature sensitivity, and relatively high control voltage range.

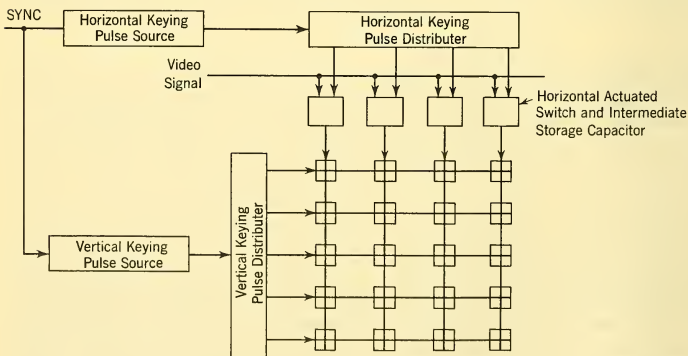


FIG. 12-46 An Electronic Distribution System for ELF.

The major difficulty in mechanizing the ELF screen is in the signal distribution problem of placing a pulse voltage at the proper ferroelectric element without resorting to an extremely expensive matrix array of diodes. An electronic distribution system such as is shown in Fig. 12-46 illustrates the elements necessary to perform the operation of placing a signal at the proper point at the proper time. A keying pulse source for the horizontal and vertical keying pulse distributor must be added to address the video pulse to the matrix ELF point. In a radar display of the B scope type, the horizontal pulse distributor could possibly be a mechanical commutator, while the vertical distributor could take the form of a delay line. The method illustrated can be considered a line-sequential system in that intermediate storage capacitors are made simultaneously to charge up the corresponding screen element as the horizontal switch commutates on a line-by-line basis. It is, of course, important to add erase capability to this circuit. This is accomplished by developing a pulse such that the previous line of information is removed before new video information is placed on the intermediate storage capacitor. The construction of the ferroelectric control and the distribution system is a formidable problem. The fabrication techniques being applied include evaporation for forming the electrode pattern on the electroluminescent phosphor layer and printed circuit techniques used to deposit the ferroelectric capacitors, intermediate storage capacitors, diode switches, and even the high-frequency distributor (delay line). It is anticipated that a resolution of a few hundred lines to the inch will eventually be realized.

Beam-Shaping Tubes. The ordinary CRT, as well as the storage tube version, is capable of presenting two-dimensional information. In the discussion of the various displays we have seen that the radar systems usually require a means for presenting additional information. It was mentioned that one or more additional guns in the same envelope or time sharing of a single gun is used to display the required additional information on the same screen. Additional information in the form of symbols may also be displayed by means of specially designed "beam-shaping" tubes such as the Typotron¹⁶ and Charactron.¹⁷ In this type of tube a stencil containing the symbols has been inserted in the electron path. Usually the stencil contains 64 symbols arranged in 8 lines and 8 columns. Two sets of deflection plates or coils are needed, one set for selecting the symbol and one set for positioning it in the correct place on the screen. By proper coding of the symbols, considerable information may be presented on the screen. Generally one of the symbols on the stencil is selected when the beam is required to write ordinary CRT information.

¹⁶Trade name, Hughes Aircraft Co.

¹⁷Trade name, Stromberg Carlson Co.

These tubes are useful for extremely high-speed teletype equipment. They are capable of a writing speed of 25,000 symbols per second, provided the input channels have wide enough bandwidth. A disadvantage of the tubes is their large length, which is about twice that of an ordinary CRT, to give room for the additional deflection and focusing system.

The front end of these tubes may be either direct phosphor excitation as in the CRT or indirect and with controlled persistence as in the direct view storage tube (in which case a storage surface and flood gun is incorporated). Fig. 12-47 illustrates the functioning of the latter as in the Typotron. A

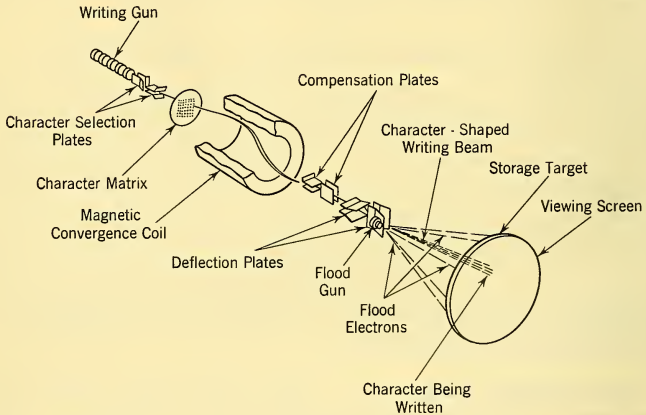


FIG. 12-47 Schematic of the Character-Writing Typotron Storage Tube.

first set of deflection plates directs the beam through the selected symbol of the stencil, whereby it is given a cross section corresponding to the symbol. A magnetic field bends the beam back to the tube axis, and compensation plates align it so as to enter the second set of deflection plates, which in turn direct the beam to the correct location on the storage mesh. A positive charge results on this spot on the storage surface caused by secondary emission in the same manner as in ordinary storage tubes. The flood gun is located next to the deflection plates. The Typotron is operated as a bistable storage tube; i.e., it has only two shades of gray — black and white.

Instead of directing a narrow beam through a single symbol, a wide collimated beam may illuminate the entire stencil as shown in Fig. 12-48. The entire beam is focused through the first set of deflection plates (selection plates) which bends the beam so that only the part containing the desired symbol penetrates a fixed aperture and subsequently is deflected

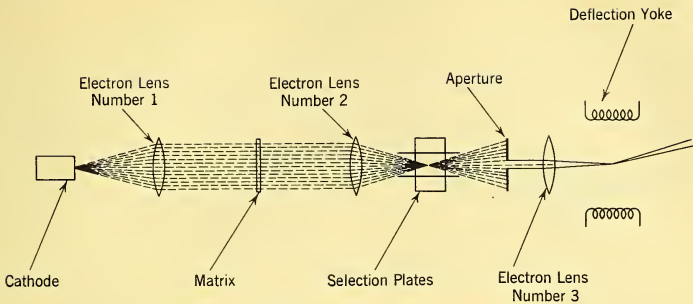


FIG. 12-48 Basic Optics of the APSEL Tube.

to the proper point on the screen. This principle is used in the APSEL version of the Charactron. The total length required is less than with the first method.

12-10 SPECIAL DISPLAYS

Special indicators have been devised to optimize information display in particular tactical situations.

Terrain Avoidance. Attack against ground objects may be conducted at a low altitude. Avoidance of high obstacles as well as detection of targets becomes the concern of the radar operator and pilot in such a tactical situation. A radar system designed for terrain avoidance operation requires high resolution, a short frame time, and an easily interpreted alarm display. High resolution is necessary to recognize the terrain objects in the search sector. A short frame time is necessary to up-date the information quickly to allow maximum aircraft operation decision time. Both resolution and frame time are limited by other units of the radar, such as the antenna, rather than the indicator.

Typical indicator configurations for terrain avoidance make use of either B, C, or sector-scan PPI presentations. Such displays give a maplike picture of the terrain obstacles which must be avoided. The ideal display should show range as well as angular position. The B, C, or sector-scan PPI are, therefore, compromise displays which leave out some information which is useful to the operator. In the C scope, Fig. 12-49, a profile of the terrain can be visualized from the display and therefore the pilot can adequately judge the aircraft maneuver necessary to evade the obstacle; however, he has no knowledge of the terrain behind the displayed terrain, and therefore idealized flying along the valleys between terrain obstacles is difficult. In the B (or sector-scan PPI), as shown in Fig. 12-49, a better

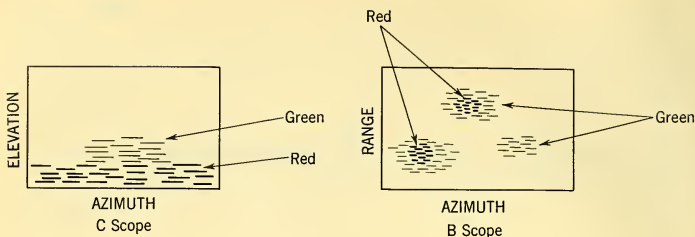


FIG. 12-49 Terrain Avoidance Displays.

flight plan can be developed from the display, but information of the height of the obstacles is not readily presented.

Mechanization of indicators has been accomplished using both monochromatic and two-color tubes. In a monochromatic display only the object of danger is displayed. Those objects are shown which, because of their range and altitude position, are an immediate barrier to aircraft. Using multicolor displays, differentiation of objects in near or far range and above and below the line of flight is graphically displayed. In the color display, quantizing of the radar information is mechanized to show the objects which constitute a danger in red color. Objects which are not barriers to the aircraft flight are displayed in green.

Improvement of these displays to show more subtle changes in terrain will come from use of the advanced tubes which offer greater color differentials.

Windshield Projection. Much work is being done to relieve the cockpit and instrument display complex that is apparent in modern aircraft. Simplification of instrumentation and windshield projection of the radar information are approaches to the problem.

Windshield projection, in its simple form, presents the typical displays which have been discussed onto a surface where the pilot is naturally looking. Much more sophisticated windshield displays are being developed, in which information of a tactical nature as well as radar data are presented in a more easily understandable form. Windshield display is developed by use of either transparent tubes like the Aiken tube or mirrors. Fig. 12-50 illustrates a configuration of the use of a diachroic mirror to effect a windshield display. With the usable brightness level of the storage tube, the attenuation of this system is not of great consequence.

Contact Analogue. The symbolic display of the AI radar is a coded language which the operator must learn. Although it is not complex, it has no analogy in human experience. Experimental efforts are being made by

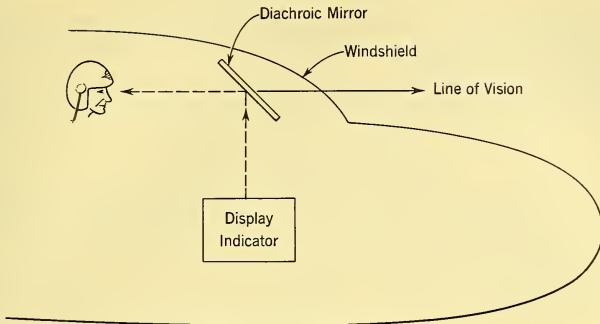


FIG. 12-50 Windshield Projection Using Diachroic Mirror.

the Office of Naval Research and the Bureau of Aeronautics to provide visual signals which are easier to interpret by the operator and thus to improve his decision-making capability. These symbols are intended to give to the operator a presentation similar to his view if he were to look out of his windshield.

The contact analogue display considers that external references, terrain texture, size, motion, shape, and perspective are the elements needed in the display to most readily present his environment to the operator. The information obtained by the contact analogue method is intended not only to improve the 'strike' presentation but also to provide information for takeoff and land, navigation, and aircraft instrument display.

As shown in Fig. 12-51 a display of grid structure simulates increasing or decreasing terrain texture as the aircraft gains or loses altitude. The convergence of the grid lines toward the center of the display gives perspective to the presentation. Distance can easily be imagined. Aircraft speed is shown in such a display by the relative motion of the horizontal grid lines toward the base of the display.

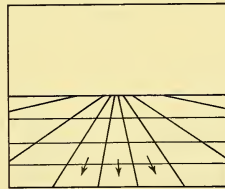


FIG. 12-51 Contact Analogue Display.

Variation of motion of the grid lines easily illustrates a variation in aircraft speed relative to the ground. Aircraft roll, as shown in Fig. 12-52, is represented by a rotation of the grid structure in the same direction as the aircraft's attitude. Pitch is displayed by displacing the horizon grid line away from the center of the tube as shown in Fig. 12-52. Yaw, in turn, is displayed by moving the vertical center of the scope picture

in the same direction as the aircraft slippage. All these visual cues appear to the operator as if he had an external reference on the ground to which he can refer his own position.

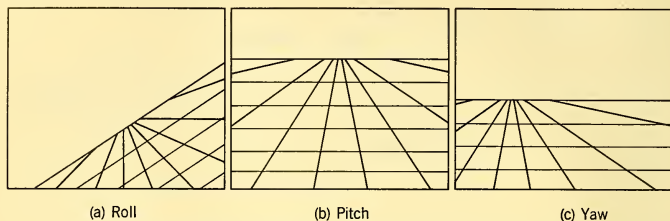


FIG. 12-52 Motion Display — Contact Analogue Indicator.

The ideal radar video information displayed on such an indicator would not only indicate the angular position of the objects of interest but would picture the information by size and shading to indicate range and bulk. The B type display of course, being cartesian, would not serve as a contact analogue display, since the antenna search cone is made in a solid angle. The limitation of this display is not in the CRT. The complexity of the information handling requires a sophisticated computer.

Aural Indicator. To relieve the burden imposed on the operator's visual sensors — his eyes — consideration has been given to using the operator's aural perception as an aid in gathering radar information and control data. The ability to use the hearing sensors is limited, however, since the ear is capable of resolving information of relatively low bandwidth content. It works best when there is only one type of information present, and it does not have the integrating capability of the eye. For example, it is difficult to listen to crew command conversation and radar-processed data at the same time. Because of these limitations not many airborne presentations, in the past, utilized the operator's hearing ability.

If the operator's aural perception is to be used, the parameters that can be distinguished are amplitude, frequency, and duration of tone. The audible range in loudness from threshold up to the pain level covers some 120 db over the medium frequencies and tapers off towards the high as well as towards the low tones. Because of the noise level in the airplane, the lower part of the loudness range cannot be used for conveying information. To prevent disturbance and fatigue the louder range cannot be used. The loudness range usable as an information parameter is consequently greatly reduced.

A frequency spectrum from about 30 cps to 12,000 cps can be heard by most people. The sensitivity, however, rapidly decreases towards both

ends. Many people can hear only a narrower band. The ear has essentially a logarithmic response. It is a sensitive comparator of differences or variations in loudness and frequency but is a poor indicator of absolute values. For this reason, absolute values cannot be given by audible signals. Deviation from otherwise determined procedures, direction of flight, approach, etc., may be given audibly, however.

The use of the dynamics of loudness or frequency variation can be adapted to define parameters such as range to a target or aid in steering toward a tracked target. They have not been used to any degree to date because of the above-mentioned shortcomings and the ease with which the visual sensors and the electrical-to-light transducers accommodate the transfer of this intelligence to the brain.

Some particular radar system signals are readily handled by audible methods. Most of these take the form of alarm or "go-no-go" signals. Instead of lights blinking, buzzers can be made to sound. The disturbance factor of such an audio signal, if loud enough, probably has more impact than visual alarms. Such alarms can be used for engagement breakaway in aircraft interceptors, guidance path directors in landing, and collision avoidance in low-flying aircraft. In most cases, even though the alarm is given in audible processes, the operator must revert to his visual perception to understand the cause of alarm and the best way to overcome his danger.

Newer radar systems which use doppler techniques to separate moving targets of varying speed by frequency separation can readily make use of an audio tone to locate and maintain track on a target. AMTI (airborne moving target indicator) radars used for air-to-ground search and track can be mechanized to sense the presence of moving targets by audible methods if the antenna dwell time on an area is long enough to obtain an audio tone. Manual track on such targets can be maintained (until the automatic lock-up processes are complete) by the differentiation of the frequency of a particular target from other audio tone. Again, in such a system the operator reverts to his indicator scope to understand his hearing with relationship to the acquired target. However, in this radar the hearing sensor can serve as a valuable adjunct to the visual sensor rather than as a second-best substitute.

REFERENCES FOR CHAPTER 12

1. T. Soller, M. A. Starr, and G. E. Valley, *Cathode Ray Tube Displays*, Massachusetts Institute of Technology Radiation Laboratory Series, Vol. 22, McGraw-Hill Book Co, Inc., New York, 1948.
2. F. E. Terman, *Radio Engineer's Handbook*, "Electron Optics and CRT's," Sec. 4, Par. 15 and 16, McGraw-Hill Book Co., Inc., New York, 1943.

3. L. N. Ridenour, *Radar System Engineering*, Chap. 6, Massachusetts Institute of Technology Radiation Laboratory Series, Vol. 1, McGraw-Hill Book Co., Inc., New York, 1947.
4. J. R. Pierce, *Theory and Design of Electron Beams*, 2nd ed., D. Van Nostrand Co., Inc., Princeton, N.J., 1954.
5. D. G. Fink (Ed.), *Television Engineering Handbook*, Chap. 5, "Cathode-Ray Devices," McGraw-Hill Book Co., Inc., New York, 1957.

CHAPTER 13

MECHANICAL DESIGN AND PACKAGING*

The preceding chapters have discussed the principles of airborne radar system design from the points of view determined by system requirements, by electromagnetic wave propagation, and by electronic considerations. The thorough application of these principles to a problem should result in development of a sound system concept and suitable circuits. The reduction of these concepts and circuits from diagrams on paper and from breadboards in a laboratory to a production radar set requires a large amount of engineering effort, much of which is mechanical in nature. The major portion of this mechanical engineering work is devoted to the many phases of what is commonly called "electronic packaging." Given good circuit design, the reliability of a system is controlled in large part by the quality of its packaging.

The term *packaging*, as used here, refers to the selection, mounting, assembly, and interconnection of the component parts of an airborne radar for its required service. Packaging design usually includes at least the design of a chassis supporting the components and of protection from some of the other environmental factors. (The design of the antenna mount, especially for a tracking radar, is a separate major project in machine design, not normally considered as packaging.) Among the major influences upon this phase of design are the following.

1. Circuit requirements
2. Space and weight limitations
3. Environment
4. Necessity for maintenance and ease of installation
5. Transportation and supply
6. Flexibility
7. Reliability
8. Economy of manufacture

The mechanical design of any electronic package is primarily determined by the circuit requirements, such as the number and type of components. While this factor must be included foremost in a list of factors influencing

*Paragraphs 13-1 through 13-14 are by R. M. Sando and J. W. Titus. Paragraph 13-15 is by M. Goetz.

electronic packaging design, it is so evident that it requires little discussion. Practically all design is influenced by space and weight limitations, and the packaging of an airborne radar is no exception. The available space, generally well defined by the airframe designer, is invariably so small as to require a considerable amount of ingenuity of the equipment designer if he is to fit it in while meeting all of the other requirements. Weight limitations are equally important. In an airborne application anywhere from 5 to 20 pounds of airframe, engine, and fuel weight are required to support each pound of payload. This ratio — defined as *growth factor* — increases with the speed and altitude for which the aircraft is to be designed. For space vehicles, growth factors from 50 : 1 to 100 : 1 are common. In a particular application the radar designer should be thoroughly familiar with the trade-offs between the weight of the radar equipment and the performance of the weapons system. The specification "as light as possible," while laudable in intent, is not a sufficient incentive to design ingenuity.

13-1 THE INFLUENCE OF ENVIRONMENT UPON DESIGN

An airborne radar which contributes to the control of guided missiles often is installed in high-performance aircraft operating at extreme speeds and altitudes. It must perform reliably according to specifications in spite of acoustic noise, vibration, heat, and other adverse environments. Failure of a single component of the radar to operate satisfactorily when subjected to the environment of a mission may mean failure of the aircraft to accomplish its task. Thus the design of an airborne radar, like the design of other military equipment, must enable it to survive the undesirable conditions imposed by the various environments encountered during its life. The information presented in this chapter includes a description of the various categories of environment, some of their adverse effects upon the equipment, and some of the design principles that will assist in obtaining optimum performance from the electrical equipment. An extensive treatment of guided missile design in the expected environment is to be found in the volume *Aerodynamics, Propulsion, Structures, and Design Practice* (see Bibliography at the end of this chapter: item 1 under Books).

Environment, for the purposes of this chapter, may be defined as the aggregate of all the external conditions and influences affecting the life and performance of an equipment. It may be classified by type and condition as was done in Paragraph 1-9, i.e. as the *physical* environment and the *airframe* environment. The elements of the physical environment likely to cause difficulty include temperature, pressure, humidity, precipitation, salt spray, radiation, sand, dust, etc., while those of the airframe include dynamic forces, acoustic noise, explosion, etc. During its life, the equipment will be subjected to the many rigors of flight conditions, as well as takeoff

and landing, both while operating and while secured. The life of the equipment also spans the conditions of manufacture, testing, storage, transportation, handling, maintenance, and possibly shipboard operation. An environment or envelope of environments is associated with each of these conditions. Thus while the flight conditions are most important, all conditions must be considered when the critical design values are selected for the various types of environment. For example, mechanical shocks received during ground handling and transportation may exceed in severity any shocks imposed by the aircraft in which a radar is ultimately installed.

The environment classifications mentioned above are closely related to two categories called *natural environment* and *induced environment*. The natural environment includes the influences which would be present at a point in space before the arrival of the aircraft. These ever-present factors include ambient wind, temperature, pressure, density and other specific properties of the atmosphere, solar radiation, humidity, salt spray, and moisture in other forms. The effects of these factors will be discussed in detail in this chapter. There are also several types of natural environment which must at some time be considered by the designer of an airborne radar, but do not warrant a detailed discussion here. Among these are sand and dust. The induced environment is the aggregate of the conditions and influences (affecting an equipment) generated by the vehicle in which the equipment is installed. One of the most troublesome examples of induced environmental conditions affecting an airborne radar is the temperature resulting from aerodynamic heating. Other examples are acceleration loads from aircraft maneuvers, acoustic noise, vibration and shock, and nuclear radiation. In most respects, an airborne radar actually lives in the induced environment which is, therefore, the one of chief concern.

The induced environment is, of course, influenced by the natural environment in which the aircraft flies. For example, the shock and vibration acting on the equipment may be influenced by wind gusts and air density. In a piece of electronic equipment cooled simply by ram air, the temperature of the air entering the equipment is roughly the temperature of the ambient or "natural" air plus the temperature rise incident to compression of the air entering the plane.

The location of airborne electronic equipment in its parent aircraft has an important effect upon the severity of the environment which it must withstand. The types of environment that change significantly in magnitude as a function of location in the aircraft are temperature, vibration, shock acceleration, acoustic noise, and (if present) induced nuclear radiation. A designer has a very limited opportunity to locate the units of an airborne radar so as to take advantage of this effect. For example, in the case of an AI radar the indicator and controls have to be in the cockpit, the antenna in a particular place, depending upon the type of target, and weight and

maintenance factors will count heavily in the location of the other units. The principle should, however, be borne in mind.

Another important factor controlling the severity of a particular type of environment is its duration. For example, the temperature rise of the aircraft skin may be a transient phenomenon rising to a significant magnitude only during a particularly high-speed flight condition. If this high-speed flight condition exists for a sufficiently short period of time, the effect upon the radar may be small, because of thermal lag. Failure from vibration and shock is also very much determined by the duration of exposure.

It is necessary that the effect of environment be considered from the very beginning of the packaging design effort. Initial engineering decisions on configuration and many details will be directly related to this problem. The design for vibration, shock, heat, and some other factors is so difficult that it is often necessary to supplement engineering judgment by the testing of models and prototypes at the earliest possible stage. This point may seem too evident to require emphasis. In practice, however, a designer is often under pressure to meet a time schedule, and it is not necessarily easy to obtain a clear and consistent description of all phases of the environment. Thus there exists a severe temptation to proceed without due regard to the basic environment problem. Such a procedure can be expected to produce equipment which will fail to pass the acceptance tests to which it is ultimately subjected. Experience with radar and other airborne electronic equipment shows that this chain of events happens all too frequently. Since these final tests usually occur very late in a large and lengthy development program, the correction of design errors then becomes costly and difficult.

In principle, there are two methods of designing to meet the environmental challenge. One is to obtain components and to use design factors that will enable the radar to function properly in spite of any specified environment. At the other extreme is the selection of components and completion of all design assuming an ideal environment, that is, ignoring the environmental problem. The equipment is then isolated from all undesirable factors and put in the ideal environment required for its operation. It is doubtful if any useful airborne radar could be designed on the basis of either method alone, at this time. Practical design will produce reliable equipment by a suitable combination of these two approaches. This point will be expanded in the following discussions of the various types of environment.

13-2 MILITARY SPECIFICATIONS

To obtain an overall system design that will provide optimum operation while being subjected to the various categories of environment, it is necessary to have design criteria and methods of testing the individual

component parts. The design philosophies and environment requirements are set forth in documents known as *military specifications* (MIL Specs). The documents also establish procedures for testing airborne equipment under simulated and accelerated environmental conditions.

Types of information contained in the military specifications, and listed therein under the heading "Requirements," are environmental tests, modification of test limits, test facilities, test measurements, and sequence of tests. Listed under "Test Procedures" are tests of reliability in face of temperature and altitude, vibration, shock, humidity, salt spray, explosion, sand and dust, fungus, and acceleration.

Recommended materials, parts and processes are sometimes outlined in the military specifications in order to achieve optimum space, weight, and interchangeability and still allow the designers to meet the performance and environment requirements. Methods of mechanical locking, fastening devices, structural mounting of component parts, electrical connections, wiring, etc. are included in order to aid designers in creating products that can resist the imposed environmental conditions.

The major documents that outline the general philosophy of the design and testing of airborne electrical equipment are: MIL-E-5400A(ASG), General Specification Electronic Equipment (aircraft); MIL-T-5422C(ASG), Aircraft Electronic Equipment, Testing Environmental; MIL-E-5272A, General Specification for Environmental Testing, Aeronautical and Associated Equipment; and MIL-E-8189A, General Specification for Electronic Equipment, Guided Missiles. Other applicable specifications defining detail requirements of items such as equipment mounting bases, vibration and shock mounts, electrical connectors, installation of wirings, etc. are all referred to in the above-mentioned specifications.

Military specifications are necessary in order to obtain uniformity of design, high reliability, operation and maintenance suitability, economy, manufacturability, and weapon compatibility. Conformance to the specifications will aid in achieving an optimum airborne electronic system which is required to operate while being subjected to the various environmental conditions.

13-3 TEMPERATURE

Plainly the performance of an airborne radar system will be affected and limited by the temperature of its components. This is also true of the other equipment installed in the aircraft, and indeed, of the aircraft itself. Temperature has an important effect upon the materials problem. The variation in structural properties of metals and plastics with temperature (see Fig. 13-1) may have to be considered in designs for installations in high-speed aircraft. More critical, however, are the characteristics of the

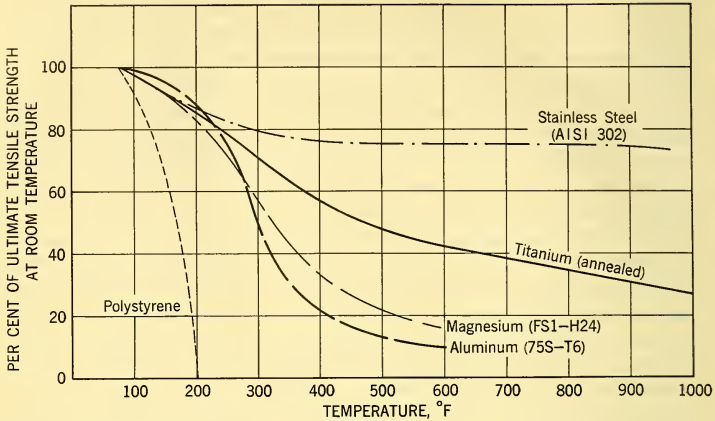


FIG. 13-1 Decrease in Strength of Common Structural Materials at Elevated Temperatures.

nonstructural components. The life expectancy of vacuum tubes decreases rapidly with increasing envelope temperatures. The transistors now available are even more sensitive to temperature than the vacuum tubes. Wattage ratings of resistors must be reduced as ambient temperatures increase. Magnetic properties of materials change with temperature. Other examples could also be cited.

The radar equipment will normally gain heat (*a*) from internal sources, (*b*) from the environment, or (*c*) from the entering "cooling" fluid, if used.

Although internally produced heat is, technically, not an environmental factor, it is convenient and logical to consider it at this point. In craft moving at subsonic speeds, this heat, produced by vacuum tubes and resistive components, is the principal factor contributing to temperature rise. It will still be important wherever use is made of heat-dissipating components.

As long as the equipment is cooler than its environment, heat may flow into it by conduction through the supporting structure, by direct radiation from the hot surfaces which it can "see," and by natural convection of air. To the radar packaged inside the airframe, the principal environmental sources of heat are the aircraft skin and the other electrical and electronic components. In the forward part of the aircraft, the skin temperature will approach the aerodynamic stagnation temperature. This means that aerodynamic heating becomes serious at supersonic speeds. The relation-

ship between the *stagnation temperature*, speed, and altitude is

$$T_s = T_a + \frac{1}{2} \frac{V^2}{\mathcal{J}gC_p} \quad (13-1)$$

where T_s = stagnation temperature, degrees Rankine

T_a = ambient air temperature at altitude, degrees Rankine

V = aircraft velocity, ft/sec

\mathcal{J} = Joule's constant, 778 ft lb/Btu

g = acceleration of gravity, 32.2 ft/sec²

C_p = specific heat of air at constant pressure, Btu/lb-°F. The specific heat is a variable, dependent upon the temperature of the air. A value of $C_p = 0.24$ Btu/lb-°F can be used for air temperature below 1000°F without introducing a noticeable error.

To give an example of the serious nature of aerodynamic heating, the stagnation temperature will be calculated for a Mach number of 2.5 (velocity = 2905 ft/sec) at sea-level conditions of the standard hot atmosphere. Ambient temperature, from Reference 1, is 103°F.

Then, $T_a = 460 + 103 = 563^\circ\text{Rankine}$

$$T_s = 563 + \frac{1}{2} \frac{2905^2}{(778)(32.2)(0.24)} = 1266^\circ\text{R}$$

$$t_s = 1266 - 460 = 806^\circ\text{F.}$$

Fig. 13-2 shows the variation in stagnation temperature with Mach number for the standard hot atmosphere.

Another source of heat is solar radiation. For ground operations, solar radiation can subject aircraft compartments to temperatures in the vicinity of 160°F. These high temperatures can occur at times when the equipment cooling system is inoperative and for periods of time long enough to render useless the natural thermal lag capabilities of the electronic equipment in its installed position. For flight conditions, the effects of solar radiation are of greatest relative magnitude for those portions of the flight profile encompassing low-speed, high-altitude combinations, since aerodynamic heating is small under these conditions. Therefore, time, as well as exposed body surface area, is a determining factor in obtaining an appreciable heat load from solar radiation. The higher the altitude of the body within the earth's atmosphere, the greater the solar radiation intensity, because less scattering and absorption of the sun's rays take place farther away from the

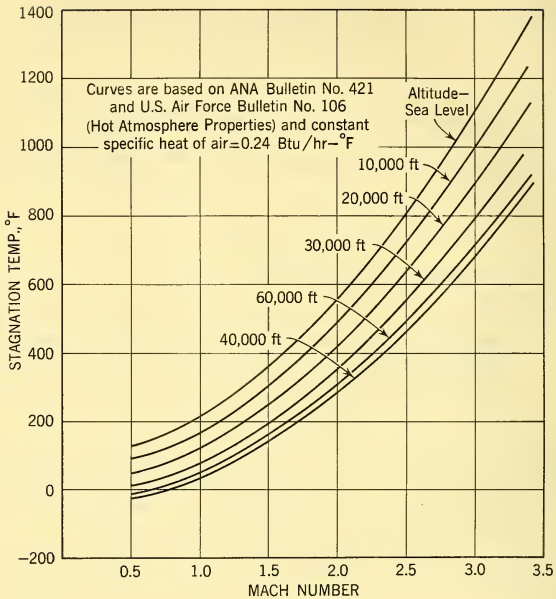


FIG. 13-2 Stagnation Air Temperature vs. Flight Speed and Altitude.

earth's surface. Since, however, the factor of transmissivity of solar radiation through the earth's atmosphere is a variable, it is feasible and conservative to consider the maximum solar radiation constant of 429 Btu/ft²/hour to obtain for all design conditions, ground operations as well as flight conditions. Therefore, the actual rate of heat transfer to the aircraft or body skin by solar radiation (resulting in an increase of internal compartment temperatures) is the product of the solar radiation constant, the absorptivity of the body surface, and the surface area. For purposes of calculation, the surface area used should be only that area exposed to the sun's rays. The temperature effects of solar radiation should be added to the effects of aerodynamic heating and equipment heat in order to establish the total temperature of the compartment housing the electronic equipment. Solar radiation is a large percentage of the total load for that phase of the equipment life concerned with ground operation. This is especially true during the nonoperating phase of the equipment, since the total compartment ambient temperature is then determined by the effects of solar radiation.

The principal thermal design problem is to keep the temperature of each component low enough to allow satisfactory operation of the component during the required period of time. The principal approaches to this problem are

- (a) Control of internal heat sources
- (b) Cooling (provision of heat sinks)
- (c) Increase in allowable component temperature.

Control of internal heat sources consists of the use of components generating a minimum of heat and the separation of heat sources from temperature sensitive equipment. This approach is fundamental but limited in applicability by the availability of components with the desired characteristics. In the design of an airborne radar, every effort must be made to use components tolerant of the highest temperatures, insofar as this can be done without compromising other qualities. In most cases, however, these two techniques alone will not produce equipment which can operate reliably in the thermal environment. Resort to some types of cooling will therefore be required.

The equipment as a whole may normally lose heat in several of the following ways:

- (a) By radiation
- (b) By conduction to structural members of the airframe
- (c) By convection, i.e. forced cooling using a fluid
- (d) By the conversion of sensible heat to latent heat, i.e. by the evaporation of a liquid or sublimation of a solid.

As discussed above, the structure surrounding the radar in a supersonic aircraft may well be hotter than the allowable temperature of some of the components, or at least too hot for direct radiation and conduction to provide significantly useful cooling. Thermal insulation techniques may indeed be desirable to check the flow of heat into the equipment.

Many systems, both simple and complex, have been developed for the removal of heat from aircraft electronic equipment, using the principles (c) and (d) above, alone or in combination. A few typical systems are depicted in Fig. 13-3 in schematic form. In Fig. 13-3a the critical components are cooled by a direct supply of ram air (forced air). In Fig. 13-3b, the electronic unit is sealed and pressurized. The sealed-in air is cooled by being circulated through a heat exchanger cooled, in turn, by the ram air. In Fig. 13-3c (another pressurized system), the design provides a conductive path adequate to bring the heat from the critical elements to the stream of ram air. This system is commonly called *cold plate cooling*. In any of these three systems, if the ram air is not substantially cooler than the lowest

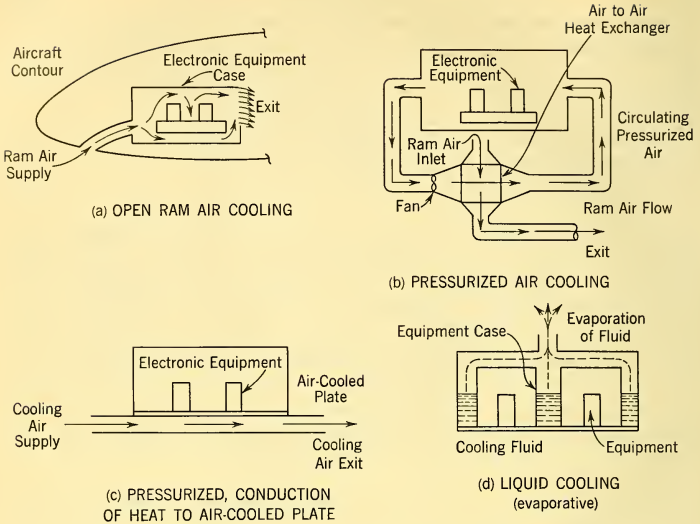


FIG. 13-3 Cooling Schemes for Electronic Equipment.

maximum allowable component temperature, it will have to be cooled before entering the equipment heat exchanger. Schemes *b* and *c* are more critical than *a* in this respect. In Fig. 13-3d, which may also be pressurized, the internal heat is conducted to an enclosed tank of liquid. The evaporation of the liquid absorbs heat, preventing the sensitive components from experiencing temperatures greatly above the boiling temperature of the liquid.

Cooling and temperature-control systems can operate in a transient or a steady state. If the thermal environment and internal heating remain constant for a sufficient time, a steady-state heat transfer condition is reached. The flow of thermal energy for this condition may be indicated as shown in Fig. 13-4.

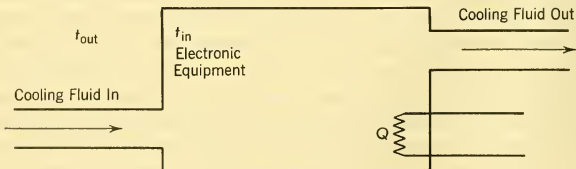


FIG. 13-4 Steady-State Cooling.

Let \mathcal{Q} = heat production rate of equipment, Btu/min (approximately equal to input power)

U = heat transfer coefficient of equipment enclosure for combined radiation and conduction, Btu/sq ft-min-°F

A = area of equipment housing, sq ft

t_{out} = external temperature of equipment housing, °F

t_{in} = temperature inside equipment housing, °F

W = flow rate of cooling fluid, lb/min

h_{in} = specific enthalpy of entering fluid, Btu/lb

h_{out} = specific enthalpy of fluid at exit, Btu/lb.

Then the steady-state heat flow can be expressed by equating the heat removed by the cooling fluid to the total heat load:

$$W(h_{out} - h_{in}) = \mathcal{Q} + UA(t_{out} - t_{in}). \quad (13-2)$$

If the cooling fluid undergoes no change of state, the equation becomes:

$$WC_p\Delta t = \mathcal{Q} + UA(t_{out} - t_{in}) \quad (13-3)$$

where C_p = average specific heat of cooling fluid over the operating temperature range Btu/lb-°F

Δt = temperature rise in cooling fluid, °F

Equation 13-2 applies to any system using fluids for cooling. If applied to a system similar to Fig. 13-3d, W would still represent the flow rate of cooling fluid in lb/min although the flow would be only that of the evaporated fluid. The h_{in} would be the specific enthalpy of the saturated liquid; therefore, $h_{out} - h_{in}$ would equal h_{fg} , the latent heat of vaporization.

Equation 13-3 is useful for obtaining a preliminary estimate of the required rate of flow of cooling fluid in any proposed system similar to those of Fig. 13-3a, b, or c. In such a study, the quantities on the right-hand side would be calculated. The value for Δt would be the result of an estimate of the differential temperature required between the hottest component and the cooling fluid. This is often obtained by assuming a heat-transfer efficiency η defined as follows

$$\eta = \frac{\Delta t}{t_{max} - t_f} \quad (13-4)$$

where t_{max} = maximum allowable component surface temperature

t_f = temperature of entering cooling fluid.

As stated above, the objective of a cooling system is to keep the temperature of each component low enough to allow its satisfactory operation during the required period of time. While Equation 13-2 gives the necessary condition for the steady state, i.e., that the heat lost equal the heat gained, it is not a sufficient condition for the proper cooling of each component. The latter depends upon achieving satisfactory paths for the heat flow from the sources (including environment) to the cooling fluid which acts as a heat sink. The means for accomplishing this varies with the type of cooling used. For example, in systems similar to those in Fig. 13-3a and b, the problem essentially is to meter the correct proportion of the available air to each heat-producing component. The provision of a good conductive path is essential in systems similar to Fig. 13-3c and would be helpful to systems such as in Fig. 13-3a and b. In a system such as Fig. 13-3d, the natural convection is negligible; the flow of heat to the cooling fluid depends upon conduction, forced fluid circulation, or a combination of the two.

The circulating cooling fluid in a sealed or closed-cycle system or portion of a cooling system must be chosen by the designer. In principle, a variety of gases, liquids, and mixtures is available. Air is most often used, since it has adequate thermodynamic and dielectric characteristics for most purposes, aside from being the most available fluid. Helium and hydrogen have superior heat-transfer properties which may justify their consideration in specific cases. Various liquids have favorable characteristics; their weight, however, restricts airborne cooling applications. Although there are evident disadvantages to such a scheme, the use of air with entrained liquid has been seriously advocated for the cooling of airborne equipment. This would combine the light weight and low viscosity of a gas with the high latent heat of evaporation of a liquid, producing the effect of a gas with exceedingly high specific heat.

Since entrained liquid is usually not desired in cooling air, it is necessary, whenever refrigerated ram air is to be used, for a designer to ascertain whether water vapor in the air can be expected to condense. It is shown in various treatises on the thermodynamics of moist air that the maximum amount of water vapor which will remain in a stable mixture with air is given by the following expression:

$$W_s = 18/29(P_s/p - p_s). \quad (13-5)$$

In this equation, W_s is the number of pounds of water vapor per pound of dry air, in a saturated mixture. The coefficient 18/29 is the ratio of the molecular weight of water to the "apparent molecular weight" of air. The quantity p is the total atmospheric pressure. The quantity p_s is the saturation pressure of water and is a function of temperature which is tabulated in engineering handbooks. To determine whether condensation may take place, the cooling-system designer calculates the worst tempera-

ture and pressure combinations which can be expected to occur in cooling air during various flight conditions. For each combination of temperature and pressure, a value of W_s can be calculated from Equation 13-5. Whenever W_s is less than the number of pounds of water vapor per pound of dry air in the ambient atmosphere, condensation will occur. A cooling system design similar to Fig. 13-3a will then need to be changed, either to provide removal of water, or to avoid combinations of pressure and temperatures which cause frost or water to appear. Closed cooling systems, such as in Fig. 13-3b and 13-3c should operate very well with entrained condensate, unless ice forms and blocks the air passage. The prevention of icing in air passages requires very careful design.

In many flight conditions, especially with higher-speed aircraft, the steady-state heat transfer condition is never reached. For the transient case, the heat flow may be described by

$$W(h_{out} - h_{in}) + MC_p(dt/d\theta) = \mathcal{Q} + UA(t_{out} - t_{in}) \quad (13-6)$$

where M = weight of equipment, lb

C_p = average specific heat of equipment

t = average temperature of equipment at an instant of time

θ = time, in minutes.

Thus the heat inflow is balanced by the combined effect of the heat removed with the cooling fluid and the heat required to raise the temperature of the equipment. The factor $dt/d\theta$ is proportional to the temperature difference between the equipment and internal environment and can therefore be expected to be an exponential function. The analytical evaluation of this function is not feasible for most practical equipment; it may be obtained, if desired, by experiment.

To look at the transient heat flow in another way, first integrate Equation 13-6 with respect to time between the limits θ_0 and θ , letting t_0 be the value of temperature at the initial time θ_0 .¹ Then, solving for t ,

$$t = t_0 + \frac{1}{MC_p} \left[\mathcal{Q}(\theta - \theta_0) + UA \int_{\theta_0}^{\theta} (t_{out} - t_{in}) d\theta - W \int_{\theta_0}^{\theta} (h_{out} - h_{in}) d\theta \right] \quad (13-7)$$

This equation shows the dependence of equipment temperature upon the components of heat flow and upon its mass and specific heat. The second and third terms in the bracket are exponentials; this variation in temperature with time is therefore often called *thermal lag*. Full use should be made of this principle when transient conditions exist; otherwise, a cooling system may be designed with unnecessarily large capacity.

¹In the heat flow calculations, θ is used to represent *time* to avoid confusion with the temperature symbols.

13-4 SOLAR RADIATION

Solar radiation is included in the category of natural environments. The intensity of the solar radiation acting on a body is directly dependent upon the altitude of the body, the type of atmosphere (the amount of dust, smoke, moisture, etc. which affect the absorption, scattering, and reflection of the rays) and the angle at which the rays enter the atmosphere. Thus a body can receive solar radiation either by direct rays or indirect rays of varying intensity, depending upon the above-mentioned conditions.

Special design considerations are involved for electronic equipment exposed directly to the sun's rays. However, this is not the normal situation unless the electronic equipment is placed in an exposed place prior to the actual installation in the aircraft. Special surface finishes, coatings, and material would be required for directly exposed parts, since solar radiation deteriorates materials like rubber, plastics, textiles, paints, etc.

A body subjected to solar radiation will experience an increase of temperature in the enclosed areas. As was pointed out in the discussion on temperature, the design of electronic equipment housed in aircraft compartments subjected to a solar radiation environment must consider the increase of temperature for all phases of the equipment life. In general, the thermal effects of solar radiation for airborne electronic equipment are appreciable for ground operations and will dictate the volume and temperature limits of the required ground cooling system to keep a satisfactory ambient temperature within the electronic equipment compartment. However, the effects of solar radiation are small during flight when compared to the other forms of heat generation and do not necessarily influence the type of equipment cooling system required for satisfactory operation.

The critical areas of the electronic equipment for solar radiation should be determined, and their influence on the overall temperature problem must be considered in assessing their effect on the equipment design criteria.

13-5 NUCLEAR RADIATION

The use of electronic equipment in nuclear-powered aircraft will subject the equipment to the induced environment of nuclear radiation. This environment is due to a high-intensity field of nuclear radiation emitted from the nuclear reactor of the aircraft power plant. Radiation from a nuclear reactor suitable for an aircraft power plant can be divided into three types — beta particles, gamma rays, and neutrons. Beta particles are easily dealt with, requiring very thin layers of material to stop penetration; they thus present no major problem. Gamma rays are of high energy content; the amount of attenuation of these rays depends directly upon the density of the material. A very dense material like lead is often used for radiation

shields. Attenuation of neutrons differs from that of gamma rays in that neutrons are slowed down most effectively by using light materials for shielding. Water is a commonly used substance to attenuate neutrons. The actual amount of destruction or deterioration of either the mechanical or electrical component parts of a system depends on the type of radiation and distribution of energy in the immediate area.

Electrical materials are affected in many ways upon being subjected to a nuclear radiation field. Studies indicate that components like transistors, condensers, rectifiers, and semiconductors in general are unstable at certain levels of radiation. Moreover, serious trouble can occur in tubes because the irradiation process deteriorates the various types of tube seals, thus allowing air into the tube. Various changes in the characteristics of oil subjected to nuclear radiation (for example, changes in the viscosity) can result in the breakdown of oil-filled capacitors. Resistors may malfunction as an effect of nuclear radiation acting on the insulation used on the cover.

Insulation materials used for most electrical components are not able to resist nuclear radiation. The fatigue life or wear rate of some electrical components appears to be accelerated in a nuclear or ionized atmosphere. Relay and switch contacts are examples of this type of effect. Changes in electrical resistivity have been demonstrated in electrical insulators.

Structural materials used in the fabrication of electronic equipment are affected to varying degrees when subjected to a nuclear radiation environment. For example, the aluminum alloys indicate an increase in the tensile, yield, and shear strengths and a decrease in the percentage of elongation, resulting in a reduction in ductility. This leads to an increase in the notch sensitivity or stress concentration effect and a decrease in the ability of the material to resist an impact loading. The corrosion resistance of material is improved, when subjected to certain levels of radiation intensity, because of the chemical effects on the surface of the material.

The elasticity of materials similar to rubber is reduced. Fluids used in hydraulic systems, lubricants of different types, and fuel for aircraft power plants all exhibit unfavorable changes in their characteristics when subjected to nuclear radiation. Plastics tend to change in color and become increasingly brittle with increase of radiation intensity.

To produce an airborne radar system capable of satisfactory operation in a nuclear radiation environment requires the maximum of effort in the study of basic materials, electrical component fabrication techniques, and shielding methods. Close coordination between the electronic equipment designer and airframe designer is an absolute necessity in achieving the optimum design.

It is not possible to shield the nuclear reactor completely and thus eliminate this environment external to the shield. For ground-installed nuclear power plants, the large bulk or weight and surface area necessary

for substantially complete shielding can be tolerated, but volume and weight are primary factors in the design of an aircraft. Therefore, airborne electronic equipment must exist and operate satisfactorily in a nuclear radiation environment which will vary in intensity depending upon the amount of shielding and type of shielding used and the distance and direction of the equipment relative to the nuclear reactor.

Various techniques can be used to reduce the intensity of this environment acting on the equipment. Three methods of shielding can be utilized; and the total shield volume can be distributed around specific items or compartments in the aircraft. Partial protection from the effects of nuclear radiation can also be accomplished by locating items such as fuel cells, bomb bays, landing gear, etc. between the nuclear reactor and the compartments housing the electronic equipment, thus achieving a blanketing effect. Concentrating the entire required shield volume and weight around the nuclear reactor results in a heavy weight design. Distribution of the shield to specific areas affords maximum protection per unit weight since the required amount of shielding is governed by the distance and direction from the nuclear reactor; further, the nuclear radiation energy, when deflected from its original direction, loses strength. However, the disadvantage of this method is the fact that sections of the aircraft not chosen to be shielded will receive maximum nuclear radiation doses beyond those of the concentrated shield design. A distributed shield system in combination with the blanketing concept appears to have merit. Since nuclear radiation may be scattered by items external to the power plant and may therefore approach a given location from several directions, the problem of shielding design is further complicated. This fact influences the design of the shield to the extent that, for maximum protection of electronic equipment, the entire compartment must be enclosed. However, for optimum weight distribution of the shielding material, various degrees of shielding thickness can be used, depending on the relative distances and directions of the compartment walls with respect to the nuclear reactor. For example, aircraft compartments located forward of the nuclear reactor will sustain maximum radiation intensity on the aft wall of the compartment. The forward wall will be subjected to the lowest level of radiation intensity, and therefore a relatively thin, even transparent, shield can be installed there.

13-6 VIBRATION AND SHOCK

The possibility of mechanical failure under dynamic load presents one of the most serious threats to reliability in the complex electronic components of an airborne radar. In military specifications it is customary to handle the general topic of dynamic loading in terms of the four environments of vibration, shock, acoustic noise, and aircraft acceleration. In particular,

vibration and shock constitute a challenge which engages much of the attention of designers, specification writers, and environmental test specialists. Vibration and shock problems are the subject of several books and a vast number of technical papers and articles.

A *vibration* is an oscillatory motion continuing over an extended period of time. This oscillation may be sinusoidal, having a definite frequency and amplitude; it may be a complex periodic motion which can be described by a Fourier series; or it may be a random motion describable only in statistical terms. When an oscillatory force supplies to a vibrating system sufficient energy to compensate for that lost by damping, the vibration is termed *forced vibration*. When vibrations are maintained only by the elastic and kinetic energy initially in the system, they are called *free* or *natural vibrations*. Since some friction will always be present in a mechanical system, free vibration is a transient condition. Forced vibration, on the other hand, is normally a steady-state condition.

Mechanical shock is a transient condition wherein a body or system is distributed by a sudden impulse of force or a sudden change in velocity. This disturbance is followed by a damped free vibration in which the system returns to equilibrium or to its steady-state vibration, depending on the environment.

The dynamic loads of shock and forced vibration can occur during the transportation and handling of the radar and during the various phases of aircraft flight operation, including taxiing, takeoff, flight maneuvers, and landing. Aerodynamic effects and power-plant reactions in flight contribute to the generation of potentially dangerous vibration and shock loads. The firing of missiles, rockets, and guns adds further excitation. Equipment destined for naval use will usually be subjected also to the vibrations and shocks characteristic of ships, catapult takeoffs, and arrested landings.

Vibration may cause failures in a radar that are either temporary or permanent. The temporary failures are the malfunctions which persist during vibration but leave no mechanical damage. A common example of a temporary failure would be a vibrating relay contact. The permanent failures occur in components, structural members, and fasteners which sustain various types of mechanical damage leading to equipment failure. The loosening of threaded fasteners during vibration may be considered a third class of failure. Since the reliability of the radar system operation is a measure of design success, it must be remembered that a temporary malfunction of the equipment, even though caused by deflections well within the elastic limit of the materials used, is a real failure of mechanical design or construction, just as surely as a broken bolt or a cracked casting.

In general, vibration failures of a permanent nature are fatigue failures resulting from the repetition of many cycles of stress within the elastic limit of the material used. Thus vibration failures are a function of (a) the

environment forcing the vibration, (b) the response characteristics of the equipment and its parts, and (c) the fatigue characteristics of the materials used. Design for reliability in a vibrating environment accordingly requires constant awareness and consideration of these three factors. Vibration testing machines are of use in simulating a known or specified environment in order to obtain response characteristics or otherwise evaluate tentative or finished designs.

The character of the vibration environment of an installed airborne radar can be measured in flight by means of vibration pickups and suitable recording equipment. The prediction of this environment for a radar under design is difficult, however, because the aircraft for which it is destined will probably not be available at this stage. Even if a similar aircraft were available, the results of vibration measurements must be used with caution, for they will be completely valid only when the aircraft is loaded with a dynamically similar model mounted in the exact manner planned for the radar. The mechanical design and the early phases of vibration testing, therefore, must usually depend upon an environment estimated from past experience, possibly described in a military specification. Periodic vibrations in aircraft have been observed at frequencies ranging from about 5 cps to high acoustic frequencies; however, many authorities now agree that the structure of a high-speed airframe usually vibrates in a random manner forced by the cumulative effect of a very large number of independent actions.

Random motion means nonperiodic motion which is the result of a very large number of impulses occurring by chance and which may be described by a normal probability law. The instantaneous acceleration of a typical random environment is plotted against time in Fig. 13-5. It can be seen

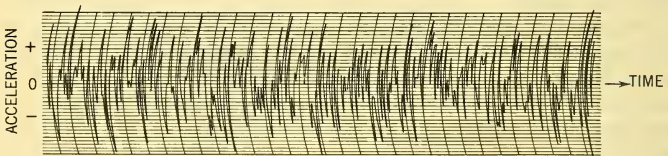


Fig. 13-5 Instantaneous Acceleration of Typical Random Motion.

that some acceleration peaks have a greater magnitude than others and that the period of time between values of zero acceleration varies irregularly. Although the precise acceleration which will occur at any future instant cannot be calculated, the probability of that acceleration occurring can be stated.

The characteristics of a random acceleration can be described in terms of the mean square acceleration spectral density, or as it is often called,

acceleration density. The quantity is defined by

$$G = \lim_{B \rightarrow 0} \frac{a^2}{B} \quad (13-8)$$

where G = acceleration density; the usual units for acceleration density are g^2 per cycle per second

a^2 = mean square acceleration within a specified bandwidth

B = the bandwidth within which the acceleration a occurs.

In this chapter, the acceleration density function is defined for positive frequencies only. Thus, it differs slightly from the power density functions of Chapter 5. It is evident from the defining equation that for a given random motion this property is a function of frequency. A plot of the acceleration density of Gaussian noise as a function of frequency, known as an *acceleration density spectrum*, defines a steady-state random motion as completely as it can be known. Two such plots are shown in Fig. 13-6. The solid

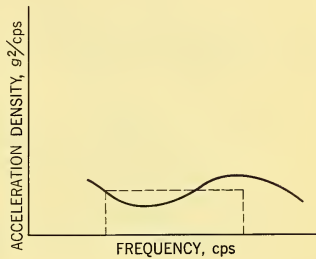


FIG. 13-6 Acceleration Density Spectra.

curve shows an acceleration density spectrum which might be measured in a random environment. The dashed curve represents a special case called a *flat random motion spectrum* or a *white random motion spectrum*. Such a curve is often chosen for a test specification, being a convenient compromise between many different actual environments.

If the acceleration density spectrum is known, the rms acceleration may be calculated for any frequency band. For the band lying between frequencies f_1 and f_2 , the rms acceleration may be calculated by Equation 13-9a

$$a = \sqrt{\int_{f_1}^{f_2} Gdf}. \quad (13-9a)$$

For the case of white random motion spectrum, this expression simplifies to the following:

$$a = \sqrt{BG}. \quad (13-9b)$$

Before considering the response of a system to random excitation, it is helpful to review briefly the case of sinusoidal vibration. If a sinusoidal

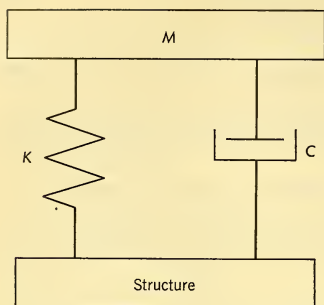


FIG. 13-7 Mechanical System with a Single Degree of Freedom.

force is applied to a linear mechanical system, it will vibrate (in the steady state) at the frequency of the applied force, with an amplitude depending upon the magnitude and frequency of the force, and upon the characteristics of the system. Let Fig. 13-7 represent a single degree of freedom translational system consisting of a mass mounted upon a structure through a spring and a viscous damper. The motion of the structure provides the excitation for the system.

In the figure, let

M = mass, slugs or $\text{lb sec}^2/\text{ft}$

c = damping coefficient, $\text{lb sec}/\text{ft}$

K = spring rate, lb/ft

$c_c = 2\sqrt{KM}$ = critical damping coefficient, $\text{lb sec}/\text{ft}$

$\Omega = \sqrt{KM}$ = undamped natural frequency, rad/sec .

The response of the mass M to a steady sinusoidal motion of the supporting structure is shown in Fig. 13-8 as a function of frequency and of the ratio of the given damping to the critical damping. The ordinate of the figure represents the *displacement transmissibility*, that is, the ratio of the maximum displacement of the mass M to the maximum displacement of the structure. It can also be taken to represent *force transmissibility*, the ratio of the maximum transmitted force to the maximum applied force. The abscissa is the ratio of the forcing frequency to the undamped natural frequency of the mass spring system (obtained by assuming infinite structural mass). This figure shows that at low frequencies the spring acts in a rigid manner, causing the mass nearly to reproduce the motion of the structure. At high frequencies the spring more or less isolates the mass from the excitation. At intermediate frequencies, close to the undamped natural frequency, the maximum displacement reaches values large compared to the excitation. This condition is known as *resonance*; the frequency at which maximum displacement is reached, incident to a sinusoidal force of constant amplitude, is known as the *resonant frequency*. When the damping is small, as is usually the case with mechanical vibrations, the resonant frequency nearly equals the undamped natural frequency.

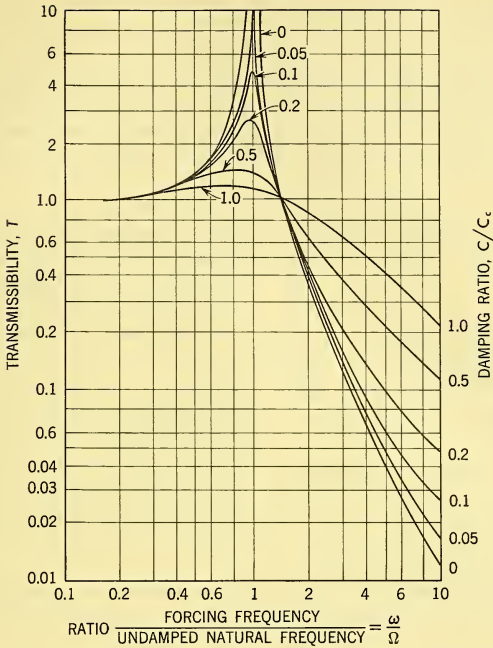


FIG. 13-8 Force and Displacement Transmissibility for a Viscously Damped, Single-Degree-of-Freedom with Sinusoidal Excitation System Excited by Random Vibration. (After C. E. Crede, "Vibration and Shock Isolators," *Machine Design*, Aug. 1954)

The maximum transmissibility for sinusoidal excitation is now often called the \mathcal{Q} of the resonance, by analogy with the usage which has long been standard in electrical networks. This quantity is conveniently used in a discussion of random vibration. It is important to remember that the peak amplitudes plotted in Fig. 13-8 apply for steady-state vibration. It requires a number of cycles to reach this peak.

Let n be the number of cycles required to permit the vibration of mass M to reach 95 per cent of its steady-state amplitude.

$$\text{Then:} \quad n \cong 3\mathcal{Q}/\pi \quad (13-10)$$

The following expression is useful when c/c_c is less than 0.2:

$$\mathcal{Q} = \frac{1}{2c/c_c} = \frac{1}{c} \sqrt{KM}. \quad (13-11)$$

If the single mass and spring of Fig. 13-7 were to be replaced with a more complex system, the single resonant "mode" of Fig. 13-8 would be replaced by a number of resonant modes equal in number to the degrees of freedom possessed by the system. Each mode of vibration would be characterized by a resonant frequency and a \mathcal{Q} . It can be seen that for complex electronic equipment in an environment of sinusoidal vibration, the most damaging structural effects will occur at the resonant frequencies of the various modes.

If the excitation of the system (i.e. the motion of the structure) of Fig. 13-7 is random, the response of the mass will necessarily be random; its time history, however, will look quite unlike that of the excitation. For

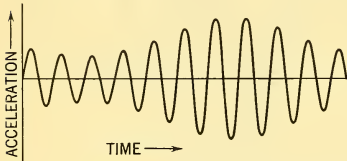


FIG. 13-9 Vibration of Single-Degree-of-Freedom Resonant System Excited by Random Vibration.

a single degree of freedom resonant system with little damping, the time plot of the response is rather similar to that of the excitation as viewed through a narrow-band filter. A brief sample of the acceleration response of such a system is shown in Fig. 13-9. It can be seen that the curve looks much like a sine wave with the resonant frequency of the system, but with a continually

changing amplitude and phase.

If the acceleration density of the structure motion is known in the region of the resonant frequency, the rms acceleration of the mass may be obtained.

Let a_M = rms random acceleration of the mass, g

G = acceleration density of structure, g^2/cps

Ω = undamped natural frequency of the system of Fig. 13-7 (or of a normal mode of a more complex system), rad/sec

\mathcal{Q} = maximum transmissibility for the corresponding mode.

Then $a_M = \frac{1}{2}\sqrt{G\mathcal{Q}\Omega}$.

To sum up, the response of a resonant system to random vibration is a motion very similar to a sine wave at the resonant frequency of the system. The rms acceleration of the response is directly proportional to the square roots of the acceleration density of the excitation, the \mathcal{Q} of the system, and the undamped natural frequency.

It was stated in the introduction to this discussion that most fractures resulting from vibration are *fatigue failures*. The term *fatigue*, by common engineering usage, refers to the behavior of materials under the action of

cyclic stress, that is, under vibration, as distinguished from their behavior under steady stress. It is well established that metals can be made to fail under the action of cyclic stress even though the maximum stress of the cycles is far below their static breaking strength. Fatigue failure appears to start with the development of a crack in or near the surface of a part. Under the influence of the cyclic stress, this crack propagates steadily through the part until the remaining material is unable to sustain the maximum load. At this instant, a rapid failure occurs.

Design for cyclic stresses depends, in part, upon the results of fatigue tests made upon numerous identical specimens of the material in question. The usual graphical presentation of these results is known as an *S-N diagram*. One form of S-N diagram is shown in Fig. 13-10. Here the logarithm of the maximum stress in a load cycle is plotted versus the logarithm of the number of cycles required to produce a failure.

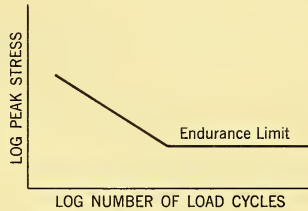


FIG. 13-10 S-N Diagram.

Many materials, when tested, give a curve similar in shape to that shown in Fig. 13-10. The stress corresponding to the horizontal part of the diagram is called the *endurance limit* for the conditions of the test. It is evident from inspection of the figure that for parts which may be stressed above their endurance limits, there is a relationship between peak cyclic stress (or vibration amplitude) and probable life. The S-N diagram gives this relationship for the specimens tested. Application to an actual design may require consideration of many additional factors affecting the fatigue strength of a part. Some of these are now listed:

- | | |
|------------------------------------|--|
| 1. Mean stress | 8. Overstressing (before service) |
| 2. Residual stress | 9. Other environmental factors —
temperature, atmosphere, external pressure, etc. |
| 3. Size of specimen | 10. Surface finish |
| 4. Stress concentration | 11. Surface cold work |
| 5. Frequency of stress cycle | 12. Surface coating or treatment |
| 6. Heat treatment of specimen | |
| 7. Understressing (before service) | |

Analytical design for vibration is further complicated by the fact that the available data on fatigue characteristics of metals are based upon sinusoidal tests and therefore require interpretation before being applied to random stresses.

If an environment with severe vibration is expected, the avoidance of damage will require a design based upon consideration of the expected

vibrations, the principles of mechanical response, and the fatigue characteristics of the contemplated materials. It has been shown above that the most damaging structural effects from vibration occur at the various resonant frequencies of the equipment. These frequencies can often be calculated for simple designs; but in the later stages of equipment development they are usually determined by the observation of a prototype on a vibration testing machine.

If the frequencies of environmental vibration are known to lie within certain narrow bands, damage will be avoided by the obvious technique of designing equipment all of whose resonant frequencies lie well outside of these known bands. Unfortunately, this approach is not feasible for random vibration or for wide ranges of sinusoidal frequencies. In such cases, the maximum accelerations to which components and parts are subjected can be limited only by the environment, the Q 's of the various modes, and, for random excitation, the frequencies of the natural modes. It is rarely possible for the designer to exercise much control over the Q 's in electronic packages. He is thus left with the alternatives of designing equipment which can operate reliably in spite of resonant vibration or of protecting the equipment from its environment by means of vibration isolators.

In packaging for vibration, therefore, certain principles must govern the detail design. Every attempt should be made to have all natural frequencies as high as possible, even though the acceleration of a system may be somewhat more violent at higher frequencies. This course is justified by the relation between amplitude and frequency. As frequency increases, the amplitude of a vibration with given acceleration decreases rapidly. The corresponding stress, proportional to amplitude of vibration, will decrease similarly. A second principle is design for efficient storage of energy. For example, every member subject to stress should be efficiently proportioned to give a uniform maximum stress level. All sources of stress concentration should be eliminated, where possible. These procedures contribute to efficient storage of energy under shock as well as vibration.

An aircraft radar may experience shock, especially in its shipment and handling prior to installation, during ordinary and arrested landings, during catapult takeoffs, and in the form of blast when in action. Shock failures may be brittle-state fractures caused by immediate impact, or they may be low-cycle fatigue failures caused by the brief but severe vibration which follows the first impulse of a shock. Shock can also hasten the complete failure of parts in which fatigue cracks have previously formed. It is believed that a shock of magnitude insufficient to cause total failure could initiate unnoticed cracks which would cause stress concentration, reducing the fatigue strength of the part.

The design techniques which tend to give equipment reliability in an

environment of vibration are generally useful in shock as well. When the application of isolators is considered, on the other hand, shock and vibration sometimes seem to give conflicting requirements. This topic will now be treated briefly. Crede² has stated, "Isolation of vibration or shock signifies the temporary storage of energy, and its subsequent release . . . in a different time relation. Isolation is thus distinct from the absorption or dissipation of energy. The effectiveness of an isolator is sometimes enhanced by limited dissipation of energy, but this is a secondary consideration in its function."

To consider a simple application of vibration isolators, let Fig. 13-8 represent the response of one vibrational mode of a chassis or assembly mounted upon isolators (assuming a geometrical arrangement which decouples the rotational and translational modes). The figure shows that the transmissibility of the system is less than unity for all frequencies greater than 1.4 times the undamped natural frequency, and becomes very small at high frequencies when damping rates are small. The isolators thus protect the equipment from damaging forces which might be transmitted through a rigid connection to the structure, provided that the natural frequencies of the resulting spring-mass system (loosely called the *isolator natural frequency*) are low compared with the environmental vibration frequencies. The low natural frequency of such a system requires relatively "soft" isolator springs, permitting large deflections. In compact aircraft installations, it is often difficult to provide the required clearance between isolated equipments and the structure.

It is not necessary to isolate for all possible environmental frequencies, if the equipment can withstand the transmitted vibrations at isolator resonance. For this situation, the isolators are required to provide significant protection only at the lowest resonant frequency of the equipment and at all higher frequencies. A general rule that appears workable is that the isolator natural frequency should be less than 40 per cent of the lowest resonant frequency within the equipment.

The function of a shock isolator is to reduce large acceleration loads by increasing the time interval during which the change in velocity of the protected equipment occurs. Looking at this in another way, it can be said that shock isolators are required to protect the mounted equipment by storing much of the energy that would be transmitted to the equipment in a shock impulse. This energy is then supposed to be released in a form that does not damage the mounted equipment.

Now let Fig. 13-7 represent a body mounted on shock isolators of stiffness K lb/ft. If the supporting structure undergoes a shock velocity change of V ft/sec and damping is negligible, the maximum acceleration of the mass M is

$$X_M = V\Omega. \quad (13-13)$$

²Charles E. Crede, *Vibration and Shock Isolation*, John Wiley & Sons, Inc., New York, 1951.

Let δ_{\max} be the maximum deflection of the shock isolator. Then

$$\delta_{\max} = V/\Omega. \quad (13-14)$$

Equation 13-13 indicates that a low natural frequency favors good shock isolation, as it favors good vibration isolation. Equation 13-14, however, shows that substantial velocity shocks will be accompanied by large deflections unless Ω is high. For example, with an isolator natural frequency of 15 cps (a common value for vibration isolation), a drop from a height of only 5 inches with inelastic impact will cause a maximum isolator deflection of more than $\frac{5}{8}$ inch if isolator damping is small. Such a large deflection would not normally be permissible, and such a shock would cause a limiting action in the isolator known as *snubbing*. This action increases the severity of the shock by a factor which approaches 2 as the isolator clearance is reduced. Therefore, in many cases an isolator which is beneficial during vibration may be worse than useless during shock.

If the natural frequency of the isolator is sufficiently high to prevent snubbing and if it thus provides some shock protection, the mounted equipment will probably undergo severe vibration at this frequency. Unless the shock isolator is heavily damped, this will require that the structure and components have resonant frequencies about 2 to $2\frac{1}{2}$ times that of the isolator.

It has been pointed out that in order to produce a design that will be reliable under conditions of shock and vibration, it is necessary to consider, from the beginning until the end of the packaging design effort, the expected environment, the response characteristics of the equipment and its parts, and the fatigue characteristics of the materials used. The topic of testing is beyond the scope of this chapter; however, it may be stated that the vibration and shock testing of early units and assemblies is now an essential engineering technique for the development of an airborne radar. Isolators may in some degree protect delicate equipment from the environment; their application, however, has pitfalls and is to be undertaken with caution. A design with high natural frequencies is often necessary in connection with isolator application. This is in any case the generally applicable approach to successful design for shock and for vibration.

13-7 ACOUSTIC NOISE

Supersonic speeds and high engine power are responsible for an induced environmental condition termed acoustic noise. *Acoustic noise* can be defined as vibration waves carried through the air.

In aircraft the two major sources generating these sound pressure waves are the power plant and the aerodynamic effects of the air passing over the various surfaces of the aircraft. Most of the power plant noise from

turbojets and rockets is created by the turbulence of combustion and exhaust. Even higher levels of acoustic noise are generated when afterburners are used on turbojet engines. Fig. 13-11 is an example of the sound

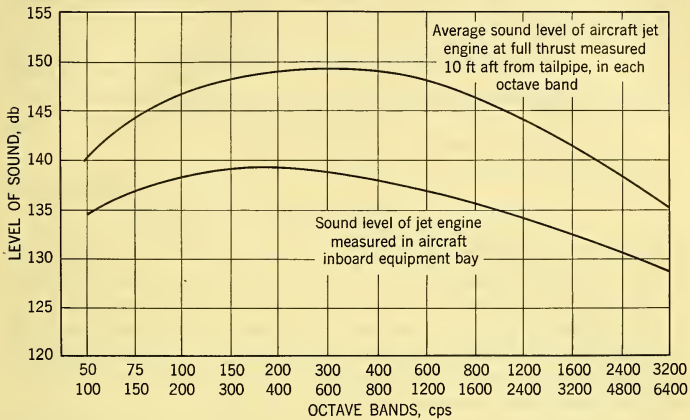


FIG. 13-11 Typical Sound Spectra for Operating Jet Aircraft. "The Scream of a Jet in a Six-Inch Test Cell" by R. H. Jacobson, Research Engineer, Armour Research Foundation, Chicago, Ill.

levels that can be expected from the operation of jet aircraft. (The measure of sound level is defined below.) In propeller-driven aircraft, the major source of noise is usually the propeller.

In-flight conditions do not produce the maximum levels of power-plant noise. The highest levels are generated during engine run-up and acceleration of the engine for takeoff and climb. Tests have indicated that the magnitude of the sound pressure caused by the power plant will not be greater than the magnitude of sound created by the aerodynamic effects for an in-flight condition.

The aerodynamic effects on the level of acoustic noise are not critical for aircraft flight speeds in the subsonic range. However, high magnitudes of sound pressure in the supersonic speed range are created by the air passing over the external aircraft surfaces. The magnitude of sound pressure generated by the aerodynamic effect is established primarily by the air-speed. The effects of ambient pressure and temperature at various altitudes and the actual profile shape and aircraft size are secondary effects on the magnitude of sound pressure.

Sound pressure is specified in terms of decibels. The *sound pressure level in db* is defined as $20 \times \log_{10}$ of the ratio of the induced sound pressure to

the reference sound pressure. The reference pressure usually used in referring to pressure waves transmitted through air is 0.0002 microbar. This sound pressure level provides a good reference since it is audible to less than 1 per cent of the human population. The general equation for the sound pressure level is:

$$\text{Sound pressure level (db)} = 20 \log_{10} p_2/p_1 \quad (13-15)$$

where p_2 = rms induced or measured pressure, microbars

p_1 = reference pressure = 0.0002 microbar.

Some useful conversion factors are:

$$1 \text{ microbar} = 1 \text{ dyne/cm}^2 = 1.45 \times 10^{-5} \text{ lb/inch}^2$$

Acoustic noise levels reaching the area housing the electronic equipment are reduced by the atmosphere primarily by two mechanisms, absorption and deflection. Absorption losses can result from many effects, but the most important of these is the reduction by water vapor and oxygen gas in the air of the sound energy generated by the source. The deflection losses of major importance are caused by the scattering of the sound rays by air gusts and the general turbulent condition of the air.

The acoustic noise environment is responsible for erratic operation of airborne equipment, especially tubes and certain types of relays and accelerometers. High-frequency resonance of tube grid supports and relay contact arms is due to acoustic noise. Sound pressure levels between 110 and 140 db appear to be critical. Frequencies below 500 cps are serious for tubes. Some frequency bands existing above 1000 cps cause resonance on parts of relays, resulting in unsatisfactory performance of the relay. Besides the malfunctioning of electronic equipments caused by this environment, structural failure of the support structure can occur, especially at riveted and bolted locations.

Two methods of approach can be taken to provide satisfactory operation of airborne electronic equipment subjected to the acoustic noise environment. Control of the sound pressure level at the equipment can be accomplished by sound absorption or sound isolation techniques. The use of mechanical and electrical components not sensitive to this environment provides a second approach.

Transistors, magnetic amplifiers, and ceramic tubes are examples of components that can be subjected to high sound-energy levels before malfunction occurs. Improvements in fabrication techniques of the support chassis for electrical components will result in increasing the resistance to acoustic noise. For example, the use of material bonding techniques in place of riveted and bolted assemblies appears to be a satisfactory method of fabrication.

When it is not possible to substitute rugged components for more sensitive components for either electrical or mechanical reasons, it is desirable to group the sensitive components together. The area housing these components can then be partially protected from the acoustic noise by various sound absorption and sound insulation techniques. Sound-absorbing materials are light and usually porous. Sound insulation materials are massive and therefore impervious. Either or both techniques can be efficiently used depending upon restrictions of weight, volume, form factor, etc. for a particular design problem.

13-8 ACCELERATION

When an aircraft deviates from constant velocity flight, the equipment on board is necessarily subjected to complex dynamic forces. When the principal frequencies of these forces are low enough to be considered steady, the environment is discussed in terms of *acceleration*. Such forces are experienced notably in controlled maneuvers. Launching, landing, and air turbulence will also induce dynamic forces, some of which may have components of frequency low enough to be treated as acceleration, although shock and vibration frequencies are more to be expected from these flight conditions.

Since the load induced by an acceleration is proportional to its magnitude it is customary to measure loads in terms of acceleration. The ratio of the magnitude of acceleration at a point to the magnitude of the acceleration of gravity is called the *load factor*. The weights of all components and this resultant combined with the weights are multiplied by this load factor to obtain the net load. Load factors are often given for the center of gravity of an aircraft. Since an aircraft is of finite size and is effectively a rigid body at the frequencies under consideration here, the location of the equipment must be considered before applying a load factor. It will be recalled from the mechanics of rigid bodies that the acceleration of a point on such a body is a function of the velocity and acceleration of the body, in all six of its degrees of freedom, and of the position of the point on the body.

Acceleration loads, when considered alone, do not present, for most components of a radar, so severe a requirement as shock and vibration loads, although acceleration is a difficult environment for certain instruments, such as gyroscopes, which are auxiliary to radars. Because of the comparatively long duration of an acceleration load, it may in some cases affect the design of a major structural member.

Steady acceleration probably has the greatest effect on packaging design when considered in combination with vibration or shock. Accelerations of an aircraft tend to increase the static deflection of installed vibration isolators and may even cause full deflection, or *snubbing*. If this occurs, the

mounted equipment is no longer effectively isolated from the vibration or shocks transmitted through the airframe structure. It has been seen that vibration isolators are most effective when they have large static deflections and concomitant low natural frequencies. To avoid snubbing during acceleration, it is necessary either to sacrifice space by providing for these large isolator deflections or to sacrifice isolation effectiveness by using stiffer mounts with smaller static deflections. If the duration of a high load factor is expected to be short compared with the duration of severe vibration, it may be best to permit snubbing under certain circumstances in order to obtain the best isolation during most of the time. On the other hand, the effect of isolators may be so badly degraded by this type of environment that a more reliable radar will be obtained with no isolators whatsoever. Careful consideration of the expected environment must precede the choice of a specific design.

13-9 MOISTURE

Moisture is classified as a natural environment and can exist whenever humidity, fog, ice, snow, sleet, rain, dew, hail, or salt spray are present or when a part is immersed in water. The principal problems resulting from the presence of moisture are corrosion, reduction in the effectiveness of insulation, and the possible growth of fungus.

Moisture generates two major types of corrosion, chemical and electrochemical. Corrosion caused by the combined action of the atmosphere and moisture on the surface of a body is termed *chemical corrosion*. Corrosion between two different metals existing in a conductive medium is known as *electrochemical corrosion*. The corrosion of the metals used in electrical systems results in a decrease in structural strength of individual components and assemblies, as well as the sticking and binding of moving parts.

Moisture on electrical parts will affect the insulation resistance. If the insulation material has a porosity that will allow absorption of the water, the resistance of the insulation will be decreased and a power loss will result. Insulation that allows a complete water film to form on its surface is undesirable. If the insulation surface is such that water collects in drops and has a tendency to run off, the effect on the insulation resistance is small.

Protecting electrical equipment from the effects of the moisture environment can be accomplished by control of design techniques, material selection and surface finishes. Although the exclusion of moisture from parts and equipment is elementary in theory, it is surprisingly difficult in practice and has required the development of many special techniques. Humidity control by mechanical or chemical means may be used on hermetically sealed containers, free-breathing containers, or containers with controlled venting. Other techniques in moisture control include the

use of vapor-phase inhibitor wrap, plastic film barriers, and dry nitrogen atmosphere.

Chemical corrosion can be prevented by the application of protective coatings of certain metals or organic finishes over the base metal susceptible to the corrosive effects of the atmosphere. The three most common forms of surface protection are anodizing processes, chemical film treatment, and electroplating. Anodizing consists of the formation of an aluminum oxide layer on the surface of aluminum alloys by an electrochemical treatment. Formation of a chromate film on an aluminum surface is an example of chemical film treatment. However, this method of surface protection does not offer protection from electrochemical corrosion. The electroplating method of surface protection is employed when it is necessary to provide a good electrical current path or when soldering is necessary. Surface protection for magnesium parts requires methods similar to those used for aluminum alloys. For steel parts, zinc or cadmium plating is used. Copper is usually protected by nickel or silver plating.

Electrochemical corrosion can be avoided by preventing contact between dissimilar metals. If it is necessary to use different metals in contact with each other, the following design principles should be followed. At the junction, one of the metals can be plated with a film which is similar to the other adjacent metal. This, in effect, will decrease the electrolytic potential differences. The plating of cadmium or zinc on steel in contact with a material such as aluminum will alleviate the damaging effects of electrolytic corrosion. An inert material as a gasket or washer will act as a mechanical insulator. The use of organic coatings or a zinc chromate primer coat on the contact faces of the metals is satisfactory for most metals except magnesium. The application of a coat of clear lacquer over the joint after assembly is necessary when using an organic coat. The application of corrosion inhibitors to the faces of each of the metals and the sealing of all joints containing dissimilar metals are other forms of protection. Perhaps the best course to follow when it is necessary to have dissimilar metals in contact is to assemble the piece parts in such a way that the larger piece is the anodic (corroded area) whereas the smaller part will be cathodic or protected. This will allow a higher degree of structural safety beyond the above-mentioned protection methods employed in the design.

The reduction in the damaging effects from corrosion can be accomplished by careful design and use of proper materials and finishes. For example, electrical insulators constructed from ceramic material should have glazed surfaces or the surfaces should be sealed from moisture. Terminal strips and boards should be fabricated from material with low moisture absorption. Techniques employing the principles of impregnating and potting will provide a measure of protection from the effects of moisture and associated

environments. The elimination of moisture traps, wells, and pockets from the design is also desirable.

Fungus is closely associated with moisture. To prevent fungus from damaging equipment, parts should be fabricated from materials which are not nutrients for fungus — metals, ceramics, mica, teflon, nylon, rubber, etc. Materials not resistant to fungus are cotton, linen, wood, leather, cork, etc. These can be used when protected from a fungus environment — for example, in hermetically sealed assemblies.

13-10 STATIC ELECTRICITY AND EXPLOSION

Static electricity is defined as electricity at rest — electric charges — as compared with *dynamic electricity*, which is electricity in motion — electric current. An electrically charged environment creates electrical noise which may cause echo loss by the radar and malfunction of the radio or other electronic equipment. Static electricity may arise from local thunderstorms, which generate intermittent electrical impulses of high intensity. These can damage the external surface of the radome. Distant thunderstorms can generate continuous noise levels and cause radio interference. A voltage differential incident to static electricity can be produced by rubbing together dissimilar surfaces. Thus, during its flight an aircraft may build up on its skin static charges from the action of snow, hail, dust and sand, ice crystals, or smoke or exhaust particles from the power plant. Very high voltages may be so generated. The charge leaks off in a corona discharge which causes electromagnetic noise that may obscure the radar echoes. The effect of static is less severe on equipment operating at the high frequencies; it is increasingly critical at lower frequencies of operation.

The adverse effects of static electricity may be offset to some extent by the choice of bandwidth of the radar and by the design of the airframe. In particular, it is helpful to insulate or to round off sharp points of the airframe to prevent discharge from them. For low-speed aircraft, trailing wicks give a relatively noiseless discharge.

Ground operations, especially during fueling, can result in explosion generated by static electricity. Proper grounding of the system will prevent this. During flight, explosions can be caused by electrical equipment capable of generating a spark — switches, motors, relays, etc. Hermetic sealing or the use of pressurized containers will reduce the explosion hazard for electrical components which in their normal operation could ignite an explosive gaseous mixture.

Proper protection from static electricity is necessary for the safety of maintenance personnel as well as for successful completion of missions.

13-11 PRESSURE

The operation of electrical equipment located in nonpressurized areas of airplanes and missiles may be adversely affected by the change in the ambient atmospheric pressure induced by changes in the flight altitude. One of the most serious effects of the low pressure existing at high altitudes is the reduction in dielectric strength of air. This reduction in the strength of the dielectric medium can cause a corona discharge. Corona discharge occurs when the dielectric strength of the air is exceeded by the potential gradient created by a conductor in a high-voltage circuit — for example, a power supply unit. It can be detected visually by the purplish discharge around or between the conductors or by the radio noise level that is produced.

Low ambient pressure conditions also allow a condition known as *arc-over* to take place between switch contact points, breaking points, bare terminals, uninsulated wire, etc. Erratic operation or the partial or complete breakdown of the equipment may result. Increase in the wear rate of the contact points will occur because of the increase of intensity of the arc. To eliminate induced radar noise, decrease in the insulating ability of material, and reduction in the useful life of equipment, the designer of electrical equipment must include this environment in the basic design criteria for the equipment.

Pressurization of high-voltage equipment is a positive method of combating the adverse effects of low-pressure conditions. However, this method for elimination of this particular electrical problem complicates the overall mechanical problem. For example, pressurized containers, in order to withstand required differential pressures, must have a certain degree of structural strength and stiffness, thus adding a weight penalty to the overall system weight. Sealing problems along with the possible use of a pressure pump to hold a fixed pressure level within the container add complexity to the problem. The cooling system is made more complex since the heat-dissipating electrical units cannot be cooled directly by the ambient air. Instead, an air-to-air heat exchanger must be used in order to remove the heat from the recirculated air within the pressurized container.

For nonpressurized units, the use of proper detail design techniques will allow the equipment to operate satisfactorily in low-density air. Some recommended methods follow. Since the breakdown voltage between electrical conductors is a function of the distance between conductors, the correct spacing between conductors will materially aid in the solution to the problem. For some voltage levels, the required distance between conductors may be greater than the available space. In this case, replacing the air by some other medium, gas or liquid, with a higher dielectric strength is a feasible solution to the problem. This technique, however,

introduces disadvantages similar to those of a pressurized system. The placing of a grounded shield between the conductors and the careful use of insulation techniques between terminals is sometimes of advantage. Smooth surfaces and rounded edges on conductors will help to eliminate any tendency toward excessive voltage gradients.

Protection from the damaging effects of this environment can be met satisfactorily by known conventional methods. The challenge lies in evaluating each of the methods for its effect on an overall system with regard to the electrical, mechanical, structural, and thermal requirements.

13-12 MAINTENANCE AND INSTALLATION

Another major influence upon packaging design is the necessity of providing for efficient maintenance and easy installation. Although the designers make every effort to promote reliability, it is evident that maintenance requirements cannot be altogether eliminated.

Maintenance, the practice of keeping equipment in a state of full efficiency, generally may be divided into *preventive maintenance* and *repair*. Both are dependent upon checking and testing. It is important that maintenance interfere with aircraft availability as little as possible, and that it require a minimum of work space and manpower. It is also desirable that the required inventory of parts and tools be as small as possible. Good packaging promotes these objectives mainly by minimizing the effort and skill required to gain access to all parts of the radar.

Application of several principles can assist in reducing the maintenance problem. The approach to all subassemblies and components should be reasonably direct. It should be possible to remove any major unit of the radar without first removing adjacent units. Within any subassembly (except throwaway units) it should be possible to replace any component without disturbing others. Subassemblies should be light enough to be handled safely by one man; their removal and replacement should require no more than two hands. Alignments should be positive: shimming and other difficult adjustments are undesirable. Test points for checking should be visible and accessible.

Although it is advantageous for mounting fasteners to be easily removable, it must be remembered that the primary function of fasteners is to fasten securely in spite of all environmental hazards. Reliability in fasteners must never be sacrificed to obtain quick-release features.

The use of *plug-in subassemblies* is an important packaging technique and has several good features for maintenance and installation. In this type of design, the radar is built up of a number of comparatively small assemblies. The interconnecting wiring of such an assembly is brought out from a multicontact connector which is rigidly mounted to the chassis in a fashion

permitting the subassembly to plug into a mating connector on a major assembly. Fasteners and guide pins protect the electrical connector from mechanical loads. If a plug-in subassembly fails, it can be replaced with a spare quickly and easily. Fig. 13-12 illustrates a computer system designed to use plug-in units extensively.

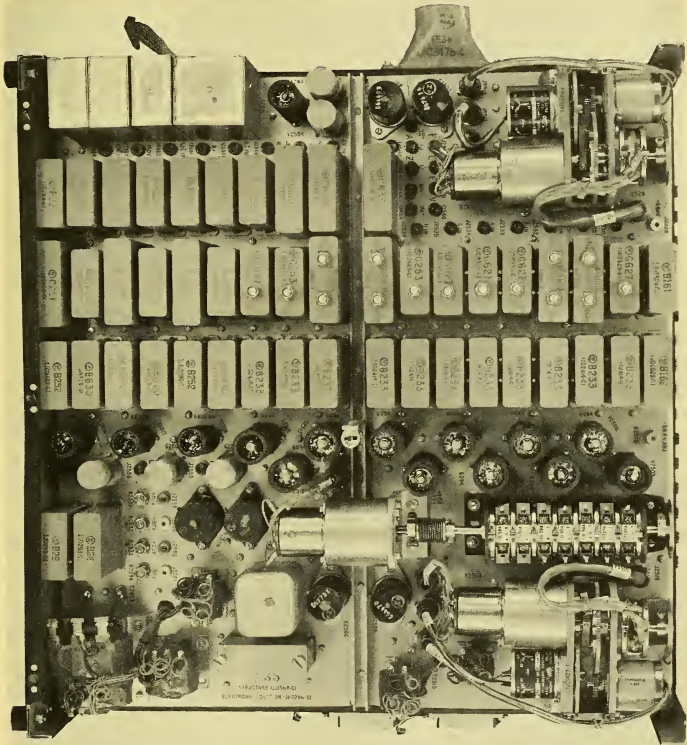


FIG. 13-12 Airborne Computer, Illustrating the Use of Plug-in Units.

The size and complexity of plug-in subassemblies may be governed in part by maintenance considerations. When a failure occurs on a radar, it is usually rather difficult to locate the particular *component* or components at fault, regardless of accessibility. On the other hand, it is comparatively

easy, in most cases, to isolate the trouble in a particular *function*. This suggests that each plug-in subassembly should provide a complete function. The breakdown of a major assembly into subassemblies can be made in the same manner that a block diagram is drawn to illustrate the functions within a major assembly. Each block corresponds to a plug-in subassembly. When trouble is isolated in a certain function, the corresponding plug-in subassembly can be replaced.

Plug-in assemblies may have a *modular design*. That is, the general design of the various subassemblies is similar, and the linear exterior dimensions are integral multiples of some lowest common denominator, such as the smallest plug-in unit. Often two linear dimensions are the same for all units of a modular design, the third dimension being the only one to vary. The use of modular design simplifies and accelerates the actual design effort as well as the manufacturing processes. Modular design also lends itself well to automatic manufacturing.

Plug-in subassemblies can be repairable, or they can be designed for discard after failure. The relative advantages of these two design philosophies depend upon the type of service expected, duration of service life, logistics, and maintenance facilities.

13-13 TRANSPORTATION AND SUPPLY

During transportation of airborne electronic equipment, severe environmental conditions may occur that may result in damage to the structural and electrical component parts of the system. The most serious environment encountered during transportation is the dynamic loading induced by vibration and shock energy input. The magnitudes of the loads depend upon the mode of transportation — airplane, railroad, truck, or ship. A good method for protecting equipment from the damaging effects of transportation is to provide shipping containers incorporating shock or vibration isolators. The equipment should have lifting lugs clearly labeled to help prevent damage during handling.

In maintaining an adequate supply of radar systems and parts, problems are encountered in the prevention of deterioration in storage. Moisture, dust, and temperature are among the principal items responsible for this deterioration. Although the equipment has presumably been designed to operate in spite of these environmental factors, it is good practice to protect the equipment as much as possible by providing clean dry storage areas which are not subject to extremes of temperature. Since the shipping containers should, in any event, be designed to protect the equipment from the transportation environment, they can, with little or no additional expense, be made suitable to act as storage containers.

13-14 POTENTIAL GROWTH

The overall design criteria for an airborne radar must take into account its future growth possibilities. Rapidly changing electrical requirements and environmental conditions tend to make equipment obsolete before it reaches the final design stage. Therefore, equipment must be made flexible in order to receive the maximum return for the engineering and manufacturing effort and cost. The limits on the potential growth of a system are directly dependent upon the flexibility designed and built into its component units. A flexible system will be able to absorb the latest types of electrical components and circuit designs. Advanced techniques in structural and mechanical design, fabrication methods, and new materials can be utilized in the system any time during its life.

Plug-in subassemblies add to flexibility, especially if they are modular in design, because one plug-in unit can be easily replaced with an improved model. High flexibility is difficult for systems that include items that cannot be easily divided into small units. Thus, major packaging problems exist with radar antennas, high-voltage equipment requiring large power supplies, modulators and transmitters with critical interconnecting wiring, and mechanical linkages between units.

In attaining a high degree of flexibility in a system, it is important that environmental considerations are not ignored. That is, the technique employed in the mounting of components within a specific package should provide for maximum structural strength and stiffness to carry the imposed dynamic loads, allow proper cooling, exclude moisture and dust, and so on. It is often a successful technique to utilize the covers of the packaged units as structural members in order to increase the structural natural frequency of the unit, thus eliminating, in some cases, the need for vibration isolators. The use of printed circuits, plug-in units, and potted or hermetically sealed units should be considered. For major electronic assemblies, proper location of mounting or installation lugs in order to carry acceleration loads should be included in the packaging design criteria.

The design of electrical equipment that depends very closely on particular performance characteristics of an airplane or missile cannot be considered to possess maximum potential growth capabilities as far as interchangeability of equipment between aircraft is concerned. However, this equipment can still be made flexible as far as its relationship to spare or improved component part replacement and maintenance is concerned.

13-15 RELIABILITY

A major problem facing the designers of any military weapons system is ensuring that the equipment will be reliable in field operation. *Reliability* is a system or equipment trait (or capability) which has been defined as *the*

probability that a system will give satisfactory performance for a given period of time when used in the manner and for the purpose intended.

From this definition, it is clear why reliability is a system parameter of paramount importance. The success or failure of a military mission turns on the ability of equipments to carry out assigned functions. Regardless of the degree of sophistication of the equipments, or the high performance of which they are theoretically capable, their potential value to a military commander is strictly a function of the probabilities that they will operate satisfactorily when called on.

In this section we shall consider various aspects of the reliability problem of which the system designer must be aware in his designs. Specifically, we shall be concerned with the influence of reliability on system planning, the factors which affect reliability of electronic equipment, how to attain satisfactory reliability, how reliability is measured, and how it may be quantitatively predicted.

Reliability and System Planning. Reliability is a basic system performance parameter. As shown in Paragraph 2-23 the specification of reliability forms a part of the system requirements as dictated by the tactical mission requirements. This material also illustrated the more-or-less direct relationship between reliability and the attainable level of mission accomplishment.

The derived reliability requirement represented an estimate of a reliability goal that would — in combination with the goals set for other aspects of system operation (detection, vectoring, etc.) — allow the system to fulfill its overall mission objectives.

With requirements generated, we are ready to examine the specific factors which influence the attainment of the reliability objective.

Factors Which Influence Reliability. The final reliability of an electronic equipment is a function of many interacting and only incompletely understood factors. Some of these factors are listed here.

1. The detail design and fabrication of the electrical and mechanical components used in the system.
2. The application of the individual components in the equipment.
3. The methods of manufacture and assembly of the equipment, including quality control and inspection techniques.
4. The simplicity of the design for field maintenance as well as the maintenance facilities available.
5. The abilities and skills of the military personnel who operate and maintain the equipments.
6. The physical environmental conditions operating on the equipment, such as temperature, shock, altitude, etc.

Each of these factors must be considered, controlled, and allowed for if the required degree of reliability is to be attained. It is obvious that regardless of how thoroughly all other factors are controlled, the neglect of any one will seriously degrade the final reliability.

Attaining High Reliability. Most manufacturers of military electronic equipment have reliability organizations whose sole responsibility is that of ensuring that the equipment they build is sufficiently reliable to meet its requirements. These organization usually serve the design engineers in a staff capacity, by providing them with information they can use in the design process. Specifically, they provide component lists and design handbooks which contain preferred components and circuits and required safety factors. The information in the handbooks is usually based upon data obtained on older equipments already used in the field. A feedback loop is thus generated between the older equipments in the field and the design of new equipments, with the result that past errors in design or component choice are not repeated.

Experience has shown that the basic design of the equipment determines the maximum reliability the equipment can possibly attain. For this reason, reliability must be designed into equipment; it cannot be provided by superior manufacturing and testing techniques alone. Many organizations have instituted a procedure known as *saturation designing*, in which the proper designs of equipments are very thoroughly and critically reviewed, in the light of past experience.

A major area of reliability attainment efforts, as presently conducted, is that of reliability testing. Such testing is normally carried out during the development of the equipment and is designed to determine the equipment's capability to meet its intended environment. This of course requires a highly specific definition of the environmental stresses, so that they may be simulated in the laboratory. Test-to-failure programs, designed to test the limits of the equipment endurance capabilities, are also widely used. Perhaps the basic purpose of all reliability testing is to determine the equipment's weak spots so that they may be remedied before the equipment is placed in large-scale production and use.

Reliability Measurement. The numerical reliability of an existing electronic equipment can be determined by actually operating the equipment in a realistic operational environment for sufficient time to attain statistically valid data. The number of hours the equipment is operated, as well as the number of failures it suffers, are recorded and used in the calculation of the average number of operating hours between successive failures, commonly referred to as *mean time to failure*. This empirically determined number can then be used to determine the probability of successful equipment operation for any specified length of time. Before

presenting the mathematics involved in this determination, it is first necessary to discuss several of the basic concepts of reliability theory.

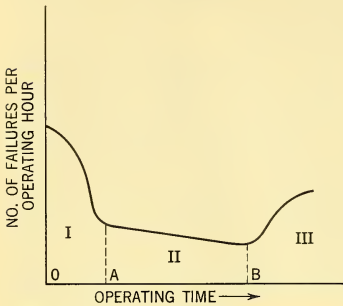


FIG. 13-13 The Failure Rate Curve.

during shipment and installation, and perhaps initial debugging of the equipment. This is analogous to the infant mortality phenomenon in humans. Obviously, the time between 0 and A is short relative to the equipment's entire life.

In Stage II (bounded by A and B), most failures occur randomly in time and at a relatively constant rate. These random failures are a result of accidental, unexpected, and unusually severe conditions arising during the operating period, as well as randomly occurring failures of parts. During this period, the probability of the accidental failure of a given equipment in any time interval is entirely independent of the time the equipment has already operated. For the most part, reliability is measured during this period and, as will be shown later, the prediction of future reliability is based upon the existence of a long period of the equipment's life in which the failure rate is constant.

The final hump in the curve beyond B represents increased wear-out failures which occur due to a large number of component parts gradually reaching a more or less failed state as a result of deterioration of the physical structure of the parts.

If a large sample of identical electronic equipments were to be operated in region II until all the equipments failed, a curve similar to the observed curve in Fig. 13-14 could be generated. Superimposed over the observed curve is a theoretical exponential decay curve with the same time constant as the observed curve. It can be seen that the two curves are so clearly comparable that it is reasonable to hypothesize that the differences between them are caused by chance factors. In most cases, of course, the degree of similarity between the observed and a theoretical reliability function are

Electronic equipment has been found to exhibit a characteristic failure behavior with time. This behavior is represented schematically in Fig. 13-13. As can be seen from the curve, the failure behavior can be divided into three stages. Each stage can be represented by a distinct mathematical model, as well as by a specific time period in the operating life of the equipment. Stage I is characterized by initial failures incident to defective component parts finding their way into the equipment, rough handling

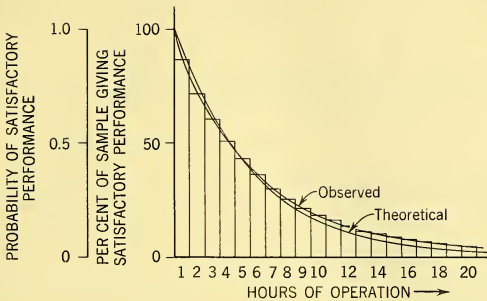


FIG. 13-14 Representative Observed and Theoretical Reliability Functions.

not as striking as in Fig. 13-14; however, the overall hypothesis of similarity has been found to be a good one. Thus the familiar exponential reliability law may be written:

$$R = e^{-t/T} \quad (13-16)$$

where R = reliability

t = time during which satisfactory operation is desired

T = equipment mean time to failure.

The quantity r (hourly failure rate) is the reciprocal of T , giving rise to the alternate form of the law:

$$R = e^{-rt}. \quad (13-17)$$

Let us return now to the measurement of reliability discussed earlier. Field testing of equipment will yield the value of T needed for Equation 13-16. This is numerically equal to the sum of the operating hours divided by the number of failures. It is important to realize that the soundness of this procedure rests on the assumption that a unit once repaired is equivalent to a new unit. It is further essential to use a sufficient number of representative samples to ensure generality of the results. Using Equation 13-16, a theoretical curve as shown in Fig. 13-14 can be generated by substituting a continuum of t values.

As an example of how this law might be used, let us consider the hypothetical AI radar problem treated in Chapter 2. The derived requirement stated that a reliability of 90 per cent was required over a 3-hour operating period.

Thus, from Equation 13-17

$$0.9 = e^{-3r}$$

or

$$r = 0.035 \text{ (system failures per hour)}$$

$$T = \frac{1}{r} = 28.5 \text{ hours (mean time-to-failure).}$$

Reliability Prediction. In weapons system analysis it is often necessary to effect a reliability prediction on equipment which has not yet been built. Reliability predictions are most useful in comparing two or more alternative systems under consideration, because of the limitations of present prediction techniques; however, they are also of some value as an indication of the reliability any given system can be expected to attain. Reliability predictions are useful as a basis for contractual negotiations of reliability goals or specifications. Further, a prediction will often point out the portions of the equipment which will require special, concentrated reliability efforts. However, in order for a prediction to have any value at all it must be arrived at in a sound and meaningful manner.

In many areas of science and engineering, an object's future behavior can be predicted with great accuracy. Unfortunately, this is not yet the case in predicting the reliability of electronic equipment. Reliability engineering is still a young science and its prediction techniques are, for the most part, still unproven. Coupled with this uncertainty of prediction techniques is the generally poor quality of the past failure data on which predictions must be based. This deficiency in the accuracy and completeness of available data can be seen to be of overriding importance, when it is considered that the only manner in which it is possible to predict future behavior is by utilizing past data and experience.

The *product rule* prediction technique is the one most widely used in reliability work today. Basically, the product rule states that the probability of a system operating satisfactorily is equal to the product of the operating probabilities of all the independent components that go to make up the system. This probability rule holds only when the performance of every component is entirely independent of the performance of every other component. This condition is very rarely encountered in real life since few electrical components operate independently of their associated components; however, the product rule does give a good first approximation of the equipment's potential reliability. Further, the product rule also assumes that the failure of any single component will result in a complete system failure. Since this is not completely true, the product rule tends to give a pessimistic estimate of reliability.

Mathematically, the multiplication of component reliabilities to obtain system reliability can be represented as

$$R_{system} = R_1 \cdot R_2 \cdot R_3 \cdot \dots \cdot R_n. \quad (13-18)$$

Now, if it is further assumed that each component in the system fails exponentially, we get from Equations 13-16 and 13-18

$$R_{system} = e^{-t/T_1} \cdot e^{-t/T_2} \cdot e^{-t/T_3} \cdot \dots \cdot e^{-t/T_n} \quad (13-19)$$

where T_n is the mean life of the n th component in the system. The mean life of the system is then given by

$$\frac{1}{T_{system}} = \frac{1}{T_1} + \frac{1}{T_2} + \frac{1}{T_3} + \dots + \frac{1}{T_n}. \quad (13-20)$$

Since $1/T$ is equal to r , the hourly failure rate, we can rewrite Equation 13-20 as

$$r_{system} = r_1 + r_2 + r_3 + \dots + r_n. \quad (13-21)$$

The actual process of predicting the reliability of electronic equipment with as many as 5000 individual electrical and mechanical component parts would be impractical if it were necessary to multiply together 5000 individual failure rates. The procedure generally used is to group all similar components into a series of ten or twelve classes — capacitors, resistors, receiving tubes, etc. Each class of components has an associated failure rate per hour, based on past history for that component class; for example, vacuum tubes in airborne equipment have shown a failure rate of 200×10^{-6} per hour. Multiplying the failure rate for any given component class by the number of components in that class will give the failure contribution of the class. Performing this for all the classes and then summing the failure contributions will give the overall system predicted failure rate. Mathematically, this can be represented as

$$r_{system} = X_1 r_1 + X_2 r_2 + X_3 r_3 + \dots + X_n r_n \quad (13-22)$$

where X is the number of individual components in a class and r is the class failure rate. Substituting the calculated system failure rate in Equation 13-17 will allow the generation of a reliability prediction.

As an example, let us consider the AI radar and fire-control system already discussed. Such systems typically have something of the order of 200 vacuum tubes. Thus the expected failure rate from vacuum tubes alone would be

$$(200)(200 \times 10^{-6}) = 0.04 \text{ failures per hour.}$$

This failure rate exceeds the allowable system failure rate (from all causes) previously derived as 0.035. Thus, it may be expected that consid-

erable difficulty will be experienced in meeting the system reliability requirement.

Several possible corrective approaches exist. First, the radar design may be simplified by sacrificing features that seem desirable but are not absolutely essential to overall mission accomplishment. This is always a worthwhile objective; a reliability deficiency provides a sharp spur to its eventual achievement.

A second approach is to examine the possible use of more reliable components such as diodes and transistors to replace vacuum tubes.

A third approach is to provide redundancy, i.e. to duplicate portions of the system where a failure is more likely.

An approach of a somewhat different character is to examine the original requirement to determine whether it represents the only possible answer to the tactical problem. For example, the three-hour operating requirement of the hypothetical AI radar and fire-control system was based on the mission time of the CAP aircraft. Actually, only 12 of the CAP aircraft are aloft at any given time, and only 6 of these are used to engage the specified raid. The remaining 36 aircraft used to combat the specified attack being deck-launched have a required operating time of less than one-half hour. Eighteen aircraft are unavailable for combat because of previous failures which have not been corrected.

Let us assume a system failure rate of 0.1 per hour and one-half hour of effective operating time for the equipment in the deck-ready interceptors. The expected number of failures that would occur among the 42 aircraft used to engage the raid would be

$$(36)(0.1)(0.50) + (6)(0.1)(3) = 3.6 \doteq 4 \text{ failures}$$

This represents slightly less than 10 per cent of the total number. Thus the 90 per cent reliability requirement can be met by a system with a 10-hour mean time-to-failure provided that this increased failure rate does not overload maintenance facilities to the point where aircraft availability is reduced below the assumed value of 48 out of 66.

As this analysis shows — and as has been pointed out in several previous cases — the original radar requirements are not sacred. They merely represent a first approach to the translation of tactical requirements into technical requirements. As more becomes known about the problem and the factors that limit the solution, it is often necessary to modify this first approach through a more intensive appraisal of the original tactical problem. This is the essence of the system approach to radar design.

BIBLIOGRAPHY FOR CHAPTER 13

MILITARY PUBLICATIONS AND SPECIFICATIONS

(latest issue should be consulted)

1. ANA Bulletin No. 421 Air Force-Navy Aeronautical Bulletin, *Atmospheric Properties — Extreme Cold and Hot; Standard for Aeronautical Design.*
2. U.S. Air Force Specification *Environmental Criteria for Guided Missile*
Bulletin No. 106 *Systems*
3. MIL-STD-202 *Military Standard, Test Methods for Electronic*
and Electrical Component Parts
4. MIL-E-5400 (ASG) *General Specification, Electronic Equipment*
(Aircraft)
5. MIL-T-5422 (ASG) *Aircraft Electronic Equipment, Testing En-*
vironmental
6. MIL-E-5272 *General Specification for Environmental Test-*
ing, Aeronautical and Associated Equipment
7. MIL-E-8189 *General Specification for Electronic Equipment,*
Guided Missiles
8. NavShips 900,190 *Guide Manual of Cooling Methods for Electronic*
Equipment
9. RD 219/3 *Fundamentals of Guided Missile Packaging*
Department of Defense
10. NavShips 250-660-30 *A Guide for Design of Shock Resistant Naval*
Equipment, BuShips (1949), Washington, D.C.

ARTICLES AND REPORTS

1. Bassler, S. G., "Electronic Circuit Packaging for Missile Applications" *Electrical Mfg.*, Vol. 61, No. 3 (March 1958).
2. Blake, R. E., and E. S. Swick, *Dynamics of Linear Elastic Structures*, NRL Report No. 4420.
3. Booth, Galt B., "Introduction to Random Motion: The Natural Environment and its Simulation," *Shock and Vibration Bulletin*, No. 23, June 1956, Department of Defense, Washington, D. C.
4. Crede, Charles E., "Vibration and Shock Isolators," *Machine Design*, August 1954.
5. Forkois, Harold, and Kenneth Woodward, *Design of Shock and Vibration Resistant Electronic Equipment for Shipboard Use*, NRL Report No. 4789.
6. Goff, John A., and Serge Gratch, "Thermodynamic Properties of Moist Air," *Am. Soc. Heating, Ventilation Eng. Trans.*, 51, 125 (1945).
7. Goff, John A., and Serge Gratch, "Low Pressure Properties of Water in the Range minus 160° to 212° F," *Heating, Piping and Air Conditioning*, February 1946, ASHVE Journal section.
8. Harwood, Robert R., "Repairable Plug-In Subassemblies," *Elec. Mfg.* 56, 124-127 (1955).
9. Jacobson, R. H., "The Scream of a Jet in a Six Inch Test Cell," Armour Research Foundation, Chicago, Illinois.
10. Kaufman, Joseph, "A Re-evaluation of Vibration Testing Techniques," *Elec. Mfg.* 56, 132 (1955).

BOOKS

1. Bonney, E. A., M. J. Zucrow, and C. W. Besserer, *Aerodynamics, Propulsion, Structure, and Design Practice*, D. Van Nostrand Co., Inc., Princeton, N.J., 1956.
2. Brown, Aubrey I., and Salvatore M. Marco, *Introduction to Heat Transfer*, McGraw-Hill Book Co., Inc., New York, 1942.
3. Crede, Charles E., *Vibration and Shock Isolation*, John Wiley & Sons, Inc., New York, 1951.
4. Keenan, J. H., and F. G. Keyes, *Thermodynamic Properties of Steam*, John Wiley & Sons, Inc., New York, 1936.
5. Lessels, John M., *Strength and Resistance of Metals*, John Wiley & Sons, Inc., New York, 1954.
6. Locke, Arthur S., et al., *Guidance*, D. Van Nostrand Co., Inc., Princeton, N.J., 1955.
7. Marks, Lionel S., *Mechanical Engineer's Handbook*, McGraw-Hill Book Co., Inc., New York, 1951.
8. Seely, Fred B. and Newton E. Ensign, *Analytical Mechanics for Engineers*, John Wiley & Sons, Inc., New York, 1941.
9. *Heating, Ventilating and Air Conditioning Guide*, American Society of Heating and Ventilating Engineers, New York, 1951.

CHAPTER 14

AIRBORNE NAVIGATION AND GROUND SURVEILLANCE RADAR SYSTEMS*

The preceding pages have concentrated on airborne radar systems which are employed for the detection, acquisition, and tracking of airborne targets (AI and AEW radars). For such systems, the operation of the radars had to be made as independent as possible of the degrading effects of weather, ground return, and sea return.

An equally important group of radar systems is composed of systems for which the radar returns from the ground, sea, and meteorological phenomena constitute the *desired* intelligence. Such systems are employed for navigation, bombing, and reconnaissance, as was briefly described in Chapter 1.

While the same basic laws of nature apply to these systems, there are sufficient differences in their importance relative to tactical requirements — as reflected by the design practices necessary to achieve the navigation and mapping function — to warrant a separate discussion of these radar system types. This chapter will discuss the essential features of the airborne radar systems designed to provide navigation and ground mapping intelligence. Particular emphasis will be given to the interrelations among radar techniques and the tactical functions. This information will provide the reader with the background knowledge necessary to derive the technical specifications for such systems, as was done for the AI and AEW systems in Chapter 2. The primary technical problems and limiting considerations of each type of system will be outlined. Examples of operational installations will be presented to acquaint the reader with the state of the art.

The discussion will cover the following radar system types:

1. Doppler navigation systems
2. Weather radar systems

*Paragraphs 14-1 through 14-8 are by W. R. Fried. Paragraphs 14-9 through 14-14 are by B. L. Cordry. Paragraphs 14-15 through 14-21 are by R. H. Laprade. Paragraphs 14-22 through 14-26 are by F. Stauffer.

3. Active ground mapping systems
 - (a) Forward look
 - (b) Side look
4. Passive ground mapping systems
 - (a) Radiometric
 - (b) Infrared

14-1 INTRODUCTION TO DOPPLER NAVIGATION SYSTEMS

Doppler radars may be designed for measuring an airborne vehicle's velocity with respect to the earth's surface. One of the primary applications of the measured velocity is its use in a self-contained navigation system in which the velocity is integrated to obtain distance traveled and combined with heading information to obtain the vehicle's present position in earth coordinates — for instance, as latitude and longitude or as deviations from a desired course. Present position coordinates may then be compared with desired destination coordinates, and the distance and the required course to the destination computed and furnished to the pilot or autopilot as a vehicle steering signal. Such a system indirectly measures the effects of unknown winds, drift errors, etc., and thus achieves an improvement in navigation.

In such an application the doppler radar becomes one of the components of a self-contained navigation system. A typical system consists of three basic components: (1) a *velocity sensor*, for instance a doppler radar, (2) a *reference* which relates the information to the earth, such as a magnetic compass, astrocompass, or gyrocompass, and (3) a *computer*, which operates on the information obtained from the velocity sensor and the reference (which is really a form of sensor also) and computes the desired output information. It is important to recognize that in such a system the overall accuracy performance is determined by the accuracies of all three of the components, and is essentially limited by the accuracy of the weakest of the three.

A basic self-contained navigation system is shown in block diagram form in Fig. 14-1. A system of this type can be considered a dead reckoning navigation system since it is based on navigating from a known point and

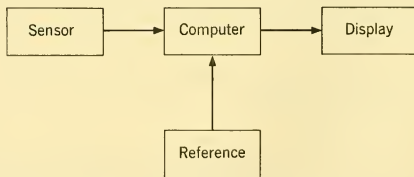


FIG. 14-1 Basic Self-Contained Navigation System.

computing the track being made good and a new position by means of airborne data, namely, ground velocity and aircraft heading. It is significant that these data are obtained without the need for visual access to the ground or the use of radio or radar stations located on the ground; hence this method has been given the name *self-contained navigation*.

One example of such a self-contained dead reckoning navigation system is the *true air speed system* shown in Fig. 14-2. Here the basic velocity

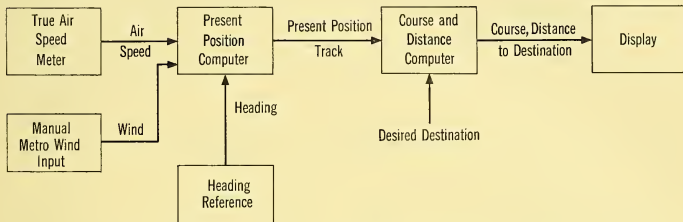


FIG. 14-2 Typical True Air Speed Dead Reckoning System.

information is composed of air speed from a true air speed meter and wind vector (magnitude and direction) information. The latter is usually obtained from a meteorological source and inserted manually. The heading reference is most frequently a magnetic compass. The weakest links in this system are probably the air speed sensor and the poorly known wind information, so that accuracies of this type of system have been quite limited. The system obviously does not employ a radar.

A doppler radar is a highly accurate sensor of aircraft ground velocity (ground speed and drift angle). It measures actual ground velocity, that is, the velocity of the aircraft with respect to the ground (not the air mass); hence all wind motion is included in this measurement and external wind data are no longer needed. Substitution of doppler radar for the true air

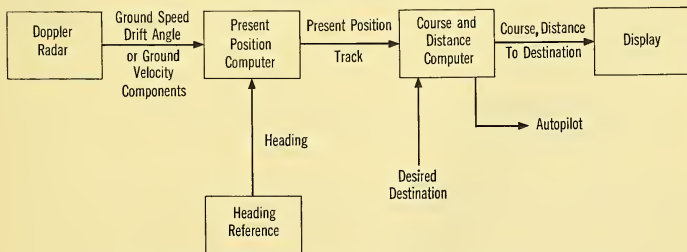


FIG. 14-3 Block Diagram of a Typical Doppler Navigation System.

speed meter in the previous system yields a block diagram of a typical *doppler navigation system* as shown in Fig. 14-3. The system operates in the following manner. The doppler radar furnishes aircraft ground velocity as ground speed and drift angle (or as along-heading and cross-heading components) to the present position computer. The computer combines this information with heading information (from the heading reference) to convert it to earth coordinates and continuously integrates the velocity to obtain miles traveled. This results in North-South and East-West miles traveled and ground track being made good. Further conversion, using the secant of latitude, finally results in the generation of present position in coordinates of latitude and longitude, assuming those are the coordinates desired. The present position information is then furnished to the course and distance computer, where it is compared with the desired destination coordinates, and the course and distance to the destination are automatically and continuously computed. These are displayed to the pilot or furnished to an autopilot for automatic control of the aircraft.

It bears emphasizing that a doppler navigation system is not a complete system unless it contains all three elements—the doppler radar, the computer, and the heading reference. The performance of the overall system is again dependent upon the performance of its weakest component. However, the component which is weakest in the true air speed system—namely the velocity source—has been replaced by a highly accurate source, the doppler radar. The navigation computer can be built with accuracies compatible with the doppler radar; thus the heading reference is usually the weakest component of the system.

A doppler velocity measuring radar may be used for many applications other than as a component in a self-contained navigation system. These applications include radar display ground speed stabilization (as in AEW) photographic stabilization, approach and landing, and north-seeking compass correction. A large number of the design criteria for a doppler velocity measuring radar are invariant, regardless of the application, while certain other ones, such as frequency tracker time constant and degree of attitude stabilization may be greatly dependent on the application intended.

14-2 BASIC PRINCIPLES OF DOPPLER RADAR NAVIGATION

If we mount a transmitter and a receiver on an aircraft, transmit energy toward the ground, and receive the back-scattered energy, the difference between the received and transmitted frequency is the doppler shift caused by the fact that the transmitter (and also the receiver) are moving with respect to the earth, from which the energy is reflected. If we measure this

doppler shift and if we know the frequency of transmission, the velocity of light, and the direction of radiation, we can determine the only unknown quantity, the ground speed of the aircraft. The velocity can then be integrated to measure the miles traveled along the ground and combined with heading information to compute present position.

A *doppler navigation radar*, then, is an airborne radar which transmits electromagnetic energy toward the earth's surface and utilizes the doppler shift of the received energy to determine two or three of the velocity components of the aircraft. This is illustrated schematically in Fig. 14-4.

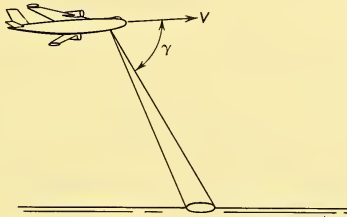


FIG. 14-4 Basic Doppler Beam Configuration.

The basic output of a doppler radar is a frequency, which is the observed doppler shift, given by the basic doppler equation

$$f_d = \frac{2Vf}{c} \cos \gamma = \frac{2V}{\lambda} \cos \gamma \quad (14-1)$$

where f_d is the doppler shift, V is the velocity of the aircraft, c is the velocity of light, γ is the angle between the velocity vector and the direction of propagation, and λ is the wavelength of transmission.

Since the antenna beam has a finite width and since the scattering from the earth is randomlike, the information received from the ground is not a single frequency, but rather is in the form of a noiselike frequency spectrum as shown in Fig. 14-5. A certain amount of smoothing time is therefore required to determine the quasi-instantaneous velocity to a given accuracy. A velocity smoothing time constant of about 1 second is usually chosen; this value is limited by the need for compatibility with the dynamics of typical aircraft. However, the effective smoothing time for navigational *distance* measurement is the total time flown and, for typical systems, it turns out that the so-called *fluctuation error* is completely overshadowed by certain other instrumentation errors after approximately 10 miles of flight.

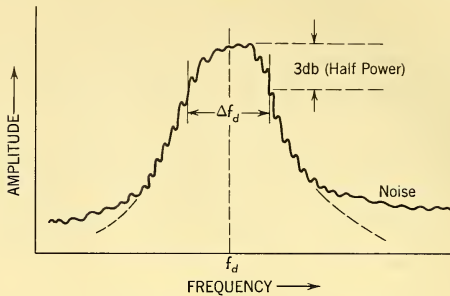


FIG. 14-5 Typical Doppler Spectrum.

It is the function of the *frequency tracker* of the doppler radar, which will be discussed later, to determine the center of area of the doppler spectrum, i.e. to determine the single frequency which is proportional to the desired velocity component. The half-power width of this spectrum Δf_d is quite important with regard to velocity accuracy considerations. It is dependent on the antenna beamwidth and several other radar parameters and can be expressed mathematically by the formula

$$\Delta f_d = \frac{2V}{\lambda} \sin \gamma \Delta \gamma \quad (14-2)$$

where Δf_d is the half power spectrum width, and $\Delta \gamma$ is the two-way beamwidth of the antenna beam in the γ -direction.

In order to measure both of the horizontal velocity components of the aircraft or both ground speed and drift angle, it is necessary to have at least two beams radiating energy toward the ground. This is shown schematically in Fig. 14-6. The along-heading velocity component (or ground speed

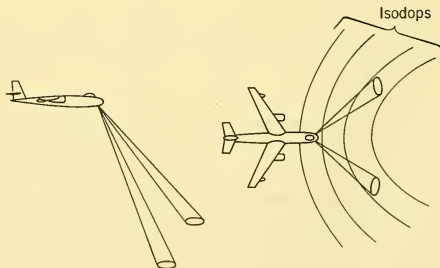


FIG. 14-6 Two-Beam Doppler System Configuration.

in a gimballed and track-directed antenna) can then be obtained by adding the doppler shift obtained from the two beams.

The cross-heading velocity component is obtained by taking the difference between the doppler shift from the two beams. In a gimballed antenna system, the drift angle can be obtained by servoing the antenna until equal doppler shifts are obtained from the two beams. It is clear that the measurement of along-heading and cross-heading velocity components is equivalent to the measurement of ground speed and drift angle, since the former two are simply components of the ground velocity vector, of which the ground speed is the magnitude and the drift angle is the angle of the vector with reference to the aircraft heading line.

The hyperbolas appearing on the right side of Figs. 14-6 and 14-7 are lines of constant doppler shift on the ground ("isodops") for a velocity vector

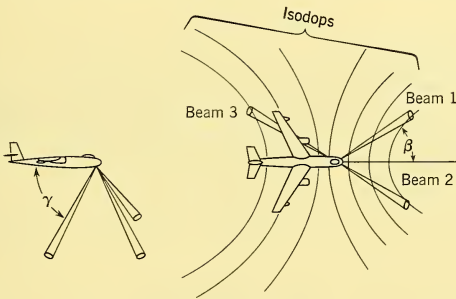


FIG. 14-7 Three-Beam Doppler System Configuration.

which is coincident with the heading line of the aircraft. By considering these lines, it is easy to see how the doppler shifts from the left and right antenna (of a fixed antenna system) will be different if the velocity vector is not coincident with the heading line, i.e. when a drift angle exists. The difference between these two doppler shifts is then proportional to the cross-heading (or drift) component of velocity.

For a number of reasons it turns out that the use of two beams is not optimum for most applications, but rather the use of three or four beams, as shown schematically in Figs. 14-7 and 14-9. This configuration has been given the name *Janus*, after the Roman god who was said to have the facility of looking forward as well as backward. Such a Janus system has the following advantages over the one-way looking (or non-Janus) system:

1. It provides for inherent cancellation of the vertical velocity components, when the forward and rearward looking beams are combined to measure horizontal velocity components.

2. The measurement of the horizontal velocity components is highly insensitive to errors in the vertical reference used.
3. Measurement of the vertical velocity component (rate of climb and descent) is possible, whereas it is not possible in the 2-beam (non-Janus) case. In fact, in the non-Janus case, it is necessary to furnish rate of climb and descent information from an external source, such as a barometric device, as a correction signal.

In a Janus system, the doppler shift obtained with a forward beam is effectively added to the doppler shift obtained with the rearward beam. For the condition of no pitch and roll (so that the forward and rearward doppler shifts are equal) Equation 14-1 for the total doppler shift from two beams takes the form:

$$f_d = \frac{4V}{\lambda} \cos \gamma \quad (14-3)$$

where the symbols are the same as those defined earlier.

The particular beam configuration shown in Fig. 14-7 has been called λ -configuration. It is perhaps the most general beam configuration for *fixed-wing* aircraft. Since it is a three-beam Janus system, all three components of aircraft velocity — the along-heading, cross-heading, and vertical components of velocity — can be obtained with it. For the condition of no aircraft pitch or roll, the mathematical expressions for the computation of the three velocity components are:

$$\bar{V}_H = \frac{(\bar{f}_{d_3} - \bar{f}_{d_1})}{4\lambda \cos \theta \cos \beta} \quad (14-4)$$

$$\bar{V}_D = \frac{(\bar{f}_{d_1} - \bar{f}_{d_2})}{4\lambda \cos \theta \sin \beta} \quad (14-5)$$

$$\bar{V}_V = \frac{(\bar{f}_{d_3} + \bar{f}_{d_2})}{4\lambda \sin \theta} \quad (14-6)$$

where \bar{V}_H = along-heading velocity component

\bar{V}_D = cross-heading (drift) velocity component

\bar{V}_V = vertical velocity component

\bar{f}_{d_n} = doppler shift of beam n

θ = depression angle of antenna beams

β = azimuth angle of antenna beams (smallest angle between projections of longitudinal axis of aircraft and antenna beam on the ground plane).

Since a condition of no pitch or roll was assumed for the above expressions, the quantities \bar{V}_H , \bar{V}_D , and \bar{V}_V can be defined as the three orthogonal velocity components in *aircraft coordinates*.

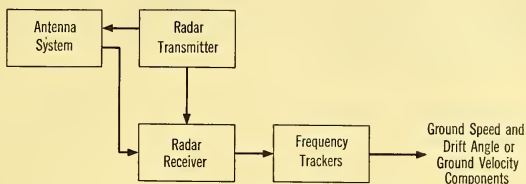


FIG. 14-8 Block Diagram of Typical Doppler Radar.

The block diagram of a typical doppler radar is shown in Fig. 14-8. It consists of four major components—the antenna system, a radar transmitter, a radar receiver, and one or more frequency trackers. The outputs of the doppler radar are frequencies, voltages, or pulse rates proportional to the orthogonal aircraft velocity components or to the ground speed, and/or a shaft angle proportional to drift angle of the aircraft.

14-3 SYSTEM CONSIDERATIONS

Typically, doppler navigation systems are employed where the navigation accuracy requirements are of the order of 1 per cent of the distance traveled. Because of certain limitations which will be explained in more detail, accuracies appreciably better than this are not obtainable, particularly for extended operations over water. Since a doppler navigation system is an *active* system, its use is restricted to applications where the transmission of electromagnetic radiations does not seriously endanger aircraft survival or mission accomplishment. Other important system considerations governing the choice of a particular type of doppler navigation system are (1) trade-offs between accuracy—as it affects mission accomplishment—and the size, weight and complexity of the system, (2) mission profiles (maximum and minimum speeds, altitudes, range, times of flight, and geographical latitudes), and (3) types of terrain and weather.

A doppler navigation system may take many forms. It may be a combination of various existing components or it may be a highly integrated system. The basic thesis, advanced earlier, that the system performance, primarily as regards accuracy, is a function of the performance of its weakest component, holds in either case.

Accuracy is the basic measuring-stick of navigation system performance. The position accuracy of a doppler navigation system depends on the following four sources of error:

1. Doppler ground speed (or along-heading velocity) error, σ_V
2. Doppler drift angle (or cross-heading velocity) error, σ_D
3. Heading reference error, σ_H
4. Computer error, σ_C

The total position error may then be expressed *approximately* on a root-sum-squared basis by

$$\sigma_P = \sqrt{\sigma_V^2 + \sigma_D^2 + \sigma_H^2 + \sigma_C^2} \quad (14-7)$$

where σ_P is the standard deviation (or rms) positional error in percent
 σ_V , σ_D , and σ_C are the standard deviation (or rms) per cent errors of the quantities defined above. When σ_D is expressed as drift angle error rather than as cross-heading velocity error, it must first be converted to an equivalent velocity or distance error through division by 57.3° (1 radian)
 σ_H is the standard deviation (or rms) heading reference error converted to an equivalent distance error through division by 57.3° (1 radian).

As an example, if we assume a standard deviation doppler ground speed error of 0.2 per cent, a doppler drift angle error of 0.15° , a heading reference error of 0.5° and a computer error of 0.25 per cent, then an overall position error of 0.97 per cent of distance traveled results. If three of the component errors are left unchanged, and only the heading reference error is changed from 0.5° to 0.25° , the overall system error now becomes 0.6 per cent of distance traveled. These two cases are typical and illustrate the importance of the heading reference error with regard to system accuracy performance.

One important aspect of error analysis is the statistical behavior of the data; that is, the question arises, What is the frequency of occurrence or what is the probability of occurrence of certain errors? If and only if one has a sufficiently large number of samples, can one realistically talk about a measured probability distribution, i.e. about *probable errors* (50 per cent probability) and *maximum errors* (95 per cent probability).

A somewhat more accurate method of calculating the system error than that indicated by Equation 14-7 is based on statistical probability considerations as a two-dimensional error problem.¹ The two dimensions chosen are the along-track and the cross-track directions, frequently called *range* and *transverse* directions, respectively. The along-track or range error σ_R

¹For a detailed treatment of statistical analysis applied to the two-dimensional error problem see Merrill, Goldberg, Helmholtz, *Operations Research, Armament, Launching*, p. 102, D. Van Nostrand, Co., Inc., Princeton, N. J., 1956.

is further assumed to consist of the error in doppler velocity (or the along-heading velocity component) σ_V and a computer error contribution to the range error σ_{C_R} . The cross-track or transverse error σ_T is assumed to consist of the error in doppler drift angle (or cross-heading velocity component) σ_D , the error of the heading reference σ_H , and a computer error contribution to the transverse error σ_{C_T} . Based on measured results, these component errors, and hence also the range and transverse errors, exhibit roughly normal or Gaussian probability distributions.

The resulting position error therefore exhibits a two-dimensional normal or elliptical probability distribution. If the standard deviations of the component errors σ_V , σ_{C_R} , σ_D , σ_{C_T} , and σ_H are known, the standard deviations of the range and transverse errors can be determined from the following expressions:

$$\sigma_R = \sqrt{\sigma_V^2 + \sigma_{C_R}^2} \quad (14-8a)$$

$$\sigma_T = \sqrt{\sigma_D^2 + \sigma_{C_T}^2 + \sigma_H^2} \quad (14-8b)$$

where all symbols are standard deviations of the quantities which were defined in the text above.

The major and minor axes of the ellipse — namely, the range and transverse error, σ_R and σ_T — having thus been determined, one can obtain the radius of the position error circle for a desired probability of occurrence as a function of the transverse error σ_T and the ratio $\sigma_R/\sigma_T = b$ and a multiplying factor a , by consulting "Tables of Integrated Probability within a Circle Centered at the Origin of Normal Elliptical Distributions."²

Expressed in general mathematical terms, if the ratio b is known, where

$$b = \frac{\sigma_R}{\sigma_T} \quad (14-9)$$

a multiplying factor a can be found from the elliptical probability tables for the determination of the radius of the position error E_p for any desired probability, such that

$$E_p = a\sigma_T. \quad (14-10)$$

For the special case for which σ_R and σ_T are equal, the distribution becomes circular (rather than elliptical) and is known as a Rayleigh distribution. As will be demonstrated later by means of measured data, in typical doppler navigation systems, σ_T and σ_R are most frequently not equal; rather, their ratio is often found to be 2 or greater ($\sigma_T/\sigma_R \geq 2$), which is usually attributable to the effect of the heading reference error. From the probability tables mentioned above, one finds, for instance, that

²Such tables have been calculated — for instance, in yet unpublished work by R. H. Bacon and others at GPL Division, General Precision, Inc., Pleasantville, New York.

for a case of $\sigma_T/\sigma_R = 2$, the 95 per cent probability (*maximum*) error circle has a radius of $2.04\sigma_T$, while the radius of the 50 per cent probability (*probable*) error circle has a radius of $0.87\sigma_T$, or a ratio of maximum to probable error of nearly 2.34. For the case of $\sigma_T/\sigma_R = 1$, the 95 per cent probability circle has a radius of nearly $2.45\sigma_T$, while the 50 per cent probability circle has a radius of nearly $1.18\sigma_T$.

Measured results, described later, have shown that the per cent range error σ_R , the per cent transverse error σ_T , and hence the per cent position error, of a doppler navigation system are inversely proportional to the square root of the distance traveled. This is expressed mathematically by

$$\sigma_R = \frac{K_R}{\sqrt{D}} \quad (14-11a)$$

$$\sigma_T = \frac{K_T}{\sqrt{D}} \quad (14-11b)$$

This behavior can be explained by the fact that the additional smoothing afforded by longer flights decreases the effect of certain slowly varying random errors and of any end-point errors. Thus if the standard deviations of the component errors listed earlier are known for a given distance (as is frequently the case for doppler radars and heading references), the constants K_R and K_T above can be determined from Equation 14-8 through 14-11. The range, transverse, and radial position errors for any desired distance can then be found by means of Equations 14-11a and 14-11b.

14-4 MAJOR CHARACTERISTICS AND COMPONENTS OF A DOPPLER RADAR NAVIGATION SYSTEM

Type of Transmission. The types of transmission for doppler radars can be divided into four general categories:

- Low-duty cycle pulse (5 per cent or less)
- Continuous wave (CW)
- High-duty cycle pulse (25 to 50 per cent)
- Frequency modulation — continuous wave (FM-CW)

The pulse systems can be further subdivided into incoherent and coherent pulse systems; low-duty cycle pulse systems are normally found in the incoherent class, while high-duty cycle pulse systems are found in both the incoherent and coherent classes. The order in which these types are listed above coincides approximately with the historic development of doppler radars. The low-duty cycle pulse system employs techniques and components most similar to those used in conventional pulse radar systems. Although this type of transmission is the least efficient of the four, it is the

one used in the earlier doppler radars largely because of the state of the art of techniques and components at the time. The continuous wave (CW) type of transmission is the simplest and inherently most efficient technique; however, it is still faced with a number of difficulties (primarily transmitter-receiver isolation) which have made the design and installations of such systems somewhat precarious and troublesome and have resulted in double-aperture antenna systems requiring larger antenna cutouts in the airframe. The high-duty cycle (25–50 per cent) pulse and the frequency-modulation-continuous-wave (FM-CW) types of transmissions combine most of the advantages of the other two systems, namely high efficiency, high transmitter-receiver isolation (permitting single antenna operation), and only moderate complexity.

The coherent type of high-duty cycle pulse system is intrinsically somewhat more efficient than the incoherent type, primarily at high altitudes. However, it tends to be more complex than the incoherent system, since in addition to special modulation techniques it normally requires several klystron tubes as compared with the single magnetron of incoherent systems. Also, at least from certain aspects, the transmitter frequency stability requirements for a coherent pulse system are more critical than for an incoherent pulse system. For the latter system the frequency needs to remain only within the receiver IF bandwidth, and the frequency deviation during a transmitted pulse must not be large compared with the reciprocal of the pulse duration.

The FM system combines many of the advantages of CW and pulse systems. It resembles a CW system in simplicity, since normally a single klystron is required in the transmitter which needs to be modulated only by a simple sine-wave oscillator. The receiver of such a system is so designed as to accept and use only the doppler shift of a particular sideband of the beat between the received and transmitted signals; thereby the return energy from nearby objects is largely rejected. This is so because the modulation index of the beat, and hence the amplitude of all but the zero-order sidebands, decreases very rapidly with decreasing range, becoming zero at the receiver terminals. Transmitter-receiver isolation and radome reflection problems (prevalent in CW systems) are eliminated here on a frequency basis, much as they are eliminated in pulse systems on a time basis. Since the energy in all other sidebands, except a particular one, is rejected in such a system, a certain spectrum utilization inefficiency results. In certain cases, coherence requirements are similar to those of CW and coherent pulse systems, and the requirement for maintaining the modulation index at a particular value is reasonably critical. In single-antenna FM-CW systems the problem of receiver crystal deterioration caused by transmitter power feed-through can be severe, preventing the use of higher transmitter powers (greater than 1 watt or so) which might be

required for high-altitude operation. Contrary to CW systems, all pulse and FM systems are plagued by so-called "altitude hole" problems which are discussed in later sections.

Coherency. In order for a doppler system to function, it is necessary for the received signal and a reference signal to be made phase-coherent with one another. This can be accomplished in a number of ways. However, there are primarily three methods which are commonly used.

In one method the transmitter is itself an incoherent pulsed transmitter and the fore and aft returns of the system are used to beat against one another, since they are themselves always of the same phase. This method has been called *external* or *self-coherence*. With this type of coherency, it is frequently necessary to employ physically stabilized antennas in order to assure range overlap at all times.

Continuous wave (CW) systems are coherent by the nature of their transmission; i.e. the transmitted, received, and reference signals are continuously present at all times.

Coherent pulse doppler systems can be made coherent by using a continuously running oscillator as driver (as well as a source of reference signal), feeding its output to an amplifier, and pulsing or chopping the transmitted signal in the transmitter output circuit. This can be accomplished in a gridded klystron tube, traveling wave tube, and/or a crystal modulation circuit, the latter having exhibited considerable advantage over the gridded klystron, primarily because of the state of the art of these tubes. A radar transmitter of this type is called a *coherent radar transmitter* and has recently found application in a number of other radar systems.

Frequency-modulation-continuous-wave (FM-CW) systems are coherent in much the same way as pure CW systems; i.e., the transmitted, received, and reference signals are continuously present at all times. However, in many FM-CW systems the fore and aft return spectra are mixed in a manner similar to that used in the incoherent pulse systems in order to eliminate the undesirable effects of incidental frequency modulation in the transmitter.

Frequency of Transmission. Doppler radar systems have been designed at X band (8800-10,000 Mc) and at K_E band (13,500 Mc). While the latter frequency band has certain advantages over the former from the viewpoint of gain and over-water operation for the same antenna aperture size, it brings with it slight degradations as regards operation in weather, particularly rain. The importance of these weather effects has yet to be determined and may well not be significant.

Polarization. The types of polarizations presently used for doppler radars are linear and circular-odd (opposite rotation received), although

cross-linear and circular-even (same rotation received) have also been considered. Originally the main advantage of circular-odd polarization was more efficient duplexing; ferrite duplexers have now removed this advantage over linear polarization. The same is true for cross-linear polarization; however, the greatly increased return power loss, particularly over water, is a considerable disadvantage of this polarization. To date, there is no experimental evidence of any appreciable difference in performance between linear and circular-odd (opposite rotation received) polarization. Circular-even polarization (same rotation received) has very desirable and well-known rain discrimination properties. However, like cross-linear polarization, it suffers from an appreciable scattering loss over the sea, which has discouraged designers from using this polarization in practical systems.

Beam Configuration. The beam configurations of doppler radars can be classed into two general categories, namely Janus (two-way looking) and non-Janus (one-way looking). Janus systems use three or four beams of radiation, while non-Janus systems use two beams of radiation. Four beams are used in some systems, primarily because of the symmetry of this configuration (which usually results in greater computer simplicity) and also because it is naturally produced by certain types of antennas, such as planar arrays. While two beams are the minimum number for determination of ground speed and drift angle (or along-heading and cross-heading velocity components), three beams are the minimum number for determination of along-heading, cross-heading, *and* vertical velocity components. However, as indicated briefly in the beginning of this section, even for determination of only the horizontal velocity components the Janus configuration has important advantages, at the cost of one extra beam, over the non-Janus configuration — namely, much less velocity error dependence on the knowledge of the vertical attitude and vertical rate of the aircraft. The mathematical expressions for this dependence are easily derived and are

$$E_v = \frac{\delta v}{v} = \tan \gamma (\delta \gamma) \quad \text{non-Janus (pitch only)} \quad (14-12a)$$

$$E_v = \frac{\delta v}{v} = 1 - \cos (\delta \gamma) \quad \text{Janus (pitch only)} \quad (14-12b)$$

$$E_v = \frac{\delta v}{v} = 1 - \cos (\delta \gamma) + \sin (\delta \gamma) \tan \alpha \quad (14-12c)$$

Janus (pitch and vertical velocity)

where E_v = fractional velocity error

δv = error in velocity

v = velocity

$\delta\gamma$ = uncertainty in pitch angle

γ = angle between horizontal velocity component and direction of radiation

α = angle of climb or descent.

It is seen from these expressions that for a γ angle of 70° (a typical value for many doppler radars) the non-Janus system suffers approximately 4.7 per cent error in velocity per degree of uncertainty in pitch angle, whereas the Janus case suffers approximately 0.014 per cent per degree of uncertainty. Similarly, the error incident to vertical velocity is very small in the Janus case, while it must be compensated for in the non-Janus case, with uncertainties in vertical velocity appearing as a direct factor.

The choice of γ angle for a doppler system represents a compromise between high velocity sensitivity (cycles-per-second per knot) which increases with smaller γ angles, and high signal return over the sea which increases with larger γ angles (or at least larger antenna depression angles, which generally increase with γ angles). Most early equipment designers adopted a γ angle of 70° as the optimum compromise value. Another disadvantage of γ angles larger than this value is a larger land-to-water bias error (discussed in more detail in Sec. 14-5); however, the use of land-sea switches has reduced the importance of this factor to a very considerable extent, resulting in the use of larger γ angles in later designs.

The choice of the β angle depends largely on the application of the system and certain technical considerations such as drift angle sensitivity and signal return. Typical β angles range all the way from 20° to 90° . Fig. 14-9

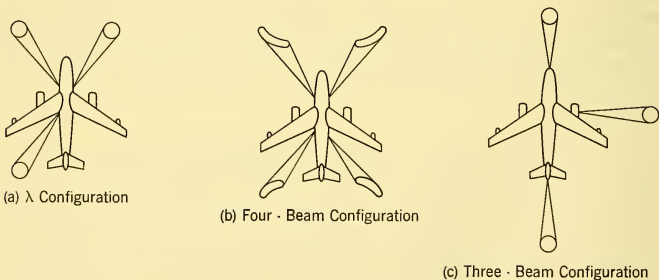


FIG. 14-9 Typical Doppler Janus System Beam Configuration.

shows three examples of azimuth beam configurations of three typical Janus system configurations. Fig. 14-9a is the λ configuration already

discussed. Fig. 14-9b is a typical four-beam configuration particularly suited for certain RF-mixed Janus systems. Fig. 14-9c is a configuration particularly suitable for a helicopter doppler system, since a complete $\pm 180^\circ$ drift angle range is required for this application; this configuration can also be applied to fixed-wing systems.

Antennas. Antennas used for doppler radars include linear arrays producing partial fan-shaped beams and a large variety of other antennas producing pencil beams or near pencil beams such as those shown in Fig. 14-9. Among the latter are parabolas and cut parabolas, dielectric and metal plate lens antennas fed by horns, and planar slotted arrays. Linear and planar arrays are normally considerably thinner than equivalent lenses and parabolas; they are also more adaptable to beam shaping and can be designed to make the doppler calibration constant completely independent of transmitter frequency. Essentially the desired beamwidth dictates the aperture or cutout size of the antenna or antennas. The size of the beamwidth largely determines the velocity fluctuation error, the over-water bias error, and hence the overall system error of the doppler navigation system. One-way beamwidths chosen for modern doppler radars range from 3° to 7° , the majority of navigation systems having adopted a value near 5° .

Misalignment of the antenna can contribute a fixed (or systematic) navigation error. Janus-type systems are generally much less sensitive to this alignment error than non-Janus systems because of the nature of Janus self-compensation. Also, since the antenna misalignment error is systematic it can be "biased out" in a flight calibration procedure. This applies as well to any error in antenna bore-sighting.

Stabilization. As indicated in the discussion in the previous sections all doppler radars are sensitive in some degree to the attitude (pitch and roll) of the aircraft — Janus systems less so than non-Janus systems. Some correction or stabilization for the pitch and roll of the aircraft must therefore be made by means of data from a vertical reference. The two ways of accomplishing this are called *antenna stabilization* and *data stabilization*. Also, systems using fan-shaped beam configurations require *physical drift angle stabilization*, because of the shape of the constant-doppler hyperbolas shown in Figs. 14-6 and 14-7. For the same reason if such systems are roll and pitch stabilized, this stabilization is best done physically by the antennas. This form of stabilization is therefore termed *antenna stabilization*. *Data stabilization*, on the other hand, as the name implies, constitutes pitch angle and roll angle correction of the data in some form of computer and allows the use of fixed antennas, usually requiring less cutout size and weight than stabilized antennas. Antenna stabilization is capable of somewhat greater accuracy in some cases, particularly for drift-angle determination over water. While fan-beam systems require the use of

antenna stabilization, pencil beam systems may use either antenna stabilization or data stabilization.

Receivers. Doppler radar receivers can be classed into the two categories of *crystal video* (zero frequency IF) and the *intermediate frequency* (IF). There are further differences, determined largely by whether the system is CW or pulse, non-Janus, RF Janus, IF Janus, or AF Janus, as discussed later in this paragraph.

Crystal video receivers heterodyne a portion of the transmitted energy directly with that of the received energy to obtain an audio doppler component at the crystal of the receiver. While this system is by far the simplest and also eliminates one of the carrier leakage problems, it suffers from the poor noise characteristics of crystals at low audio-frequencies (i.e. low velocities) resulting in an inherent reduction of signal-to-noise ratio by 15-20 db. Intermediate frequency systems require the necessary local oscillators, mixers, and filters, but exhibit a 15-20 db signal-to-noise ratio improvement over crystal video receivers.

In incoherent pulse systems the local oscillator source is usually a conventional low-power klystron. In CW, coherent pulse, and FM-CW systems the local oscillator power is usually derived directly from the transmitter or a continuously-running driver (in the pulse case) by means of a so-called side-step local oscillator crystal mixer. The latter method cannot be used in the incoherent transmitter case, since local oscillator energy would then not be present during the receiving period.

Frequency Trackers. As mentioned earlier, the function of the frequency tracker is to determine the center (of area) of the noiselike frequency spectrum obtained from the ground, which in turn results from the random nature of the radar scattering phenomenon.

The generic forms of the frequency tracker are the *axis-crossing counter* and the *closed-loop frequency filter* or *frequency discriminator*.

Practically all modern doppler radars use some form of closed-loop frequency discriminator as the frequency tracking device. Such a frequency tracker can further assume different forms, depending on whether the tracking operation is done at audio (d-c) or at some intermediate frequency and whether or not the frequency discriminator employs a sweeping operation, a comparison of two nearby filters, some form of autocorrelation technique, or a combination of these methods. Practice appears to show that there is little difference in the performance of any of the existing frequency trackers either as regards accuracy or signal sensitivity, although conclusive comparisons have not been made to date. The doppler signal is usually fed to one or more modulators which modulate it with a signal from a variable frequency oscillator (tracking oscillator) and feed the output to the device which performs the discriminator function. The output from the

latter is smoothed in an integrator and used to control the frequency of the tracking oscillator. The tracking oscillator can then provide the frequency tracker output. Mechanically tuned discriminators have also been employed, although they appear to have less accuracy capability.

The type of computation system used in conjunction with the doppler radar determines the form which the output of the frequency tracker must have. Most analogue computation systems require 400-cps voltages proportional to the velocity components. In such cases, the frequencies must be converted to shaft positions or voltages; the basic components used for this purpose have been the motor-tachometer, which has a linearity error proportional to its full-scale value, and, more recently, various types of more accurate electronic circuits, such as the so-called "bucket counter." Radars operating in conjunction with digital computers that accept pulse trains whose rates are proportional to the desired velocity components do not have this requirement for frequency-to-voltage conversion.

In a well-designed system, the frequency tracker measurement error and the conversion error are the two primary sources of random errors in the radar. (As previously shown, the stabilization error can be made negligibly small.) Typically, the frequency tracker is of the order of better than 0.05 per cent of the velocity, while the conversion error ranges from negligible to small values for digital and bucket counter outputs to 0.1 per cent of the full-scale velocity for tachometer outputs.

Ground Speed and Drift Angle Determination. In general, the ground speed or the along-heading velocity components are obtained by adding the one-directional (rearward) left and right doppler components in the non-Janus case, or by subtracting the forward and aft-looking doppler components in the Janus case. The drift angle or drift (cross-heading) velocity component is generally obtained by subtracting the left and right doppler components in either the Janus or the non-Janus case. In the antenna-stabilized system, however, the drift angle can be obtained by servoing the antenna until the left and the right doppler components are equal. Practically, the two techniques are equivalent.

Furthermore, Janus systems can be subdivided into three categories, *RF Janus*, *IF Janus*, or *AF Janus*, depending on whether the Janus mixing process is done at radio, intermediate, or audio frequencies. All three types are represented in modern doppler radars. The type used has an effect on whether or not directional sense of motion is obtained (as required in helicopters), on the signal-to-noise ratio obtained, on cancellation of incidental FM in the transmitter, and on the dynamic characteristics of the system. Generally speaking, Janus mixing after tracking results in slightly improved signal-to-noise ratios at the higher altitudes, but at the same time it results in some degradation in dynamic performance where fixed antennas

are used and does not afford any cancellation of incidental FM in the transmitter, which can result in appreciable spectrum broadening.

The Navigation Computer. The computer of the navigation system is required to do primarily two jobs:

1. Computation of present position by integration of the velocity and by transformation of the data to directionally oriented earth coordinates by use of data from a heading reference.
2. Computation of course and distance to a destination by comparing destination coordinates with present position coordinates and thus solving the navigational triangle.

The discussion which follows will treat these two functions of the computer separately.

Mathematically, all present position computers are essentially equivalent, since they must integrate the velocity components after converting them into earth coordinates. There are, however, several types of computers, which may be classed in several ways. One division is that of *generic analogue computers* and *general purpose digital computers*.

Generic analogue present position computers can be classed as:

1. Electromechanical analogue computers
2. Electrical analogue computers
3. Operational digital computers

The *electro-mechanical analogue computer* contains ball-and-disk integrators and other analogue components to perform the analogue computation. Several modern navigation computers are of this type. They are probably the oldest of the three types listed and therefore represent fairly well-established techniques. Their major characteristics include reasonable simplicity, medium size and weight advantage, good reliability and maintenance characteristics, and fair accuracy.

The *electrical analogue computer* contains various forms of electronic integrator circuits to perform the computations and uses such components as motor tachometers, synchros, and resolvers which are considered to be primarily electrical rather than mechanical. Several modern computers are of this type and their characteristics include medium size and weight, good reliability and maintenance characteristics, and medium accuracy.

The *operational digital computer*, is probably the most recent development of the three types. In a generic or conceptual sense, it is an analogue computer which performs the computation in real time. The computer converts the doppler frequencies representing the velocity components into pulse trains, the rates of which are now proportional to the velocity

components, and performs the integration by simply counting these pulses, so that the total number of pulses represent total miles traveled. The coordinate transformation (with reference to heading) and roll and pitch stabilization can be done by means of binary multipliers. The major characteristics of this type of computer include very high potential accuracy, because of the digital techniques used, and medium size and weight. Among the disadvantages of this type of computer is its high component count and the necessity of using analogue-to-digital converters for converting such quantities as heading and roll and pitch information into digital form. However, small-size analogue-to-digital converters of this type have recently been developed for this purpose.

It is the function of the course and distance computer to compare a desired destination with the aircraft's present position and thus by solving the navigational triangle, to compute the desired course and distance to destination. Mathematically there are three types of computations possible,³ and all three types are represented in modern course computers.

These are:

1. Great-circle computation
2. Rhumb-line computation
3. Midlatitude planar computation

Existing present-position computers are of either the analogue or the digital type. Their accuracies range from 0.2 per cent to 2 per cent peak error. Their weights range from 15 lb to approximately 50 lb depending largely on the accuracy, the form of outputs, and, particularly, on the number of special functions and outputs required.

Existing course and distance computers are most frequently analogue type. Great-circle, rhumb-line, and midlatitude types of computers are in use; the trend is in the direction of great-circle computation for long-distance navigation in view of the great accuracy of this computation for long distances and improvements in computer techniques, making small weights and volumes possible. Course computer accuracies are of the order of 0.5 per cent peak error, and weights vary greatly, again depending on the type of outputs required.

Presently there is considerable development effort on general-purpose as well as differential analyzer types of digital computers (and combinations of these) for airborne applications, which can be programed for the present position and course computation. Such computers promise to have high potential accuracy and versatility, although they tend to exhibit considerable complexity.

³See Locke, *Guidance*, Chap. 3 for a discussion of terrestrial and celestial navigation.

The Heading Reference. The heading reference is an extremely important component of a doppler navigation system, not only because it is the device which transforms the navigational information into directionally oriented earth coordinates but also because the heading reference error has a major effect on the total system accuracy of a navigation system. In terms of numbers it turns out that a 1° error in heading represents approximately a 1.75 per cent error in present position. It is not surprising that the heading reference error tends to swamp the other errors in most modern doppler navigation systems and therefore assumes extreme importance in the system analysis and synthesis of such systems.

The three most important types of heading reference are: (1) the gyromagnetic compass, (2) the earth-rate directional reference or north-seeking gyrocompass, and (3) the astrocompass.

The best versions of the first two devices can achieve probable heading accuracies of the order of $0.25 - 0.35^\circ$ at latitudes up to 70° . In the polar regions, both of these systems have to be operated in a free-gyro mode and the errors are somewhat larger, increasing with the flight time. Because it does not depend upon earth's magnetic field information, further improvements in the north-seeking gyrocompass system accuracy may be anticipated as the state of the art in gyros and stabilization techniques improves. The astrocompass is capable of great accuracy. However, it requires a clear sky and accurate vertical reference, time and position information, and a gyro for memory.

14-5 DOPPLER NAVIGATION SYSTEM ERRORS CAUSED BY INTERACTIONS WITH THE GROUND AND WATER

The foregoing pages have dealt with system errors originating within the navigation system. Four major sources of random error were indicated: (1) frequency measurement, (2) data conversion, (3) computation, and (4) heading reference. For Janus systems in particular, the other possible major sources of random error (stabilization and fluctuation) were shown to be negligible if sufficient care is taken in the design. Also, it was assumed that systematic errors originating in the radar (antenna misalignment, boresighting, and average transmission frequency) could be "biased out" in a calibration procedure.

There is still another error which arises from interactions between the radar system and the terrain or water illuminated by the radar. This error — commonly called the *terrain bias error* — is both small and systematic for *overland* operation. It is caused by a combination of factors: (1) a *small* change in scattering coefficient with looking angle over the finite radiated beamwidth, (2) range difference effects over the beamwidth, and (3) nonlinearities in converting ray angles to doppler frequencies, resulting in a

very slight shift of the center of gravity of the received doppler spectrum and hence of the measured velocity. These errors, though very small, depend on beamwidth, increasing with increasing beamwidth. Flight calibration can completely "bias out" the terrain bias error over land.

Over water, the situation is more serious. The velocity accuracy is decreased by three causes: (1) an increased terrain bias error, (2) an error incident to surface droplet motion of the water, and (3) an error incident to water current motion.

The terrain bias error is perhaps the most important. It results from the marked change in back-scattering coefficient with looking angle over the extent of the radiated beamwidth. Figure 14-10 is a set of plots of the

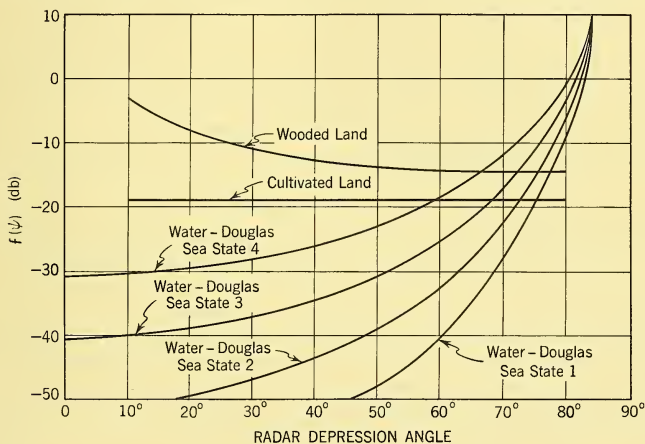


FIG. 14-10 Scattering Coefficients for Land and Water, X Band. (Courtesy, General Precision Laboratory, Inc.)

measured back-scattering coefficient $f(\psi)$ versus the radar depression angle (see Fig. 4-4) for various Douglas sea state conditions and several types of terrain at X-band frequencies. The cross section per unit area σ^0 is related to $f(\psi)$ by the expression $\sigma^0 = 2f(\psi) \cos \psi$. Douglas sea state 1 is normally defined as "smooth sea with wave heights less than 1 foot" and Douglas sea state 4 is normally defined as "rough sea with wave heights from 5 to 8 feet." (Because of the manner in which Douglas sea states as well as the so-called Beaufort sea states are defined — on the basis of wind speed and wave height — they do not appear to be quite fine enough from the viewpoint of radar scattering. A more applicable, fine scale of sea states was developed by J. Campbell of General Precision Laboratory, Inc., and is

described in reference 7 of this chapter's bibliography, along with recent measurements of radar scattering at X band over land and sea. For example, in that sea scale the Douglas sea state 1 is subdivided into four separate sea states.) The curves in Fig. 14-11 show that for a beamwidth of of

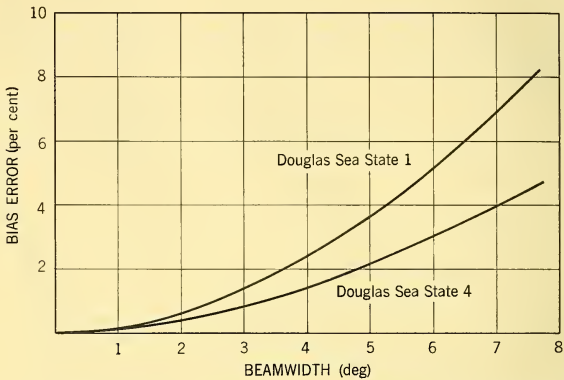


FIG. 14-11 Doppler Velocity Water Bias Error vs. One-Way Beamwidth for a γ Angle of 69° and β Angle of 0° . (Courtesy, Laboratory for Electronics)

5° , which requires an antenna diameter of approximately 18 inches at X band (3 cm) and 13.5 inches at K_E band (2.2 cm), the uncorrected bias error is 2.3 per cent at Douglas sea state 4 and 3.7 per cent at Douglas sea state 1. This is an error spread of 1.4 per cent between the two extremes. For a beamwidth of 3° (corresponding to antenna diameters of 30 inches at X band and 22 inches at K_E) the uncorrected bias errors for the two extremes of sea state reduce to 0.9 per cent and 1.4 per cent respectively, or a spread of 0.5 per cent.

Since this error is a systematic or bias error, it is possible to eliminate it either by careful calculation or (still better) by in-flight calibration for any one particular sea state. Most modern doppler systems are equipped with a *land-sea switch*, with one sea position, which should be calibrated for the average or most frequently occurring sea state. The error which remains is the spread between the errors for the sea state extremes. As can be seen from Fig. 14-11 and the examples cited above, this error spread again decreases rapidly with decreasing beamwidth (or increasing antenna aperture). For the two above examples of 5° and 3° beamwidths, the total error spread would be 1.4 per cent, or an actual peak error of ± 0.7 per cent, and 0.5 per cent, or an actual peak error of ± 0.25 per cent if the sea switch were assumed to be adjusted for the average of these sea states. These,

then, are the resulting water bias errors for typical modern doppler systems equipped with land-sea switches.

The second error mentioned above is caused by the motion of the water surface droplets (and hence scatterers) incident to the action of the surface wind. Workers in the United States (primarily at the Naval Research Laboratory) and in Great Britain have experimentally determined this error to be somewhere between 8 and 16 per cent of the surface wind speed. It should be remembered, however, that this error (unlike the bias error) is proportional to the prevailing wind speed and hence becomes smaller and smaller, percentagewise, as the speed of the aircraft increases.

The third and final water error mentioned is that incident to current motion (tidal currents, Gulf Stream, etc.). Since in this case the scattering surface below the aircraft is moving with a given rate, this rate appears as a direct error in the ground speed measurement. If this rate is known, however, as is often the case for water currents, then it is possible to correct for it by preflight calibration. In general such currents are known to move at speeds of no greater than 1 to 2 knots. And again, as in the case of the surface droplet motion error, the error decreases percentagewise as aircraft speeds increase.

It might be mentioned at this point that water "wave motion" as such produces no doppler error, since water mass is not actually transported or moved forward in the wave action.

In summary it can be said that, while water *motion* errors are appreciable at lower aircraft speeds and short distances, these errors become less significant for the higher aircraft speeds and for the longer distances. This is the case since the absolute velocity errors tend to be a small percentage of the total aircraft speed and because such errors are likely to occur in different directions and hence will tend to average out over long distances. Also, correction for a portion of these errors before or during the flight appears possible. The water *bias* error, on the other hand, can be reduced considerably through the use of a land-sea switch, and for typical doppler radars should amount to less than 0.25 to 0.7 per cent peak error in velocity. New automatic techniques have recently been developed which are designed to further reduce the over-water bias error.

14-6 MODIFYING THE RADAR RANGE EQUATION FOR THE DOPPLER NAVIGATION PROBLEM

The standard radar range equation (see Paragraph 3-2) may be expressed in the following form which is more convenient for the doppler radar problem:

$$\left| \frac{S}{N} \right|_d = \frac{P_{ave} G_o W_{RF} E_f(\psi) \lambda^2 F_j A}{N F k T \Delta f_d (4\pi h)^2 \sec^2 \psi} \quad (14-13)$$

where $\left|\frac{S}{N}\right|_d$ = ratio of total doppler signal power to the noise power in the bandwidth of the doppler spectrum, at the highest speed of operation. (The bandwidth of the doppler spectrum is measured at the 3-db points)

P_{ave} = average transmitted power per beam

G_o = one-way maximum antenna gain relative to an isotropic radiator

W_{RF} = RF attenuation in the plumbing of the transmitter and receiver paths, including duplexing and waveguide losses

E = *efficiency factor*, defined as the ratio of available doppler-signal-to-noise ratio to the doppler-signal-to-noise ratio which would be available if all the received signal were converted to doppler information. It includes sideband power loss in pulse and FM-CW systems, noise increase caused by "foldover," and any noise gating improvement. The foldover loss occurs because of the noise-bandwidth contraction and resulting noise density increase at the second detector output of CW or pulse systems not using single sideband detection techniques. It amounts to -3 db, as explained in Sec. 6-6. The noise gating improvement occurs if the receiver crystals are gated off completely from the receiver input during a portion of the transmitter time (as is done in most modern pulse systems) and is therefore proportional to the ratio of receiver-on-time to the pulse repetition period, sometimes called D_0 . The definition of E above applies to coherent non-Janus and coherent post-tracker-mixed Janus systems only. See additional j factor for incoherent and coherent pre-tracker-mixed Janus systems

$f(\psi)$ = scattering coefficient at the given ψ angle, where ψ is the angle between the beam and the normal to the surface of the sea or terrain [$f(\psi) = \frac{1}{2}\sigma^0 \sec \psi$]

λ = wavelength of transmission

NF = RF noise figure

k = Boltzmann's constant (1.38×10^{-23})

T = absolute temperature ($300^\circ \text{K} = 27^\circ \text{C}$)

Δf_d = bandwidth in which the noise is measured — the bandwidth of the doppler spectrum at the 3-db points, at the maximum

ground speed. This was shown in Equation 14-2 to be $\frac{2V}{\lambda} \sin \gamma \Delta\gamma$, where $\Delta\gamma$ is the two-way beamwidth of the antenna. In pre-tracker-mixed Janus systems, there is an additional factor of $\sqrt{2}$ in this expression due to the spectrum broadening in the Janus mixing process

h = aircraft altitude

j = Janus factor; applicable only to Janus (two-way looking) systems and consisting of the following three factors, any or all of which may exist and be practically significant: (1) j_s = signal suppression factor, which exists in any pretracker mixed Janus system but which becomes significant only at extremely small IF S/N ratios; it can cause the S/N ratio altitude fall-off to increase to as much as 12 db per octave (double the altitude); it is the result of certain undesirable signal-cross-noise and noise-cross-noise products. (2) j_t = time spread factor, which may become significant in coherent or incoherent pulse systems at high altitudes due to the time spread of the echoes (with respect to the pulse repetition period); it can cause the S/N fall-off to increase from 6 db per octave at the low altitudes to 9 db per octave. (3) j_0 = overlap noise factor, which may become significant in incoherent pulse systems at very high altitudes and which results from the so-called pulse overlap noise (noise produced by the beating of the returns of successive transmitter pulses)

A = antenna efficiency (frequently near 0.55)

ψ = incidence angle of beam at the ground, i.e. the angle between the vertical and the direction of radiation. This angle is normally somewhere between 10° and 30°

F = antenna pattern parameter. This factor depends on the type of antenna — parabola, rectangular aperture, linear array, etc. — and for an extended target such as the ground usually lies between 0.5 and 0.67.

One additional factor not included in Equation 14-13 is that of atmospheric attenuation, i.e. absorption in or back-scattering from rain, clouds, etc. The data presented in Chapter 4 can be used to account for these phenomena in calculating the theoretical signal-to-noise ratio.

Several other factors tend to degrade the signal-to-noise ratio in an actual equipment. For example, pure CW systems are affected by reflection noise

from vibrating equipment, back-scattering from weather and turbulent air, and transmitter signal leakage. Doppler spectrum broadening, caused by power supply ripple leading to transmitter amplitude or frequency modulation, local oscillator frequency drift, transmitter tube noise, and internal vibration noise, can degrade S/N ratio in all classes of doppler systems. These phenomena must receive careful attention in the design to avoid potentially severe degradations, particularly since they cannot be overcome by increasing transmitter power.

In summary, then, the signal altitude capability of a doppler radar is a function of available transmitter power and radar circuit design, the latter factor including antenna size, type of modulation, physical rigidity, etc. On this basis, most doppler radars achieve the minimum doppler signal-to-noise ratio which is required by the doppler frequency tracker for signal acquisition (2 to 4 db) at an altitude above 70,000 ft over Douglas sea state 1 and at their top operational ground speed, with an average transmitter power between 2 and 20 watts.

It is worthwhile noting at this point that many doppler navigation systems are equipped with an automatic wind memory circuit, which becomes operative as soon as the received doppler signal-to-noise ratio goes below a predetermined threshold which is selected near the minimum signal-to-noise ratio required for tracking by the frequency tracker (normally between 0 and 2 db). It is immaterial whether this signal drop occurs when flying over extremely calm water, when making a steep bank, when climbing rapidly or because of loss of transmitter power. As soon as the signal drops below this threshold level, the system goes "on memory," i.e. the ground speed data fed to the computer are obtained by vectorially adding a true air speed signal from a true air speed meter and a "last remembered wind vector" signal which has been continuously computed from the difference between doppler ground speed and true air speed and which is instantly available as soon as the system goes "on memory." It is therefore possible to continue to obtain rather accurate navigational information from such a system even if the doppler signal goes below the required level for a limited time. Generally speaking, this accuracy is maintained as long as the wind structure does not change. When it does change the remembered wind does of course become erroneous.

14-7 LOW-ALTITUDE PERFORMANCE AND THE "ALTITUDE HOLE"

A discussion of the performance capability of a doppler radar would not be complete without reference to the problems of *low-altitude performance*, inasmuch as certain systems are limited in this respect. This limitation can occur in certain pulse and FM-CW systems and takes two forms:

1. Insufficient or no usable signal because of certain range effects.
2. Some deterioration in accuracy incident to weighting and skewing of the spectrum because signal returns from certain ranges are de-emphasized.

Single-antenna continuous wave (CW) systems exhibit neither of the two low-altitude limitations mentioned above, since transmission and reception occur at all times. Dual-antenna CW systems can exhibit an effect similar to that of (2) above at extremely low altitudes because of the physical spacing of the antennas; however, the resultant error should normally be small.

Pulse systems, particularly those using TR-ATR (Transmit-Receive-Anti-Transmit-Receive) duplexing, can experience both of the limitations mentioned above. At very low altitudes the receiver gate simply does not have time to open before the entire signal pulse has returned, primarily because of the finite TR recovery time. For this reason typical pulse systems of this type are limited in low-altitude operation to near 200 ft. For long distance navigation this is normally not a serious limitation. The second limitation, the accuracy deterioration, also exists in these systems, since spectrum weighting occurs at the lower altitudes incident to the short-signal return time. The reason for this spectrum weighting is that fewer returns from the close-in ranges than from the farther-out ranges (of the beam ground intersection) have a chance to arrive at the receiver after the receiver gate has opened. Hence the doppler spectrum is unduly skewed and its center of gravity shifted somewhat, resulting in an error in velocity measurement.

The situation is somewhat different in certain coherent pulse systems and FM-CW systems. Since TR-ATR duplexing is not used in typical coherent pulse systems, it is possible to overlap intentionally the transmitting and receiving gates slightly below a given altitude (say 1000 ft) in order to assure sufficient signal return down to zero altitude. Only a small overlap is required, since the necessary transmitter power at such low altitudes is extremely small. Effectively, the system is a partial CW system in this overlap mode of operation; however, because of the low altitudes involved, transmitter-receiver leakage is small compared with the back-scattered signal. This type of system requires a very wide dynamic-range receiver, critical gating adjustments, and carrier-elimination filters. If these requirements are fulfilled, the partial CW (overlap) scheme can eliminate the low-altitude signal problem and permit zero-altitude operation. If the overlap is sufficiently long, considering the range differences of the beam extremities, the low-altitude accuracy deterioration effect can also be avoided.

In FM-CW systems, very low-altitude operation becomes difficult because such systems have the property of inherently discriminating against

nearby echoes on a frequency basis, much as pulse systems do on a time basis. In FM-CW systems the shape of the signal versus range function takes the form of the product of the Bessel function (of the same order as the selected sideband), the sine function (of a period which is half of the wavelength of the modulation frequency), and the normal range spread signal reduction (usually the inverse square of the range). Hence, operation to very low altitudes can be obtained if a sufficiently high modulation frequency is used (several megacycles per second). However, the use of such high modulation frequencies results in a relatively large number of "altitude holes" (to be discussed below) within a given operational altitude range. Due to the above-mentioned shape of the signal versus range function in FM-CW systems, the velocity accuracy can decrease somewhat through the undesirable weighting of signal returns from certain ranges, thereby affecting the shape of the doppler spectrum.

Both of the low-altitude phenomena of pulse and FM-CW systems discussed above are similar and related to the so-called "altitude hole" phenomenon in these systems. In fact, the low altitude region is often referred to as the "first-order altitude hole."

In pulse systems, the *altitude hole phenomenon* occurs at the periodic altitudes at which the return pulse arrives near or at the time of the next transmission pulse and can therefore be completely blanked out, if special steps are not taken. The spectrum weighting error again occurs at and near these altitudes, because of the gating-out process at certain ranges. Many of the pulse systems employ a wobbled (frequency-modulated) pulse repetition frequency (PRF) to reduce these altitude hole effects. In other words, the pulse period is continuously changed (at a low rate) so that some return signal is received at all altitudes. By this means, the signal effect of the altitude hole phenomenon is largely eliminated and the accuracy deterioration effect is greatly reduced.

In some of the coherent pulse systems the pulse repetition rate is changed in discrete steps at periodic altitudes, on the basis of barometric altimeter information. For instance, the repetition rate is halved at every octave of altitude (double the altitude) in order to yield a sufficiently long received return signal at all times. The dependence of the operation of such a system on barometric altitude information is a disadvantage, however, giving rise to the possibility of operational difficulties over mountainous terrains.

In FM-CW systems, altitude holes occur at the periodic altitudes corresponding to multiples of the half-wavelength of the modulation frequency as a result of the periodic signal versus range function discussed earlier in this paragraph. Similar to the case of varying the pulse repetition frequency in pulse systems, changes in the modulation frequency are used in FM-CW systems in order to eliminate the altitude hole problems.

14-8 DOPPLER NAVIGATION SYSTEM PERFORMANCE DATA

One doppler navigation system for which a large number of actual test data exist is the AN/APN-66, probably the first of all doppler navigation systems. A histogram of one set of test flights by Wright Air Development Center, Dayton, Ohio, of the AN/APN-66(XA-2) Doppler Navigation System over land is shown in Fig. 14-12. From this histogram the so-called

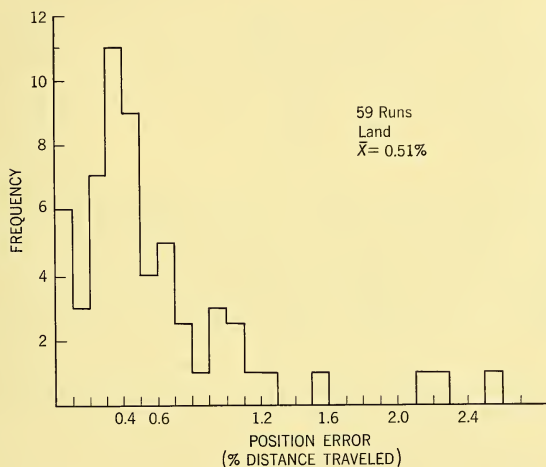


FIG. 14-12 Flight Test Histogram AN/APN-66 (XA-2). After R. M. Gustin, "Flight Test Results of Navigation Set, Radar AN/APN-66 and AN/APN-82," WADC Tech. Note 55-746.

most probable position error or the *mode* of the curve, is found to be 0.42 per cent and the mean error is 0.51 per cent of distance traveled. When these data are processed statistically, a fitted probability distribution curve can be drawn, and this is shown on Fig. 14-13. This is a plot of the cumulative probability distribution of the circular position error of the AN/APN-66(XA-2). From this curve it is found that the "probable error" (50 per cent probability) is 0.52 per cent of distance traveled, and the "maximum error" (95 per cent probability) is 1.2 per cent of distance traveled, giving a ratio of maximum to probable error of 2.31. Thus the data on the AN/APN-66(XA-2) closely fit the case of $\sigma_T/\sigma_R = 2$, which was cited as a typical example earlier.

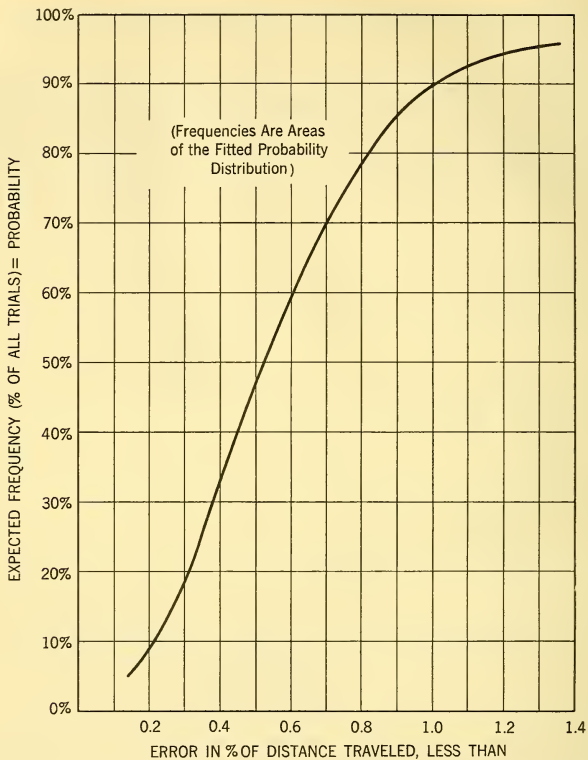


FIG. 14-13 Cumulative Frequencies of Radial Position Errors AN/APN-66.

Over water the mean position error of the AN/APN-66(XA-2) was 1.17 per cent of distance traveled, in contrast to the mean error over land of 0.51 per cent of distance traveled. The largest position error experienced over water in this series was 2 per cent. It is interesting to note that the AN/APN-66 has a beamwidth, in the γ direction, of near 3° , but does not have a land-sea switch facility. Most later systems have a wider beamwidth (usually near 5°), but they have a land-sea switch incorporated in them, which removes a large part of the over-water bias error.

One interesting and important aspect of doppler navigation system performance is the fact that the position accuracy (in percentage of distance traveled) improves for longer flight distances. Apparently the additional

smoothing afforded by these longer flights decreases the effect of certain slowly varying errors, such as compass and variation errors, and the errors in fix measurement and lack of knowledge of the flight end points. An example of this is the data shown in Table 14-1, taken from the test flights of the AN/APN-66(XA-2) discussed above, as broken down into three different ranges of flight length. It is interesting to note that the mean error for the longer flights of 1422 miles average length is approximately half as large (0.45 per cent) as the mean error for the shorter flights of 145 miles average length (0.86 per cent).

TABLE 14-1 ACCURACY CHARACTERISTICS OF AN/APN-66
FOR 59 OVERLAND RUNS

Range of Distance, Miles	Average Length, Miles	Mean Position Error, %
0-200	145	0.86
201-650	486	0.64
651-2500	1422	0.45

Another series of flights, consisting of 766 flight legs with the same equipment, provided position error data which were broken down into range and transverse errors, as defined in Equations 14-8 through 14-11 of Sec. 14-3. On the basis of these data the constants which relate these errors to the total distance traveled and which were defined as K_R and K_T in Equation 14-11 were determined (for that equipment) to be 8 and 16, respectively, for D expressed in nautical miles.

One further demonstration of the significant effect of the heading reference and magnetic variation information on system performance is shown by the results obtained by Wright Air Development Center, Dayton, Ohio, with Radar Navigation Set AN/APN-82, which is composed of the same *doppler radar* as the AN/APN-66 (the AN/APN-81) and Navigational Computer AN/ASN-6 (with only a manual variation capability) and which was tested with a standard N-1 magnetic compass. As discussed earlier in this paragraph, the mean position error over land of the AN/APN-66 (with automatic variation and an improved N-1 heading reference) was 0.51 per cent of distance traveled, whereas it was 1.5 per cent of distance traveled for the AN/APN-82. The difference can be attributed largely to the variation and heading errors, and, perhaps to a lesser degree, to the difference in the type of computer used on those two systems.

Modern doppler radars are capable of accuracy performance of the order of less than 0.1 to 0.2 per cent probable and approximately 0.5 per cent maximum error over land. Over-water performance should be improved by the use of land-sea switches and new automatic techniques, and the remaining water bias error should increase the maximum velocity error to somewhere less than 1 per cent, depending largely on antenna size. Computers

with 0.1 to 0.5 per cent accuracy are available. The heading reference appears as the weakest link in the system. Effort to improve heading reference continues, and the potential use of earth rate directional references (north-seeking gyrocompasses) for the navigation system holds great promise. Heading reference equipment with a maximum long term heading error of less than 0.75° should be available in the future. On this basis one can conclude that doppler navigation systems are now and will be capable, over reasonably long flights, of probable position errors (50 per cent probability) near 0.65 per cent of distance traveled and a maximum error (95 per cent probability) of approximately 1.5 per cent of distance traveled over land and average sea state. Neglecting nonsystematic water motion effects, which should be of importance only for low-speed aircraft, the maximum position error over extreme sea states should generally be somewhere below 2 per cent of distance traveled.

Altitude performance of doppler radar systems is a function of the available microwave power and the type of transmitter-receiver design (e.g. pulse or CW). It appears that most modern doppler radars, by a combination of these two factors (i.e. sufficiently high power and special design technique) satisfy the maximum altitude requirements of both military and civilian aircraft.

Modern doppler navigation systems compute and display ground speed, drift angle, present position, ground track being made good, course to destination, and distance to destination and can also provide an autopilot steering signal. Some of the systems also continuously compute and/or display wind speed and wind direction. Systems that do this usually have a wind memory feature which allows the system to operate from the true air speed and "last remembered wind" when the doppler signal goes below a given threshold for any reason (e.g. very smooth water). Many modern systems employ great-circle course and distance computers and are capable of operating in conjunction with either of the three types of heading references discussed earlier.

It bears re-emphasizing that the performance of a complete doppler navigation system is a direct function of the performance characteristics of all three of the major components of the system — the doppler radar, the computer, and the heading reference — and that the overall system accuracy is no better than that of the least accurate of the three components.

In some cases, the interconnection of a doppler radar and an inertial platform, in a so-called doppler-inertial system, becomes advantageous from the viewpoint of overall system performance. For example, for some applications extremely accurate short-term velocity information is required, in addition to accurate vertical attitude and long-term velocity information. In this case, the interconnection of a doppler radar and an inertial platform

represents an ideal solution. Such a system combines, in a complementary fashion, the highly accurate long-term velocity information furnished by the doppler radar, which is also useful for airborne erection and damping of the inertial system, and the highly accurate short-term velocity and vertical information of the inertial platform, which can, in turn, aid the doppler radar tracking and provide dynamic inertial memory. Hence, such a combined system is capable of providing very accurate velocity and vertical information (see references 15 and 16 of Bibliography at end of this chapter).

In another important example, if highly accurate position, and hence directional, information is required for long periods, as in the case of long-range vehicles, an ideal doppler-inertial interconnection is a system which employs a doppler radar as part of a so-called doppler inertial compass, i.e., with a platform which is interconnected so as to operate in a north-seeking gyrocompass mode. In this mode of operation, the platform is kept pointed north at all times. The necessary vehicle ground velocity information is of course available in the system from the doppler radar. Such a system provides accurate information on long-term velocity from the doppler radar and short term velocity, and verticality and long-term azimuth information from the inertial platform. Thus it permits highly accurate determination of position for long ranges and periods of time (reference 16).

Another possible combined system configuration is a so-called doppler-inertial-stellar system which uses a doppler radar for long-term velocity, an inertial platform for short-term velocity, and an astrocompass for accurate directional information and position correction. Various forms of such self-contained combined systems using doppler radar have found application in military vehicles.

Doppler radars can also be used advantageously in other forms of combined systems, such as in conjunction with certain ground-based radio guidance systems. In such systems, the doppler radar provides continuity of information but with an error which accumulates with distance, whereas the ground-based radio system provides intermittent, but instantaneously highly accurate, position fix information (references 14 and 17).

It may be said in conclusion that the next ten to twenty years should see tremendous application of doppler radar navigation systems in military as well as in civil aviation.

14-9 INTRODUCTION TO WEATHER RADAR

Generally speaking, the designer of a radar system is concerned with meteorological factors at radar wavelengths only insofar as they cause detrimental effects to the prime function of the system. The designer of weather mapping or weather avoidance radar, on the other hand, can profit a great deal from an understanding of some of the basic meteorological

factors as they affect radar signals. Even where weather is not related to a system's prime function, the designer would do well to have some appreciation of radar meteorology, if only to safeguard his design from its detrimental effects.

This section, therefore, although primarily concerned with the airborne weather radar system, will include some discussion of the nature of the meteorological disturbances which are its targets and will describe the utilization of special circuitry designed to display most efficiently the information available in these targets.

14-10 METEOROLOGICAL EFFECTS AT MICROWAVE FREQUENCIES

Meteorological Scatterers. One of the effects of precipitation particles suspended in the atmosphere (see Paragraph 4-16) is back-scattering of the radar energy.⁴ The degree of backscattering is usually given in terms of the *back-scattering coefficient* n . By definition n is the area of an isotropic scatterer which returns an amount of energy in the direction of the receiving antenna equivalent to that returned by a unit volume of the specific weather target. When n is multiplied by the volume of precipitation particles illuminated by the radar beam, the resultant gives the total back-scattered energy. When wavelength (λ) is long compared with the scatterer diameter D (droplets are assumed spherical), the Rayleigh scattering approximation gives n in terms of

$$n = \frac{\pi^6 |K|^2 \Sigma D^6}{\lambda^4} \quad (14-14)$$

The expression ΣD^6 is the summation of the sixth power of the diameters of the scatterers being illuminated and $|K|^2$ is a factor related to the complex refractive index of the particles. A number of important considerations can be drawn from this expression.

1. Target returns increase as a fourth power of frequency.

2. Target returns vary as the sixth power of drop diameter for spherical droplets; therefore it is to be expected that rain would provide a much better target than clouds or fog, particularly at frequencies where $\lambda \gg D$. Since the determination of drop diameters becomes extremely cumbersome (see Paragraph 4-16) it is desirable to convert the expression to an equivalent one in terms of R , rainfall rate in mm hr^{-1} , a parameter much more readily measured:⁵

$$\Sigma D^6 = kR^{1.6} \quad (14-15)$$

⁴For an excellent study of meteorological echoes the reader is referred to D. E. Kerr, (Ed.), *Propagation of Short Radio Waves*, McGraw-Hill Book Company, Inc., New York, 1951.

⁵D. Atlas and J. S. Marshall, *Weather Effects on Radar*, Air Force Surveys in Geophysics No. 23, Air Force Cambridge Research Center, December 1952.

This presumes a given size distribution of drops based on the work of Laws and Parsons⁶ and indicates n varying with the 1.6 power of rainfall rate for this one specific case. There is some evidence to indicate that this figure represents a common rainfall condition. The symbol k is a proportionality constant.

3. The factor $|K|^2$ is about five times larger for water than for ice. This results in much stronger returns from rain than from dry snow, for equivalent masses. However, as ice particles fall through warmer layers and melting water forms on their surface, they very rapidly take on the $|K|^2$ value of water. Particularly in the case of hail falling through the melting zone, very high values of back-scattering occur because of the large diameters (D) involved. This gives rise to the so called *bright band* effect mentioned in Sec. 4-16.

Another interesting facet of meteorological scatterers is the effect of shape on back-scattering. Equation 14-14 is based on spherical hydrometeors. As these particles become elongated or flattened out, the value of n increases, assuming a random orientation of the spheroids. As a result, snow or ice particles having distorted shapes often give strong returns as they fall through the melting band, before they form into more uniform rain droplets.

Rain droplets, if large enough in size, flatten as they fall, resulting in somewhat greater target returns for horizontally polarized energy impinging upon them than from vertically polarized energy.

Meteorological Attenuators. Weather hydrometeors are not only to be considered from the standpoint of reflective scatterers, but also as attenuators of the radar signal when they happen to lie between the radar antenna and a desired target. Part of this attenuation is caused by diffusion and part by absorption of the microwave energy. This attenuation is usually designated *round trip attenuation* (γ) and is in terms of decibels per unit distance of attenuating media. γ can be expressed as a function

$$\gamma = f(\lambda, M) \quad (14-16)$$

where M is in terms of water density, gm/m³. In the case of rainfall

$$\gamma = AR^b \quad (14-17)$$

where R is the equivalent rainfall rate in mm hr⁻¹ and A is a parameter that varies with wavelength and temperature. For the case of rain, the exponential b is essentially unity for wavelengths of 10 cm and longer and increases to about 1.3 at $\lambda = 3$ cm.

As was the case with the scattering coefficient n , attenuation γ increases rapidly with frequency and is much greater for rain than for snow, fog, or

⁶J. O. Laws and D. A. Parsons, "The Relation of Raindrop Size to Intensity," *Trans. Am. Geophys. Union* **24**, part 2, 1943.

clouds (owing to the lesser equivalent water density of the latter hydrometeors). Values of attenuation have been calculated, but experimentally determined values have been difficult to collect and correlate either with each other or with the theoretical values. Fig. 14-14 gives some typical

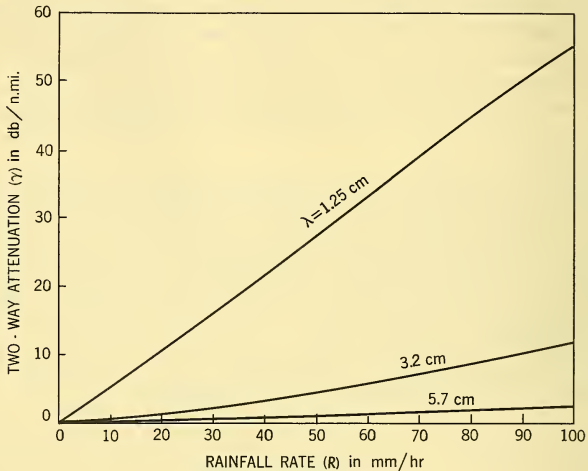


FIG. 14-14 Theoretical Curves of Two-Way Attenuation γ Versus Rainfall Rate (R). Summarization of Data Derived by Ryde and Ryde "Attenuation of Centimeter and Millimeter Waves by Rain, Hail, Fog, and Clouds," GEC Report No. 8670 (1954).

theoretical values for two-way attenuation γ and illustrates the interrelation of attenuation and rainfall rate. It is implied in these curves that a given theoretical drop size distribution is typical for weather targets.

The nature of meteorological scatterers has just been discussed in a general way. It is also important in the design of weather radar to have some knowledge of the nature of weather disturbances and how they might affect the utilization of the radar. Most of this discussion will be devoted to the thunderstorm problem for two reasons. It creates one of the most severe turbulence problems for the pilot, and it is quite commonplace. Tornadoes and hurricanes are obviously more dangerous, but not nearly so often encountered.

Thunderstorms and heavy showers normally stand out well, even among large stable rainfall areas because of the high back-scattering of heavy

rainfall rates. They often group themselves in bands of cells along a frontal or squall line, building up and dissipating rapidly, only to build up in new cells. Within a cell it is quite normal to have extremely high rainfall gradients between the dry air just outside the leading edge of the cell and the heavy rain within. This is often accompanied by high vertical shear winds, updrafts predominating on the leading edges of the storm cell and downdrafts in the trailing areas of the cell.⁷ For this reason, the most dangerous turbulence, insofar as aircraft are concerned, tends to be found where this rainfall gradient is the sharpest. Another distinguishing characteristic of the thunderstorm is the tremendous heights they often reach (as much as forty or fifty thousand feet) during their maturing stage. This type of storm activity can often be identified by radar by tilting the antenna upwards. Because of the preceding factors it is highly desirable for the radar system to be capable of distinguishing varying levels of target returns and to possess sufficient definition of display to make it possible for the observer not only to identify storm cells but also to pinpoint areas where high turbulence is most likely.

Areas of general rainfall, often associated with warm fronts, do not normally provide such good targets as do thundershowers, since n is so much less owing to the considerably lower rainfall rates and scatterers of generally smaller diameter. These regions likewise cause much less attenuation of the microwave energy. At wavelengths of 10 cm these scatterers will often be indiscernible with the exception of small cells of more intense activity, which will sometimes occur within the overall rain region. At 3.2 cm this same weather activity will often give an area of snowy appearance on the radar indicator.

Because of the damage that hail is capable of doing on the aircraft, it is only natural for the pilot to be extremely interested in a method of avoiding this meteorological phenomenon through use of weather radar. Unfortunately, although hail normally gives good radar returns thanks to the large diameters D of its scatterers as already noted, it is always associated with storm activity which likewise gives strong returns, thus no accurate differentiation is possible on a back-scatter basis. There is some evidence, however, collected during an evaluation of airborne radar by United Airlines,⁸ to indicate that hail is often accompanied by a distinctive hook or finger extending out of the displayed parent storm cell. Sometimes the display exhibits a scalloped edge. The utilization of such radar pictures for hail avoidance would obviously require a great deal of operator "feel" for interpreting the display.

⁷R. F. Jones, *Radar Echoes from and Turbulence within Cumulus and Cumulonimbus Clouds*, Prof. Notes 109, Meteorological Office, Air Ministry, London, 1954.

⁸H. T. Harrison and E. A. Post, *Evaluation of C-Band (5.5-cm) Airborne Weather Radar*, United Airlines, Inc., Denver, Colorado, 1954.

14-11 DESIGNING AIRBORNE RADAR SYSTEMS EXPLICITLY FOR WEATHER MAPPING

The designer of airborne weather radar systems must be fully cognizant of a very important principle. With the one exception of purely weather reconnaissance radar, the airborne system is a highly valuable but not indispensable tool to the aircraft operator. Unless the added weight, space, and power consumption taken up by the system can be justified through a realization of an improvement in flight safety, flight time, passenger comfort, or the accomplishment of an otherwise doubtful mission, the economics of flight make its inclusion prohibitive. This implies a system which is versatile, accurate, reliable, and capable of providing the pilot with sufficient usable weather information, while at the same time eliminating all unnecessary frills or refinements.

This principle militates against the design of a limited system which might be only an impending weather warning device, or a very limited-range system. To be of value, the radar must have a range commensurate with the speed of the user aircraft. Information must be made available in sufficient time to allow the pilot to make well-considered decisions and take corrective actions. This may require a range of only 30 miles, for aircraft with cruising speeds of the order of 120 mph, or a range of 150 miles for speeds approaching Mach 1.

Generally, airborne weather radars require reasonably good definition, accurate range calibration, simplicity of operation, and inherent high reliability under adverse service conditions. Also highly desirable is antenna stabilization against pitch and roll, climb, and bank. In order to increase its versatility and utilization, considerable attention has been directed toward means of adapting the radar to ground mapping, terrain avoidance, drift angle determination, and radar beaconry.

Among the refinements normally given up in the interest of simplicity and small size are special scope displays, variable scan rates, variable pulse lengths, and the high-peak transmitter power levels often associated with ground weather radars. Operating frequency must also be compromised in the interest of size, weight, and power conservation. In order to obtain a better understanding of the interrelation of the many factors involved in optimizing a system design, it is well to review some of the basic radar equations.

14-12 MODIFYING THE RADAR RANGE EQUATION FOR THE WEATHER PROBLEM

The form of the radar equation most generally used (see Paragraph 3-2) equates the power reflected to the receiver to the balance of the radar parameters:

$$P_R = \left(\frac{P_T G}{4\pi R^2} \right) \left(\frac{\sigma}{4\pi R^2} \right) \left(\frac{G\lambda^2}{4\pi} \right) \quad (14-18)$$

where P_R = reflected power returned to the receiver

P_T = power delivered to the antenna from the transmitter

R = range to radar target

G = antenna gain (assuming that the same antenna is used for both transmitting and receiving, as is usually the case)

λ = wavelength

σ = scattering cross section of target.

The antenna gain G can be related to the physical dimensions of the antenna through the expression

$$G_0 = \frac{4\pi A m}{\lambda^2} \quad (14-19)$$

where A is the aperture area, G_0 is maximum gain at nose of beam, and m is a fractional quantity relating the actual gain to that of a theoretical antenna having a spherical phase front and a uniform amplitude field impinging on the whole aperture. Therefore, Equation 14-18 can be simplified and rearranged in terms of maximum attainable range for a specified minimum usable power into the receiver (P_{Rmin}):

$$R_{max} = \sqrt[4]{\frac{P_T (A_{eff})^2 \sigma}{4\pi (P_{Rmin}) \lambda^2}} \quad (14-20)$$

A moment's study of Equation 14-20 would indicate that the designer has control over at least four variables — transmitter power, antenna size, minimum usable receiver power (MDS), and frequency; with a fifth variable σ not quite so well understood, since it is the contribution made by the target — in this case a meteorological disturbance. These targets take on special significance for two reasons. First, the radiated energy may have to pass through accumulations of hydrometeors, suffering attenuation during its round trip as discussed in Paragraph 14-10. Then the target itself always has appreciable cross-sectional area and depth; storms can never be classed as point targets. If the area of a meteorological target exceeds that of the intersected radar beam (a beam-filling target) and the target depth exceeds half the distance occupied by the pulse as it propagates through space, σ can be expressed in terms of⁹

$$\sigma = R^2 \phi \theta \frac{\tau c}{2} \cdot n = \frac{n}{2} \times \text{volume occupied by radiated pulse at range } R \quad (14-21)$$

⁹D. E. Kerr, (Ed.), *Propagation of Short Radio Waves*, McGraw-Hill Book Co., Inc., New York, 1951.

where $R^2\phi\theta$ = illuminated area at range R for an antenna of half power beamwidths of ϕ and θ

τ = pulse length

c = velocity of propagation of light

n = scattering cross section per unit volume of target (Equation 14-14).

If attenuation due to intervening media γ is now given as the summation of attenuations of i number of individual cells of path lengths L_i , Equation 14-18 may be rewritten¹⁰:

$$P_R = \frac{P_T G^2 \lambda^2}{(4\pi)^3 R^4} \left(R^2 \phi \theta \frac{\tau c}{2} n \right) \left(10^{-0.1 \sum_i \int \gamma dL_i} \right). \quad (14-22)$$

Several important conclusions can be drawn from this equation. First, reflected power varies directly with pulse length τ . This derives from the purely incoherent nature of the replies from weather scatterers. Total return energy will continue to increase for a period equal to τ as back-scattering from within the target reinforces that from the leading edge. Also, as target size approaches or exceeds that of the illuminated area (a beam-filling target) the R^2 term in the numerator reduces the equation for received power to an inverse function of the square of the range rather than the fourth power of range usually associated with radar targets. Thus we see that the designer does have some control over σ in Equation 14-20, through the relation of frequency and antenna size to beamwidths and through control of pulse width τ .

14-13 RELATIVE IMPORTANCE OF DESIGN VARIABLES IN AIRBORNE WEATHER RADAR

Frequency. This factor is considered first because it figures in so many facets of the problem. Antenna gain G varies inversely with λ^2 ; beamwidths become narrower with increased frequency; and back-scattering n increases as a fourth power of frequency. Likewise, losses caused by attenuating media are highly frequency-sensitive (note Fig. 14-14). The designer must also consider the size and weight increase of RF components at lower frequencies. Because optimizing each of these factors is not mutually compatible, the choice of frequency of necessity becomes a compromise. As a result, most air-weather radars today operate at either C or X band (usually 5400 Mc and 9375 Mc respectively).

¹⁰J. C. Johnson, *Meteorological Factors and Their Effects on Microwave Propagation*, Technical Memorandum 412, Hughes Aircraft Co., 1955.

To illustrate the extent of this compromise, let us assume that transmitter power and antenna size are fixed by other considerations. Utilizing Equations 14-14, 14-19, and 14-22, we see that a 2-to-1 increase in frequency provides a 64-to-1 (18 db) increase in return power for clear air propagation ($\gamma = 0$). Since we are interested in maximum system range, and storm cell areas are generally a fraction of that required for beam filling at such ranges, (target area $\ll R^2\phi\theta$), we can set

$$R_{\max} \propto \sqrt[4]{P_R}. \quad (14-23)$$

In other words, doubling the frequency increases the range 2.8 times for clear-air propagation. This conclusion is modified appreciably when intervening meteorological attenuators are considered (see Fig. 14-14).

Obviously if system and antenna size, weight, and power consumption are of little consequence, a lower frequency would be more desirable. Since this cannot be the case for airborne radar, the band of frequencies between C band and X band are found to be the best compromise. The lower frequency end is dictated principally by the required target definition (nominally 8° maximum beamwidth) which is a function of antenna dish size and frequency; and the size, weight, and power consumption of a transmitter required to give sufficient range. On the other hand, the high frequency end is limited by the significance attached to the ability to penetrate extensive heavy-rainfall areas, and the distortion in displayed storm cells, caused by two-way attenuation through the cell. This attenuation could reach as high as 30 to 40 db at a wavelength of 3.2 cm for a particularly severe storm situation. The comparable attenuation at 5.7 cm would be 8 db.

Peak Transmitter Power. This parameter must be construed as peak magnetron output power modified by losses in the duplexer and other waveguide components. Because high power output can only be purchased through larger, heavier magnetrons, modulators, and power supplies, this is normally a highly undesirable means of improving airborne radar performance. Equation 14-20 indicates that maximum range varies with the fourth root of peak power (P_T). Inasmuch as the weather targets at the maximum range of most systems probably lie in a classification somewhere between that of a point target and a beam-filling target, this expression may be more nearly represented as a cube root function. A peak power increase of 2 to 1 therefore might only be expected to improve range by 19 to 26 per cent. In the discussion above on frequency it was shown that return power drops very rapidly with lowering of frequency. If peak transmitted power (P_T) is to be increased to counteract this decrease in P_R it is readily seen that the maximum permissible size and weight of the transmitter soon limits how low in frequency the designer may go.

C-band weather radars are usually designed with at least 3 db more peak transmitted power than their X-band equivalents in order to match more nearly the X-band range under light attenuation conditions, while further enhancing their superiority under very heavy rainfall conditions.

Antenna Size. Because the antenna size enters into the range equation twice, affecting both transmitting antenna gain and receiving antenna cross-sectional area, it becomes one of the most effective parameters for increasing range. From Equation 14-18 we see that power into the receiver is proportional to the square of antenna gain G , which is in turn proportional to the square of the antenna diameter, thus making overall system gain vary with the fourth power of the diameter (d^4). Since maximum range is proportional to the fourth root of power, it is apparent that system range varies directly with dish diameter (again assuming non-beam-filling targets).

Pulse Length. As seen in Equation 14-21, back-scattered energy returned to the antenna from weather targets is directly proportional to pulse length. Doubling the pulse length should increase the power return by 3 db, assuming that the target is greater than $\tau c/2$ in depth. A practical limit of about 5 μ sec is placed on this factor by the minimum range resolution that can be tolerated.

Minimum Usable Receiver Power (Often referred to as *Minimum discernible signal*). This is principally a function of the noise figure of the receiver. Because of the requirements for a simple, rugged receiver it is not generally feasible to obtain a noise figure for the airborne equipment comparable with its large ground-based counterpart. A great deal of work has been done and considerably more needs to be done to reduce this figure. IF strips can now be designed having noise figures sufficiently low to make no sizable contribution to the overall noise figure. Crystal mixer characteristics and waveguide losses are being improved to the point where a 9-db receiver noise figure is feasible on a production basis utilizing conventional techniques. Furthermore, the development of small solid-state devices, capable of amplification at microwave frequencies, is making possible still greater reductions in noise figure.

Another factor over which the designer has cognizance is receiver bandwidth. Reducing receiver bandwidth, so long as it is wide enough to pass the main components of the pulse spectrum, improves the signal-to-noise ratio.

One last factor to be considered for its effect on radar performance is the ability of the scope indicator to discriminate minimum discernible targets from noise incident to the integration effect provided by tube persistence. Noise, being random, does not integrate to the same extent as do repetitious

targets. This action can be heightened by increasing pulse rate or slowing scan rate, the first being limited by the maximum range to be displayed, the second by tube persistence and rate at which the display changes. The new storage tube indicators will give some small improvement incident to their greater integration action (see Chapter 12).

14-14 DESIGN FEATURES

Inasmuch as servicing and adjustment cannot normally be made in the aircraft in flight and reliability is of utmost importance (dual installations being highly impractical), the design must be simple and straightforward in its approach. Critical adjustments and circuits should be avoided if at all possible. For this reason virtually all present-day weather radars utilize more or less standard design techniques. However, a few special design features are highly desirable in order to enhance the system's utilization, principally for weather mapping. Some of the more important of these are the following:

Iso-Echo Contour Circuitry. As mentioned in Sec. 14-13, one of the most valuable pieces of weather information from the pilot's standpoint is knowledge of areas of highest rainfall gradient and rainfall rate. The *iso-echo contour circuit* is a video inversion device which causes a black hole to occur within a displayed storm cell wherever the rainfall rate is above a predetermined level, as determined from relative back-scattering (see Fig. 14-15). The thinner the line between the outer edge of the displayed cell and the contour "hole," the higher is the rainfall gradient, thus indicating areas the pilot definitely should avoid.

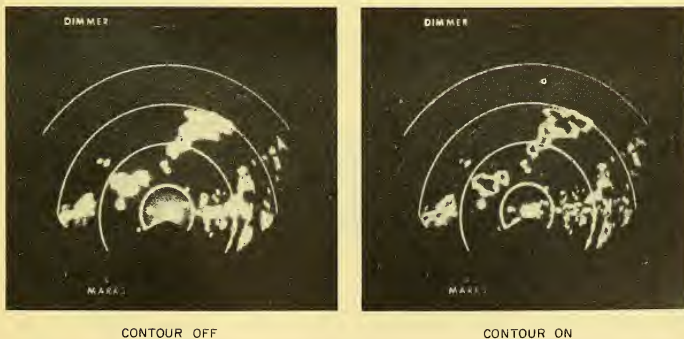


FIG. 14-15 Effect of Iso-Echo Contouring On Storm Display.

STC. *Sensitivity time control* is extremely useful in weather mapping. This circuit operates on the receiver sensitivity in such a manner as to maintain constant returns from a constant target for a moderate range of distances. Its inclusion is virtually mandatory if iso-echo contouring is to have any significance.

AFC. *Automatic frequency control* is generally desirable, particularly in that it frees the operator from receiver tuning for more important tasks. It needs to be sufficiently precise and fast to follow any predictable transmitter frequency variation. The advent of small ferrite load isolators (see Sec. 10-16), as well as ferrite duplexers with their inherent load isolation, has alleviated this problem by buffering the magnetron against load mismatch fluctuations caused by the antenna as it rotates.

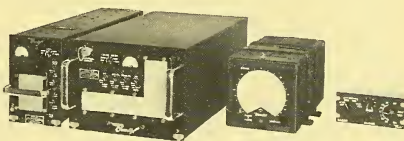
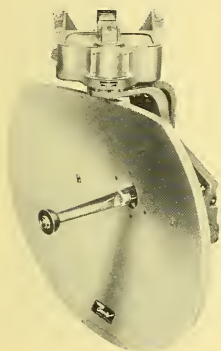
Antenna Stabilization. The stabilization of the antenna against pitching and rolling of the aircraft is generally desirable. It prevents a smearing of the display when the aircraft is in a climb or turn and is particularly valuable if the radar is also to be used for ground mapping. The two systems most generally used for this function are *line of sight stabilization* and *platform stabilization*. Line-of-sight stabilization utilizes the information derived from the plane's vertical gyro to operate on the antenna tilt servo system in such a manner as to compensate for variations in the plane's aspect. Platform stabilization essentially floats the active portion of the antenna free of the airframe and, through its own separate gyro, tends to maintain this platform level with the horizon regardless of the motion of the plane. Of the two systems, platform stabilization is inherently more accurate, but is considerably more complex and requires more costly parts and components. Stabilization accuracy must be related to antenna vertical beamwidth, which is in turn a function of dish size and frequency. If care is exercised in design and utilization, line-of-sight stabilization is generally considered adequate for beamwidths down to 3° .

Csc² Antenna Pattern. In order to improve the radar's utilization for ground mapping, several X-band systems have included means of spoiling the antenna's normal pencil beam into a fan-shaped beam in the elevation cross section, following essentially a cosecant squared radiation pattern over the vertical angle of interest. This provides approximately equal power return from ground scatterers over a considerable distance to give a uniform mapping of ground. A typical method of accomplishing this is to provide the antenna with a spoiler grid and a means of rotating the polarization of the microwave energy relative to the direction of the grid lines. When polarization is perpendicular to the grid lines, the spoiler is essentially invisible to the radiated energy and hence a pencil beam results. When parallel, the spoiler is seen by the radiated energy and the pattern is spoiled downward.

Indicators. Considerations of space and simplicity of design fairly well limit the display system to a PPI type indicator, most often utilizing a 5-inch-diameter display tube. It is often found desirable to have dual indicators in a cockpit; these may be completely parallel systems or may have independently selectable ranges. A variety of ranges and accurate range marks are a necessity in any weather radar. Perhaps the most severe problem encountered in the use of the PPI is brought about by the high ambient light conditions so often found in the cockpit. As a result of this, airborne radar has done much to stimulate the development of the storage type indicator tube with its tremendous increase in light output.

C-BAND-5.7 CM

NOMINAL RANGE:	150 N. MI.
WEIGHT:	113 LBS
NOMINAL POWER OUTPUT:	80 KW PEAK
POWER CONSUMPTION:	APPROX. 800 W
ANTENNA BEAMWIDTH:	5.5 DEG



X-BAND-3.2 CM

NOMINAL RANGE:	150 N. MI.
WEIGHT:	89 LBS
NOMINAL POWER OUTPUT:	18 KW PEAK
POWER CONSUMPTION:	500 W
ANTENNA BEAMWIDTH:	4 DEG

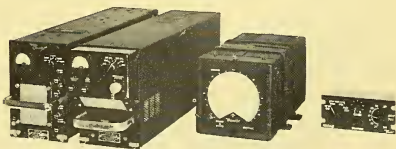
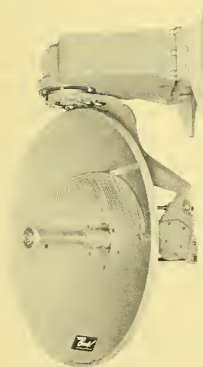


FIG. 14-16 Typical C-Band and X-Band Commercial Air Weather Radar Systems.
(Courtesy, Bendix Radio Division of Bendix Aviation Corporation)

As in the design of any equipment involving compromises, the designer of airborne weather radar must weigh carefully the relative importance of all the many factors concerned in the light of his own specific application, in order to arrive at the optimum design. Figs. 14-16a and 16b give some idea of the relative size, weight, and power consumption of two typical commercial weather radars — one a 3.2-cm system, the other a 5.7-cm system.

14-15 INTRODUCTION TO ACTIVE AIRBORNE GROUND MAPPING SYSTEMS

The mapping of large areas and the capability of providing data which permit rapid and accurate identification, location, and interpretation of all important features within the area surveyed are among the primary requirements of ground mapping systems. Airborne radar ground mapping systems have the potential to satisfy many of these requirements. They combine the advantage of radar, with its ability to operate during the night or day under adverse weather conditions, with the mobility and range of an airborne platform.

The following discussion will outline the basic principles of active airborne ground mapping systems and examine system considerations as well as the major system characteristics and components. The basic radar range equation will be considered as it applies to active ground mapping systems, and factors which determine the limits of resolution will be discussed. Finally, the future possibilities of this type of mapping system will be considered briefly.

14-16 BASIC PRINCIPLES

The basic principle of an active radar ground mapping system is the same as that of any active radar; it transmits energy and detects the part of it scattered back from a target. However, instead of the usual point target, the target in this case is the ground, which can be considered as an extended array of scatterers. The radar map is obtained by scanning or "painting" the ground and displaying the return on a cathode ray tube or photographic film. Since the scattering characteristics of the ground will vary from point to point, the map will be in the form of a varying brightness pattern. Variations of intensity in this brightness pattern can be interpreted in terms of the topographical and man-made features of the terrain. For example, the energy back-scattered from a smooth surface such as calm water will be much less than that from a rough surface such as the ground. As a result land-water boundaries are usually among the most easily identifiable features on a radar map. The degree of correspondence between the brightness pattern and the features of the terrain depends to a considerable

extent on the characteristics of the antenna beam pattern used to paint the ground.

The antenna beam pattern usually employed in ground mapping systems is narrow in one dimension and has wide angular coverage in the other dimension. (This type of beam is known as a *fan beam*; see Fig. 10-6.) It is usually oriented so that the narrow dimension is horizontal, thus illuminating a long narrow strip of ground from beneath the airplane to some maximum range (see Figs. 1-5a and 1-5b.) Thus, for a given pointing direction, the radar beam illuminates (or "paints") targets at many different ranges and depression angles. Variations in the brightness are therefore functions of range and the angle at which the ground is viewed as well as the reflective properties of the terrain. This condition — if not corrected — would complicate the correlation between the radar map and terrain. To compensate for the effects of range and viewing angle, the vertical gain pattern of a radar ground mapping antenna is designed to be a function of the depression angle at which a given patch of ground is viewed. The characteristics of this type of pattern, known as a *cosecant-squared beam*, are discussed in Chapter 10 and Paragraph 14-18 below.

Scanning of the antenna beam is usually accomplished either by rotating the antenna about a vertical axis or by positioning the antenna along the fuselage of the aircraft, so that the motion of the airplane provides the scanning. When the beam is rotated a full 360° about the vertical axis, the map is usually in the form of a plan position map with the airplane at the center. Most systems employing this type of scan use a sector scan, i.e. less than the full 360° . When the sector is directed forward of the airplane, the system is known as a *forward-looking mapping radar*. Systems employing the velocity scanning technique, where the beam is directed to the side of the airplane, are known as *side-looking radars*. The maps obtained with this system are in the form of strip maps along each side of the airplane track (see Fig. 1-5) and are especially adapted to the use of photographic techniques to obtain a permanent record.

Fig. 14-17 shows the simplified block diagram of a typical mapping system. In addition to the usual components — i.e. antenna, transmitter, receiver, and display — of the typical radar system, a recorder is included. In some systems the recorder may be used merely to preserve a permanent record of the

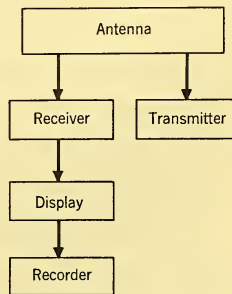


FIG. 14-17 Block Diagram of a Typical Ground-Mapping Radar System.

map as viewed on the display. In other systems the map may be viewed directly on film from a high-speed recorder.

14-17 SYSTEM CONSIDERATIONS

As with all radar systems, the design of a ground mapping radar system involves a series of compromises between requirements and state-of-the-art limitations. Some instances represent a simple choice; others involve a number of system parameters. The choice may represent a trade-off between a desired capability and system size, weight, and complexity, or it may represent a fundamental limitation imposed by laws of nature. Some of the most important system considerations for ground mapping radar systems involve:

1. Resolution
2. Accuracy
3. Range and operational altitude
4. All-weather capability

Just as accuracy is the basic measuring stick of navigation systems performance, resolution is the basic measuring stick of performance for radar ground mapping systems. *Resolution*, a measure of the system's ability to distinguish between closely spaced objects or to delineate the details of a large area, is defined usually in terms of range resolution and transverse or azimuth resolution. While the ultimate resolution attained by the system is a function of many parameters, the single criterion most commonly used to judge it is the pulse packet size as projected on the ground. The system parameters which determine the pulse packet size are antenna beamwidth and pulse length, as measured at the half-power points.

Beamwidth is a function of the transmitted wavelength and the dimensions of the antenna, being expressed approximately by

$$\theta = \frac{K\lambda}{D} \quad (14-24)$$

where θ is the beamwidth defined by the half-power points

λ is the transmitted wavelength

D is the pertinent dimension of the antenna aperture

K is the constant dependent on the particular aperture. A typical value for this constant is 70 where θ is expressed in degrees, and λ and D are measured in the same units.

The system designer is confronted at once with a compromise, both in the selection of the transmitted wavelength and the dimensions of the antenna. To narrow the beamwidth, either the wavelength must be decreased or the dimensions of the antenna must be increased. In decreasing the wavelength

the problems of atmospheric attenuation (see Chapter 4) and the generation of large amounts of power become increasingly difficult. The maximum size of the antenna obviously will be limited for airborne installations.

Range resolution may be improved by decreasing the pulse length. Here again there is a minimum limit, since the average power transmitted is a direct function of pulse length. Also to be considered in this connection is the altitude at which the system will operate, since the length of the pulse, as projected on the ground, is a function of the radar altitude as shown in Fig. 14-18. The packet dimensions for a typical beamwidth and pulse

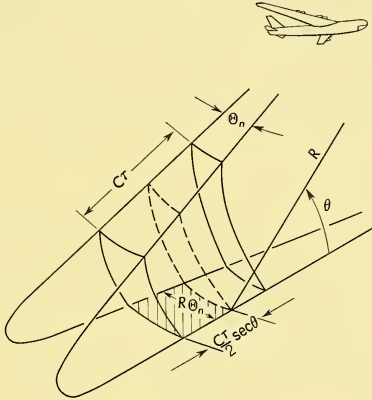


FIG. 14-18 Pulse Packet Dimensions.

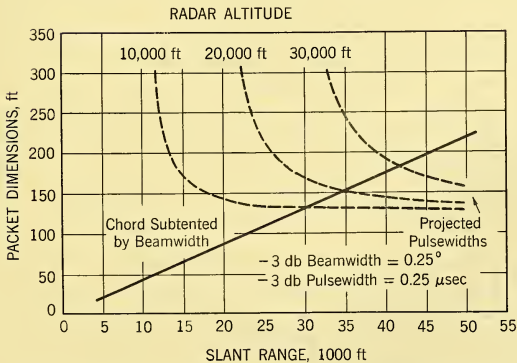


FIG. 14-19 -3-dB Packet Dimensions vs. Slant Range.

length are shown in Fig. 14-19 as a function of slant range to the ground and radar altitude.

Another important area with regard to system resolution concerns the receiver and display. Here, one of the basic considerations is their dynamic range, i.e. the range of signal amplitudes which they can accommodate without distortion. Since a radar map is a brightness pattern, variations of intensity within the map contribute to the resolution of certain features. The range of signal amplitudes encountered in a given mapping operation may be quite large. Recent data taken from a number of different types of terrains indicate that the range will be of the order of 30 db or more.¹¹ If the receiver or display system cannot accommodate such a range of signal amplitudes without distortion, loss of much detail within the map will occur. If the gain or intensity is set at the noise threshold, strong targets will "bloom," thus obscuring nearby weaker targets. If the gain or intensity is set too high, the weaker targets will not be mapped. In connection with the display, the minimum spot size of the CRT will be an important factor also in determining the overall system resolution.

Another consideration in connection with system resolution is the manner in which the antenna is scanned. When the beam is scanned relative to the aircraft, as in the conventional forward-looking system, the return from a given point will tend to fluctuate from scan to scan, except at very long ranges. This is a result of the change in aspect angle between scans. Fluctuations of the return from a target on successive scans will reduce the detail with which it is painted. In the side-looking system where scanning is accomplished by the motion of the airplane all return from a given point is within a single beamwidth and fluctuations in the return are greatly reduced. This produces a more integrated map with greater detail. "Not all the factors involved in this integrating effect are yet fully understood, but the performance of comparable scanning and side-looking systems has demonstrated the superior resolving capabilities of the latter type of scan.

System accuracy is equally important in many respects as system resolution. Distorted or "smeared" maps make it difficult to obtain a true measure of ground distances or to resolve details within the map. Types of errors that can occur are:

1. Display system errors
2. Altitude errors
3. Drift errors
4. Stabilization errors

Display errors can result from a number of factors — nonlinear sweeps, irregularities in the start of the sweep trace, imperfect stabilization of the display. Most errors of this type can be eliminated or minimized by good system design. Errors in the measurement of the altitude of the mapping

¹¹C. R. Grant and B. S. Yapple, "Back Scattering from Water and Land at Centimeter and Millimeter Wavelengths," *Proc. IRE* 45 (July 1957).

airplane can cause distortions in the map, since in most systems altitude is used to set in the range scale factor. Angular distortion of the map can occur also if there is no compensation made for airplane drift. This effect can be compensated for in several ways if the drift angle is known. In forward-looking systems the sector scan can be repositioned along the ground track of the airplane. In side-looking systems where the orientation of the antenna may not be changed easily, correction for the drift angle can be obtained, for example, by rotating the CRT trace line by an amount equal to the drift angle.

Stabilization errors occur when the antenna is not stabilized for aircraft pitch and roll displacements. Such errors produce distortion and smearing of the map. The degree of distortion or smearing that will occur is difficult to define analytically. However, some idea of the nature of the problem can be obtained by considering the range vector to some point on the ground, as shown in Fig. 14-20, and the displacement produced in it by

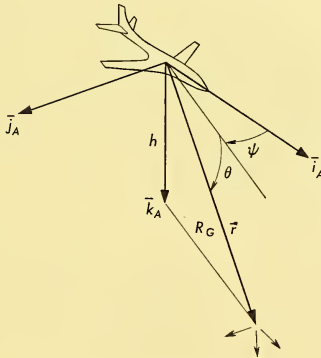


FIG. 14-20 Airplane Coordinates.

airplane pitch and roll displacements. It can be shown that when the airplane pitch angle θ_A and roll angle ϕ_A are small, the components of the displacement Δr_i , Δr_j , and Δr_k along axes parallel to the airplane set are:

$$\Delta r_i = \theta_A[\phi_A R_G \sin \psi + h] \quad (14-25)$$

$$\Delta r_j = \phi_A h \quad (14-26)$$

$$\Delta r_k = R_G[\phi_A \sin \psi - \theta_A \cos \psi] \quad (14-27)$$

where the various quantities are defined in the figure. These equations apply to forward-looking systems; for side-looking systems, where $\psi = 90^\circ$, the above equations reduce to

$$\Delta r_i = \theta_A[\phi_A R_G + h] \quad (14-28)$$

$$\Delta r_j = -\phi_A h \quad (14-29)$$

$$\Delta r_k = \phi_A R_G. \quad (14-30)$$

Examination of the last set of equations indicates that the roll stabilization of the antenna in the side-looking system will reduce the displacement to the single term $\theta_A h$ along the i axis. As already noted, the maximum displacement that can be tolerated without undue distortion is difficult to define analytically; let us specify arbitrarily that the maximum along the i axis shall not exceed 1 beamwidth. Referring to the beam of Fig. 14-19, at an altitude of 10,000 ft and a slant range of 35,000 ft this represents a distance of approximately 150 ft. Thus the maximum allowable pitch angle will be less than 1° . Studies of airplane motions in turbulent air indicate that displacements of this magnitude will probably occur with sufficient frequency to make stabilization in pitch desirable. Stabilization would be necessary in both axes for forward-looking systems also.

The range of a radar ground mapping system is affected by several factors in addition to the customary signal-to-noise ratio considerations which will be discussed in Paragraph 14-19. The maximum range for a given radar altitude will be limited by the angular coverage in the vertical dimension of the fan beam. In seeking to extend the range of coverage by mapping from higher altitudes, the increase in packet dimensions and its effect on resolution must be considered (see Fig. 14-19). Another problem in connection with mapping at long ranges and low grazing angles is the shadowing effect produced by projections above the terrain such as mountain ridges.

The problem of attaining all-weather capability has been alluded to previously in connection with the effect of wavelength on the beamwidth. Selection of wavelength must, in the final analysis, represent a compromise based upon the considerations of atmospheric and weather attenuation, resolution requirements, and the state of the art.

14-18 MAJOR CHARACTERISTICS AND COMPONENTS

Type of Transmission. In general, present-day radar ground mapping systems are ordinary pulse radars employing a low-duty cycle pulse, the duty cycle being of the order of 10^{-4} . Side-looking systems can employ coherent-pulse doppler techniques to obtain a narrow beam by rejecting all return except from points at or near the perpendicular to the airplane ground track (see Sec. 6-6). The transmission periodicity must be compatible with the maximum and minimum mapping ranges. For example, a mapping radar might have a maximum slant range of 30 n.mi. and a minimum slant range of 10 n.mi. Based on the maximum range only, the maximum pulse repetition frequency would be approximately 2700 pps.

This value could be increased to about 4000 pps if use is made of the fact that no returns occur in the 0-10-mile range interval; thus, this interpulse space could be used for returns from a 20-30-n.mi. interval.

Frequency of Transmission. Resolution requirements and weather attenuation effects combine to restrict pulse-type ground mapping radar transmission frequencies to X band and K band. The latter frequency has the advantage of producing a smaller beamwidth for a given antenna size; this fact must be weighed against increased susceptibility to weather effects. Unfortunately, the effect of weather on the tactical utility of a ground-mapping system is most difficult to evaluate. The issue is further complicated by the fact that both X-band and K-band operations are adversely affected by heavy cloud cover (it will be recalled that X band is used for weather mapping). For a given application, the designer is well advised to obtain as much experimental information as possible, because experience has shown that a purely analytical approach (using the type of data presented in Chapter 4) can be misleading in many cases.

In considering doppler techniques for mapping, the choice of frequency becomes more straightforward; the premium on a very narrow beamwidth is reduced. In general, X band is preferable for such applications.

Polarization. Most present-day radar mapping systems employ linearly polarized beams. Certain types of terrain, particularly rough seas at small depression angles, exhibit some difference between their reflective properties for vertically and for horizontally polarized radiation. This difference has been suggested as the basis of a possible means of identifying different types of terrain, and some experimental investigations have been made. However, the results were inconclusive and it appears that considerably more study is necessary to determine the possibilities of this approach.

Beam Configuration. The fan beam is used in most systems designed primarily for mapping. Its characteristics make it particularly suited for the side-look system since the motion of the aircraft provides the scanning. In most systems the beam power pattern is so shaped in the vertical plane that the intensity of the ground return is nearly independent of range. The power gain within this pattern, known as a cosecant-square beam, is expressed usually as

$$G(\theta) \propto \csc^2\theta\sqrt{\cos\theta}. \quad (14-31)$$

Additional terms are sometimes included to account for the change in atmospheric attenuation due to the varying range to the ground.

The derivation of Equation 14-31 assumes that the ground is an array of closely packed isotropic scatterers, the incident radiation being scattered equally in all directions. It assumes also that the area to be considered as

the effective "echoing" area is the projection normal to the beam of the pulse packet on the ground.¹² These assumptions raise two problems if one uses the back-scattering cross section σ^0 defined in Chapter 4. They are:

- (1) σ^0 is a function of the depression angle θ
- (2) The effective echoing area to be used with σ^0 is the actual area on the ground covered by the pulse packet, as expressed in Equation 4-59.

The expression obtained for the gain pattern when these factors are considered is

$$G(\theta) \propto \frac{\csc^{3/2}\theta \sqrt{\cos \theta}}{[\sigma^0]^{1/2}}. \quad (14-32)$$

Since the dependence of σ^0 on the depression angle is a function of the type of terrain, exact mechanization of Equation 14-32 for all cases is not possible. However, a σ^0 averaged for several types of terrain should be satisfactory.

Antennas. The antennas used in ground mapping systems include linear arrays which produce fan beams and various parabolic shapes to produce pencil beams. Doppler and correlation techniques which produce the effect of a narrow beam without the large dimensions necessary in conventional antennas are being investigated. These are in the developmental stage at present.

In the forward-looking fan beam system the linear array must be positioned perpendicular to the airplane fuselage to obtain the correct orientation of the beam. Scanning of the beam is accomplished by various mechanical or electronic techniques. The maximum dimensions of the array are limited by mounting problems. The possibility of placing the array in the leading edges of the wings has been suggested as a means of increasing the maximum length of the array that can be accommodated. However, the thinness of the wings of present-day high-speed aircraft seems to preclude this possibility. The side-looking system offers a considerable advantage over the forward-looking system in this respect, since quite long arrays of the order of 10 to 15 feet or more can be mounted along the sides of the airplane's fuselage without undue difficulty.

Stabilization. As already noted, distortion and smearing within the radar map can occur as a result of aircraft motions about the flight path. The analysis of stabilization errors in Paragraph 14-17 indicates that stabilization is probably required both in roll and pitch. While drift-angle errors can be corrected in the display system, correction for roll and pitch requires stabilization of the antenna in those axes. Roll stabilization in the

¹²*Microwave Antenna Theory and Design*, Radiation Laboratory Series, Vo. 12, Chap. 14, McGraw-Hill Book Co., Inc., New York, 1947.

side-looking systems should not be too difficult, but stabilization in pitch for long arrays will present more difficult mechanization problems. In forward-looking systems using linear arrays, stabilization in roll will be more difficult. One approach to the stabilization problem is to stabilize the aircraft itself along the axes where antenna stabilization is most difficult.

Receivers. In general, receivers used in radar ground mapping systems are of the superheterodyne type because of the very high frequencies at which such systems operate and the requirements for sensitivity, fidelity, and selectivity. While many radar systems which operate at high frequencies employ receivers of this general type, there are several considerations of particular interest for mapping systems. One of these is the effect of the requirement for good range resolution on receiver design.

To obtain a given resolution in range, ground mapping systems usually employ a very narrow pulse. Therefore, since the optimum IF amplifier bandwidth is given approximately by the reciprocal of the pulse width, the receiver bandwidth required will be comparatively wide. For example, airborne fire-control pulse radar systems, which usually employ pulse widths of the order of 1 or 2 μ sec, have receiver bandwidths of the order of several megacycles. Mapping systems, which may use pulse widths of the order of several tenths of a microsecond, will require much wider bandwidths; a pulse width of 0.1 μ sec would dictate a bandwidth of the order of 10 Mc.

Another consideration, which has been mentioned in Paragraph 14-17, is the dynamic range of the receiver; as noted, the range of signal amplitudes may vary by 30 db or more. The dynamic range of present linear amplifiers, which is of the order of 20 db, will not encompass such a range of signal amplitudes without limiting. Various approaches to the solution of this problem have been considered. One is the *lin-log amplifier*. It has linear amplification characteristics at low and moderate signal levels, but a logarithmic characteristic for large input amplitudes. Thus, although large signals are compressed, saturation of the receiver will not occur. Systems using this type of amplifier have obtained good results.

Display and Data Processing. This part of the system usually includes the cathode ray tube with its power supplies, a camera and optical system, and associated circuits which may be used to obtain computer inputs.

The CRT represents an important factor in determining the overall resolution of the system. Two characteristics of particular importance in determining its resolving capabilities are spot size and contrast. Spot size depends upon tube type, acceleration voltage, beam current, and the type of focusing used (as was explained in Paragraph 12-6). Minimum spot sizes obtainable at present are of the order of 0.1 mm or less, where the spot size

is usually so defined that two spots separated by a spot diameter are just resolvable. In terms of resolvable lines this represents about 10 lines per millimeter or more. Usable diameters for tubes which achieve spot sizes of this magnitude are about 3 to 5 inches. For a usable tube diameter of say 3.5 inches and a spot size of 0.1 mm the available number of elements or lines would be about 890 along the tube diameter. For a sweep range of 10 n.mi. this would represent a minimum resolvable range element of about 69 ft. In terms of system requirements the maximum resolution in range required for a pulse width of 0.25 μ sec would be about 123 ft; for a pulse width of 0.1 μ sec this would be reduced to about 50 ft. In azimuth the minimum required resolution as determined by the beamwidth varies with slant range. For a beamwidth of 0.25° the arc subtended by the beam on the ground would vary from approximately 27 ft at a slant range of 1 n.mi. to 265 ft at a slant range of 10 n.mi. For a spot size of 0.1 mm the CRT would limit the minimum resolvable distance in azimuth at slant ranges less than about 2.6 n.mi. Thus spot size may or may not limit system resolution, depending on various system parameters.

Display contrast depends on the characteristics of the cathode ray tube screen and the manner in which the tube is excited. The dynamic range obtainable with present cathode ray tubes is of the order of 10 db. This represents even more of a limitation than does the range of the receiver. One technique that has been used with some success to retain detail is the *three tone* method which enhances low-level signals while retaining to some extent the detail in strong signals.¹³ Where the display is viewed directly from the scope by an observer, resolution will be affected also by physiological considerations — viewing distance, levels of luminance, gradients between areas of high and low luminance, and background levels of illumination.

The order of resolution that can be obtained with typical photographic systems is such that this should not represent a limiting factor on overall system resolution. This is considered in more detail in Paragraph 14-20.

14-19 MODIFYING THE RADAR RANGE EQUATION FOR THE ACTIVE GROUND MAPPING PROBLEM

The radar range equation is conveniently considered in terms of the signal-to-noise ratio for mapping systems as

$$\left| \frac{S}{N} \right| = \left[\frac{P_{AVG} \lambda^2 c W_{RF} \Theta_n}{2(4\pi)^3 f_r h^3 KTB(NF)} \right] \left[\frac{\sigma_0 G_\theta^2}{\alpha \csc^3 \theta \cos \theta} \right] \quad (14-33)$$

where P_{AVG} = average transmitted power (= peak power \div $f_r \tau$)

¹³*Cathode Ray Tube Displays*, Radiation Laboratory Series, Vol. 22, p.38, McGraw-Hill Book Co., Inc., New York, 1948.

λ = wavelength of transmission

c = velocity of transmitted energy, customarily taken as the velocity of light in vacuum

Θ_n = horizontal beamwidth, defined by the half-power points

W_{RF} = RF attenuation losses in the system

f_r = repetition frequency

h = aircraft altitude

K = Boltzmann's constant (1.38×10^{-23} joule/°C)

T = absolute temperature (°K)

B = receiver bandwidth

(NF) = RF noise figure. A typical value for this parameter is about 12 db

σ^0 = back-scattering cross section of earth's surface

G_θ = antenna gain at a depression angle θ

α = round trip atmospheric attenuation.

As noted earlier, for uniform mapping the returned power should be independent of the depression angle θ . If we use the customary expression for the cosecant-squared beam given in Equation 14-31 and add a factor α to account for the change in atmospheric attenuation with range, the signal-to-noise ratio becomes

$$\left| \frac{S}{N} \right| = \left[\frac{P_{AVG} \lambda^2 c W_{RF} \Theta_n}{2(4\pi)^3 f_r h^3 K T B (NF)} \right] \left[\sigma^2 G_{\theta_0}^2 \csc^2 \theta \right] \quad (14-34)$$

where G_{θ_0} is the antenna gain in the direction θ_0 . The signal-to-noise ratio will still be dependent on the depression angle to some extent.

Using the expression for antenna gain given in Equation 14-32, where a term has also been included to account for atmospheric attenuation, Equation 14-33 becomes

$$\left| \frac{S}{N} \right| = \left[\frac{P_{AVG} \lambda^2 c W_{RF} \Theta_n}{2(4\pi)^3 f_r h^3 K T B (NF)} \right] \left[G_{\theta_0}^2 \right] \quad (14-35)$$

Thus the signal-to-noise ratio is now independent of the depression angle. Of course this is an ideal condition since, as already noted, the scattering cross section σ^0 will be a function of the type of terrain being mapped.

14-20 RESOLUTION LIMITS IN GROUND MAPPING SYSTEMS¹⁴

The systems factors which determine the resolution of ground mapping systems can be grouped roughly into two categories; those that determine pulse packet size and those that deal with the receiver and display.

Pulse packet size, as defined by the half-power beamwidth and pulse length, is used frequently to describe the resolution capability of a mapping system. It is a useful concept for obtaining a relative comparison of the capabilities of two systems equal in other respects. However, care must be exercised since the values assigned to the packet dimensions can be very misleading if they are used to determine randomly observed resolvability. One of the major problems in applying the packet concept results from the wide range of amplitudes of signal returns. In general the resolution of targets from which the return is very weak approaches the limits defined by the half power packet. However, for targets from which the return is much stronger the effective beamwidth and pulselength are increased. In effect, strong targets are mapped with a packet size proportional to their amplitudes as well as their relative positions and ranges from the airplane. Experience with ground mapping systems indicates that the packet size which should be used to determine resolution is somewhat undefinable unless strictly qualified. However, a qualitative estimate of the validity of the concept can be obtained by examining the form of the signal return as the antenna scans by two closely spaced point targets.

Freedman has used this approach to study the azimuth resolution of two closely spaced point targets.¹⁵ His results, expressed in terms of the ratio of power midway between the targets and the power at one of the targets as a function of target separation in beamwidths, indicate that resolution will occur for a target separation slightly greater than 1 beamwidth. The exact separation required depends on the criterion used to define resolution. Using Rayleigh's familiar rule for resolution, the required separation is about 1.1 beamwidths; using a half-power criterion, the required separation is about 1.2 beamwidths. It should be noted, however, that in the vector addition of the integrated signals from the two targets an average phase difference of 90° has been assumed. In most practical cases the number of pulse returns from each target will not be sufficient for the statistical average to apply. As Freedman notes, the average phase difference will vary between the extremes of "in phase" and "out of phase" so that in some instances targets would be resolved and in other instances the same targets would not be resolved. This phenomenon is observed in the operation of mapping systems.

¹⁴See Chapter 3 for a more detailed discussion of resolution.

¹⁵J. Freedman, "Resolution in Radar Systems," *Proc. IRE* **39** (July 1951).

When the targets differ in their reflective strengths, the same type of analysis indicates that the minimum resolvable separation is increased. Fig. 14-21 shows the envelope of the square of the received signal voltage as

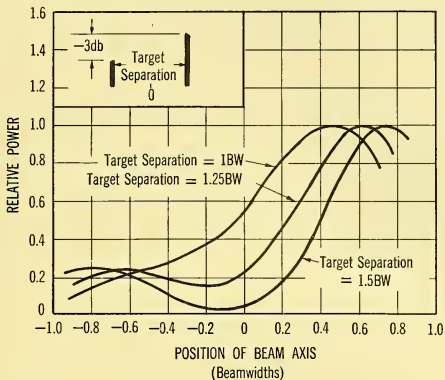


FIG. 14-21 Distribution of Power Between Unequal Point Targets.

the antenna scans past point targets which differ in reflective strength by only 3 db. Fig. 14-22 shows the ratio of the minimum power between the targets and the power at the smaller target. The minimum resolvable

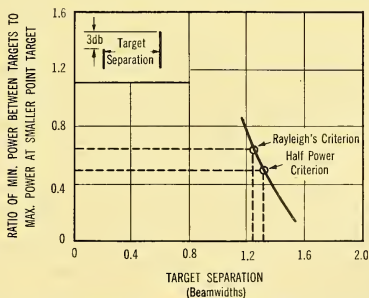


FIG. 14-22 Resolution of Unequal Point Targets.

separation has increased about 0.12 beamwidth over that for the equal point targets. This does not represent a large increase but the assumed difference in target size is quite small also.

The presence of sidelobes can also have a degrading effect on azimuth resolution. For targets of sufficient strength sidelobes could result in an ultimate resolution of 4 beamwidths¹⁶.

The same type of analysis can be used to study range resolution if the modulator pulse form can be suitably described. In general the slope of the modulator pulse will not be as steep as that radiated from the antenna; the minimum resolvable separation in range will be more affected by differences in target size.

The other area of consideration regarding resolution is the receiver and display system. The importance of the dynamic range of the receiver and display and of the spot size of the CRT has been pointed out. In general the resolution of the display will be limited by the minimum spot size.

Modern mapping systems usually employ a camera to photograph the display. This can also affect the overall system resolution. The resolution of the overall photographic process can be determined by an equation of the form

$$\frac{1}{N} = \sqrt{\left(\frac{1}{N_1}\right)^2 + \left(\frac{1}{N_2}\right)^2 + \left(\frac{1}{N_3}\right)^2 + \dots} \quad (14-36)$$

where the N 's refer to the number of resolvable lines arising from each independent element of the system¹⁷. There are four major independent factors which combine to determine the photographic system resolution. These are:

1. *Lens Resolution* — The resolution of the lens is determined by such factors as lens diameter, f number, and image size. For a flux response factor of 0.5 a typical value of N for high-quality lenses is about 70 lines per millimeter.

2. *Depth of Field* — The image formed by the lens will be in focus only for objects at a given distance. Objects at a slightly greater or lesser distance will not be sharply focused upon the film. However, because of lens aberrations even a point in the focal plane will be imaged as a small disk, called the *disk* or *circle of confusion*. The circles of confusion for points at other distances will be larger. Depending upon the sharpness of focus required, there is a certain range of object distances, called the *depth of field*, within which all objects are in satisfactory focus. The depth of field can be expressed in terms of resolvable lines and for a typical system could be of the order of 100 lines per millimeter.

3. *Film Resolution* — The resolving power of photographic films depends upon such factors as grain size and emulsion thickness. A conservative

¹⁶Freedman, *loc. cit.*

¹⁷O. H. Schade, "Electro-Optical Characteristics of Television Systems," Parts I through IV, *RCA Review* 9 (1949).

estimate of the rated resolving power of present photographic films for large-signal response is about 100 lines per millimeter.

4. *Enlargement Resolution*—It can be assumed that the enlarger resolution is equal to that of the camera lens.

For these typical values the overall resolution of the photographic system is found to be about 40 lines per millimeter. Since the resolving limit of the CRT is of the order of 10 lines per millimeter, the photographic system should not represent a limiting factor in determining the number of resolvable lines. The above considerations have been in terms of the number of resolvable lines to be obtained in the photographic process. The final resolution will be affected also by the dynamic range of the film. If the final map is to be a positive print, the total range of signal amplitudes must be compressed within the rather narrow range of present photographic papers, with a resultant loss of contrast. If the negative is to be viewed with a projector, a much wider range of contrast will be available.

As the reader has probably already concluded, the ultimate resolution of a mapping system involves such a large number of different parameters that the actual resolution that will be obtained must be determined by actual operation of the system. However, knowledge of the effect of the various parameters enables the system designer to make a reasonable estimate.

14-21 FUTURE POSSIBILITIES IN AIRBORNE ACTIVE GROUND MAPPING SYSTEMS

Two areas that appear to be among those most likely to provide significant improvements in future mapping systems are antennas, and display and data processing systems. The maximum dimensions of physical arrays are obviously limited, but synthetic arrays show promise of making it possible to achieve large antenna dimensions effectively without large physical dimensions (see Paragraph 6-7). New advances in display and data processing systems include techniques for image enhancement through video data processing and optical filtering.

14-22 INTRODUCTION TO INFRARED RECONNAISSANCE

Reconnaissance by the reception of naturally radiated infrared energy provides intelligence that is difficult to obtain by other means. Photography in the visible spectral region and radar at microwave frequencies "see" the purely physical aspects of targets and terrain. Infrared mapping "sees" a scene because of its heat content and contrast. While infrared reconnaissance will generally provide poorer resolution than photographic reconnaissance, it can produce several unique types of tactically useful information. Experienced interpretation of infrared pictures yields

information concerning physical features, heat distribution, targets, present and recent activity of targets, rate of activity, and the presence of camouflaged targets such as underground factories. The ability of IR systems to provide information which can complement radar and photographic intelligence has resulted in considerable development effort in this field.

All objects above absolute zero radiate electromagnetic energy. An appreciable portion of this energy lies in the infrared spectrum. Within the broad assumption that an object is a perfect absorber and perfect emitter (a black body) it radiates infrared energy in accordance with the Planck radiation law. It is this infrared energy which, obeying the laws for visible radiation, is received and focused, thus creating a thermal image.

14-23 BASIC PRINCIPLES CONCERNING IR GROUND MAPPING

Infrared mapping requires a scanning procedure in which the sensing element or elements are exposed to radiation from a small area (or areas) of the ground at any one instant. The angle thus subtended by this ground area at the optics is the *instantaneous field of view*. Fig. 14-23 shows this

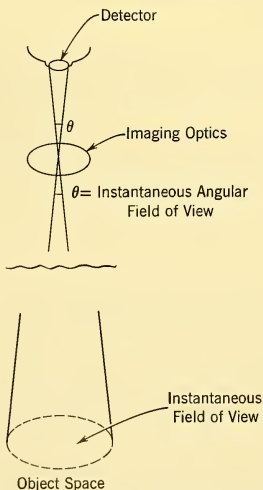


FIG. 14-23 Instantaneous Field of View.

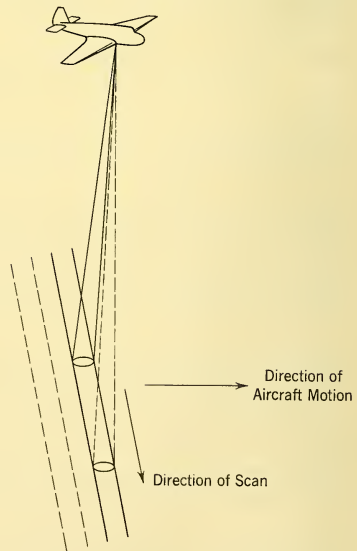


FIG. 14-24 Linear Scan Pattern.

to be the same as the angle subtended by the detector at the real or effective optical element located a distance equal to the focal length from the detector. Scanning the sensor across the desired mapping area in a direction perpendicular to the flight path while the aircraft is in motion yields a continuous line-by-line thermal map of the region below. Fig. 14-24 illustrates this type of scan pattern.

The width of a strip map (w) as measured on the ground is given by

$$w = 2h \tan \alpha \quad (14-37)$$

where w = object space width of map

h = aircraft altitude

α = total field half-angle.

Table 14-2 shows approximate strip map widths for several altitudes for 60° , 90° , and 150° total field of view systems. As can be seen, total map width increases rapidly as angular coverage is increased beyond 90° .

TABLE 14-2

h , ft	2α , n.mi.		
	60°	90°	150°
1000	0.19	0.33	1.23
12,000	2.28	4.0	14.76
72,000	13.68	24.0	

The upper limit on angular coverage is generally governed by resolution requirements. The size of the ground area defined by the instantaneous field varies with the slant range. The ratio of the ground area filling the instantaneous field when looking at an angle θ as measured from the vertical to the corresponding ground area when looking directly below the vehicle is equal to $\sec^3\theta$. Thus

$$\frac{(\text{area})_{\theta=\theta}}{(\text{area})_{\theta=0}} = \sec^3\theta \quad (14-38)$$

Because of this resolution degradation, total fields of view are generally less than 150° .

The instantaneous angular field of view and electronic bandwidth normally provide the limit to system resolution, given optics with better resolution capability than is required. Therefore, the size of the smallest anticipated targets exerts an influence on the chosen instantaneous field of view. If detection alone is required, the instantaneous field need not be small, and indeed may, depending on the target and background heat characteristics, be substantially larger than the target itself. Reconnaissance usually demands identification of the type of target, thus requiring

smaller instantaneous fields. One to three milliradians is representative of present capabilities, with requirements becoming more stringent for good detail from very high altitudes. Insufficient detector sensitivity and the Rayleigh limit are not working in favor of instantaneous fields of view much below 0.5 milliradian with 4- or 6-inch optics.¹⁸

The radiation from the earth, considered as a black body, peaks at about 10 microns. In this respect the problem of ground mapping is significantly different from that of detection of "hot" targets such as engines of jet aircraft or high-speed missiles (see Paragraph 6-8) whose radiation peaks

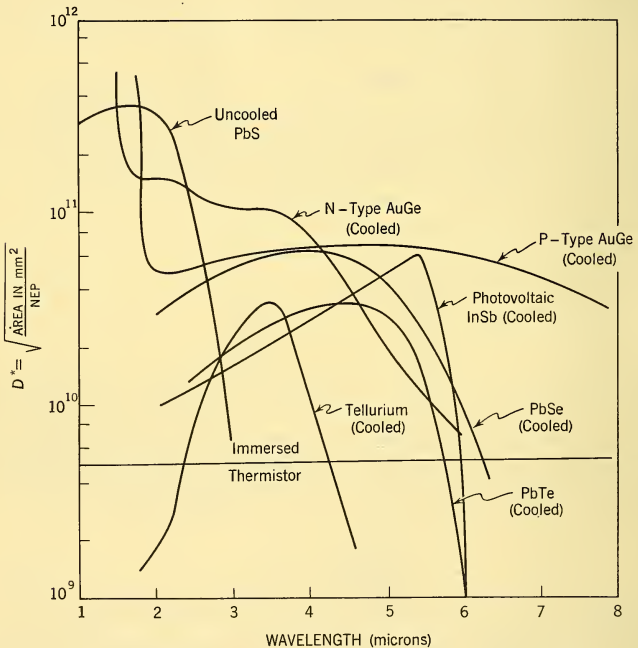


FIG. 14-25 Spectral Response of Several Infrared Detectors.

¹⁸According to the Rayleigh criterion, two point sources are at the limit of resolvability when the center of the diffraction pattern of one object coincides with the first minimum of the other. The minimum point source angular separation (θ_{\min}) for such resolvability with a circular aperture is given by $\theta_{\min} = 1.22\lambda/a$, where λ is the wavelength of the radiation and a is the aperture diameter. For a more complete description of resolution and methods of improving resolution by means of apodization, see J. Strong, *Concepts of Classical Optics*, W. H. Freeman and Co., San Francisco, 1958.

occur in the 2- to 5-micron region. It is necessary to choose the most sensitive detector and the one with the best spectral response in order that as much of the radiation as possible may be utilized efficiently. Assuming other important considerations (discussed below) to be favorable, the combined value of good absolute and wide spectral response determines the degree of contrast that may be obtained. A picture is created *not* because of the temperature of object space as a whole, but by the changes of temperature (and emissivity, of course; this may be neglected if an *effective* temperature is considered) between instantaneous field of view regions in object space. Fig. 14-25 shows a spectral and absolute response comparison of the common detectors.

This paragraph has dealt with the fundamental considerations of thermal mapping, namely, the angular coverage and picture quality. The criteria for quality are resolution and contrast. The effect of these basic considerations on system parameters is discussed in the following paragraph.

14-24 SYSTEM CONSIDERATIONS

The elements of a typical IR ground mapping system are shown in block diagram form in Fig. 14-26. The main parameters associated with each

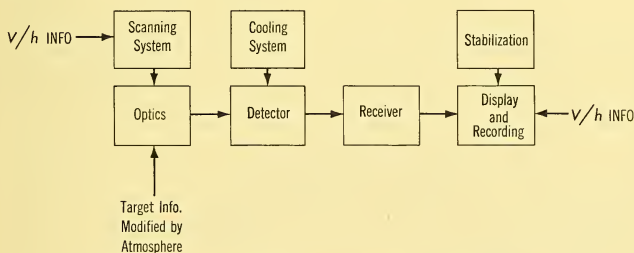
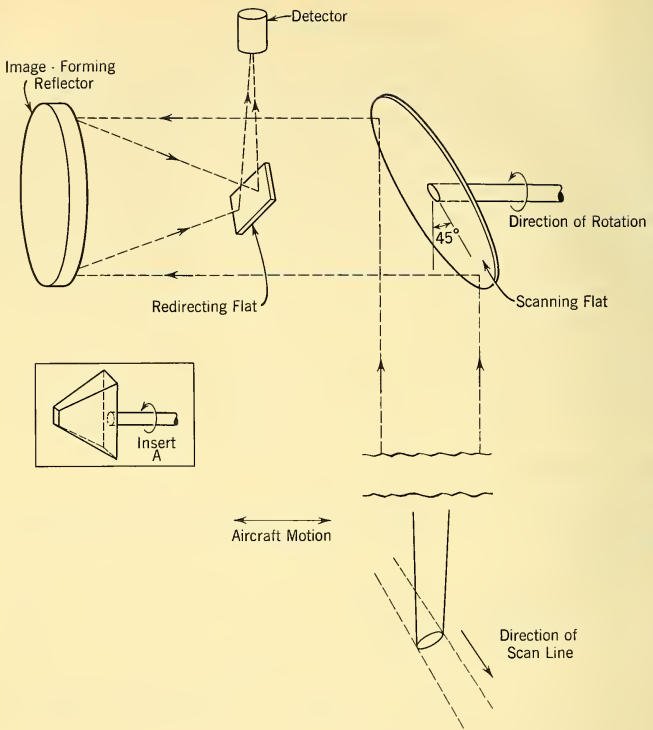


Fig. 14-26 Typical IR Mapping System and Parameters.

system element are also listed. The choice of these parameters determines the system characteristics. The modification of any one parameter is never independent of all of the others. This paragraph discusses the interrelationships of these parameters and shows how the choice of them is influenced by operating requirements and environment (speed, altitude, etc.).

The total lateral angular coverage required determines the number of scan lines attainable per revolution of the scanning optics. High-speed scanning precludes anything other than rotatory motion. Fig. 14-27 illustrates perhaps the simplest method for obtaining a line scan in object space. Energy from object space is incident upon the scanning flat and



SENSITIVITY PARAMETERS

Scanning System
 Type of Scan
 Scan Rate
 Total Field of View

Optics
 Aperture
 Focal Length
 Resolution
 Efficiency

Detector
 Spectral Response
 Sensitivity
 Time Constant
 Noise
 Size and Shape
 Impedance

Cooling System
 Temperature
 Capacity

Receiver
 Noise
 Bandwidth
 Controlled Nonlinearity

Display and Recording
 Dynamic Range
 Resolution
 Synchronism with Scan
 Linearity

Stabilization
 Correction Signal Generation

FIG. 14-27 Simple Line Scan Generation.

redirected into the image-forming or primary reflector. The resultant converging beam is redirected by a second flat and is finally incident on the detector. Rotation of the scanning flat, as shown, provides a linear scan pattern (neglecting distortions incident to change in slant range and deviation of the scanning flat from the 45° position shown) in object space. If the total field is to be 60° , five scan lines per revolution are possible with a single element sensor. If 150° is to be scanned, two lines per revolution is the maximum. This becomes more clear by referring to Fig. 14-27 and picturing (for example) the scanning flat as replaced by a four-sided pyramid reflector with the base attached to the rotating shaft as shown in insert A. If an opening is provided so that the system may look down over a total field angle of $360^\circ/4$ or 90° ($\pm 45^\circ$ from vertical), then one of the four pyramid sides is always receiving radiation through the opening and directing it into the system. With the understanding that the other three mirrors are providing no fluctuating radiation to the system, the total signal variations enter by way of the reflector that is "looking" through the opening. It is clear, then, that in one revolution of the shaft each of the four mirrors will in turn provide a 90° total field scan transverse to the flight path. Four scan lines per revolution are thus generated. The so-called dead time between scan lines may conveniently be used for processing and transmission via telemetry links.

The required resolution in general establishes the instantaneous field of view. Assuming no other system limitations (most notably the optics and recorder), the instantaneous field and electronic bandwidth determine system resolution. At this point the concept of data rate is quite useful. Data rate is the number of picture elements per second that are collected and used by the system. While a given instantaneous field of view may make possible a given data rate, the effect of the detector response, electronic bandwidth, and recorder range will generally limit the system data rate. In terms of the present discussion, the instantaneous field of view is not necessarily equivalent to the system resolution (actually, system resolution can be better than the instantaneous field of view, although this more sophisticated approach is not discussed in this text). With the instantaneous angular field established, the angular width in the direction of flight path of one scan line is fixed. A fixed angular width corresponds to nonuniform linear width in object space because of change in slant range with change in scan angle. In order that there may be no gaps in object space mapping, the scan line width used in calculations is the width measured directly below the aircraft.

With fixed angular scan line width, the aircraft altitude then determines the linear scan line width in object space (see above). The velocity determines the number of scan lines of this given width required per unit time in order that each line may lie adjacent to the preceding one. Therefore,

the velocity and altitude, together with the instantaneous field, determine the scan line rate. In turn, this rate, together with the number of scan lines per revolution of the scanning element, as fixed by the total angular field of view, determines the rotation frequency of the scanner.

For example, consider the reasonable case where the mission requirements have in turn determined the following parameters:

Aircraft velocity $v = 1000$ ft/sec (relative to ground)

Aircraft height $h = 30,000$ ft

Instantaneous field $\Delta\phi = 1$ milliradian

Total field of view $= 150^\circ$

Let $n = \text{scan lines/revolution} = 2$, in view of the total field size. Then the rotation frequency of the scanner (Ω) in revolutions per minute is clearly

$$\Omega = 60 \frac{v}{h} (n\Delta\phi)^{-1} \text{ rpm.} \quad (14-39)$$

Assuming the above conditions to impose the most severe v/h requirements the system must meet, then the scanner must be capable of

$$\Omega = 60 \times \frac{1000}{3 \times 10^4} (2 \times 10^{-3})^{-1}$$

$$\Omega = 1000 \text{ rpm.} \quad (14-40)$$

Note that the effect of detector time constant and electronic bandwidth have not been included at this point, and therefore nothing can be said concerning data rate or resolution until these effects are evaluated. These considerations are discussed below.

The scan rate (revolutions per unit time) must be variable with velocity and altitude as the above dependency indicates. The time required per scan line to assure adjacent scan lines in object space is readily shown to be inversely proportional to the velocity-to-height ratio. The most stringent requirements on the sensor and scanning optical element are encountered for a high-velocity aircraft flying at low altitude. In this case it may be desirable to somehow increase the instantaneous angular field of view. Carried to the other extreme, high velocities at very high altitudes pose less of a problem from the scan rate point of view. It is also noteworthy that the idea of range is meaningless in its usual sense when applied to reconnaissance. If sufficient system sensitivity is available to detect the ambient temperature ground while at ground level, then except for atmospheric attenuation there will be just as much energy falling into the system at one altitude as at another until the instantaneous field of view includes the entire earth. The important point is that the energy from a target within the instantaneous field is averaged over the whole instantaneous field, and when this average is not distinguishable from the average over an adjacent

field element, contrast disappears. There is no detection of the target from this altitude or higher. The determination of this altitude is dependent on the radiative characteristics and size of the target as well as on detector sensitivity and altitude. Therefore, a maximum altitude is determinable only for each individual target with its characteristics.

The optical system must have sufficient aperture to assure enough energy into the sensor for a usable signal-to-noise ratio. In general, infrared optical systems are reflective and thus spectrally nonselective. If refractive elements are used, their transmission characteristics must be compatible with the detector's spectral response. Since for most detectors (P-type gold-doped germanium is a notable exception) the detectivity varies inversely as the square root of the sensitive area, small cells should be used. This permits small instantaneous fields of view with relatively "fast" (low f number, or small diameter-to-focal-length ratio) optical systems, thus contributing to compactness.

With typically small instantaneous fields of view, either spherical or parabolic primary reflectors may be used. More important is the assurance that the circle of least confusion is sufficiently small for the resolution not to be limited by the optics. This, of course, becomes impossible as an approach of within a factor of 2 to the Rayleigh resolution limit is made. High-speed scanning requires careful evaluation of scanning element deformation. Slight deformation will seriously reduce resolution and picture quality. The method of scan must be such that the instantaneous image always falls at a fixed point in image space. This permits the sensor to be fixed in space, an extremely important feature as far as noise is concerned. In conjunction with the scanning operation, it is necessary to assure constant aperture for all scan angles and to maintain constant optical efficiency. Otherwise, nonuniform intensity across the picture results.

A basic consideration in the selection of detector characteristics is that of spectral response. It is necessary to determine the spectral distribution of the energy entering the system. Water vapor and carbon dioxide are the two constituents of the earth's atmosphere that are most effective in absorbing infrared radiation. By far the greatest amount of absorption occurs between ground level and about 30,000 ft. Fig. 14-28 shows a typical transmission curve for the atmosphere. While conditions change considerably and thus affect such transmission curves seriously, nevertheless the effect is in absolute value rather than spectral distribution. The absorption must be taken into account to the best approximation that available data and expected weather environment will permit.

The parameters thus far determined are basic in determination of the "best" infrared detector to be used. Conversely, final choice of the cell limits the range of values these parameters may assume. The spectral

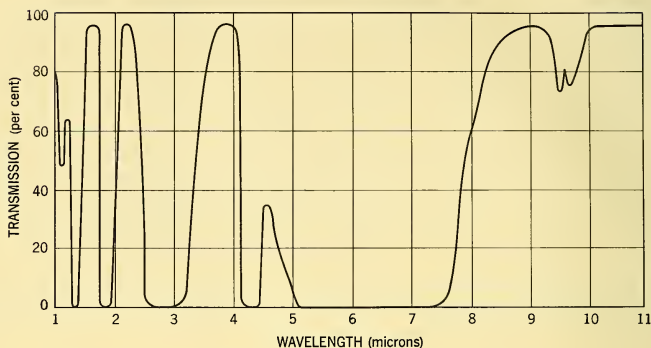


FIG. 14-28 Infrared Vertical Transmission from 0 to 30,000 Feet.

response requirement is determined by the spectral distribution of the received radiation. In general the broadest spectral response available with present cells, and within the following requirements, is best. This statement is not categorically true, but holds for present detectors. Absolute response to received radiation is equally important. The greater the detector sensitivity, the greater its capability of detecting small signal variations — and contrast is all-important in mapping. The required scan rate and instantaneous field of view determine what the detector time constant should be. This may well be an impossible figure. Time constants range from about 1 msec for thermistors and 25 to 100 μ sec for cooled lead selenide and lead telluride to 0.2 μ sec for P-type gold-doped germanium. The detector time constant may provide the limit on scan rate and resolution and thus the limit on maximum value of velocity-to-height ratio with given instantaneous field of view. In other words, the data-rate capability of the detector is limited in part by its time constant and in part by its sensitivity.

Another possible limitation on scanning parameters is that of system sensitivity. The sensitivity of the detector *as degraded by the remainder of the system* determines system sensitivity. The sensitivity of detectors is usually given in terms of noise equivalent power (NEP) or detectivity ($D^* = \frac{\sqrt{\text{area of cell}}}{\text{NEP}}$). Required supplementary information includes center frequency at which measurements were taken, bandwidth, detector size, and source temperature. *Noise equivalent power* is defined as the incident power required to produce a signal-to-noise ratio of 1. It might be noted that the type of limiting detector noise must be investigated to assure optimum operating conditions.

The signal-to-noise ratio of the system (a measure of system sensitivity) is inversely proportional to NEP and the square root of bandwidth:

$$S/N \propto (\text{NEP})^{-1}(\Delta f)^{-1/2}. \quad (14-41)$$

Neglecting any other degradation, particularly of the recording system, system sensitivity may well be orders of magnitude less than detector sensitivity. For example, if the scan rate and resolution demands are such as to require 1-Mc bandwidth, then

$$S/N \propto (\text{NEP})^{-1}(10^6)^{-1/2} = (\text{NEP})^{-1}(10^3)^{-1} \quad (14-42)$$

and if the cell NEP is 10^{-11} watt, then

$$S/N \propto 10^{11} \times 10^{-3} = 10^8 \quad (14-43)$$

whereas it would have been 10^{11} , given a normalized 1-cycle bandwidth. Thus the price of bandwidth is system sensitivity. A compromise in resolution, scan rate, and system sensitivity must be decided upon in the light of the particular requirements of each system.

Among the considerations for the electronic circuitry into which the detector operates, the most important is that of noise. The preamplifier noise level must be below that of the detector. Matching the detector to its load may not be possible because of the extremely high impedance shown by some cells (up to 150 megohms for lead telluride). Such high impedances are extremely conducive to noise pickup. In the case of cells such as the 1.5-megohm P-type gold-doped germanium, matching is not desirable from the bandwidth point of view. A cascode preamplifier is ideally suited to this cell. Low noise and elimination of the Miller effect are features of this circuit.

Numerous techniques have been used for recording and displaying infrared mapping information. Whatever the method, care must be taken to assure full use of the resolution capabilities of the remainder of the system. The resolution of the recorder should not provide the limiting resolution. If it does, sensitivity is being sacrificed. Furthermore, the dynamic range of the recorder should match that of the foregoing system as much as possible.

As mentioned earlier, the scan rate must be variable in accordance with the velocity-to-height ratio. Identical v/h information must be given to the display system in order that it may be in synchronism with the scanning element.

Generally some provision is made for correction of errors caused by roll, pitch, and yaw. Roll is usually a problem even if pitch and yaw may be neglected. A stabilized platform may be used for mounting of the system. An alternate scheme is to provide the correction electronically. This would be done in the indicator and may be the more convenient, especially when size and weight are critical factors.

All detectors (except thermistor bolometers) that currently find use in infrared mapping systems are cooled to liquid nitrogen temperature (90° K). Cooling is no longer a handicap, and may be accomplished by the use of available miniature cryostats or by manually pouring the liquid coolant into the cell. This latter method may, of course, be a forced flow system in which the liquid flow is pressure-regulated instead of actually requiring a manned operation. Particularly for P-type germanium it is important that good thermal contact be made between the detector's heat sink and the coolant. Otherwise, minor temperature fluctuations will give rise to additional noise.

14-25 MAJOR SYSTEMS FEATURES

No one set of criteria provides an adequate basis for comparison of infrared mapping systems. However, some characteristics of a system do describe it sufficiently well for meaningful evaluation:

- | | |
|------------------------------|-------------------------|
| (a) Spectral response | (c) Resolution |
| (b) System Sensitivity | (d) Total field of view |
| (e) Maximum v/h capability | |

An overall figure of merit for system performance is sorely needed. In view of the different types of application this may not be completely feasible. An effective and standard method for comparison of infrared reconnaissance systems must include the important features of scanning systems and must do the comparison in a completely fair manner. These requirements have so far precluded a useful approach to the problem. Most systems, however, are sufficiently similar in purpose to warrant an attempt at a standard method of comparison. It would seem that a proposed method open to critical comments would provide a good starting point. With this in mind, let us consider the problem in more detail.

The following pages propose a standard method which should be fully applicable to mapping systems. Special-purpose systems may be subjected to the comparison procedure, but due caution must be observed. These system characteristics are inherent either explicitly or implicitly in the comparison scheme:

- | | |
|---------------------------------|----------------------------------|
| (a) Instantaneous field of view | ture difference |
| (b) Resolution (system) | (f) Sensitivity (system) |
| (c) Time constant (system) | (g) Scan rate |
| (d) Spectral response | (h) Effect of multiple detectors |
| (e) Minimum detectable tempera- | (i) Size of optics |

A plot of the reciprocal of the product of minimum detectable black body temperature difference and instantaneous field of view versus effective scan rate is proposed. At this point it should be noted that scanning systems

may be compared in two basic ways. First, a more or less ideal response may be calculated when only the system parameters are known. Thus system comparison on a theoretical basis may be accomplished with system evaluation in terms of chosen parameters resulting. This type of evaluation is useful as a preparation for final systems design. However, the comparison value is open to question because of the latitude in assigning values to the numerous parametric quantities involved, and because of the complexity of preparing a concise form of comparison. The comparison at best provides an indication of what the systems involved *should do* under various circumstances rather than what they actually *can do*.

The second basic method of comparison is that of experimentation. In this method each system would be subjected to experimental evaluation, with the system itself as the measuring instrument. The entire system, including display and recording units, is therefore included. Experimental comparison leaves no room for doubt as far as the difference between theory and practice is concerned.

The choice of parameters for the proposed evaluation scheme is justified by the following line of reasoning. The ordinate value is the reciprocal of the product of the minimum measurable temperature difference and the instantaneous field of view. The latter parameter is a measured value and can be readily determined. In itself it has little meaning, but its value together with bandwidth determine in part the minimum measurable temperature difference. In view of this, and because of the direct relationship with detector size and focal length of the optics, the instantaneous field of view ($\Delta\phi$) is a basic scanning-system parameter. Being a geometrically measured quantity, there can be little more than negligible error associated with its determination. The choice of minimum measurable temperature difference (ΔT) is clear. Given any number of systems, each with fully acceptable characteristics (resolution, etc.) for a particular application, the best system will be the one capable of the smallest ΔT measurements over the temperature range in question. Effective scan rate in conjunction with the ordinate parameters determines the velocity and height limitations under which the system will be effective.

The experimental procedure would be as follows:

Two black body apertures are considered to lie in a uniform and much cooler object space. Each subtends a fixed angle at the optics; probably 0.25 milliradian is a good choice. This value is chosen so that the instantaneous field of view will almost assuredly be larger than the black body apertures. The black body apertures are placed so that they may be readily resolved by the scanner. Then one black body is adjusted to a specific temperature while the other is adjusted to a temperature just sufficiently higher than the first for the scanner recording system to identify a change in gray level between black bodies. This ΔT , or difference in black body

temperatures, is multiplied by the instantaneous field of view of the system $\Delta\phi$, and the reciprocal of the product is plotted against effective scan rate. The same is done for different black body temperatures, so that a family of curves each of a given temperature T is obtained. The $\Delta\phi$ for a multi-element system is the $\Delta\phi$ of any one detector.

The effective scan rate is the actual revolutions per minute (for convenience) times the number of scan lines per revolution. For a front- and back-surface scanning mirror (double primary) the effective scan rate would be 2 times the actual revolutions per minute; for a 10 element linear array, the effective rate would be 10 times the actual rate, and so on.

The system itself is doing the actual measuring, so that any data are modified in accordance with system characteristics. If bandwidth is insufficient, $(\Delta T \times \Delta\phi)^{-1}$ drops off at smaller effective scan rates. If the bandwidth is too large, the ΔT will be so affected that the ordinate value $(\Delta T \times \Delta\phi)^{-1}$ is smaller than it otherwise would be. If $\Delta\phi$ is decreased too much, the additional bandwidth requirement and possibly time constant limitation will reduce sensitivity, thus increase ΔT , and cause the curve to break at lower effective scan rates. If $\Delta\phi$ is increased, ΔT will increase and thus the ordinate value will again be lower. If a slow detector is used, the break will of course occur at low effective scan rates. The use of fast detectors will allow the curve to remain flat to high effective scan rates with both the ordinate value and deviation from flatness occurring as a result of

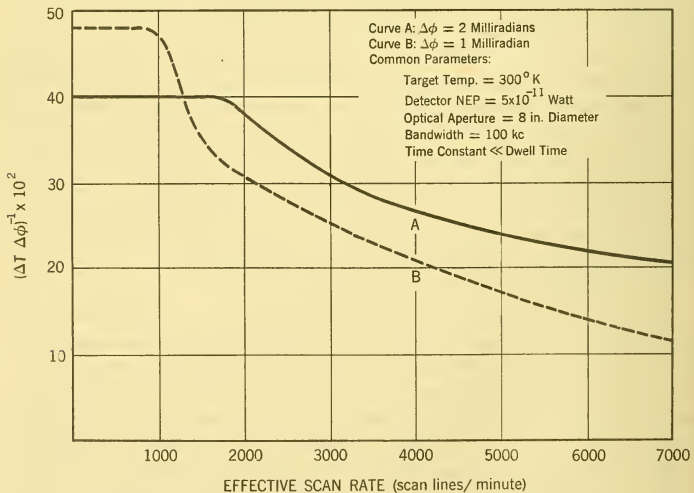


FIG. 14-29 A Method of Comparison for IR Mapping Systems.

bandwidth. The same effect is obtained by using multiple-element detectors even though the time constant of each element may be low. In all cases for any use whatever, the best system will be that which, for any fixed parameters, exhibits the break in the curve at the proper point and has the highest ordinate value over the required temperature range.

Such curves should be extremely useful from the design point of view. For example, if v/h requirements are not compatible with the location of the break in the curve, the obvious factor available for trading purposes is the $\Delta\phi$. An increase in $\Delta\phi$ will cause ΔT to increase, but either the bandwidth may be decreased or the scan rate increased. Choosing the latter, the value of v/h is accordingly increased. In leaving the bandwidth alone, the decrease in $(\Delta T \times \Delta\phi)^{-1}$ may be found, and a new plot will show a decreased ordinate value, but the break in the curve has been pushed to higher effective scan-rate values. Thus the optimum point of operation may be determined, with comparisons for a change in any one parameter or combination of parameters. A plot of the type discussed appears in Fig. 14-29. The values chosen and curves drawn are not at all representative of actual values but are used merely to indicate the kind of information that may be obtained.

14-26 NEW DEVELOPMENTS

Research and development in the infrared field promise many improvements and new features that will extend the tactical and strategic usefulness of IR ground mapping techniques. Major advances in detector sensitivities, both spectral and absolute, will be realized. Evolutionary improvements in optical systems, cooling systems, and materials will contribute to the increasing efficiency of present high-speed scanning systems.

The advanced development of matched line arrays and two-dimensional mosaics of detector elements offers aid in the quest for higher scan rates for the immediate future. These arrays are simply sets of single elements placed very close together, each with its own electrical connections. Adjacent elements thus look at adjacent regions in object space. The effect is to increase the instantaneous field of view at the cost of system resolution. The advantage thus purchased may be applied as greater total field coverage, slower scanning speed, less bandwidth, greater data rate, etc., or perhaps more profitably, as optimum compromises among these parameter variations. The major development problems are matching between elements, and adequate (i.e. sufficiently noise-free) switching among elements.

Revolutionary changes in IR ground mapping may be anticipated with the introduction of highly sensitive image tubes. Research is being done on both photoconductive and thermally sensitive electron-scanned image

tubes. The former must have improved spectral response, the latter require improved absolute response before full application to ground mapping may be realized. Nevertheless, the not too distant future should bring IR image tubes into serious contention with the so-called "point" detectors for mapping operations.

Equally important with new components are new techniques. Among the most promising are the recent developments leading to stereoscopic infrared pictures with simultaneous moving-target indication, multicolor response (response to several well-defined spectral regions simultaneously but with each response separated from the others and individually treated), and advanced variable contrast control. Infrared reconnaissance was born about ten years ago. The advances during this time and the work in both research and development of the present make the outlook for the future an optimistic one indeed.

BIBLIOGRAPHY FOR CHAPTER 14

DOPPLER NAVIGATION

1. J. S. Hall, *Radar Aids to Navigation*, Massachusetts Institute of Technology Radiation Laboratory Series, Vo. 2, p. 110, McGraw-Hill Book Co., Inc., 1947.
2. J. L. R. Cudaback, *Earth Rate Directional Reference for Aircraft*, Master's Thesis, United States Air Force Institute of Technology, 13 March 1956.
3. F. B. Berger, "The Nature of Doppler Velocity Measurement," *IRE Trans. ANE-4*, 103-112 (September 1957).
4. F. B. Berger, "The Design of Airborne Doppler Velocity Measuring Systems," *IRE Trans. ANE-4*, 157-175 (December 1957).
5. W. R. Fried, "Principles and Performance Analysis of Doppler Navigation Systems," *IRE Trans. ANE-4*, 176-196 (December 1957).
6. C. R. Grant and B. S. Yapple, "Back Scattering from Water and Land at Centimeter and Millimeter Wavelengths," *Proc. IRE* 45, 976-982 (1957).
7. J. P. Campbell, "Back-Scattering Characteristics of Land and Sea at X-Band," *Proc. 1958 National Conference on Aeronautical Electronics, IRE*, Dayton, Ohio.
8. W. R. Fried, *Doppler Navigation Systems — Principles and Practice*, Wright Air Development Center Technical Note 57-175, ASTIA Document No. AD-118271, December 1957.
9. F. B. Berger, "Summary of AN APN-66 Flight Test Accuracy Data," General Precision Laboratory Internal Report 161, May 17, 1955.
10. W. R. Fried, "Performance Profiles and Future Outlook of Doppler Navigation Systems," *IRE Trans. ANE-5*, (December 1958).
11. A. Shapiro, "Precision Azimuth Reference Systems," *Proc. 1957 National Conference on Aeronautical Electronics, IRE*, Dayton, Ohio, May 1957.
12. W. L. Harmon and C. C. Goldman, "Gyro Heading Reference for High Latitudes," Wright Air Dev. Center Tech. Note WCLN 54-4, Wright-Patterson Air Force Base, Dayton, Ohio, May 1954.
13. D. J. Atwood and F. A. Best, "Airborne Earth Rate Directional Heading Reference," pp. 457-462, *1958 Proc. National Conference on Aeronautical Electronics*, May 1958.

14. H. J. Galbraith and N. Braverman, "The Practical Combination of Air Navigation Techniques," *IRE Trans. on Aeronautical and Navigational Electronics*, March 1956.
15. B. D. Duncan, "Combined Doppler Radar and Inertial Navigation Systems," *Navigation*, vol. 6, no. 5 (Spring 1959).
16. L. M. Dworetzky and A. Edwards, "Introduction to Doppler-Inertial System Design," *ARS Journal*, vol. 29, no. 12, pp. 967-972 (December 1959).
17. W. R. Fried, "On the Application of Doppler Navigation Equipment in the Air Traffic Control Environment," pp. 502-517, *Proc. 1959 National Aeronautical Electronics Conference, IRE*, Dayton, Ohio, May 1959.
18. M. Y. Silberberg and J. P. Campbell, "Space Vehicle Applications of Self-Contained Doppler Radar," pp. 104-109, *Proc. 1959 National Aeronautical Electronics Conference, IRE*, Dayton, Ohio, May 1959.
19. J. P. Campbell, "Doppler Techniques for Space Vehicle Guidance," *Proc. Third National Convention on Military Electronics, IRE*, Washington D. C., July 1959.
20. F. B. Berger, "Application of Doppler Techniques to Space Navigation," *Navigation*, vol. 6, no. 7 (Autumn 1959).

WEATHER RADAR

1. D. E. Kerr (Ed.), *Propagation of Short Radio Waves*, McGraw-Hill Book Company, Inc., New York, 1951.
2. D. Atlas and J. S. Marshall, *Weather Effects on Radar*, Air Force Surveys in Geophysics, No. 23, Air Force Cambridge Research Center, December 1952.
3. J. O. Laws and D. A. Parsons, "The Relation of Raindrop Size to Intensity," *Trans. Am. Geophys. Union* 24 (part II), 1943.
4. J. D. Ryde and D. Ryde, *Attenuation of Centimeter and Millimeter Waves by Rain, Hail, Fog and Clouds*, GEC Report No. 8670, 1945.
5. R. F. Jones, *Radar Echoes from and Turbulence within Cumulus and Cumulonimbus Clouds*, Prof. Notes 109, Meteorological Office, Air Ministry, London, 1954.
6. H. T. Harrison and E. A. Post, *Evaluation of C-Band (5.5-cm) Airborne Weather Radar*, United Airlines, Inc., Denver, Colorado, 1954.
7. D. E. Kerr (Ed.), *Propagation of Short Radio Waves*, McGraw-Hill Book Co., Inc., New York, 1951.
8. J. C. Johnson, *Meteorological Factors and Their Effects on Microwave Propagation*, Technical Memorandum 412, Hughes Aircraft Co., 1955.

ACTIVE AIRBORNE GROUND MAPPING SYSTEMS

1. C. R. Grant and B. S. Yaplee, "Backscattering from Water and Land at Centimeter and Millimeter Wavelengths," *Proc. IRE* 45 (July 1957).
2. Samuel Silver (Ed.), *Microwave Antenna Theory and Design*, Massachusetts Institute of Technology Radiation Laboratory Series, Vol. 12, McGraw-Hill Book Co., Inc., New York, 1947.
3. J. Freedman, "Resolution in Radar Systems," *Proc. IRE* 39 (July 1951).
4. S. N. Van Voorhis (Ed.), *Microwave Receivers*, Massachusetts Institute of Technology Radiation Laboratory Series, Vol. 23, McGraw-Hill Book Co., Inc., New York, 1948.

5. O. H. Schade, "Electro-Optical Characteristics of Television Systems," Parts I through IV, *RCA Review* 9 (1949).
6. T. Soller, M. A. Starr, and G. E. Valley, Jr. (Eds.), *Cathode Ray Tube Displays*, Massachusetts Institute of Technology Radiation Laboratory Series, Vol. 22, McGraw-Hill Book Co., Inc., New York, 1948.
7. H. W. Hake and E. Averbach, *Apparent and Real Resolution in Radar Visibility*, WADC Technical Report 55-459, Wright Air Development Center, Ohio, 1955.
8. J. W. Ogland, *Behavior of a Cathode Ray Tube Under Pulsed Operation*, AA Report 1364, Westinghouse Electric Corp., Air Arm Division, 1956.
9. Grayson Merrill (Ed.), *Principles of Guided Missile Design*, D. Van Nostrand Co., Inc., Princeton, N.J., 1955.

INDEX

- absorbers; narrow-band, 566
acceleration, mechanical design and, 707-708
acceleration density, 696-697
a-c error signal, 265-266
accuracy requirements, beamwidth and, 90; error measurement and, 80-84; regulatory circuits and, 405-407
acoustic noise, 704-707; *see also* noise
active system, defined, 293; vs. passive, 2
Adler, F. P., 232
aerial photographs, 10
aerodynamic heating, 686
AEW, *see* airborne early warning
AF Janus receiver, 742-743
Aiken, W. R., 658
Aiken tube, 658-659
AI, *see* airborne intercept
airborne early warning (AEW), 1, 46, 50, 53, 725; accuracy of, 80-84; angular resolution in, 168-173; antennas for, 527-528; beamwidth in, 89-92; detection requirements in, 63-65; information-handling capacity of, 84-85; preliminary system design in, 67-69; sensitivity of, 76; summary of requirements for, 96-98; system analysis in, 69; system logic and fixed elements in, 70-73; target resolution in, 75-79
airborne electronic equipment, failure rate in, 720-722; maintenance and repair of, 712-714; reliability of, 715-722; transportation and supply of, 714
airborne ground mapping system, 772-787; *see also* ground mapping
airborne intercept (AI) radar, 46, 725; angle stabilization in, 433; conversion problem analysis in, 116-129; summary of arguments for, 136-137; airborne intercept (AI) radar (*Cont.*)
vectoring considerations and, 111-116; *see also* interceptor
airborne moving target, aural display and, 676-677; *see also* target
airborne navigation, 725-802; *see also* navigation; doppler navigation system
airborne radar, 3; beacon system and, 11-12; design problem of, 27-30, 766-769; equipment development in, 34-35; human operator's role in, 634-651; mechanical design and packaging of, 33-34, 679-723; operating frequencies, 27; primary mission of, 224; system approach to, 30-32; technical analysis of, 32-33; weather mapping and, 766-769 (*see also* weather radar)
aircraft, radar cross section of, 192-197
aircraft carrier, 11; attack diagram for, 64; task force and defense system, 50-55
aircraft motion, errors in, 442-444; regulatory circuits and, 433-439
aircraft radar system, 3; *see also* airborne radar
aircraft rate error, 457-458
aircraft transfer functions, 438
airframe environment, 35
airplane drift, ground mapping and, 776-777
air surveillance system, 15
air-to-air intercept system, 41-43
air-to-air missile system, 50
air-to-missile system, 54-55
air-weather radar systems, 759-772; *see also* weather radar
Allen, P. J., 511, 550 n.
Alluisi, E. A., 649 n.
altitude, *see* height-finding
"altitude hole" problems, 738, 752-754
altitude line, 523

- altitude return, 222-224
 aluminas, 566
 Ament, W. S., 192 n.
 amplifier, high-gain, 601-602; IF, *see* IF amplifier; parametric, 356; power transfer function, 144; RF, *see* RF amplifier; variable parameter, 356; *see also* gain (adj.)
 amplitron, 603-604, 606
 amplitude distribution, cumulative, 193-196
 amplitude fluctuations, 200
 amplitude modulation (AM), 17-18
 amplitude noise, 198-203, 395-396, 479
 AMTI systems, 317
 analogue computers, 744
 analogue model, 37-38
 Anderson, Tore N., 546 n.
 Andrews, E. R., 233
 angle demodulator, 389-390, 414
 angle error, 303
 angle measurement stabilization problem, 429-433, 741
 angle noise, 203-207, 396
 angle tracking, accuracy requirements in, 478-480; in pulsed doppler system, 332
 angle tracking error, 300
 angle tracking loop, 264, 267, 483-484; mechanization of, 492-498; position errors in, 489-491; rate errors in, 486-489
 angle tracking modulation, 424
 angle tracking noise, 398-401; application to evaluation of, 264-269
 angle tracking stabilization, 457-468, 473
 angle tracking system, 474, 480-485
 angular error, 265-266, 290
 angular resolution, calculation of, 138-173; defined, 138; factors affecting, 168-173; in high-resolution radar systems, 333-335
 angular scanning frequency, 296
 anisotropic transmission line, 560
 anode, magnetron, 582-583
 antenna aperture, 334
 antenna beam, rotation of, 9
 antenna beamwidth, 90, 337, 515; in high-resolution radar systems, 333-334; *see also* beamwidth
 antenna control loops, errors in, 442-448
 antenna corner frequencies, 467
 antenna directivity, 514
 antenna disturbance, aircraft motion and, 437
 antenna drift, signal fade and, 462
 antenna drives, 450-451, 481
 antenna gain, 514, 518, 766
 antenna pattern, characteristics of, 169-170; detection probability and, 142-143; ground mapping and, 770
 antenna position, 11
 antennas, 511-579; aircraft motion and, 464; angle tracking error and, 301; anomalous effects in, 523-524; beamwidth and, 90, 333-334, 337, 515; characteristics of, 513, 518; corporate feed structure, 527; db difference, 522; detection probability and, 156; directional, 10; directional couplers, 548; doppler navigation system, 741; dual-feed, 203; duplexing schemes and, 567-573; hybrid junctions, 548-550; fan beam, 524-528 (*see also* fan beam radar); fundamental concepts, 513-515; Gaussian shape, 289; ground mapping, 773, 780; linear array, 524-528; lobe structure, 187-188; lock on, 520; lossless matched, 513; omnidirectional, 17; parabolic reflectors and, 169; paraboloidal, 515-518; pattern simulation and, 520-523; pencil-beam, 188, 515; power pattern of, 161; power requirements, 135; PPI system, 617; radar wave propagation and, 174; radomes and, 531-535; rearward-looking, 15; stabilization of, 430-432, 741, 770; system requirements for, 518-520; tracking function of, 519-520; tracking stabilization and, 440-452; transmission lines and, 535-540; as transmitters, 138-139, 514; two-arm spiral, 528-531; types of, 169; weather radar, 768
 aperture illumination, 516
 APSEL tube, 673
 aquadag, 625
 arc-over, 711
 Atlas, D., 760 n.
 atmosphere, attenuation in, 227-230; refraction of radar waves by, 175, 233-237; standard, 234
 atmospheric duct, 234

- atmospheric noise, 395; *see also* noise
- ATR tubes, duplexing and, 568
- attack diagram, 64
- attack display, 654-655
- attack doctrine, fire-control system and, 120-123
- attack phase system logic, 117-118
- attenuation, 175, 227-230, 391; precipitation and, 230-231, 767; by propellant gases, 231-233
- attenuators, meteorological, 761-762, 767
- A-type presentation, 613-614
- aural indicator, in display system design, 676-677
- autocorrelation, defined, 305
- autocorrelation function, 247-259, 288
- automatic angle tracking, 474
- automatic frequency control (AFC), 350-352, 394, 401-402; in air-weather radar, 770; continuous-correction type, 407-411; design of, 542; instantaneous, 415-416; limit-activated, 412-413
- automatic gain control (AGC), 265, 379-380, 394, 416-428
- automatic space stabilization system, 395
- automatic tracking, general problems of, 471-474
- automatic tracking circuits, 471-510; accuracy requirements in, 479-480; angle tracking and, 480-489; design considerations, 508-510; external inputs to, 475-479; general problem of, 471-474; range and velocity tracking and, 498-510; range tracking design example, 505-508; servo system transfer function, 502-504; tracking loop design, 485-486; tracking loop network and amplifier, 496-497
- Axelby, George S., 446 n., 458 n., 467 n., 471 n.
- Axford, W. E., 233 n.
- back-scattering, 10, 174
- back-scattering coefficient, 760-761, 766
- back-scattering cross section, 176
- back-up coverage, 74
- backward wave oscillator, 602
- Bacon, R. H., 735 n.
- balance point, 303
- bandpass characteristic, pulsed doppler system, 331-332
- bandpass detection filters, 167
- bandwidth, angular resolution and, 333; angular scanning and, 319; automatic tracking and, 474; carrier frequency and, 260; detection minimum, 299; discriminator design and, 415; dynamic response of receiver and, 375-376; filter, 327, 335; gain product, 365-367; of IF amplifier, 365; of infrared sensing systems, 799; in intensity-modulated display, 618, 633; of Janus and non-Janus systems, 750-751; noise and, 260, 268, 285; predetection and postdetection and, 167-168; radome design and, 533; scan modulation and, 290; in tracking-loop design, 485; transmitted, 295; voltage reflector and, 543; wide, 336-338
- bandwidth capability, 543-545
- Barnes, J. F., 242 n.
- Barnett, E. F., 555 n.
- battle-control system, 53, 66; *see also* fire-control system
- beacon system, 11-12
- beam attacks, accuracy requirements for, 125-129
- beam deflection, cathode ray tube, 624-625
- beam power pattern, ground mapping, 774-780
- beam-shaping tubes, 671-673
- beam splitting, 96
- beam-type amplifiers, 599-601; *see also* klystron; magnetron
- beamwidth, 151; accuracy of, 80-82; angle noise and, 207; angular resolution and, 169, 333; antenna pattern and, 143; cathode ray tube and, 632; defined, 170, 515; detection probability and, 156; effective, 173; ground mapping and, 774; height-finding system and, 95; scanning and, 159-160; sea return and, 212; tactical problem and, 89-92; velocity bias error and, 748; vertical, 91-92
- beating frequency, 361
- Bessel function, 169, 261, 521-522
- Bessemer, C. W., 634 n.
- bias errors, angle tracking loop, 489-490; automatic tracking, 479;

- bias errors, (*Cont.*)
 predictable, 123, 128
- bidirectional coupler, 548
- binocular acuity, 646
- Birdsall, T. G., 308 n.
- black body temperatures, 799-800
- Blake, L. V., 162 n.
- blind spot, 637
- blip-scan curve, 329
- blip-scan ratio, 144
- Boltzmann's constant, 353, 750
- boxcar detectors, 257
- branched duplexer, 568
- breakaway display, 655
- Brewster angle, 183
- bright band effect, 761
- brightness, contrast and, 631, 641;
 defined, 630, 635; electroluminescent,
 666; range of, 618
- bright spot wander noise, 396, 400
- Brown, R. M., 554 n.
- brute force jamming, 15
- B type presentation, 615-616
- Burdine, B. H., 530 n.
- Burrows, C. R., 182 n.
- Cacheris, J., 563 n.
- Calcote, H. F., 232 n.
- camouflaged target, 10
- Campbell, J., 747
- CAP, *see* combat air patrols
- capacitor, electroluminescent, 669; ferroelectric, 669-670
- Caplan, P. J., 229 n.
- Capps, F. M., 182 n.
- Carcinotron, 602-603, 605
- Carl, J. M., 643 n.
- carrier, aircraft, *see* aircraft carrier
- carrier deck space, 74
- carrier frequency, 26-27
- cascade circuit, 370
- cascode input amplifier, 380
- cascode preamplifier, 797
- cathode ray tube, beam deflector in, 624-625; beam-shaping type, 671-673; in display system design, 621; in ground mapping, 781-782; intensity modulated, 630
- cavity resonator, 416, 582, 590-591
- channels, number of, 297-299
- Charactron tube, 671-672
- Chestnut, Harold, 492 n.
- chromatic aberration, 173
- Chromatron tube, 663-664
- "chunks," of information, 620
- CIC, *see* combat information centers
- circle of confusion, 786
- circulators, 559-560, 572
- clamping circuits, 257-258
- Clavin, Alvin, 563 n.
- Clementson, G. C., 438
- cloud clutter, 10-11, 35
- clutter power, 271
- clutter problem, 35, 322; solutions to, 224-227; *see also* ground clutter; sea clutter
- clutter spectra, pulsed doppler system, 322
- coaxial lines, 537
- coherency, in doppler navigation systems, 738
- coherent detector, 298
- coherent radar transmitter, 738
- Cohn, Seymour B., 543 n.
- cold plate cooling, 687
- collision course, 132
- collision guidance, 120
- collision vectoring, 72-73, 108-111, 120
- color differentiation, human eye, 10, 638
- color transposition, mapping and, 10
- color tubes, in display system design, 661-665
- combat air patrols (CAP), 53, 63, 76
- combat information centers (CIC), 15, 71-72, 85; height finding and, 95; target resolution and, 76
- comb filter, 279, 376
- communications systems, 53, 66; radar system and, 12-13
- conical scanning, 515, 519, 523
- contact analogue display, 675-676
- continuous correction AFC, 407-411; *see also* automatic frequency control
- continuous wave (CW) systems, 20, 293, 311-320, 737-737; duplexing in, 567; *see also* FM/CW systems
- contrast, brightness and, 631; defined, 641
- control electrode, 600
- control loop errors, 442-448
- conversion probability, 132-134
- conversion problem, analysis of, 116-129; lock-on range and look-angle requirements, 130-135

- cooling system, 686-691, 711
 corner frequency, bandwidth and, 485
 corporate feed structure, antenna array, 527
 correlation radar, 273, 293, 305-308
 corrosion, chemical and electrochemical, 708
 cosecant-square beam, 773, 779
 countermeasures system, 2
 course dynamics, 475-477
 Cox, C., 216 n.
 Crandell, P. A., 554 n.
 Crede, Charles E., 699, 703
 cross-correlation, defined, 305
 cross-correlation detector, 307-308
 cross coupling, angle tracking loop and, 492-495; tracking problem, 474
 cross modulation, radar receiver, 350
 cross polarizing, 179
 crystal diode, 357
 crystal mixers, 355, 357-362; *see also* mixer
 crystal video receiver, 742
 C type presentation, 616
 cutoff bias, 428
 Cusano light amplifier, 667
 CW radar system, 312, 315-316, 324; *see also* continuous wave radar
 cycle time, aircraft, 74
 cyclic stress, 701
- Damonte, J. B., 521 n.
 darkness, 10
 data processing, detection system and, 297-300; ground mapping and, 781-782; tactical problem and, 99
 data quantization, 80-83
 data stabilization, 80-83, 429-430, 741-742
 data storage and read-out, 612-613
 Davis, I. L., 308 n.
 Davis, R. C., 308 n.
 db difference, antenna design and, 522
 deceptive jamming, 15
 decibels, sound pressure measurement and, 705-706; *see also* db
 decision element, 144, 147-149
 decoupling circuits, 378
 deflecting tubes, 658
 Delano, R. H., 205 n.
 delay-line video integration, 309
 delta functions, *see* impulse functions
 demodulation, 25, 285-286, 374, 492-494
 demodulator, 266, 479; antenna drive and, 451
 demodulator circuits, 390-391
 density modulation, 591-592
 depression angle, sea return and, 212, 216
 depth of field, 680-682, 786
 design, environment and, 680-682; mechanical, *see* mechanical design and packaging
 detectability, pedestal amplitude and, 645
 detection, 25
 detection and data processing system, 297-300
 detection and identification, display systems and, 609
 detection probability, 138-173; cumulative, 75, 144, 151; multiple, 156-162; nonfluctuating target and, 155; of pulsed doppler radar, 162-168; of pulse radar, 141-156; single-scan, 149-151
 dielectric constant, 181-182
 digital computers, 744
 directional couplers, 548
 discrimination, IF, *see* IF discrimination
 discriminator design, 414-415
 disk of confusion, 786
 displacement transmissibility, 698
 display errors, ground mapping radar, 776-777
 display information, uses of, 608-613
 display integration, 113, 631-632
 display requirements, 112-113
 display size, resolution and, 627-629
 display system design, attack display and, 654-655; aural indicators in, 676-677; beam-shaping tubes and, 671-673; breakaway display, 655; cathode ray tube in, 621-627; color tubes in, 661-665; electroluminescence and, 665-671; flat tubes in, 658-659; human operator factors in, 634-651; man-machine interrelationships in, 645-651; physical environment and, 634; requirements for, 651-655; search and vectoring display, 651-654; special display devices, 655-677; terrain avoidance and, 673-674; windshield projection displays, 674

- display systems, amplitude-modulated, 613-615; combination, 619; control loop and, 611; design problems of, 607-677; ground mapping and, 781-782; input information types, 619-621; types of, 613
- divergence factor, 191
- doppler equation, 729
- doppler effect, 6, 15; discrimination by, 10
- doppler filter, 167, 324, 331
- doppler frequency, 26, 181; FM deviation from, 314; noise and, 270
- doppler navigation systems, 725-759; accuracy in, 734-736; "altitude hole" and, 752-754; basic principles, 728-833; coherency in, 738; computations possible in, 745-746; errors caused by interaction with ground and water, 746-749; flight test results of, 755-759; forms of, 733-736; frequency trackers in, 742-743; heading reference in, 746; low-altitude performance of, 752-754; major characteristics and components of, 736-746; navigation computer and, 744-746; performance data for, 755-759; polarization used in, 738-739; radar range equation and, 749-752; stabilization of, 741-742; transmission frequency of, 738; transmission types, 736-738; *see also* pulsed doppler systems
- doppler radar systems, 22, 162-168, 293, 742; clutter problem and, 227; described, 727-728; design of, 280; sea return in, 217-219; target motion and, 180-181; *see also* doppler effect; doppler navigation system
- doppler shift, 22, 26, 166, 308, 729, 731
- doppler signal-to-noise ratio, 750
- doppler spectrum, 217
- doppler techniques, use of, 317
- Dorr, R., 663-664
- double-ring strapping, magnetron anode, 583
- Dougherty, J. P., 502 n.
- Douglas sea state, 747-748
- drift, 394; antenna, 462
- drift angle, doppler navigation system, 743-744
- drift errors, ground mapping, 776
- Dunn, J. H., 395 n., 422 n., 479 n.
- duplexed-active operation, 317
- duplexing problem, 566-573; doppler navigation system and, 753
- dynamic lag, automatic tracking, 479
- dynamic lag errors, 124
- dynamic system, 2, 41
- early warning detection and interception vectoring, 52; *see also* airborne early warning (AEW)
- earth-flattening procedure, 233
- earth's curvature, effect of, 190-191
- East, T. W. R., 230 n.
- echo, radar, 174-175; multiple-time-around, 220-221
- echoing area, 176
- eclipsing, 162-166; in pulsed doppler systems, 328-329
- ECM, *see* electronic countermeasures
- Ektron cell, 346
- electrical-to-light transducers, 627-634
- electrochemical corrosion, 709
- electroluminescence, in display system design, 665-671
- electromagnetic radiations, 2, 35-36; *see also* radio waves
- electron beam, focussing of, 658-659
- electronic equipment, airborne, *see* airborne electronic equipment; cooling of, 687-688
- electron gun, 622-624; storage tubes and, 656
- electronic countermeasures (ECM), 2, 55-56
- electronic environment, 35-36
- electron motion, magnetrons and, 580-582
- electrostatic storage, 310
- elevated duct, 235
- elevation coverage, 114
- Emerson, W. H., 566 n.
- endurance limit, stress and, 701
- enemy target, detection of, 5; *see also* target
- energy density spectrum, 240
- energy transfer function, 242
- envelope detector, 382-385
- environment, effects of on human operation, 644-645; launching zone parameters and, 120; natural vs. induced, 681
- environmental factors, 35-36; frequency deviations and, 402-403; mechanical

- environmental factors, (*Cont.*)
 design and, 680-682
 equipment, airborne radar, 34-35
 error, ground mapping and, 776; measurement of, 39-40
 error signal, angle tracking and, 480-482; noise and, 265; tracking problem and, 473
 error specification, in fire-control system, 124-129
 error voltage, 204
 explosion, mechanical design and, 710
 eye, characteristics of, 635-644
- fading, in automatic tracking, 476-477
 failure, reliability measurements and, 717-720
 false-alarm number, 145
 false-alarm probability, 144, 148-150; noise and, 279-280
 false-alarm rate, 318
 fan-beam antenna, 524-528
 fan-beam radar, 71, 95-96, 773; ground mapping and, 779-780
 fan beamwidth, *see* beamwidth
 Faraday effect, 558-559
 Faraday rotator gyrotor, 559-561
 fatigue failures, vibration and, 700
 Federal Communications Commission, 26
 feedback, 34, 350; antenna loop and, 450; in automatic gain control, 417-420; circuit stability of, 309; in IF amplifiers, 366; in IF preamplifiers, 371; local oscillator characteristics and, 413
 feedthrough, 318
 Fellars, R. G., 544 n.
 Ferret application, radiation detection, 13
 ferrite circulators, 331
 ferrite devices, 557-566
 ferrite isolator, 563-565
 ferrite phase shifters, 562-563
 ferrite switches, 331, 561-562
 ferroelectric capacitors, 669-670
 field of view, 786, 788
 fighter-bomber fire-control problem, 9
 fighter vectoring guidance, 72-73
 film resolution, 786-787
 filter amplifier, 362
 filter-bank detector, 312
 filtering techniques, 167-168, 297-300, 375-376
 fire-control problem, 8
 fire-control system, 104; antenna and, 461, 512-515; error specification in, 123-129; height-finding and, 94; parameters of, 120-123
 Fitts, P. M., 649 n.
 flat random motion spectrum, 697
 Fleet Center, 15
 flexibility, mechanical design and, 715
 fluctuating target, detection of, 151-156
 fluctuation error, 729
 fluorescent screens, cathode ray tube, 625-626
 flyability error, 125
 FM/CW systems, 311-320; in doppler navigation systems, 737-738, 752-754; *see also* continuous wave systems; frequency modulation
 "FM-ing," 295; frequency of, 313
 folded-tie hybrids, 549
 folding effect, 323-324, 327
 foldover, in doppler navigation system, 750
 foot lambert, 630, 635, 656
 forced vibration, 695
 force transmissibility, 698
 forward-looking mapping radar, 9, 773
 Fourier analysis, 238-242, 434
 Fourier integral, 244
 Fourier series, 18 n., 24, 239
 Fourier transform, 244, 249-250, 255, 273 n.
 Fourier transform pair, 239, 241
 fovea, 636-638
 Fox, A. G., 555 n., 562 n.
 Freedman, J., 168 n., 784
 Freeman, J. J., 198
 frequencies, sampling, 295-297
 frequency, doppler, *see* doppler frequency; mapping and, 10; operating range of, 26-27; radar, 174; weather mapping, 766-767
 frequency control, *see* automatic frequency control
 frequency deviations, 432-433
 frequency discriminator, 415, 742
 frequency domain, transformation to, 502-503
 frequency modulation, 18-20; in CW systems, 313-315; *see also* FM/CW

- frequency modulation, (*Cont.*)
 systems
- frequency-performance parameters, 28
- frequency search and acquisition, 416
- frequency shift, 22; *see also* doppler shift
- frequency spectrum, 18
- frequency tracker, 730, 742-743
- Fresnel integrals, 178
- Fried, W. R., 725 n.
- "friendly" aircraft, IFF and, 5
- Fris, H. T., 514 n.
- fungus damage, 710
- gain bandwidth product, 365
- gain variation and setting, 375
- Gardner, M. F., 242 n.
- Gardner, R. E., 643 n.
- gases, atmospheric, 175
- gas tube duplexer, 567-570
- gating, range, *see* range gating
- gating circuits, radar receiver, 386-387
- Gaussian noise process, 246, 253, 309;
see also noise
- George, S. F., 292 n., 502 n.
- Gerwin, H. L., 198 n., 398 n.
- ghosts, pulse repetition frequency and,
 330-331
- gimbal angle, maximum, 104, 112-113
- gimbal system, angle tracking and, 480
- glint noise inputs, 478-479; *see also* angle
 noise
- Goetz, M., 679 n.
- Goldberg, H. B., 62 n., 550 n., 734 n.
- Goldstein, H., 211 n.
- Gould, Lawrence, 545 n.
- Gow, J., 663-664
- Graham, R. E., 503 n.
- Grant, C. R., 215 n., 216, 222
- grass and weeds, 221
- Greenberg, Harold, 41 n., 80 n.
- gross aspect variation, 195
- ground, as pure dielectric, 183; reflection
 coefficient for, 186; scattering co-
 efficients for, 747
- ground-based radar system, 3
- ground clutter, 35, 181, 219, 317; can-
 cellation of, 270
- ground mapping, 219, 764; angular
 resolution and, 168
- ground-mapping display, 628
- ground mapping system, active airborne
 systems and, 772-787; basic principles,
 ground mapping system, (*Cont.*)
 772-774; characteristics and com-
 ponents of, 778-782; design considera-
 tions and, 774-778; future of, 787;
 new developments in, 801-802; radar
 range equation in, 782-784; receivers
 for, 781; resolution limits in, 784-787;
see also infrared reconnaissance
- ground return, 175, 181-190, 219-222;
 antennas and, 523; vertical beam-
 width and, 92
- ground return systems, 725-802
- ground speed, doppler navigation sys-
 tem, 743-744
- guidance phase, in interception, 101
- guidance systems, AI radar require-
 ments and, 135-137; semiactive, *see*
 semiactive guidance systems
- guided missile launching zone param-
 eters, 118-120
- guided missile systems, semiactive sys-
 tems and, 14
- Guillemin, E. A., 502 n.
- Gunn, K. L. S., 230 n.
- gust disturbance, 438
- gyration, radius of, 284-285
- gyrator, microwave, 559
- gyromagnetic resonance, 558
- gyroscopes, hermetic integrating, 456;
 mounting of, 469; rate, *see* rate gyro
- Haddock, F. T., 231
- hail, avoidance of, 763
- halation, 629
- half-wave plates, 555
- Hall, W. M., 148 n.
- Hansen, W. W., 273 n.
- Harman, J. C., 182 n.
- harmonic generation, 524
- Harrison, H. T., 763 n.
- Hartree voltage, 585
- Hastings, A. E., 198 n., 398 n.
- haze, 10
- heading, measurement of, 99
- heading commands, in display system
 design, 651-653
- heading error, 108-111, 132; *see also*
 error; steering error
- heading reference, doppler navigation
 systems, 746
- head-on attacks, accuracy requirements
 for, 125-128

- Healy, D. J., III, 394 n.
 heat exchanger, air-to-air, 711
 heat-flow calculations, 689-691
 heat sinks, 687
 height-finding radar, 70-71; factors affecting, 92-96, 181
 Helmholtz, Robert H., 41 n., 62 n., 80 n., 734 n.
 heterodyne receivers, 511
 heterodyning, types of, 25
 high-resolution radar systems, 293, 333-338
 Hippel, A. R. von, 565 n.
 histogram, flight test, 755
 Hoel, P. G., 147 n.
 Hogan, C. L., 558-559 n.
 hold-in, 409, 411
 homing system, 13
 Hopkins, L., 292 n.
 "hot" targets, infrared sensing of, 790-791
 Howard, D. D., 198 n., 394 n., 395 n., 398 n., 400 n., 422 n., 479 n.
 Huggins, W. H., 363 n.
 Hull, J. F., 565 n.
 human eye, characteristics of, 635-644
 human operator, in display system design, 634-651; environment effect on, 644-645; reaction time of, 648-649; role of, 147-148, 607
 hybrid junction, 548-550
 iconic model, 37-38
 identification friend or foe (IFF), 4-6
 IF admittance, 361
 IF amplifier, automatic gain control and, 417; control characteristics, 427-429; design considerations, 363-368; gating of, in pulsed doppler systems, 326; mixer and, 550-551
 IF bandwidth, 351-352, 406
 IF characteristics, dynamic error and, 414
 IF discriminator, 409-410
 IF envelope detection, 374, 379
 IFF information, displays and, 619; *see also* identification friend or foe
 IF Janus receiver, 742-743
 IF noise figure, 356
 IF preamplifier design, 368-373
 IF signal, error signal and, 265
 Illinois, University of, 218
 illumination, radar systems and, 13-15
 illumination effectiveness, 514
 image frequency, 359-360
 image intensifiers, 667-668
 impedance, nonlinear, 23-24
 impedance matching, 540-543
 impulse functions, 243-245
 impulse response, 241
 indicator brightness, 629-631
 information, "chunks" of, 620; spurious, 5; types of, 294-295
 information handling capacity, AEW system, 84-85
 information relay, radar systems and, 15
 information subcarriers, 20-21
 information utilization, 300
 infrared detection, evaluation of, 798; sensitivity of, 796
 infrared reconnaissance (IR), 11, 787-802; airborne radars, 338-346; basic principles of, 339-343; 788-791; system features, 791-801
 input information, 619-621
 inputs, vs. outputs, 40-41
 input variables, 43
 insertion losses, 331
 installation, maintenance and, 712-714
 installation environment, 3
 instantaneous field of view, 788
 instantaneous frequency, 415
 integrating gyro, 455-457, 466, 469
 intensity, defined, 635
 intensity-modulated displays, 615-618, 633
 interception, three phases of, 101
 interceptor aircraft system (AI), 1, 14, 50, 53-54; *see also* airborne intercept radar
 interceptor-bomber duel, 44
 interceptor kill probability, *see* kill probability
 interceptors, back-up, 66; number of, 63, 68, 72, 80; time delay in launching, 66
 interceptor system, air-to-air, 41-43; beamwidth and, 90; interrelated characteristics of AEW/CIC system and, 79; reliability of, 103-104; study model of, 100-103
 interceptor velocity, 65
 interference, direct and reflected, 187; filter-mixer combination, 362
 internal noise inputs, 396; *see also* noise

- interstage coupling network, IF amplifier, 364-365
 Irdome, 339
 isodops, 731
 iso-echo contour circuitry, 769
 isolator natural frequency, 703
- Jacobson, R. H., 705
 James, R. W., 225 n.
 jamming, radar systems and, 2, 15-16
 Janus system, 731-732, 739-741; errors caused by interactions with ground and water, 746-749
 jittered range gate method, 501
 Johnson, C. M., 216 n.
 Johnson, J. B., 353
 Johnson, J. C., 766 n.
 Jones, R. F., 763 n.
 Jones, R. N., 182 n.
 J type display, 615
- Kagan, B. 668 n.
 Kahn, A., 607 n., 649 n.
 Kales, M. L., 562 n.
 Katzin, M., 192 n., 212 n., 216
 Kay, A. F., 533 n.
 K band, 779; *see also* transmission frequency
 Kerr, D. E., 211 n., 760 n., 765 n.
 kill probability, interceptor, 58-59, 62-68, 100-103; fire-control system and, 120; vectoring probability and, 105
 King, A. M., 395 n.
 Kirby, R. S., 182 n.
 klystron amplifier, 590-592
 klystron oscillator, 592-593; control characteristics of, 408
 klystron tubes, 512; frequency deviations of, 402-403; high-power characteristics of, 593-596; vs. magnetrons, 596-597
 Kraus, John D., 169 n., 513 n.
 K type display, 615
- lambert, defined, 635
 land, reflection coefficient for, 186; *see also* ground
 land-sea switch, 748
 Laplace integral, 503
 Laprade, R. H., 725 n.
 launching errors, maximum allowable, 123
 launching zone parameters, 118-120
 Laws, J. O., 231, 761 n.
 Lawson, J. L., 147 n., 193 n., 238 n., 261 n., 376 n.
 Lax, Benjamin, 558 n.
 lead-collision system, 120-124, 130-132
 lead-pursuit course inputs, 476
 LeCraw, R. C., 562 n.
 Lewis, B. L., 198 n., 398 n., 400 n.
 Lewis, W. D., 514 n.
 light amplifiers, 667-668
 limit-acuated automatic frequency control, 412-413
 linear array, 524-528
 linear balanced duplexer, 571
 linear receivers, 273
 lin-log amplifier, 781
 load factor, acceleration and, 707
 loading mismatch, 403
 lobe, 17; interference and, 187-188
 lobing frequency, 477-478
 lobing system, simultaneous, 203
 local oscillator characteristics, 413-414
 Locke, A. S., 14 n., 504 n., 745
 lock-on, 320; antenna design and, 520; conversion probability and, 134-135; false, 520, 524; on image, 207; search volume and, 113
 lock-on range, 104, 110, 130-135
 logic, in weapons system model, 43, 67
 logistics environment, 36
 look-angle, 104-106, 112-113, 130-135
 low-altitude performance, doppler navigation system, 752-754
 L type display, 615
 luminance, 635
 luminescence, 626
- MacDonald, F. C., 192 n., 214-215 n.
 McElvery, R. M., 530 n.
 maintenance and installation, 712-714
 magnetic circuits, 557-565
 magnetic tape storage, 310
 magnetron pulling, 403-405
 magnetrons, energy reflection and, 541-542; ferrite isolators and, 565; fixed-frequency, 416; frequency deviations of, 402; vs. klystrons, 596-597; loading effect in, 588; mode skip in, 587; moding in, 585-587; noise in, 588-589; as oscillator, 580-590; resonant system and, 582-583; typical perform-

- magnetrons, (*Cont.*)
 ance of, 589-590
maneuver capability, of target, 55
mapping radar, 9-11; forward-looking, 773; *see also* ground mapping
Marcum, J. I., 146 n., 150 n.
Marshall, J. S., 760 n.
masers, 574-576
Massachusetts Institute of Technology, 530
master oscillator power amplifier (MOPA), 406
matched filter radar, 337-338; application to analysis of, 272-281
matching elements, 543
mavars, 576-578
Mayer, Arthur, 454 n., 495 n.
Mayer, Robert, 492 n.
Mazina, M., 649 n.
M curve, standard, 234-236
Meade, J. E., 398 n.
Meade, J. W., 198 n.
measurement errors, 39-40; standard deviation or rms, 81; *see also* errors
mechanical design and packaging, 679-723; acceleration and, 707-708; explosion hazard and, 710; military specifications and, 682-683; moisture and, 708-710; potential growth and, 715; pressure and, 711-712; reliability in, 715-722; solar radiation and, 692; sound pressure and, 705-706; static electricity and, 710; temperature and, 683-691; vibration and shock, 694-704
mechanization requirements, 33-34
Merrill, Grayson, 41 n., 62 n., 80 n., 734 n.
meteorological attenuators, 761-762
meteorological effects, 227-237
meteorological scatterers, 760-761
Meyer, M. A., 550 n.
Middleton, D., 273 n.
microwave circulator, 572
microwave components, miscellaneous, 546
microwave energy, mapping and, 10-11
microwave ferrite devices, 557-565
microwave frequencies, 760-763
microwave materials, 565-566
microwave noise sources, 553-554; *see also* noise
microwave power, generation of, 580-606
microwave power, (*Cont.*)
 606
microwave propagation, 13; *see also* radio waves
microwave thermal mapping, 10-11
microwave transmission lines, 545-546
military objective, operational requirement and, 31
military specifications, design and, 682-683
minimum discernible signal, 768
missile guidance, 135-137
missile performance, 119
missile radar system, 3
missile salvo kill probability, 63
mission accomplishment, airborne radar design and, 27-28; goals and, 43, 57; level of, 31; operational requirements and, 33, 49, 57; time response and, 41
mixers and mixing, 512; balanced, 551-552; coherent and noncoherent, 25; coupling of, 361-362; purpose of, 550-553
models, weapons system, 36-39
mode multiplexing, 571
modes, primary and secondary, 3-4
mode shift, 587
modular design, 714
modulating anode, 600-601
modulation and demodulation processes, 3; target, 20-23; types of, 16-25
moisture, mechanical design and, 708-710
monopulse angle tracking, 300-305, 332
monopulse radar, 293, 301
Montgomery, C. G., 548 n., 567 n.
Moore, Richard K., 224 n., 544 n.
moving target indicator (MTI), 10, 238, 293, 601-602; application to analysis of, 269-272
MTR, *see* microwave thermal mapping
Mullen, J. A., 544 n.
Muller, P. F., Jr., 649 n.
multimode operation, 3
multiple-time-around echo, 220-221
multiplexing, 571
Munk, W., 216 n.

narrow-band noise, 258-264
narrow-band receiving system, 337
naval carrier task force, 1
Naval Research Laboratory, 192, 207,

- Naval Research Laboratory, (*Cont.*)
 400 n., 749
- navigation, airborne, *see* airborne navigation; doppler, *see* doppler navigation system; radar systems and, 11-12; self-contained, 726-727
- navigation computer, 744-746
- network function, 363
- network weighting function, 241
- Neumann, G., 225 n.
- Neyman, J., 308 n.
- Nicoll, F. H., 668 n.
- nodding, 92
- noise, amplitude, angle and range, 198-208; in magnetrons, 588; mathematical, 245 *ff.*; microwave sources, 553-554; narrow-band, 258-264; receiver, 353-355; "white," 244
- noise analysis, 238-291
- noise equivalent power, infrared detection, 796
- noise figure, radar receivers, 353-355, 390-391
- noise inputs, automatic tracking, 478-479; external and internal, 395-401
- noise power, 140, 278
- noise processes, 246; *see also* Gaussian noise processes; signal-to-noise ratio
- noncoherent detector, 298
- nonequilibrium writing, 657
- non-Janus system, 731, 739-741
- nonlinear impedances, 23-24
- nonlinear operations, 253-258
- North, D. O., 273 n.
- notch networks, 451 n.
- nuclear radiation, mechanical design and, 692
- nuclear weapons, 56
- null point, 301, 426
- Oliver, B. M., 419 n.
- O'Neal, N. V., 198 n.
- open center PPI, 617
- operating frequency, airborne radar system design and, 28; range of, 26-27
- operational requirement, 49; airborne radar design and, 31-32; mission accomplishment and, 33, 57
- operator factor, 147-148; *see also* human operator
- oscillation, in magnetrons, 585-590
- oscillator, local, *see* local oscillator
- outputs, vs. inputs, 40-41
- overshoot, allowable, 411, 485
- oxygen, attenuation due to, 228
- oxygen deprivation, effect on human operator, 645-647
- packaging, mechanical design and, 679-723
- Page, R. M., 292 n., 398 n.
- "painting," *see* scanning
- Palmer scan, 160
- Pappas, N. L., 545 n.
- parabolic reflectors, antennas and, 169; properties of, 516
- paraboloidal antenna, 515-518
- paramagnetic amplifiers, 575
- parameters, choice of, 11, 28-29, 50; guided missile launching zone, 118-120; IF preamplifier design, 371-372; infrared mapping systems, 343-344, 791-793, 795; radar range requirements, 138-141; radar target, 319; in signal/noise analysis, 279-280; summary of, 97-100; target, 281 *ff.*; "trade-off" between, 62, 67; of vectoring system, 106-108
- parametric amplifiers, 356, 576
- Parseval's equality, 240, 250
- Parsons, D. A., 231, 761 n.
- Passerini, H. J., 192 n.
- passive systems, vs. active, 2, 13; defined, 293
- Pearson, E. S., 308 n.
- pedestal amplitude, detectability and, 645
- pedestal amplitude voltage, 643
- pencil-beam antenna, 185, 515
- periodic functions, Fourier series and, 239
- permeability, ferrite devices, 557-565
- Peterson, W. W., 308 n.
- phase modulation, 20
- phase shifters, 555-556, 562-563
- phase shifts, 424
- phasors, 175, 179
- phosphor, photoluminescent, 665-666
- phosphor films, transparent, 659-661
- photographs, aerial, 10
- physical environment, 35; *see also* environment; environmental
- Pierson, W. J., Jr., 235 n.
- Pisa indicator, 304

- pitch, roll and yaw, 434, 797
 plan position indicator (PPI), 216-217, 616-617, 673, 771
 platinotron, 603-606
 polarization, circular, 554-556; in doppler navigation systems, 738-739; effect of on reflection, 179-180; ground mapping and, 779; ground reflection and, 182; sea return and, 212
 polarization dependence, 187
 position data, scan-to-scan correction of, 88
 position error, 87; *see also* error
 Post, E. A., 763 n.
 postdetection filtering, 167-168
 postdetection integration, 309
 Potter, R. S., 550 n.
 Pound, R. V., 550 n.
 Povejsil, D. J., 292 n.
 power, noise and, 247, 259, 278
 power density, 289; transmitter and receiver, 138-139
 power density spectrum, 248-253, 263
 power divider, 548
 power envelope, 170
 power transfer function, 144, 242
 PPI indicator, air-weather radar systems, 771
 PPI presentation, 673
 PPI range and bearing, 216-217
 preamplifier, cascode, 797; IF, *see* IF preamplifier
 precipitation, attenuation and backscattering by, 230-231, 760-761, 767
 predetection amplifier, 143
 predetection doppler filter, 167
 predetection filter, 145
 prediction, in guidance system, 110; *see also* target position
 pressure, mechanical design and, 711-712
 preventive maintenance, 712
 Pritchard, W. L., 544 n.
 probability curves, 75
 probability density functions, 146-147, 155, 245, 253, 261-262
 probability detection, *see* detection probability
 probability distribution, 39-43
 probability of kill, *see* kill probability
 procurement agencies, 3
 product building blocks, 23
 product demodulator, 256
 product rule prediction, 720
 propagation anomalies, 36
 propagation factor, 175; oscillation of, 188
 propellant gases, attenuation by, 231-233
 propeller modulation, 192, 200-202
 pull-in, 403, 409-410
 pulsed doppler systems, 293, 309, 317, 320-333; applications of, 332-333; design problems of, 331; detection probability in, 162-168; eclipsing in, 328-329; filtering systems in, 331; range gating in, 325-327; range performance of, 327; range measurements in, 329-331; *see also* doppler (adj.)
 pulse integrator, 143, 277-278, 376
 pulse length, 80-81, 151, 333, 411, 768
 pulse modulation, 18, 293; clutter problem and, 224; detection probability calculation in, 141-156; duplexing in, 567; ground return in, 219-220; scanning and, 161-162; *see also* pulsed doppler systems
 pulse packet size, resolution and, 784
 pulse radar spectrum, 336
 pulse repetition frequency (PRF), 18, 296, 323, 326-327, 330, 371, 617, 754; brightness and, 632-633
 pulse stretchers, 257, 265-266, 268, 387-388
 pulse-time modulation, 20; *see also* pulse modulation
 pulse trains, 19, 21, 277-278
 pulse-width modulation, 20
 Pyroceram materials, 566
 quantization errors, 82, 95
 quarter-wave plates, 554
 quarter-wave transformer, 542
 radar, basic capabilities of, 10; characteristics of, 4; derivation of word, 2; dynamics of, 2; frequencies used in, 174; height-finding in, *see* height-finding; interceptor-borne, 14; maintenance and repair, 712-714; matched filter, 272-281; as part of balanced system, 2; state-of-the-art, 27, 49, 69; transportation and supply, 714; two-

radar, (*Cont.*)

way transmission characteristic of, 15
 radar area or cross section, 176, 180
 radar beacon system, 11-12
 radar beamwidth, *see* beamwidth
 "radar buster," 13
 radar cross sections, 209-210
 radar designer, first task of, 27
 radar detection probability distribution, 75
 radar echo, 174-175, 180; *see also* echo;
 radar waves, reflection of
 radar equation, 175
 radar field, 178
 radar frequency components, *see* RF
 components
 radar information, 609; *see also* display
 system
 radar length, 176
 radar mapping, *see* ground mapping
 radar pictures, 10
 radar range, idealized, 141; problem of,
 749-752
 radar range equation, 138-141; ground
 mapping and, 782-784; weather prob-
 lem and, 764-766
 radar range tracking, *see* range tracking;
 tracking
 radar receiver, 347-393; amplifier gain,
 373-374; angle demodulation in, 389-
 390; automatic gain control and, 379-
 380; bandwidth and dynamic response
 in, 375-377; characteristics of, 350-
 351; envelope detector in, 382-385;
 gain variation and gain setting in,
 375; gating circuits in, 386-387;
 general design principles, 347-352;
 generic and product-building blocks
 in, 23; ground mapping with, 781; at
 high-input power levels, 380-381; in-
 terdependence of components in, 352-
 353; measurement problems in, 390-
 392; mixer in, 550-553; noise and
 sensitivity measurements in, 390-393;
 pulse stretching circuits in, 387-388;
 receiver noise figure of, 353-355;
 regulating and tracking circuits in,
 388-389; RF components of, 511-512;
 signal modulation in, 16-25
 radar systems, airborne design problems
 and, 27-30; basic principles of, 293-
 300; classifications of, 2-3; communi-

radar systems, (*Cont.*)

cations and, 12-13; complexity of, 34;
 coverage of, 4; detection performance
 by, 272; decision element in, 144; dis-
 play system design and, 619-621;
 environments of, 35-36; extraction of
 target information from, 23; func-
 tional characteristics of, 3-16; generic
 types and techniques of, 292-346;
 high-resolution, 333-338; identifica-
 tion friend or foe (IFF), 4; illumina-
 tion and, 13-15; information matrix
 in, 294; information relay and, 15;
 installation environment, 2-3; jam-
 ming and, 15-16; mapping by, 9-11
 (*see also* ground mapping); modula-
 tion of, 16-25; multimode operation,
 3; navigation and, 11-12 (*see also*
 doppler navigation systems); noise
 inputs to, 395-401; operating carrier
 frequency of, 26-27; passive vs. active,
 13; primary and secondary modes,
 3-4; pulsed-doppler, 320-333; radia-
 tion detection by, 13; reliability of, 34
 (*see also* reliability); scientific re-
 search and, 16; search and detection
 by, 4; subcarriers and, 20-21; tracking
 by, 6-9 (*see also* tracking); types of,
 293
 radar waves, reflection of, 181-190
 radiation, heat transfer through, 687;
 nuclear, 692-694; solar, 692
 radiation detection, 13
 radio frequency (RF), 4, 26; *see also*
 RF (adj.)
 radio hole, 236
 radiometers, mapping, 173
 radio waves, propagation of, 174; reflec-
 tion and transmission of, 174-237;
 scattering of, 174
 radomes, 531-535
 rainfall, attenuation through, 230-231,
 761-762; *see also* precipitation
 random error, 124-125; angle tracking
 loop position and, 489-490
 random motion, 696
 random noise, 245-248, 273, 316; *see*
also noise
 random tracking errors, 480
 random vibration, 699-700
 range aperture, 80
 range gating, 163, 325-327, 332, 499-501

- range measurement, 499; in CW systems, 313-315; in pulsed-doppler systems, 329-331; *see also* radar range
- range noise, 207-208, 395
- range rate, 6-7
- range resolution, 313; ground mapping and, 774-775
- range tracking, automatic, 498-510
- range tracking noise, 397-398
- range tracking system, 502-504
- rate errors, 486-489
- rate gyros, 454-455, 464, 494-495
- rate noise errors, 487-488
- Raven, R. S., 394 n., 607
- raw video, 620
- Rayleigh distribution and criteria, 142, 193-194, 735, 760, 784, 790 n.
- RC compensating networks, 451
- RC filter circuit, 241-242, 252
- RC tracking network, 496-497
- read-out and storage system, 612-613
- receiver noise, 396
- receiver tuning control, 394
- reflection coefficient, 182, 541
- refractive index, modified, 233
- Reggia, F., 566 n.
- regulatory circuits, 394-470; aircraft motions and, 433-439; accuracy requirements and, 405-407; angle track stabilization and, 464-468; automatic gain control loop, 416-428; modulation transmission requirements, 424-425; need for, 394-395; search stabilization, 440-441; space stabilization, 439; stabilization problem and, 429-433; static and dynamic control loop and, 442-448
- Reich, E., 308 n.
- reliability, factors influencing, 34, 716; measurement of, 717-720; mechanical design and, 715-722; prediction of, 720-722; probability of, 103-104
- repair, maintenance and, 712
- resolution, beamwidth and, 89-92; defined, 168; display size and, 627-629; in ground mapping, 774, 784-787; infrared sensing, 793; *see also* angular resolution
- resolution capability, 75-80
- resonance, mechanical, 698
- resonance absorption isolator, 558, 563-564
- resonant frequency, vibration and, 698
- resonant system, magnetron and, 582
- resonator array, 527
- retina, 636
- retrodirectivity, 222
- RF amplification, 355; low-noise figure devices for, 356-357
- RF breakdown, at altitude pressures, 545
- RF carrier frequency, ground clutter and, 270
- RF circuits, antenna and, 301
- RF components, air-weather radar, 766; antennas and, 511-579; microwave ferrite devices and, 557-565
- RF filter, 356
- RF generator, standing waves and, 541
- RF Janus receiver, 742-743
- RF pulse train, 321-322
- RF reference phase, 315
- RF transmission lines, 535-540; *see also* transmission lines
- RF voltage, angle noise and, 203-204; *see also* radio frequency
- Riblet, H. J., 549 n.
- Rice, S. O., 238, 385 n.
- Ridenour, L., 222, 615 n.
- Rieke diagram, 404-405
- rising sun anode, 583, 586
- Rizzi, Peter A., 544 n.
- rms errors, *see* root mean square errors
- roll, pitch and yaw, 433, 797
- root mean square errors, 81, 84, 87; noise and, 286-290, 346
- rotary joints, 546-547
- round-trip attenuation, 761
- Sack, E. A., 669
- Sagar, A., 550 n.
- sampling frequencies, 295-297
- Sando, R. M., 679 n.
- Saxton, J. A., 103 n., 187 n.
- scanning, 4; angular, 319; conical, 405, 414, 515; effect of, 156-162; infrared, 344; mapping and, 9 (*see also* ground mapping); *see also* scanning (adj.)
- scanning frequency, 296-297
- scanning loss, 161
- scanning patterns, types of, 160
- scanning power modulation, 289-290
- scanning rate, infrared, 794
- scanning system, infrared characteristics

- scanning system, (*Cont.*)
 and, 343-344; pulling characteristics of, 405
- scanning time, optimum, 159-160; total, 294
- scan-to-scan correction, 88
- scattering, 174
- scattering coefficients, 747; *see also* back-scattering
- Schade, O. H., 786 n.
- Schetne, H. A., 533 n.
- Schlesinger, S. P., 216 n.
- Schwarz's inequality, 274-275, 282-284
- Schweibert, H., 546 n.
- scientific research, radar systems and, 16
- scintillation, 199; *see also* target noise
- sea clutter, 35, 211, 220, 222; *see also* sea return
- search and detection, 4
- search display, 651-654
- search loop mechanization, 448-452
- search pattern elevation, 114
- search pattern stabilization, 116, 440-441
- search pattern velocity errors, 444-447
- search stabilization system, 430-432
- search volume requirements, 113-115
- sea return, 92, 181-190, 211-219, 725-802
- sea surface, radar area and, 211
- sea water, dielectric properties of, 183
- second-time-around echoes, 221
- seeker frequency, 135
- semiactive system, 13-14, 135, 318, 340; defined, 293
- semiconductor diodes, 512
- self-contained navigation, 727
- sensitivity measurements, 392-393
- sensitivity parameters, infrared sensing system, 792
- sensitivity time control (STC), 790
- servo noise, 396
- servo systems, 502-505
- sferics, 395; *see also* atmosphere
- shadow-mask colot tube, 661-662
- shielding, nuclear radiation and, 693-694
- ship-based radar system, 3
- ship weapons system, 50
- shock, vibration and, 694-704
- shock isolators, 703-704, 708
- short-pulse radar system, 335-336
- shot noise, 246, 353, 360
- sidebands, 17-19
- sidelobes, antenna, 323, 521
- side-look system, radar mapping, 9-10, 773
- Silver, Samuel, 516 n.
- sigma zero, 140
- signal, high and low-frequency, 285-286; minimum discernible, 768; time of arrival of, 281-291
- signal analysis, techniques for, 238-291
- signal correction, 376
- signal display integration, 631-633
- signal fading, 461, 476-477
- signal "folding," 323-324, 327
- signal location error, 282, 284
- signal power, 139, 260
- signal-to-clutter ratio, 271, 325, 333
- signal-to-noise ratio, 260, 265, 269, 274-276, 288, 290, 318, 329, 346; detection and data processing and, 297, 300; in display system design, 621; doppler navigation systems and, 750; loss in, 165; maximum, 276-277; predetection of, 168, 310; in pulsed-doppler systems, 325; receiver design and, 350-351; resolution of, 172; *see also* noise
- Simmons, A. J., 554 n., 555 n., 566 n.
- single-glimpse probability, 144, 149-151, 157, 162-163
- slot hybrid duplexer, 568
- slot hybrid mixers, 552-553
- slot hybrid phase shifter, 556
- slotted waveguide, 524-528
- smoke, 10
- smoothing schemes, 99-100, 329
- Smullin, L. D., 567 n.
- S-N diagram, 701; *see also* signal-to-noise ratio
- sneak circuits, 377-380
- snow, attenuation by, 231
- snubbing, acceleration and, 707-708
- solar radiation, mechanical design and, 692; temperature and, 685
- solid state amplifiers, 574-578
- sound pressure level, 705-706
- spaced-active operation, 318
- space modulation, 17
- space stabilization systems, 395, 439, 452-453; cross coupling from, 494
- sparkling, in magnetrons, 587
- spectral response, 790, 792, 798
- spectrum, in Fourier series, 239
- spot brightness, 630

- spot detection and resolution, in human eye, 636
- spot noise figure, 368
- spurious information, 5, 352
- square-law detector, 143, 145, 203, 254-255, 265, 286-287
- stabilization, 429-433, 741-742, 780-781
- stabilization errors, 82, 458; ground mapping and, 777
- stabilization loop, accuracy requirements of, 457-464; automatic tracking and, 473-474; configurations of, 453-454; mechanization of, 468-470; search volume and, 116; transfer ratio designs of, 466
- stabilotron, 604
- stable local oscillator (STALO), 373
- STAE, *see* second-time-around echoes
- stagnation temperature, 685-686
- standing waves, 540-543
- static electricity, in mechanical design, 710
- stationary processes, 246
- Stauffer, F., 725 n.
- steady-state cooling, 688
- steady-state error, 504; stabilization loop and, 456
- Stecca, A. J., 198 n., 400 n.
- steering error, 125-132
- Steinberg, Bernard D., 308 n.
- step-twist rotary joint, 547
- Stoddard, D. J., 521 n.
- storage radar, 309-311
- storage tubes, 656-658, 672
- straddling factor, 163-166
- Straiton, A. W., 229 n.
- Stratton, J. A., 182 n.
- stresses, cyclic, 701
- strip lines, 540
- strip map, 789
- Strom, L. D., 551 n.
- Strong, J., 790 n.
- Strum, P. D., 361 n.
- Sturley, K. S., 384 n.
- subcarriers, information, 20-21
- suboptimization, of weapons system, 44
- Sugden, T. M., 233
- sulfur hexafluoride gas, 545
- Summers, R. A., 438
- superheterodyne circuit, radar receivers, 347-349
- superhigh frequency mixers, 357-362
- surface-to-air missile system, height-finding and, 95
- Sutherland, J. W., 550 n.
- Swerling, P., 155, 290 n., 308 n.
- switches, ferrite, 561-562
- symbolic model, 37-39
- synchronous detector, 256
- system concept, military objective and, 32, 49; in weapons system study, 57-59; *see also* weapons system
- system configuration, 43
- system effectiveness models, in weapons system, 61-67
- system logic, 43, 67, 70-73
- system planning, reliability and, 716
- system study plan, 47-50, 59-60
- tactical problem, assessment and monitoring in, 609-611; AEW beamwidth and, 89-92; design parameters and, 98-100; environment and, 35; geometry of, 77
- tangential signal measurement, 393
- target, angular bearing of, 7; back-scattering from, 174; detection and evaluation of, 67; instantaneous velocity of, 72; maneuver capability of, 55; number and importance of, 75; radar area of, 192-197; viewing probability of, 104-106
- target aircraft, number and spacing of, 57
- target altitude, 73; vertical beamwidth and 91-92
- target assignment procedure, 63, 66
- target complex, 55-57
- target detection, decision element and, 144
- target echo, 274; *see also* echo
- target fluctuation, effect of, 151-156
- target heading, 71, 85-89; measurement of, 99
- target information, extraction of, 23; tracking circuits and, 477-478
- target modulations, 20-23, 180-181
- target motion, 6-7; effects of, 21-22; signal modulations by, 180-181
- target noise, 199; *see also* noise; signal-to-noise ratio
- target points, vectoring and, 94
- target position, 6, 71; prediction of, 72, 89

- target power ratio, 171-172
 target radar characteristics, 208-210
 target range, 6
 target reflection, phase changes and, 21
 target resolution requirements, AEW, 75-79
 target spectra, pulsed-doppler system, 322
 target tracking, accuracy of, 88; detection of, 79; infrared system and, 344-346; vertical beamwidth and, 91-92; *see also* tracking
 target velocity, 85-89; relative, 7-8
 Teflon, 565
 temperature, mechanical design and, 683-691
 temperature control, 687-688
 Terman, F. E., 182 n.
 terrain avoidance, 764; in display system design, 673-674
 terrain bias error, 746-747
 Theissing, H. H., 229 n.
 thermal noise, 140, 353, 373; folding and, 324
 thin phosphor films, 659-661
 Thompson, B. J., 368
 threat evaluation, height finding and, 93-94
 thunderstorms, 762-763
 time delay errors, 65-66, 80-83
 time-dependent operations, 253-258
 time discriminator, 500
 time domain effects, 377
 time modulation, 500
 time response system, 41
 Titus, J. W., 679 n.
 Tolbert, C. W., 229 n.
 Tompkins, R. D., 550 n.
 total reflection, 541
 Townes, C. H., 574
 tracking, automatic, *see* automatic tracking; defined, 513; by radar system, 6-9; *see also* target tracking
 tracking error, guidance considerations and, 135-137
 tracking noise, 396-397
 tracking phase, in interception, 101
 tracking stabilization, 432-433, 452-453
 tracking system, human operator and, 649-650
 track-while-scan operation, 513
 "trade off," between system parameters, "trade-off," (*Cont.*)
 62, 67
 train angle, *see* look angle
 transconductance, 428
 transducers, display devices as, 608; electrical-to-light, 627-634
 transfer function, 23, 241, 363; AGC loop design and, 423-426; airborne radar range and, 509-510; aircraft, 438; angle tracking and, 483-484; energy and power, 242
 transmission frequency, ground mapping and, 779
 transmission line, anisotropic, 560; antennas and, 535-540; for broadband systems, 544; strip lines as, 540; types of, 536-540
 transmission periodicity, 296
 transmission shaping unit, 501
 transmit-receive tubes, 331
 transmitted-waveform storage, 310
 transmitter power, air-weather radars, 767-768
 transmitter stability, problems of, 331
 TR-ATR duplexing, 753
 traveling wave tube (TWT), 356, 597-599
 trees, ground return from, 221
 trimode turnstile junction, 550
 TR tube, duplexing and, 568
 true air speed system, 727
 Truxal, J. G., 502 n., 504 n.
 tube size, in display system design, 633-634
 tuning characteristics, 350
 tuning control, 394
 tuning errors, 409, 414; accuracy requirements and, 405-407
 turboprop propulsion system, 56
 turnstile waveguide junction, 550
 two-color tubes, in display systems, 665
 Typotron tube, 671
 Tyrell, W. A., 549 n.
- Uhlenbeck, G. F., 147 n., 193 n., 238 n., 361 n., 376 n.
- vacuum tubes, cutoff bias in, 428; life expectancy of, 684
 Van Vleck, J. H., 228-229, 273 n., 646-664
 vectoring display, 651-654

- vectoring errors, 108-111
- vectoring guidance, 65-66, 72-73; aircraft interceptor considerations and, 111-116; analysis of, 106-111; height-finding and, 94-95; in interception, 101
- vectoring probability, 104-106
- vectoring system information, 70-71, 106
- vector transfer function, 23
- velocity discrimination, 10; *see also* doppler effect
- velocity estimates, 85
- velocity modulation, 591
- velocity sensor, 726
- velocity spectrum, 218
- vertical beamwidth, *see* beamwidth, vertical
- vibration, defined, 695; environmental, 701-702; shock and, 694-704
- video, artificial, 621
- video bandwidth, vs. IF bandwidth, 406
- video integrator, delay-line, 309
- video pulses, IF discriminator and, 410
- video signals, in display system design, 619-621; noise and, 260, 263, 265-266, 271, 286-287; *see also* signal-to-noise ratio
- video voltage, representation of, 145
- viewing probability, 104-106; look angle and, 114
- visual acuity, defined, 635; requirements for, 636-637
- visual angle, defined, 635
- voltage envelope, 170-171
- voltage pulse train, 286
- voltage standing wave ratio (VSWR), 541, 546-547
- voltage-to-brightness transfer function, 618
- water, dielectric properties of, 182-183; scattering coefficients for, 747
- water bias error, 748-749
- water drops, attenuation by, 175
- water motion errors, doppler navigation system and, 748-749
- water vapor, attenuation due to, 229; ground return and, 222
- Watkins, P. L., 192 n.
- waveguides, 473; circular and square, 539-540; directional couplers, 548; waveguides, (*Cont.*)
- hybrid junctions, 548-550; pressurized, 545-546; rectangular, 537-539; ridge, 539; rotary joints for, 546-547; slotted, 524
- waves, standing, *see* standing waves
- weapons system analysis, master plan for, 48
- weapons system models, 36-39; construction and manipulation of, 41-44; parameters of, 43, 60-61; statistical character of, 39-41
- weapons system requirements, development of, 46-137
- weapons systems, active and passive types, 2; airborne radar and, 27-35; defined, 41; integration environment of, 36; suboptimization of, 43-44; system concept and, 57-67
- weather radar, 759-760; design features and variables in, 764, 766-772; isoecho contour circuitry in, 769-770; peak transmitter power and, 767-768; radar range equations and, 764-766; *see also* air-weather radar systems
- weeds and grass, reflections from, 221
- Weiderborn, N. M., 533 n.
- Weidner, R. T., 544 n.
- Weiner, N., 273 n., 471 n.
- Weiss, Max T., 558 n.
- Wheeler, H. A., 546 n.
- Wheeler, M. S., 511, 531 n.
- White, C. F., 394 n., 426 n., 471 n., 505 n.
- "white noise," 244
- white random motion spectrum, 697
- wide bandwidth coherent systems, 336-338
- Williams, C. S., Jr., 224 n.
- Wiltse, J. C., 216 n.
- windshield displays, 674-675
- Woodward, P. M., 238 n.
- Wright Air Development Center, Dayton, Ohio, 755
- X band, 779; *see also* transmission frequency
- Yaplee, B. S., 215 n., 216, 222
- Yates, H., 292 n.
- yaw, pitch and roll, 434, 797
- Zener diodes, 497

