AMC PAMPHLET AMCP 706-127

AD 763495

ENGINEERING DESIGN HANDBOOK

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
U S Department of Commerce
Springfield VA 22151

INFRARED MILITARY SYSTEMS

PART ONE

INCLUDES CHANGE 1
DATED 7 MAY 1973

HEADQUARTERS, U.S. ARMY MATERIEL COMMAND APRIL 1971

569

REPRODUCTION QUALITY NOTICE

This document is the best quality available. The copy furnished to DTIC contained pages that may have the following quality problems:

- Pages smaller or larger than normal.
- Pages with background color or light colored printing.
- Pages with small type or poor printing; and or
- Pages with continuous tone material or color photographs.

Due to various output media available these conditions may or may not cause poor legibility in the microfiche or hardcopy output you receive.

☐ If this block is checked, the copy furnished to DTIC contained pages with color printing, that when reproduced in Black and White, may change detail of the original copy.

DEPARTMENT OF THE ARMY
HEADQUARTERS UNITED STATES ARMY MATERIEL COMMAND
5001 Eisenhower Ave, Alexandria, VA 22304

AMC PAMPHLET
No. 706-127
CHANGE 1

7 May 1973

Engineering Design
Handbook

INFRARED MILITARY SYSTEMS

PART ONE

AMCP 706-127, 21 April 1971, is changed as follows:

a. Remove pages and insert new pages as indicated below.

Remove pages--	Insert pages--
	Title
xxvii through xxix	xxvii through xxx

b. On all revised pages, changed portions of the text are indicated by vertical lines in the left margins.

(AMCRD-TV)

FOR THE COMMANDER:

OFFICIAL:

JOSEPH W. PEZDIRTZ
Major General, USA
Chief of Staff

JOHN LYCAS
Colonel, GS
Chief, HQ Admin Mgt Ofc

DISTRIBUTION:
Special

AMCP 706-127

ENGINEERING DESIGN HANDBOOK
ON
INFRARED MILITARY SYSTEMS,
PART ONE

Editor
Khalil Seyrafi
Electro-Optical Research Company

1971

Headquarters, US Army Materiel Command

HEADQUARTERS
UNITED STATES ARMY MATERIEL COMMAND
WASHINGTON, D.C. 20315

AMC PAMPHLET
No. 706-127

21 April 1971

ENGINEERING DESIGN HANDBOOK
INFRARED MILITARY SYSTEMS
PART ONE

TABLE OF CONTENTS

Paragraph		*Page*
	List of Illustrations	xvi
	List of Tables	xxv
	Preface	xxvii
	CHAPTER 1 INTRODUCTION	
1-1	Definition of Infrared Spectrum	1-1
1-2	Milestones in the Development of Infrared Technology	1-1
1-3	Military Application	1-2
1-3.1	IR Imaging Systems	1-3
1-3.2	Infrared Missiles	1-3
1-3.3	Infrared Fire Controls	1-3
1-3.4	Tail-warning Systems	1-4
1-3.5	Space Applications	1-4
1-3.6	Spectroscopy	1-4
1-4	Advantages and Disadvantages of Infrared Systems	1-4
	References	1-5
	CHAPTER 2 INFRARED PHYSICS	
2-1	Basic IR Symbols and Definitions	2-1
2-2	Radiation Laws	2-1
2-2.1	Kirchhoff's Law	2-1
2-2.2	Planck's Law	2-5
2-2.3	Rayleigh-Jeans Law	2-5
2-2.4	Wien's Law	2-8

TABLE OF CONTENTS (Cont)

Paragraph		*Page*
2-2.5	Stefan-Boltzmann Law	2-8
2-2.6	Wien's Displacement Law	2-8
2-3	Radiant Energy Transmission	2-8
2-3.1	Radiant Intensity from a Point Source	2-8
2-3.1.1	Isotropic Point Source	2-8
2-3.1.2	Lambertian Point Source	2-8
2-3.2	Radiant Energy Density	2-9
2-3.2.1	Energy Density for Collimated Irradiance	2-9
2-3.2.2	Energy Density Within an Isothermal Enclosure	2-9
2-3.3	Transfer of Radiant Power	2-9
2-3.3.1	An Object in Space	2-9
2-3.3.2	Transfer Between Two Infinite Planes	2-11
2-3.4	Irradiance As a Function of Range	2-12
2-3.4.1	Point Source	2-12
2-3.4.2	Extended Source	2-12
2-3.5	Calculation Aids	2-13
2-3.5.1	Use of GE Radiation Calculator	2-14
2-3.5.2	Sample Calculations	2-16
2-4	Atmospheric Transmission	2-17
2-4.1	Extinction Coefficient	2-17
2-4.2	Molecular Absorption	2-18
2-4.2.1	Absorption Coefficient	2-18
2-4.2.2	Absorber Constituents	2-20
2-4.2.3	Absorption Models	2-23
2-4.2.3.1	Absorption Due to a Single Line	2-23
2-4.2.3.2	Absorption Due to an Assembly of Independent Lines	2-25
2-4.2.3.3	The Elsasser Band Model	2-25
2-4.2.3.4	The Goody Model	2-29
2-4.2.3.5	Other Less Frequently Used Models	2-30
2-4.2.3.6	Selective Absorption of Plume Radiation	2-30
2-4.2.4	Absorption Calculation	2-30
2-4.2.4.1	Narrow-band Absorption Data	2-31
2-4.2.4.2	Wide-band Absorption Data	2-31
2-4.2.4.3	Atmospheric Windows	2-31
2-4.2.4.4	Equivalent Sea-level Path	2-47
2-4.2.4.4.1	Equivalent-path Absorption Calculations	2-47
2-4.2.4.4.2	Equivalent-path Absorber Contents	2-47
2-4.2.4.5	Examples of Atmospheric Transmission Calculations	2-51
2-4.3	Scattering	2-51
2-4.3.1	Scattering Coefficient	2-52
2-4.3.2	Meteorological Range	2-55
2-4.3.3	Atmospheric Scintillation	2-55
2-5	Transmission of Infrared Radiation Through Optical Materials	2-58
2-5.1	Maxwell's Equations	2-59
2-5.2	Units	2-59

TABLE OF CONTENTS (Cont)

Paragraph		*Page*
2-5.3	Boundary Relations	2-60
2-5.4	Plane Waves	2-61
2-5.5	Poynting Vector	2-61
2-5.6	Refractive Index	2-61
2-5.7	Snell's Law	2-62
2-5.8	Reflection Coefficient	2-62
2-5.9	Brewster's Angle	2-63
2-5.10	Polarization	2-63
2-5.11	Anti-reflection Coatings	2-64
2-5.12	Dielectric Waveguides and Cavities	2-65
2-5.13	Spectral Transmission of Optical Materials	2-65
2-6	Sources of Natural Radiation	2-68
2-6.1	Background Radiation	2-68
2-6.1.1	Terrain	2-71
2-6.1.2	Sea	2-79
2-6.1.3	Clouds	2-81
2-6.1.4	Celestial Background	2-90
2-6.2	Target Radiation	2-106
2-6.2.1	Rockets and Missiles	2-107
2-6.2.2	Plume Radiation	2-107
2-6.2.2.1	Chemical Composition	2-107
2-6.2.2.2	Plume Structure	2-108
2-6.2.2.3	Infrared Emission	2-115
2-6.2.3	Re-entry Vehicles	2-119
2-6.2.4	Aircraft	2-121
2-6.2.5	Clear Air Turbulence	2-122
2-6.2.6	Ground Targets	2-122
	References	2-125

CHAPTER 3
IR SYSTEM COMPONENTS

3-1	Introduction	3-1
3-2	Optics	3-3
3-2.1	Optical Materials	3-3
3-2.1.1	Material Types	3-3
3-2.1.1.1	Refractive Materials	3-3
3-2.1.1.2	Reflective Materials	3-3
3-2.1.2	Material Properties	3-16
3-2.1.3	Anti-reflection Coatings	3-16
3-2.2	Optical Components	3-18
3-2.2.1	Types of Lenses and Principal Characteristics	3-18
3-2.2.2	Types of Mirrors	3-18
3-2.2.3	Filters	3-18
3-2.2.4	Prisms	3-22
3-2.3	Optical Design	3-27
3-2.3.1	Definition, Notations, and Sign Conventions	3-27
3-2.3.1.1	Optical Systems	3-27

TABLE OF CONTENTS (Cont)

Paragraph		*Page*
3-2.3.1.2	Object	3-27
3-2.3.1.3	Image	3-27
3-2.3.1.4	Object and Image Space	3-27
3-2.3.1.5	Point Image	3-27
3-2.3.1.6	Extended Image	3-27
3-2.3.1.7	First-order (Gaussian) Optical Theory—Paraxial Region	3-27
3-2.3.1.8	Ray—Slope of a Ray	3-28
3-2.3.1.9	Sign Conventions	3-28
3-2.3.1.10	Nomenclature	3-29
3-2.3.2	First-order Theory—Formulations	3-29
3-2.3.2.1	Cardinal Points of Optical Elements	3-29
3-2.3.2.2	Multiple Element Systems	3-31
3-2.3.2.3	Image Position, Magnification, and Virtual Image	3-32
3-2.3.2.4	Sine Condition—Lagrange Law of Invariance	3-34
3-2.3.3	Limitation of Rays	3-35
3-2.3.3.1	Entrance Pupil—Exit Pupil	3-35
3-2.3.3.2	Relative Aperture Speed, f/no., and Numerical Aperture	3-35
3-2.3.4	Aberrations in Third-order Optics	3-38
3-2.3.4.1	Third-order Theory	3-38
3-2.3.4.2	Monochromatic Aberrations	3-39
3-2.3.4.3	Chromatic Aberration	3-43
3-2.3.4.4	Correction of Aberrations	3-43
3-2.3.4.5	Relation of Aberrations to Optical System Parameters	3-44
3-2.3.5	Ray Tracing	3-47
3-2.3.5.1	Use of Computers for Ray Tracing	3-47
3-2.3.5.2	Graphical Ray Tracing	3-47
3-2.3.6	Image Quality in Terms of Resolution, Spot Size, and Energy Distribution	3-51
3-2.3.6.1	The Diffraction Effect—Airy Disc	3-51
3-2.3.6.2	Geometrical Effects—Image Blur	3-51
3-2.3.6.3	Overall Effect	3-51
3-2.3.6.4	Resolution	3-51
3-2.3.6.5	Geometrical Energy Distribution—Spot Diagrams	3-52
3-2.3.6.6	Energy Distribution—Spread Function	3-52
3-2.3.6.7	Frequency Response	3-52
3-2.3.6.8	Image Evaluation	3-52
3-2.3.6.9	Depth of Focus	3-52
3-2.4	Optical System Descriptions and Engineering	3-53
3-2.4.1	Afocal Systems	3-53
3-2.4.2	Relay and Field Lenses	3-53
3-2.4.3	Aplanatic Systems	3-53
3-2.4.4	Symmetrical Combinations	3-53
3-2.4.5	Multiple-element Systems	3-57
3-2.4.6	Reflective IR Optical Systems	3-57

TABLE OF CONTENTS (Cont)

Paragraph		*Page*
3-2.4.6.1	Simple Mirror	3-57
3-2.4.6.2	Folded Systems	3-58
3-2.4.6.3	Off-axis Configuration	3-60
3-2.4.6.4	Compound Reflectors	3-60
3-2.4.6.5	Paired Conics	3-60
3-2.4.7	Catadioptric Systems	3-60
3-2.4.7.1	Mangin Mirror	3-61
3-2.4.7.2	Schmidt System	3-61
3-2.4.7.3	Maksutov—Bouwers System	3-61
3-2.4.7.4	Hybrid Configurations	3-61
3-2.4.8	Rapid Estimation of Blur Size	3-61
3-2.4.8.1	Spherical Mirror, On-axis	3-61
3-2.4.8.2	Spherical Mirror, Off-axis	3-62
3-2.4.8.3	Paraboloidal Mirror	3-62
3-2.4.8.4	Off-axis Paraboloidal Mirror	3-62
3-2.4.8.5	Schmidt System	3-62
3-2.4.8.6	Bouwers Concentric Design	3-62
3-2.4.8.7	Single Refractive Element	3-62
3-2.4.9	Mechanical Stability of Large Optical Systems	3-63
3-2.4.10	Thermal Stability	3-63
3-2.4.11	Establishing Optical Tolerances	3-63
3-2.4.11.1	Surface Quality	3-63
3-2.4.11.2	Thickness and Spacing	3-64
3-2.4.11.3	Optomechanical Centering	3-64
3-2.4.11.4	Prism Angles and Dimensions	3-64
3-2.4.11.5	Materials	3-64
3-2.4.11.5.1	Index of Refraction	3-65
3-2.4.11.5.2	Annealing	3-65
3-2.4.11.5.3	Transmission Range	3-65
3-2.4.11.5.4	Imperfections	3-65
3-2.4.11.6	Summation of Tolerances	3-65
3-2.5	Testing of Optical Systems	3-65
3-2.5.1	Material Inspection Tests	3-65
3-2.5.1.1	Index of Refraction	3-65
3-2.5.1.2	Stresses and Strains	3-66
3-2.5.1.3	Imperfections	3-66
3-2.5.2	Calibration	3-66
3-2.5.2.1	Focal Length Calibration	3-66
3-2.5.2.2	Curvature Measurements	3-66
3-2.5.2.3	Reflectance and Transmittance	3-67
3-2.5.3	Image Quality Measurements	3-67
3-2.5.3.1	Knife-edge Autocollimation	3-67
3-2.5.3.2	Knife-edge Testing of Complex System	3-70
3-2.5.3.3	Knife-edge Testing of Hyperbola	3-70
3-2.5.3.4	Knife-edge Testing of Large, Flat Mirrors	3-70
3-2.5.3.5	Ronchi Grating	3-70
3-2.5.4	Resolution Targets	3-73
3-3	Emitters and Illuminators	3-73

TABLE OF CONTENTS (Cont)

Paragraph		*Page*
3-3.1	Sources of Illumination	3-73
3-3.1.1	Flares	3-73
3-3.1.2	Lamps	3-76
3-3.1.2.1	Tungsten Filament	3-76
3-3.1.2.2	Carbon Arc	3-77
3-3.1.2.3	Gaseous Arc	3-77
3-3.1.2.4	Flashlamps	3-83
3-3.1.3	Lasers	3-84
3-3.1.3.1	Introduction	3-84
3-3.1.3.2	Laser Theory	3-84
3-3.1.3.2.1	Pumping	3-86
3-3.1.3.2.2	Optical Cavities	3-87
3-3.1.3.2.3	Condition for Threshold of Oscillation	3-89
3-3.1.3.2.4	Heat Dissipation	3-89
3-3.1.3.2.5	Cavity Q and Width of Resonance	3-89
3-3.1.3.2.6	Q-switching	3-90
3-3.1.3.3	Laser Types	3-90
3-3.1.3.3.1	Solid-state Lasers	3-90
3-3.1.3.3.1.1	Ruby Lasers	3-91
3-3.1.3.3.1.2	Neodymium Ion Lasers	3-91
3-3.1.3.3.1.3	Other Ions and Hosts	3-92
3-3.1.3.3.2	Gas Lasers	3-92
3-3.1.3.3.2.1	Neutral Atom Laser (He-Ne)	3-93
3-3.1.3.3.2.2	Ion Laser (Ionized Argon)	3-93
3-3.1.3.3.2.3	Molecular Laser (Carbon Dioxide)	3-94
3-3.1.3.3.3	Injection Lasers	3-94
3-3.1.3.4	Beam Control Devices	3-95
3-3.1.3.4.1	Modulators	3-95
3-3.1.3.4.1.1	Mechanical Modulator	3-95
3-3.1.3.4.1.2	Frustrated-internal-reflection Modulator	3-95
3-3.1.3.4.1.3	Acoustic Modulator	3-96
3-3.1.3.4.1.4	Electro-optical Modulator	3-97
3-3.1.3.4.1.5	Magneto-optical Modulator	3-97
3-3.1.3.4.1.6	Cavity Length Modulator	3-97
3-3.1.3.4.2	Q-switches	3-97
3-3.1.3.4.2.1	Mechanical Shutter Q-switch	3-97
3-3.1.3.4.2.2	Rotating Reflector Q-switch	3-97
3-3.1.3.4.2.3	Electro-optical and Magneto-optical Shutter Q-switch	3-97
3-3.1.3.4.2.4	Passive Q-switching	3-98
3-3.1.3.4.3	Beam Deflection	3-98
3-3.1.3.5	Nonlinear Optics	3-100
3-3.1.3.5.1	Harmonic Conversion	3-100
3-3.1.3.5.2	Raman Conversion	3-100
3-3.2	Design Considerations	3-100
3-3.2.1	Coherence	3-100
3-3.2.2	Optical Design for Incoherent and Coherent Sources	3-103

TABLE OF CONTENTS (Cont)

Paragraph		*Page*
3-4	Detectors	3-104
3-4.1	Detector Terminology	3-104
3-4.2	Detector Classifications	3-109
3-4.2.1	Quantum Detectors	3-109
3-4.2.1.1	Photoemissive Detectors	3-109
3-4.2.1.2	Photoconductive Detectors	3-109
3-4.2.1.3	Photovoltaic Detectors	3-115
3-4.2.1.4	Photoelectromagnetic Detectors	3-116
3-4.2.2	Thermal Detectors	3-116
3-4.2.2.1	Liquid Thermometer	3-116
3-4.2.2.2	Golay Cell	3-116
3-4.2.2.3	Calorimeter	3-117
3-4.2.2.4	Thermocouple	3-117
3-4.2.2.5	Thermopile	3-117
3-4.2.2.6	Bolometer	3-117
3-4.2.3	Imaging Devices	3-118
3-4.2.3.1	Photographic Film	3-118
3-4.2.3.2	IR Vidicon	3-125
3-4.2.3.3	Image Converter	3-125
3-4.3	Detector Performance and Test Procedures	3-125
3-4.3.1	Detector Figures of Merit	3-125
3-4.3.1.1	Noise Equivalent Power (*NEP*)	3-125
3-4.3.1.2	Detectivity (D)	3-125
3-4.3.1.3	Specific Detectivity (D^*)	3-125
3-4.3.2	Theoretical Background Limited Detectors	3-126
3-4.3.3	Measurement of D^*	3-127
3-4.3.4	Spectral Responsivity	3-129
3-4.3.5	Input-output Relationship	3-133
3-4.3.5.1	Frequency Response	3-133
3-4.3.5.2	Time Response	3-133
3-4.3.6	Detector Noise	3-135
3-4.3.6.1	Johnson Noise (also Nyquist or Thermal)	3-135
3-4.3.6.2	Temperature Noise	3-135
3-4.3.6.3	Generation-recombination Noise	3-135
3-4.3.6.4	Shot Noise	3-135
3-4.3.6.5	Current Noise	3-137
3-4.4	Detector Fabrication	3-137
3-4.4.1	General	3-137
3-4.4.2	Photon Detectors	3-137
3-4.4.3	Thermal Detectors	3-137
3-4.5	Comparative Performance of Infrared Detectors	3-138
3-5	Cooling Systems	3-141
3-5.1	Requirements	3-141
3-5.1.1	Detector Cooling	3-141
3-5.1.2	Cooling of IR Optics	3-141
3-5.2	Types of Cooling Systems	3-141
3-5.2.1	Direct Contact	3-141
3-5.2.1.1	Liquid Coolant Systems	3-141

TABLE OF CONTENTS (Cont)

Paragraph		*Page*
3-5.2.1.2	Solid Coolant Systems	3-146
3-5.2.1.3	Passive Cooling	3-148
3-5.2.2	Joule-Thomson	3-148
3-5.2.2.1	Closed Cycle	3-148
3-5.2.2.2	Open Cycle	3-150
3-5.2.3	Expansion Engine	3-150
3-5.2.3.1	Stirling Cycle	3-150
3-5.2.3.2	Claude Cycle (Reversed Brayton)	3-150
3-5.2.3.3	Solvay Cycle (Gifford-McMahon)	3-150
3-5.2.3.4	Vuilleumier Cycle	3-154
3-5.2.4	Thermoelectric	3-155
3-5.2.5	Other Cooling Systems	3-155
3-5.2.5.1	Pulse Tube	3-155
3-5.2.5.2	Vortex Tube	3-157
3-5.2.5.3	Adsorption	3-157
3-5.2.6	Comparison of Cooling Systems	3-157
3-5.3	Cooling System Selection	3-157
3-5.3.1	Design Criteria	3-160
3-5.3.1.1	Weight	3-160
3-5.3.1.2	Volume	3-160
3-5.3.1.3	Durability	3-160
3-5.3.1.4	Logistics	3-160
3-5.3.2	Reliability	3-160
3-5.3.2.1	Mechanical Design	3-160
3-5.3.2.2	Structural	3-161
3-5.3.2.3	Vibrational	3-161
3-5.4	Insulation	3-161
3-5.4.1	Vacuum Insulation	3-161
3-5.4.2	Radiation Shields	3-161
3-5.4.3	Superinsulation Materials	3-161
3-5.4.4	Support of Insulating Materials	3-161
3-6	Signal Processing	3-162
3-6.1	Introduction	3-162
3-6.2	Spatial Filtering	3-162
3-6.2.1	Optical Transfer Function (OTF)	3-163
3-6.2.1.1	Point Spread Function (PSF)	3-163
3-6.2.1.1.1	Modulation Transfer Function (MTF)	3-163
3-6.2.1.1.2	One-dimensional Image Analysis	3-165
3-6.2.1.2	Line Spread Function (LSF)	3-165
3-6.2.1.3	Resolution Criteria and MTF	3-165
3-6.2.2	Detector Spatial Filtering	3-165
3-6.2.2.1	Analyses	3-165
3-6.2.2.2	Analysis in the Time Domain	3-165
3-6.2.2.3	Analysis in the Frequency Domain	3-166
3-6.2.2.4	Spatial Filtering Neglecting the Time-effect	3-166
3-6.2.2.5	Spatial Filter of a Detector Scanning in One Dimension Only	3-167
3-6.2.2.6	Rectangular Detector of Width a and Length b	3-167

TABLE OF CONTENTS (Cont)

Paragraph		*Page*
3-6.2.2.7	Plus-minus Detectors	3-169
3-6.2.2.8	Circular Detectors	3-169
3-6.3	Electronic Signal Processing	3-170
3-6.3.1	Low-noise Amplification	3-170
3-6.3.1.1	Thermal Noise	3-171
3-6.3.1.2	Shot Noise	3-171
3-6.3.1.3	1/f Noise	3-173
3-6.3.1.4	Performance Characterization	3-173
3-6.3.1.4.1	Classic Approach for Conventional Transistors	3-173
3-6.3.1.4.2	Classic Approach for Field-effect Transistors	3-175
3-6.3.1.4.3	$\bar{e}_n, \bar{i}_n$ Approach for Conventional Transistors	3-176
3-6.3.1.5	IR Application Problems	3-178
3-6.3.1.5.1	Field-effect vs Conventional Transistor Characteristics	3-178
3-6.3.1.5.2	True Noise-figure	3-180
3-6.3.1.6	Cryogenic Amplifiers	3-180
3-6.3.2	Electronic Filtering	3-184
3-6.3.2.1	Application and Response Analysis	3-184
3-6.3.2.2	Detection Filters	3-185
3-6.3.3	Electronic Multiplexers	3-188
3-7	Data Display and Recording	3-189
3-7.1	Introduction	3-189
3-7.2	Terminology	3-189
3-7.3	Types of Displays	3-193
3-7.3.1	Cathode Ray Tube (CRT)	3-193
3-7.3.1.1	Resolution	3-195
3-7.3.1.2	Brightness	3-196
3-7.3.1.3	Contrast	3-200
3-7.3.1.4	Deflections	3-200
3-7.3.1.5	Focus	3-201
3-7.3.1.6	Bandwidth	3-201
3-7.3.2	Storage Type Cathode Ray Tubes	3-201
3-7.3.3	Electrical-readout Storage Tubes	3-203
3-7.3.4	Character Generation Tubes	3-204
3-7.3.4.1	Matricon	3-204
3-7.3.4.2	Monoscope	3-205
3-7.3.4.3	Shaped-beam Tube	3-206
3-7.3.4.4	Stroke Technique	3-207
3-7.3.4.5	Lissajous Techniques	3-207
3-7.3.4.6	Alphanumeric Indicator Tubes	3-208
3-7.3.4.7	Comparison of Character Generating Devices	3-208
3-7.3.5	Image-converter Tubes	3-208
3-7.3.6	Tape Recording	3-212
3-7.3.6.1	Magnetic Tape	3-212
3-7.3.6.2	Thermoplastic Film	3-213
3-7.3.7	Photography	3-213
3-7.3.8	Holography	3-217
3-7.3.9	Special Purpose Displays	3-220

TABLE OF CONTENTS (Cont)

Paragraph		*Page*
3-7.3.9.1	Projection CRT	3-220
3-7.3.9.2	Multi-gun CRT	3-222
3-7.3.9.3	Electroluminescent (EL) Panels	3-224
3-7.3.9.4	Photo-emitter Diodes	3-227
3-7.4	Display Equipment Design Parameters	3-232
3-8	Testing IR and Associated Equipment	3-233
3-8.1	Test Categories	3-235
3-8.2	Laboratory Instruments and Techniques	3-235
3-8.2.1	IR Collimator	3-235
3-8.2.2	Blackbody	3-238
3-8.2.3	Chopper-modulator	3-241
3-8.2.4	Blackbody Usage	3-241
3-8.2.5	Modulation Transfer Function (MTF) Measurements	3-243
3-8.2.5.1	Method 1	3-244
3-8.2.5.2	Method 2	3-246
3-8.2.5.3	Method 3	3-246
3-8.3	On-board Calibration	3-247
3-8.3.1	Radiometric Source	3-247
3-8.3.1.1	Quantitative Calibration	3-247
3-8.3.1.2	Qualitative Calibration	3-247
3-8.3.2	Simulated Electronic Signal	3-247
3-8.3.3	Type of BITS	3-248
3-8.3.3.1	Lamps	3-248
3-8.3.3.2	Emitters	3-248
3-8.3.3.3	Blackbody Sources	3-249
3-8.3.3.4	Modulation of Blackbody Radiation	3-249
3-8.3.4	Calibration of BITS	3-250
3-9	Ancillary IR Components and EMI Rejection Techniques	3-251
3-9.1	Gimbals	3-252
3-9.1.1	Coordinate System	3-254
3-9.1.2	Servo Loop	3-254
3-9.1.3	Specifications	3-256
3-9.1.3.1	Steady-state Errors	3-256
3-9.1.3.2	Transient Response	3-259
3-9.1.4	Design Considerations	3-260
3-9.1.4.1	Example Design Problem	3-261
3-9.1.4.1.1	Moment of Inertia	3-261
3-9.1.4.1.2	Acceleration	3-261
3-9.1.4.1.3	Component Selection	3-261
3-9.1.4.1.4	Analysis	3-264
3-9.1.4.1.4.1	Rate Loop	3-264
3-9.1.4.1.4.2	Position Loop	3-267
3-9.1.4.1.4.3	Response to Torque Disturbance	3-269
3-9.1.5	Components	3-273

TABLE OF CONTENTS (Cont)

Paragraph		*Page*
3-9.1.5.1	Bearings	3-273
3-9.1.5.2	Motors	3-273
3-9.1.5.3	Gears	3-274
3-9.1.5.4	Amplifiers	3-274
3-9.1.6	Gimbal Associated Measurements	3-274
3-9.1.6.1	Angle Measurement	3-274
3-9.1.6.2	Rate Measurement	3-275
3-9.2	IR System Monitors	3-275
3-9.2.1	Requirements for Monitor Devices	3-275
3-9.2.2	Thermocouples	3-275
3-9.2.3	Thermistors and Resistance Thermometers	3-277
3-9.2.4	Voltage and Current Monitors	3-277
3-9.2.5	Calibration of IR Sensors	3-277
3-9.3	Sun Shutters	3-278
3-9.4	Electromagnetic Interference	3-287
3-9.4.1	Sources of Electromagnetic Radiation	3-287
3-9.4.2	Elimination and Rejection	3-287
3-9.4.2.1	Methods	3-288
3-9.4.2.1.1	Bonding	3-288
3-9.4.2.1.2	Grounding	3-288
3-9.4.2.1.3	Shielding	3-289
	References	3-290

CHAPTER 4
IR SYSTEMS DESCRIPTION

4-1	Passive Systems	4-7
4-1.1	Imaging Systems	4-7
4-1.1.1	Mechanical Scanners	4-9
4-1.1.1.1	Scanner Parameters	4-9
4-1.1.1.1.1	Noise Equivalent Temperature (Sensitivity)	4-9
4-1.1.1.1.2	Optical Gain and Resolution	4-13
4-1.1.1.1.3	Scan Rate and Bandwidth	4-13
4-1.1.1.1.4	Performance Requirements and Instrumentation Parameters	4-14
4-1.1.1.2	Thermographs	4-16
4-1.1.1.3	Down-looking Line Scanners	4-16
4-1.1.1.4	Forward-looking Infrared Systems	4-18
4-1.1.2	Image Tubes	4-18
4-1.2	Search Equipment	4-19
4-1.2.1	Spectral Filtering	4-21
4-1.2.2	Spatial Filtering	4-21
4-1.2.3	Temporal Filtering	4-21
4-1.3	Trackers	4-22
4-1.3.1	The IR Scanner As an Optical Angle Transducer	4-24
4-1.3.1.1	Small Instantaneous Field-of-view Scanning	4-24

TABLE OF CONTENTS (Cont)

Paragraph		*Page*
4-1.3.1.2	Large Instantaneous Field-of-view Reticle Scanning	4-26
4-1.3.2	Tracking Scanner Sensitivity	4-27
4-1.3.3	Tracking Loop Servo Parameters	4-28
4-1.4	Radiometers	4-29
4-1.5	Spectrometers	4-30
4-1.6	Interferometers	4-31
4-1.7	Hybrid Systems	4-31
4-1.7.1	Track-while-scan	4-31
4-1.7.2	Scanning Radiometer/Spectrometers	4-31
4-2	Active Systems	4-32
4-2.1	Illuminators	4-32
4-2.1.1	Searchlights	4-32
4-2.1.2	Lasers	4-32
4-2.2	Rangefinders	4-33
4-2.3	Communications and Data Transmission	4-35
	References	4-37

CHAPTER 5
IR SYSTEM DESIGN

5-1	System Approach to Infrared Design	5-1
5-2	Requirements	5-1
5-2.1	System Analysis	5-2
5-2.1.1	Operational Environment	5-2
5-2.1.2	Functions	5-8
5-2.1.3	Requirement Analysis	5-8
5-2.1.4	Block Diagrams	5-12
5-2.2	Target Definition	5-13
5-2.2.1	Spectral Radiant Intensity Bounds	5-13
5-2.2.2	Radiance Gradients	5-14
5-2.3	Background Definition	5-14
5-2.3.1	Terrain, Sea	5-14
5-2.3.2	Clouds	5-14
5-2.3.3	Stellar Backgrounds	5-15
5-2.4	Transmission	5-15
5-2.4.1	Absorption	5-15
5-2.4.2	Obscuration	5-16
5-2.5	Countermeasures	5-17
5-2.5.1	Passive Countermeasures	5-17
5-2.5.2	Active Countermeasures	5-17
5-3	Baseline Design Concept	5-18
5-3.1	Spectral Optimization	5-18
5-3.1.1	Trade-offs	5-18
5-3.1.2	Detector Selection	5-22
5-3.2	Frame-time Optimization	5-24

TABLE OF CONTENTS (Cont)

Paragraph		*Page*
5-3.3	Parametric Design	5-25
5-3.3.1	Collecting Aperture Diameter	5-25
5-3.3.2	Number of Detectors	5-27
5-3.3.3	Cooling Requirements	5-30
5-3.3.4	Scanning Systems	5-31
5-4	Sample IR System Designs	5-32
5-4.1	Missile Seeker	5-32
5-4.1.1	Missile Seeker Requirements	5-33
5-4.1.1.1	Operational Environment	5-33
5-4.1.1.2	Seeker Functions	5-33
5-4.1.1.3	Functional Description	5-35
5-4.1.1.4	Apparent Radiant Intensity of Targets and Backgrounds	5-35
5-4.1.2	Baseline Design Concept	5-35
5-4.1.3	Seeker Design	5-39
5-4.1.3.1	Optical and Mechanical	5-39
5-4.1.3.2	Detector	5-40
5-4.1.3.3	Acquisition-mode Programmer	5-40
5-4.1.3.4	Track-loop-mode Servo	5-40
5-4.1.3.5	Sensitivity	5-41
5-4.1.3.6	Signal Processing	5-42
5-4.2	Active Imaging Systems	5-42
5-4.2.1	Sample Design of Active IR Imaging Systems	5-42
5-4.2.1.1	Spatial Resolution and Field of View	5-42
5-4.2.1.2	Data Rate and System Bandwidth	5-44
5-4.2.1.3	Laser Power and Signal-to-noise Characteristics	5-45
5-5	Trade-off Analyses	5-46
5-5.1	General Principles	5-46
5-5.2	Use of Advanced Components and Concepts	5-47
5-5.3	Cost, Space, Weight, and Power Trade-offs	5-47
5-5.4	Human Factors	5-48
5-5.4.1	Visible Spectrum	5-48
5-5.4.2	Visual Acuity	5-48
5-5.4.3	Other Factors Affecting Acuity	5-49
5-5.4.4.	Flicker	5-49
5-5.4.5	Supplementary Auditory Displays	5-50
5-5.4.6	Human Reaction Time	5-50
5-5.4.7	Human Engineering	5-52
5-5.5	Reliability	5-53
5-5.5.1	The Importance of Reliability	5-53
5-5.5.2	Reliability Program	5-53
5-5.5.3	Definitions	5-53
5-5.5.4	Reliability Prediction	5-53
5-5.5.5	What the IR Design Engineer Should Know About Reliability	5-54

TABLE OF CONTENTS (Cont)

Paragraph		*Page*
5-5.5.6	Design Practices for Achieving High Reliability	5-54
5-5.5.6.1	Borrowing and Buying	5-54
5-5.5.6.2	Simplification	5-54
5-5.5.6.3	Selection of Reliable Part Types	5-54
5-5.5.6.4	Derating	5-55
5-5.5.6.5	Redundancy	5-55
5-5.5.6.6	Providing for Satisfactory Heat-transfer Properties	5-55
5-5.5.6.7	Mechanical Strength and Stability	5-56
5-5.5.6.8	Formulating Alternative Designs	5-56
5-5.5.6.9	Miscellaneous Considerations	5-56
5-5.6	Maintainability	5-57
5-5.6.1	The Importance of Maintainability	5-57
5-5.6.2	Maintainability Program	5-57
5-5.6.3	Definitions	5-58
5-5.6.4	Maintainability Predictions	5-58
5-5.6.5	What the IR Design Engineer Should Know About Maintainability	5-58
5-5.7	Producibility	5-59
5-5.7.1	Definition and Importance	5-59
5-5.7.2	Producibility Program	5-60
5-5.7.3	What the IR Design Engineer Should Know About Producibility	5-60
5-5.8	Mobility	5-61
5-5.8.1	Definitions and Importance	5-61
5-5.8.2	Transportability Design Program	5-61
5-5.8.3	What the IR Design Engineer Should Know About Mobility	5-61
5-5.9	Integrated Logistic Support	5-62
5-5.10	Ground Support Equipment	5-63
5-5.10.1	The Importance of Ground Support Equipment	5-63
5-5.10.2	Ground Support Equipment Functions	5-63
5-5.10.3	Ground Support Equipment Design Factors	5-63
5-5.10.4	Selection of Aerospace Ground Equipment	5-63
5-5.10.5	Applicability of Government Equipment	5-64
5-5.10.6	Multi-function Concept	5-64
5-5.10.6.1	General	5-64
5-5.10.6.2	Advantages	5-64
5-5.10.6.3	Disadvantages	5-64
5-5.10.6.4	Unitized Multi-purpose Equipment	5-64
5-5.10.7	Quality	5-65
5-5.10.7.1	Criteria	5-65
5-5.10.7.2	Evaluation Factors	5-65
5-5.10.8	Commercial Counterparts	5-65
5-5.10.9	System Safety	5-65

TABLE OF CONTENTS (Cont)

Paragraph		*Page*
5-5.10.10	General Military Documents	5-66
5-5.10.11	What the IR Engineer Should Know About Ground Support Equipment	5-66
5-6	Test Requirements	5-67
5-6.1	Quality Control and Inspection	5-67
5-6.2	Inspection and Testing	5-67
5-6.2.1	Test and Inspection Planning	5-68
5-6.2.2	Inspection and Testing of Infrared Components and Systems	5-68
5-6.2.2.1	Optical Elements, Lenses, Prisms, and Mirrors	5-68
5-6.2.2.2	Optical Materials	5-68
5-6.2.2.3	Optical Coatings and Filters	5-69
	References	5-70
	Appendix	5-72
	Bibliography	B-1

LIST OF ILLUSTRATIONS

Figure No.	*Title*	*Page*
2-1	Angle θ in Radians	2-4
2-2	Solid Angle Ω in Steradians	2-4
2-3	Blackbody Curves	2-7
2-4	Geometry of a Lambertian Source	2-10
2-5	The GE Radiation Calculator	2-15
2-6	The Near-Infrared Spectra of Solar Irradiation and of CO, CH_4, N_2O, O_3, CO-, and H_2O	2-19
2-7	Seasonal Variation of the Vertical Distribution of Ozone	2-20
2-8	Annual Variation of Total Ozone for Each 10° of N. Latitude	2-21
2-9	Water Vapor Profile in the Atmosphere	2-22
2-10	Lorentz Line Shape	2-24
2-11	The 6.3-micron Band of H_2O	2-26
2-12	Experimental Fit of Burch and Williams Data to Error Function Absorption at 2.7 Microns	2-45
2-13	Variation of Continuum Absorption Coefficient vs Wavelength	2-46
2-14	Geometrical Relation Between Observer A and Object B	2-48
2-15	IR Absorber Content Above a Given Altitude	2-49
2-16	Water Vapor Content Above a Given Altitude	2-50
2-17	Normalized Particle Density As a Function of Altitude	2-53
2-18	Scintillation vs Telescope Aperture	2-56
2-19	Frequency of Intensity Scintillation	2-57
2-20	Dependence of the Diameter d_o of the Collecting Optics, Upon Wavelength λ, and Altitude at a Zenith Angle of 70° for Both Nighttime and Daytime Conditions	2-57
2-21	The Electromagnetic Spectrum	2-58
2-22	Incident, Refracted, and Reflected Light Beams at the Interface of Two Media (drawn in the plane of incidence)	2-64
2-23	Transmission Curves for Representative IR Optical Materials	2-66
2-24	Comparisons of Material Properties	2-67
2-25	Idealized Background Spectrum	2-69
2-26	Comparison of Calculated Earth Radiance with Measurements for Downlooking System at 37 km Altitude	2-70
2-27	Radiance of 50° and 15°C Blackbodies	2-72
2-28	Hourly Variation in the 10-micron Radiance of Various Backgrounds	2-73
2-29	Daytime Spectral Radiance of Miscellaneous Targets	2-74
2-30	Relative Contrast of Deciduous Trees to Short Grass	2-77

LIST OF ILLUSTRATIONS (Cont)

Figure No	Title	Page
2-31	Relative Contrast of Wooden Catwalk Atop a Dam to the River Upstream from the Dam	2-78
2-32	Relative Contrast of a Small Asphalt-Surfaced Concrete Bridge to an Asphalt Roadway	2-79
2-33	Reflectivity of Solar Radiation vs Sun Angle for Various Sea-Surface Roughnesses	2-80
2-34	Reflectance of a Water Surface at 0°, 60°, and 80° Angle of Incidence	2-81
2-35	Typical 1.2- to 2.5-micron Cirrus Cloud Spectrum	2-84
2-36	Average Spectrum of Gladys I	2-85
2-37	Cloud Spectrum	2-86
2-38	Probability of Clear Lines-of-sight over Northern Hemisphere for All Seasons Combined	2-87
2-39	Probability of Clear Lines-of-sight Between Aircraft at ~30,000 ft and the Horizon	2-88
2-40	Probability of Clear Lines-of-sight Between Aircraft at ~30,000 ft and Space at an Angle of 30° Above the Horizon	2-89
2-41	Spectral Irradiance of the Brightest Visual and Red Stars As a Function of Wavelength	2-90
2-42	Major Planet Solar Reflection and Emission	2-91
2-43	Normalized Spectral Irradiance of R. Monocerotis As a Function of Wavelength	2-93
2-44	Normalized Spectral Irradiance of the Orion Infrared Star As a Function of Wavelength	2-94
2-45	Spectral Irradiance of the Cygnus Infrared Star As a Function of Wavelength	2-94
2-46	Stellar Irradiance As a Function of Visual Magnitude	2-95
2-47	Visual Fraction of Total Blackbody Irradiance As a Function of Temperature	2-96
2-48	Relative Spectral Distribution of Stellar Radiation As a Function of Star Class	2-98
2-49	Detector Response Fraction $F_e(T)$ for Ge:Hg Detectors As a Function of Spectral Band and Target Temperature	2-99
2-50	Number of K and M Stars in the 8- to 14-micron Band Exceeding a Given Irradiance	2-101
2-51	Lunar Temperature Measured by Surveyor I	2-102
2-52	Solar Spectral Irradiance As a Function of Wavelength	2-103
2-53	Infrared Map of the Galactic Center	2-104
2-54	Flow Chart for Plume Radiation Calculations	2-106
2-55	Low Altitude Plume Structure	2-109
2-56	Plume Structure As a Function of Altitude	2-110
2-57	Typical Line-of-sight Variations on the Saturn S-II Stage	2-111

LIST OF ILLUSTRATIONS (Cont)

Figure No.	*Title*	*Page*
2-58	Limiting Particle Streamlines in the Near Field	2-112
2-59	Trajectories of 0.79-micron Radius Particle External to the Gas Plume As a Function of Altitude	2-112
2-60	Near Field Isotherms for 0.79-micron Radius Particle	2-113
2-61	Far Field Temperature Profile, Vacuum Profile	2-114
2-62	Water Vapor and Carbon Dioxide Regions	2-116
2-63	Emission Spectrum of HF for 0.5 cm-atm at 2640°K	2-117
2-64	Theoretical Radiance Spectrum for a Solid Metallized Propellant Rocket Plume	2-117
2-65	Spectral Radiance of RP1/O_2 Rocket Exhaust	2-118
2-66	Spectral Radiance of Solid Propellant Rocket	2-119
2-67	Flow Field Around Blunt and Slender Vehicles	2-120
2-68	Typical Radiant Signature of a Jet Aircraft	2-121
2-69	Idealized Reflectance Curve of a Uniform Fabric to Afford Camouflage Protection Over a Spectral Range from 0.4 to 1.2 Microns	2-123
2-70	Reflectance from Typical Uniform Cloth	2-123
2-71	Reflectance from Human Skins	2-124
3-1	Functional Relationship of the Major Disciplines and Associated Design Specialties	3-2
3-2	Refractive Index vs Wavelength for Several Optical Materials	3-12
3-3	Dispersion vs Wavelength for Several Optical Materials	3-13
3-4	Effects of Anti-reflection Coatings on Ge Transmission	3-17
3-5	Basic Lens Forms	3-19
3-6	Basic Mirror Forms	3-20
3-7	Typical Spectrophotometric Filter Curves	3-21
3-8	Prism Configurations and Functions	3-23
3-9	Conventional Geometry in First-order Optics	3-28
3-10	Cardinal Points for Single-surface Element	3-30
3-11	Cardinal Points for Double-surface Thick Lens Element	3-31
3-12	System Cardinal Points by Ray Tracing	3-32
3-13	Meridian Plane Section of a Refractive Optical System	3-33
3-14	Meridian Plane Section of a Reflective System	3-33
3-15	Virtual Image Formation by an Optical System	3-34
3-16	Application of Sine Condition	3-35
3-17	IR Optical System Showing Arrangement of Stops and Baffles	3-36
3-18	Display Pickup System	3-36
3-19	Relationship of Clear Aperture and Effective Focal Length	3-37

LIST OF ILLUSTRATIONS (Cont)

Figure No.	Title	Page
3-20	Geometrical Presentation of Refracted and Reflected Rays	3-38
3-21	Optical Diagram Illustrating Spherical Aberration	3-39
3-22	Optical Diagram Illustrating Coma Aberration	3-40
3-23	Optical Diagram Illustrating Astigmatism	3-41
3-24	Field Curvature	3-42
3-25	Examples of Pincushion and Barrel Distortion	3-43
3-26	Lens Bending for Minimum Spherical Aberration	3-44
3-27	Elementary Ray Trace Required for Computing Seidel Aberrations	3-45
3-28	Graphical Ray Trace	3-48
3-29	Graphical Ray Tracing. Surface May Be Convex, Flat, or Concave	3-49
3-30	Graphical Ray Tracing. Surface May Be Concave, Flat, or Convex	3-50
3-31	Afocal Systems	3-54
3-32	Optical Relay System	3-55
3-33	Example of Simple IR Radiometer	3-55
3-34	Geometry of an Aplanatic System	3-56
3-35	Symmetrical Relay System	3-57
3-36	Reduction of Spherical Aberration in Series Lens Combinations	3-58
3-37	Reflective Systems	3-59
3-38	Catadioptric Systems	3-60
3-39	Setup for Determining the Curvature of Concave Spherical Mirrors	3-67
3-40	Knife-edge Test Setup for Spherical and Paraboloidal Mirrors	3-68
3-41	Knife-edge Shadow Patterns	3-69
3-42	Knife-edge Test Setup for Complex and Hyperboloidal Systems	3-71
3-43	Ritchie Setup for Testing a Large Flat Mirror Using a Highly Accurate Spherical Mirror as Reference	3-72
3-44	Ronchi Grating for Determining Duration of Rays from Perfect Focus	3-72
3-45	Resolution Chart	3-74
3-46	Spectral Distribution of Xenon Discharge With Resolution of 0.1 Micron for Selected Input Powers	3-79
3-47	Spectral Distribution of 10-kilowatt Xenon Discharge With Resolution of 0.01 Micron	3-80
3-48	Spectral Radiant Power of Experimental dc Lamp (1.7-atm Xe, amp cm^{-2}) Normalized to Input Power. Spectral Resolution Equal to 0.01 Micron	3-81
3-49	Diagram of the Experimental 20-kw Lamp	3-82
3-50	Spectral Radiant Energy of FZ-47A Flashtube Normalized to Input Power. Spectral Resolution Equal to 0.01 Micron	3-83

LIST OF ILLUSTRATIONS (Cont)

Figure No.	*Title*	*Page*
3-51	Simplified Laser Energy-level Diagram	3-87
3-52	Outline of Solid-state Laser	3-91
3-53	Injection Laser	3-95
3-54	Frustrated-internal-reflection Modulator	3-96
3-55	Electro-optical Modulator	3-96
3-56	Digital Deflector	3-99
3-57	Schematic Diagram Showing Basic Types of Detectors	3-110
3-58	Electron Excited Into Vacuum by IR Radiation Impinging on Photocathode	3-111
3-59	Semiconductor Energy Levels at Absolute Zero	3-112
3-60	Semiconductor With Thermally Excited Electrons	3-112
3-61	Electron Excited by a Photon	3-113
3-62	Semiconductor (Intrinsic Detector) With Electric Field $\vec{E}$	3-113
3-63	Semiconductor (N-type Extrinsic Detector) With Donors	3-114
3-64	Semiconductor (P-type Extrinsic Detector) With Acceptors	3-114
3-65	P-n Junction (Photovoltaic Detector)	3-115
3-66	Photo-induced Hall Effect (Photoelectromagnetic Detector)	3-116
3-67	Spectral Range and Sensitivity of IR Photographic Material	3-118
3-68	Density vs Exposure for Various IR Film Sheets	3-122
3-69	Density vs Exposure for Various Types of 35-mm IR Film	3-123
3-70	Density vs Exposure for Various Types of High-speed IR Film	3-124
3-71	Block Diagram of D^* Measurement Apparatus	3-127
3-72	Modulation Factor for Sine and Square Wave Chopping	3-129
3-73	Idealized Spectral Response of Photon and Thermal Detectors	3-130
3-74	Detectivity of Long-wavelength Detectors	3-131
3-75	Detectivity of Short-wavelength Detectors	3-132
3-76	Equivalent Circuit Illustrating Detector-associated Capacitance	3-134
3-77	Typical Noise Voltage Spectrum and Signal Response of a Photoconductor	3-136
3-78	Detector Performance vs Temperature	3-142
3-79	Liquid Cryogen Cooling System	3-143
3-80	Vapor Pressure of Liquid Cryogens	3-145
3-81	Solid Cryogen Cooling System	3-146
3-82	Vapor Pressure of Solid Cryogens	3-147
3-83	Closed Cycle Joule-Thomson System	3-149
3-84	Joule-Thomson Closed Cycle System. Temperature-entropy Diagram	5-150
3-85	Closed Cycle Stirling System	3-151

Figure No.	*Title*	*Page*
3-86	Refrigeration Efficiency vs Temperature	3-152
3-87	Closed Cycle Claude (Reversed Brayton) System	3-153
3-88	Closed Cycle Solvay (Gifford-McMahon) System	3-154
3-89	Vuilleumier Cycle Cooling System	3-155
3-90	Point Spread Function and Modulation Transfer Function of a Rectangular Slit	3-164
3-91	Spatial Filter Amplitude Response of Rectangular Detector	3-167
3-92	Spatial Frequency Amplitude Response (P) of a Plus-minus Detector Spaced One Detector Width Apart ($c = 2a$)	3-168
3-93	Spatial Amplitude Response of Circular Detector	3-170
3-94	Characterization of Thermal Noise from (A) Voltage Source, and (B) Current Source	3-171
3-95	Characterization of Shot-noise from (A) Current Source, and (B) Voltage Source	3-172
3-96	Classic Conventional Transistor Noise-equivalent Circuit	3-173
3-97	Low-frequency (1/f) Noise Characteristics	3-175
3-98	FET Noise Equivalent Circuit	3-176
3-99	$\bar{e}_n, \bar{i}_n$ Noise Equivalent Circuit	3-177
3-100	$\bar{e}_n, \bar{i}_n$ Approach Noise Figure Computational Model	3-178
3-101	Noise Figure As a Function of Frequency from (A) Transistor, and (B) FET	3-179
3-102	Gross Comparison of Transistor and FET Noise Figures vs Source Resistance	3-179
3-103	Transconductance and Drain Current As a Function of Temperature	3-181
3-104	Junction FET Noise As a Function of Temperature	3-182
3-105	FET Noise As a Function of Frequency	3-183
3-106	Model of Filter Relating Output Response $e_o(t)$ to Input $e_i(t)$ and output Noise n_o to Input Noise n_i	3-185
3-107	Rectangular Pulse Input and Simple RC Filter	3-186
3-108	Peak S/N for Various Filters, Compared With Matched Filter (Rectangular Pulse Input)	3-186
3-109	Input Sinusoidal Pulse Train and Second-order Filter	3-187
3-110	Signal-to-noise Performance for a Single Tuned Circuit of Quality Q-filtering a Train of n-sinusoidal Pulses	3-187
3-111	Detector Output for One Frame	3-189
3-112	Standard Luminosity Curve for the Average Human Eye	3-193
3-113	Two Basic Types of Cathode Ray Tubes	3-194

LIST OF ILLUSTRATIONS (Cont)

Figure No.	*Title*	*Page*
3-114	Change in Spot Size With Vg_2, Spot Size vs Vg_1; Anode Voltage V_B = 7000 V	3-195
3-115	Spot Size for Several Beam Currents	3-196
3-116	Variation of Luminous Intensity With Anode Potential	3-197
3-117	Variation of Luminous Intensity With Electron Beam Current Density	3-198
3-118	Efficiency of (A) Aluminized, and (B) Unaluminized Screens	3-198
3-119	Display Storage Tube	3-202
3-120	Typical Scan-conversion Tube	3-204
3-121	Matricon Tube	3-205
3-122	Monoscope Tube	3-206
3-123	Shaped-beam Tube	3-207
3-124	Typical Image-converter Tube in Use	3-210
3-125	Screen Luminous Efficiency	3-210
3-126	Photocathode Sensitivities	3-211
3-127	Image Quality of Image Converter Tube in Terms of Distortion	3-212
3-128	Spectral Sensitivity of Various Types of Film	3-214
3-129	Typical Characteristic Curve of a Photographic Emulsion	3-215
3-130	Typical Variation of Gamma With Development Time	3-216
3-131	Typical Arrangement for Making Holograms	3-219
3-132	Reconstruction of a Hologram	3-220
3-133	Projection CRT Characteristics	3-221
3-134	Exploded View of All-glass EL Readout Panel and Sectional View of Metal-glass Unit	3-223
3-135	Example of Electroluminescent (EL) Panel	3-224
3-136	Brightness vs Voltage With 60% Transmission Glass for Both Metal-glass and All-glass EL Readout Panels at 115 and 250 V rms, Respectively	3-225
3-137	Voltage-current Curve for Nonlinear Resistor	3-226
3-138	CdSe Hysteresis Curve	3-227
3-139	Illuminating Panel Consisting of GaAs Photo-emitter Diode Array	3-228
3-140	Spectral Line Width for GaAsP Emitter Diode	3-229
3-141	Characteristic Curve for Typical GaAsP Emitter Diode	3-230
3-142	Forward Current vs Forward Bias Voltage of a Typical GaAsP Emitter Diode	3-231
3-143	Spatial Distribution of Radiation Emitted by GaAsP Diode	3-232
3-144	Beam-deviating Mirror Schematic Diagram	3-237
3-145	Lambertian Radiation	3-239
3-146	Density of Saturated Water Vapor (100% Relative Humidity) at 1 atm, 760 mm Hg	3-240

LIST OF ILLUSTRATIONS (Cont)

Figure No.	*Title*	*Page*
3-147	Section of Knife-edge Recording Characteristic Curve	3-244
3-148	Typical Knife-edge Record for Different Focus Positions	3-245
3-149	Conceptual Layout of Two-axis Gimbal Assembly	3-251
3-150	Conceptual Layout of Three-axis Gimbal Assembly	3-252
3-151	Relationship of Gimbal and Target Coordinates	3-253
3-152	Servo Loop for Search or Measurement System	3-254
3-153	Servo Loop for Rate Search	3-255
3-154	Servo Loop for Tracker	3-255
3-155	Servo Loop Diagram	3-256
3-156	Error Constants and Responses of Type 0, I, and II Servos	3-257
3-157	Servo Response of $\frac{\omega_n^2}{s^2 + 2\xi\omega_n s + \omega_n^2}$ to Unit Step Input for Various Damping Factors	3-258
3-158	Overshoot vs Damping Factor for Second-order System	3-260
3-159	Block Diagram of dc Torque Motor	3-262
3-160	Equivalent Circuit of Tachometer Generator	3-263
3-161	Block Diagram of Tachometer Generator	3-263
3-162	Block Diagram of System in Example Problem	3-264
3-163	Root Locus for Inner (Rate) Loop	3-266
3-164	Output Rate and Torque Response for Step-function Input to Rate Loop	3-268
3-165	Root Locus for Outer Loop	3-270
3-166	Output Angle and Torque Response for Step-function Input to Position Loop	3-272
3-167	Thermocouple Circuit Schematic	3-276
3-168	Resistance Thermometer Circuit	3-276
3-169	Block Diagram of Voltage and Current Monitors	3-278
3-170	Diagrams of Typical IR Calibration Systems	3-279
3-171	Mechanical Sun Shutter	3-280
3-172	Electrical Sun Shutter	3-282
3-173	Window-shade Type Protection for Focal Plane Shutter Camera	3-283
3-174	Hood-type Sun Shutter	3-284
3-175	Four-leaf Clover Sun Shutter	3-285
3-176	Rotary Damper Sun-shutter Mechanism	3-285
3-177	Classic Schmidt Sun-shutter System	3-286
3-178	Blocking Mechanism in Schmidt System	3-286
3-179	Schematic Illustrating Electromagnetic Coupling Parameters	3-288
4-1	Typical Imaging System Block Diagram	4-8
4-2	End of Scan Distortion	4-17

LIST OF ILLUSTRATIONS (Cont)

Figure No.	*Title*	*Page*
4-3	Typical Search System Block Diagram	4-20
4-4	Typical Tracker Block Diagram	4-23
4-5	Reticles	4-25
4-6	LWIR Radiometer	4-30
4-7	Typical Laser Rangefinder Block Diagram	4-34
5-1	Infrared Missile Seeker Block Diagram	5-12
5-2	Spectral Characteristics of Aircraft vs Viewing Aspect	5-13
5-3	Typical Sunlit Cloud Radiances vs Wavelength	5-15
5-4	Relative Target Energy and Transmission vs Wavelength for Plume Emission	5-16
5-5	Estimate of Probability of Cloud-free Line-of-sight from (or to) Surface Level (Washington, D. C.—Summer)	5-17
5-6	Example Spectral Background and Target Radiant Intensity	5-19
5-7	Sample Spectral Signature of Target and Background	5-20
5-8	Spectral Transmission Curves	5-20
5-9	Apparent Target and Background Signatures	5-21
5-10	Target and Background Cumulative Irradiance vs Upper Wavelength Cutoff	5-21
5-11	Signal-to-background Plus Noise Ratio for Various Upper and Lower Wavelength Cutoffs	5-22
5-12	Cumulative Signal-to-noise Ratios for Two Detector Types	5-23
5-13	Optical Diagram	5-26
5-14	Constraints on System Aperture Selection	5-27
5-15	Sensitivity of Detectors and Number Required vs Detector Aspect Ratio	5-29
5-16	Detector Overlap	5-30
5-17	Basic Scan Modes	5-31
5-18	Line Scanning Technique	5-32
5-19	Scanning Image System	5-32
5-20	Typical Missile Guidance Geometries	5-34
5-21	Seeker Servo Block Diagram	5-36
5-22	Apparent Target Intensity	5-37
5-23	Tracker Signal Processing	5-43
5-24	System Geometry	5-44
5-25	System Block Diagram	5-47
5-26	Relation Among Brightness Contrast, Brightness of Background, Exposure Time, and Visual Acuity	5-50
5-27	Critical Flicker Frequency As a Function of Brightness	5-51
5-28	Performance on a Visual, Auditory, and Combined Visual-auditory Vigilance Task	5-52

LIST OF TABLES

Table No.	Title	Page
2-1	Standard Symbols	2-2
2-2	Basic Radiometric Terms	2-3
2-3	Physical Constants	2-5
2-4	Experimental Values of Emissivity	2-6
2-5	Absorber Concentrations in the Atmosphere	2-23
2-6	Band Parameters S/d and $2\pi\gamma_o/d$ for CO_2	2-27
2-7	Band Model Parameters S/d and $S/(2\pi\gamma,)$ for H_2O	2-32
2-8	Summary of Band-model Methods for Computing Atmospheric Absorption	2-39
2-9	Summary of Laboratory Measurements of Homogeneous-path Absorption Spectra	2-41
2-10	Wide-band Absorption of Carbon Dioxide	2-43
2-11	Wide-band Absorption of Water Vapor	2-44
2-12	Conversion Table of Physical Quantities	2-60
2-13	Reflectance ρ and Emissivity ϵ of Common Terrain Features	2-75
2-14	Total Emissivity of Snow and Ice Particles	2-76
2-15	Cloud Types and Characteristics	2-82
2-16	0.5- to 0.7-micron Albedos for Various Cloud Types	2-83
2-17	Calculated Thick Cloud Emissivities	2-83
2-18	Visual Magnitudes and Effective Temperatures of Planets and the Brightest Visual and Red Stars	2-92
2-19	Estimated Number of Stars Brighter Than a Given Magnitude	2-97
2-20	Calculations of Irradiance of K and M Spectral Class Stars	2-100
2-21	Number of Stars Per Square Degree Brighter Than Photographic Magnitude M, As a Function of Galactic Latitudes	2-105
2-22	Combustion Products for a Selection of Liquid Propellants	2-107
2-23	LOX/RP-1 Combustion Products in Mole Percent Calculated at Chamber Pressure 1000 psia, Exit Pressure 14.696 psia With Shifting Equilibrium	2-108
2-24	Combustion Products for Metallized Solid Propellant Polyurethane +13% Al	2-108
2-25	Major Emission Bands for Common Molecular Combustion Products	2-115
3-1	Salient Characteristics of Optical IR Materials	3-4
3-2	Mirror Materials	3-14
3-3	Reflectivity Characteristics of Common Mirror Coatings	3-15
3-4	Standard Symbols for Par. 3-3	3-75

LIST OF TABLES (Cont)

Table No.	*Title*	*Page*
3-5	Spectral Emissivity of Tungsten at Temperatures Between 1600° and 2800°K	3-78
3-6	Selected Solid Lasers	3-92
3-7	Commonly Used Gas Lasers	3-93
3-8	IR Detector Terminology for Par. 3-4	3-105
3-9	Filters for Infrared Photography	3-119
3-10	Exposures of Infrared Photography Assuming Open Landscape, Summer Sun, Closeups, and Bright Sun	3-121
3-11	Comparative Performance of Infrared Detectors	3-139
3-12	Standard Symbols for Par. 3-5	3-144
3-13	Comparison of Cooling Techniques	3-156
3-14	Standard Symbols for Par. 3-7	3-191
3-15	Phosphor Data	3-197
3-16	Phosphor Screen Chart	3-199
3-17	Multi-gun Tubes	3-222
3-18	Standard Symbols for Par. 3-8	3-234
3-19	Window Regions in the Infrared	3-242
3-20	Methods of Protecting IR Detectors, Mirrors, and Lenses from Direct Solar Illumination	3-281
4-1	Infrared System Definitions	4-1
4-2	Comparison of 23-in. Xenon Arc, 30-in. Carbon Arc, and 30-in. Xenon Arc Searchlights	4-33
5-1	List of Symbols	5-2

PREFACE

This publication is one of a group of handbooks prepared under the auspices of the Engineering Handbook Office, Duke University, as part of the Engineering Design Handbook Series. Presented in this handbook are the basic information and fundamental principles essential to the design and development of infrared systems for military applications.

A great deal of information already has been published describing time-proven IR systems and technology; whereas, little or no information has appeared in open literature, such as this, describing the significant recent advances in IR technology and systems development. Therefore, the approach to this topic departs from the time-honored treatment of the subject of infrared technology. No attempt is made to present a complete exposition of the overall infrared discipline, but rather, most of the material is devoted to the significant technological advances of recent years.

Recent requirements for tactical nighttime surveillance and detection capabilities have introduced new challenges. The result has been the successful development and application of multi-element arrays of IR detectors for reconnaissance and surveillance. This approach has increased considerably the effectiveness of IR systems in the field of combat as well as in global defense systems. Noteworthy advances have been made in the areas of low-light-level television and long wavelength infrared. The implementation of IR searchlight and laser sources has further improved the range capability of IR systems, and has added a new capability to infrared technology—direct measurement of range or distance. It is in the light of these advances that the bulk of this book has been prepared. This handbook, therefore, is intended to complement the previously-published IR literature by bridging the gap between the historically-proven, well-documented technology and the advancing state-of-the-art.

The material is presented in a form which will be most useful to the graduate engineer who must become informed about the technology and operational performance of present-day IR systems and who appreciates their attendant advantages and limitations. This handbook is also intended to aid the professional engineer concerned with the design and development of new systems.

The information contained in this handbook consists of contributions from many infrared specialists engaged in the design and development of IR systems and associated hardware at the Electronics Division of Aerojet-General Corporation, Azusa, California. A list of contributors is given on page xxx. Special credit should be given to I. M. Maine, Program Manager, and to the following principal investigators: M. L. Bhaumik and M. A. Levine for Chapter 2, S. J. Halasz for

Chapter 4, and S. T. Braunheim for Chapter 5. A credit should be given to J. A. Lopez, who edited the various sections of the book and made it possible that a consistent list of symbols be available for each chapter and for the entire book.

Particular credit goes to Mr. R. R. Entwhistle, Mr. S. L. Hall, and Miss Peggy Nash who patiently and accurately edited the entire manuscript.

The material is organized in a logical structure, which, I hope, will result in maximum usefulness of the information. The reader is introduced to the fundamental elements of IR radiation, the basic laws governing the nature of infrared radiation and transmission. The target and background infrared signatures are then outlined and the techniques for signal detection and background discrimination are described. Each discrete element of the conventional IR system is discussed including optics, detectors, signal-processing electronics, and associated support systems. Passive as well as active IR systems presently in existence are described. Design considerations and optimization techniques are presented.

The essential equations which describe systems operation are drawn from sound and proven sources and are presented without proof. References are included at the end of each chapter. A selected bibliography is presented at the end of this handbook.

Chapter 1 gives a short history of infrared technology and the significant military applications.

Chapter 2 introduces basic infrared terminology and describes the sources of IR radiation, laws governing this radiation, atmospheric transmission and absorption, radiation from targets of military interest, and background radiation.

Chapter 3 describes the basic tools needed for the transmission and detection of infrared signals. The optical elements are analyzed as are their characteristics and basic design parameters, thermal as well as mechanical. Sources of target illumination, including lasers and their applications, are noted. The characteristics of IR receiving systems, including various infrared detectors, are studied in terms of fabrication techniques, cooling requirements and, finally, performance. Signal-processing techniques (including spatial filtering, scanning aperture, and temporal filtering) and display also are discussed in this chapter.

Chapter 4 describes the operation of the most commonly-used passive and active infrared systems. The discussion includes the principles of imaging systems (including scanners), image tubes, viewers, and sights, Search equipments discussed include trackers, radiometers, interferometers, spectrometers, and hybrid systems. Active systems include illuminators, viewers and sights, range finders, communication, data transmission, and weapon applications.

Chapter 5 discusses system design approaches and operational requirements. Systems analyses entail target and background definition, baseline design concepts, and trade-off techniques. The design of a sample infrared system is analyzed at the end of this chapter.

A separate publication, AMCP 706-128, has been prepared in order to present classified information dealing with target signatures of military interest, IR technology, and classified military systems.

K. Seyrafi

CONTRIBUTORS*

Chapter 2

T. F. Fahlen
W. J. Helm
G. A. Morton, Jr.
C. H. Parry
W. J. Pickering
A. Shulsinger
G. Weyl

Chapter 3

J. E. Ahern
P. O. Burk
G. L. Clark
P. L. Fox
K. O. Fugate
J. D. Graham
G. J. Hoover
J. Joyce
D. J. Larivee
M. A. Levine
R. C. Pitts
N. N. Richman
R. A. Spaulding
C. D. Winston
R. Y. Yagami

Chapter 4

D. J. Larivee
P. L Stoddard

Chapter 5

E. F. Davis
E. H. Dryden
R. G. Hill
T. Kotoniaš
I. M. Maine
J. C. Ranck
R. G. Richards
H. E. Walker

*All personnel are from Aerojet-General Corporation, Azusa, California.

The Engineering Design Handbooks fall into two basic categories, those approved for release and sale, and those classified for security reasons. The Army Materiel Command policy is to release these Engineering Design Handbooks to other DOD activities and their contractors and other Government agencies in accordance with current Army Regulation 70-31, dated 9 September 1966. It will be noted that the majority of these Handbooks can be obtained from the National Technical Information Service (NTIS). Procedures for acquiring these Handbooks follow:

a. Activities within AMC, DOD agencies, and Government agencies other than DOD having need for the Handbooks should direct their request on an official form to:

Commander
Letterkenny Army Depot
ATTN: AMXLE-ATD
Chambersburg, PA 17201

b. Contractors and universities must forward their requests to:

National Technical Information Service
Department of Commerce
Springfield, VA 22151

(Requests for classified documents must be sent, with appropriate "Need to Know" justification, to Letterkenny Army Depot.)

Comments and suggestions on this Handbook are welcome and should be addressed to:

Commander
US Army Materiel Command
ATTN: AMCRD-TV
Alexandria, VA 22304

(DA Forms 2028 (Recommended Changes to Publications), which are available through normal publications supply channels, may be used for comments/suggestions.)

CHAPTER 1

INTRODUCTION*

1-1 DEFINITION OF INFRARED SPECTRUM

The infrared (IR) region of electromagnetic radiation consists of that portion of the spectrum located between the longest visible wavelengths and the shortest microwave wavelengths. The IR spectral band is many times as broad as the visible optical spectrum which ranges from about 0.3 to 0.72 micron. The IR band is, therefore, divided somewhat arbitrarily into the following three regions for convenience:

1. The near IR between 0.72 - 1.2 microns
2. The intermediate IR between 1.2 - 7.0 microns
3. The far IR between 7.0 - 1000 microns

Recently, many IR systems have been developed for operation in the 8- to 30-micron region. This region, which is a subclass of the far IR region, is conventionally referred to as the Long-wavelength (LWL) Infrared Region.

1-2 MILESTONES IN THE DEVELOPMENT OF INFRARED TECHNOLOGY

This chapter contains a cursory survey of the significant milestones in the advancement of infrared technology and its application. A more detailed account of the development and applications of IR technology is given by Smith, Jones, and Chasmer[1] † and by Arnquist[2].

The actual discovery of infrared radiation was the result of Sir William Herschel's pioneering experiments in 1800[1]. While investigating the distribution of thermal energy among various colors of solar radiation, Herschel found that thermal energy increased toward the red end of the visible spectrum and continued beyond the visible region. He concluded that radiant energy exists beyond the visible region. He called this radiation "invisible radiation". Further experiments by Herschel indicated that IR radiation obeys the same laws as does visible light.

Herschel's discovery opened a new frontier in optical science. It did not lead to any further advancement for almost twenty years, however, due primarily to the lack of detectors more sensitive than the thermometer. Although progress during the nineteenth and early twentieth centuries was not dramatic, significant advancement has been made during the past 30 years.

In 1830, L. Nobili developed the thermocouple which detects IR radiation with a higher degree of sensitivity than does the thermometer. Within five years, M. Melloni developed an even more sensitive sensor by integrating a large number of thermocouples and called it a "thermopile". The development of the thermocouple and thermopile are considered to be the first important steps in the advancement of IR technology.

During 1830-1840, Sir John Herschel, the son of Sir William Herschel, continued his father's work and supported his father's conclusion regarding the nature of IR radiation. He maintained that IR radiation and visible light are similar in many basic respects. This opinion was the cause of considerable controversy until 1847 when Fizeau, Foucault, and Knoblauch illustrated that IR radiation exhibits interference effects in exactly the same way as does visible light.

In 1843, E. Becquerel discovered the photographic and phosphorescent effects of IR radiation. In 1880, S. P. Langley invented the first bolometer which is considerably more sensitive than the thermopile. A bolometer consists of a thin wire whose ohmic resistance changes as the result of the heat generated by the incident radiation.

In 1920, T. W. Case[3] developed the photoconductive "Thalofide Cell" detector which is more sensitive and has faster response characteristics than thermocouples and bolometers. Although the principles of photoconductivity were discovered by Willoughby Smith[4] in 1873 using selenium, Case should be credited for actually developing and implementing the photoconductor detector for use as an active IR infrared

---

* Written by K. Seyrafi.

† Superscript numbers refer to References at the end of each chapter.

transmitter and receiver system having an operating range of over 18 miles[2]. His device stimulated widespread interest in IR in this country and abroad.

In 1876, Adams and Day discovered the photovoltaic detector (selenium) and in 1934, Kikoin and Noskov developed the first photoelectromagnetic detector (cuprous oxide).

In 1904, Bose* discovered the photosensitive property of lead sulfide (galena). Later on, in 1917, during a routine investigation of 162 materials, Case[5] reported the photoconductor properties of lead sulfide (PbS).

In 1944, Cashman developed the first practical PbS detectors in this country[6]. Lead sulfide detectors had been previously developed by Gudden[7] in Germany during the 1930's and were used in some of their IR systems during World War II. Cashman's development marked the beginning of a rapid expansion in IR technology and systems application in this country.

Later on during the late 1940's and early 1950's, Cashman, McFee, and Levinstein extended PbS technology into lead selenide (PbSe), lead telluride (PbTe), and indium antimonide (InSb) detectors.

Another significant technological advance was the development of a pneumatic IR detector by Marcel Golay in 1946. The Golay detector is still considered the best detector for long-wavelength IR applications because of its *uniform* spectral sensitivity.

During the late 1940's, the first extrinsic photoconductor detectors were discovered by Burstein[8]. The spectral response of detectors such as gold-doped germanium (Ge:Au), zinc-doped germanium (Ge:Zn), and copper-doped germanium (Ge:Cu) was found to extend to about 40 microns. Later on, the discovery of a mercury-doped germanium detector (Ge:Hg), having a high-sensitivity response up to 14 microns, was reported by Borrello and Levinstein[9]. Most of these detectors required cooling to temperatures ranging from 4° to 40°K.

In 1959, Lawson[10] described the first pseudo-binary detector (mercury-cadmium telluride) as having a spectral response that could be extended to 40 microns. This detector, in contrast to other LWLIR detectors, required cooling only to about 77°K (liquid nitrogen).

Today, because of the availability of highly-sensitive detectors in the range of about 0.7 to 40 microns, the most useful part of the IR spectrum can be detected with almost the same sensitivity as visible light.

1-3 MILITARY APPLICATION

Although IR has been part of the scientific world for over 150 years, its application for military purposes has only taken place within the last 40 years. Secure signaling, detection of objects in the dark, and detection of and homing on military targets by their natural IR radiation are a few of the many military applications of IR systems. During World War I, the Americans, British, and Germans produced IR devices which were, for the most part, experimental.

In 1920, S. Huffman[11] described a passive imaging system which utilized a thermopile and galvanometer. This system could detect a man at 600 feet and an airplane at nearly 1 mile.

During the 1930's, IR systems found their way slowly into the military arsenals. The advent of World War II, however, accelerated the tempo of activity in this area. During 1940-1941, the Optics and Camouflage Branch of the National Defense Research Committee and Office of Scientific Research and Development was assigned the responsibility for developing the military potential of optics and IR systems.

During 1947-1955, the growing military demand for IR systems necessitated close technical coordination between the military and contract groups of the rapidly developing Southern California infrared community. Consequently the Office of Naval Research (ONR) Branch Office, Pasadena, California, sponsored a series of meetings under the leadership of W. N. Arnquist[12]. Initially these gatherings were called "The Conference on IR Instrumentation", and then, as the emphasis shifted to systems, the "Guided Missile Infrared Conference". An effective information exchange and discussion forum were thus provided for the relatively new work-

* Patent — 755,840. "Detector For Electrical Disturbances," Jagadis C. Bose, Calcutta, India, assignor of one-half to Sara Chapman Bull, Cambridge, Mass., filed Sept. 30, 1901, Serial No. 77,028 (No model).

ing groups in the area. The attendance at these conferences grew at such a rapid rate that by the mid-1950's it was no longer possible to continue on such an informal basis. Instead, formal meetings, called "Symposiums", replaced the informal conferences within the necessary security regulations. In November 1955, the name "Infrared Information Symposium" (IRIS), was formally adopted. IRIS symposiums have been convened regularly at least once a year ever since. IRIS has provided an effective means for exchanging information and stimulating new ideas for the advancement of IR research and technology. Other activities of ONR that stimulated developments in IR, especially during 1950-55, were Project Metcalf, a comprehensive review of the Navy's IR program and special liaison with the British through ONR's London Branch Office[12].

Today, various basic types of IR systems are used in conjunction with tactical weapons in military arsenals throughout the world. The brief summary which follows is a description of some of the more notable applications of IR systems.

1-3.1 IR IMAGING SYSTEMS

Development of the RCA infrared image tube by Morton, Ramberg, and Zyorykin[2] is considered to be the most significant IR development in the United States during the 1930's. The device converts IR radiation into visible light.

The first and most notable military application of near-IR technology during World War II involved use of the IR imaging tube mounted on small arms for use by foot soldiers[13]. Named the Sniperscope, it was used successfully during night operations. Infrared illuminators and receiving devices were used for night driving and battlefield surveillance by all the major powers.

1-3.2 INFRARED MISSILES

Chronologically, the next step in the advancement of IR occurred as a result of its successful application for air-to-air and air-to-surface missile guidance. The Germans developed an IR antiaircraft missile detection system during the early 1930's, to the point where piston-engine bombers could be observed at distances up to five km. However, they were not successful in implementing this latter capability in an operational system.

In this country, the use of IR in the missile field reached its peak after World War II. The most successful developments in the 1950's were the Navy SIDEWINDER air-to-air missile and the Air Force FALCON homing missile[14]. The SIDEWINDER missile is 5 in. in diameter, 9 ft long, and weighs 150 lb. It can be carried by a variety of aircraft, including the F84, F-104, and FJ-4.

The FALCON missile was developed by the Air Force to complement a radar-guided missile. These missiles are about 6-1/2 in. in diameter, about 6-1/2 ft long, and weigh slightly more than 120 lb. Used in F102A and F-89H interceptors, the missile's tracking capability is such that it can be launched many miles from the target.

Beginning about 1958 and extending to the early 1960's, the REDEYE missile was developed for the Army, to provide the foot soldier with a defense against low-flying aircraft. The 20-lb missile, which is less than 3 in. in diameter and 4 ft long, is aimed and fired from a shoulder-mounted launch tube.

1-3.3 INFRARED FIRE CONTROLS

The first IR fire control system for search, acquisition, and tracking became operational in the mid-1950's. A gunsight was developed at this time for the F-104 aircraft, and the AAA-4 IR fire control system was developed in the early 1960's by the Navy.

An early version of a down-looking (3 to 5 microns) IR reconnaissance system was developed by the Air Force in the early 1960's.

Recent breakthroughs in the fabricating of large arrays (consisting of LWIR sensors, cryogenic cooling, and LWL optics) have further increased the potential capability of IR high-resolution fire-control systems, unmatched by any other system at this time[15,16]. Forward-looking IR Reconnaissance (FLIR) systems, designed for mounting on airborne platforms, provide armament system operators with IR target detection, acquisition, recognition, and angle information. The FLIR systems, developed specially for nighttime use, provide real-time display of the

terrestrial scene within the field of view of the sensors.

1-3.4 TAIL-WARNING SYSTEMS

Infrared systems are also used for detecting missile plumes. An IR search system for use in the detection of attacking missiles was developed by the Air Force as a tail-warning system for a fighter aircraft and became operational in the mid-to-late 1960's.

1-3.5 SPACE APPLICATIONS

The first spaceborne high-resolution IR temperature mapping system was launched aboard the Nimbus 1 satellite into an earth orbit in 1964[17]. The Nimbus infrared detector, sensing in the 3.4- to 4.2-micron region, provided the first nighttime cloud pictures with a ground resolution of about two nmi.

Nimbus 1 performed temperature measurements with a resolution of about 1°K making it possible to map gradients in ocean currents, ice caps, land masses, and cloud formations. Furthermore, because of the correlation between cloud temperature and cloud altitude (a 1°K temperature change is comparable to a 1000-ft change in altitude), the IR picture provided a simple and effective method of determining cloud altitude.

1-3.6 SPECTROSCOPY

Infrared spectroscopy plays a key role in detection systems for military applications. Evaluation of targets and background in terms of spectral intensity has provided considerable information for use in the design of effective detection and homing systems. It has also provided an enormous amount of information about the sun, planets, and stars. New low-temperature stars have been discovered[18]. Terrestrial atmospheric phenomena and the atmospheres of other planets are being investigated.

In biology and medicine, infrared techniques are continually finding new uses and applications. Spectroscopy has made possible the study of plant diseases as well as the characterization and identification of fossils.

1-4 ADVANTAGES AND DISADVANTAGES OF INFRARED SYSTEMS

Infrared systems offer a distinct advantage over other detection devices, such as radar or visible optics, by being able to operate in the passive mode. This makes IR systems impervious to detection and countermeasures by methods which are effective against active systems such as radar. In addition, the passive IR systems are less complex by the absence of transmitter hardware. The fact that most natural objects radiate in the IR region makes IR wavelengths most attractive for passive systems.

A summary of the advantages of IR systems includes:

1. Small size and lightweight compared to comparable active systems
2. Low cost compared to active systems
3. Capable of passive or active operation
4. Effective against targets camouflaged in the visible region of the optical spectrum
5. Day or night operation
6. Greater angular accuracy than radar
7. No minimum range limitation
8. Minimum requirement for auxiliary equipment.

The performance limitations of IR systems are imposed mostly by atmospheric conditions. Humid atmosphere, fog, and clouds present serious limitations. The problems can be briefly summarized as follows:

1. Lack of all-weather capability (in operation within the atmosphere)
2. Line-of-sight detection capability *only*
3. Requirements for cryogenic cooling during LWL operation.

Notwithstanding these inherent limitations, IR technology faces an ever-expanding future made possible by the recent dynamic advances in the development of components such as solid-state detectors and detector arrays, cooled optics, cryogenic electronics, and IR lasers.

REFERENCES

1. R. A. Smith, et al., *The Detection and Measurement of Infrared Radiation*, Oxford at the Clarendon Press, 1957.
2. W. N. Arnquist, "Survey of Early Infrared Developments", Proc IRE 47, No. 9, 1420-1431 (1959).
3. T. W. Case, "Thalofide Cell—A New Photo-Electric Substance", Phys. Rev. 15, 289 (1920).
4. W. Smith, "Effect of Light on Selenium", Nature 7, 303 (1873).
5. T. W. Case, "Notes on the Change of Resistance of Certain Substances in Light", Phys. Rev. 9, 305 (1917).
6. R. J. Cashman, "New Photo-Conductive Cells", J.D.S.A. 36, 356 (1946).
7. J. A. Jamieson, et al., *Infrared Physics and Engineering*, McGraw-Hill Book Co., New York, 1963.
8. E. Burstein, J. J. Oberly, J. W. Davisson, and B. W. Henvis, "The Optical Properties of Donor and Acceptor Impurities in Silicon", Phys. Rev. 82, 764 (1951).
9. S. R. Borrello and H. Levinstein, "Preparation and Properties of Mercury-doped Germanium", J. Appl. Phys. 33, No. 10, (1962).
10. W. D. Lawson, S. Neilson, E. H. Putley, and A. S. Young, "Preparation and Properties of HgTe and Mixed Crystals of HgTe-CdTe", J. Phys. Chem. Solids 9, 325 (1959).
11. S. D. Hoffman, "The Detection of Invisible Objects by Heat Radiation", Phys. Rev. 14, 163-166 (1919).
12. W. N. Arnquist, Proc. IRIS 1, 5 (June 1956).
13. R. S. Wiseman and M. W. Klein, "Infrared Viewing Systems", Proc. IRE 47, 1617 (1959).
14. L. W. Nichols, "Missile Seekers and Homers", Proc. IRE 47, No. 2, 1611-1614 (1959).
15. G. J. Hoover, *Fabrication and Statistical Evaluation of High Density Ge:Hg Linear Arrays*, presented at 16th National Meeting of IRIS, 1968.
16. Hodges, et al., *System Optimization of Long-Wavelength Infrared Detection Systems*, Aerojet Report No. 3009, Azusa, Calif., March 1965.
17. *TRW Space Log*, 4—23-24 (Winter 1964-65).
18. R. C. Ramsey, "Spectral Irradiance from Stars and Planets Above the Atmosphere from 0.1 to 100 μ", Appl. Opt. 6, No. 4, 465 (1962).

CHAPTER 2

INFRARED PHYSICS*

2-1 BASIC IR SYMBOLS AND DEFINITIONS

During the advance of IR technology, various symbols have evolved as "standard" symbols while others have been used at the whim of the authors. The system of symbols used in this handbook is based on the recommendations of the Office of Naval Research[1]. Special care has been given to distinguish *intrinsic* material properties such as absorp*tivity* α from total sample properties such as absorp*tance* a. Some symbols occasionally have two functions. For instance, t can mean either time or total transmittance. Where confusion might possibly occur, the terms are carefully explained in the accompanying text. The standard symbols are given in Table 2-1.

Table 2-2 is a dictionary of the basic radiometric terms used in infrared physics. *Spectral* radiometric terms are the same as the corresponding radiometric terms but are defined per unit wavelength, per unit frequency or per unit wavenumber. They are evaluated at a specific wavelength, frequency or wavenumber. Symbols for spectral radiometric quantities are formed by adding subscripts λ, ν, or $\tilde{\nu}$, respectively, referring to where the quantities are to be evaluated. For example, H_λ is the irradiance per unit wavelength evaluated at wavelength λ. Typical units would be w cm^{-2} μ^{-1}. The relation between H and H_λ is

$$H = H_\lambda \Delta\lambda \quad \text{or} \quad H_\lambda = \frac{\partial H}{\partial \lambda}$$

where $\Delta\lambda$ is a small wavelength interval about λ.

Table 2-3 is a list of the most commonly used physical constants for infrared physics. The standard symbols for the constants is also given. A complete list and discussion are given in Ref. 2.

*Written by M. L. Bhaumik and M. A. Levine

2-2 RADIATION LAWS

2-2.1 KIRCHHOFF'S LAW

Kirchhoff's Law states that, for any temperature and any wavelength, the emissance of an opaque body in an isothermal enclosure is equal to its absorptance. Thus

$$e(\lambda, T) = a(\lambda, T) \tag{2-1}$$

This law is a consequence of the Conservation of Energy which requires that the energy emitted by the body be equal to the energy absorbed by it under isothermal conditions. Thus

$$W_{emitted} = eW_{BB} = aW_{BB} = W_{absorbed} \tag{2-2}$$

For a blackbody, $e = a = 1$ by definition. For real materials emissance depends on the material and the finish (see Table 2-4).

More generally, the Conservation of Energy says that light incident on a surface is either reflected, transmitted, or absorbed. Thus

$$r + t + a = 1 \tag{2-3}$$

For an opaque object, $t = 0$. Therefore

$$r + a = 1 \tag{2-4}$$

Using Kirchhoff's Law, Eq. 2-1 gives

$$r = 1 - e \tag{2-5}$$

for an opaque object.

TABLE 2-1. STANDARD SYMBOLS

SYMBOL	NAME	TYPICAL UNITS
a	absorptance	%
A	area	cm^2
BB	blackbody	—
BG	background	—
e	emissance	%
E	photon energy	erg
GB	graybody	—
H	irradiance	$w\ cm^{-2}$
J	radiant intensity	$w\ sr^{-1}$
N	radiance	$w\ sr^{-1}\ cm^{-2}$
P	power	w
r	reflectance	%
S	line intensity	$(absorber\ content)^{-1}\ cm^{-1}$
T	absolute temperature	°K
t	transmittance	%
U	energy	J
u	energy density	$J\ cm^{-3}$
V	volume	cm^{-3}
W	radiant emittance (flux density)	$w\ cm^{-2}$, $(J\ sec^{-1}\ cm^{-2})$
w	absorber content	"atm-cm", "pr-cm"
α	absorbtivity	cm^{-1}
ϵ	emissivity	%
λ	wavelength	cm
ν	frequency	Hz, (sec^{-1})
$\bar{\nu}$	wavenumber	cm^{-1}
ρ	density	$g\ cm^{-3}$
τ	transmissivity	cm^{-1}
Ω	solid angle	sr (steradian)

TABLE 2-2. BASIC RADIOMETRIC TERMS

TERM	DEFINITION
Absorber content	The equivalent pathlength through a gaseous absorber (par. 2-4)
Absorptance	The fraction of irradiance that is absorbed by a sample placed in the path of the incident light
Absorptivity	The absorptance per unit pathlength through a material
Blackbody	An ideal radiator or absorber with unit emissivity
Emissance	The fraction of radiant emittance of a real surface relative to a blackbody surface
Emissivity	The fraction of radiant emittance of an ideal surface (opaque, optically smooth, flat) relative to a blackbody surface
Energy density	The energy per unit volume contained in the electromagnetic fields
Graybody	A radiator or absorber with constant emissivity less than one; i.e., $\epsilon_{GB} < 1$ for all λ
Irradiance	The power per unit area incident upon a surface
Point source	A radiating surface both characteristic dimensions of which are small compared with the source-to-observer distance
Power	Energy per unit time
Radian	The unit of angular measure, which is the angle for which the subtended arc length of a circle is equal to the radius of the circle (Fig. 2-1)
Radiance	Radiant power per unit solid angle per unit area of source projected normal to the solid angle
Radiant emittance	The power per unit area, or the energy per unit time per unit area, radiated from a surface. Radiant emittance is an energy flux.
Radiant intensity	Radiant power per unit solid angle from a point source
Reflectance	The fraction of irradiance that is reflected from a real surface
Reflectivity	The fraction of irradiance that is reflected from an ideal surface (perfectly smooth and flat)
Steradian	The unit of solid angular measure, being the subtended surface area divided by the radius squared for a solid angle at the center of a sphere (Fig. 2-2)
Transmittance	The fraction of irradiance that is transmitted through a sample placed in the path of the incident light
Wavelength	The distance between two successive crests in the electromagnetic field of light traveling through a vacuum
Wavenumber	Reciprocal of wavelength in centimeters. Wavenumbers are proportional to the photon energy of the light ($E = hc\bar{\nu}$) where $\bar{\nu} = 1/\lambda$.

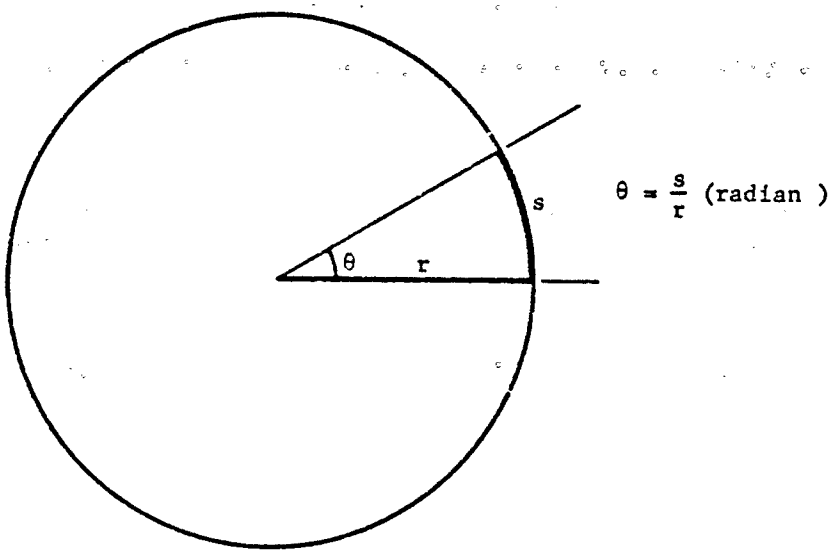


FIGURE 2-1. Angle θ in Radians

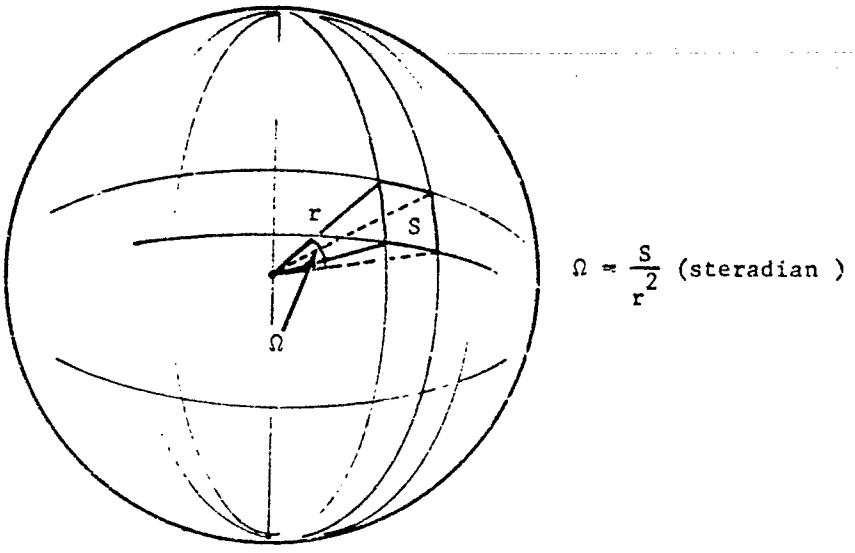


FIGURE 2-2. Solid Angle Ω in Steradians

TABLE 2-3. PHYSICAL CONSTANTS

Planck's constant	h	6.6256×10^{-34} w sec^2
Speed of light	c	2.9979×10^{10} cm sec^{-1}
Boltzmann's constant	k	1.3805×10^{-23} w sec °K^{-1} 8.617×10^{-5} eV) °K^{-1}
Stefan-Boltzmann constant	σ	5.6697×10^{-12} w cm^{-2} °K^{-4} 1.354×10^{-12} cal sec^{-1} cm^{-2} °K^{-4} 3.658×10^{-11} w in.$^{-2}$ °K^{-4}
Wein's constant	a	0.28978 cm °K
Electronic charge	e	1.6021×10^{-19} C
Capacivity of vacuum	ϵ_o	8.8543×10^{-12} F m^{-1}
Permeability of vacuum	μ_o	$4\pi \times 10^{-7}$ H m^{-1}
Electron mass	m_e	9.109×10^{-28} g
Proton mass	m_p	1.673×10^{-24} g
Avogadro's number	N_A	6.023×10^{23} per g mole
1 electron volt	=	1.602×10^{-19} w sec
Energy of 1°K	=	8.617×10^{-5} eV)
λ (1 eV)	=	1.24 μ

2-2.2 PLANCK'S LAW

The spectral energy flux, or spectral radiant emittance W_λ from a blackbody was derived empirically by Planck[8] who had to postulate that radiation consisted of discrete quanta of energy hc/λ in order to fit a smooth curve to the experimentally measured spectral distributions of radiant emittance from blackbodies. The expression he derived was

$$W_\lambda = \frac{C_1}{\lambda^5} \frac{1}{\left[e^{C_2/(\lambda T)} - 1\right]}, \text{ w cm}^{-3} \qquad (2\text{-}6)$$

The constants are

$C_1 = 2\pi c^2 h = 3.7415 \times 10^{-12}$ w cm^2

$C_2 = hc/k = 1.4388$ cm °K

c = speed of light
h = Planck's constant
k = Boltzmann's constant
λ = wavelength, cm

A plot of W_λ vs λ will result in the familiar blackbody radiation curves for various temperatures (Fig. 2-3).

2-2.3 RAYLEIGH-JEANS LAW

For long wavelengths (far IR) at not too low a temperature,

$$hc/\lambda kT \ll 1 \qquad (2\text{-}7)$$

which permits the exponential to be expanded in a power series. Thus

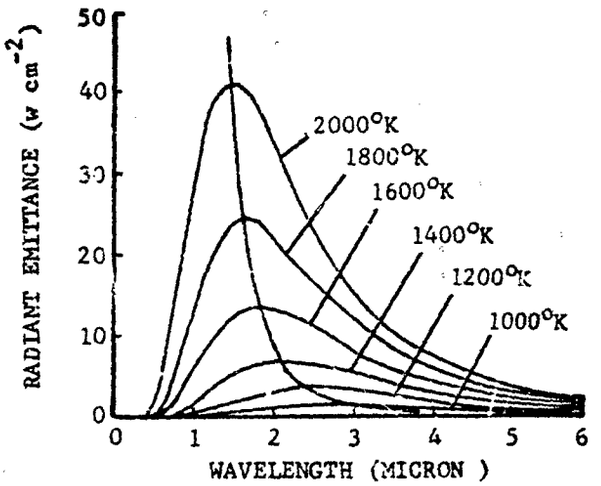
$$W_\lambda = 2\pi ckT/\lambda^4 \qquad (2\text{-}8)$$

which is the Rayleigh-Jeans Law. Notice that it does not depend on h and therefore is not of quantum mechanical origin.

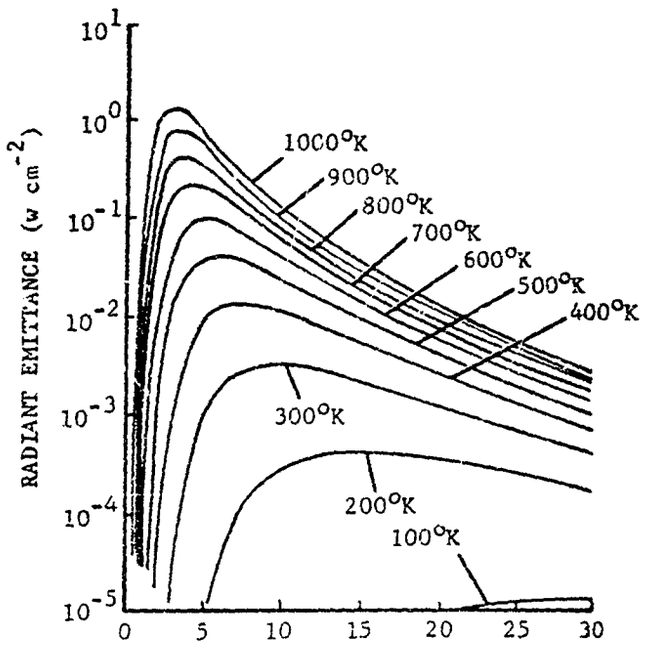
TABLE 2-4. EXPERIMENTAL VALUES OF EMISSIVITY*

MATERIAL	300°K RADN. ON 78°K SURFACE[3]	ROOM TEMP.[4]	14μ RADN. ON 2°K POLISHED SURFACE[5]	293°K RADN. ON 90°K SURFACE[6]	273°K RADN. ON 77°K SURFACE[7]
Al-clean polished foil	0.02	0.04	0.011	0.055	0.043
Al-plate	0.03				
Al-highly oxidized		0.31			
Brass-clean polished	0.029	0.03	0.018	0.046	0.10
Brass-highly oxidized		0.6			
Cu-clean polished	0.015-0.019	0.02	0.0062-0.015	0.019-0.035	
Cu-highly oxidized		0.6			
Cr-plate	0.08	0.08		0.065	0.084
Au-foil	0.010-0.023	0.02-0.03		0.026	
Au-plate	0.026				
Monel		0.2			0.11
Ni-polished		0.045			
Rh-plate	0.078				
Ag-plate	0.008	0.02-0.03		0.023-0.036	
Stainless steel	0.048	0.074			
Sn-clean foil	0.013	0.06	0.013	0.038	
Soft solder	0.03				0.047
Glass		0.9		0.87	
Wood's metal					0.16

* Note: Reference is made in column heads to references listed at end of Chapter 2.



(A) BLACKBODY CURVES, 1000° TO 2000°K



(B) BLACKBODY CURVES, -100° TO 1000°K

FIGURE 2 3. Blackbody Curves

2-2.4 WIEN'S LAW

For short wavelengths,

$$hc/\lambda kT \gg 1, \tag{2-9}$$

hence Planck's Law reduces to

$$W_\lambda = \frac{2\pi c^2 h}{\lambda^5} \exp\left[-hc/(\lambda kT)\right] \tag{2-10}$$

which is known as Wien's Law.

2-2.5 STEFAN-BOLTZMANN LAW

The radiant emittance from a blackbody W_{BB} can be obtained by integrating Eq. 2-6 over all wavelengths. Thus

$$W_{BB} = \int_0^\infty W_\lambda d\lambda = \left(\frac{2\pi^5 k^4}{15\, h^3 c^2}\right) T^4 = \sigma T^4 \tag{2-11}$$

which is the Stefan-Boltzmann Law. The Stefan-Boltzmann constant σ for various units is listed in Table 2-3.

2-2.6 WIEN'S DISPLACEMENT LAW

The maximum value of W_λ occurs at some λ called λ_{max}. W_λ in Eq. 2-6 is of the form

$$W_\lambda = \lambda^{-5} f(\lambda T). \tag{2-12}$$

Setting

$$\left.\frac{d\, W_\lambda}{d\lambda}\right|_{\lambda = \lambda_{max}} = 0 \tag{2-13}$$

gives

$$\lambda_{max} T = \left.\frac{5}{\dfrac{d\,[\ln f(\lambda T)]}{d(\lambda T)}}\right|_{\lambda = \lambda_{max}} \tag{2-14}$$

For a given temperature T, evaluating λ at λ_{max} is equivalent to evaluating λT at $\lambda_{max} T$. Hence

$$\lambda_{max} T = \left.\frac{5}{\dfrac{d\,[\ln f(\lambda T)]}{d(\lambda T)}}\right|_{\lambda_{max} T} = a \tag{2-15}$$

where a is a constant. Eq. 2-15 is Wien's Displacement Law. Solving the transcendental equation numerically for $\lambda_{max} T$ gives a = 0.2898 cm °K.

2-3 RADIANT ENERGY TRANSMISSION

2-3.1 RADIANT INTENSITY FROM A POINT SOURCE

A point source is a radiator all the dimensions of which are small compared to the source-to-observer distance. There are two basic types, the isotropic point source and the Lambertian point source.

2-3.1.1 Isotropic Point Source

The isotropic point source radiates uniformly in all directions, thus the radiant intensity J is

$$J = \frac{P}{4\pi}, \text{ w/ sr}^{-1} \tag{2-16}$$

where P is the total power radiated by the source.

Geometrically, the importance of an isotropic point source is that it presents the same radiating area to the viewer when looked at from any direction. It is assumed, of course, that the source, being small, has a uniform temperature.

2-3.1.2 Lambertian Point Source

A Lambertian "point" source is flat and does not present the same area to the viewer from all directions. Since the color does not change with the position of the viewer, he must assume that the radiant emittance is proportional to the apparent, or projected, area of the source. Thus

$$J = \frac{dP}{d\Omega} = kWA\cos\theta \tag{2-17}$$

where A is the area of the source, θ is the angle of the viewer with respect to the normal, and k is a proportionality constant (Fig. 2-4(A)). Integrating Eq. 2-17 over a hemisphere above the plane of dA (Fig. 2-4(B)) one gets the total radiated power

$$P = \int_{hemisphere} kWA \cos\theta \, d\Omega \qquad (2\text{-}18)$$

$$= kWA \int_0^{\frac{\pi}{2}} \cos\theta \, (2\pi \sin\theta \, d\theta) = (\pi k) WA$$

Since $P = WA$, therefore $k = \pi^{-1}$. Thus

$$J = \frac{WA \cos\theta}{\pi} \text{ , w sr}^{-1} \qquad (2\text{-}19)$$

This is called Lambert's Law.

2-3.2 RADIANT ENERGY DENSITY

Radiant energy density is the quantity of radiant energy per unit volume.

2-3.2.1 Energy Density for Collimated Irradiance

Collimated irradiance means that the flow of radiant energy is neither divergent nor convergent. Hence the energy flow down a tube of cross section dA is uniform and the energy density is constant. Therefore, for a tube of length ct, where t is the length of time during which energy entered the tube, and c is the speed of light,

$$u = \frac{U}{V} = \frac{HtdA}{ctdA} = \frac{H}{c} \qquad (2\text{-}20)$$

Thus the irradiance on any cross section of the tube is

$$H = uc \qquad (2\text{-}21)$$

This is also the radiant emittance from the same cross section.

2-3.2.2 Energy Density Within an Isothermal Enclosure

The temperature within an isothermal enclosure is a constant. Therefore the energy flux uc is a constant regardless of direction. Otherwise, there would be a net transfer of energy along some direction indicating a temperature difference somewhere. Therefore a calculation of the energy density next to a wall of the enclosure is sufficient.

The radiant emittance is given by

$$dW = \left(\frac{1}{2}u\right) c \frac{d\Omega}{4\pi} \qquad (2\text{-}22)$$

The factor 1/2 is because half the energy density is from irradiance and half from emittance as noted above. Integrating Eq. 2-22 over the hemisphere of Fig. 2-4(B) gives

$$W = \frac{uc}{4} \qquad (2\text{-}23)$$

Thus the energy density within an isothermal enclosure is

$$u = \frac{4W}{c} \qquad (2\text{-}24)$$

As an example, consider a 300°K isothermal enclosure.

$$u = \frac{4W}{s} = \frac{4\sigma T^4}{c} = \frac{4 \times 5.67 \times 10^{-12} \times (300)^4}{3 \times 10^{10}}$$

$$= 6.1 \times 10^{-12} \text{ J cm}^{-3}$$

As a second example, consider the energy density above the earth's atmosphere due to the sun. The solar irradiance is about 0.13 w cm^{-2} and is nearly collimated. Therefore

$$u = \frac{H}{c} = \frac{0.13}{3 \times 10^{10}} = 4.3 \times 10^{-12} \text{ J cm}^{-3}$$

2-3.3 TRANSFER OF RADIANT POWER

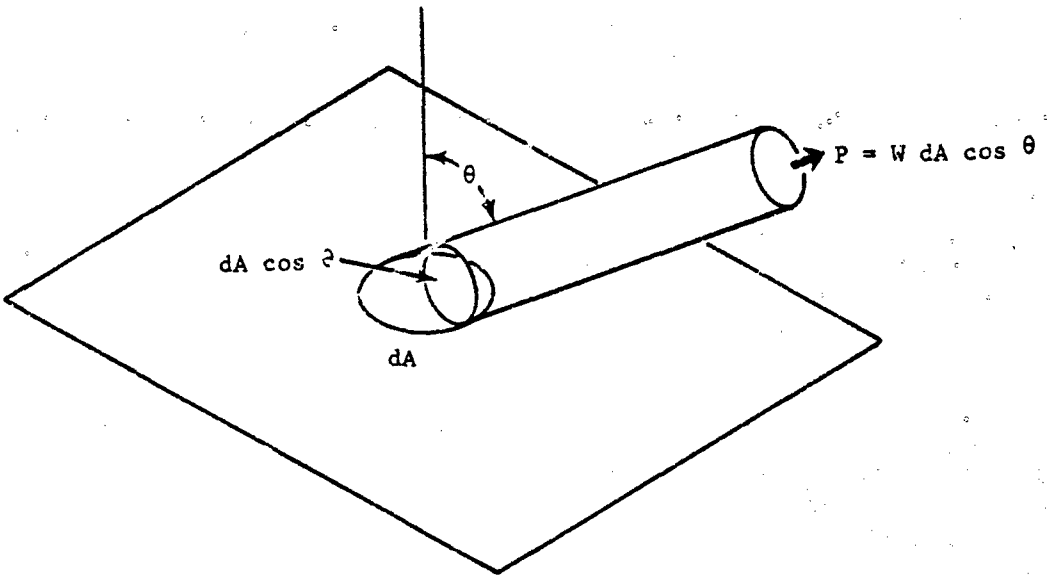
2-3.3.1 An Object in Space

An object in space receives energy from warm celestial bodies and radiates energy into space. Space can be considered a perfect absorber with no emittance, i.e., space is cold. Therefore, for a uniformly painted metal plate suspended in space near the earth and normal to the sun, the incident energy is $eA \times 0.13$ w cm^{-2}, neglecting radiation from other sources. The emitted energy is $2eA\sigma T^4$, where the 2 occurs because both sides of the plate emit. Thus, at equilibrium,

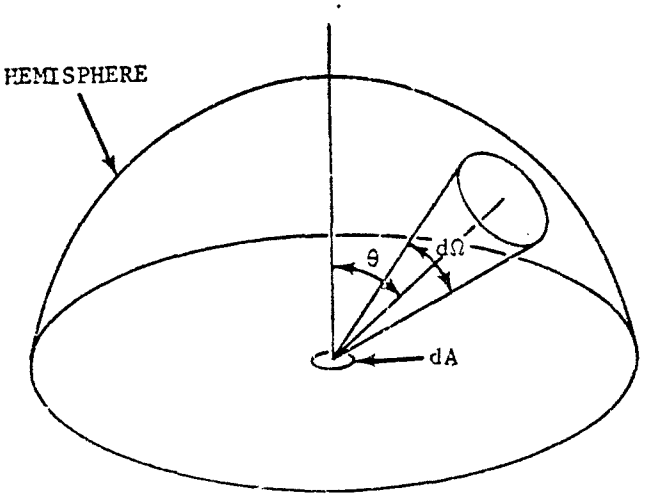
$$eA \times 0.13 = 2eA\sigma T^4$$
$$0.13 = 2 \times 5.67 \times 10^{-12} \times T^4$$
$$T = 327°\text{K}$$

If one side of the plate is black ($e = 1$) and the other is unpainted ($e = 0.1$) then, with the black side facing the sun,

$$1 \times A \times 0.13 = (1 + 0.1) A\sigma T^4$$
$$0.13 = 1.1 \times 5.67 \times 10^{-12} \times T^4$$
$$T = 380°\text{K}$$



(A)



(B)

FIGURE 2-4. Geometry of a Lambertian Source

With the shiny side facing the sun,

$$0.1 \times A \times 0.13 = (1 + 0.1)\, A\sigma T^4$$
$$0.013 = 1.1 \times 5.67 \times 10^{-12} \times T^4$$
$$T = 213^\circ K$$

2-3.3.2 Transfer Between Two Infinite Planes

For two opaque plane-parallel infinite surfaces with emissances e_1 and e_2 and temperatures $T_2 > T_1$, one can calculate the *net* radiant emittance $W_{2 \rightarrow 1}$ from surface 2 to surface 1. Surface 2 has a radiant emittance $e_2 \sigma T_2^4$. Of that emittance $e_1 e_2 \sigma T_2^4$ is absorbed by surface 1 and $(1 - e_1) e_2 \sigma T_2^4$ is reflected. Then $(1 - e_2)(1 - e_1) e_2 \sigma T_2^4$ is reflected back toward surface 1. Of that radiant emittance, $e_1 (1 - e_2)(1 - e_1) e_2 \sigma T_2^4$ is absorbed, etc. Thus

$$W_{2 \rightarrow 1} = \sigma T_2^4 \left[e_1 e_2 + e_1 (1 - e_2)(1 - e_1) e_2 + e_1 (1 - e_2)^2 (1 - e_1)^2 e_2 + \ldots \right] \tag{2-25}$$

Summing the series gives

$$W_{1 \rightarrow 2} = \sigma T_2^4 \left[\frac{e_1 e_2}{1 - (1 - e_1)(1 - e_2)} \right] \tag{2-26}$$

Similarly,

$$W_{1 \rightarrow 2} = \sigma T_1^4 \left[\frac{e_1 e_2}{1 - (1 - e_1)(1 - e_2)} \right] \tag{2-27}$$

Therefore $W_{2 \rightarrow 1}$ is given by

$$W_{2 \rightarrow 1} = W_{2 \rightarrow 1} - W_{1 \rightarrow 2} = \sigma (T_2^4 - T_1^4) \left[\frac{e_1 e_2}{1 - (1 - e_1)(1 - e_2)} \right] \tag{2-28}$$

A more extensive study of the transfer of radiant power is given in Ref. 4.

As an example, consider a black plate ($e_1 = 1$) at 2°K facing an electropolished copper plate ($e_2 = 0.01$) at 4°K. The net radiant emittance from the warmer copper plate to the cooler black plate is

$$W_{2 \rightarrow 1} = \frac{5.67 \times 10^{-12} (4^4 - 2^4) \times 1 \times 0.01}{1 - (0)(0.99)} = 1.36 \times 10^{-11} \text{ w cm}^{-2}$$

Note that the commonly accepted formula, $W_{2 \rightarrow 1} = \sigma (e_2 T_2^4 - e_2 T_1^4)$, yields $W_{2 \rightarrow 1} = -7.6 \times 10^{-11}$ w cm^{-2} indicating that the cooler plate is heating up the warmer one.

2-3.4 IRRADIANCE AS A FUNCTION OF RANGE

2-3.4.1 Point Source

A small receiving area dA is oriented normal to the line of sight at a distance R from a point source. It subtends a solid angle $d\Omega = dA/R^2$. From the definition of radiant intensity J as the power radiated per unit solid angle from a point source, the power incident on the receiving area is

$$dP = Jd\Omega = J\left(\frac{dA}{R^2}\right) \tag{2-29}$$

and the power per unit area is the irradiance H given by

$$H = \frac{dP}{dA} = \frac{J}{R^2} \tag{2-30}$$

As an example, consider the irradiance at 400 cm from a 2 cm radius sphere with emittance $W = 3$ w cm^{-2} if the receiving area is slanted 60° to the line-of-sight.

$$H = \frac{J\,dA/R^2}{dA/\cos\theta} = \frac{J\cos\theta}{R^2} = \frac{WA_{source}\cos\theta}{4\pi R^2}$$

Thus,

$$H = \frac{3 \times 4\pi\, 2^2 \times \frac{1}{2}}{4\pi(400)^2} = 3.75 \times 10^{-5} \text{ w cm}^{-2}$$

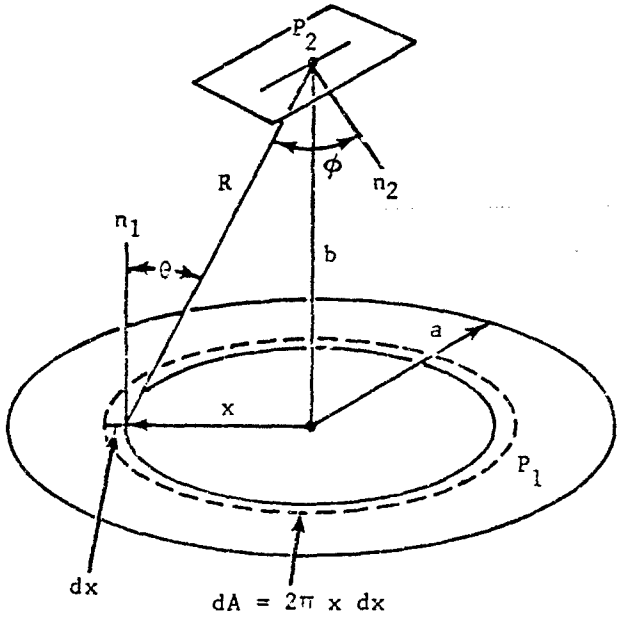
2-3.4.2 Extended Source

The concept of radiant intensity cannot be usefully applied to an extended source (one that subtends a finite solid angle to the viewer). However, in calculating irradiance H the extended source is divided into elemental areas dA and an expression is obtained for the radiant intensity of each. Contributions from each elemental source area to the irradiance are then integrated to obtain the total irradiance.

If θ is the angle between the line of sight and the normal n_1 to one of these small areas, and ϕ is the angle between the line of sight and the normal n_2 to the surface at which the irradiance is being determined, the irradiance is calculated as

$$H = \int_A \frac{dJ}{R^2}\cos\phi = \int_A \frac{W dA\cos\theta}{\pi R^2}\cos\theta \tag{2-31}$$

For example, assume that a plane circular area P_1 of radius a has an emittance W. What is the irradiance at a point located on the axis of the circle at a distance b, in the plane P_2 parallel to the source:



Since the receiving surface is parallel to the emitting surface, $\theta = \phi$. The elemental area consisting of an annular ring of diameter x and width dx has an area $dA = 2\pi x dx$ therefore

$$H = \int_o^a \frac{W}{\pi} \cos^2 \theta \frac{2\pi x dx}{x^2 + b^2}$$

$$= W \int_o^a \left[\frac{b^2}{(x^2 + b^2)}\right] \frac{2x dx}{(x^2 + b^2)} \tag{2-32}$$

Let $y = x^2 + b^2$; then $dy = 2x dx$. When $x = 0$, $y = b^2$; when $x = a$, $y = a^2 + b^2$. The integral becomes

$$H = W \int_{b^2}^{a^2 + b^2} b^2 \frac{dy}{y^2} = \left| - \frac{Wb^2}{y} \right|_{b^2}^{a^2 + b^2}$$

$$= W \left(1 - \frac{b^2}{a^2 + b^2}\right) = W \left(\frac{a^2}{a^2 + b^2}\right) \tag{2-33}$$

In the case where the receiving surface is very close to the emitting surface ($b \ll a$), $H = W$ which simply illustrates that all the radiation emitted by unit area of the source passes through unit area of the receiver. At the opposite extreme, where $b \gg a$

$$H = \frac{Wa^2}{b^2} = \frac{WA}{\pi R^2} \tag{2-34}$$

Using Eq. 2-19 with $\theta = 0$ gives

$$H = \frac{J}{R^2} \tag{2-35}$$

which is the same as Eq. 2-30 as expected.

2-3.5 CALCULATION AIDS

Since equations such as Planck's are difficult to use directly in actual calculations, several devices have been specifically devised. These include, in order of increasing accuracy, nomographs, radiation slide rules, blackbody tables, or computer storage data.

Nomographs are multi-scaled graphs designed so that a straight line drawn through a known point on each of two scales will provide an unknown's value on a third scale (Ref. 9, p. 20). Nomographs are not used extensively because of their inherent inaccuracy.

The General Electric Radiation Calculator (Ref. 1, pp. 11-17) is the most commonly used of various radiation slide rules. Radiation slide rules are sufficiently accurate for preliminary design calculations.

Blackbody data (Ref. 1, p. 21) are available for use in calculating more precise values than are possible with slide rules, but these data tables are more difficult to use.

Precise radiometric calculations require the use of highly accurate computerized data and direct, conventional calculation methods.

2-3.5.1 Use of GE Radiation Calculator

Many of the blackbody expressions given in pars. 2-2 and 2-3 can be calculated by means of a special slide rule such as the General Electric Radiation Calculator shown in Fig. 2-5. At a single setting, the following data can be read:

INDEX TEMPERATURE

(1)	CENTIGRADE	°C
(2)	KELVIN	°K = °C + 273
(3)	FAHRENHEIT	°F = (°C + 40) $\frac{9}{5}$ - 40
(4)	RANKINE	°R = °F + 460

RADIANT EMITTANCE

(5)	w = WATTS/SQ CM	Total emittance, $W_{0-\infty}$ for various emissivities (emissances) and in various units
(6)	WATTS/SQ IN.	
(7)	BTU/SQ FT/HR	

SPECTRAL EMITTANCE

(8)	$W_{\lambda max}$ = WATTS/SQ CM/MICRON $\Delta\lambda$ AT MAXIMUM	Spectral emittance at λmax with $e = 1$.
(9)	$\frac{W_\lambda}{W_{\lambda max}}$ vs λ	Multiplying (8) by (9) gives spectral emittance at λ with $e = 1$.
(10)	$\frac{W_{0-\lambda}}{W_{0-\infty}}$ vs λ	Multiplying (5), (6) or (7) by (10) gives $\int_0^\lambda W_\lambda d\lambda$ for any given emissivity used in (5), (6) or (7).
(11)	MAX vs λ	Wavelength λ_{max} at which W_λ is a maximum. (Note: This is maximum energy flux per unit wavelength interval. It is *not* maximum photon flux per unit wavelength interval.)
(12)	WAVES/CENTIMETER	This scale converts λ_{max} to $\bar{\nu}_{max} = \lambda^{-1}_{max}$.

IRRADIANCE

(13)	INCIDENT ENERGY IN WATTS/CM² FOR 1 CM² SOURCE AT INDEX TEMPERATURE vs RANGE (CENTIMETERS) (NAUTICAL MILES)	Multiply (11) by source area (cm²) to find irradiance H at ranges from 1 meter to 1000 nautical miles.
(14)	TRANSMISSION SPECTRA OF THE ATMOSPHERE 2000 YD (1 SEA MILE) OF 17 MM PRECIPITABLE WATER	Transmission coefficient t vs wavelength λ over 1 nautical mile horizontally at approximately 80% relative humidity, 80° F. This graph is independent of index temperature setting.

PHOTON EMITTANCE

(15)	PHOTONS/SEC/CM²	Total photon flux for a blackbody at index temperature.
(16)	PHOTON ENERGY AT λ_{max} IN ELECTRON VOLTS	$E = \frac{hc}{\lambda_{max}}$ in electron-volts.

There is also a C, D scale slide rule for simple calculations.

FIGURE 2-5. The GE Radiation Calculator

2-3.5.2 Sample Calculations

The following sample problems are stated and solved by direct calculation or by using the GE Radiation Calculator.

Example 1

Setup: A blackbody being used for testing is set to a temperature of 1000°C. (A blackbody aperture of 0.6 in. dia (approx. $2/\pi$ in. dia) is assumed.)

Problem A: Calculate the radiant emittance of the blackbody.

1. Solution (by direct calculation):
   a. Convert °C to °K:
   $1000°C = 1000 + 273$
   $= 1273°K$
   b. Using the formula for the Stefan-Boltzmann Law:
   $W = \sigma T^4$
   $= 5.67 \times 10^{-12} \times (1273)^4$

Answer: $W = 14.9$ w cm^{-2}

2. Solution (using the GE Radiation Calculator):
   Align CENTIGRADE pointer of TEMPERATURE scale with 1000 scale indicating W = WATTS/SQ CM. Since the radiation source is a blackbody, read the indication at 1 on the EMISSIVITY scale.

Answer: $W = 14.9$ w cm^{-2}

Problem B: Calculate the radiant power of the source.

Solution (by direct calculation):

a. Convert aperture size in inches to area in cm^2.

$$A = \pi r^2$$
$$= \pi \left(\frac{1}{\pi}\right)^2$$
$$A = \frac{1}{\pi} \text{ sq in.}$$

Since 1 sq in. = 6.45 cm^2

$$A = 2.05 \text{ cm}^2$$

b. Using the formula for radiant power as a function of radiant emittance:

$P = WA$ (all other factors remaining constant)
$= 14.9$ w cm^{-2} $\times$ 2.05 cm^2

Answer: $P = 30.6$ w

Problem C: Calculate the spectral radiant emittance W_λ in the wavelength interval between 2.5 and 3μ.

1. Solution (by direct calculation):
   Determine the total amount of radiant energy falling below each wavelength by integrating Planck's equation over the interval $\lambda = 0$ to each given wavelength ($\lambda = 2.5$ or $\lambda = 3.0$). The answer is the difference between the two radiant quantities.
2. Solution (using the GE Radiation Calculator):
   Set the CENTIGRADE pointer of the TEMPERATURE scale to 1000°C position. Observing the $\frac{W_{0-\lambda}}{W_{0-\infty}}$ scale which indicates the "Percentage increment, of energy falling below any wavelength for a blackbody at temperature T", note that 45 percent falls below 3μ and 31.5 percent falls below 2.5μ.

   Since the total energy = 14.9 w cm^{-2}

$$W_\lambda = 14.9 \times (0.450 - 0.315)$$
$$= 14.9 \times 0.135$$

Answer: $W_\lambda = 2.00$ w cm^{-2} (approx.)

Problem D: Determine the wavelength at the point of peak radiation of the 1000°C source.

1. Solution (by direct calculation):
   Using Wien's Displacement Law:

$$\lambda_{max} = \frac{2897}{T(°K)}$$
$$= \frac{2897}{1273}$$

Answer: $\lambda_{max} = 2.27\mu$

2. Solution (using the radiation calculator):
   Retain TEMPERATURE at the same setting. Read MAX point on $\frac{W_{0-\lambda}}{W_{0-\infty}}$ scale.

$$\lambda_{max} = 2.27\mu$$

Example 2

Setup: An oxidized steel graybody is heated to 2000°K.

Problem: Calculate the radiant emittance of the object.

1. Solution 1 (by direct calculation):
   a. Calculate the blackbody emittance of 2000°K using the Stefan-Boltzmann Law.

$$W = \sigma T^4$$
$$W = 90.8 \text{ w cm}^{-2}$$

b. Using emissance tables, determine the emissance of oxidized steel. (e of oxidized steel ≈ 0.8).

c. Transpose the formula for emissance:

$$e = \frac{W}{W_{BB}} \text{ to } W = e\,W_{BB}$$

$$= 0.8 \times 90.8$$

$$W = 72.6 \text{ w cm}^{-2}$$

Solution 2 (using the radiation calculator):

Set the KELVIN pointer to the 2000° position. The radiant emittance at the 0.8 position of the EMISSIVITY (emissance) scale is approximately 72 w cm^{-2}.

Example 3

Setup: The irradiance H at a point 5 ft from a radiation source is 20 w cm^{-2}.

Problem: Determine the irradiance at a point 20 ft from the source (disregarding atmospheric attenuation).

Solution (using the Inverse Square Law):

$$H_2 = H_1\left(\frac{D_1}{D_2}\right)^2$$

$$= 20\left(\frac{5}{20}\right)^2$$

$$= 20\left(\frac{1}{16}\right)$$

$$H_2 = 1.25 \text{ w cm}^{-2}$$

2-4 ATMOSPHERIC TRANSMISSION

In passing through the atmosphere, infrared energy is attenuated before it is detected and measured. The two main causes of attenuation are molecular absorption by several minor constituents of the atmosphere and scattering due to the presence of particles of matter in the atmosphere (aerosols). Molecular absorption occurs mainly in several more or less narrow absorption bands, and is due to the ability of certain molecules to go from one state of vibration-rotation to another, thereby absorbing (or emitting) a photon. In addition, scattering causes attenuation of an incident beam of radiation because in the scattering process the energy is redistributed into all directions of propagation and lost to the observer.

2-4.1 EXTINCTION COEFFICIENT

The spectral transmittance $t(\lambda)$ through a path x of uniform atmosphere is

$$t(\lambda) = \exp\,[-K(\lambda)x] \qquad (2\text{-}36)$$

where $K(\lambda)$ is the extinction coefficient at the discrete wavelength λ. $K(\lambda)$ is the sum of the molecular absorption coefficient $\alpha(\lambda)$ and the scattering coefficient $\beta(\lambda)$. Thus Eq. 2-36 can be written as

$$t(\lambda) = \exp\left\{-\left[\alpha(\lambda) + \beta(\lambda)\right]x\right\}$$

$$= \exp\left[-\alpha(\lambda)x\right] + \exp\left[-\beta(\lambda)x\right] \qquad (2\text{-}37)$$

The scattering and absorption coefficients, and therefore the extinction coefficient, depend on wavelength, atmospheric density, and the atmospheric composition. Care must be used when applying Eqs. 2-36 or 2-37 since the properties of the atmosphere, and hence its absorption, may change over the path x. To account for these changes, Eq. 2-36 can be written in integral form as

$$t(\lambda) = \exp\left[-\int_{x_1}^{x_2} K_x(\lambda)dx\right] \qquad (2\text{-}38)$$

If a finite wavelength interval, $\Delta\lambda = \lambda_2 - \lambda_1$, is considered, then the average transmittance t in the interval $\Delta\lambda$ is

$$t = \frac{1}{\lambda_2 - \lambda_1}\int_{\lambda_1}^{\lambda_2} t(\lambda)\,d\lambda \qquad (2\text{-}39)$$

Eq. 2-39 implicitly contains Eq. 2-38 and is hard to evaluate exactly. Therefore, various approximations to Eq. 2-39 are used, depending on the wavelength region and attenuating medium considered. In certain cases, either scattering or molecular absorption is negligible, and one need only consider one attenuation process. Since the transmittance due to scattering is a slowly varying function of wavelength, it will often be justified to take that part of the transmittance due to scattering out of the integral and replace it by an average value, especially if the interval $\lambda_1 - \lambda_2$ is small.

2-4.2 MOLECULAR ABSORPTION

2-4.2.1 Absorption Coefficient

Radiation traveling through the atmosphere undergoes attenuation, defined here by absorption coefficient α, which is a function of the wavenumber $\bar{\nu}$ (or wavelength $\lambda = 1/\bar{\nu}$). The transmittance of a beam of light passing through a given amount of absorber w in the atmospheric path is given by

$$t(\bar{\nu}) = \exp\,[\,-\alpha(\bar{\nu})w\,] \qquad (2\text{-}40)$$

For gases, the units of $\alpha(\bar{\nu})$ and w are unusual. The generic term for the units of w is "absorber content", a unit of length. The simplest absorber content unit is the "atmosphere-kilometer". An atm-km is one kilometer of pathlength through the atmosphere at standard temperature and pressure with the "normal" amount of absorber present—such as O_3, CO_2, H_2O, etc. Since "normal" is difficult to ascertain, the atm-km is usually normalized to the "atmosphere-centimeter". The atm-cm is one centimeter of pathlength at standard temperature and pressure (STP) through the absorber alone. For example, "normal" atmospheres contain 3×10^{-4} parts by volume of CO_2. Thus 1 atm-km of air with the normal amount of CO_2 is equivalent to 30 atm-cm of CO_2 alone.

Water vapor is still further normalized to "precipitable-centimeters". Water vapor in units of precipitable centimeters is the thickness in centimeters of the water along the path if it were condensed to liquid. Thus

$$w(\text{pr-cm}) = (\text{pathlength in cm}) \times [\rho(H_2O\ \text{vapor})\ \text{in g cm}^{-3}] \qquad (2\text{-}41)$$

Water vapor concentration is also found in units of $g(H_2O)/kg(air)$ at STP, called mixing ratio. The conversion from g/kg to pr-cm(H_2O)/km(pathlength) is

$$\frac{g(H_2O)}{kg(air)} = 35.2\left[\frac{P(\text{atm})}{T(^\circ\text{K})}\right]\left[\frac{\text{pr-cm}\,(H_2O)}{\text{km(pathlength)}}\right] \qquad (2\text{-}42)$$

The dimensions of α are (absorber content)$^{-1}$ such as (atm-km)$^{-1}$, (atm-cm)$^{-1}$, (pr-cm)$^{-1}$. Note that these are actually units of reciprocal length.

The wavenumber dependence of the absorption coefficient α is extremely complex in the IR. IR absorption bands consist of many narrow

absorption lines each corresponding to a particular transition of the absorbing molecule from one vibration-rotation state to another. Transmission calculations will usually rely on band models which take into account the band structure parameters such as line width, spacing between lines, and line intensities.

There are three basic methods of calculating atmospheric absorption values. The first requires high-resolution spectral data or theoretical calculations of exact line positions on a high-speed computer[10,11,12]. Computers are programmed to perform high-resolution transmission calculations which can then be averaged to provide transmission curves for lower resolution. A detailed discussion of this method is contained in Ref. 10. The application of this method to H_2O and CO_2 transmission is described in Refs. 11 and 12. The second method entails estimating, from a theoretical standpoint, the

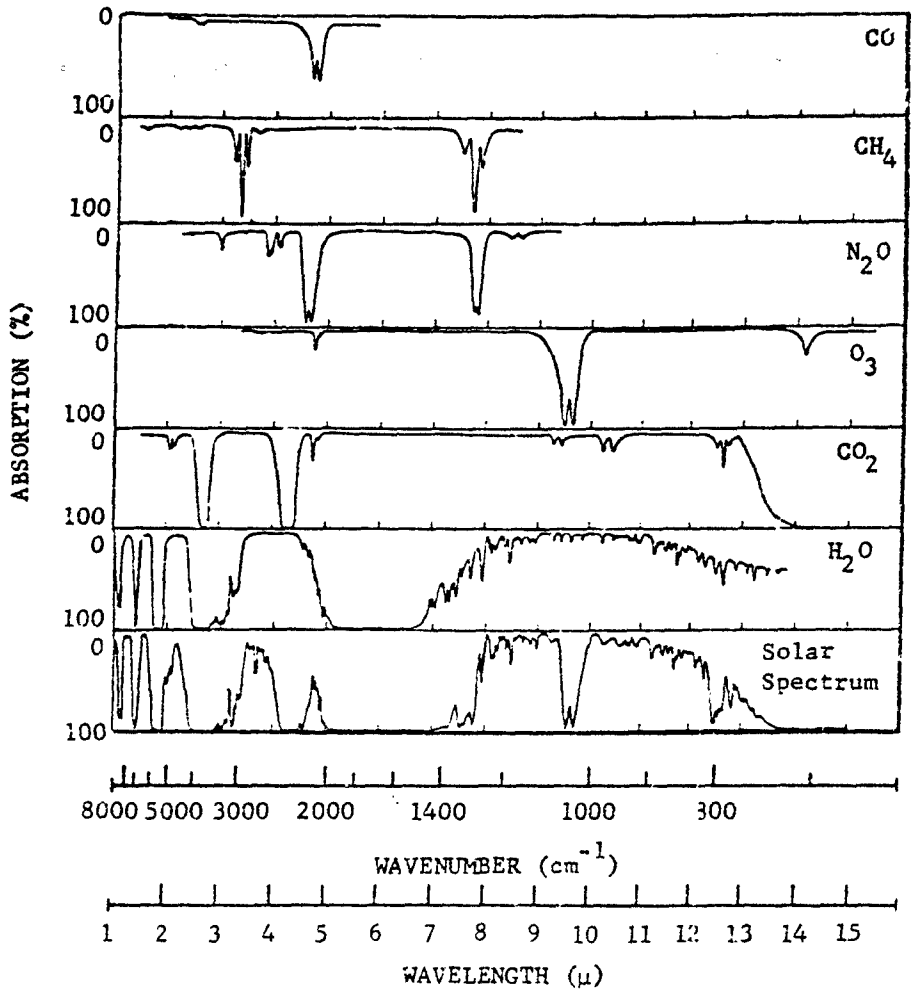


FIGURE 2-6. The Near-infrared Spectra of Solar Irradiation and of CO, CH_4, N_2O, O_3, CO_2, and H_2O

average transmission in a small wavelength interval containing many absorption lines as a function of certain band parameters[13]. The numerical values of the band structure parameters can be determined from experimental observations, and the resultant absorption effects then determined for any path length. This method is particularly useful for estimating medium-resolution ($\Delta\bar{\nu}$ X 10 to 50 cm^{-4}) transmission curves and provides the envelope of the absorption bands rather than their fine structure. The theory behind this method is discussed in par. 2-4.2.3, while the data and its application are considered in par. 2-4.2.4. The third method involves the use of formulas which are derived empirically to fit available data. It is applicable to low resolution systems for which the spectral bandpass completely encompasses one or more molecular bands. These formulas, which can be used to predict transmissions for a wide range of conditions, are also presented in par. 2-4.2.4.

2-4.2.2 Absorber Constituents

Molecular absorption in the atmosphere is primarily due to carbon dioxide (CO_2), water vapor (H_2O), and ozone (O_3). Other minor constituents which also contribute to the absorption include nitrous oxide (N_2O), carbon monoxide (CO), and methane (CH_4). Fig. 2-6 illustrates the IR transmission characteristics of the atmosphere due to these constituents. Between 14 microns and the microwave region of the spectrum, water vapor is a strong absorber and practically no transmission would be apparent in this region. Ref. 14 surveys the measurements of concentration of the minor absorbing constituents in the atmosphere.

For most problems of practical interest, atmospheric concentrations of CH_4, CO, N_2O, and CO_2 are assumed to be constant. Small variations of a few percent occur for CO_2, especially near the ground[15]. Concentrations of N_2O, CO, and CH_4 are found to be the most variable (deviations ranging from 50 to 100% are not unusual). However, because of the relatively low density of these minor constituents, the variation does not significantly affect transmission calculations. The concentration of ozone, which is variable, peaks between the altitudes of 20 and 30 km where it is produced through the photodissociation of oxygen by ultraviolet radiation[16]. Ozone diffuses and is convected downward by atmospheric turbulence and winds[17]. Upon reaching the ground, it reacts upon organic materials. Ozone can also be created near the ground due to various chemical

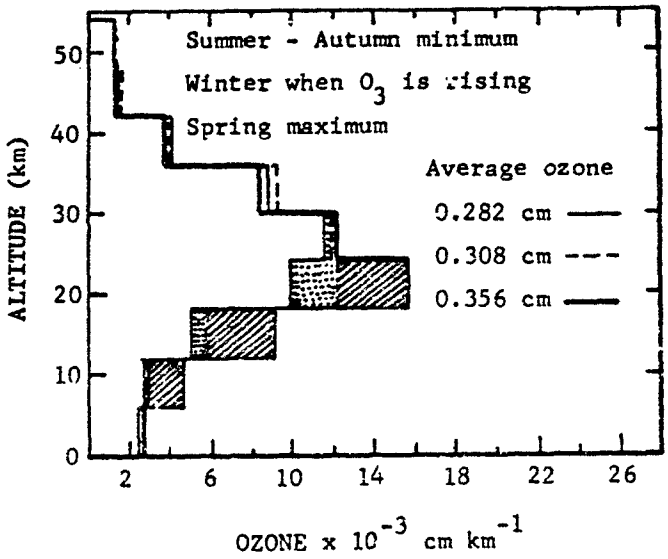


FIGURE 2-7. Seasonal Variation of the Vertical Distribution of Ozone

agents. Typical ozone profiles are shown in Fig. 2-7. Total ozone content depends on latitude and season (Fig. 2-8).

The water vapor concentration is highly variable, especially near the ground. The water vapor profile is closely related to the temperature profile of the atmosphere. These two profiles (temperature and mean H_2O) for the Gutnick[18] standard atmosphere are presented in Fig. 2-9. The third profile included in Fig. 2-9 corresponds to a saturated atmosphere. The water vapor content decreases rapidly with altitude up to the tropopause (≈ 12 km) above which both the temperature and the water vapor content cease to decrease. Good measurements of water vapor content above the tropopause are difficult to make because this content is so small. It therefore is not surprising that the measurements above the tropopause seldom agree. It is assumed that the volume mixing ratio of water vapor above the tropopause is either constant (dry stratosphere) or increases as the temperature rises (wet stratosphere).

Table 2-5 lists the mean values of absorber concentrations in the atmosphere and the absorber content along a 1 km pathlength at sea level.

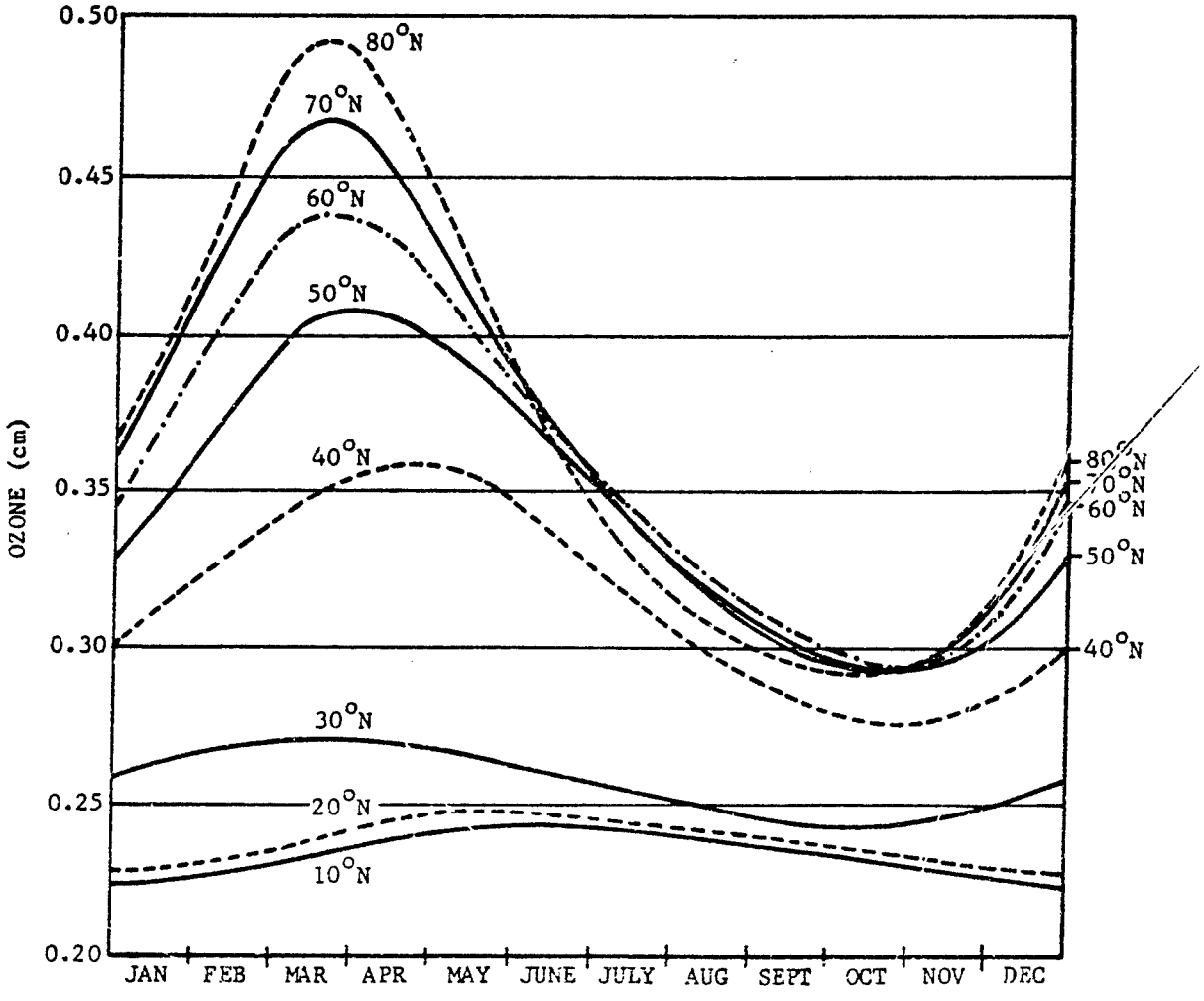


FIGURE 2-8. Annual Variation of Total Ozone for Each 10° of N. Latitude

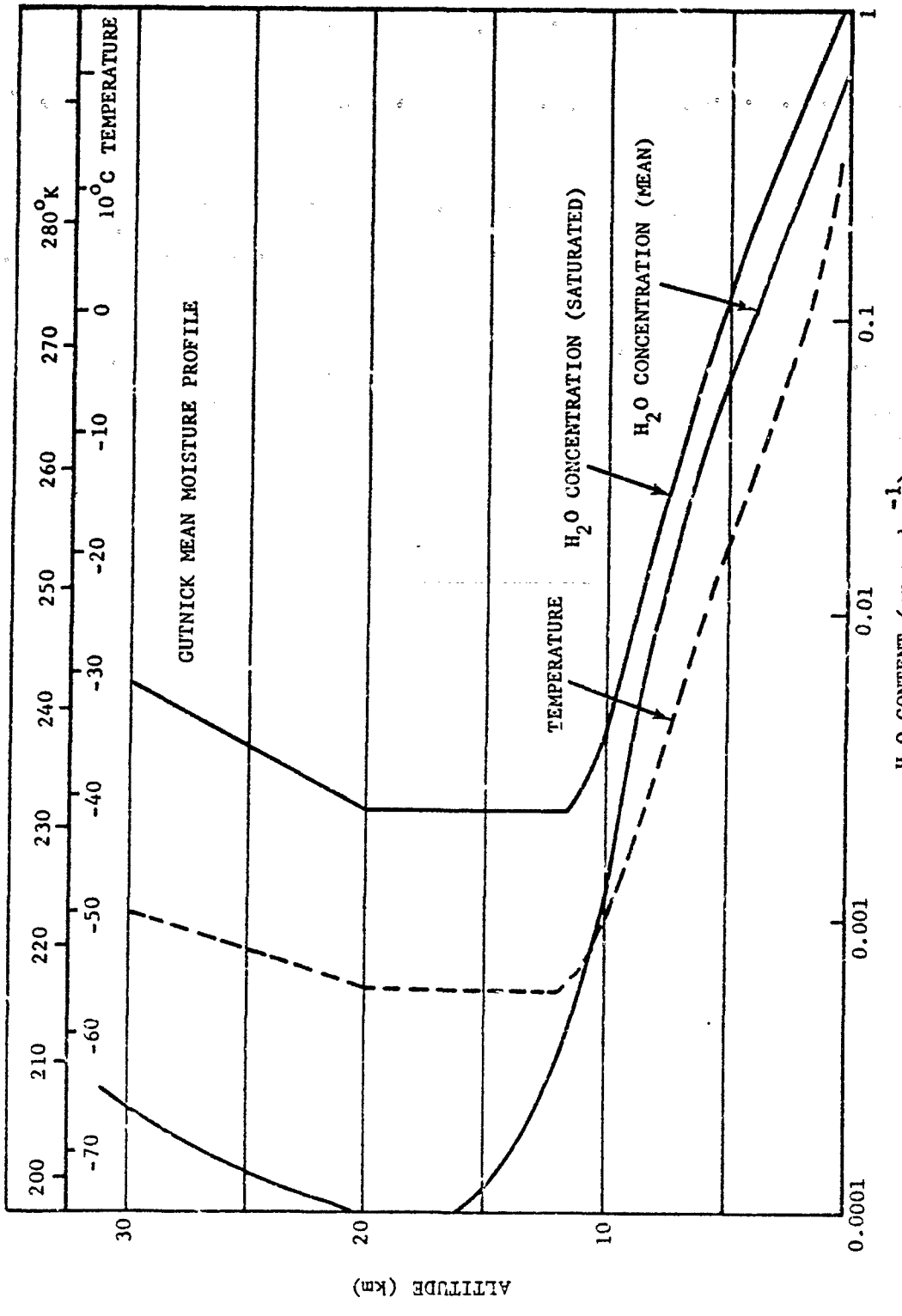


FIGURE 2-9. Water Vapor Profile in the Atmosphere

TABLE 2-5. ABSORBER CONCENTRATIONS IN THE ATMOSPHERE

REF.	CONSTITUENT	MOLECULAR MASS, G/MOLE	CONCENTRATION, % BY VOLUME	CONTENT IN 1 KM PATH AT SEA LEVEL
15	O_3	48	Variable (10^{-6})	$1 - 5 \times 10^{-3}$ atm-cm
14	CO_2	44	3.2×10^{-2}	32 atm-cm
14	H_2O	18	Variable ($1 - 10^{-3}$)	0.1 - 2 pr-cm
14	CH_4	16	1.7×10^{-4}	1.6×10^{-1} atm-cm
14	CO	28	1.2×10^{-5}	1.2×10^{-2} atm-cm
15	N_2O	44	5×10^{-5}	5×10^{-2} atm-cm

2-4.2.3 Absorption Models

The theoretical models used for the computation of medium resolution transmission ($\Delta\bar{\nu} = 10$ to $50\ cm^{-1}$) are discussed in the following paragraphs:

2-4.2.3.1 — Absorption Due to a Single Line
2-4.2.3.2 — Absorption Due to an Assembly of Independent Lines
2-4.2.3.3 — The Elsasser Band Model
2-4.2.3.4 — The Goody Model
2-4.2.3.5 — Other Less Frequently Used Models
2-4.2.3.6 — Selective Absorption of Plume Radiation

2-4.2.3.1 Absorption Due to a Single Line

In the lower atmosphere, all lines constituting an IR band are assumed to have the Lorentz line shape for which the absorption coefficient is given by

$$\alpha(\bar{\nu}) = \frac{S}{\pi}\left[\frac{\gamma}{(\bar{\nu} - \bar{\nu}_0)^2 + \gamma^2}\right] \qquad (2\text{-}43)$$

where

S = line intensity, (absorber units)$^{-1}$ - cm^{-1}
γ = half-width at half maximum, cm^{-1}
$\bar{\nu}_0$ = position of centerline, cm^{-1}

The Lorentz line shape is shown in Fig. 2-10. The integral of $\alpha(\bar{\nu})$ over all wavenumbers is equal to the line intensity S. The line half-width γ depends on the pressure P and the absolute temperature T as[13]

$$\gamma = \gamma_0\left(\frac{P}{P_0}\right)\left(\frac{T_0}{T}\right)^{1/2} \qquad (2\text{-}44)$$

For most H_2O and CO_2 lines at STP, γ_0 ranges from 0.03 to 0.15 cm^{-1}.

The Lorentz line shape, arising from molecular collisions, is valid only for low altitudes. Above 30 km, the line shape for CO_2 and H_2O become Gaussian[19]. For CO_2 it is questionable whether the wings of Eq. 2-43 are valid even at low altitudes, i.e., when $(\bar{\nu} - \bar{\nu}_0)^2 \gg \gamma$.

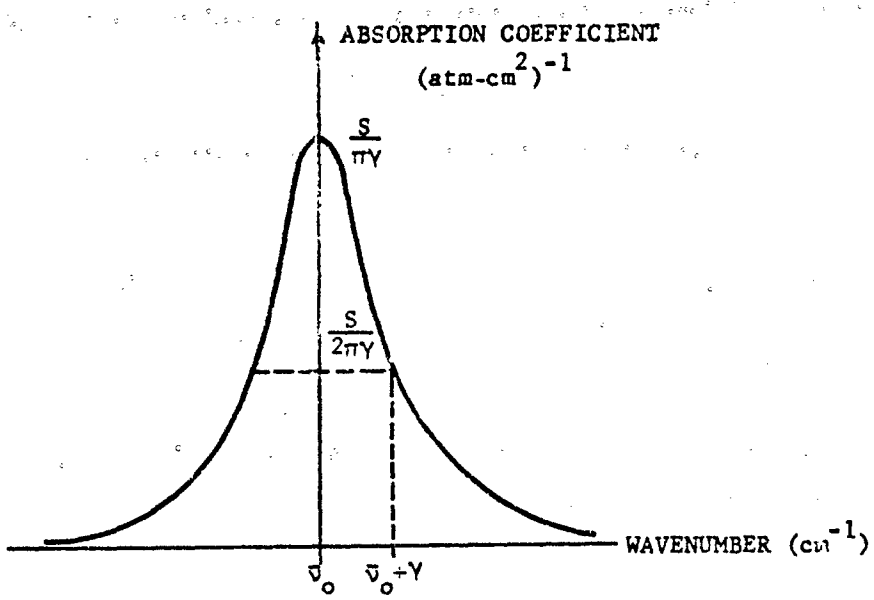


FIGURE 2-10. Lorentz Line Shape

The total absorption $\alpha\Delta\bar{\nu}$ in a band $\Delta\bar{\nu}$ centered at $\bar{\nu}_0$ —and due to a single Lorentz line viewed through a constant temperature, constant pressure pathlength containing an amount w of absorber—is

$$\alpha\Delta\bar{\nu} = \int_{\bar{\nu}_0-\frac{\Delta\bar{\nu}}{2}}^{\bar{\nu}_0+\frac{\Delta\bar{\nu}}{2}} \left\{1 - \exp\left[\frac{-Sw\gamma/\pi}{(\bar{\nu}-\bar{\nu}_0)^2+\gamma^2}\right]\right\} d\bar{\nu}, \text{ cm}^{-1} \qquad (2\text{-}45)$$

where α represents the average absorption over the interval $\Delta\bar{\nu}$. If $\Delta\bar{\nu} \gg \gamma$, then Eq. 2-45 can be approximated by

$$\alpha\Delta\bar{\nu} = \int_{-\infty}^{+\infty} \left\{1 - \exp\left[\frac{-Sw\gamma/\pi}{(\bar{\nu}-\bar{\nu}_0)^2+\gamma^2}\right]\right\} d\bar{\nu}, \text{ cm}^{-1} \qquad (2\text{-}46)$$

Although Eqs. 2-45 and 2-46 can be evaluated in terms of Bessel's functions[20], it is more usual to make weak and strong line approximations.

The weak line approximation corresponds to pathlengths and line intensities such that the exponent at the center line $Sw/\pi\gamma$ is much less than one. In that case, the exponential can be expanded in a power series. Keeping only the first two terms, the integral then easily yields

$$\alpha\Delta\bar{\nu} = Sw, \text{ cm}^{-1} \qquad (2\text{-}47)$$

Thus, for the weak line approximation, the total absorption depends linearly on pathlength w.

In the strong line approximation $Sw/\pi\gamma \gg 1$, the line saturates for several half-widths about its center. The absorption is so great that only on the wings of the line can any shape be seen. Since $(\bar{\nu}-\bar{\nu}_0)^2 \gg \gamma^2$ before the exponential becomes important, the γ^2 in the denominator of the exponent can be dropped. Eq. 2-46 is then integrated to give*

$$\alpha\Delta\bar{\nu} = 2(S\gamma w)^{1/2}, \text{ cm}^{-1} \qquad (2\text{-}48)$$

Thus, for the strong line approximation, the total absorption depends on the square root of pathlength w.

For pressures less than 1 atmosphere, and for most *IR* absorption bands of interest, the lines constituting the bands may be considered as strong under path conditions resulting in any appreciable absorptance (20% or more). The noted exception is ozone whose lines cannot be assumed to be strong even at 25 km altitude except for very long pathlengths.

2-4.2.3.2 Absorption Due to an Assembly of Independent Lines

If the total absorption of a band is due to the sum of the total absorptions of the lines in the band, then the lines are considered independent[21] . Thus independent means

$$\alpha\Delta\bar{\nu} = \sum_{i=1}^{n} (A\Delta\bar{\nu})_i = \sum_{i=1}^{n} \alpha_i \Delta\bar{\nu} \qquad (2\text{-}49)$$

Summed over all n lines in the band. The individual lines themselves may be either strong or weak and still may be independent. Thus

$$\alpha\Delta\bar{\nu} = \begin{cases} \sum_{i=1}^{n} S_i w & \text{(weak lines)} \\ \sum_{i=1}^{n} (S_i\gamma_i w)^{1/2} & \text{(strong lines)} \end{cases} \qquad (2\text{-}50)$$

If the lines have an average line strength S and an average spacing d in cm^{-1}, then the average absorptance is

$$\alpha = \begin{cases} \dfrac{wS}{d} & \text{(weak lines)} \\ \left(\dfrac{S\gamma w}{d^2}\right)^{1/2} & \text{(strong lines)} \end{cases} \qquad (2\text{-}51)$$

2-4.2.3.3 The Elsasser Band Model

The Elsasser absorption band model[22] is comprised of a series of regularly spaced, identical Lorentz lines. It is applicable to the IR bands of symmetric top molecules[23] ; i.e., CO, CO_2, N_2, CH_4. The lines comprising the IR bands of these molecules are evenly spaced. Their intensities, however, vary[20] . Hence only portions of the bands can be represented accurately by the Elsasser model. The absorption coefficient of an Elsasser band is obtained from Eq. 2-43 by summing over all the lines. The average absorptance in an interval containing many lines, in the strong line approximation $\left(\dfrac{Sw}{\pi\gamma} \gg 1\right)$ is found[22] to be

$$\alpha = erf\left(\frac{\sqrt{\pi S\gamma w}}{d}\right) \qquad (2\text{-}52)$$

where the error function (erf) is defined as

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt$$

where d is the spacing between lines. Eq. 2-52 is valid only if the overlap parameter† $\dfrac{2\pi\gamma}{d}$ is much smaller than 1. The error function is tabulated in many standard mathematical tables[24, 25]. A fit of H_2O transmission measurements to the error function is shown in Fig. 2-11. If the argument

---

* The integration is done by: (1) letting $x = \bar{\nu}-\nu_0$; (2) changing to $y = \dfrac{1}{x}$; (3) inserting an integrating factor α in exponent; (4) differentiating with respect to α; (5) integrating with respect to y; and, (6) integrating with respect to α over the range 0 to 1.

† $\dfrac{2\pi\gamma}{d}$ is referred to as the overlap parameter because it contains the ratio of the line half width γ to the (average) spacing between lines d and expresses the extent of "overlap" between lines.

of the error function is much smaller than 1, there is little overlap between lines and Eq. 2-52 reduces to

$$\alpha = 2\left(\frac{\sqrt{S\gamma w}}{d}\right) \qquad (2\text{-}53)$$

Thus, the absorptance varies proportionately with the square root of the path length. The condition $\frac{2\pi\gamma}{d} \ll 1$ holds for the IR bands of CO, CO_2, N_2O, CH_4 if the pressure is less than 1 atmosphere. The condition $\frac{Sw}{\pi\gamma} \gg 1$ will hold for any appreciable absorption ($\alpha \approx 20\%$ or greater for most bands at sea level). Therefore, Eq. 2-52 is valid in most cases of interest. When the condition $\frac{Sw}{\pi\gamma} \gg 1$ does not apply, the transmittance will be a function of the two parameters $\frac{2\pi\gamma}{d}$ and $\frac{S}{d}$ (Ref. 22). Table 2-6 lists these band parameters for CO_2 transmission. The coefficient γ_0 that appears in Table 2-6 is defined by Eq. 2-44 and corresponds to a pressure $P_0 = 1$ mm of Hg and to a temperature $T_0 = 300°K$.

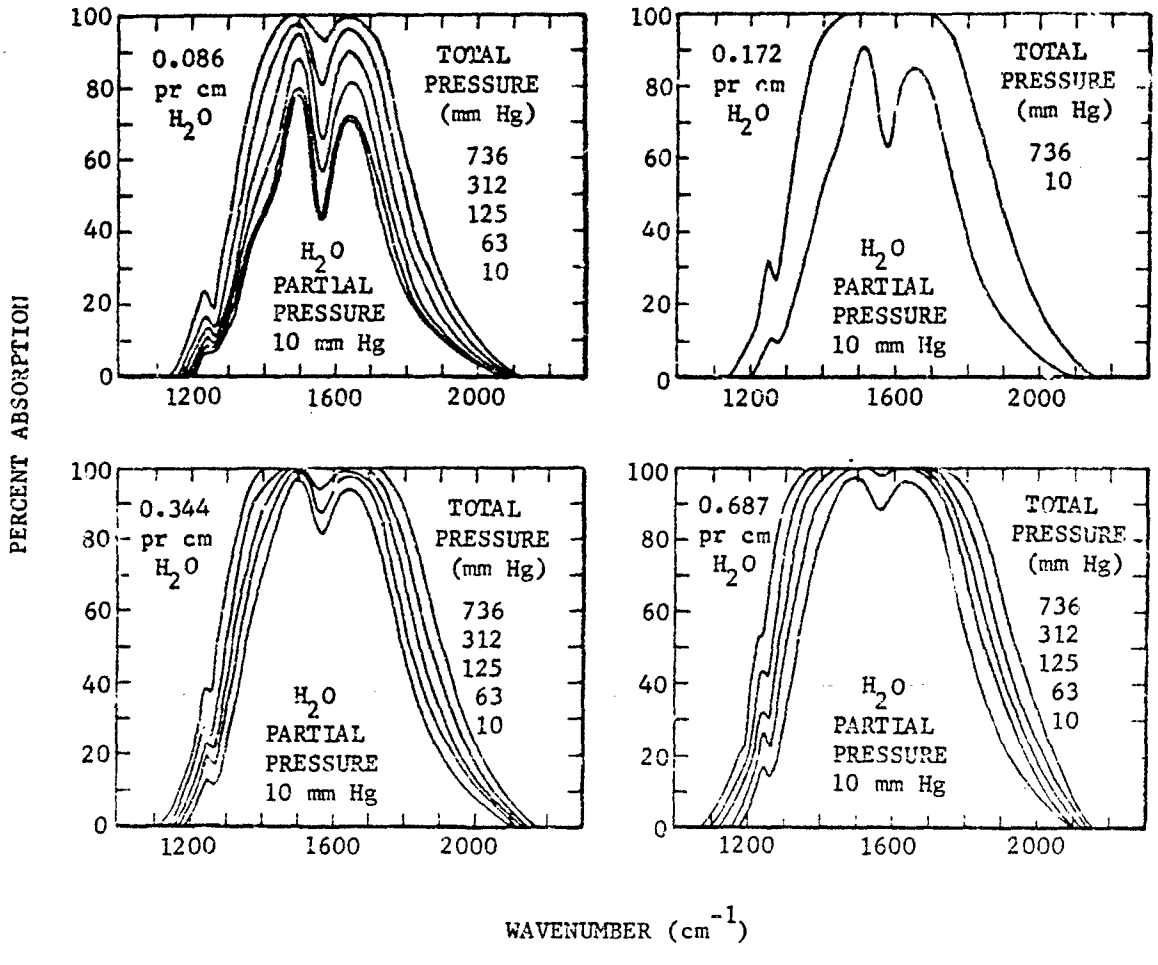


FIGURE 2-11. The 6.3-micron Band of H_2O

TABLE 2-6. BAND PARAMETERS S/d AND $2\pi\gamma_o/d$ FOR CO_2

WAVELENGTH, μ	S/d	$\frac{2\pi\gamma_o}{d}$ *
2.6434	0.3300E-03 **	0.4550E-05
2.6504	.1600E-02	.3750E-04
2.6674	.3700E+00	.1351E-03
2.6724	.5000E+00	.5000E-03
2.6738	.5000E+00	.6300E-03
2.6802	.5500E+00	.1000E-02
2.6831	.8800E+00	.6136E-03
2.6882	.8600E+00	.3256E-03
2.6911	.5400E+00	.3611E-03
2.6940	.8800E+00	.2955E-03
2.6969	.5200E+00	.6250E-03
2.7027	.7000E+00	.4929E-03
2.7086	.7400E+00	.3784E-03
2.7137	.6200E+00	.7177E-03
2.7211	.4400E+00	.8864E-04
2.7322	.1500E+00	.6667E-04
2.7397	.2500E-01	.6000E-03
2.7473	.7600E-01	.9868E-03
2.7548	.3600E+00	.7222E-03
2.7579	.5500E+00	.5455E-03
2.7609	.5800E+00	.4397E-03
2.7663	.5300E+00	.2453E-03
2.7685	.3300E+00	.2455E-03
2.7739	.2600E+00	.9231E-03
2.7778	.1500E+00	.6667E-03
2.7801	.4700E+00	.5745E-03
2.7855	.4400E+00	.3750E-03
2.7917	.2900E+00	.2414E-03
2.7941	.1600E+00	.3500E-03
2.8027	.1300E+00	.2692E-03
2.8129	.7000E-01	.1200E-03
2.8153	.3700E-01	.1568E-03
2.8177	.2700E-01	.1444E-03
2.8241	.1700E-01	.1176E-03
2.8281	.5000E-02	.3300E-03
2.8345	.3000E-02	.5167E-03
2.8433	.3900E-02	.5041E-03
2.8514	.3300E-02	.5000E-03
2.8555	.2400E-02	.5417E-03
2.8620	.1000E-02	.4500E-03
2.8686	.3000E-04	.3233E-02
2.8752	.0000E+00	.0000E+00
2.8810	.0000E+00	.0000E+00
4.1490	.0000E+00	.0000E+00
4.1580	.2600E-03	.1000E-05
4.1670	.5900E-02	.5000E-05
4.1750	.2400E+00	.8000E-04

* The values in this column correspond to a pressure of 1 mm Hg. For sea level conditions the values in this column should be multiplied by 760.

** 0.3300E-03 = 0.3300×10^{-3}.

TABLE 2-8 (Continued)

WAVELENGTH, μ	S/d	$\frac{2\pi\gamma_0}{d}$ *
4.1840	0.9940E-01	0.1815E-02
4.1930	.9800E+00	.7368E-03
4.2020	.5046E+01	.7632E-03
4.2110	.1464E+02	.8026E-03
4.2190	.2728E+02	.1158E-02
4.2280	.3363E+02	.1500E-02
4.2370	.3363E+02	.1553E-02
4.2460	.2562E+02	.1605E-02
4.2550	.1344E+02	.1684E-02
4.2640	.2316E+02	.1576E-02
4.2740	.2592E+02	.1421E-02
4.2830	.2472E+02	.1355E-02
4.2920	.2070E+02	.1211E-02
4.3010	.1698E+02	.1435E-02
4.3100	.1219E+02	.1263E-02
4.3190	.7300E+01	.1316E-02
4.3290	.5000E+01	.1316E-02
4.3380	.2915E+01	.1448E-02
4.3480	.2020E+01	.1342E-02
4.3570	.1298E+01	.1448E-02
4.3670	.6860E+00	.1290E-02
4.3760	.3762E+00	.1302E-02
4.3860	.4599E+00	.8289E-03
4.3960	.6090E+00	.5527E-03
4.4050	.5890E+00	.5000E-03
4.4150	.5610E+00	.4342E-03
4.4250	.3833E+00	.4800E-03
4.4350	.2496E+00	.6851E-03
4.4440	.1855E+00	.9218E-03
4.4540	.8624E-01	.1013E-02

* The values in this column correspond to a pressure of 1 mm Hg. For sea level conditions the values must be divided by 760.

2-4.2.3.4 The Goody Model

The statistical, or Goody, absorption band model[26, 27] is best suited for asymmetric top molecules[23], such as ozone and water vapor, whose line positions appear randomly distributed within a band, and whose line intensities vary widely. This model assumes that both absorption line positions and intensities must be specified by probability functions. The absorption in an interval $\Delta\bar{\nu}$ is then calculated by a statistical averaging process. The selected interval $\Delta\bar{\nu}$ must be sufficiently large to give validity to statistical averaging. However, since the statistical properties of an absorption band will usually vary from one absorption region of the band to another, the $\Delta\bar{\nu}$ interval selected must not be too large. It is assumed that there is equal probability of finding a line centered at any wavelength inside the interval $\Delta\bar{\nu}$; that all lines have the same half width γ; and that the normalized probability of a line having intensity S is $p(S)$.

If the line intensities have an exponential distribution with average value S_0, then $p(S) = \frac{1}{S_0} \exp -\left[\frac{S}{S_0}\right]$ and the average absorptance in the interval $\Delta\bar{\nu}$ is found to be[26]

$$\alpha = 1 - \exp\left[-\frac{wS_0}{d\left(1+\frac{wS_0}{\pi\gamma}\right)^{1/2}}\right] \qquad (2\text{-}54)$$

where d is the average space between lines.

If the line intensities are all equal,

$$\alpha = 1 - \exp\left\{-\frac{1}{d}\int_{-\infty}^{+\infty}\left[1-\exp\left[-\frac{S_0 wd}{\pi(\bar{\nu}^2+\gamma^2)}\right]\right]d\bar{\nu}\right\} \qquad (2\text{-}55)$$

When the lines saturate at their centers, Eqs. 2-54 and 2-55 reduce to

$$\alpha = 1 - \exp\left[-\frac{2}{d}\sqrt{S_0\gamma w}\right] \qquad (2\text{-}56)$$

When there is little overlap between lines,

$$\sqrt{\frac{S_0\gamma w}{d^2}} \ll 1$$

and Eq. 2-56 is reduced to Eq. 2-53.

The band parameters relevant to the Goody model Eqs. 2-54 and 2-56 are $\frac{S_0}{d}$ and $\frac{S_0^*}{2\pi\gamma}$. These parameters for H_2O and O_3 are tabulated in Ref. 14. These values for H_2O are reproduced in Table 2-7 (for sea level conditions). The coefficient γ_0 that appears in Table 2-7 is defined by Eq. 2-44 and corresponds to a pressure P_0 = 1mm Hg and to a temperature T_0 = 300°K.

2-4.2.3.5 Other Less Frequently Used Models

Other band models developed are variations of those mentioned above and do not differ sufficiently to justify a detailed discussion here.

The random Elsasser model is a model in which several Elsasser bands with different line spacings and intensities overlap in a given wavelength interval.

In the doublet model, one considers the absorption due to two Lorentz lines of the same intensity which overlap.

The Curtis model is a variation on the Elsasser band model in which the line intensities are not equal.

Another model takes into account occasional gaps which may occur between lines in a band and which have a strong influence on the transmission behavior of the band for very long paths. In this model, a gap of varying width between two random arrays of lines is considered and the transmission characteristics are related to the gap width.

The quasi random model is the most complicated in that it takes into account the fact that lines are neither uniformly spaced nor completely randomly spaced. A detailed knowledge of the band structure is required and the aid of an electronic computer is essential[10].

Most of the band models are discussed in Ref. 13.

2-4.2.3.6 Selective Absorption of Plume Radiation

The previous discussion applies only to the absorption of graybody radiation. In some applications, however, the emitter (target) is a hot gas (i.e., plume or flame radiation), and the emission spectrum is composed of many narrow emission lines. If the emitting gas is the same as one of the minor constituents of the atmosphere (CO_2 or H_2O, for example), the target radiation will be subjected to selective absorption by the atmosphere. The result will be higher average absorptance for this radiation than for blackbody radiation. If the emitter is sufficiently hot ($T > 600°K$ for CO_2 and $T > 1200°K$ for H_2O), or thick enough so that its emissance approaches one; then (1) this selective absorption of the atmosphere will be small[28]; (2) the emitting gas may be assumed to radiate as a graybody insofar as transmission calculations are concerned; and (3) the results of the previous paragraphs may then be applied.

2-4.2.4 Absorption Calculation

In order to apply the theoretical analysis of par. 2-4.2.3 to the actual calculation of molecular absorption in the atmosphere, two more points need examination: (1) spectral location and bandwidth of detection system, and (2) conversion of actual paths through the atmosphere to equivalent sea-level paths.

The ideal situation would be that in which the spectral resolution of the detection system is the same as that of the laboratory data available. The direct use of laboratory data is then possible. But this is rarely the case. In general, one must refer to laboratory data of higher resolution than that of the detection system. If the bandwidth of the system is larger than 50 cm^{-1}, the band models discussed in par. 2-4.2.3 can be used to calculate the (average) transmittance at each wavenumber inside the bandwidth of the detection system and the average transmittance over the band can be derived by averaging. If the spectral bandwidth of the detecting system is so large that it completely encompasses one or more absorption bands, transmission calculations can be simplified by calculating the total absorptions of the bands rather than the absorptances at each wavelength.

The narrow band absorption data to which band models can be applied are discussed in par. 2-4.2.4.1. The use of wideband absorption data for the determination of total band absorption is presented in par. 2-4.2.4.2. Justification for the use of equivalent sea-level paths and their method of calculation are given in par. 2-4.2.4.4. Some sample problems are worked out in par. 2-4.2.4.4.

* Note that the overlap parameter $\frac{2\pi\gamma}{d}$ is the ratio of these two parameters.

Extensive tabulations of CO_2, and H_2O transmittances for various atmospheric paths and absorber contents have been compiled[11,12,29].

2-4.2.4.1 Narrow-band Absorption Data

Narrow-band absorption is that for which the spectral resolution of the measuring instrument is 10 to 50 cm^{-1}. This resolution is sufficiently narrow such that only a small portion of an absorption band is measured, but sufficiently wide such that the theoretical considerations of par. 2-4.2.3 (Absorption Models) apply. A great amount of laboratory data on absorption by CO_2[29], H_2O[30], and by other minor atmospheric constituents[29,31] except ozone are available. Unfortunately, only little laboratory data on ozone are available[32,33]. Other data for the atmosphere as a whole[34,35] are available for sea-level paths of various lengths.

Many attempts have been made since the publication of laboratory data[29,30,31] to fit the data to the band models discussed here. Tables 2-8 and 2-9 give a summary of the data and references to investigators who reduced the data.

The band parameters derived for CO_2 and H_2O from the experiments in Refs. 29 to 31 are listed in Tables 2-6 and 2-7. From these parameters, it is possible to calculate the coefficients entering in Eqs. 2-52 (CO_2) and 2-54 (H_2O), and to compute the transmission at any wavelength if the amount of absorber w is known.

An example of the data available on the 6.3-micron vapor band, taken from Ref. 30, is shown in Fig. 2-11. The curves in the figure represent absorption across the band for a single water-vapor path-length as a function of total pressure. Table 2-9 from Ref. 14 lists the absorption bands and absorption data source references.

An example of the use of these data for finding the transmission curve is given in Fig. 2-12[38]. (Note that units on the abscissa of Fig. 2-12 are expressed in "atmospheric kilometers" rather than "atmospheric centimeters" as expressed in Table 2-3.) The experimental data on CO_2 at 2.7 microns in Fig. 2-12 is fitted to an error function curve.

Ref. 14 contains a description of a computer program which can be used for calculating transmissions between 1 and 20 microns.

2-4.2.4.2 Wide-band Absorption Data

Wide-band absorption is that which is measured by an instrument with a spectral bandwidth completely encompassing an absorption band. For example, if all lines of a given band are between the wavelength limits of 1.75 and 2.00 microns, the wide-band absorption would be measured by an instrument with a spectral passband between 1.5 and 2.5 microns. To avoid problems with instrument passband width, which for the above-mentioned band could extend in width from 0.25 micron to infinity, wide-band absorption is measured in absolute units, microns. An absorption of 0.1 micron would mean that the integrated absorption in the band would be equivalent to total absorption in a 0.1-micron interval. Thus, for the 1.75- to 2.00-micron band above, total absorption can range from 0 to 0.25 micron. The wide-band absorption data for the various bands of carbon dioxide and water vapor are listed in Tables 2-10 and 2-11. These tables list the effective pressure p_e as total pressure plus the absorber partial pressure. For atmosphere transmission problems, the absorber partial pressure can be neglected.

2-4.2.4.3 Atmospheric Windows

As illustrated in Fig. 2-6, attenuation by the atmosphere is weak in the following window regions:

Microns
1 – 1.1
1.2 – 1.3
1.5 – 1.8
2.1 – 2.4
3.5 – 4
4.5 – 5
8 – 14

TABLE 2-7. BAND MODEL PARAMETERS S/d AND $S/(2\pi\gamma_0)$ FOR H_2O

WAVELENGTH, μ	S/d	$S/(2\pi\gamma_0)$*
2.16000	0.7030E-03**	0.8675E+00
2.18000	.9430E-03	.1157E+01
2.20000	.1680E-02	.2078E+01
2.24000	.7590E-03	.9356E+00
2.28000	.3070E-01	.3791E+02
2.33100	.5650E-01	.6790E+02
2.33645	.5660E-01	.6800E+02
2.34192	.7930E-01	.8210E+02
2.34742	.7980E-01	.7880E+02
2.35294	.8390E-01	.8240E+02
2.35849	.8820E-01	.8640E+02
2.36407	.9310E-01	.8670E+02
2.36967	.9890E-01	.9270E+02
2.37530	.1040E+00	.9910E+02
2.38095	.1090E+00	.9920E+02
2.38663	.1150E+00	.1030E+03
2.39234	.1210E+00	.1050E+03
2.39808	.1250E+00	.1050E+03
2.40385	.9580E-01	.5260E+02
2.40964	.1370E+00	.9790E+02
2.41546	.1500E+00	.1030E+03
2.42131	.1620E+00	.9810E+02
2.42718	.1860E+00	.1060E+03
2.43309	.2180E+00	.1320E+03
2.43902	.2490E+00	.1690E+03
2.44499	.2910E+00	.1760E+03
2.45098	.3970E+00	.2490E+03
2.45700	.6460E+00	.4390E+03
2.46305	.7490E+00	.4850E+03
2.46970	.4380E+01	.5110E+05
2.47400	.2530E+01	.2320E+05
2.48020	.6650E+01	.2110E+05
2.48320	.5280E+01	.3920E+05
2.48760	.5630E+01	.3140E+05
2.49380	.1000E+02	.2150E+05
2.50130	.9860E+01	.1670E+05
2 50500	.1030E+02	.5090E+05
2.50880	.1590E+02	.3630E+05
2.51180	.1390E+02	.3070E+05
2.51890	.2600E+02	.2480E+05
2.52400	.3450E+02	.1730E+06
2.52840	.5160E+02	.9290E+05
2.53610	.9200E+02	.1150E+06
2.53940	.7220E+02	.1030E+06
2.54450	.1440E+03	.1520E+06
2.54970	.9060E+02	.5650E+06
2.55620	.1120E+03	.2700E+06

* The values in this column correspond to a pressure of 1 mm Hg.
For sea level conditions the values in this column should be multiplied by 760.
** 0.7030E-03 = 0.7030×10^{-3}.

TABLE 2-7 (Continued)

WAVELENGTH, μ	S/d	$S/(2\pi\gamma_0)$*
2.56280	0.1410E+04	0.6690E+07
2.56610	.4570E+03	.8500E+06
2.57070	.2590E+03	.1310E+06
2.57730	.2320E+03	.1150E+06
2.58400	.8130E+03	.1300E+07
2.58930	.5980E+03	.7880E+06
2.59200	.1020E+04	.1310E+07
2.59740	.2880E+03	.1310E+06
2.60210	.4950E+03	.2020E+06
2.60890	.2940E+03	.1270E+06
2.61440	.4070E+03	.1360E+06
2.62190	.2510E+03	.1740E+06
2.62880	.2380E+03	.6220E+05
2.63370	.9650E+02	.4900E+05
2.63990	.1260E+03	.5670E+05
2.64340	.1130E+03	.7030E+05
2.65040	.1930E+03	.7880E+05
2.66740	.6180E+03	.1530E+06
2.67240	.4890E+03	.1190E+06
2.67380	.6090E+03	.1850E+06
2.68020	.3050E+03	.8040E+05
2.68310	.3090E+03	.8640E+05
2.68820	.2150E+03	.7070E+05
2.69110	.2590E+03	.7980E+05
2.69400	.3070E+03	.8880E+05
2.69690	.1470E+03	.4340E+05
2.70270	.1310E+03	.4830E+05
2.70860	.2490E+03	.3050E+05
2.71370	.2130E+03	.1200E+06
2.72110	.5790E+03	.3090E+06
2.73220	.1240E+03	.7180E+05
2.73970	.4950E+03	.2020E+06
2.74730	.1390E+03	.6150E+05
2.75480	.2780E+03	.6890E+05
2.75790	.2730E+03	.1100E+06
2.76090	.2660E+03	.1520E+06
2.76630	.2940E+03	.6690E+05
2.76850	.2500E+03	.6640E+05
2.77390	.1710E+03	.6460E+05
2.77780	.1900E+03	.7200E+05
2.78010	.1800E+03	.7630E+05
2.78550	.2300E+03	.8460E+05
2.79170	.1130E+03	.6270E+05
2.79410	.8740E+02	.4970E+05
2.80270	.2550E+03	.1320E+06
2.81290	.1000E+03	.1150E+06
2.81530	.1050E+03	.9170E+05
2.81770	.1360E+03	.9230E+05

* The values in this column correspond to a pressure of 1 mm Hg. For sea level conditions the values must be divided by 760.

TABLE 2-7 (Continued)

WAVELENGTH, μ	S/d	$S/(2\pi\gamma_0)$*
2.82410	0.8170E+02	0.1300E+06
2.82810	.6120E+02	.9380E+05
2.83450	.1070E+03	.6000E+05
2.84330	.3740E+02	.6240E+05
2.85140	.6000E+02	.5740E+05
2.85550	.8030E+02	.6880E+05
2.86200	.3870E+02	.3490E+05
2.86860	.2830E+03	.5200E+05
2.87520	.1500E+02	.1800E+05
2.87770	.1490E+02	.1600E+05
2.88100	.1500E+02	.1780E+05
2.88600	.1540E+02	.7930E+04
2.89440	.5910E+01	.8930E+04
2.90280	.3890E+02	.5700E+05
2.91040	.8030E+01	.1130E+05
2.91460	.8730E+05	.1070E+05
2.91630	.8160E+01	.1550E+05
2.92400	.2600E+02	.2480E+05
2.93080	.4140E+01	.4920E+04
2.93260	.3810E+01	.9710E+04
2.94120	.7170E+01	.6610E+04
2.94810	.1560E+02	.3020E+05
2.95070	.1210E+02	.5430E+05
2.95510	.1160E+02	.1720E+05
2.96120	.2320E+01	.2470E+04
2.97180	.1570E+02	.1240E+05
2.97620	.1340E+02	.2130E+05
2.98950	.2050E+01	.6270E+04
2.99400	.8180E+01	.1630E+05
3.00300	.6850E+01	.3000E+05
3.00750	.8010E+01	.8370E+05
3.01390	.4740E+01	.1500E+05
3.02110	.1120E+02	.9920E+04
3.02480	.7720E+01	.5220E+05
3.03030	.8140E+01	.1490E+05
3.03580	.9810E+01	.6650E+04
3.03950	.8430E+01	.2450E+05
3.04410	.1010E+02	.2390E+05
3.04880	.1670E+02	.1770E+05
3.05340	.7600E+01	.3210E+05
3.06000	.8280E+01	.1940E+05
3.06750	.9680E+01	.1190E+05
3.07220	.7190E+01	.4480E+05
3.07790	.1280E+02	.2190E+05
3.08360	.7030E+01	.1050E+05
3.08830	.8860E+01	.7010E+04
3.09410	.8630E+01	.3130E+05
3.09890	.6250E+01	.4900E+04

* The values in this column correspond to a pressure of 1 mm Hg. For sea level conditions the values must be divided by 760.

TABLE 2-7 (Continued)

WAVELENGTH, μ	S/d	$S/(2\pi\gamma_o)$*
3.10750	0.9860E+05	0.1660E+05
3.11520	.6430E+01	.1150E+06
3.12110	.9780E+01	.1440E+05
3.12500	.8120E+01	.9550E+04
3.12990	.2200E+01	.2480E+04
3.13480	.1420E+01	.1470E+04
3.13970	.7670E+01	.9380E+04
3.14470	.7100E+01	.7250E+04
3.14960	.4280E+01	.3540E+04
3.15460	.9240E+00	.8140E+03
3.15960	.4570E+00	.2300E+03
3.16460	.2700E+00	.1260E+03
3.16960	.2290E+00	.1020E+03
3.17460	.4260E+00	.3300E+03
3.17970	.1300E+01	.1310E+04
3.18470	.1560E+01	.1280E+04
3.18980	.7340E+01	.8010E+04
3.19490	.1130E+02	.1380E+05
3.20000	.1100E+02	.7690E+04
3.20510	.1000E+02	.8070E+04
3.21030	.1290E+02	.1100E+05
3.21540	.1110E+02	.1000E+05
3.22060	.1290E+02	.1070E+05
3.22580	.1450E+02	.7960E+04
3.23100	.1110E+02	.6420E+04
3.23620	.8780E+00	.5260E+03
3.24150	.4870E+01	.8410E+04
3.24680	.1020E+02	.1400E+05
3.25200	.6470E+00	.5070E+03
3.25730	.4350E+01	.9140E+04
3.26260	.1500E+02	.4610E+05
3.26800	.3330E+01	.4670E+04
3.27330	.6400E+00	.4150E+03
3.27870	.1210E+01	.1070E+04
3.28400	.6470E+00	.5820E+03
3.28950	.1940E+01	.3610E+04
3.29490	.1480E+02	.1590E+05
3.30030	.8470E+01	.8880E+04
3.30580	.5160E+01	.4640E+04
3.31130	.3830E+01	.4090E+04
3.31670	.2010E+01	.3860E+04
3.32230	.1370E+01	.1560E+04
3.32780	.1710E+00	.9100E+02
3.33330	.1090E+00	.6100E+02
3.33890	.4560E+00	.5050E+03
3.34450	.1870E+01	.2780E+04
3.35000	.2880E+00	.1670E+03
3.35570	.5690E+00	.5050E+03

* The values in this column correspond to a pressure of 1 mm Hg. For sea level conditions the values must be divided by 730.

TABLE 2-7 (Continued)

WAVELENGTH, μ	S/d	$S/(2\pi\gamma_0)$*
3.36130	0.5300E+01	0.9430E+04
3.36700	.1120E+01	.1150E+04
3.37270	.1130E+01	.1090E+04
3.37840	.2020E+00	.1310E+03
3.38410	.8970E+00	.1220E+04
3.38980	.1790E+01	.2220E+04
3.39560	.1730E+01	.2150E+04
3.40140	.2480E+00	.1640E+03
3.40720	.1680E+00	.1030E+03
3.41300	.8310E+00	.1050E+04
3.41880	.7170E+00	.8240E+02
3.42470	.3320E+00	.2800E+03
3.43050	.2030E+00	.1330E+03
3.43640	.3060E+00	.1980E+03
3.44230	.5290E+00	.5510E+03
3.44830	.2090E+00	.1590E+03
3.45420	.1720E+00	.9160E+02
3.46020	.1130E+00	.5260E+02
3.46620	.1110E+00	.8360E+02
3.47220	.2280E+00	.1600E+03
3.47830	.1190E+00	.5210E+02
3.48430	.5310E-01	.3610E+02
3.49040	.8090E-01	.4460E+02
3.49650	.1130E+00	.7350E+02
3.50260	.9140E-01	.7050E+02
3.50880	.8840E-01	.3830E+02
3.51490	.8620E-01	.3520E+02
3.52110	.1290E+00	.4770E+02
3.52730	.8300E-01	.5230E+02
3.53360	.8390E-01	.5400E+02
3.53980	.8380E-01	.5270E+02
3.54610	.5250E-01	.3780E+02
3.55240	.1490E+00	.6220E+02
3.55870	.7630E-01	.4830E+02
3.56510	.4390E-01	.2310E+02
3.57140	.1550E+00	.5310E+02
3.57780	.1300E+00	.7100E+02
3.58420	.1140E+00	.7830E+02
3.59070	.1900E+00	.1270E+03
3.59710	.3280E-01	.3700E+02
3.60360	.4260E-01	.4600E+02
3.61010	.7680E-01	.4670E+02
3.61660	.3830E-01	.2620E+02
3.62320	.3470E-01	.4360E+02
3.62980	.5920E-01	.5410E+02
3.63640	.4990E-01	.4060E+02
3.64300	.3820E-01	.2830E+02
3.64960	.5970E-01	.3140E+02
3.65630	.4700E-01	.3650E+02

* The values in this column correspond to a pressure of 1 mm Hg. For sea level conditions the values must be divided by 760.

TABLE 2-7 (Continued)

WAVELENGTH, μ	S/d	$S/(2\pi\gamma_0)$*
3.66300	0.5570E-01	0.6290E+02
3.66970	.3780E+00	.2020E+03
3.67650	.3920E+00	.1070E+03
3.68320	.1420E+00	.4180E+02
3.69000	.8360E-01	.2240E+02
3.69690	.5330E-01	.1730E+02
3.70370	.3540E-01	.2590E+02
3.71060	.5580E-01	.4030E+02
3.71750	.3530E-01	.3110E+02
3.72440	.2320E-01	.3740E+02
3.73140	.6800E-01	.5990E+02
3.73830	.5100E-01	.4260E+02
3.74530	.5890E-01	.1230E+03
3.75230	.6220E-01	.5760E+02
3.75940	.5880E-01	.4920E+02
3.76650	.2320E-01	.3890E+02
3.77360	.4420E-01	.4390E+02
3.78070	.7120E-01	.5250E+02
3.78790	.2710E-01	.5420E+02
3.79510	.2290E-01	.2650E+02
3.80230	.5340E-01	.4740E+02
3.80950	.4120E-01	.4130E+02
3.81680	.2190E-01	.2760E+02
3.82410	.3480E-01	.3410E+02
3.83140	.4780E-01	.3550E+02
3.83880	.2030E-01	.2970E+02
3.84620	.2070E-01	.2980E+02
3.85360	.3790E-01	.3000E+02
3.86100	.2900E-01	.2920E+02
3.86850	.1940E-01	.3000E+02
3.87600	.2380E-01	.3280E+02
3.88350	.3180E-01	.2620E+02
3.89110	.2400E-01	.3350E+02
3.89860	.1940E-01	.4070E+02
3.90630	.2220E-01	.2830E+02
3.91390	.2270E-01	.2450E+02
3.92160	.2700E-01	.3510E+02
3.92930	.1990E-01	.5500E+02
3.93700	.2120E-01	.2770E+02
3.94480	.2010E-01	.3180E+02
3.95260	.2230E-01	.5290E+02
3.96040	.1960E-01	.3400E+02
3.96830	.2120E-01	.3420E+02
3.97610	.1940E-01	.3870E+02
3.98410	.2340E-01	.5750E+02
3.99200	.2180E-01	.3870E+02
4.00000	.2440E-01	.3930E+02
4.00800	.2110E-01	.4200E+02
4.01600	.2000E-01	.4120E+02

* The values in this column correspond to a pressure of 1 mm Hg. For sea level conditions the values must be divided by 760.

TABLE 2-7 (Continued)

WAVELENGTH, μ	S/d	$S/(2\pi\gamma_0)$*
4.02400	0.2090E-01	0.4160E+02
4.03200	.3400E-01	.6320E+02
4.04000	.2570E-01	.4680E+02
4.04900	.2150E-01	.4190E+02
4.05700	.2260E-01	.4090E+02
4.06500	.2200E-01	.4000E+02
4.07300	.4970E-01	.7810E+02
4.08200	.3280E-01	.5280E+02
4.08900	.2200E-01	.3910E+02
4.09800	.2610E-01	.3880E+02
4.10700	.4970E-01	.7320E+02
4.11500	.9700E-01	.1660E+03
4.12400	.2490E-01	.3670E+02
4.13200	.2340E-01	.4160E+02
4.14100	.3190E-01	.4770E+02
4.14900	.2740E-01	.3870E+02
4.15900	.2380E-01	.3770E+02
4.16700	.2560E-01	.4810E+02
4.17500	.3550E-01	.5820E+02
4.18400	.4990E-01	.6110E+02
4.19300	.2750E-01	.3910E+02
4.20200	.2800E-01	.3570E+02
4.21100	.2530E-01	.4050E+02
4.21900	.5360E-01	.6970E+02
4.22800	.1170E+00	.2070E+03
4.23700	.3480E-01	.3700E+02
4.24600	.3360E-01	.3860E+02
4.25500	.2980E-01	.3360E+02
4.26400	.3190E-01	.3490E+02

* The values in this column correspond to a pressure of 1 mm Hg. For sea level conditions the values must be divided by 760.

TABLE 2-8. SUMMARY OF BAND-MODEL METHODS FOR COMPUTING ATMOSPHERIC ABSORPTION[14]

AUTHOR	GAS	BAND MODEL	WAVELENGTH RANGE, μ	EXPERIMENTAL DATA
Bradford	CO_2	Elsasser	4.184 - 4.454	Bradford, McCormack, and Selby (Ref. 36)
Thomson and Downing	CO_2	Elsasser	4.0 - 4.5	Theoretical
Elsasser	CO_2	Empirical	11.77 - 18.2	Cloud, Howard, et al. (Ref. 30)
	H_2O	Statistical	4.39 - 250	Palmer, Howard, et al.
	O_3	Empirical	8.85 - 16.4	Summerfield, Walshaw (Ref. 32)
Oppel	CO_2	Elsasser	2.64 - 2.88	Burch, et al. (Refs. 29, 31)
			4.15 - 4.52	
	H_2O	Statistical	1.08 - 4.0	Burch, et al.; Howard, et al.;
			4.0 - 5.0	Stull, Wyatt, and Plass
	N_2O	Elsasser	4.36 - 4.74	Burch, et al.
	CH_4	Statistical	3.02 - 4.16	Burch, et al.
	CO	Statistical	4.4 - 5.0	Burch, et al.
Altshuler	CO_2	Empirical	1.0 - 15.0	Howard, et al.
	H_2O	Statistical	1.0 - 9.0	Howard, et al.
			9.1 - 20.0	Taylor and Yates (Refs. 34, 35)
			20 - 40	Palmer, et al.
			40 - 250	Stanevich
	O_3	Elsasser	4.4 - 16.5	Strong (Ref. 33)
	N_2O	Elsasser	3.8 - 18.0	Shaw
Zachor	CO_2	(a) Elsasser	1.0 - 5.3	Howard, et al.
		(b) Two overlapping Elsasser bands	1.0 - 5.3	

TABLE 2-8. (Continued)

AUTHOR	GAS	BAND MODEL	WAVELENGTH RANGE, μ	EXPERIMENTAL DATA
	O_3	Elsasser	9.3 - 10.2	Walshaw, et al.
Howard, Burch, and Williams	H_2O	Statistical	1.0 - 10.6	Howard, et al.
Carpenter	CO_2	Elsasser	4.167 - 4.51	Theoretical
Lindquist	H_2O	Statistical	1.0 - 9.2	Howard, et al.
Green and Griggs	CO_2	Statistical	1.0 - 10.0	Howard, et al.
	H_2O	Statistical	1.0 - 10.0	Burch, et al. Howard, et al.
	O_3	Statistical	9.3 - 10.2	Burch, et al. Walshaw, et al.
	N_2O	Statistical	4.4 - 4.6	Burch, et al.
	CH_4	Statistical	3.15 - 3.45	Burch, et al.
	CO	Statistical	4.15 - 4.85	Burch, et al.
Stull, Wyatt, and Plass	CO_2	Quasi-random	1.0 - 20.0	Theoretical*
	H_2O	Quasi-random	1.0 - 20.0	Theoretical*
Elder and Strong	H_2O	Empirical	0.7 - 5.9**	See appropriate paragraph
Stauffer and Walsh	H_2O	Empirical	14 - 20	Stauffer and Walsh[35A]

* The results of Stull, Wyatt, and Plass were normalized against the laboratory data of Howard, et al. for each absorption band.
** The results of Elder and Strong give average window transmission for seven windows in the wavelength range.

TABLE 2-9. SUMMARY OF LABORATORY MEASUREMENTS OF HOMOGENEOUS-PATH ABSORPTION SPECTRA[14]

GAS	BAND, μ	OBSERVED INTERVAL, cm^{-1}	RANGE OF PRESSURES* mm Hg	RANGE OF ABSORBER, w**	NO. OF CURVES	RESOLUTION† APPROX.	RESEARCHERS
CO_2	1.6 & 1.4	6000 - 7200	75 - 760	540 - 8100	13	0.12μ	Howard, et al.[30]
	2.0	4600 - 5400	10 - 760	108 - 8630	32	0.9μ	Howard, et al.
	2.7	3300 - 4100	1 - 755	11 - 1619	67	0.07μ	Howard, et al.
	2.7	3450 - 3850	14.2 - 2065	0.164 - 24.4	32	10 - 15 cm^{-1}	Burch, et al.[29, 31]
	4.3	2200 - 2450	3.8 - 2115	0.0108 - 22.8	53	5 - 10 cm^{-1}	Burch, et al.
	4.3 & 4.8	2000 - 2500	1 - 735	9 - 1570	43	0.1μ	Howard, et al.
	4.3	2250 - 2450	38 - 760	1.0 - 300	10	10 cm^{-1}	Bradford[36]
	5.2	1800 - 2000	10 - 735	104 - 1570	9	0.1μ	Howard, et al.
	9.398 & 10.41	900 - 1100	103 - 3800	48 - 11,200	25	5 - 10 cm^{-1}	Burch, et al.
		720 - 875	103 - 3800	305 - 11,200	14	5 - 10 cm^{-1}	Burch, et al.
	15	500 - 900	20 - 745	1 - 863	37	0.5μ	Howard, et al.
		495 - 875	0.26 - 3190	0.0118 - 2470	30	5 - 10 cm^{-1}	Burch, et al.
H_2O	1.1	8250 - 9500	9.8 - 740	0.03 - 1.93	41	0.13μ	Howard, et al.
	1.38	6500 - 8000	3 - 740	0.026 - 3.85	62	0.12μ	Howard, et al.
	1.875	4000 - 6000	3 - 740	0.026 - 3.85	62	0.1μ	Howard, et al.
	1.875	4950 - 5800	27.5 - 862	0.0033 - 0.101	14	20 cm^{-1}	Burch, et al.
	2.7 & 3.2	2800 - 4400	2 - 750	0.017 - 2.1	114	0.07μ	Howard, et al.
	2.7	3000 - 4300	27.5 - 862	0.0033 - 0.101	5	20 cm^{-1}	Burch, et al.
	6.3	1000 - 2200	2.5 - 742	0.021 - 1.49	69	0.4μ	Howard, et al.

* Pressure designated is generally the equivalent pressure.
** w is expressed in precipitable centimeters for H_2O and in atmospheric centimeters for the other gases.
† Resolution specified is ordinarily center-band resolution and is used only to give an approximate designation.

TABLE 2-9. (Continued)

GAS	BAND, μ	OBSERVED INTERVAL, cm^{-1}	RANGE OF PRESSURES,* mm Hg	RANGE OF ABSORBER, w**	NO. OF CURVES	RESOLUTION† APPROX.	RESEARCHERS
H_2O	6.3	1200 - 2200	14.0 - 805	0.0041	15	6 cm^{-1}	Burch, et al.
		200 - 500	0.76 - 600	0.0041 - 0.143		5 - 10 cm^{-1}	Palmer
O_3	9.6	1000 - 2560	11.2 - 744	0.00278 - 1.968		7 cm^{-1}	Walshaw[32]
CO	4.666	2000 - 2250	1 - 3210	0.00096 - 22.2	147	25 cm^{-1}	Burch, et al.
	2.347	4100 - 4400	54 - 756	36.6 - 1140	26	15 - 20 cm^{-1}	Burch, et al.
CH_4	3.311	2700 - 3200	2 - 3085	0.015 - 188	88	25 cm^{-1}	Burch, et al.
	6.452 & 7.657	1100 - 1800	3.8 - 3050	0.026 - 188	86	10 cm^{-1}	Burch, et al.
N_2O	4.0	2400 - 2650	22.4 - 746	0.38 - 18.6	8	20 cm^{-1}	Burch, et al.
	4.5	2100 - 2300	1.0 - 3120	0.00016 - 76.4	177	25 cm^{-1}	Burch, et al.
	4.5	2100 - 2300	99.8 - 849	0.01 - 2.3	65		Abels[37]
	7.78 & 8.57	1100 - 1400	3.2 - 3035	1.37 - 46.7	11	10 cm^{-1}	Burch, et al.
	14.45 & 16.98	500 - 800	4.5 - 851	1.89 - 359	7	6 cm^{-1}	Burch, et al.

* Pressure designated is generally the equivalent pressure.
** w is expressed in precipitable centimeters for H_2O and in atmospheric centimeters for the other gases.
† Resolution specified is ordinarily center-band resolution and is used only to give an approximate designation.

TABLE 2-10. WIDE-BAND ABSORPTION OF CARBON DIOXIDE [30]

Band and Limits	Total Absorption ($\alpha\Delta\bar{\nu}$ in μ^{-1})	Remarks
15μ; 11.7 - 18.5μ	$-0.15 + 1.24 \log w + 1.06 \log p_e$	Also see Refs. 29, 31
10.4μ and 9.4μ		For data see Ref. 29
5.2μ; 5.05 - 5.35μ	$6.5 \times 10^{-5}\ w^{1/2} p_e^{0.4}$	Also see Ref. 31
4.8μ; 4.65 - 5.05μ	$2.75 \times 10^{-4}\ w^{1/2}\ p_e^{0.37}$	Also see Ref. 31
4.3μ; 4.0 - 4.65μ	$0.051 + 0.063 \log w + 0.058 \log p_e$	Also see Refs. 29, 31
2.7μ; 2.66 - 2.82μ	$-0.100 + 0.056 \log w + 0.050 \log p_e$	$w < 4$ atm-km
	$-0.126 + 0.078 \log w + 0.067 \log p_e$	$w > 4$ atm-km
	$-0.012 \log w \log p_e$	Also see Refs. 29, 31
2.0μ; 1.92 - 2.1μ	$+ 1.97 \times 10^{-4}\ w^{1/2}\ p_e^{0.39}$	Total absorption $< 0.02\mu$
	$-0.2144 + 0.0552 \log w + 0.0456 \log p_e$	Total absorption $> 0.02\mu$ Also see Ref. 31
1.6μ; 1.52 - 1.66μ	$1.61 \times 10^{-5}\ w^{1/2}\ p_e^{0.38}$	Also see Ref. 31
1.4μ; 1.38 - 1.47μ	$1.14 \times 10^{-5}\ w^{1/2}\ p_e^{0.41}$	Also see Ref. 31

NOTE: 1. All logs to the base 10.
2. w = atm-cm; p_e = mm *Hg*.

TABLE 2-11. WIDE-BAND ABSORPTION OF WATER VAPOR [30]

Band and Limits	Total Absorption ($\alpha\Delta\beta$ in μ^{-1})	Remarks
6.3μ; 4.75 - 8.7μ	$1.21 + 0.87 \log w + 0.62 \log p_e$	Questionable validity for $w < 0.001$ pr-cm or $p_e < 10$ mm Hg. Also see Ref. 31
3.2μ; 3.08 - 3.57μ	$0.041\, w^{1/2}\, p_e^{0.3}$	Also see Ref. 31
2.7μ; 2.3 - 3.08μ	$0.23\, w^{1/2}\, p_e^{0.32}$	Valid for total absorption $< 0.15\mu$
	$0.246 + 0.180 \log w + 0.109 \log p_e$	Valid for total absorption $> 0.15\mu$ Also see Ref. 31
1.87μ; 1.67 - 2.08μ	$0.053\, w^{1/2}\, p_e^{0.3}$	Valid for total absorption < 0.097
	$0.0445 + 0.0810 \log w + 0.0505 \log p_e$	Valid for total absorption > 0.097 Also see Ref. 31
1.38μ; 1.19 - 1.56μ	$0.0304\, w^{1/2}\, p_e^{0.3}$	Valid for total absorption < 0.067
	$0.0384 + 0.0875 \log w + 0.0376 \log p_e$	Valid for total absorption > 0.067
1.1μ; 1.05 - 1.19μ	$0.00375\, w^{1/2}\, p_e^{0.26}$	
0.94μ; 0.89 - 1.00 μ	$0.00335\, w^{1/2}\, p_e^{0.27}$	

NOTE: 1. All logs to the base 10.
2. w = pr-cm; p_e = mm Hg.

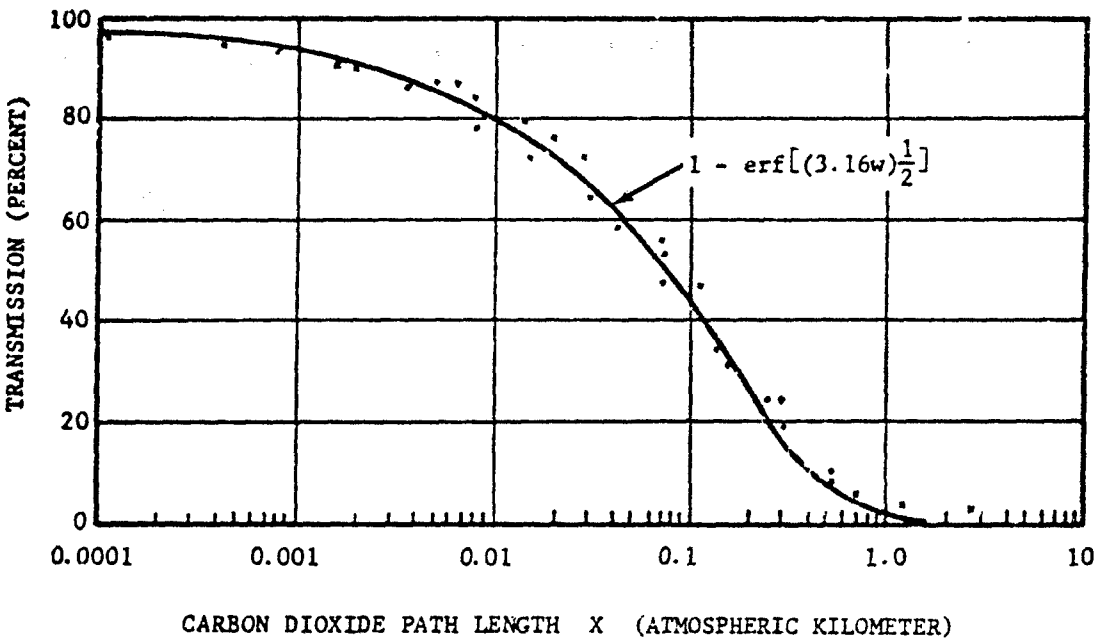


FIGURE 2-12. Experimental Fit of Burch and Williams Data to Error Function Absorption at 2.7 Microns

By far the largest window extends from 8 to 14 microns. Elder and Strong[39] and Streete[39A] have related the wide-band transmission in each window to the amount of water vapor in the path. Their data should be used with caution since it is difficult to eliminate the effect of aerosol scattering (especially at wavelengths less than 5 microns) from absorption due to water vapor. The transmittance inside a window region is expected to vary as follows:

$$t = (1 - C\sqrt{w}\,) \exp\,[-(\alpha w + \beta_a X)] \quad (2\text{-}57)$$

where

C = constant to be determined
α = continuum absorption coefficient due to water vapor
w = amount of water vapor in the path
β_a = aerosol scattering coefficient (par. 2-4.3.1)
X = path length

The second term on the right hand side of Eq. 2-57 is due to the few water vapor lines that may exist in the window and is only important for short paths. The continuum absorption coefficient α is the sum of the contributions of all the distant water vapor lines in the neighboring bands and is a slowly varying function of wavelength. The measurements of Ref. 39 clearly show the effect of the exponential in Eq. 2-57 and the exponential variation of transmittance with w. Careful measurements in the 8- to 14-micron window have been performed in Refs. 40 and 41 to separate the effect of water vapor absorption from aerosol scattering. A plot of α in Eq. 2-57 vs wavelength λ is shown in Fig. 2-13. Transmission curves in the 8- to 14-micron window taking into account all these effects can be found in Ref. 42. Recent unpublished measurements by Burch indicate that absorption due to water vapor in the 8- to 14-micron window is strongly dependent on water vapor partial pressure. Thus at altitudes above a few kilometers where this partial pressure is down by an order of magnitude as compared to sea level, absorption due to water vapor should be negligible. This should apply also to the other windows.

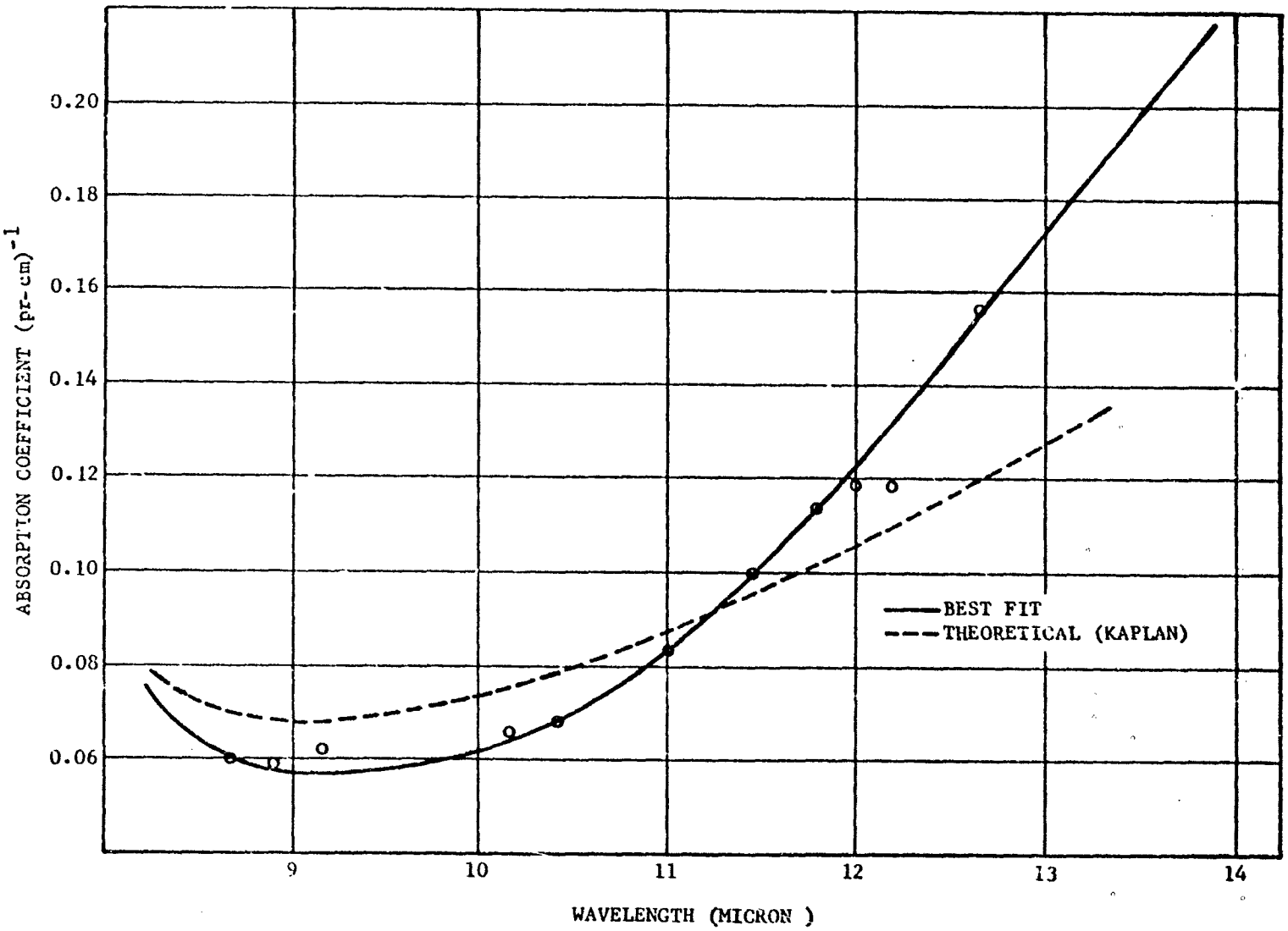


FIGURE 2-13. Variation of Continuum Absorption Coefficient vs Wavelength

2-4.2.4.4 Equivalent Sea-level Path

The problem of atmospheric absorption is complicated by the fact that in most cases the paths considered are slant paths along which the temperatures and pressures vary. The discussion and results in the previous paragraphs apply only to the case of constant pressure and temperature paths.

2-4.2.4.4.1 Equivalent-path Absorption Calculations

It is possible under certain conditions, however, to reduce a slant path through the atmosphere to an equivalent sea level path in which pressure and temperature are constant. The reduction can be easily justified if either the weak line or the strong line approximation holds[27]. If the line intensities are not too dependent on temperature, the results are as follows:

(1) Weak Line Approximation:

The equivalent absorber content is w given by

$$w = \int_{Path} \zeta(h) \frac{dh}{\cos \theta} \tag{2-58}$$

where $\zeta(h)$ is the absorber concentration (atm-cm km^{-1}), h is the altitude (km), and θ is the zenith angle. The absorptance will be given by the first of Eq. 2-50.

(2) Strong Line Approximation:

The pressure and temperature corrected absorber content in a slant path is defined by[43]

$$\bar{w} = \int \zeta(h) \left(\frac{dh}{\cos \theta}\right) p(h) \left[\frac{273}{T(h)}\right]^{1/2} \tag{2-59}$$

and the average pressure in the path is

$$\bar{p} = \frac{\bar{w}}{w} \tag{2-60}$$

where $p(h)$ is the pressure in atmospheres at altitude h, T is the temperature in degrees Kelvin and w is given by Eq. 2-58. Transmission through a slant path can then be obtained from any of the models discussed above (see the second part of Eqs. 2-50 and 2-56) either by replacing w wherever it occurs by $\bar{w}$, or by replacing γ by $\gamma\bar{p}$ where γ is the line half-width at sea level.

To evaluate the integrals in Eqs. 2-58 and 2-59, it is necessary to know $\zeta(h)$, $T(h)$, and $p(h)$. A discussion of absorber concentration was given in par. 2-4.2.2 and $\zeta(h)$ can be obtained from Table 2-5. A typical temperature profile is given in Fig. 2-9 and the pressure, which can be assumed to decrease exponentially with altitude, is given by

$$p = \exp(-h/7.5), \text{ atmospheres}$$

where h is in km and 7.5 km represents the average scale length of the atmosphere.

In the case of the Goody model, it can be shown[44] that the absorptance is given by an equation analogous to Eq. 2-54 which holds in both the weak and the strong line limits.

$$\alpha = 1 - \exp\left[-\frac{wS_0}{d\left(1 + \frac{wS_0}{\pi\gamma\bar{p}}\right)^{1/2}}\right] \tag{2-61}$$

where $\bar{p}$ is defined by Eqs. 2-59 and 2-60, and w by Eq. 2-58.

2-4.2.4.4.2 Equivalent-path Absorber Contents

A typical atmospheric measurement situation is illustrated in Fig. 2-14, in which observer A, at an altitude h_1, views object B, at altitude h_2, along a slant path, at zenith angle θ. The exact calculations of absorber content along this path are quite complex. There are certain simplifications however, which, for most problems, will permit rapid estimates of absorber contents to be made. If the zenith angle θ is less than 85 deg, the curvature of the earth can be ignored and absorber contents estimated by dividing vertical path contents by cos θ. Values for w (uncorrected curve) and $\bar{w}$ (pressure-corrected curve), for a vertical path extending from any altitude to a point outside the earth's atmosphere, may be obtained from Fig. 2-15 for any gas with constant mixing ratios (CO_2, CH_4, etc.). Values for w and $\bar{w}$ are given in atm-km but may be converted to atm-cm by using Table 2-5. Water vapor w and $\bar{w}$ values can be obtained from Fig. 2-16 for a dry model atmosphere and a wet model atmosphere[45].

Another situation which lends itself to simplified calculations involves slant paths near horizontal and path altitudes which vary only slightly (they can only occur for path lengths

less than 100 km). In this case, for a uniformly distributed absorber such as CO_2,

$$\bar{w} = w_o \exp\left[-h/7.5\right] \qquad (2\text{-}62)$$

and

$$\bar{w} = w_o \exp\left[-2h/7.5\right] \qquad (2\text{-}63)$$

where

h = path altitude, km
w_o = absorber content for the same path length at sea level

Absorber values w for ozone and water vapor may be approximated from Figs. 2-7 and 2-9, respectively. It should be understood, however, that values for any actual problem can vary considerably from those given in these figures. Approximate values for $\bar{w}$ for ozone and water vapor can be obtained by multiplying the w values by exp $(-h/7.5)$.

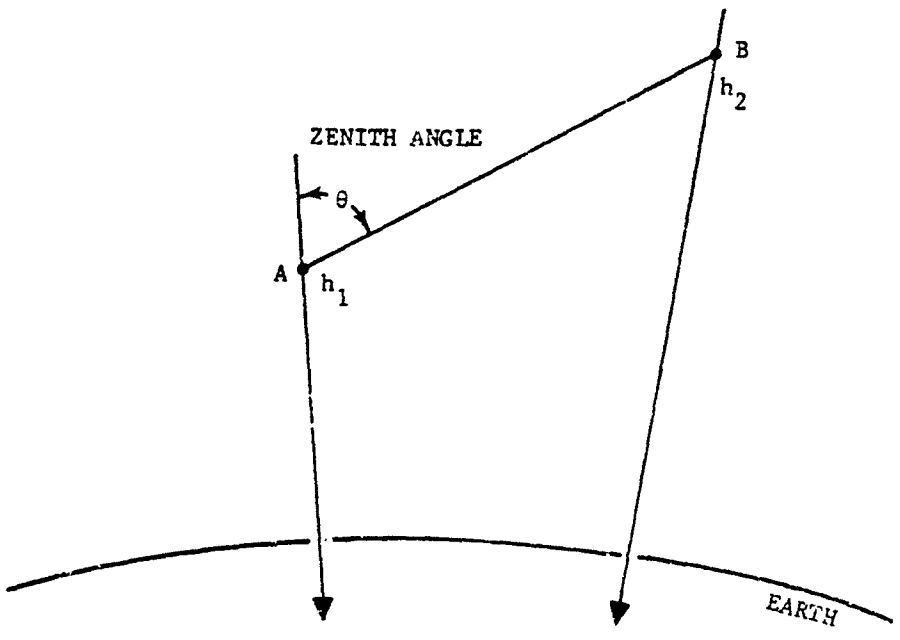


FIGURE 2-14. Geometrical Relation Between Observer A and Object B

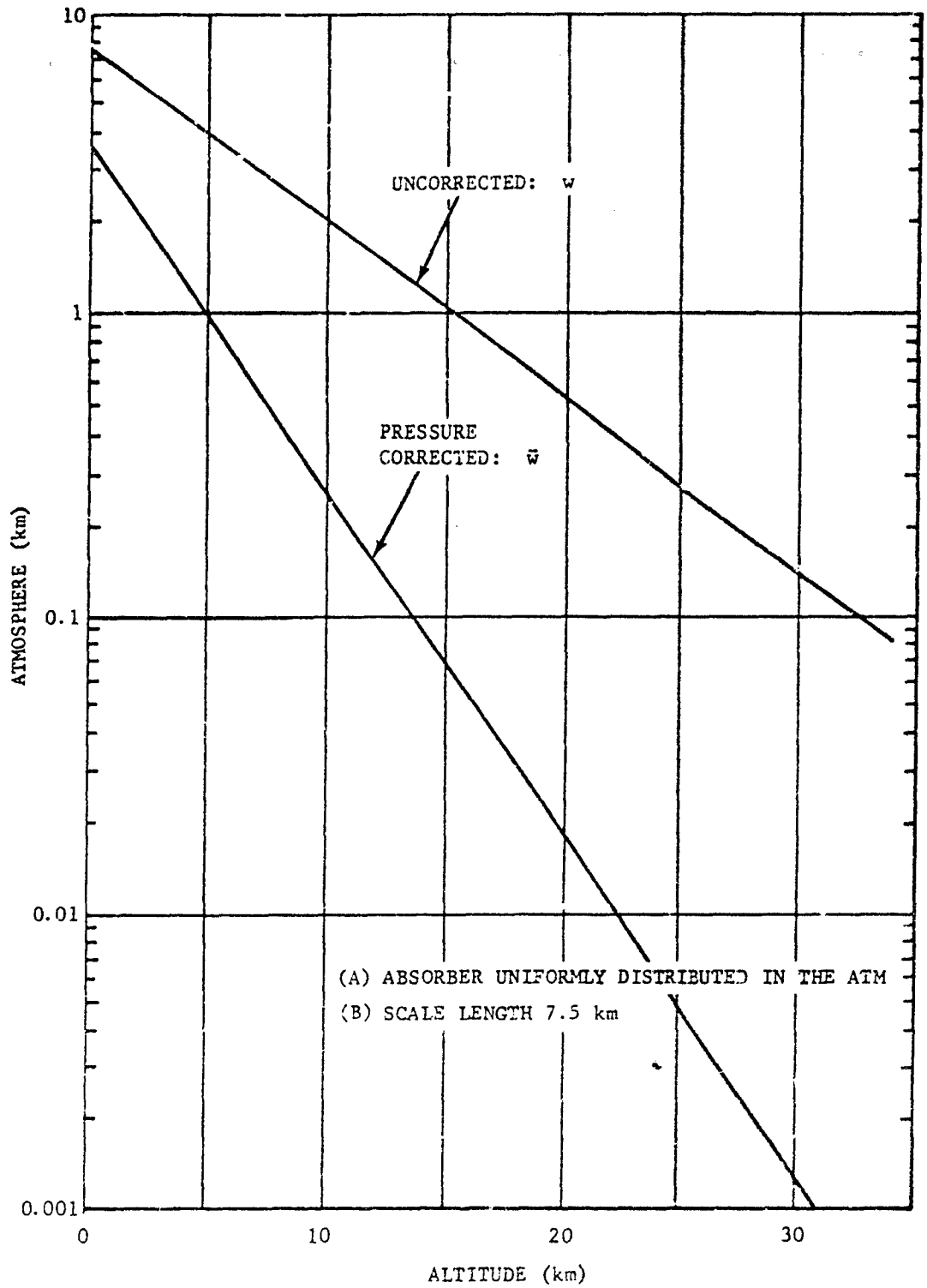


FIGURE 2-15. IR Absorber Content Above a Given Altitude

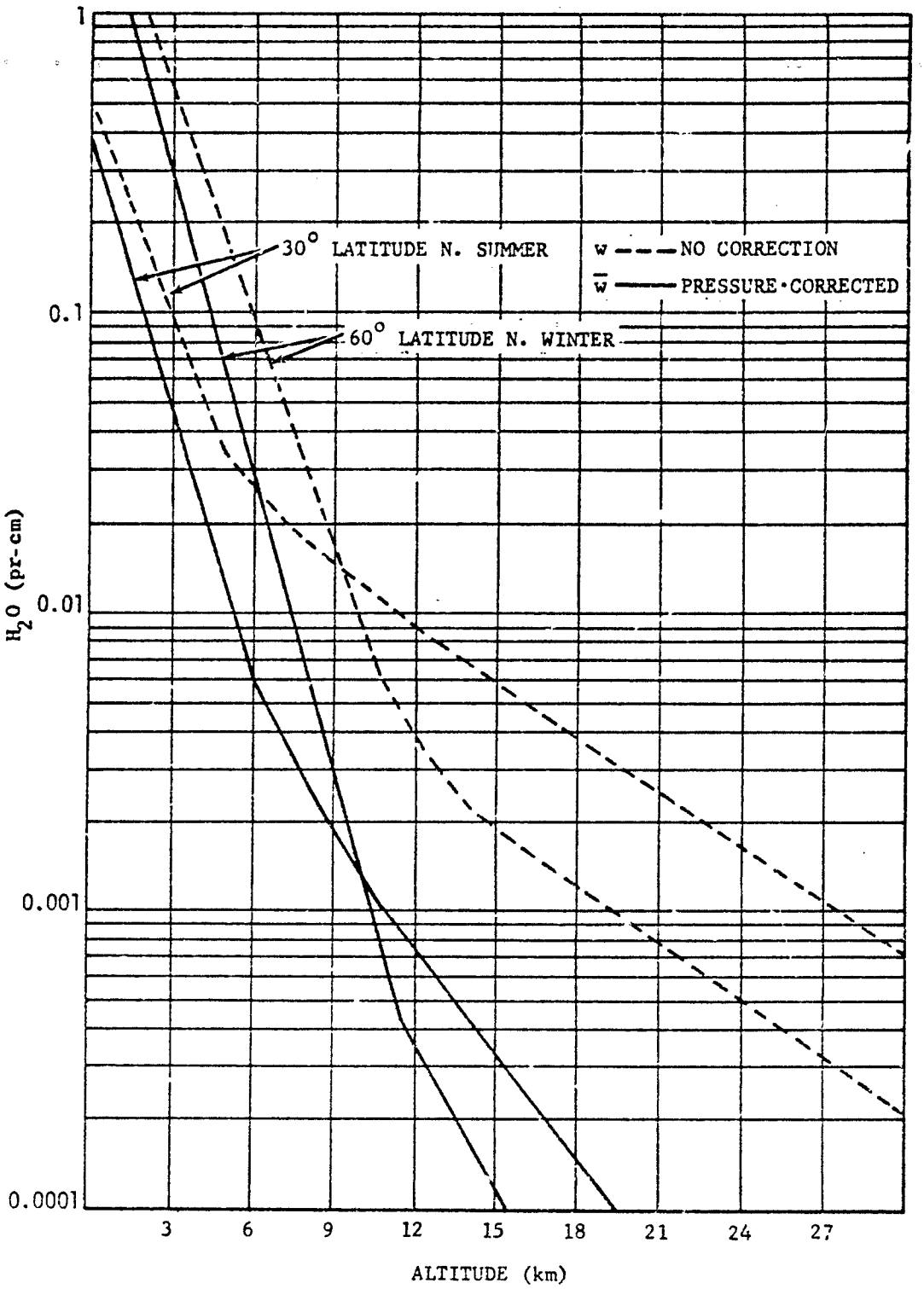


FIGURE 2-16. Water Vapor Content Above a Given Altitude

2-4.2.4.5 Examples of Atmospheric Transmission Calculations

Problem 1: Compute the absorptance of CO_2 at 2.7 microns along an atmospheric slant path extending between 15 and 20 km altitude for a zenith angle of 60 degrees.

Solution: The discussion of CO_2 transmission (par. 2-4.2.3.3) indicates that for pressures less than atmospheric and for appreciable absorption (i.e., the strong line approximation), the error function defines the transmission curve. In Fig. 2-12 just such a function is implemented using 2.7-micron, CO_2 transmission data obtained from laboratory measurements. The problem is, therefore, reduced to finding the appropriate CO_2 path length.

The equivalent sea level paths $\bar{w}$ to be used in the strong line approximation is given by Eq. 2-59. As discussed in par. 2-4.2.4.4.2, the $\bar{w}$ values can be determined from the pressure corrected curve in Fig. 2-15. Over a vertical path between 15 and 20 km

$$\bar{w}_v = 0.068 - 0.018 = 0.05 \text{ atm-km}$$

Since the path in question has a zenith angle of 60°, $\bar{w}$ for the path will be

$$\bar{w} = \bar{w}_v/\cos 60^\circ = 0.1 \text{ atm-km}$$

Referring to Fig. 2-12, then, the absorptance of CO_2 at 2.7 microns is 0.42. Alternatively, the transmittance t can be determined from the band parameters for CO_2 listed in Table 2-6. At 2.7 microns, S/d equals 0.7 and $2\pi\gamma_0/d = 4.93 \times 10^{-4}$. From the last column of Table 2-5 $\bar{w}$ can be converted to atm-cm, $\bar{w} = 0.1 \times 32 = 3.2$ atm-cm. The argument of the error function in Eq. 2-52 is

$$\frac{\pi S \gamma_0}{d^2} \times p \times \bar{w} = \frac{0.7 \times 4.93 \times 10^{-4}}{2} \times 760 \times 3.2 \qquad (2\text{-}64)$$
$$= 0.42$$

Therefore,

$$t = 1 - erf\left(\sqrt{0.42}\right)$$
$$= 0.35$$

Problem 2: Determine the transmittance of the atmosphere between 1.65 and 2.15 microns over a 20-km path at an altitude of 10 km (no clouds).

Solution: The broad band approximation of Tables 2-10 and 2-11 provides the simplest method of computation. (Tables 2-10 and 2-11 show that the 1.65- to 2.15-micron region completely encompasses the 2-micron CO_2 band and the 1.8-micron H_2O band.) At 10 km, the atmospheric pressure is 200 mm of Hg, the CO_2 path length is given by Eq. 2-62; the path can be transformed into atm-cm by means of Table 2-5.

$$w = 20 \exp(-10/7.5) = 5.28 \text{ atm-km}$$
$$= 5.28 \times 32 = 169 \text{ atm-cm}$$

These values are used in the approximation in Table 2-10 to determine the total absorption

$$\alpha\Delta\bar{\nu} = 1.97 \times 10^{-4} w^{1/2} (p_e)^{0.39}$$
$$= 1.97 \times 10^{-4} (169)^{1/2} (200)^{0.39}$$
$$= 0.0202 \text{ micron}^{-1}$$

This value for $\alpha\Delta\bar{\nu}$ is sufficiently close to the absorption limit of 0.02 micron^{-1} wherein this approximation is valid, thereby, making this an acceptable value.

The mean water-vapor content for a 20-km path at 10 km is determined from Fig. 2-9 as as 0.022 pr-cm and from Table 2-11

$$\alpha\Delta\bar{\nu} = 0.053 (0.022)^{1/2} (200)^{0.3}$$
$$= 0.0385 \text{ micron}^{-1}$$

Since $\alpha\Delta\bar{\nu}$ is smaller than 0.097 and, therefore, in the range of the approximation used, it is acceptable. The total absorption for the two bands is $0.0385 + 0.0214 \approx 0.060$ micron. The transmittance for the 1.65- to 2.15-micron region is thus

$$t = 1 - \frac{0.060}{(2.15 - 1.65)} = 0.88$$

2-4.3 SCATTERING

As a beam of radiation traverses a medium, it is attenuated by scattering processes as well as the absorption processes previously discussed. Scattering occurs whenever a scattering center (drop of water, particle of dust, molecule, etc.)

causes the spatial redistribution of the radiation incident on it. Scattering processes which occur include: (1) nonresonant scattering by bound electronic or molecular systems, (2) diffraction of radiation by the scattering centers, (3) reflection of radiation from the surface of the scattering centers, and (4) refraction of radiation through the scattering center. The first two processes deal primarily with scattering centers that are small compared to the incident radiation wavelength, while the second two processes are applicable to large scattering centers. The two general categories of molecular scattering and aerosol scattering are discussed in this paragraph.

2-4.3.1 Scattering Coefficient

The scattering coefficient is related to atmospheric transmission in the following manner. Neglecting all absorption processes, transmission $t(\lambda)$ defined as the ratio of the transmitted radiation intensity to the incident radiation intensity over a path length X is given by

$$t(\lambda) = \exp\left[-\beta_s(h,\lambda)\,X\right] \tag{2-65}$$

where $\beta_s(h,\lambda)$ is the total scattering coefficient. The total scattering coefficient β_s depends on the altitude h and the radiation wavelength λ. It can be considered as the sum of the two separate scattering coefficients, $\beta_m(h,\lambda)$ the molecular, scattering coefficient, and $\beta_a(h,\lambda)$ the aerosol scattering coefficient. Since molecular scattering deals with light scattered by particles much smaller than the radiation wavelength, the Rayleigh scattering theory is used to determine the value of β_m. However, aerosol scattering which involves scattering particles that may be comparable in size to the radiation wavelength follows the Mie theory[46].

The molecular scattering coefficient β_m is the product of the Rayleigh scattering cross section σ_R and the number density of molecules M. A table of M for various altitudes is given in Ref. 47. The scattering cross section σ_R is given by

$$\sigma_R = \frac{8\pi^3(n_s^2-1)^2}{3\lambda^4 M^2}\left(\frac{6+3\delta}{6-7\delta}\right) \tag{2-66}$$

where

n_s = index of refraction of air at STP
M = molecular number density at STP, cm^{-3}
λ = wavelength of the radiation, cm
δ = depolarization factor for the medium

The most recent depolarization data (a table of δ is given in Ref. 53) gives a value of 0.035. Due to the λ^{-4} dependence, molecular scattering becomes negligible compared to aerosol scattering for wavelengths greater than approximately 1 micron[35].

Theoretically, aerosol scattering can be calculated by means of the Mie scattering theory in the following manner:

The aerosol scattering coefficient β_a is given by[46]

$$\beta_a(h,\lambda) = \pi\int_{r_{min}}^{r_{max}} r^2 M(h,r)K(x,n)dr \tag{2-67}$$

where

r = particle radius, cm
$M(h,r)$ = number density of particles at altitude h as a function of r, cm^{-3}
$K(x,n)$ = efficiency factor defined as the ratio of scattering cross section to geometrical cross section
x = size parameter defined as the ratio of the circumference of the particle to the wavelength of the radiation
n = index of refraction of the particles

The value of $M(h,r)$ is discussed in Refs. 48, 49, and 50. The value of $K(x,n)$, given by the Mie theory[46], is the exact mathematical solution to an infinite, plane, monochromatic, electromagnetic wave scattered from a perfect sphere of any radius and any index of refraction. Since $K(x,n)$ is a very complex function, this integral is generally not evaluated (Ref. 16 is an exception). Assuming that the size distribution of the particles remains constant with altitude[51], an experimental approximation for determining $\beta_a(h,\lambda)$ is given by

$$\beta_a(h,\lambda) = \beta_a(o,\lambda)\,\frac{M(h)}{M(o)} \tag{2-68}$$

where

$\beta_a(o,\lambda)$ = scattering coefficient at sea level, cm^{-1}
$M(h)$ = aerosol number density at altitude h
$M(o)$ = aerosol number density at sea level

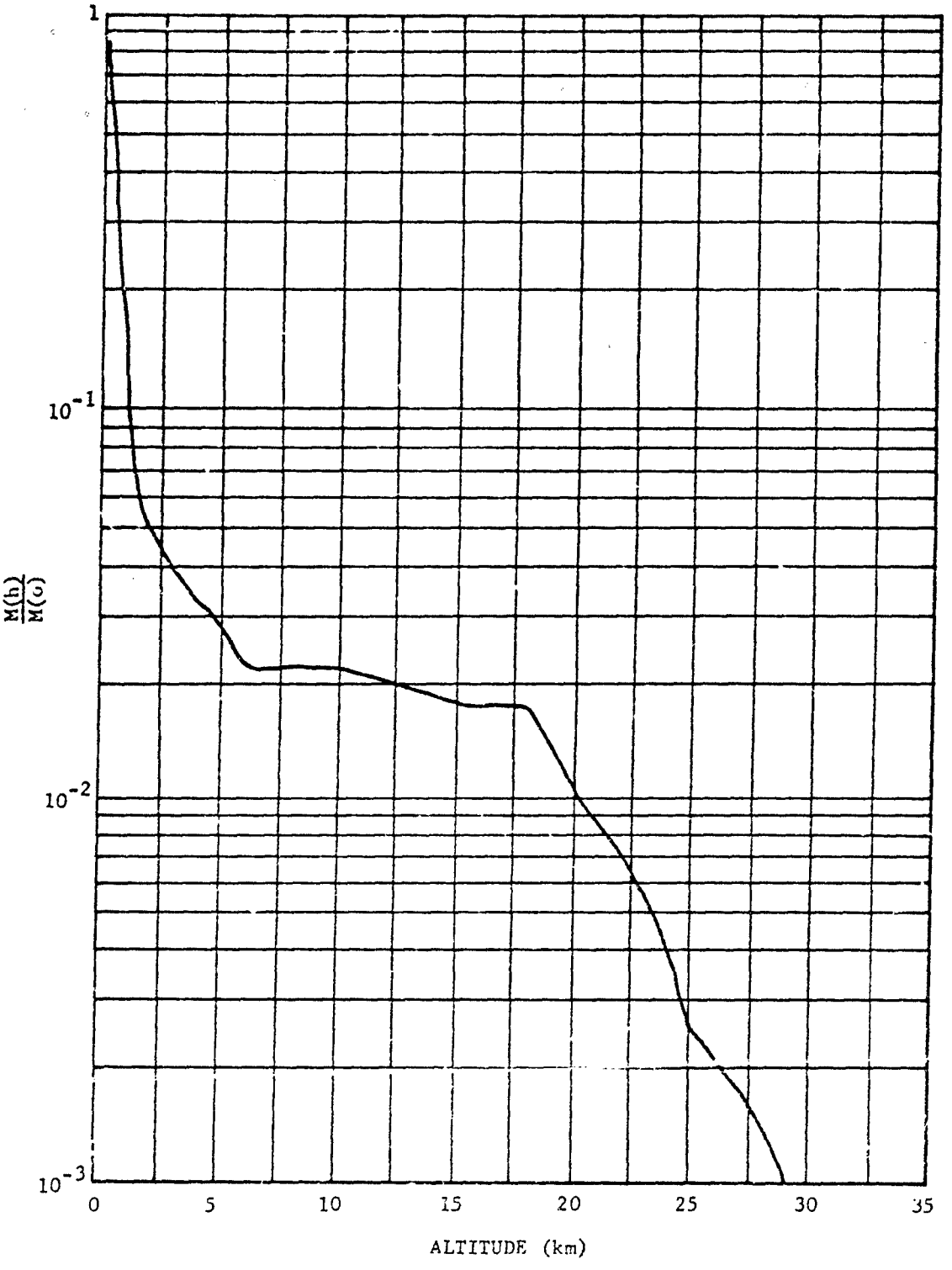


FIGURE 2-17. Normalized Particle Density as a Function of Altitude

Since both $\beta_a(o,\lambda)$ and $\frac{M(h)}{M(o)}$ depend on the geographical and meteorological conditions, only representative values of $\beta_a(h,\lambda)$ are determined and some general trends noted. A representative value of $\beta_a(o,\lambda)$, derived from data given in Ref. 48 for a clear atmosphere (meteorological range = 25 km), is given by

$$\beta_a(o,\lambda) = 0.124\ \lambda^{-0.43},\ \text{km}^{-1} \quad (2\text{-}69)$$

where λ is in microns. This wavelength dependence of β_a is only applicable for wavelengths less than 2.3 microns and meteorological ranges greater than 9 km. For wavelengths between 2.3 and 10 microns, the wavelength dependence seems to disappear. Beyond 10 microns, the scattering coefficient decreases with wavelength[48]. The constant, 0.124, increases as the meteorological range decreases (par. 2-4.3.2). For a light fog, the value of the scattering coefficient is about 0.7.

Representative values of the normalized aerosol number density $\frac{M(h)}{M(o)}$ can be determined from experimental data found in Ref. 52 for a clear atmosphere and are shown as a function of altitude in Fig. 2-17.

As an example, for $\lambda = 2$ microns, $\beta_a(o,2) = 0.092\ \text{km}^{-1}$. At an altitude of 5 km

$$\frac{M(h)}{M(o)} = 3 \times 10^{-2}.$$

Therefore,

$$\beta_a(5,2) = 2.76 \times 10^{-3}\ \text{km}^{-1}.$$

For comparison, the molecular scattering coefficient is calculated from Rayleigh's theory of scattering under the same conditions. Here, the molecular number density M at 5 km is $1.53 \times 10^{19}\ \text{cm}^{-3}$ and the index of refraction n is 1.0002729 for $\lambda = 2$ microns[53]. Therefore, since $M = 2.547 \times 10^{19}\ \text{cm}^{-3}$,

$$\beta_m = \frac{8\pi^3 (5.46 \times 10^{-4})^2}{(3)(2 \times 10^{-4})^4 (2.55 \times 10^{19})^2} \left(\frac{6.105}{5.755}\right)(1.53 \times 10^{19})$$

$$= 3.86 \times 10^{-10}\ \text{cm}^{-1}$$

$$= 3.86 \times 10^{-5}\ \text{km}^{-1}$$

Since molecular scattering is negligible compared to aerosol scattering (and ignoring absorption) the transmission $t(\lambda)$ over a 10-km path X at 5-km altitude is given by, Eq. 2-65,

$$t(\lambda) = \exp[-\beta_a X]$$
$$= \exp[-\ 0.028]$$
$$= 97\%$$

2-4.3.2 Meteorological Range

Meteorological range is defined as a 2 percent contrast between a distant nonreflecting, nonemissive target and the horizon sky. Contrast is the absolute difference in sky and target intensities divided by their sum. The condition of 2 percent contrast corresponds to the average person's capability for distinguishing a large black object from the horizon sky. The meteorological range R_m can be further defined as

$$R_m = \frac{3.9}{\beta_s} \qquad (2\text{-}70)$$

which relates the meteorological range R_m in km to the scattering coefficient β_s in km^{-1} (Ref. 1) Thus, meteorological range is a convenient parameter to describe atmospheric scattering in the visible region.

2-4.3.3 Atmospheric Scintillation

Localized as well as large-volume inhomogeneities in the index of refraction of the atmosphere will affect a beam of radiation passing through the atmosphere. Small or localized variations in the refractive index cause portions of the beam to be refracted away from the direction of the main portion of the beam. These small index variations also affect the wavefront (cross-sectional relative phase) of the propagating radiation. Large scale variations in the atmospheric index of refraction cause the entire beam to be refracted.

The spatial variations of atmospheric index of refraction, caused by atmospheric temperature and pressure gradients, fluctuate in time and space. This condition of a turbulent inhomogeneous atmosphere causes intensity fluctuation as well as angular variation in a collimated beam of light propagating through it. This time variation in intensity of the received beam is termed scintillation, while the degradation of the image formed with the received radiation, due to the angular variation, is termed image dancing. Perhaps the two most common manifestations of these atmospheric effects are the blurring of stellar images and the propagation and detection of laser beams.

The angular deviation of starlight due to its passage through the atmosphere causes a blurring of the telescope image of that star (independently of the optical resolution of the telescope itself). This phenomenon of image dancing is evident when high-resolution optical systems are used and long atmospheric paths exist. Short exposure times tend to improve image quality by "freezing" the temporal fluctuations of the atmosphere. With this technique, an image may be shifted from the optical axis of the imaging system or possibly distorted but the image blur will be reduced. The effects of atmospheric turbulence on the formation of images of exo-atmospheric objects can also be minimized by operating at high altitudes so that the atmospheric path is reduced. Also, atmospheric turbulence is diminished at night when thermal gradients are less pronounced. The atmospheric limitation to optical resolution of exo-atmospheric objects is about 1 sec of arc[54].

In addition to image dancing, the intensity of received radiation will fluctuate as inhomogeneities of the atmospheric index of refraction bend portions of the radiation away from the finite collecting aperture. The larger the collecting aperture, the less will be the relative effect of atmospheric scintillation. Fig. 2-18 illustrates the dependence of stellar intensity scintillation on aperture size[55].

A detailed theoretical analysis of wave propagation in the turbulent atmosphere can be found in Ref. 52. The degradation of images by the atmosphere and the restoration of such images is thoroughly treated in Ref. 56.

These same atmospheric turbulence effects also cause destructive interference to take place within a collimated laser beam as it propagates through the atmosphere[57] and, therefore, introduces an intensity modulation on the beam. Here, atmospheric index of refraction heterogeneousness disturbs the wavefront of a propagating laser beam so that some portions of the beam interfere with other portions. Therefore, instead of receiving a beam whose cross section has a uniform intensity, the beam is spatially broken up into many filaments of light[58]. These filaments vary in intensity and position with time so that the energy received depends on exposure time and on the size of the collecting optics, i.e., on the amount of temporal and spatial integration of the received energy. If

integration time is short (or if a short pulse of energy is to be received), turbulence may break up the beam so that no energy is incident on the receiver. The temporal fluctuation of laser energy received over a 90-mile atmospheric path by a 3-in. aperture is depicted in Fig. 2-19[59]. This amplitude modulation due to atmospheric turbulence must be considered when attempting to detect information carried as an amplitude modulation on a light beam.

Since atmospheric turbulence distorts the phase of a propagating wave, a receiving system employing heterodyne detection of an optical signal will also be adversely affected by turbulence (heterodyne detection is discussed in Chapter 3). If there were no atmospheric turbulence, a plane wavefront would be received, and the signal-to-noise ratio would increase in proportion to the amount of energy collected. However, since heterodyne detection depends on a definite phase relation between the signal and local oscillator beam, the random phase distortion of the signal beam caused by atmospheric turbulence increases the noise in the heterodyne process. The smaller the collecting aperture, the more uniform the phase of the radiation across it; therefore, increasing the diameter of the collecting aperture d beyond a critical diameter d_o results in very little improvement in the signal-to-noise ratio[60]. Fig. 2-20 illustrates the dependence of d_o upon wavelength λ and altitude h of the collecting aperture for heterodyne detection of coherent exo-atmospheric radiation. The scale on the left ordinate of Fig. 2-20 pertains to nighttime conditions, while the scale on the right pertains to daytime conditions.

Ref. 61 contains the results of an intensive survey of the literature dealing with scintillation.

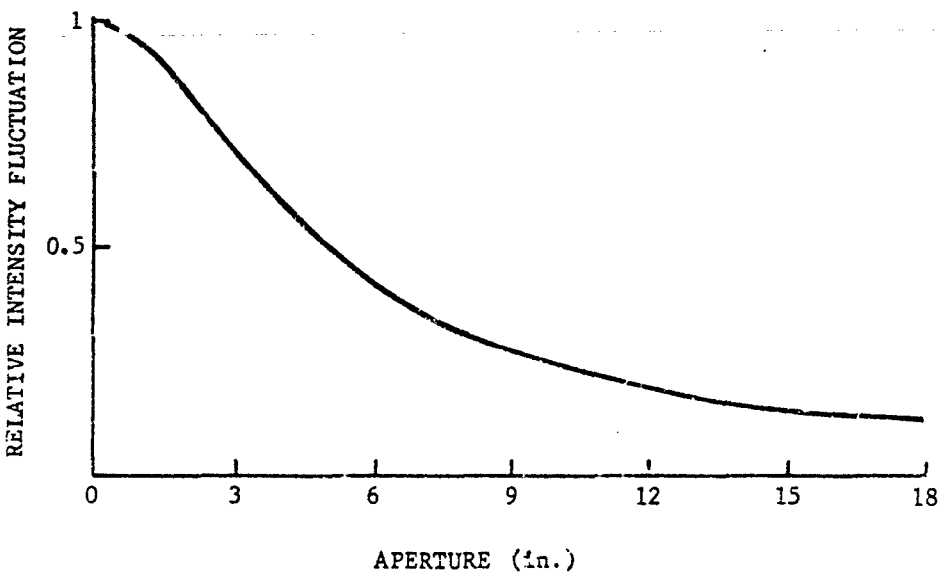


FIGURE 2-18. Scintillation vs Telescope Aperture

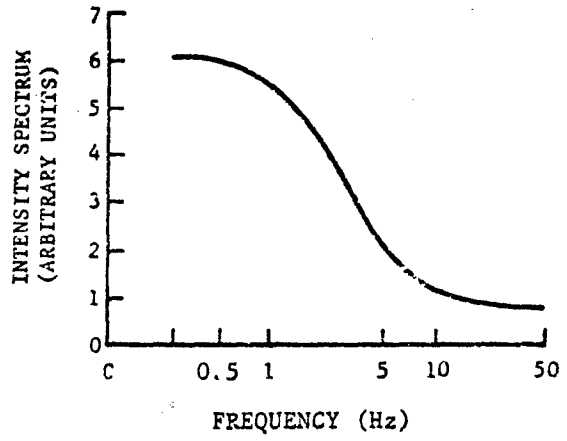


FIGURE 2-19. Frequency of Intensity Scintillation

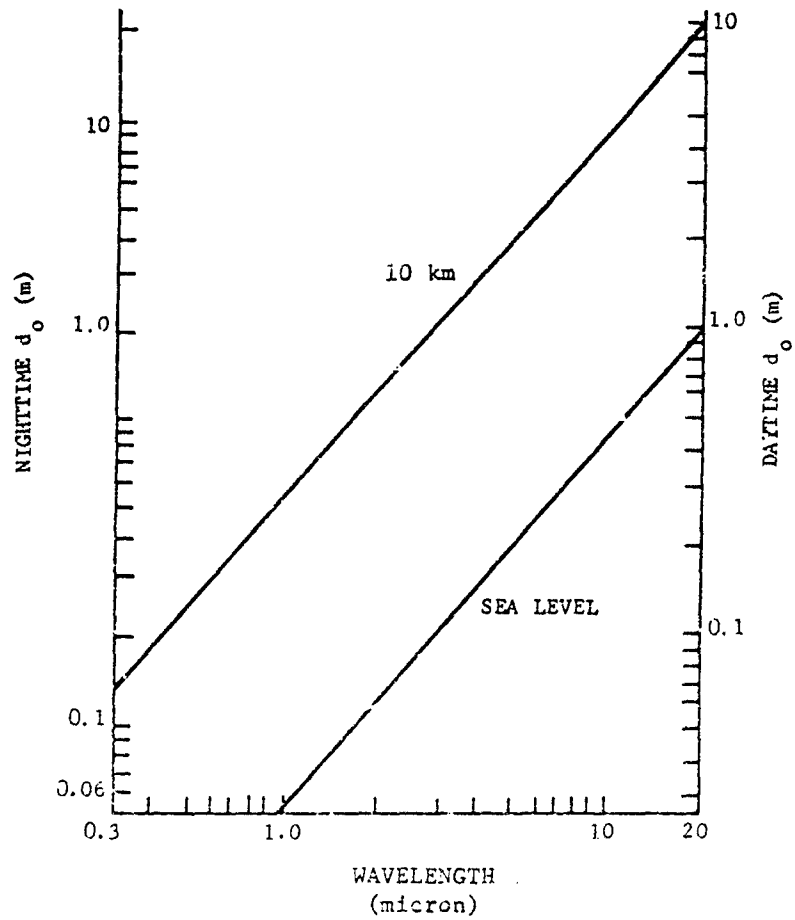


FIGURE 2-20. Dependence of the Diameter d_o of the Collecting Optics, Upon Wavelength λ and Altitude at a Zenith Angle of 70° for Both Nighttime and Daytime Conditions

2-5 TRANSMISSION OF INFRARED RADIATION THROUGH OPTICAL MATERIALS

Infrared radiation can be described as an electromagnetic wave phenomenon. The spectrum of infrared radiation is between the visible and microwave regions. (See Fig. 2-21.)

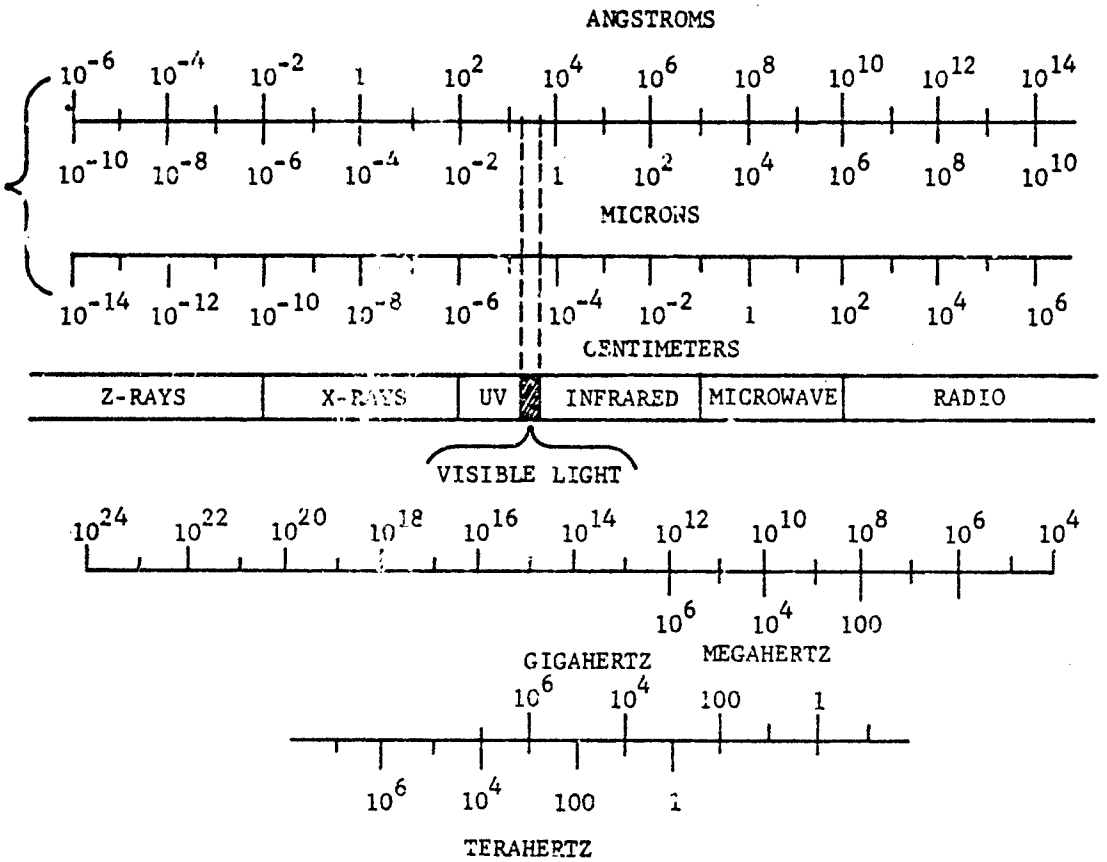


FIGURE 2-21. The Electromagnetic Spectrum

2-5.1 MAXWELL'S EQUATIONS

Electromagnetic waves are described mathematically by Maxwell's Equations[62,63,64]. These equations are

$$\nabla \cdot \boldsymbol{D} = \rho \tag{2-71}$$

$$\nabla \times \boldsymbol{H} = \boldsymbol{J} + \frac{\partial \boldsymbol{D}}{\partial t} \tag{2-72}$$

$$\nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t} \tag{2-73}$$

$$\nabla \cdot \boldsymbol{B} = 0 \tag{2-74}$$

where

$$\boldsymbol{D} = \epsilon_0 \boldsymbol{E} + \boldsymbol{P} = \epsilon \boldsymbol{E} \tag{2-75}$$

$$\boldsymbol{H} = \frac{1}{\mu_0} \boldsymbol{B} - \boldsymbol{M} = \frac{1}{\mu} \boldsymbol{B} \tag{2-76}$$

and

$\boldsymbol{D}$ = displacement (vector), coulomb meter^{-2} (C·m^{-2})

$\boldsymbol{P}$ = polarization (vector), coulomb meter^{-2} (C·m^{-2})

$\boldsymbol{E}$ = electric field (vector), volt meter^{-1} (V·m^{-1})

$\boldsymbol{B}$ = magnetic induction (vector), weber meter^{-2} (Wb·m^{-2})

$\boldsymbol{M}$ = magnetization (vector), weber meter^{-2} (Wb·m^{-2})

$\boldsymbol{H}$ = magnetic field (vector), ampere-turn meter^{-1}

$\boldsymbol{J}$ = electric current density (vector), ampere meter^{-2} (A·m^{-2})

ρ = charge density (scalar), coulomb meter^{-3} (C·m^{-3})

ϵ_0 = permittivity of free space (scalar) $= \frac{10^7}{4\pi c^2}$ farad meter^{-1} (F·m^{-1})

μ_0 = permeability of free space (scalar) $= 4\pi \times 10^{-7}$ henry meter^{-1} (H·m^{-1})

c = speed of light $= 2.998 \times 10^8$ meter sec^{-1} (m·sec^{-1})

The Lorentz force $\boldsymbol{F}$ on a charge q, moving with velocity $\boldsymbol{v}$, is given by

$$\boldsymbol{F} = q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}) \tag{2-77}$$

where

$\boldsymbol{F}$ = force (vector), newton = kilogram-meter second^{-2} (N = kg·m·sec^{-2})

q = charge (scalar), coulomb (C)

$\boldsymbol{v}$ = velocity (vector), meter second^{-1} (m·sec^{-1})

For a medium with no free electrical charges, $\rho = 0$ and $\boldsymbol{J} = 0$. Thus Eqs. 2-71 through 2-74 reduce to

$$\nabla \cdot \boldsymbol{D} = 0 \tag{2-78}$$

$$\nabla \times \boldsymbol{B} = \mu\epsilon \frac{\partial \boldsymbol{E}}{\partial t} \tag{2-79}$$

$$\nabla \times \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t} \tag{2-80}$$

$$\nabla \cdot \boldsymbol{B} = 0 \tag{2-81}$$

2-5.2 UNITS

Maxwell's Equations in par. 2-5.1.1 are in "rationalized mks units". Other systems of units are possible and, of course, all give the same physical results. The other popular system is Gaussian units. In Gaussian units, Eqs. 2-71 through 2-77 are written as

$$\nabla \cdot \boldsymbol{D} = 4\pi\rho \tag{2-82}$$

$$\nabla \times \boldsymbol{H} = \frac{4\pi}{c} \boldsymbol{J} + \frac{1}{c} \frac{\partial \boldsymbol{D}}{\partial t} \tag{2-83}$$

$$\nabla \times \boldsymbol{E} = \frac{1}{c} \frac{\partial \boldsymbol{B}}{\partial t} \tag{2-84}$$

$$\nabla \cdot \boldsymbol{B} = 0 \tag{2-85}$$

$$\boldsymbol{D} = \boldsymbol{E} + 4\pi\boldsymbol{P} = \epsilon \boldsymbol{E} \tag{2-86}$$

$$\boldsymbol{H} = \boldsymbol{B} - 4\pi\boldsymbol{M} = \frac{1}{\mu} \boldsymbol{B} \tag{2-87}$$

$$\boldsymbol{F} = q(\boldsymbol{E} + \frac{1}{c} \boldsymbol{v} \times \boldsymbol{B}) \tag{2-88}$$

The conversion between the units for the various quantities is given in Table 2-12.

TABLE 2-12. CONVERSION TABLE OF PHYSICAL QUANTITIES

PHYSICAL QUANTITY	SYMBOL	RATIONALIZED MKS	GAUSSIAN*
Capacitance	C	1 farad (F)	$= 9 \times 10^{11}$ cm
Charge	q	1 coulomb (C)	$= 3 \times 10^{9}$ statcoulombs
Charge density	ρ	1 C m^{-3}	$= 3 \times 10^{3}$ statcoul cm^{-3}
Conductivity	σ	1 mho m^{-1}	$= 9 \times 10^{9}$ sec^{-1}
Current	I	1 ampere (A), (C sec^{-1})	$= 3 \times 10^{9}$ statamperes
Current density	J	1 A m^{-2}	$= 3 \times 10^{5}$ statamp cm^{-2}
Displacement	D	1 C m^{-2}	$= 12\pi \times 10^{5}$ statvolt cm^{-1} 3×10^{5} statcoul cm^{-2}
Electric field	E	1 volt m^{-1} (V m^{-1})	$= 1/3 \times 10^{-4}$ statvolt cm^{-1}
Energy	w, u	1 joule (J)	$= 10^{7}$ ergs
Force	F	1 newton (N)	$= 10^{5}$ dynes
Inductance	L	1 henry (H)	$= 1/9 \times 10^{-11}$ stathenry
Length	l	1 meter (m)	$= 10^{2}$ centimeters (cm)
Magnetic field	H	1 ampere-turn m^{-1}	$= 4\pi \times 10^{-3}$ oersted
Magnetic flux	ϕ, F	1 weber (Wb)	$= 10^{8}$ gauss cm^{2} (maxwells)
Magnetic induction	B	1 Wb m^{-2}	$= 10^{4}$ gauss
Magnetization	M	1 Wb m^{-2}	$= 1/4\pi \times 10^{4}$ gauss
Mass	m	1 kilogram (kg)	$= 10^{3}$ grams (g)
Polarization	P	1 C m^{-2}	$= 3 \times 10^{5}$ statcoul cm^{-2} $12\pi \times 10^{5}$ (statvolt cm^{-1})
Potential	Φ, V	1 volt (V)	$= 1/300$ statvolt
Power	P	1 watt (w)	$= 10^{7}$ ergs sec^{-1}
Resistance	R	1 ohm	$= 1/9 \times 10^{-11}$ sec cm^{-1}
Time	t	1 second (sec)	= 1 second (sec)
Work	W	1 joule (J)	$= 10^{7}$ ergs

*Except for exponents, all factors of 3 should be replaced by 2.997930 ± 0.000003 for accurate work.

2-5.3 BOUNDARY RELATIONS

In solving electromagnetic problems, it often is necessary to relate vectors in medium 1 with vectors in medium 2 across an interface with normal $\boldsymbol{n}$. These can be derived from Eqs. 2-71 through 2-74 or from Eqs. 2-78 through 2-81. For uncharged media, the relations are

$$(\boldsymbol{E}_2 - \boldsymbol{E}_1) \times \boldsymbol{n} = 0 \qquad (2\text{-}89)$$

$$(\boldsymbol{D}_2 - \boldsymbol{D}_1) \cdot \boldsymbol{n} = 0 \qquad (2\text{-}90)$$

$$(\boldsymbol{H}_2 - \boldsymbol{H}_1) \times \boldsymbol{n} = 0 \qquad (2\text{-}91)$$

$$(\boldsymbol{B}_2 - \boldsymbol{B}_1) \cdot \boldsymbol{n} = 0 \qquad (2\text{-}92)$$

For media with free charges present, an unusual case for infrared work, the charges at the interface, both moving and static, affect the boundary relations. The relations are

$$(\boldsymbol{E}_2 - \boldsymbol{E}_1) \times \boldsymbol{n} = 0 \qquad (2\text{-}93)$$

$$(\boldsymbol{D}_2 - \boldsymbol{D}_1) \cdot \boldsymbol{n} = \sigma \qquad (2\text{-}94)$$

$$(\boldsymbol{H}_2 - \boldsymbol{H}_1) \times \boldsymbol{n} = \boldsymbol{J}_S \qquad (2\text{-}95)$$

$$(\boldsymbol{B}_2 - \boldsymbol{B}_1) \cdot \boldsymbol{n} = 0 \qquad (2\text{-}96)$$

where σ is the surface charge density and J_S is the surface current density at the interface.

2-5.4 PLANE WAVES

To see the wave nature of Maxwell's Equation, substitute Eq. 2-79 into the curl of Eq. 2-80 thus obtaining

$$\nabla \times \quad (\nabla \times \boldsymbol{E}) = -\mu\epsilon \frac{\partial^2 \boldsymbol{E}}{\partial t^2} \qquad (2\text{-}97)$$

In one dimension, this is easily seen to be

$$-\frac{\partial^2 \boldsymbol{E}}{\partial x^2} = -\mu\epsilon \frac{\partial^2 \boldsymbol{E}}{\partial t^2} \qquad (2\text{-}98)$$

which is a wave equation with propagation velocity $1/\sqrt{\mu\epsilon}$.

The plane wave solution to Eq. 2-98 is

$$\boldsymbol{E} = \boldsymbol{E}_0 e^{j\boldsymbol{k}\cdot\boldsymbol{x} - j\omega t} \qquad (2\text{-}99)$$

where

$$k = \frac{\omega}{v} = \sqrt{\mu\epsilon}\,\omega \qquad (2\text{-}100)$$

$$j = \sqrt{-1}$$

By analogy, in three dimensions the plane wave solutions of $\boldsymbol{B}$ and $\boldsymbol{E}$ are:

$$\boldsymbol{E}(\boldsymbol{x},t) = \boldsymbol{\epsilon}_1 E_0 e^{j\boldsymbol{k}\cdot\boldsymbol{x} - j\omega t} \qquad (2\text{-}101)$$

$$\boldsymbol{B}(\boldsymbol{x},t) = \boldsymbol{\epsilon}_2 B_0 e^{j\boldsymbol{k}\cdot\boldsymbol{x} - j\omega t} \qquad (2\text{-}102)$$

where $\boldsymbol{\epsilon}_1$ and $\boldsymbol{\epsilon}_2$ are the unit vectors denoting the polarization. Substituting Eqs. 2-101 and 2-102 into Eqs. 2-78 and 2-81 gives

$$\boldsymbol{\epsilon}_1 \cdot \boldsymbol{k} = 0, \quad \boldsymbol{\epsilon}_2 \cdot \boldsymbol{k} = 0 \qquad (2\text{-}103)$$

Thus $\boldsymbol{E}$ and $\boldsymbol{B}$ are transverse waves, i.e., their polarizations are normal to the direction of propagation $\boldsymbol{k}$.

Substituting Eqs. 2-101 and 2-102 into Eq. 2-80 gives

$$j\left[(\boldsymbol{k} \times \boldsymbol{\epsilon})\boldsymbol{E} - \omega\,\epsilon\boldsymbol{B}\right] e^{j\boldsymbol{k}\cdot\boldsymbol{x} - j\omega t} = 0 \qquad (2\text{-}104)$$

Thus

$$\frac{\boldsymbol{k} \times \boldsymbol{\epsilon}_1}{k} = \boldsymbol{\epsilon}_2 \qquad (2\text{-}105)$$

and

$$B_0 = \frac{k}{\omega} E_0 = \sqrt{\mu\epsilon}\, E_0 \qquad (2\text{-}106)$$

Eq. 2-105 says that $\boldsymbol{\epsilon}_1$, $\boldsymbol{\epsilon}_2$, and $\boldsymbol{k}$ are mutually perpendicular.

2-5.5 POYNTING VECTOR

The Poynting vector $\boldsymbol{S}$ is essentially the energy flux. In an unchanged medium,

$$\boldsymbol{S} = \boldsymbol{E} \times \boldsymbol{H}^* = \frac{1}{\mu} \boldsymbol{E} \times \boldsymbol{B}^* \qquad (2\text{-}107)$$

where $*$ denotes the complex conjugate. Substituting Eqs. 2-101 and 2-102 gives

$$\boldsymbol{S} = \frac{1}{\mu}(\boldsymbol{\epsilon}_1 \times \boldsymbol{\epsilon}_2)\, E_0 B_0 = \frac{1}{\mu} E_0 B_0 \frac{\boldsymbol{k}}{k} \qquad (2\text{-}108)$$

Thus, the time average value S of $\boldsymbol{S}$ is

$$S = \frac{1}{2\mu} E_0 B_0 = \frac{1}{2}\sqrt{\frac{\epsilon}{\mu}}\, E_0^2 \qquad (2\text{-}109)$$

Note that since S is power per unit area and E_0 is in volts per unit length, $\sqrt{\frac{\mu}{\epsilon}}$ is an impedance.

2-5.6 REFRACTIVE INDEX

The index of refraction n is the ratio of the speed v of waves propagating through a real material relative to the speed c of waves propagating through free space. Thus

$$n = \frac{c}{v} = \frac{ck}{\omega} \qquad (2\text{-}110)$$

In a slightly conductive medium, the electromagnetic waves are damped due to joule heating of the medium. To show this, Eqs. 2-101 and 2-102 are rewritten as:

$$\boldsymbol{E}(\boldsymbol{x}, t) = \boldsymbol{\epsilon}_1 e^{-\beta\boldsymbol{n}\cdot\boldsymbol{x}}\, E_0 e^{j\alpha\boldsymbol{n}\cdot\boldsymbol{x} - j\omega t} \qquad (2\text{-}111)$$

$$\boldsymbol{B}(\boldsymbol{x}, t) = \boldsymbol{\epsilon}_2 e^{-\beta\boldsymbol{n}\cdot\boldsymbol{x}}\, B_0 e^{j\alpha\boldsymbol{n}\cdot\boldsymbol{x} - j\omega t} \qquad (2\text{-}112)$$

where $\boldsymbol{n}$ is a unit vector in the direction of $\boldsymbol{k}$, and

$$k = \alpha + j\beta \qquad (2\text{-}113)$$

Using Ohm's Law,

$$\boldsymbol{J} = \sigma \boldsymbol{E} \qquad (2\text{-}114)$$

for isotropic media, and substituting it into Eq. 2-72, one can solve Eqs. 2-71 through 2-74 with $\rho = 0$ to get

$$\left[k^2 - (\mu\epsilon\omega^2 + j\mu\sigma\omega)\right] \boldsymbol{E} = 0 \qquad (2\text{-}115)$$

Thus

$$k^2 = (\mu\epsilon\omega^2) + j(\mu\sigma\omega) = \mu\epsilon\omega^2\left(1 + j\frac{\sigma}{\epsilon\omega}\right). \tag{2-116}$$

Therefore

$$\left.\begin{matrix}\alpha\\ \beta\end{matrix}\right\} = \sqrt{\mu\epsilon}\,\omega\left[\frac{\sqrt{1+\left(\frac{\sigma}{\epsilon\omega}\right)^2}\pm 1}{2}\right]^{1/2} \tag{2-117}$$

Sometimes, the index of refraction is considered the complex quantity. The index of refraction is then written as

$$n(1 - j\kappa) = \frac{c}{\omega} = (\alpha + j\beta) \tag{2-118}$$

where κ is the *absorption index*. Therefore

$$n = \frac{c\alpha}{\omega} = c\sqrt{\mu\epsilon}\left[\frac{\sqrt{1+\left(\frac{\sigma}{\epsilon\omega}\right)^2}+1}{2}\right]^{1/2} \tag{2-119}$$

$$\kappa = \frac{-\beta}{\alpha} = \frac{-\frac{\sigma}{\epsilon\omega}}{\sqrt{1+\left(\frac{\sigma}{\epsilon\omega}\right)^2}+1} \tag{2-120}$$

2-5.7 SNELL'S LAW

Consider light incident at an angle i upon an interface of two media, 1 and 2, with the incident light being in medium 2. (See Fig. 2-22.) For the *incident* light, both the $\boldsymbol{E}_i$ vector and the $\boldsymbol{B}_i$ vector vary as

$$\left.\begin{matrix}\boldsymbol{E}_i\\ \boldsymbol{B}_i\end{matrix}\right\} \sim e^{j\boldsymbol{k}_i\cdot\boldsymbol{x}_i - j\omega t} \tag{2-121}$$

Similarly, the *refracted* light varies as

$$\left.\begin{matrix}\boldsymbol{E}_r\\ \boldsymbol{B}_r\end{matrix}\right\} \sim e^{j\boldsymbol{k}_r\cdot\boldsymbol{x}_r - j\omega t} \tag{2-122}$$

Taking the x-direction parallel to the interface, the y-direction also parallel but normal to the direction of incidence, and the z-direction normal to the interface, then Eqs. 2-121 and 2-122 can be written explicitly as

$$\left.\begin{matrix}\boldsymbol{E}_i\\ \boldsymbol{B}_i\end{matrix}\right\} \sim \exp\left[j\frac{\omega}{v_2}(x\sin i + z\cos i) - j\omega t\right] \tag{2-123}$$

$$\left.\begin{matrix}\boldsymbol{E}_r\\ \boldsymbol{B}_r\end{matrix}\right\} \sim \exp\left[j\frac{\omega}{v_1}(x\sin r + z\cos r) - j\omega t\right] \tag{2-124}$$

At the interface $z = 0$, the incident vectors and the refracted vectors must have the same phases. (Otherwise the boundary relations, Eqs. 2-89 to 2-92, could not possibly be satisfied everywhere on the interface.) Therefore

$$\frac{\sin i}{v_2} = \frac{\sin r}{v_1}. \tag{2-125}$$

Rewriting Eq. 2-125 in terms of η_2 and η_1 gives Snell's Law:

$$\frac{\sin i}{\sin r} = \frac{\eta_1}{\eta_2} \tag{2-126}$$

If $\sin i > \eta_1/\eta_2$, then $\sin r > 1$ which cannot be. Therefore, for $\sin i > \eta_1/\eta_2$ there is no refracted light beam. This is called total internal reflection.

2-5.8 REFLECTION COEFFICIENT

In the notation of par. 2-5.7, the *reflected* vectors are of the form

$$\left.\begin{matrix}\boldsymbol{E}_{i'}\\ \boldsymbol{B}_{i'}\end{matrix}\right\} \sim \exp\left[j\frac{\omega}{v_2}(x\sin i' - z\cos i') - j\omega t\right] \tag{2-127}$$

At $z = 0$, this expression must match Eq. 2-123. Thus

$$i' = i \tag{2-128}$$

To find the amplitudes of reflected vectors $\boldsymbol{E}_{i'}$ and $\boldsymbol{B}_{i'}$, the boundary relations, Eq. 2-89 to Eq. 2-92, must be used. There are two distinct cases that must be considered; when the electric vector $\boldsymbol{E}$, of the incident light is (1) in the plane of incidence, and (2) perpendicular to the plane of incidence. Fig. 2-22 is drawn in the plane of incidence.

The results are [62,63,64].

(1) $\boldsymbol{E}_i$ parallel to the plane of incidence:

$$\frac{E_{i'}}{E_i} = \frac{\left(\frac{\mu_2}{\mu_1}\right)\sin 2i - \sin 2r}{\sin 2r + \left(\frac{\mu_2}{\mu_1}\right)\sin 2i},$$

$$\frac{E_r}{E_i} = 2\sqrt{\frac{\mu_2\epsilon_2}{\mu_1\epsilon_1}}\,\frac{\sin 2i}{\sin 2r + \left(\frac{\mu_2}{\mu_1}\right)\sin 2i} \tag{2-129}$$

(2) $\boldsymbol{E}_i$ perpendicular to the plane of incidence:

$$\frac{E_i'}{E_i} = \frac{1 - \frac{\mu_2 \tan i}{\mu_1 \tan r}}{1 + \frac{\mu_2 \tan i}{\mu_1 \tan r}},$$

$$\frac{E_r}{E_i} = \frac{2}{1 + \frac{\mu_2 \tan i}{\mu_1 \tan r}} \qquad (2\text{-}130)$$

The B vectors can be obtained from Eqs. 2-105 and 2-106 for each medium.

In general, for infrared optical materials, $\mu_2 = \mu_1$. i.e., infrared materials are usually non-magnetic. Then Eqs. 2-129 and 2-130 reduce to:

(1) *Case 1:*

$$\frac{E_i'}{E_i} = \frac{\tan (i - r)}{\tan (i + r)},$$

$$\frac{E_r}{E_i} = \frac{2 \cos i \sin r}{\sin (i + r) \cos (i - r)} \qquad (2\text{-}131)$$

(2) *Case 2:*

$$\frac{E_i'}{E_i} = \frac{\sin (i - r)}{\sin (i + r)},$$

$$\frac{E_r}{E_i} = \frac{2 \cos i \sin r}{\sin (i + r)} \qquad (2\text{-}132)$$

For normal incidence, $i = r = 0$. Then both cases reduce to

$$\frac{E_i'}{E_i} = \frac{\eta_1 - \eta_2}{\eta_1 + \eta_2}, \quad \frac{E_r}{E_i} = \frac{2\eta_2}{\eta_1 + \eta_2} \qquad (2\text{-}133)$$

where the incident light is in medium 2.

The reflectivity ρ is defined as

$$\rho = \frac{S_i'}{S_i} = \left(\frac{E_i'}{E_i}\right)^2 \qquad (2\text{-}134)$$

Thus, for $\mu_2 = \mu_1$,

$$\rho_{\parallel} = \frac{\tan^2 (i - r)}{\tan^2 (i + r)}, \quad \rho_{\perp} = \frac{\sin^2 (i - r)}{\sin^2 (i + r)} \qquad (2\text{-}135)$$

For unpolarized light, i.e., light with an equal probability per unit time of E_i being parallel or perpendicular to the plane of incidence,

$$\rho = \frac{1}{2} (\rho_{\parallel} + \rho_{\perp}) \qquad 2\text{-}136$$

If Eq. 2-135 is substituted into Eq. 2-136, then Eq. 2-136 becomes the Fresnel formula for reflectivity.

For normal incidence, $i = r = 0$. Thus from Eq. 2-133,

$$\rho = \left(\frac{\eta_1 - \eta_2}{\eta_1 + \eta_2}\right)^2 \qquad (2\text{-}137)$$

2-5.9 BREWSTER'S ANGLE

There is the possibility of incident light being refracted with no reflected wave. From Eq. 2-131 it is clear that

$$\frac{E_i'}{E_i} = 0 \text{ for } i + r = \pi/2 \qquad (2\text{-}138)$$

for incident light with the electric vector E_i parallel to the plane of incidence. This particular angle of incidence is called the Brewster's angle i_B. From Snell's Law, Eq. 2-126, the Brewster's angle is

$$i_B = \arctan\left(\frac{\eta_1}{\eta_2}\right) \qquad (2\text{-}139)$$

Since at $i = i_B$ the parallel component is entirely transmitted, the reflected component must be entirely perpendicular. Thus, reflection off a surface at the Brewster's angle is a means of obtaining linearly polarized light, or light with the electric vector E_i, in a single, fixed direction.

2-5.10 POLARIZATION

A light wave is linearly (or plane) polarized if the end of the electric vector moves in a straight line (the electric vector moves in one plane). A light wave is circularly (elliptically) polarized if the end of the electric vector moves in a circle (ellipse). If, when looking along the direction of propagation from a fixed point in space, the electric vector rotates clockwise it is right-handed: if counterclockwise, then it is left-handed.

This is described mathematically by separating the electric vector into components as

$$E = (E_+ \epsilon_+ + E_- \epsilon_-) e^{j\mathbf{k}\cdot\mathbf{x} - j\omega t} \qquad (2\text{-}140)$$

where

$$\epsilon_{\pm} = \frac{1}{\sqrt{2}} (\epsilon_1 \pm j\epsilon_2) \qquad (2\text{-}141)$$

and ϵ_1, ϵ_2 are orthogonal and perpendicular to $\epsilon_3 = \frac{k}{|k|}$, the direction of propagation of E. The standard types of polarization then are

linear:	$E_+ = E_-$
right circular:	$E_+ = 0$
left circular:	$E_- = 0$
right elliptic:	$E_+ < E_-$
left elliptic:	$E_- < E_+$

2-5.11 ANTI-REFLECTION COATINGS

In par. 2-5.5 it was noted that $\sqrt{\frac{\mu}{\epsilon}}$ had the dimensions of an impedance. With this analogy to transmission lines established, one can insert a quarter-wave matching section between the two media of Fig. 2-22 and obtain a proper impedance match with no reflections. The three impedances are

$$\eta_2 = \sqrt{\frac{\mu_2}{\epsilon_2}},\ \eta_1 = \sqrt{\frac{\mu_1}{\epsilon_1}},\ \eta_3 = \sqrt{\frac{\mu_3}{\epsilon_3}}. \qquad (2\text{-}142)$$

To be correctly matched,

$$\eta_3 = \sqrt{\eta_2 \eta_1} \qquad (2\text{-}143)$$

The quarter wave refers to one-quarter wavelength in medium 3.

This approach is discussed in Ref. 63, Chapter 7.

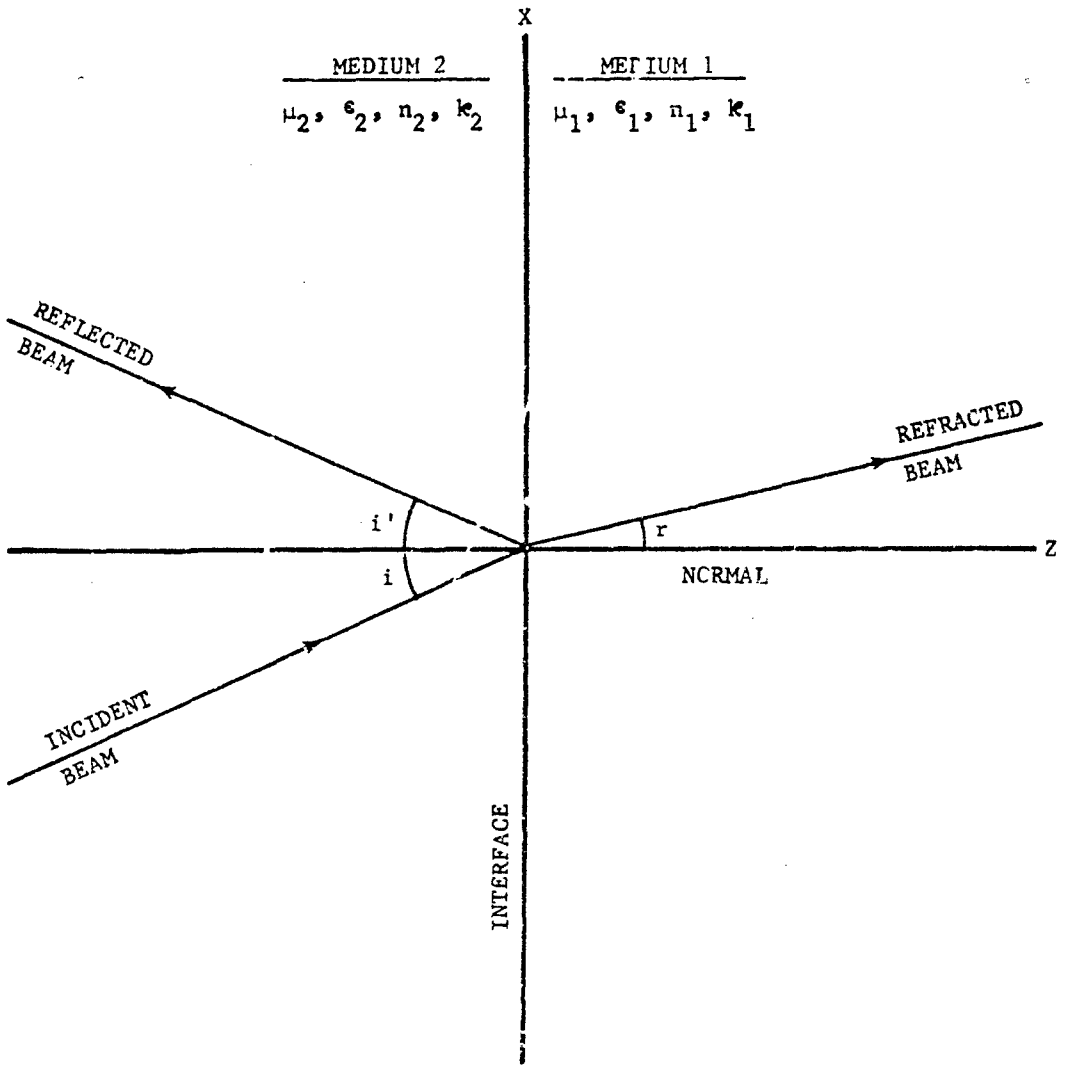


FIGURE 2-22. Incident, Refracted, and Reflected Light Beams at the Interface of Two Media (drawn in the plane of incidence)

2-6.12 DIELECTRIC WAVEGUIDES AND CAVITIES

For most applications, the sizes and/or tolerances of the boundaries of the media through which the infrared light is propagating are, in general, large compared to the wavelengths of the light. Two important exceptions are dielectric waveguides (optical fibers) and laser cavities.

Infrared light propagates down a dielectric waveguide by internal reflections off the outside surface of the dielectric fiber. This requires a nonabsorbing dielectric with a high index of refraction. The nonabsorbing characteristic is required so that the transmission will be high. The high index of refraction is needed so that the light must hit a surface at a small angle to the normal in order to escape.

If two dielectric waveguides are placed within about one wavelength (in the external medium) of each other, then there will be radiative coupling and energy will be lost from one fiber to the other. In optical fiber imaging devices, this is one source of optical crosstalk.

The study of dielectric waveguides is basically by analogy with conventional waveguides. Maxwell's Equations are solved subject to the boundary conditions of the guide. Modes are determined, propagation modes numbered, etc. Ref. 62, Chapter 8, contains a good introductory discussion.

In a relatively short length of dielectric waveguide, if the ends are reflecting, then the guide becomes a dielectric, or optical, cavity. Optical cavities form the resonant circuits of lasers.

2-6.13 SPECTRAL TRANSMISSION OF OPTICAL MATERIALS

Typical spectral transmission curves for representative, useful IR optical materials are shown in Fig. 2-23. Additional information on these and other materials is given in Fig. 2-24.

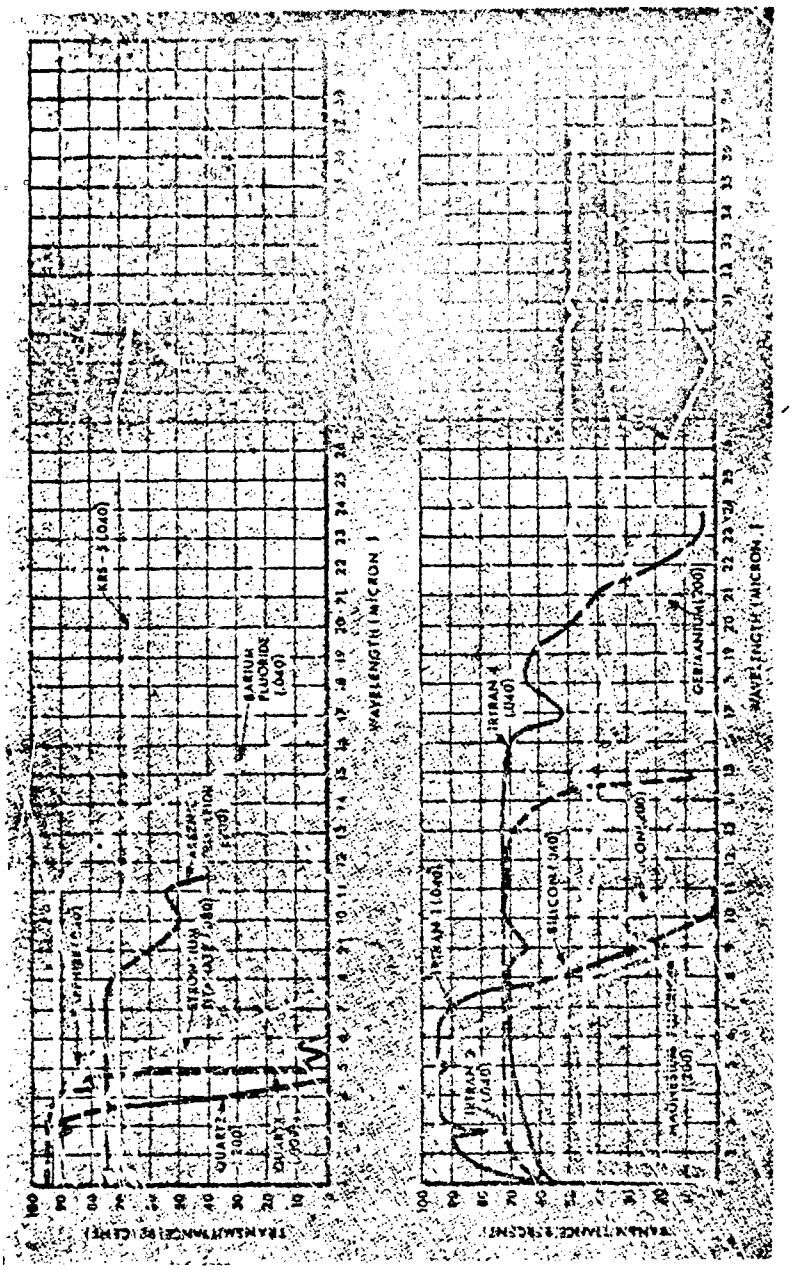
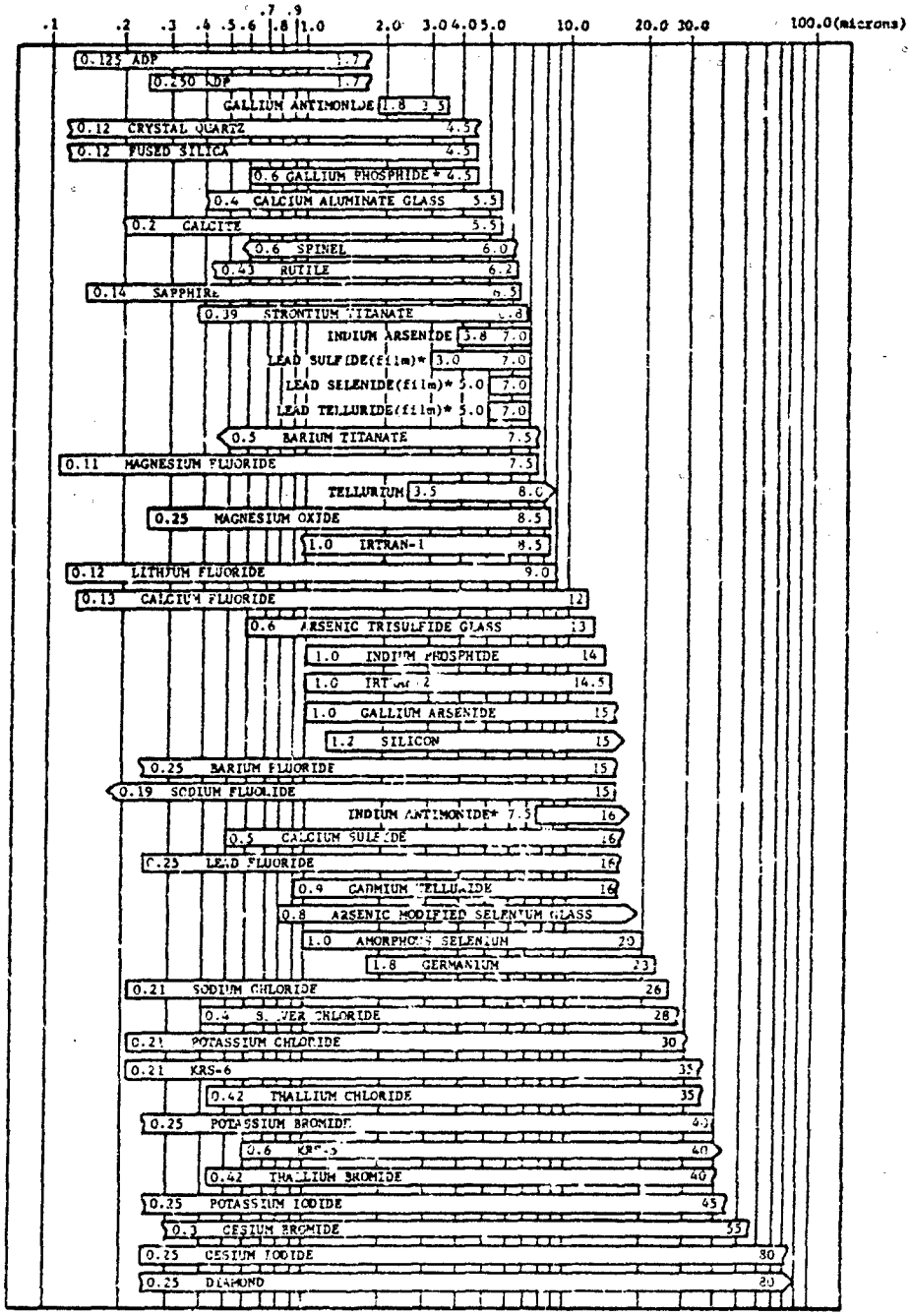


FIGURE 2-13. Transmission Curves for Representative IR Optical Materials



*Maximum external transmittance of less than 10%

Transmission regions of optical materials, 2 mm thickness; cutoff is defined as 10 percent external transmittance, and materials marked with an asterisk never have external transmittance as high as 10 percent.

FIGURE 2-24. Comparisons of Material Properties

2-6 SOURCES OF NATURAL RADIATION

2-6.1 BACKGROUND RADIATION

Background radiation can be due to self-emission and reflected or scattered radiation from terrain, sea surface, the atmosphere, and atmospheric objects (aerosols) or celestial objects, or it can be due to various combinations of these.

The gross features of background radiation from earth, air, and sea sources are shown in an idealized fashion in Fig. 2-25. Below 3 microns, the background is dominated by reflected and scattered solar radiation, In this region then, the spectral distribution of radiation will approximate that of a 6000°K blackbody, and the actual radiance will depend upon the reflection and scattering properties of the particular background. Beyond 4.5 microns, the background is dominated by thermal emission from earth and air objects (near 300°K). Between 3 and 4.5 microns, background radiation is at a minimum. The radiance of the earth beyond 5 microns, as seen from a balloon at 37 km, is shown in Fig. 2-26. Atmospheric effects are seen to have a strong influence on the shape of the radiance curve beyond 5 microns. The radiance is larger in the atmospheric window (8 to 14 microns) than in the atmospheric absorption bands at 6 to 15 microns. This is because the earth is warmer than the surrounding atmosphere; therefore, emission from the atmosphere in its strong emission bands (at 6- and 15-micron bands) is weaker than emission of radiation from the earth, which escapes with hardly any attenuation in the 8- to 14-micron window.

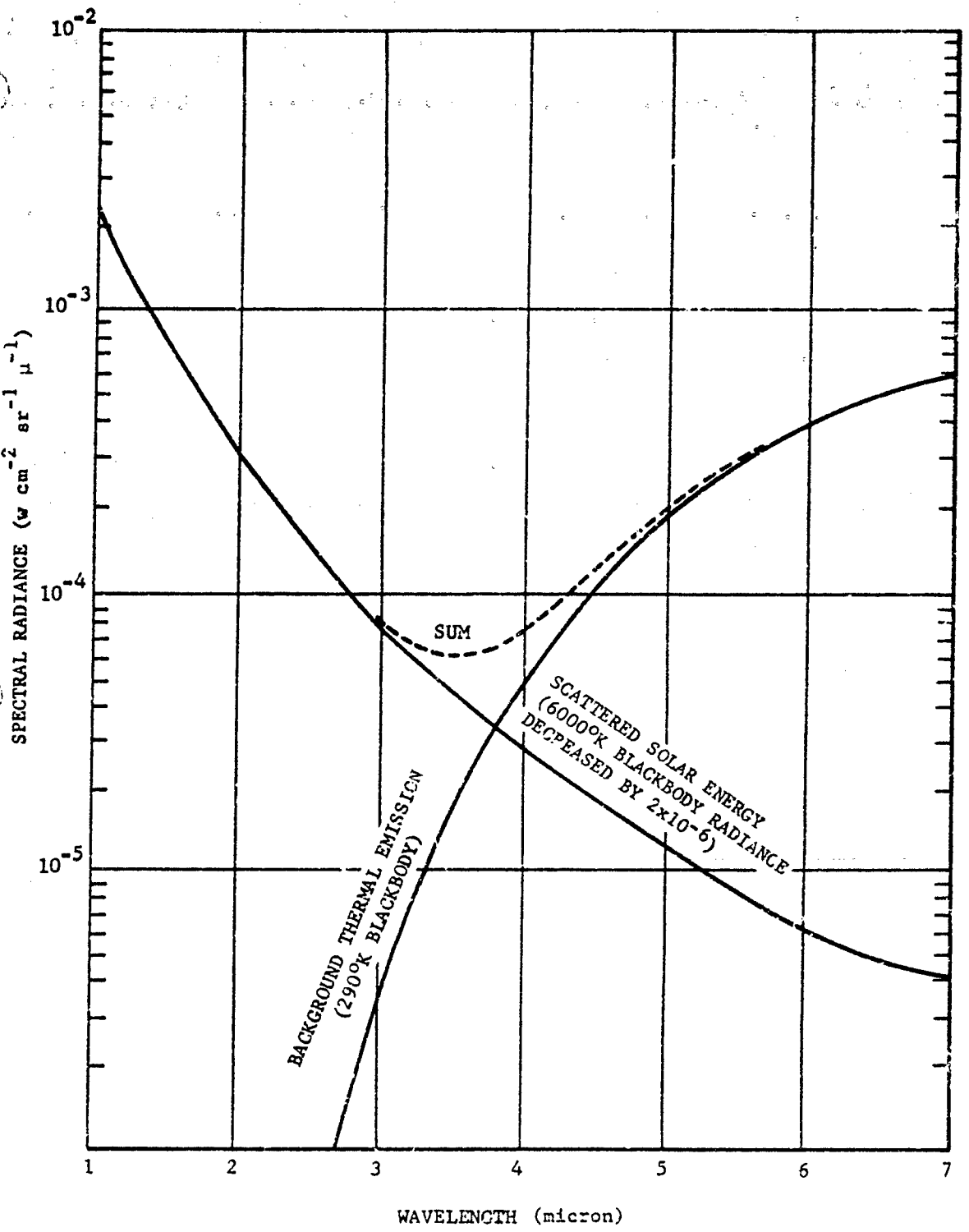


FIGURE 2-25. Idealized Background Spectrum

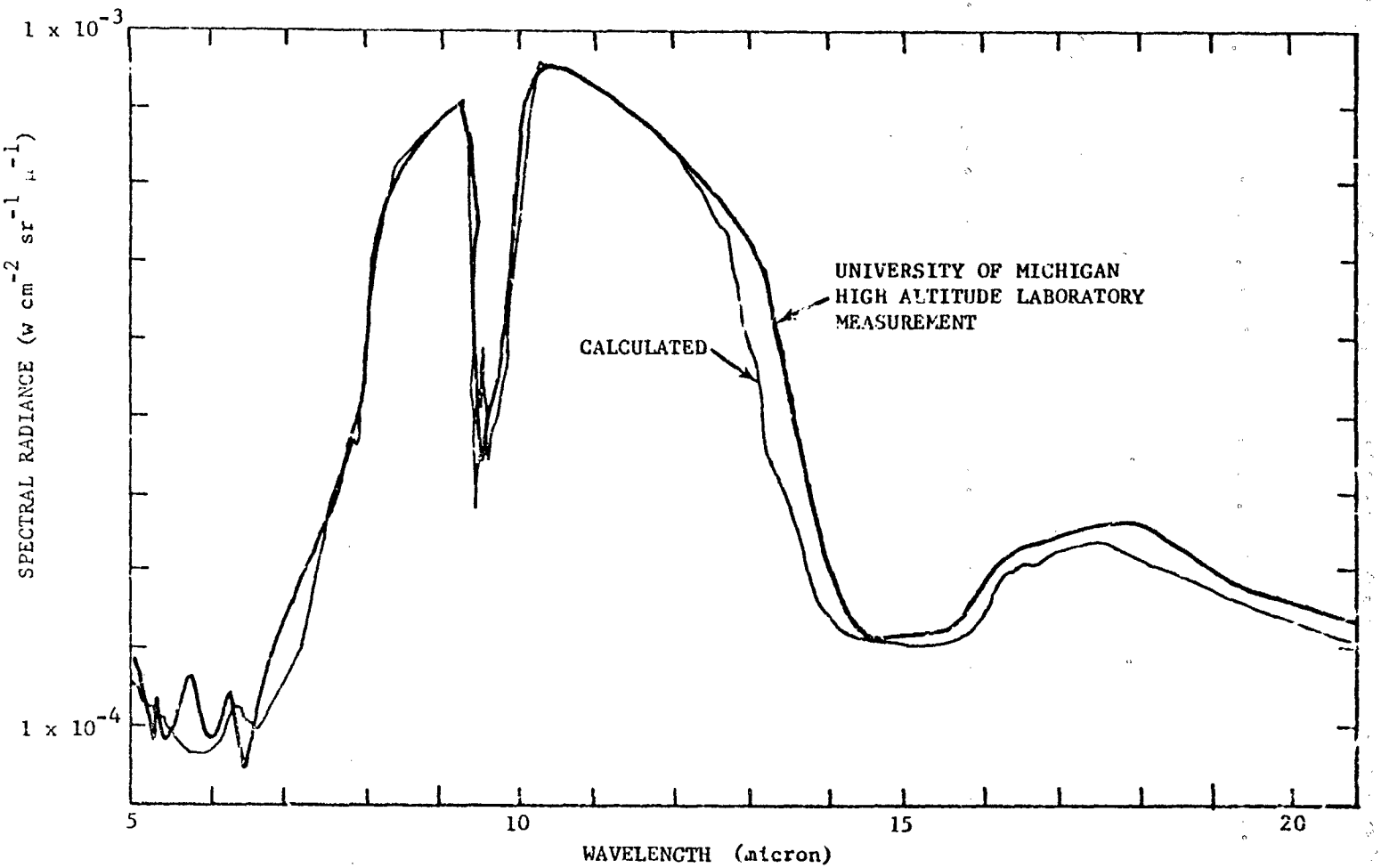


FIGURE 2-26. Comparison of Calculated Earth Radiance with Measurements for Downlooking System at 37 km Altitude

2-6.1.1 Terrain

Terrain radiation in the daytime and at wavelengths shorter than 4 microns is dominated by sunlight and by the reflectivity of the objects constituting the terrain. The reflectivity of background objects at wavelengths shorter than 3 microns ranges from 0.03 (bare ground) to 0.95 (fresh snow)[65,68] . Beyond 4 microns the radiation from the terrain is dominated by terrain self-emission which depends on the temperature of the terrain objects and on their emissivities. Table 2-13 lists emissivity and reflectance values of some typical background objects in the IR[65] . This table clearly shows that most terrain objects beyond 3 microns will have an emissivity larger than 0.8. Table 2-14 lists infrared emissivities of snow and ice particles[68] . In the daytime the temperature of the background objects is related to their optical properties in the visible and IR regions (6 to 15 microns), their thermal contact with the air, and their heat conductivity and capacity[69] . Assuming that the maximum irradiance of a background object due to sunlight is 0.10 w cm^{-2}, and that the background object is a perfect absorber in the visible and a perfect emitter in the IR, the object can be expected to reach a temperature of 90°C. In reality since few objects face directly into the sun and because of heat conduction in the body and thermal contact with the surrounding air, a maximum temperature of 50°C is not at all unusual. Assuming that the surrounding air temperature is 15°C, the expected maximum contrast of a background is shown in Fig. 2-27. (The contrast can be somewhat greater in the 8- to 12-micron window where an object with high reflectivity and, therefore, low emissitivity would reflect the cold radiation from the sky but make no thermal contribution of its own.) The cooling rate of background objects at nighttime will depend on their heat capacity, heat conductivity, thermal contact with the surrounding air, IR emissivity in the 8- to 12-micron band, atmospheric humidity, and cloud cover[69,70] . In a very dry location when there is no cloud cover, all the thermal radiation in the 8- to 12-micron band will be radiated into space and objects will cool down rapidly. Vegetation which is in close contact with the surrounding air, and water surfaces (lakes, rivers, ocean) which have large heat capacities will radiate fairly evenly during the day and night. Fig. 2-28 shows typical diurnal variation in radiance from various terrain backgrounds[71] . Daytime spectral radiances from 1 to 6 microns of various terrain features[72] are shown in Fig. 2-29, which also shows a large spread in radiance values at wavelengths shorter than 3 microns where scattering of solar radiation predominates. Beyond 4 microns the radiance values of the various terrain features differ little. Concrete and a brick wall have a higher temperature than grass and radiate more at wavelengths longer than 4 microns. At wavelengths shorter than 3 microns snow is an excellent scatterer of solar radiation and gives the highest radiance values while grass, on the contrary, has the smallest reflectivity to sunlight at wavelengths shorter than 3 microns and yields the smallest radiance values.

Relative radiance measurements are often of more importance to system designers than the measurement of radiance itself. Of particular importance, will be to know the times during the day and night when the spectral radiances of different terrain features are identical, and when contrasts are therefore at a minimum. These times are called crossover times. Measurements of crossover times of various terrain features considered two by two have been made and these times are found to vary widely depending on the objects considered[70] . Several examples are presented in Figs. 2-30, 2-31, and 2-32. Contrasts in the spectral region extending beyond 4 microns are at a minimum on very cloudy and on rainy days.

What is considered as a background or as terrain to one person may be considered as a target or a field of operations to another. It is, therefore, not surprising that a considerable amount of measurements and information on the infrared radiance of terrain is classified. Among this classified information are infrared photographs of terrain during the day and night at low and high altitudes, studies of the diurnal and seasonal variation of radiation from terrain, and studies of the seasonal dependence of the crossover points[70,73-79] .

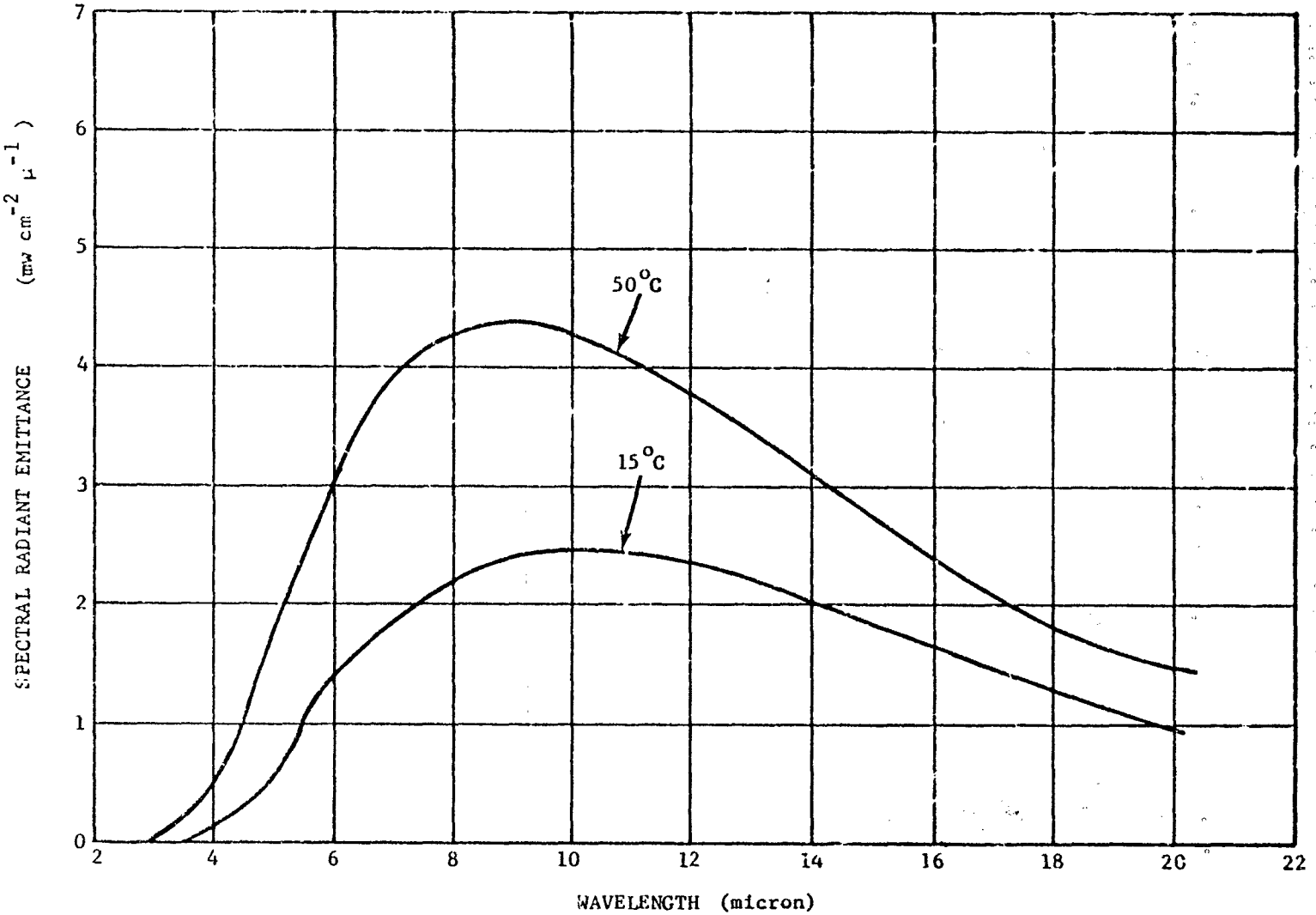
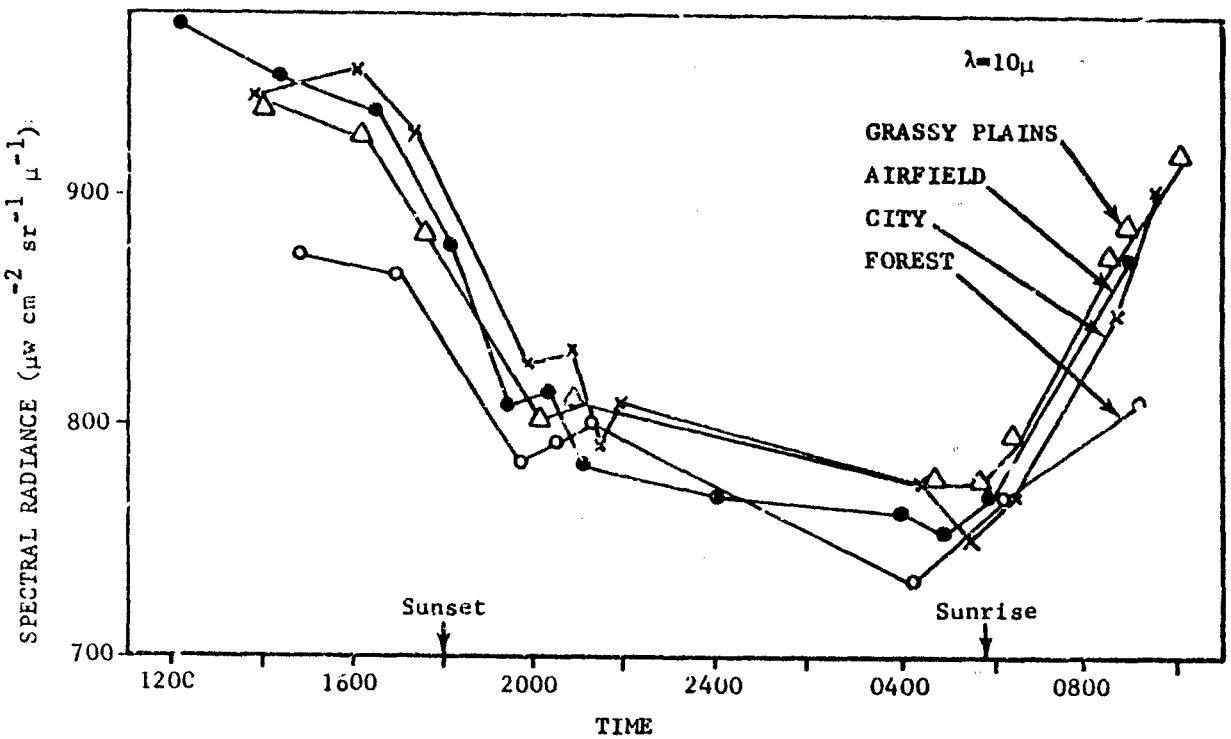


FIGURE 2-27. Radiance of 50° and 15°C Blackbodies



September - Pikes Peak, Colorado

Line-of-Sight Distances:

Grassy Plains	- 21 Miles
Airfield	- 19 Miles
City	- 15 Miles
Forest	- 30 Miles

FIGURE 2-28. Hourly Variation in the 10-micron Radiance of Various Backgrounds

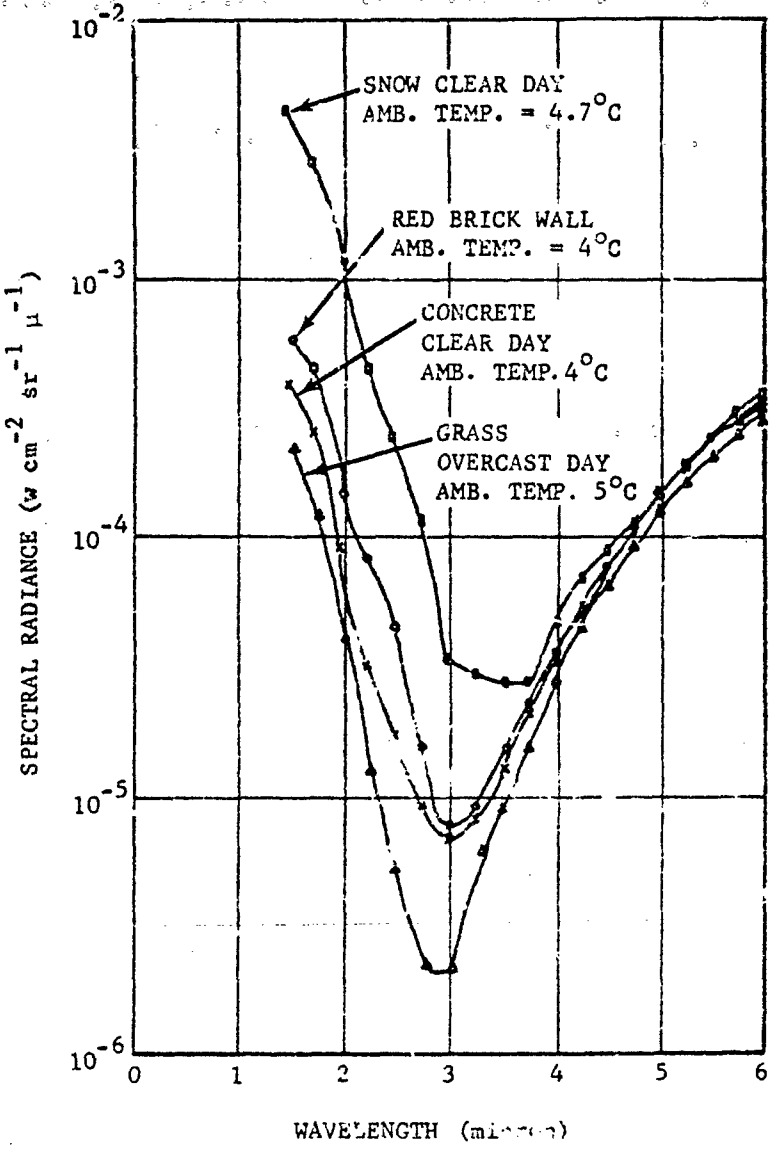


FIGURE 2-29. Daytime Spectral Radiance of Miscellaneous Targets

TABLE 2-13. REFLECTANCE ρ AND EMISSIVITY ϵ OF COMMON TERRAIN FEATURES*

	0.7 - 1.0 μ	1.8 - 2.7 μ	3 - 5 μ	8 - 13 μ
Green Mountain Laurel	ρ = 0.44	ϵ = 0.84	ϵ = 0.90	ϵ = 0.92
Young Willow Leaf (dry, top)	0.46	0.82	0.94	0.96
Holly Leaf (dry, top)	0.44	0.72	0.90	0.90
Holly Leaf (dry, bottom)	0.42	0.64	0.86	0.94
Pressed Dormant Maple Leaf (dry, top)	0.53	0.58	0.87	0.92
Green Leaf Winter Color — Oak Leaf (dry, top)	0.43	0.67	0.90	0.92
Green Coniferous Twigs (Jack Pine)	0.30	0.86	0.96	0.97
Grass — Meadow Fescue (dry)	0.41	0.62	0.82	0.88
Sand — Hainamanu Silt Loam — Hawaii	0 15	0.82	0.84	0.94
Sand — Barnes Fine Silt Loam — So Dakota	0.21	0.58	0.78	0.93
Sand — Gooah Fine Silt Loam — Oregon	0.39	0.54	0.80	0.98
Sand — Vereiniging — Africa	0.43	0.56	0.82	0.94
Sand — Maury Silt Loam — Tennessee	0.43	0.56	0.74	0.95
Sand — Dublin Clay Loam — California	0.42	0.54	0.83	0.97
Sand — Pullman Loam — New Mexico	0.37	0.62	0.78	0.93
Sand — Grady Silt Loam — Georgia	0.11	0.58	0.85	0.94
Sand — Colts Neck Loam — New Jersey	0.28	0.67	0.90	0.94
Sand — Mesita Negra — lower test site	0.38	0.70	0.75	0.92
Bark — Northern Red Oak	0.23	0.78	0.90	0.96
Bark — Northern American Jack Pine	0.18	0.69	0.88	0.97
Bark — Colorado Spruce	0.22	0.75	0.87	0.94

* Estimated average values of reflectance ρ, or emissivity ϵ, = 1 - ρ.

TABLE 2-14. TOTAL EMISSIVITY OF SNOW AND ICE PARTICLES

MATERIAL	TEMP, °F	EMISSIVITY	NO. OF MEASUREMENTS
Snow			
Fine particles similar to frost particles in size.			
(1) Depth 3/8 in. on aluminum foil	11.5 - 19.5	0.825	20
(2) Depth 3/8 in. on aluminum foil	15 - 19.5	0.820	9
Ice			
Plane sheet 1/16-in. thick	8.5	0.964	2
Artificial	0.8 - 18.9	0.97	10
Coarse particles 1/32-in. diameter (approx) granular snow	13.0 - 21.0	0.887	12
Artificial ice crushed and refrozen particle size approx 1/16-in. diameter	12.9	0.950	4
Frost			
Fine	0 - 20.0	0.985	—

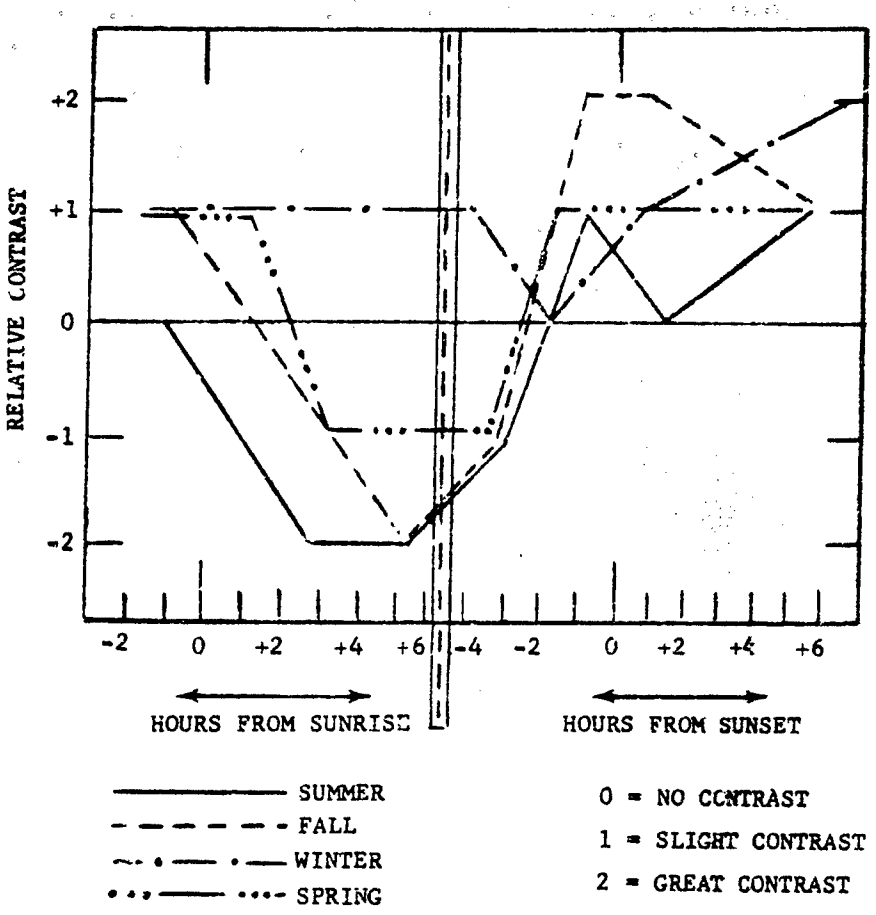


FIGURE 2-30. Relative Contrast of Deciduous Trees to Short Grass

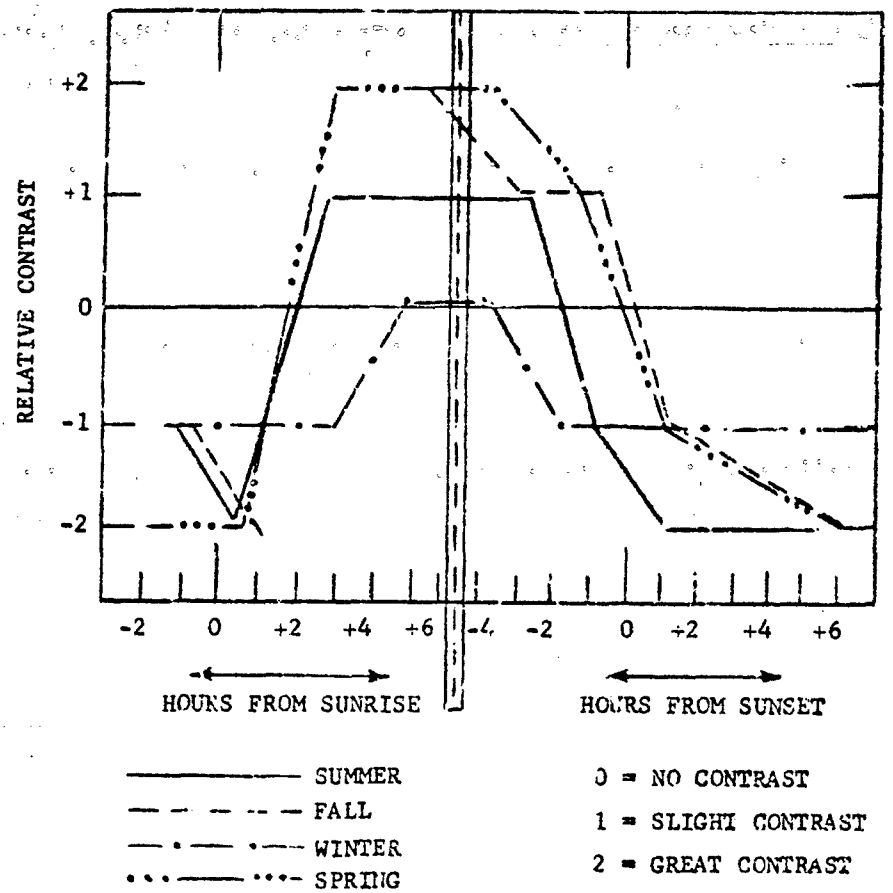


FIGURE 2-31. Relative Contrast of Wooden Catwalk Atop a Dam to the River Upstream from the Dam

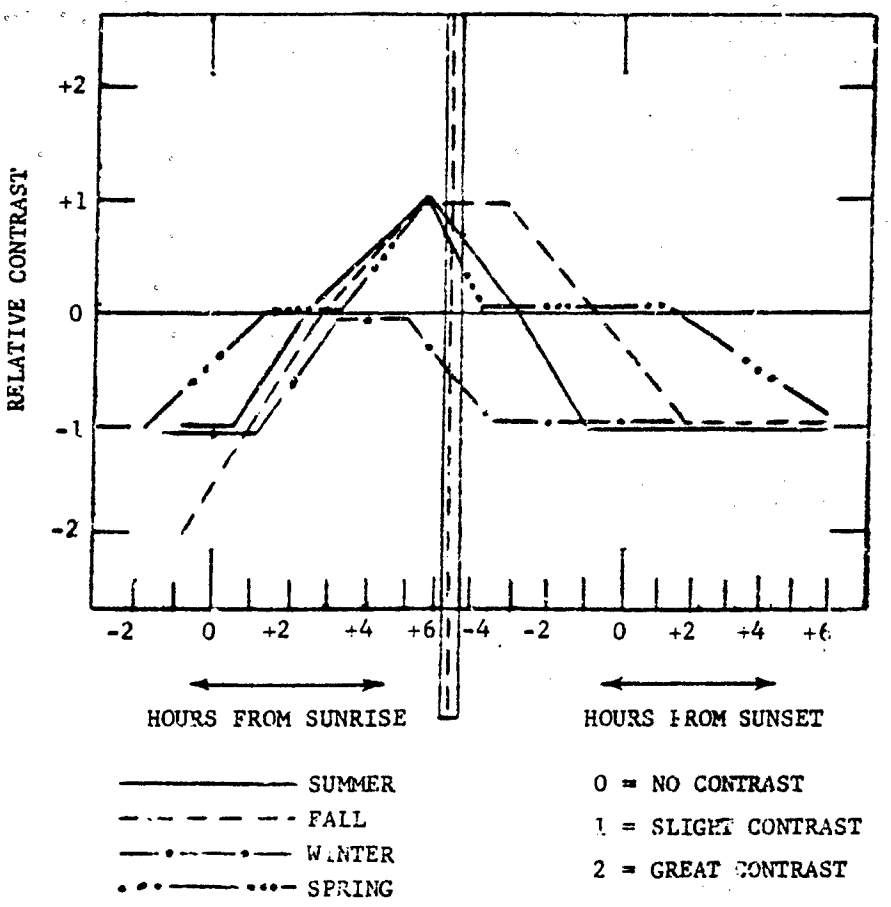
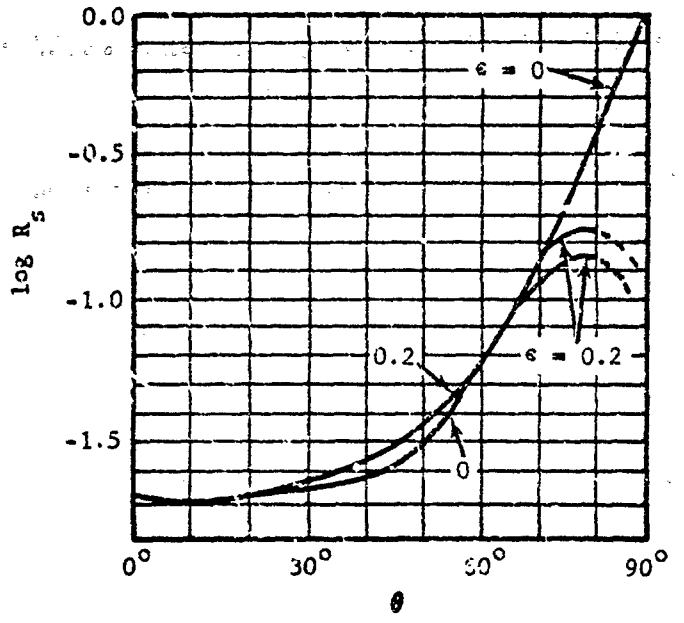


FIGURE 2-32. Relative Contrast of a Small Asphalt-surfaced Concrete Bridge to an Asphalt Roadway

2-6.1.2 Sea

The gross features of the spectral radiance curve for the sea background is similar to the background spectrum shown in Fig. 2-25. Beyond 4 microns the radiation is due to thermal emission and, because of the opaqueness of water[80], the effective blackbody temperature will be that of the first few millimeters of the sea surface. The emissivity of the sea is substantially unity for all viewing situations except those in which the line of sight is nearly parallel to the sea surface[1]. Below 4 microns, sea background radiance is due to reflected sky radiance. The reflectivity of the ocean (albedo), as measured by various satellites, has been found to be about 7 percent[81,82,83]. It should be noted, however, that reflectivity and emissivity, particularly near the horizon, will be a function of the smoothness of the water surface. This can be seen in Fig. 2-33[84] which gives the albedo of the sea versus solar angle. Since the average reflective properties of water in the 2 to 15-micron band approximate closely those in the visible, Fig. 2-33 can be used to estimate reflected sky radiance from the sea in the infrared. Fig. 2-34 shows the reflection coefficient versus wavelength of a plane water surface for various angles of incidence. This coefficient was calculated from normal reflectivity data.



NOTE:

Reflection of solar radiation from flat surface, $\epsilon = 0$.

Reflection from surface roughened by a Beaufort 4 wind: $\epsilon = 0.2$

Albedo R varies from 0.02 for a zenith sun, $\theta = 0^{o}$, to unity for sun at horizon $\theta = 90^{o}$ on flat sea surface.

For a rough surface, shadowing and multiple reflections become important factors when sun is low.

The upper and lower branches of curve $\epsilon = 0.2$ represent assumptions regarding the effect of multiple reflection.

True values are expected to lie between the indicated limits.

FIGURE 2-33. Reflectivity of Solar Radiation vs Sun Angle for Various Sea-surface Roughnesses

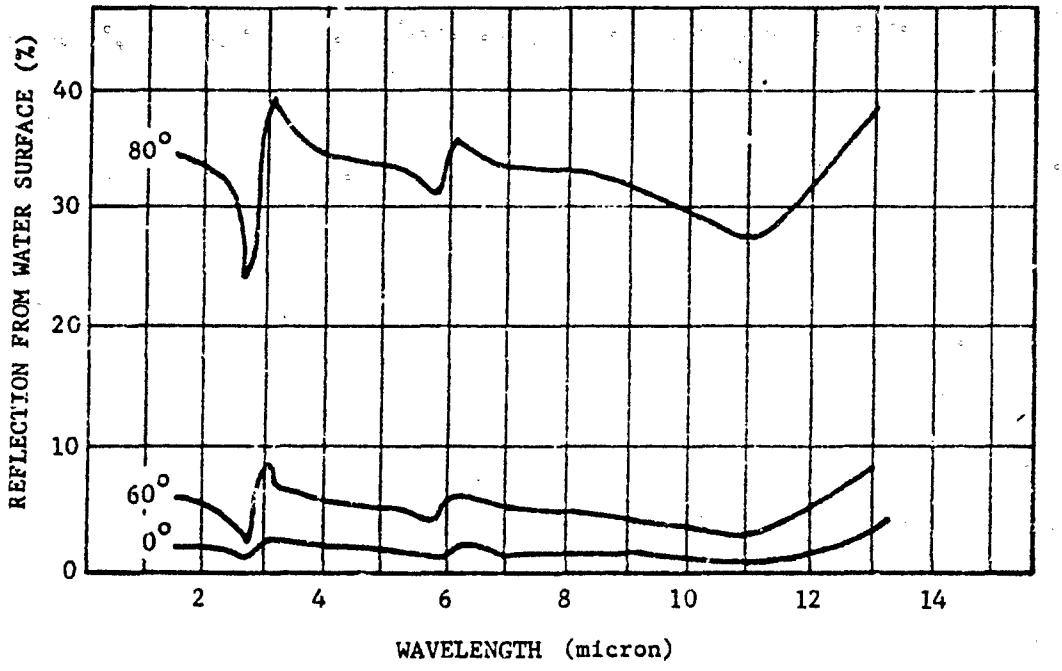


FIGURE 2-34. Reflectance of a Water Surface at 0°, 60°, and 80° Angle of Incidence

2-6.1.3 Clouds

Clouds can affect the operations of optical systems in two important ways. The radiance of the clouds due to scattered or emitted radiation can be sufficient to prevent proper operation of the system, or the physical presence of the clouds can mask the radiation from the objects which the system was designed to detect.

There are ten main classifications of clouds: cirrus, cirrocumulus, cirrostratus, altocumulus, altostratus, nimbostratus, stratocumulus, stratus, cumulus, and cumulonimbus, These ten genera together with their characteristic cloud base heights, thicknesses, and microstructures are listed in Table 2-15[85] . It should be noted that the cloud base heights listed in the table vary somewhat with altitude, particularly for high clouds in polar regions where base heights can be 2 to 3 km lower than those listed[1] . Cloud top heights are extremely variable, particularly for cumulonimbus clouds, which occasionally reach to the tropopause[86] .

The general envelope of cloud radiance is illustrated in Fig. 2-25. Below 4 microns, clouds reflect the incident sunlight, and appear as 6000°K graybodies. This radiation will, of course, be modified by atmospheric, liquid water, and ice crystal absorption. Beyond 4 microns cloud radiation is principally due to the thermal radiation of the cloud itself.

Table 2-16 gives albedos, the ratio of light reflected to that incident, for the 0.5- to 0.7-micron region for various cloud types, as measured by TIROS[87] . A typical spectrum for the 1.2- to 2.5-micron region, for a high altitude ice cloud is presented in Fig. 2-35. Minor absorption in the 1.4- and 1.8-micron region is due to water vapor, and the dips at 1.5 and 2.0 microns appear to be characteristic of ice crystal clouds but not of water droplet clouds[88] . A typical cloud spectrum for the 2.5- to 3.5-micron region is given in Fig. 2-36. Atmospheric absorption can be noticed at 2.68 and around 3.3 microns and, in conjunction with ice absorption, at 2.8 microns.

TABLE 2-15. CLOUD TYPES AND CHARACTERISTICS

TYPE	BASE HEIGHT, km	THICKNESS	MICROSTRUCTURE
High Clouds			
1. Cirrus (Ci)	7 - 10	Hundreds of meters to several kilometers	Columnar crystals
2. Cirrocumulus (Cc)	6 - 8	0.2 - 0.4 km	Columnar crystals, hollow prisms, separate or combined in complexes.
3. Cirrostratus (Cs)	6 - 8	0.1 - several kilometers	Cubic crystals, sometimes combined in complexes. Occasionally thick plates.
Middle Clouds			
4. Altocumulus (Ac)	2 - 6	0.2 - 0.7 km	5 - 7μ radius droplets with fluctuations between 3 and 24μ, rarely crystalline
5. Altostratus (As)	3 - 5	1 - 2 km	Ice crystals and water droplets mixed. Raindrops or snowflakes in lower portion.
Low Clouds			
6. Stratocumulus (Sc)	0.6 - 1.5	0.2 - 0.8 km	5 - 7μ droplets with fluctuations from 1 to 60μ
7. Stratus (St)	0.1 - 0.7	0.2 - 0.8 km	2 - 5μ droplets with fluctuations from 1 to 29μ
8. Nimbostratus (Ns)	0.1 - 1	to several km	Mixed crystals: Mainly columnar at cloud top, laminar below. Droplets: 7 - 8μ with fluctuations 2 - 72μ.
Vertically Developed Clouds			
9. Cumulus (Cu)	0.8 - 1.5 can be much higher	Hundreds of meters to several km	Droplets: Radius about 11μ in center and at top, smallest near base (~ 6μ).
10. Cumulonimbus (Cb)	0.4 - 1	To several km. Sometimes to tropopause	Droplets in lower regions, crystals in upper regions. Crystals mainly plates if temperature above - 15°C and mainly columns if below.

TABLE 2-16. 0.5- to 0.7-MICRON ALBEDOS FOR VARIOUS CLOUD TYPES

TYPE	ALBEDO	COMMENTS	NO. OF MEASUREMENTS
Cumulonimbus	0.92	Large and thick	8
Cumulonimbus	0.86	Small, top ~ 6 km	1
Cirrostratus	0.74	Thick with lower clouds and precipitation	7
Cirrostratus	0.32	Alone, over land	1
Cirrus	0.36	Alone, over land	2
Stratus	0.64	Thick, ~ 0.5 km. over ocean	14
Stratus	0.42	Thin, over ocean	2
Stratocumulus	0.60	Masses within cloud sheet over ocean	4
Stratocumulus	0.68	Masses within cloud sheet over land	3
Cumulus and Stratocumulus	0.69	Masses within cloud sheet over land	4
Cumulus of fair weather	0.29	Masses within cloud sheet over land	2

TABLE 2-17. CALCULATED THICK CLOUD EMISSIVITIES

WAVELENGTH, MICRON	DROP RADIUS, MICRON	EMISSIVITY
4.6	6	0.722
7.0	6	0.809
8.5	6	0.847
10.0	1	0.983
10.0	2	0.939
10.0	6	0.897
10.0	12	0.803
11.0	6	0.960
11.9	6	0.960
13.5	6	0.944

Beyond 4 microns, cloud radiation is due to the thermal emission of the cloud itself and to reflected earth radiation. Table 2-17 presents some calculated emissivities for thick clouds composed of droplets of various sizes[89] . Even thin clouds can cause an appreciable radiation. Measurements made on cirrus clouds that were thin enough that they could not be seen visually, showed radiation in the 8- to 13.5-micron region of some $1\text{-}5 \times 10^{-4}$ w cm ^{2} sr[90]. Fig. 2-37 shows a cumulus cloud spectrum based on measurements taken from Pikes Peak[91] . The ambient temperature of the cloud was - 10°C while that of the observatory was +10°C. The effects of atmospheric absorption and emission around 6 and 15 microns are clearly indicated by the approximation of the radiation in these spectral regions to that of a +10°C cloud to reach the measuring instrument.

There are various methods of estimating the probability of cloud-free lines-of-sight through the atmosphere[92] . Most of these deal with data on cloud cover (percent of sky filled with clouds) such as that found in Ref. 1 and regular Weather Bureau reports. Some measurements are available, however, and the sometimes surprising results cast some doubt on the validity of any of the above-mentioned approximation methods[93] . Fig. 2-38 depicts the combined clear lines-of-sight data for the Northern Hemisphere. Five viewing angles, from the nadir to +90°, are represented for measuring aircraft altitudes of

from 1,000 to 40,000 ft. Figs. 2-39 and 2-40 present clear line-of-sight probabilities for winter in the Northern Hemisphere. Measurement aircraft altitude was between 25,000 and 35,000 ft. Fig. 2-39 is the probability of a clear line-of-sight to the horizon and Fig. 2-40 of a clear line-of-sight at an angle 30° above the horizon.

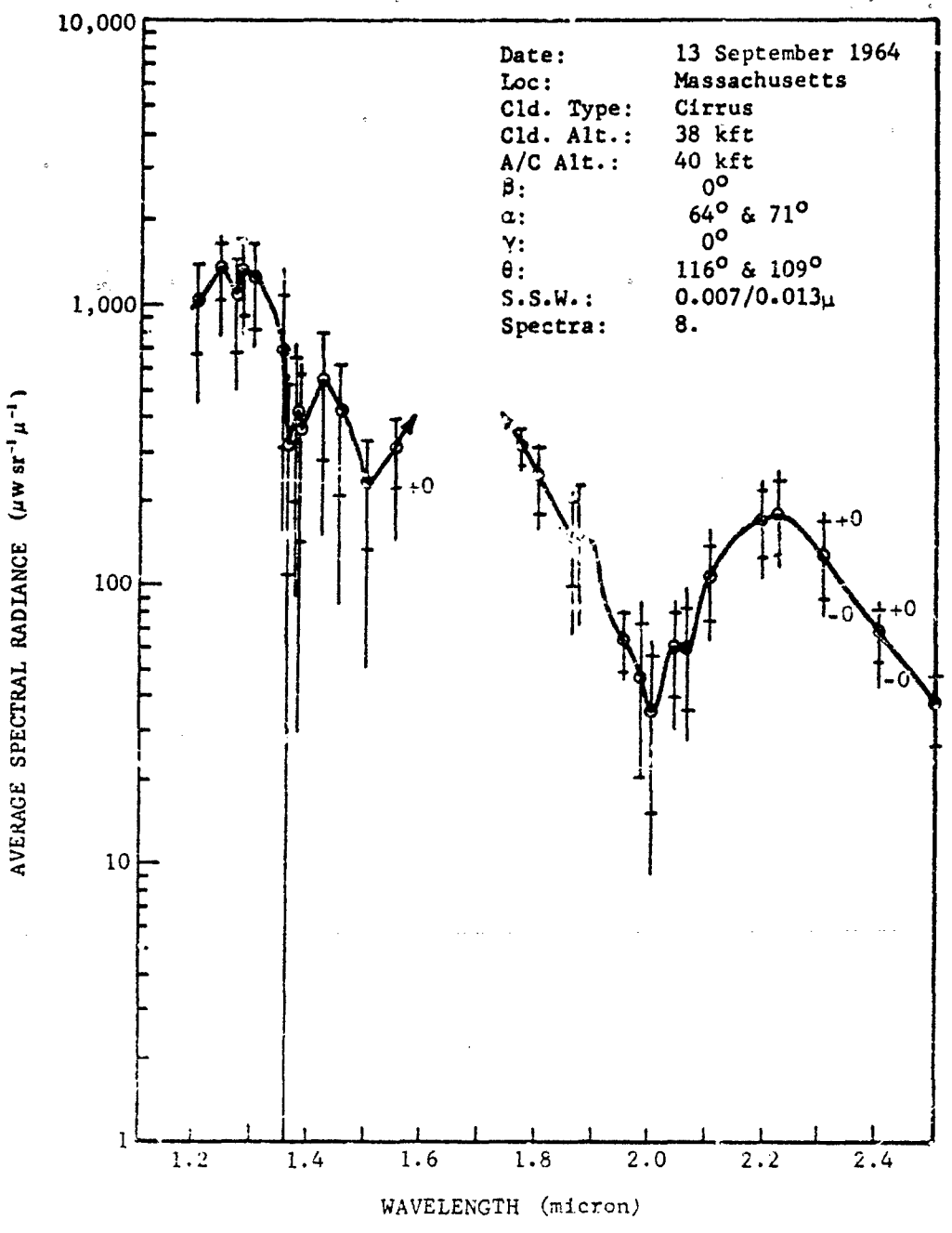


FIGURE 2-39. Typical 1.2- to 2.5-micron Cirrus Cloud Spectrum

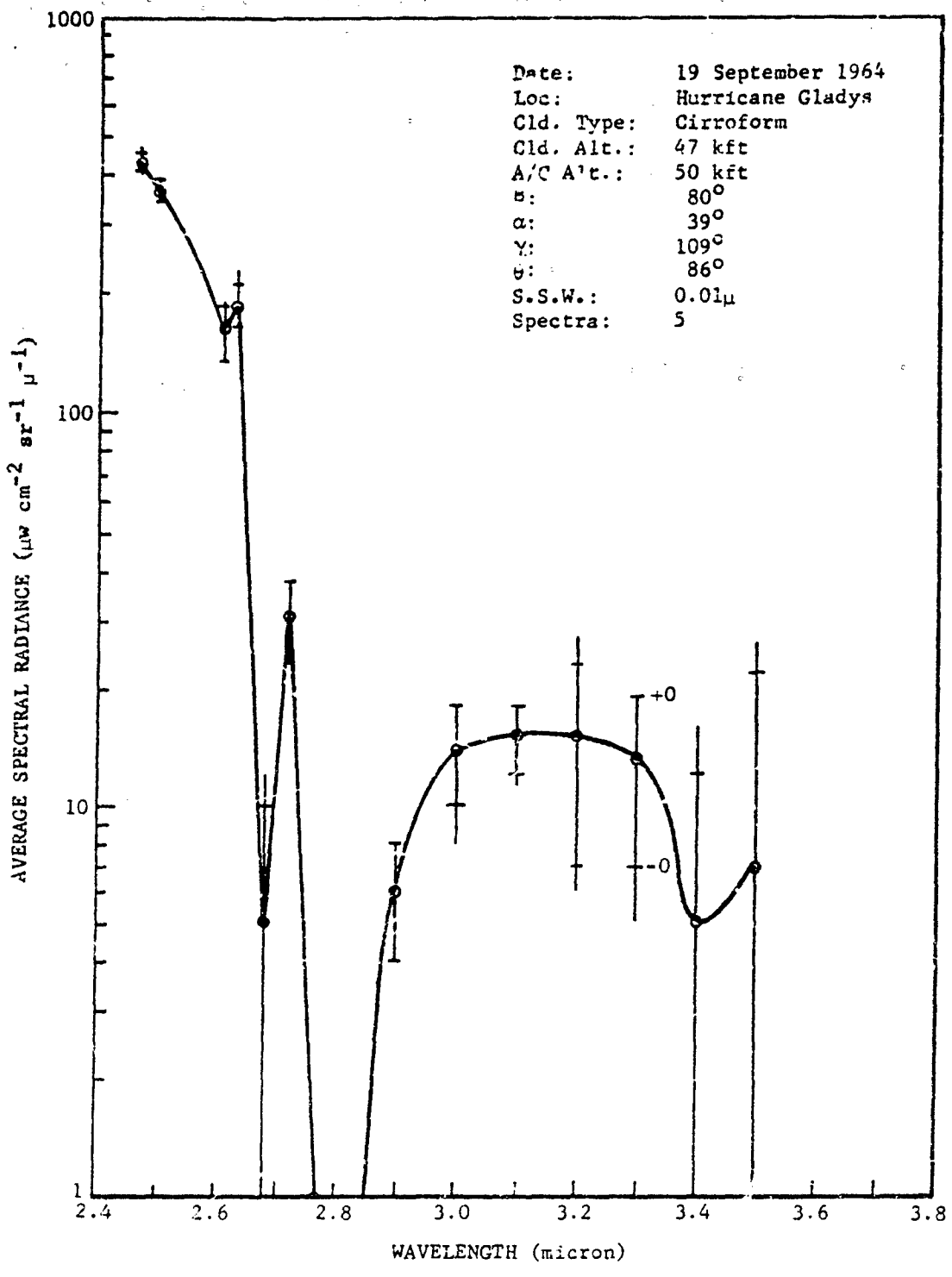


FIGURE 2-36. Average Spectrum of Gladys I

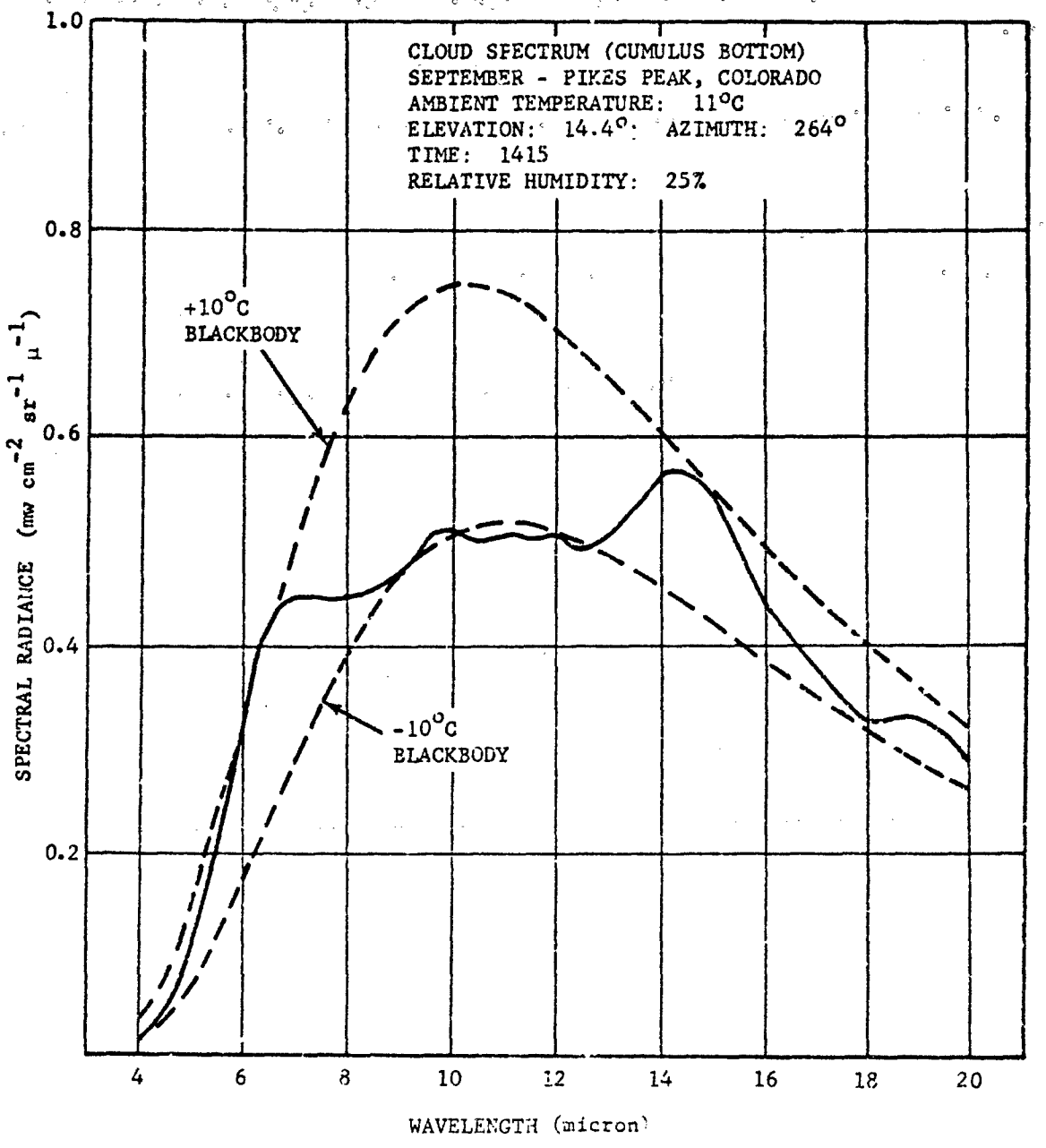


FIGURE 2-37. Cloud Spectrum

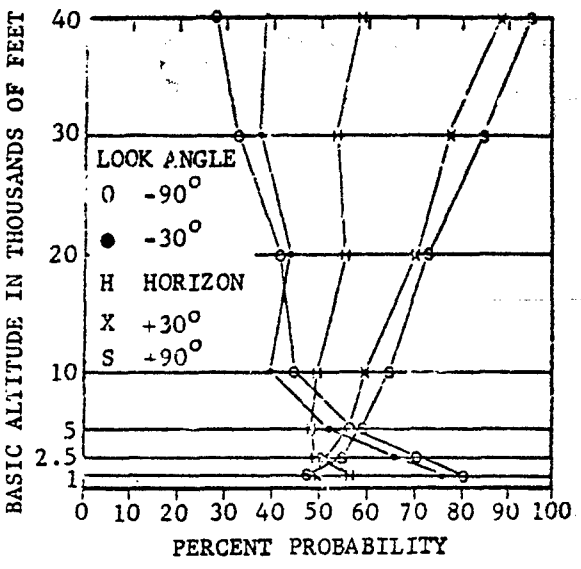


FIGURE 2-38. Probability of Clear Lines-of-sight over Northern Hemisphere for All Seasons Combined

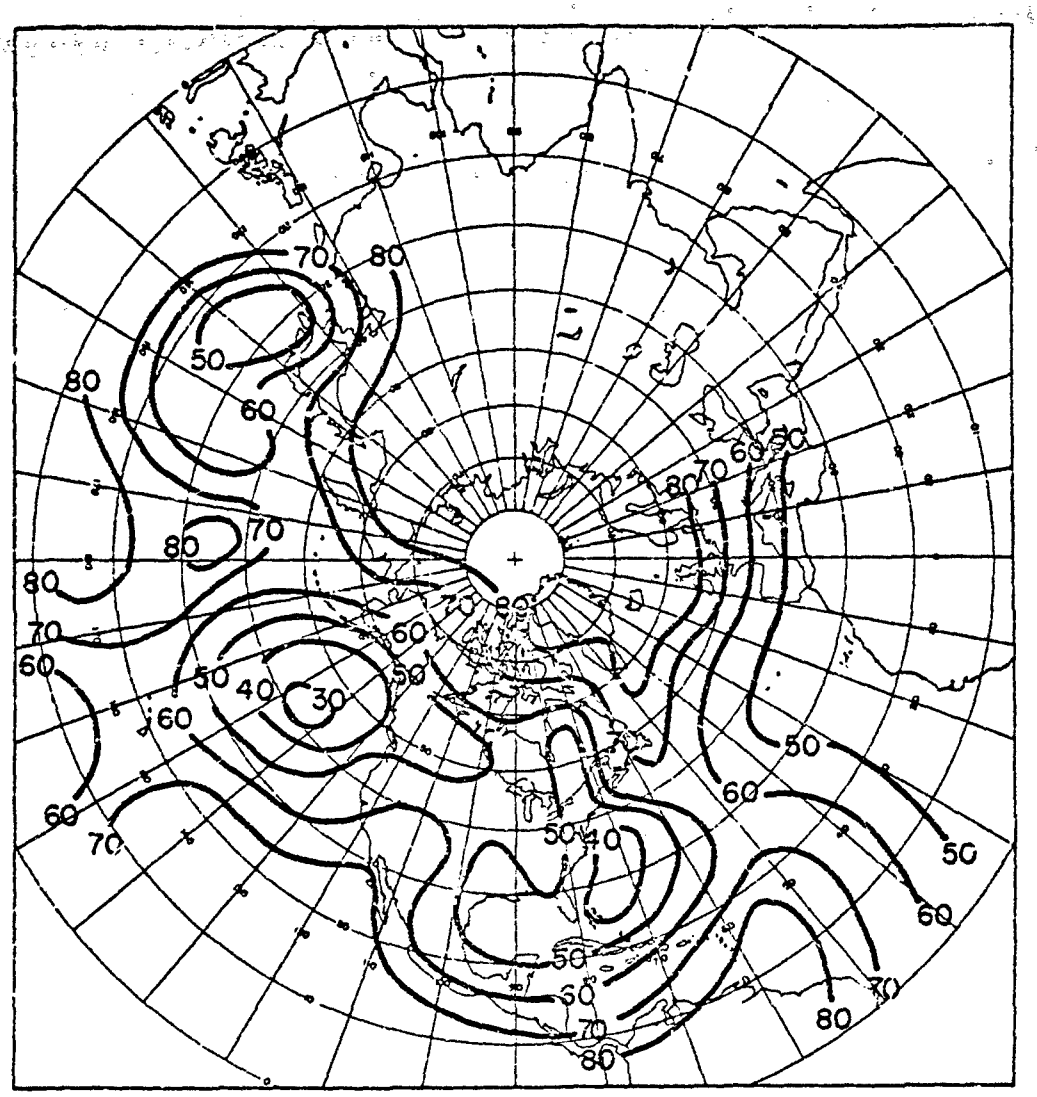


FIGURE 2-39. Probability of Clear Lines-of-sight Between Aircraft at ~30,000 ft and the Horizon

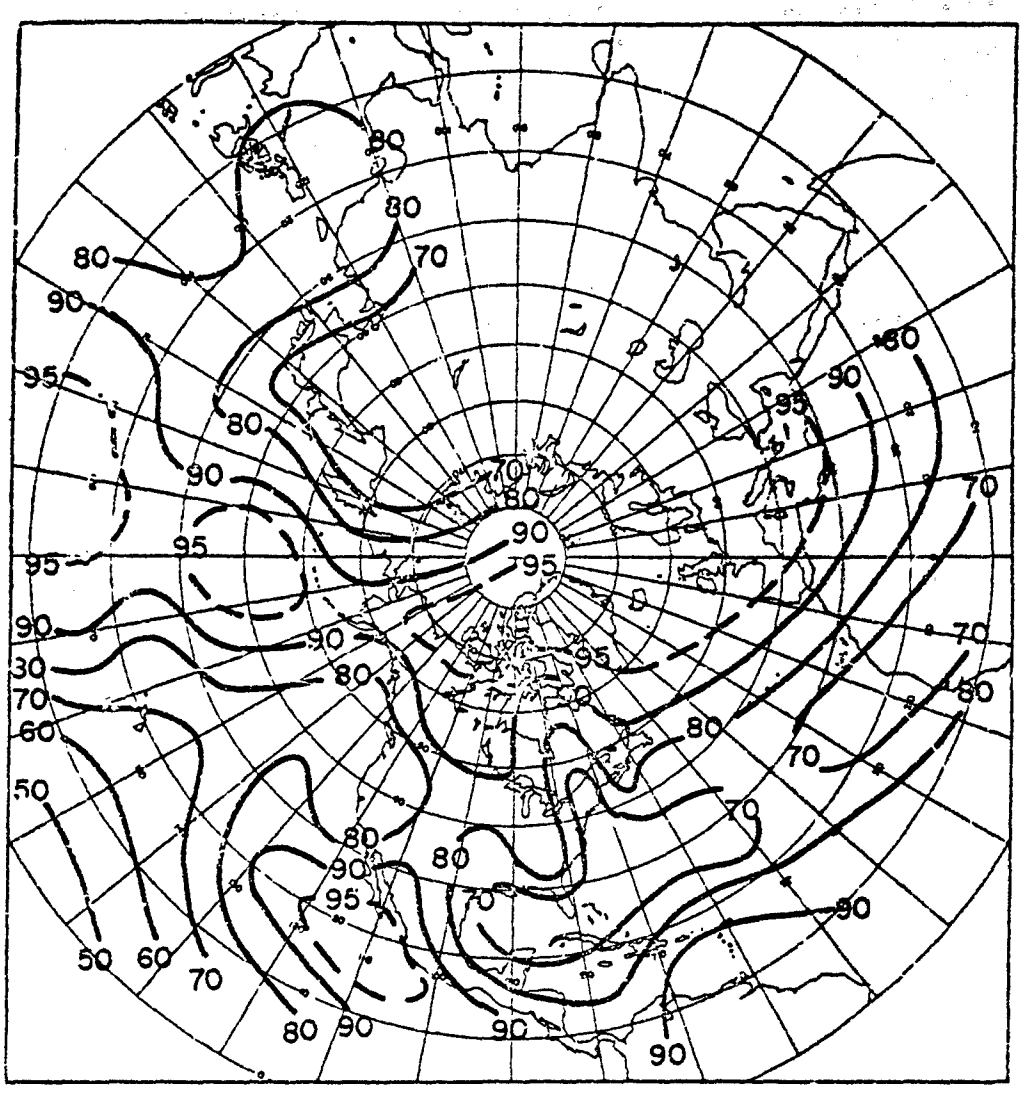


FIGURE 2-40. Probability of Clear Lines-of-sight Between Aircraft at ~30,000 ft and Space at an Angle of 30° Above the Horizon

2-6.1.4 Celestial Background

At very high altitudes, the main sources of background radiation are the stars, moon, sun, and planets. All of these sources produce radiation by self-emission. The moon and planets also reflect a relatively large amount of solar radiation due to their lower temperatures. Actual measurement data covering the entire spectral region of these sources is not available; however, such data can be derived by assuming that the source of radiation is a blackbody emitter. Radiometric observations of celestial sources at several wavelengths indicate that this is a valid and practical assumption.

Most observations performed in astronomy have been made with the eye and are classified in terms of visual magnitude. For infrared work, it is convenient to be able to convert from visual magnitudes directly into irradiance at some desired infrared bandpass. Being able to convert the measurement data in this fashion gives the infrared designer access to the great quantity of star observation data which is classified by means of visual magnitude.

The problem faced by the infrared designer in deriving his data can best be visualized by referring to Fig. 2-41[94] which shows the spectral irradiance of several of the brightest visual and red stars as a function of wavelength. The vertical line centered at 0.5 micron designates the visual magnitude m_v of the star where the vertical line intersects the irradiance curve for that star. It can be seen from Fig. 2-41 that stars having the same visual magnitude may possess greatly different irradiances at infrared wavelengths. Conversely, stars may possess the same irradiances in part of the infrared region but possess different visual magnitudes. This is true because blackbody curves are determined by two points. One point, the peak irradiance, is

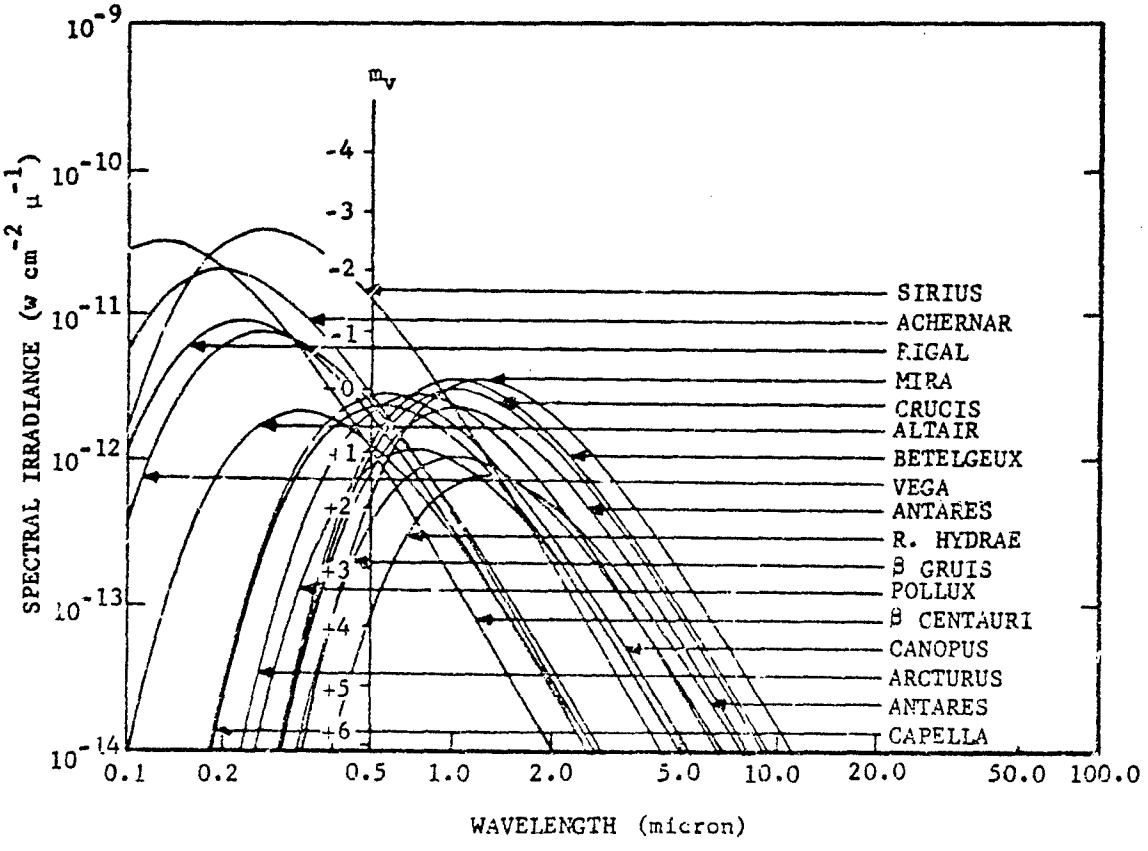


FIGURE 2-41. Spectral Irradiance of the Brightest Visual and Red Stars As a Function of Wavelength

determined by Wien's Displacement Law, Eq. 2-15, par. 2-2.6. The other point, the visual magnitude, determines the point of intersection of the irradiance curve with the vertical line at 0.5 micron. A similar curve can be drawn for the moon and the planets; however, radiation from these sources has components of self-emission and reflected sunshine as shown in Fig. 2-42. A graphical solution for the irradiance over any spectral band can be obtained by drawing a similar curve for any source. However, a more direct method which does not involve plotting a new curve for each source can be employed. This method illustrates the point that only the temperature and the visual magnitude are necessary as explained in the subsequent paragraphs.

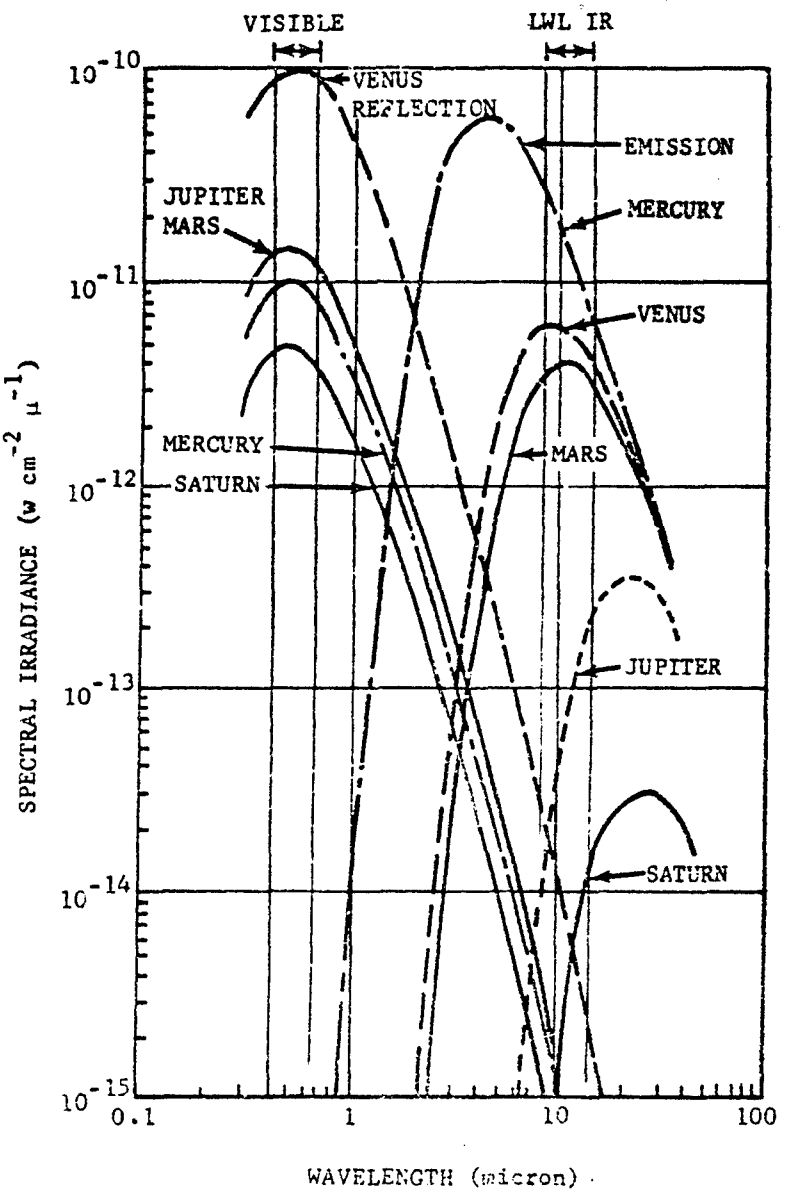


FIGURE 2-42. Major Planet Solar Reflection and Emission

TABLE 2-18. VISUAL MAGNITUDES AND EFFECTIVE TEMPERATURES OF PLANETS AND THE BRIGHTEST VISUAL AND RED STARS

NAME	VISUAL MAGNITUDE m_v	EFFECTIVE TEMPERATURE T, °K
Moon and Planets		
1. Moon (full)	-12.2	5,900
2. Venus	-4.28	5,900
3. Mars	-2.25	5,900
4. Jupiter	-2.25	5,900
5. Mercury	-1.8	5,900
6. Saturn	-0.93	5,900
Stars		
1. Sirius	-1.60	11,200
2. Canopus	-0.82	6,200
3. Rigel Kent (double)	-0.01	4,700
4. Vega	0.14	11,200
5. Capella	0.21	4,700
6. Arcturus	0.24	3,750
7. Rigel	0.34	13,000
8. Procyon	0.48	5,450
9. Achernar	0.60	15,000
10. B. Centauri	0.86	23,000
11. Altair	0.89	7,500
12. Betelguex (variable)	0.92	2,810
13. Aldebaran	1.06	3,130
14. Pollux	1.21	3,750
15. Antares	1.22	2,900
16. ∂ Crucis	1.61	2,810
17. Mira (variable)	1.70	2,390
18. B Gruis	2.24	2,810
19. R Hydrae (variable)	3.60	2,250
20. RMonocerotis[95]	—	—
21. Orion IR Star[95]	—	650
22. Cygnus IR Star[96]	—	700

A list of the brightest visual and red stars, their visual magnitude when brightest, and their effective temperatures are tabulated in Table 2-18. The data in Table 2-18 come from Ref. 94 with the addition of data of some of the more recently measured infrared stars. The spectral irradiance curves for infrared stars are shown in Fig. 2-43, 2-44[95], and 2-45[96]. The brightness of celestial bodies is designated in visual magnitudes m_v and is a function of the visible irradiance received from the source

$$m_v = -2.5 \log_{10} \frac{I_m}{I_o} \qquad (2\text{-}144)$$

where

m_v = visual magnitude

I_m = visible irradiance from a zero source, w cm^{-2}

$I_o = 3.1 \times 10^{-13}$ w cm^{-2}, visible irradiance from a zero magnitude source

Eq. 2-144 is plotted in Fig. 2-46 as a function of visible irradiance.

Since the visible irradiance refers to the capacity of radiation to produce the sensation of light on the standard eye, it is not a convenient quantity to use when applying the results to infrared detectors. It can, however, be expressed in terms of more fundamental quantities

$$I_m = [H_{BB}(T)]\,[F_v(T)],\ \text{w cm}^{-2} \qquad (2\text{-}145)$$

where

$H_{BB}(T)$ = blackbody irradiance of source, w cm^{-2}
$F_v(T)$ = visual fraction of the total blackbody irradiance
T = source temperature, °K

The function $F_v(T)$ is evaluated by computing the following integral and is plotted in Fig. 2-47 as a function of temperature:

$$F_v(T) = \frac{\int_0^\infty [W_{BB}(T,\lambda)]\,[R_v(\lambda)\,d\lambda]}{\int_0^\infty [W_{BB}(T,\lambda)]\,d\lambda} \qquad (2\text{-}146)$$

where

$W_{BB}(T,\lambda)$ = blackbody emittance, w cm^{-2}
$R_v(\lambda)$ = spectral response of standard eye
λ = wavelength, microns
T = source temperature, °K

Substituting Eq. 2-145 into Eq. 2-144 and rearranging terms, the solution for the blackbody irradiance is found by:

$$H_{BB}(T) = \frac{3.1 \times 10^{-13}}{F_v(T)\,[\text{antilog}_{10}(0.4 m_v)]},\ \text{w cm}^{-2} \qquad (2\text{-}147)$$

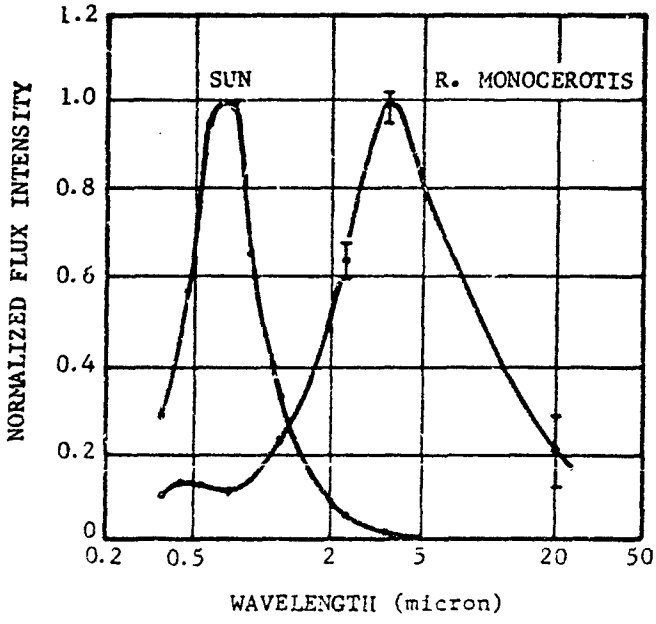


FIGURE 2-43. Normalized Spectral Irradiance of R. Monocerotis As a Function of Wavelength

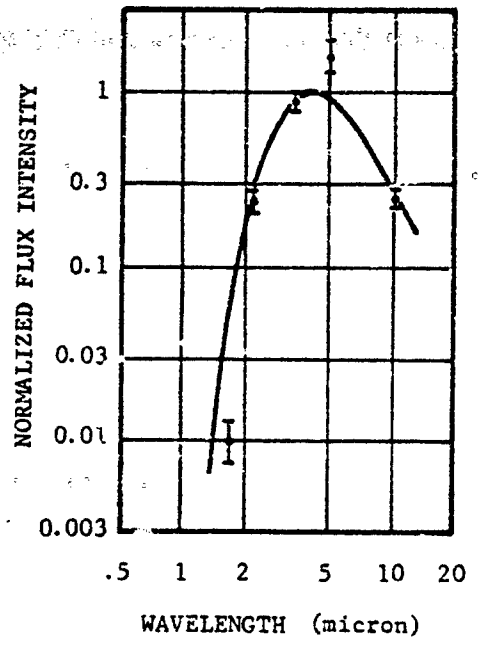


FIGURE 2-44. Normalized Spectral Irradiance of the Orion Infrared Star As a Function of Wavelength

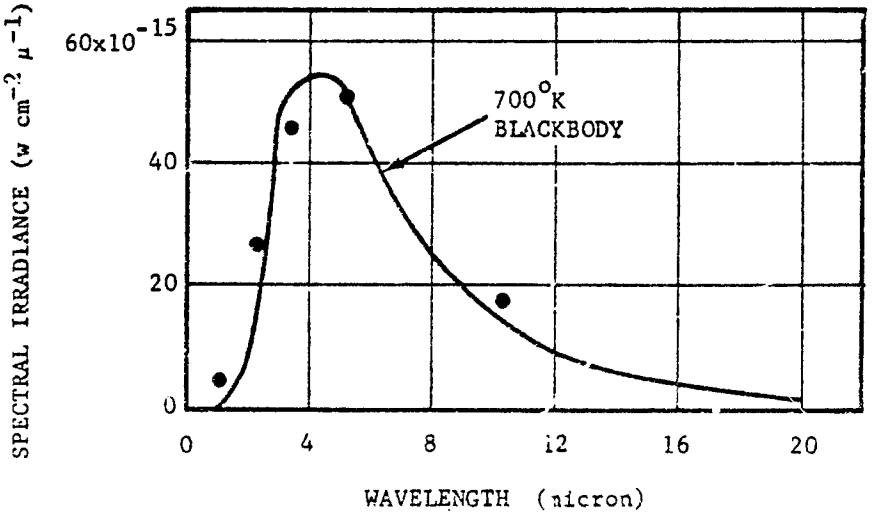


FIGURE 2-45. Spectral Irradiance of the Cygnus Infrared Star As a Function of Wavelength

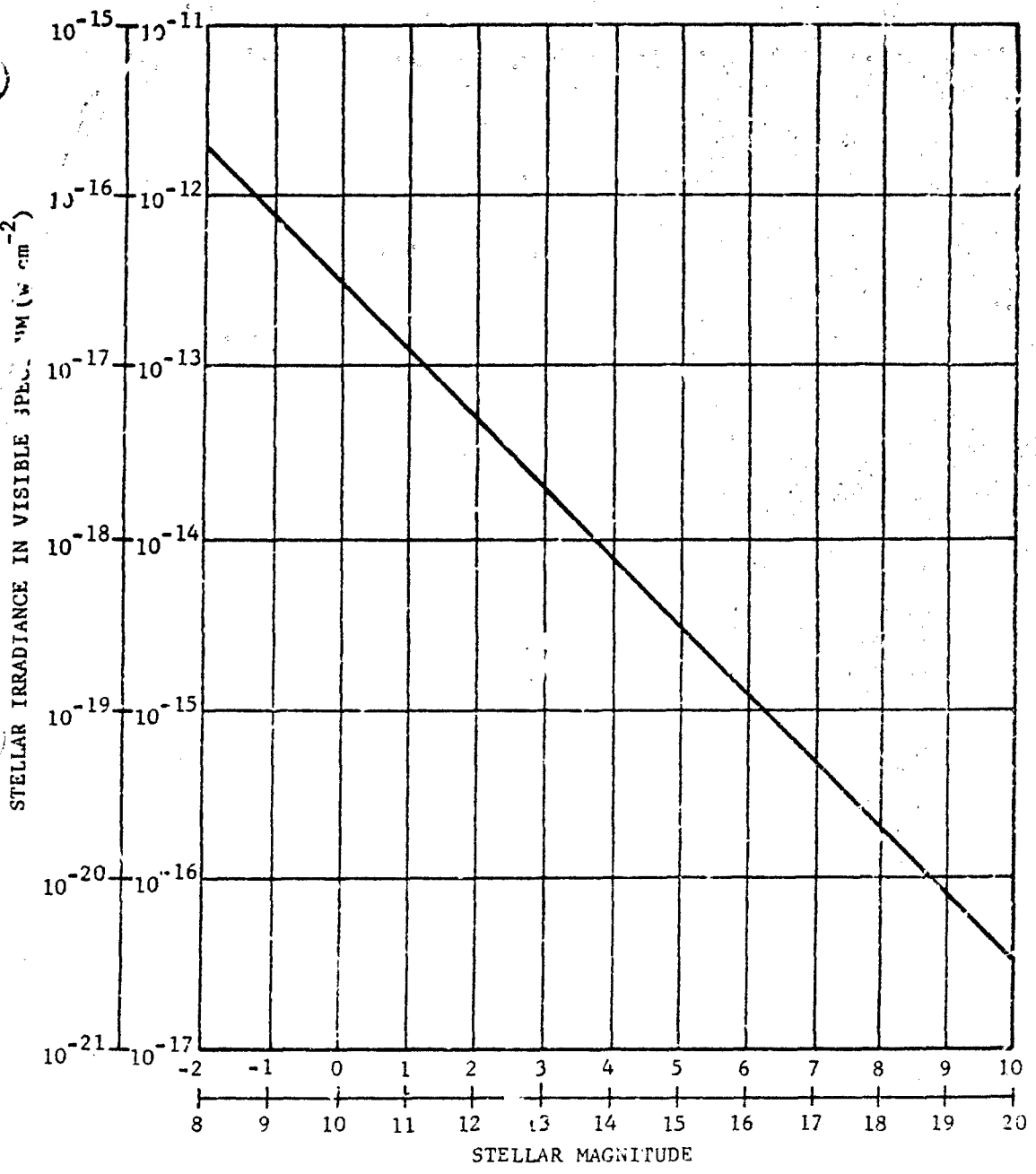


FIGURE 2-46. Stellar Irradiance As a Function of Visual Magnitude

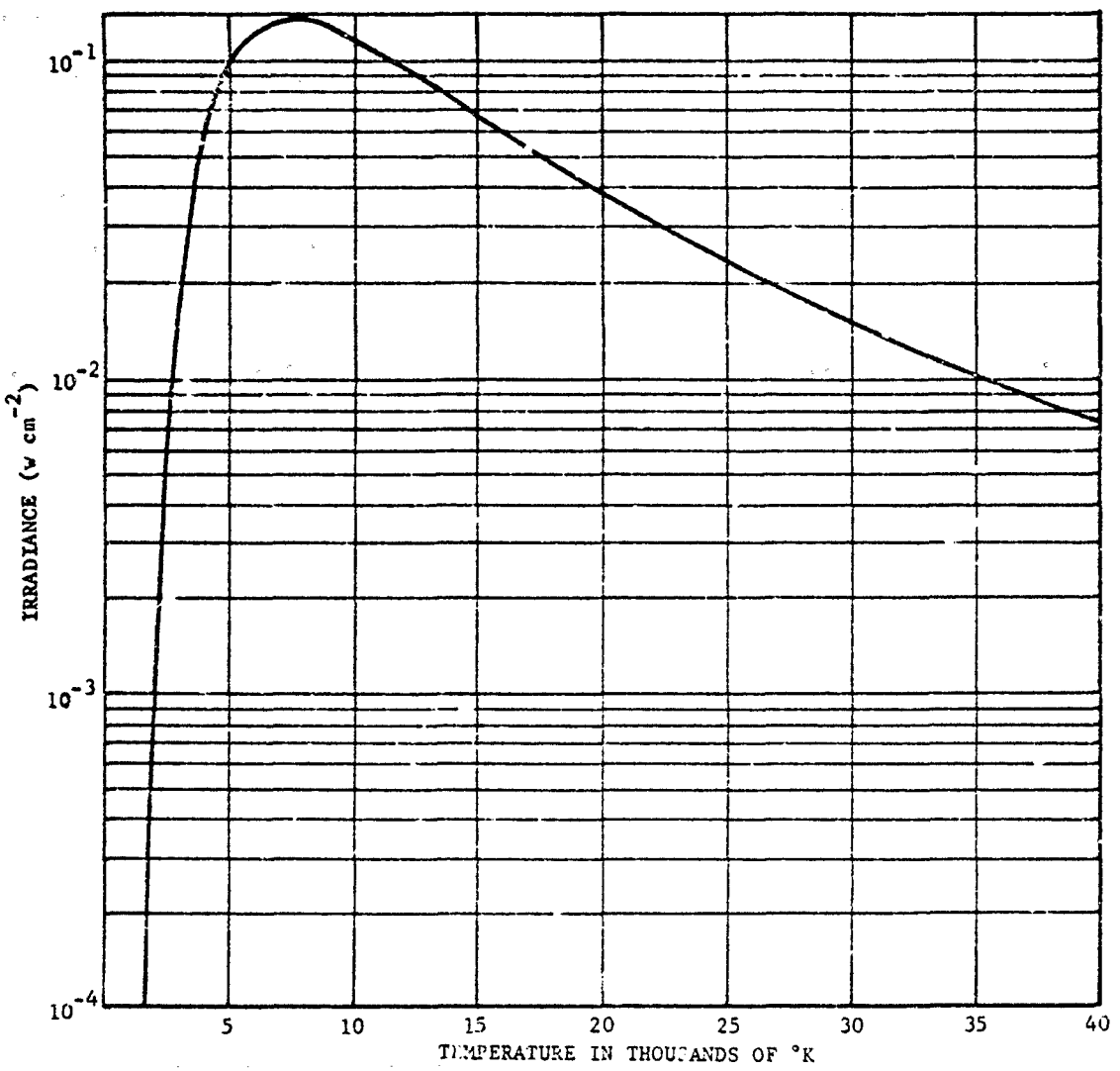


FIGURE 2-47. Visual Fraction of Total Blackbody Irradiance As a Function of Temperature

Eq. 2-147 illustrates the fact that the total blackbody irradiance is determined by the visual magnitude and the temperature of the source. The source temperature is used in conjunction with Fig. 2-47 to obtain the value of the visual fraction $F_v(T)$.

In the design of infrared systems, the designer is concerned with the irradiance in a specified spectral band in either the detector band or a narrow filter band. In order to obtain the irradiance in a spectral band, Eq. 2-147 is used in conjunction with Fig. 2-48[97]. Fig. 2-48 is a plot of the fraction of irradiance below a specified wavelength as a function of source temperature at various wavelengths. This same information can be found on a radiation slide rule. The relationship of the spectral class of the stars to star temperature is also shown in Fig. 2-48. The percent of irradiance below a specified wavelength for a source at a known temperature is found by the intersection of the temperature line with the wavelength curve. For the percent of irradiance within a spectral band, the difference is taken of the percentages corresponding to the wavelengths of the band limits. The values obtained from Fig. 2-48 are expressed in percentages. The 100-percent value in Fig. 2-48 corresponds to the value obtained from Eq. 2-147.

Table 2-19 contains a list of the number of stars brighter than a given magnitude[97]. It is also of interest to obtain a similar list of stars which applies to the infrared wavelengths. Such a list can be generated by counting the number of stars brighter than a given irradiance calculated at the infrared wavelengths. The calculation is described in the paragraphs which follow and is based on the detected irradiance.

It has been previously stated that the visible irradiance is not a convenient quantity to apply to infrared wavelengths since it refers to the capacity of radiation to produce the sensation of light on the standard eye. There is, however, an infrared counterpart which is detected irradiance H_d. Detected irradiance refers to the capacity of radiation to produce a response at the detector.

$$H_d = \left[H_{BB}(T)\right]\left[F_d(T)\right], \text{ w cm}^{-2} \qquad (2\text{-}148)$$

where

$H_{BB}(T)$ = blackbody irradiance of source, w cm^{-2}
$F_d(T)$ = detected fraction of the total blackbody irradiance
T = source temperature, °K

The function $F_d(T)$ is evaluated by computing the following integral and is plotted in Fig. 2-49 as a function of source temperature for various spectral bands using a mercury doped germanium detector:

$$F_d(T) = \frac{\int_0^{\infty} \left[W_{BB}(T,\lambda)\right]\left[R_D(\lambda)\right] d\lambda}{\int_0^{\infty} \left[W_{BB}(T,\lambda)\right] d\lambda} \qquad (2\text{-}149)$$

where

$W_{BB}(T,\lambda)$ = blackbody emittance, w cm^{-2}
$R_D(\lambda)$ = spectral response of detector and filters
λ = wavelength, microns
T = source temperature, °K

TABLE 2-19. ESTIMATED NUMBER OF STARS BRIGHTER THAN A GIVEN MAGNITUDE

MAGNITUDE	NUMBER OF STARS PHOTOGRAPHIC	VISUAL
0		2
1		12
2		40
3		140
4	360	530
5	1,030	1,620
6	2,940	4,850
7	8,200	14,300
8	22,800	41,000
9	62,000	117,000
10	166,000	324,000
11	431,000	870,000
12	1,100,000	2,270,000
13	2,720,000	5,700,000
14	6,500,000	13,800,000
15	15,000,000	32,000,000
16	33,000,000	71,000,000
17	70,000,000	150,000,000
18	143,000,000	296,000,000
19	275,000,000	560,000,000
20	505,000,000	1,000,000,000
21	890,000,000	

Having determined the blackbody irradiance $H_{BB}(T)$ from Eq. 2-147 and the detected fraction $F_d(T)$ from Fig. 2-49, we compute the detected irradiance for any star by substituting the values into Eq. 2-148. Since $H_{BB}(T)$ is determined by the temperature and visual magnitude of the star, it would be necessary to compute a new value of $H_{BB}(T)$ for each visual magnitude and each spectral class. It would also be necessary to determine a new value for $F_d(T)$ for each spectral class. It is possible, however, to simplify these calculations once the spectral region is chosen.

The 8- to 14-micron band is an infrared band of interest because of the atmospheric window at this band and because the presently available LWL detector (mercury doped germanium detector) has its maximum response over this region. Having chosen the 8- to 14-micron band, it is possible to restrict the calculations to only those stars in the K and M spectral classes. This is justified by the fact that the K and M stars will contribute the greatest part of the infrared radiation in this spectral region. By the same token, there are possibly a large number of infrared stars missing from the count by virtue of the fact that the cooler stars have so little radiation at shorter wavelengths they might have never been observed.

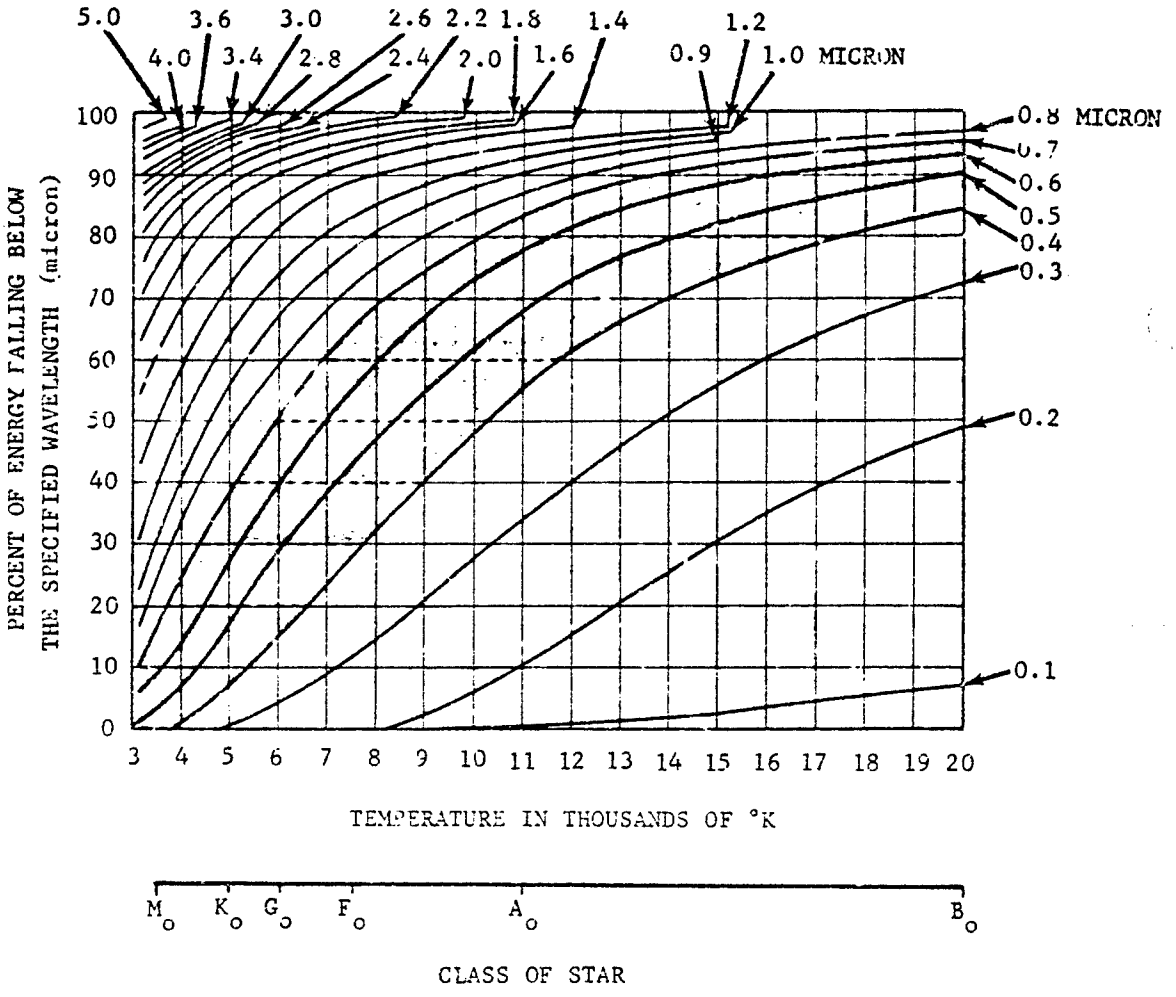


FIGURE 2-48. Relative Spectral Distribution of Stellar Radiation As a Function of Star Class

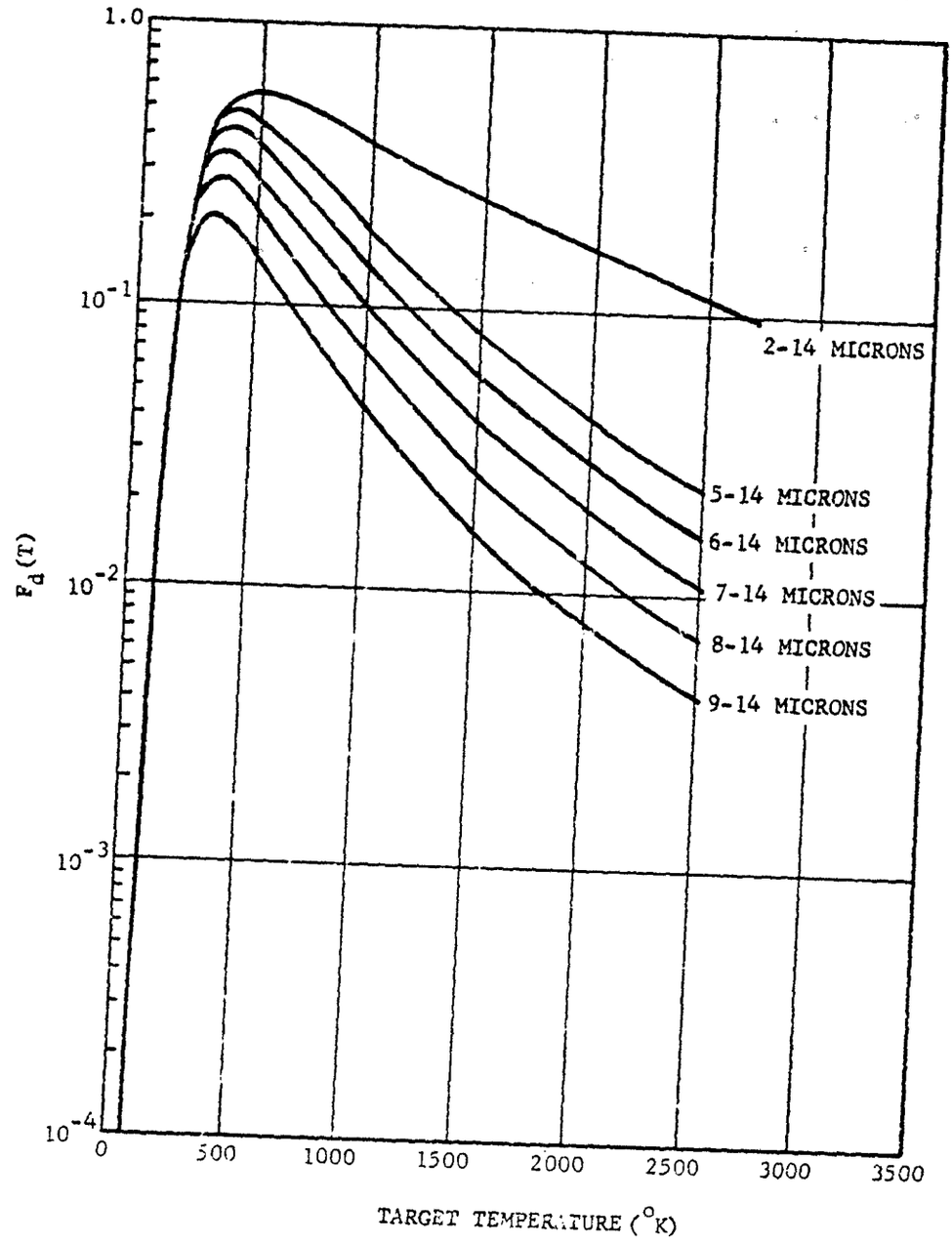


FIGURE 2-49. Detector Response Fraction $F_d(T)$ for Ge:Hg Detectors As a Function of Spectral Band and Target Temperature

TABLE 2-20. CALCULATIONS OF IRRADIANCE OF *K* AND *M* SPECTRAL CLASS STARS

STAR VISUAL MAGNITUDE		COUNT OF *K* SPECTRAL CLASS STARS	COUNT OF *M* SPECTRAL CLASS STARS	8- TO 14-MICRON IRRADIANCE OF *K* STARS OF m_v + 0.5 MAGNITUDE, w cm^{-2}	8- TO 14-MICRON IRRADIANCE OF *M* STARS OF m_v + 0.5 MAGNITUDE, w cm^{-2}
(1)		(2)	(3)	(4)	(5)
m_v		*K*	*M*	H_{dK}	H_{dM}
0.00	0.99	1	1	8.27×10^{-15}	4.90×10^{-14}
	1.00	6	2	$3.30 \times 10^{-}$	1.96×10^{-14}
	2.00	27	4	1.31×10^{-15}	7.76×10^{-15}
	3.00	102	10	5.24×10^{-16}	3.10×10^{-15}
	4.00	265	52	2.09×10^{-16}	1.24×10^{-15}
	5.00	941	137	8.27×10^{-17}	4.90×10^{-16}
	6.00	2,855	294	3.30×10^{-17}	1.96×10^{-16}
	7.00	8,815	805	1.31×10^{-17}	7.76×10^{-17}
	8.00	23,189	1,349	5.24×10^{-18}	3.10×10^{-17}
	9.00	19,538	938	2.09×10^{-18}	1.24×10^{-17}
	10.00	4,371	194	8.27×10^{-19}	4.90×10^{-18}
	11.00	231	35	3.30×10^{-19}	1.93×10^{-18}

Table 2-20 contains the results of carrying out the calculations just described. Column 1 contains the visual magnitude. Columns 2 and 3 are the star count of K and M stars from Barnhart and Mitchell[98]. Columns 4 and 5 contain the detected irradiance produced by each star having that magnitude for the K and M stars. The next step is to add up all K and M stars having a given detected irradiance or greater, the results of which are shown in Fig. 2-50.

Due to the proximity of the moon, a lot more can be said about it than simply representing it as a blackbody source. Without getting into great detail about the surface temperature, it is worth mentioning that measurement of the lunar surface was accomplished by Surveyor I which made a soft landing on the moon on 2 June 1966. Fig. 2-51 shows the temperature curve for the lunar surface at the Surveyor I site[99]. It can be seen from this figure that at local noon the temperature reaches 390°K. By one day past local sunset the temperature dropped to 130°K. The actual moon irradiance which reaches an instrument will depend upon the particular sun-moon-instrument-target geometry.

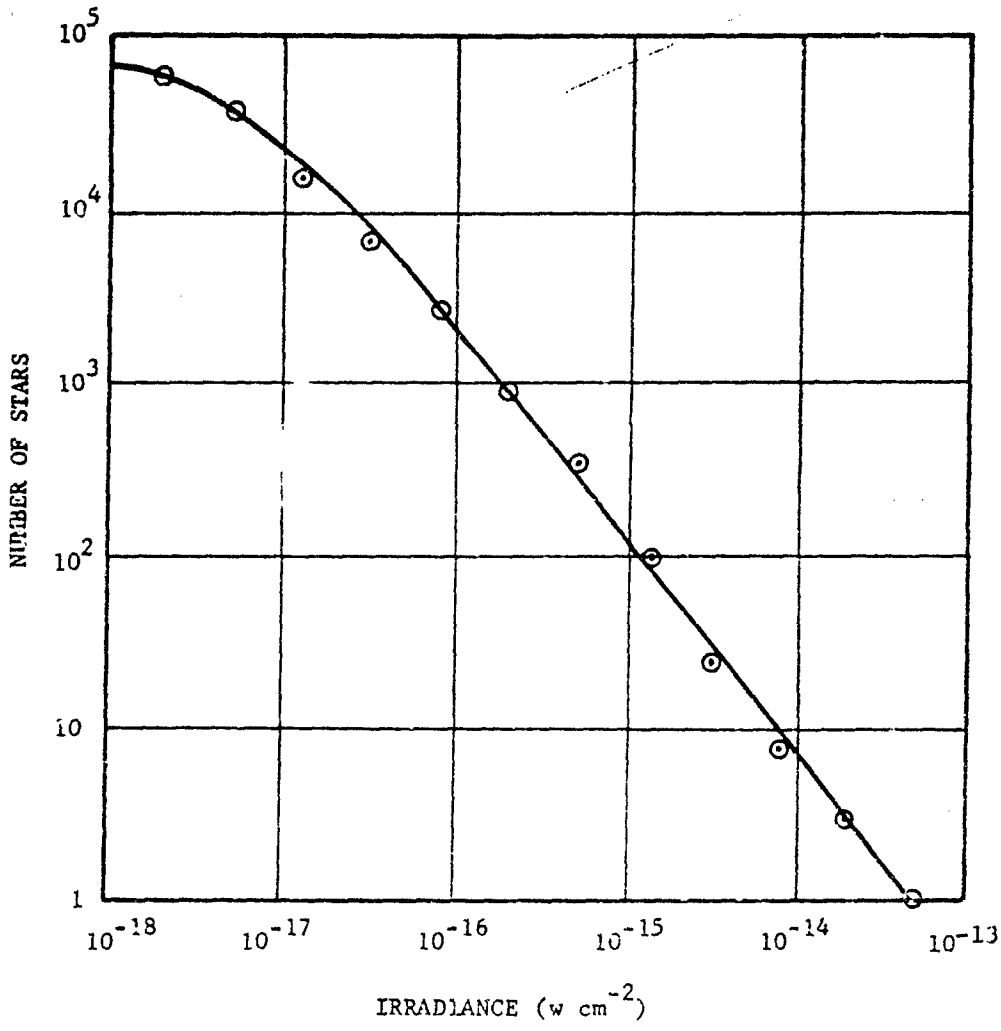


FIGURE 2-50. Number of K and M Stars in the 8- to 14-micron Band Exceeding a Given Irradiance

The solar spectral radiation curve is shown in Fig. 2-52[100]. Fraunhofer lines are observed in the visual and ultraviolet portions. In the infrared, however, very few Fraunhofer lines are present and the sun emits as a uniform 6000°K blackbody.

Extended celestial sources of infrared are of interest to the system designer as they could affect the system sensitivity by raising the background level. G. Neugebauer and Robert Leighton[95] have generated an infrared map of the galactic center measured at 2 to 2.4 micrometers. The results of their survey are shown in Fig. 2-53. Related to this point is the spatial distribution of the stars. Such a list is given in Table 2-21 which shows the number of stars brighter than a given photographic magnitude as a function of galactic latitude.

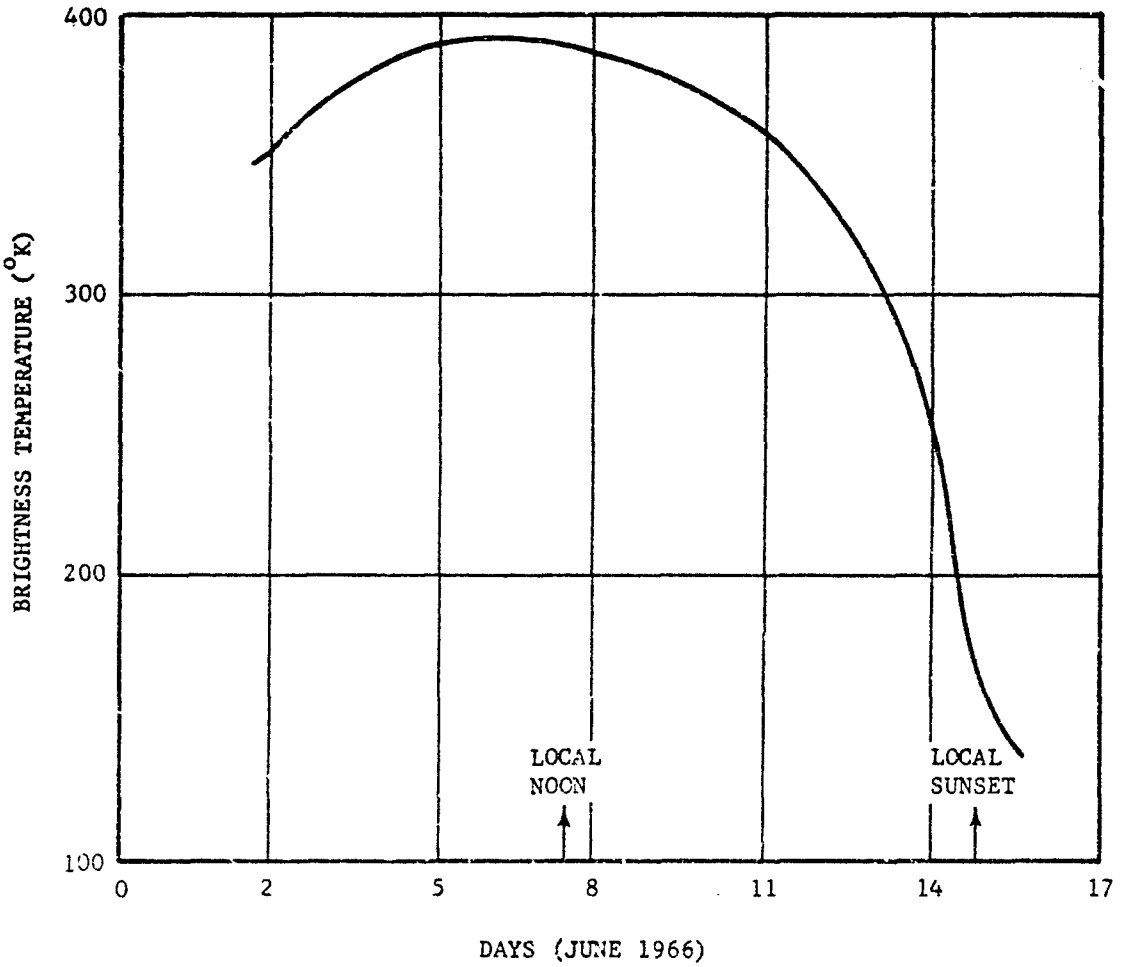


FIGURE 2-51. Lunar Temperature Measured by Surveyor I

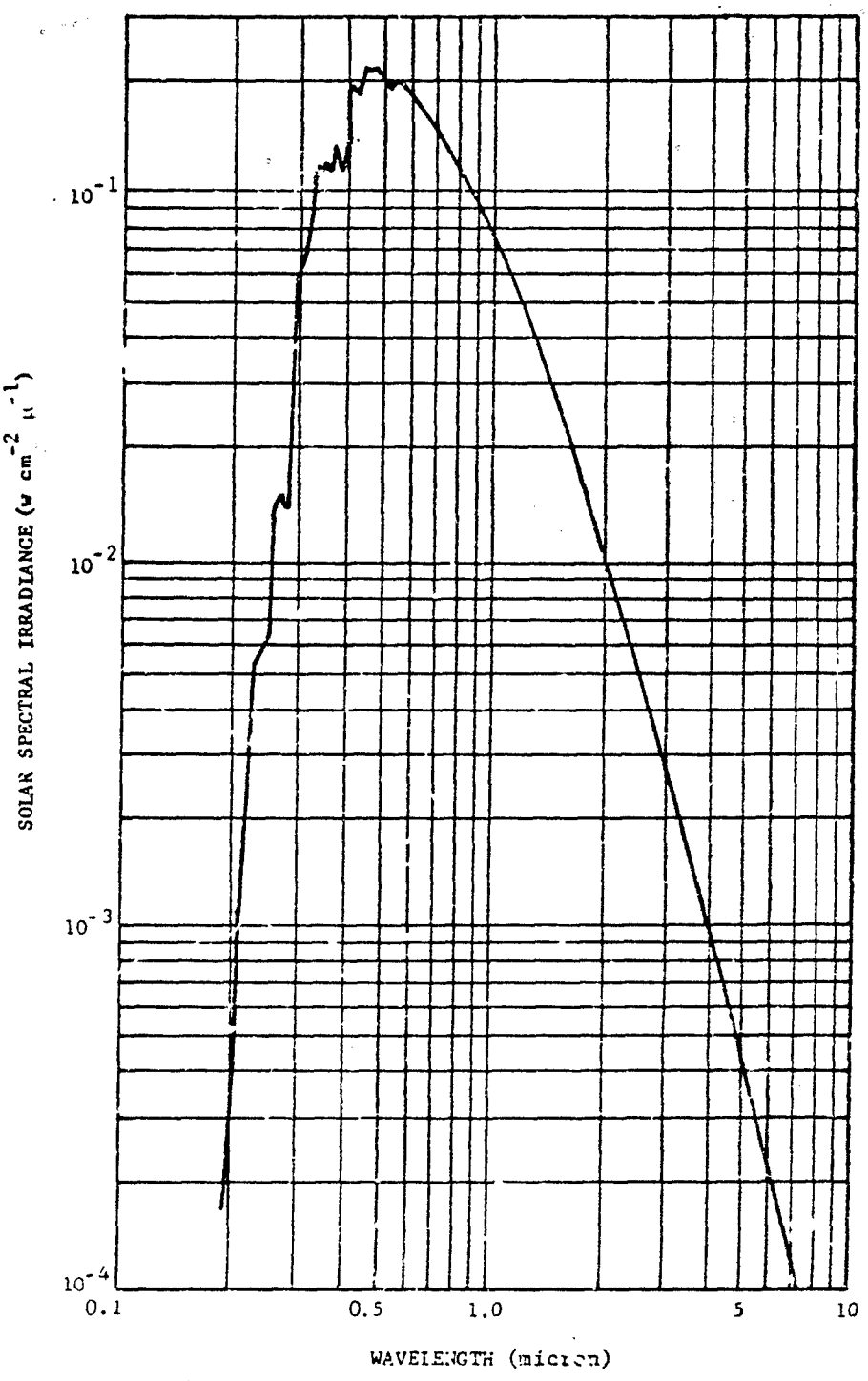
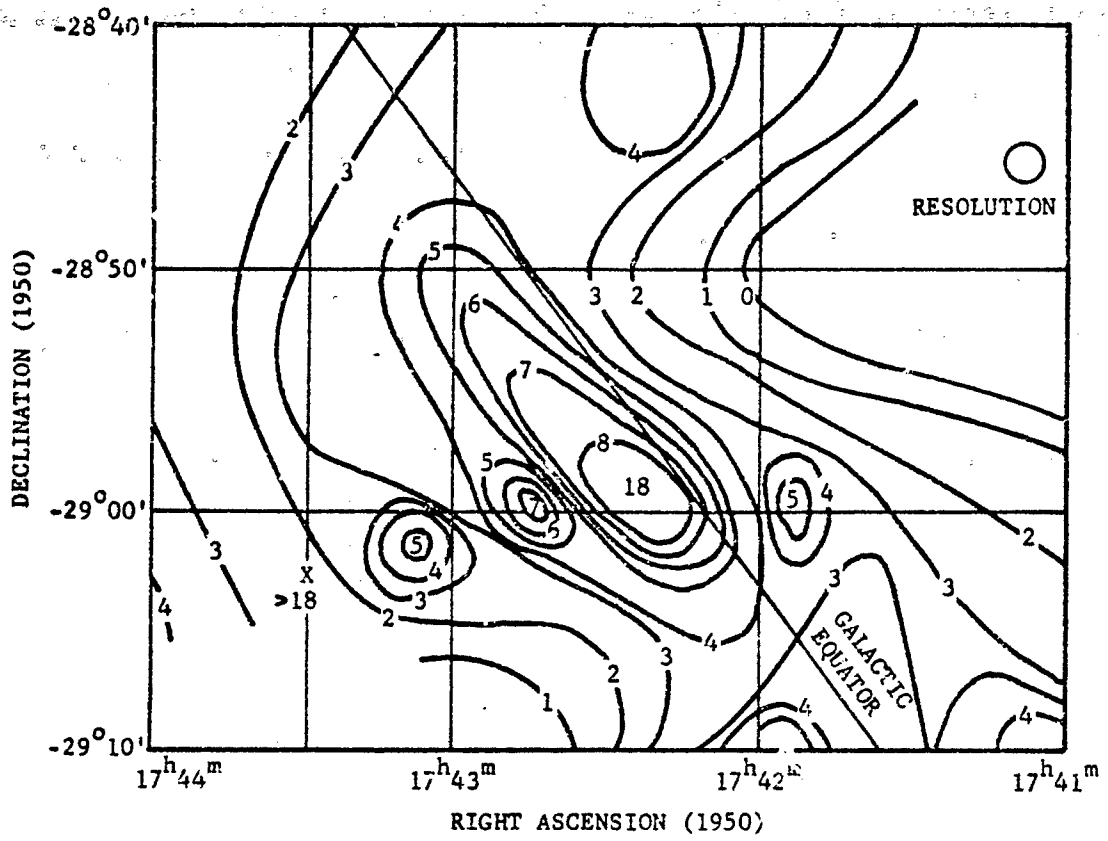


FIGURE 2-52. Solar Spectral Irradiance As a Function of Wavelength



Map shows intense emission at 2.2 microns slightly to the left of the galactic equator. To the left is an intense source marked by a cross. The zero line represents the sky background radiation. Other lines are labeled in units of 5.2×10^{-10} w cm^{-2} μ^{-1} sr^{-1}.

FIGURE 2-53. Infrared Map of the Galactic Center

TABLE 2-21. NUMBER OF STARS PER SQUARE DEGREE BRIGHTER THAN PHOTOGRAPHIC MAGNITUDE *M* AS A FUNCTION OF GALACTIC LATITUDES[97]

PHOTOGRAPHIC MAGNITUDE	+90°	+40°	+20°	+10°	0°	-10°	-20°	-40°	-90°
5.0	0.014	0.0175	0.023	0.031	0.059	0.045	0.032	0.0178	0.012
6.0	0.039	0.053	0.071	0.089	0.166	0.126	0.087	0.051	0.042
7.0	0.015	0.151	0.20	0.257	0.436	0.323	0.224	0.144	0.123
8.0	0.275	0.42	0.59	0.741	1.230	0.851	0.617	0.398	0.316
9.0	0.724	1.12	1.62	2.14	3.55	2.34	1.69	1.10	0.832
10.0	1.78	2.95	4.50	5.89	10.5	6.61	4.68	2.95	2.09
11.0	4.3	7.4	12.0	16.2	30.9	18.2	12.8	7.76	5.25
12.0	10.2	18.2	32.0	43.6	89.1	50.1	34.7	19.50	13.2
13.0	24.9	43.0	79.0	112.0	245.0	138.0	89.1	47.8	30.2
14.0	50.0	93.0	190.0	282.0	661.0	371.0	218.0	107.0	60.3
15.0	95.0	200.0	457.0	708.0	1660.0	977.0	525.0	218.0	104.0
16.0	182.0	407.0	1047.0	1778.0	3981.0	2455.0	1175.0	436.0	182.0
17.0	338.0	794.0	2291.0	4365.0	9120.0	5754.0	2512.0	832.0	302.0
18.0	616.0	1113.0	4677.0	9330.0	20890.0	12590.0	4786.0	1514.0	501.0
19.0	770.0	2180.0	6068.0	—	—	—	—	—	—
20.0	—	—	—	—	—	—	—	—	—
21.0	1670.0	5000.0	21200.0	—	—	—	—	—	—

2-6.2 TARGET RADIATION

From the standpoint of infrared physics there can be no real distinction between targets and backgrounds. The interest of the moment is the only criterion for such classification and, particularly in the case of structures and terrain features, a target in one situation may well form part of the background in another. Certain objects of military interest frequently fall into the category of targets and will be treated here insofar as security restrictions will permit. Inevitably this treatment must be of a generalized and fragmentary nature.

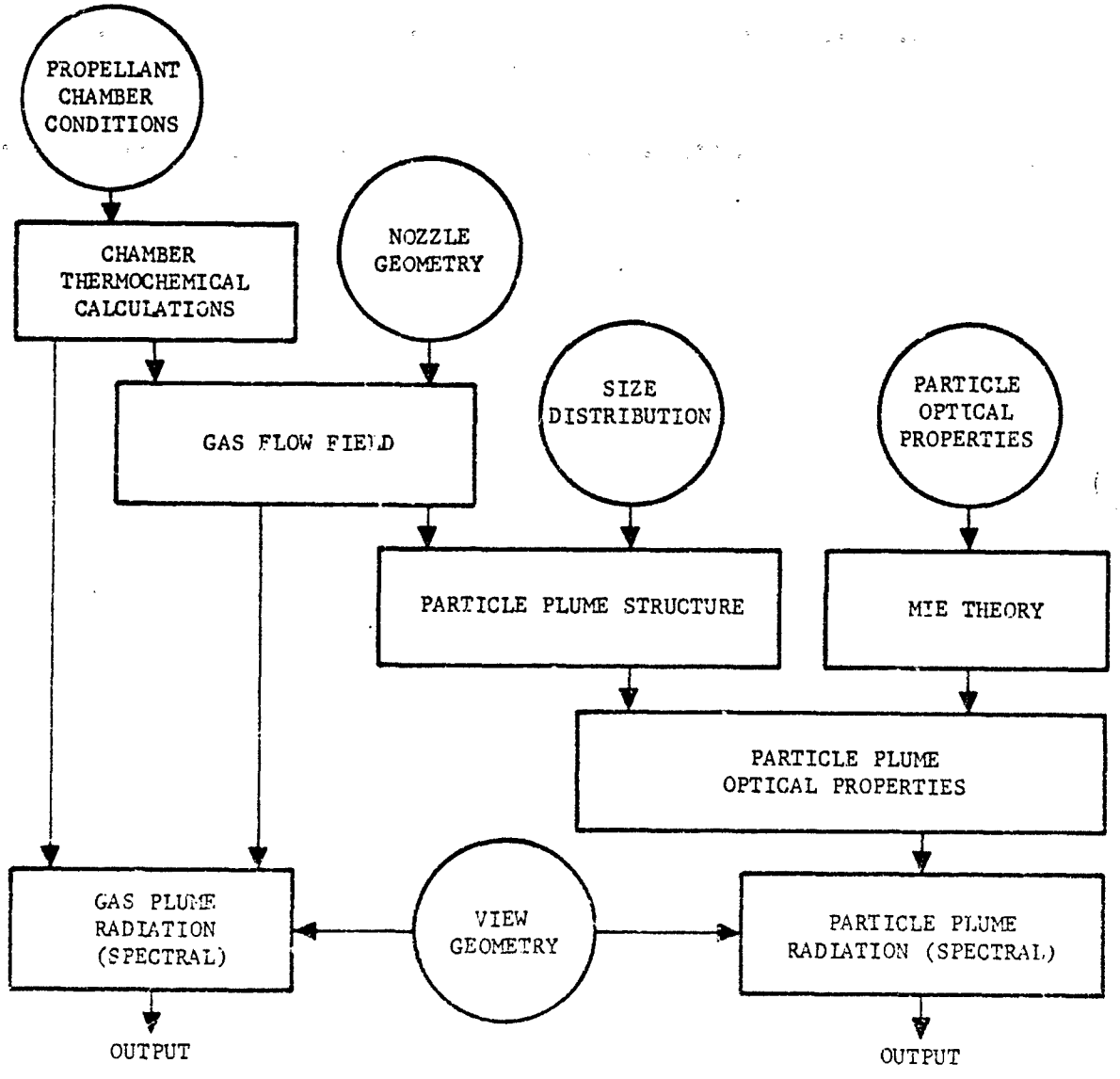


FIGURE 2-54. Flow Chart for Plume Radiation Calculations

2-6.2.1 Rockets and Missiles

Rockets and missiles act as infrared sources in a variety of ways, varying in importance according to the stage and type of flight. From launch to sustainer burnout (powered phase), the most intense source of radiation is the exhaust plume (see par. 2-6.2.2). The structure and radiance of the plume depend strongly on the ambient atmospheric pressure and velocity, a particularly relevant point when considering those devices which fire over a range of altitudes. Base heating by the plume can cause parts near the rear of a rocket to radiate significantly as graybodies. In addition, radiation from the hot exhaust nozzle and combustion chamber itself may be directly observable in the rear aspect. Scattering of this radiation by solid particles in the plume, the searchlight effect, may contribute to the radiance in other directions too, especially for solid fueled devices.

Aerodynamic heating at supersonic velocities will cause the surface of the vehicle to radiate as a graybody. During exit from the atmosphere this will be significant, if at all, only in the forward aspect in which the plume is partially obscured. During re-entry much higher temperatures are attained—aerodynamically heated surfaces, the wake of hot ablation products, and shock heater air are strong sources.

2-6.2.2 Plume Radiation

Determination of the radiant flux from a plume is a complex problem in

(1) Fluid dynamics and thermochemistry, to determine the thermodynamical structure and chemical composition at various altitudes, and

(2) In physical optics, to find the net spectral radiance of the resultant heterogeneous, nonisothermal, nonisobaric gas mass. Fortunately it appears that, except at very high altitudes, these two aspects may be considered independently. Some idea of the steps necessary in a first principle's calculation of this type are illustrated in Fig. 2-54 and some qualitative remarks are included in the subsequent paragraphs.

2-6.2.2.1 Chemical Composition

The spectral distribution of the plume radiation depends on the constituent molecular species. A representative selection of liquid fuel/oxidizer combinations is given in Table 2-22 together with their major and minor combustion

TABLE 2-22. COMBUSTION PRODUCTS FOR A SELECTION OF LIQUID PROPELLANTS

FUEL	OXIDIZER	MAJOR PRODUCTS	MINOR PRODUCTS
Hydrazine (N_2H_4)	Chlorine Trifluoride	HF, N_2, HCl	H_2, H, Cl
Hydrazine (N_2H_4)	Fluorine	HF, N_2	H_2, F, H
Hydrazine (N_2H_4)	Hydrogen Peroxide	H_2O, N_2, H_2	H, OH
Hydrazine (N_2H_4)	Oxygen	H_2O, N_2, H_2	H, OH
Hydrogen	Fluorine	H_2, HF	H
Hydrogen	Oxygen	H_2, H_2O	H, OH
RP-1 ($C_{10}H_{20}$)	Oxygen	CO, H_2O, CO_2, H_2	H, OH, O, O_2
Ethyl Alcohol (C_2H_5OH)	Oxygen	H_2O, CO, CO_2, H_2	H, OH, O
UDMH [$(CH_3)_2N_2H_2$]	IRFNA (HNO_3 + NO_2 + N_2O + HF)*	H_2O, N_2, CO, CO_2, H_2	H, OH, O, NO, HF
50/50 (UDMH + Hydrazine)	Nitrogen Tetroxide (N_2O_4)	H_2O, N_2, CO_2, H_2	H, OH, O, NO, O_2
Pentaborane (B_5H_9)	Nitrogen Tetroxide (N_2O_4)	H_2, N_2, HBO_2, B_2O_3	H, BO, H_2O, B_2O_2, OH, O, NO

* Note: UDMH = unsymmetrical dimethyl hydrazine
IRFNA = inhibited red fuming nitric acid

products under practical conditions. The exhaust composition, particularly for minor constituents depends on the oxidizer/fuel ratio and the operating conditions of the motor. As an example of this dependence, Table 2-23 gives the combustion products of LOX/RP-1 propellant for various oxidizer/fuel ratios. It should be noted that, for most efficient operation, rocket motors are run fuel rich, so that a considerable amount of the exhaust will be combustible material. This will mix and react exothermically with atmospheric oxygen to form the afterburning zone, which is characteristic of rocket plumes at low altitude, and cause an increase in plume temperatures of as much as 500°K. Afterburning does not always occur in the immediate vicinity of the nozzle. This ignition delay is due to the fact that the mixed region has to reach a finite minimum thickness in order to support combustion. As the altitude increases the degree of afterburning must decrease as atmospheric oxygen decreases.

In addition to the molecular species, there may be a number of particles present. While the amine fuels are relatively clean, the hydrocarbons produce varying amount of carbon particles. Solid fueled rocket exhausts are rich in particles. The combustion products of an aluminized polyurethane are given in Table 2-24. Note the large proportion of Al_2O_3 which may be in the form of solid particles or liquid, perhaps in a supercooled state[101]. The efflux may include other solid matter due to ablation of the nozzle; also, burnrate controllers in solid fuels can introduce particles such as ferric-oxide.

For more detailed information on the calculation of exhaust jet composition and on specific fuels the reader is referred to Refs. 101, 102, 103.

TABLE 2-23. LOX/RP-1 COMBUSTION PRODUCTS IN MOLE PERCENT CALCULATED AT CHAMBER PRESSURE 1000 PSIA, EXIT PRESSURE 14.696 PSIA WITH SHIFTING EQUILIBRIUM

	O/F RATIO			
SPECIES	3.0	2.6	2.2	1.8
H_2O	42.52	40.17	31.70	18.28
CO_2	30.92	22.96	14.94	10.70
CO	17.92	27.43	35.65	39.90
H_2	3.71	8.12	17.68	31.14
OH	2.08	0.33	—	—
H	0.84	0.41	0.03	—
O	0.38	0.01	—	—
O_2	1.64	0.03	—	—

TABLE 2-24. COMBUSTION PRODUCTS FOR METALLIZED SOLID PROPELLANT POLYURETHANE + 13% Al

Species	H_2O	CO_2	CO	H_2	HCl	Al_2O_3	N_2
Weight %	4.8	3.8	35.1	3.4	20.2	24.6	8.1

2-6.2.2.2 Plume Structure

A typical low altitude plume structure for a motor burning liquid propellant is depicted in Fig. 2-55. Immediately downstream from the nozzle is a region known as the undisturbed cone where the composition is essentially homogeneous and isothermal. Outside this cone, mixing with the ambient atmosphere occurs and is accompanied by afterburning.

As the altitude increases above the design optimum, the gases at the nozzle exit become increasingly underexpanded compared with the ambient atmosphere. Outside the undisturbed cone they expand rapidly and the kinetic temperature falls. The sort of variation to be expected is depicted in Fig. 2-56. For a clustered engine configuration the situation is complicated by the mutual interaction of the individual jets. Fig. 2-57 illustrates the temperature and pressure structure calculated for the near field of a Saturn II exhaust. In the far field the plume structure will be similar to that for a single motor.

Any efflux of particles will modify the plume to some extent, though it is believed that there is only weak coupling in two-phase flow, and that the situation is adequately represented by considering the effect of the gas plume on the particles, but not vice-versa. The temperature and velocity of the particles at the nozzle tend to lag that of the gases, i.e., the temperature will

tend to be higher and the velocity lower. In the plume they segregate according to size, which is generally of the order of microns but may vary by more than an order of magnitude. This segregation—defined by the limiting streamlines within which all particles of a given size are contained and the modification of these streamlines with altitude—is illustrated in Figs. 2-58 and 2-59 respectively. Typical near and far field isotherms are shown in Figs. 2-60 and 2-61. Note that shock heating at the Mach disc reheats the particles.

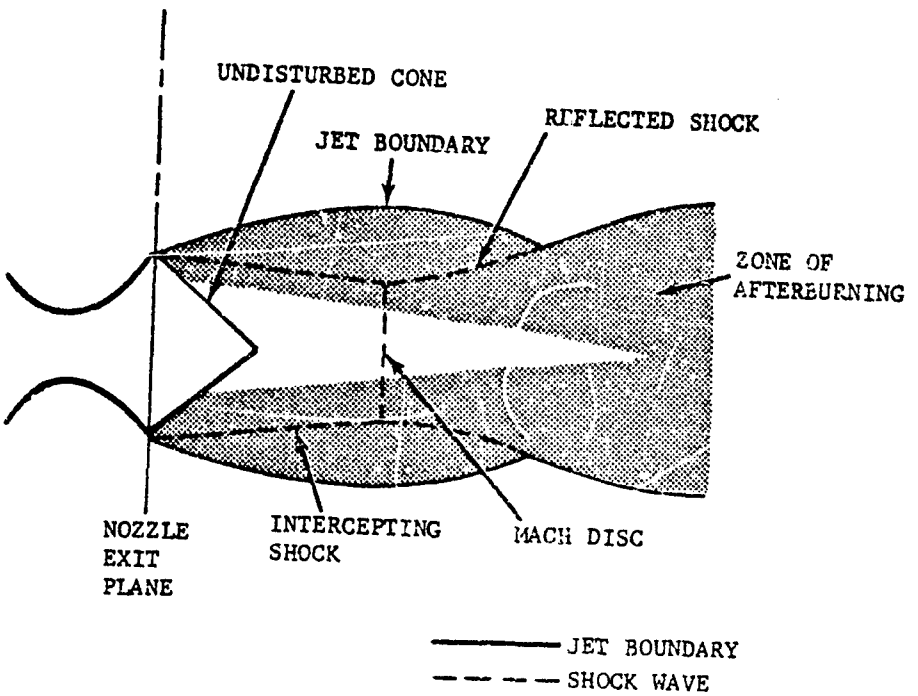


FIGURE 2-55. Low Altitude Plume Structure

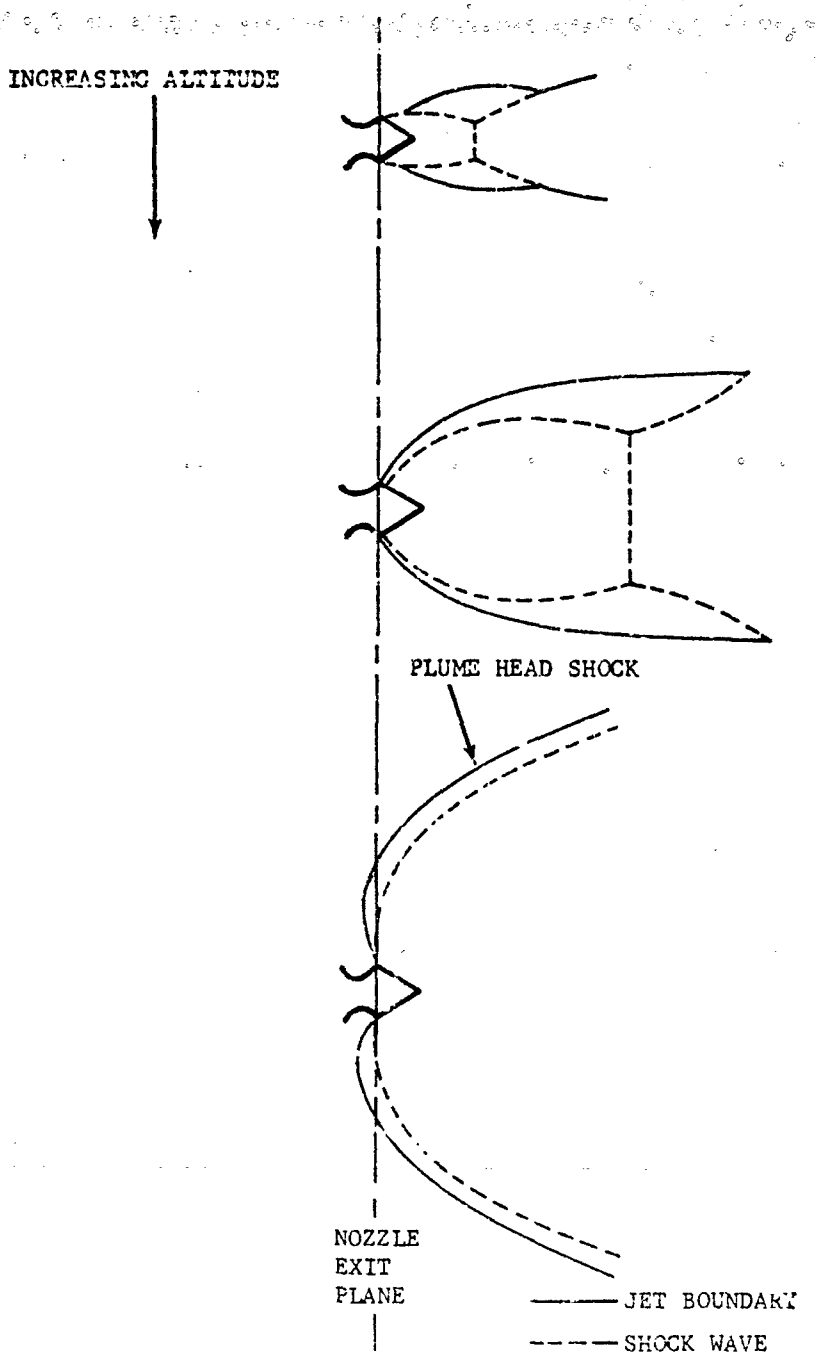


FIGURE 2-56. Plume Structure As a Function of Altitude

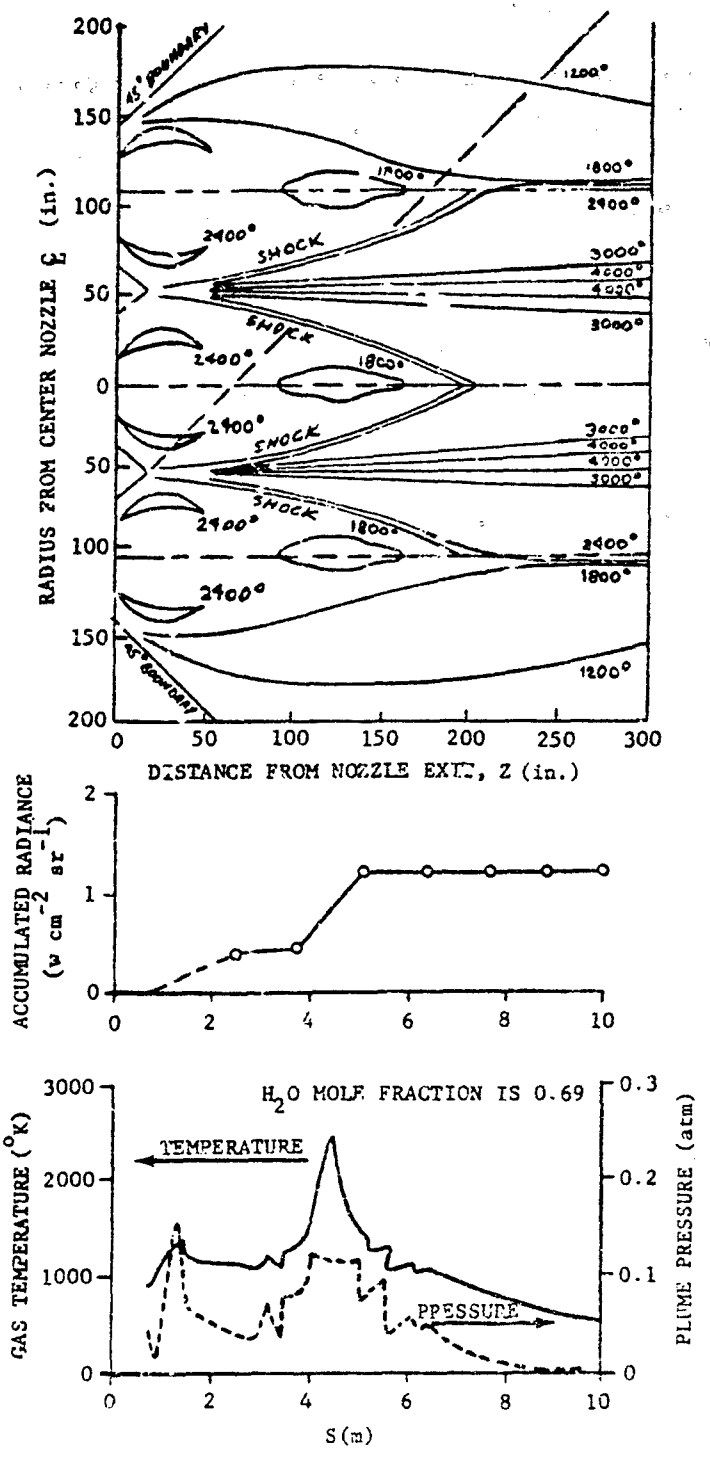


FIGURE 2-5.7. Typical Line-of-sight Variations on the Saturn S-II Stage

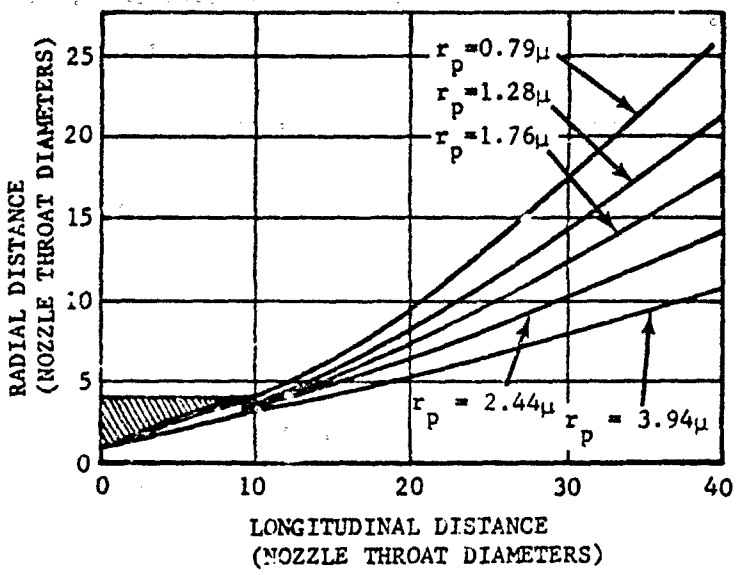


FIGURE 2-58. Limiting Particle Streamlines in the Near Field

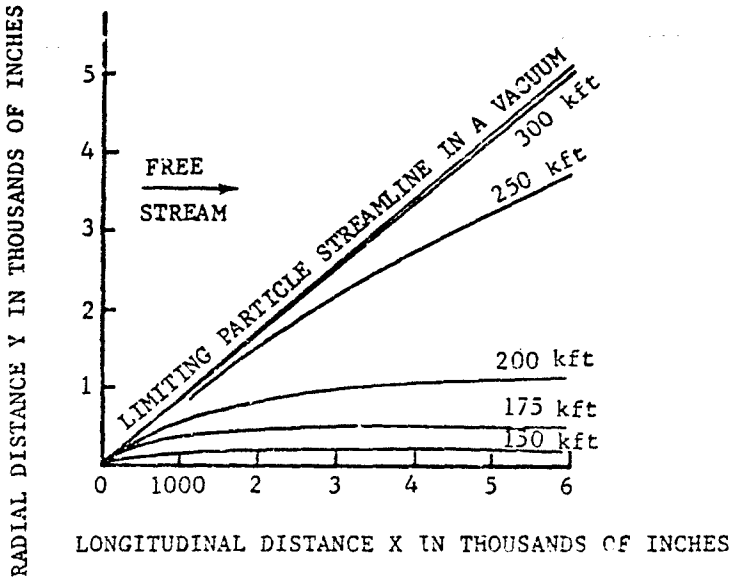
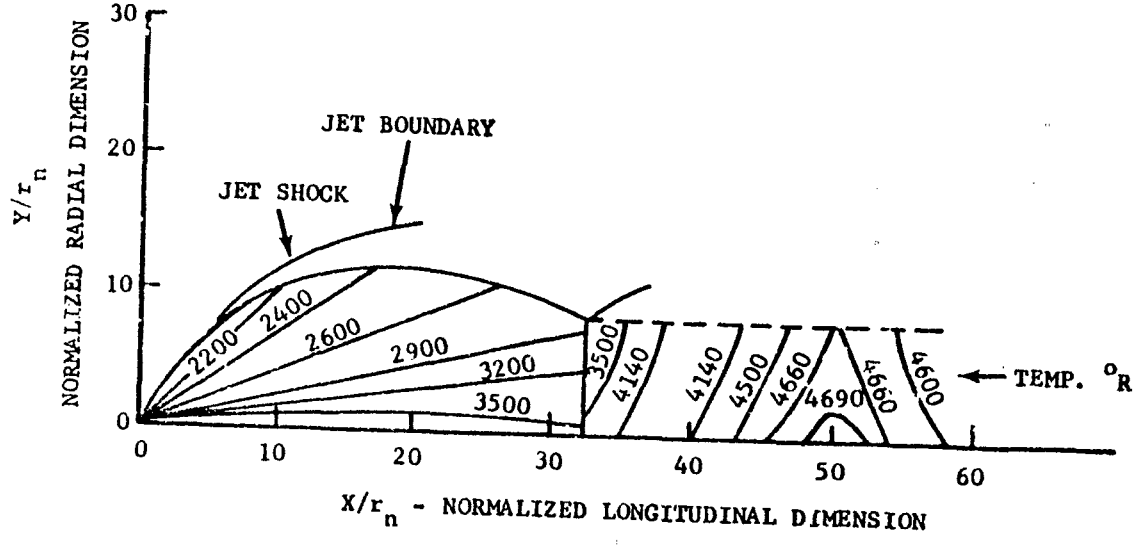


FIGURE 2-59. Trajectories of 0.79-micron Radius Particle External to the Gas Plume As a Function of Altitude



ALTITUDE - 100 kft; $T_c = 6100^{\circ}R$

(r_n = NOZZLE RADIUS)

FIGURE 2-60. Near Field Isotherms for 0.79-micron Radius Particle

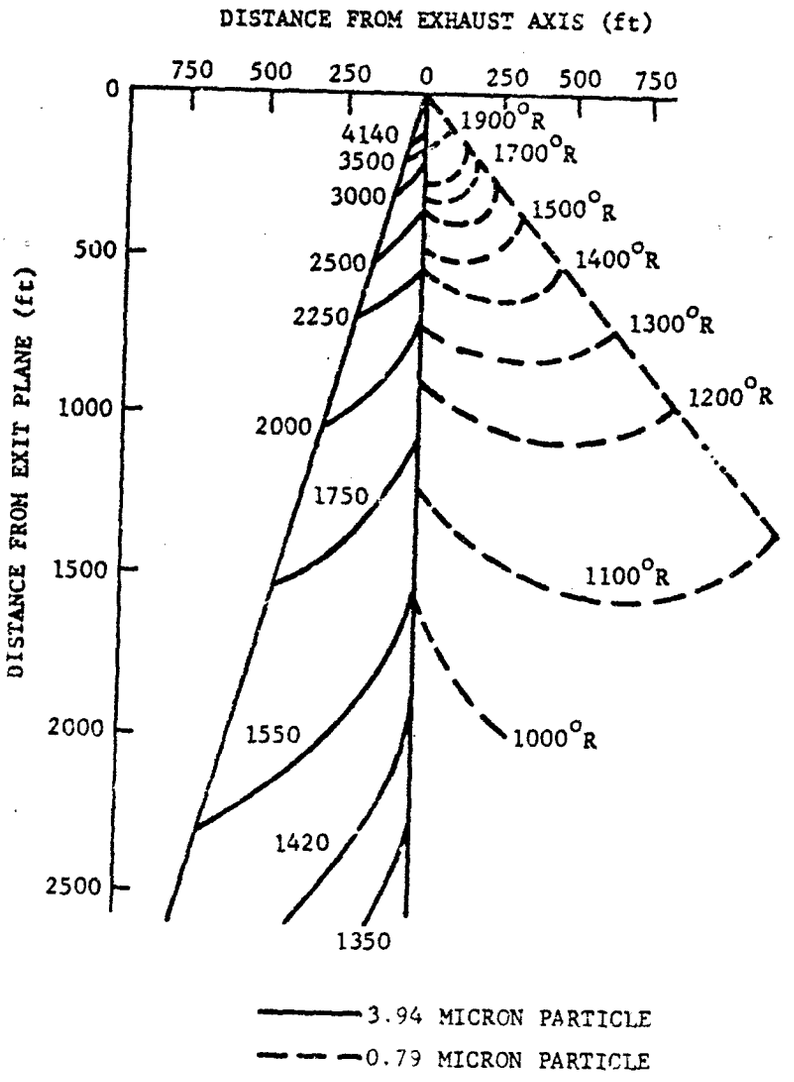


FIGURE 2-61. Far Field Temperature Profile, Vacuum Profile

2-8.2.2.3 Infrared Emission

The spectral radiance of the plume will consist of molecular band emissions superimposed upon a continuum due to emission and scattering by any particles present. The centers of the main infrared emission bands of molecular species commonly appearing as combustion products are given in Table 2-25. Fig. 2-62 is a spectra of the near field of a LOX/RP-1 exhaust showing the strong H_2O and CO_2 emission characteristic of hydrocarbon fuels. In Fig. 2-63 an emission spectrum of HF is shown. Note the presence of significant emission in the (3,2) and (2,1) bands at this temperature, in addition to the (1,0) fundamental. Of particular interest is the fact that this molecule emits quite strongly in the 2.3- to 2.4-micron spectral range where atmospheric transmission is comparatively good, especially at high altitudes.

Since their diameters are of the same order as the infrared wavelengths of interest, the particles cannot be expected to emit exactly as graybodies, but rather as Mie particles[46]. Fig. 2-64 shows a calculated spectrum for the near field of a rocket burning an aluminized solid propellant, and demostrates the importance of the particles' role in such plumes.

While it is not too difficult to predict the local radiance of small homogeneous volumes of the plume in thermodynamic equilibrium, predicition of the total spectral radiance as seen by a distant observer is extremely complex. As opposed to transmission problems which may be treated by a relatively simple absorption relation, the calculations for the emission/absorption problem require solution of the radiative transfer equation. For a discussion of this equation, and the utility of various band and plume models in its solution, the reader is referred to open literature, e.g., Refs. 104 and 105.

At high altitudes most of the radiation comes from the undisturbed core. The plume signature may be aspect independent, apart from the highly directional continuum radiation due to the nozzle and combustion chamber. The degree of anisotropy will depend on the size and number of any particles scattering radiation into other aspects. The plume head shock may also contribute. In the launch phase the afterburning zones are strong sources particularly in the vicinity of the Mach discs[106]. In general the radiance will decrease as the altitude increases. The radiation in the molecular bands, emitted by the hot zones of the plume, will be modified by absorption in the cooler outer regions. The radiance will thus depend quite strongly on the emission in the wings of the bands since absorption there decreases rapidly with temperature while increasing in the centers.

TABLE 2-25. MAJOR EMISSION BANDS FOR COMMON MOLECULAR COMBUSTION PRODUCTS

SPECIES	APPROXIMATE BAND CENTERS, MICRON
H_2O	1.14, 1.38, 1.88, 2.66, 2.74, 3.17, 6.27
CO_2	2.01, 2.69, 2.77, 4.26, 4.82, 15.0
HF	1.29, 2.52, 2.64, 2.77, 3.44
HCl	1.20, 1.76, 3.47
CO	1.57, 2.35, 4.66
NO	2.67, 5.30
OH	1.43, 2.80
NO_2	4.50, 6.17, 15.4
N_2O	2.87, 4.54, 7.78, 17.0

Figs. 2-65 and 2-66 are spectra for small liquid and solid fueled rockets, respectively, showing variation of the radiance with altitude, and also the differences between near and far field radiation[107]. Note the relative intensities of the 2.7-micron CO_2 + H_2O band and the 4.3-micron CO_2 band in the relatively cool far field as compared with the ratio in the much hotter near field, particularly at high altitude.

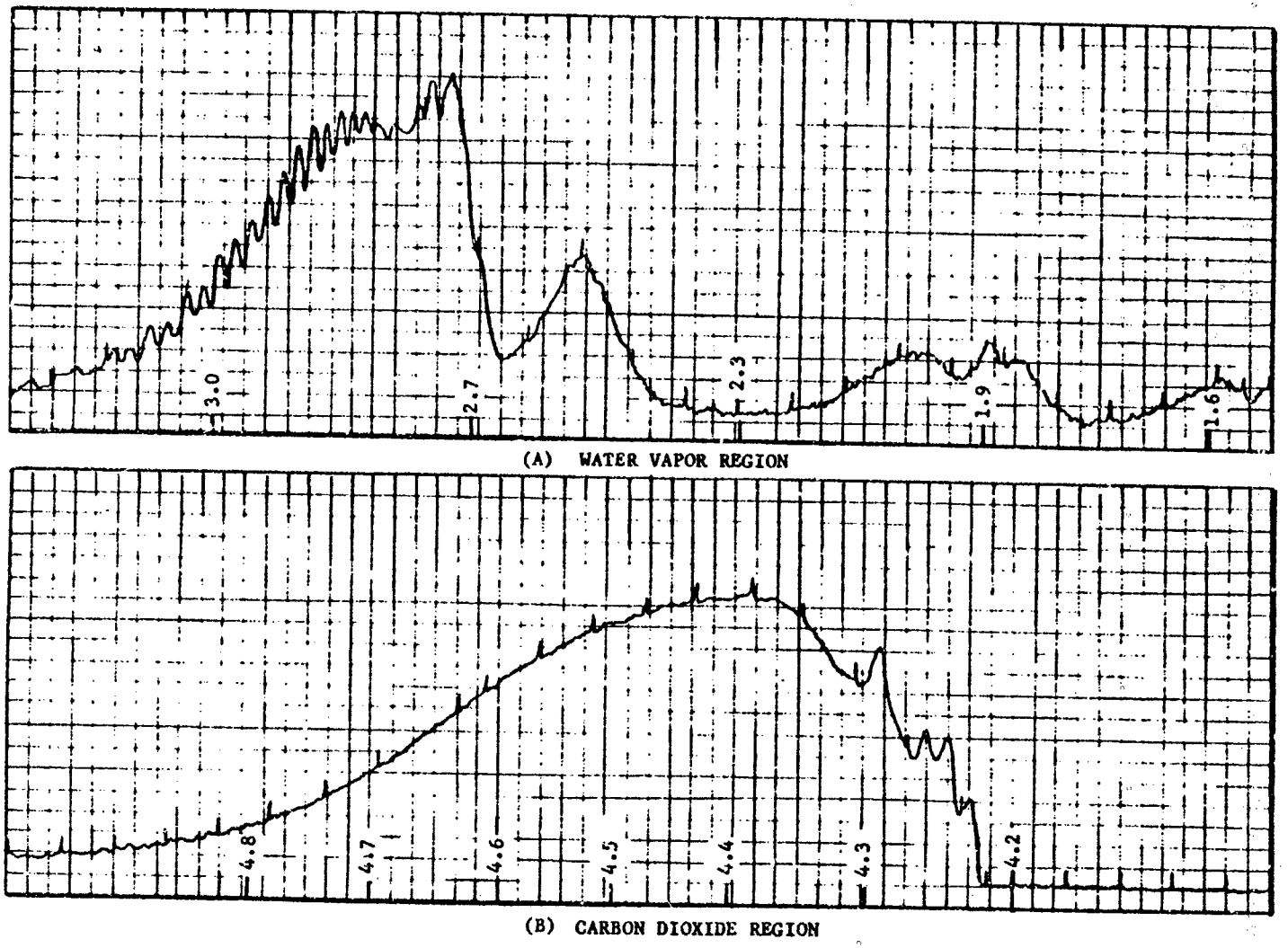


FIGURE 2-62. Water Vapor and Carbon Dioxide Regions

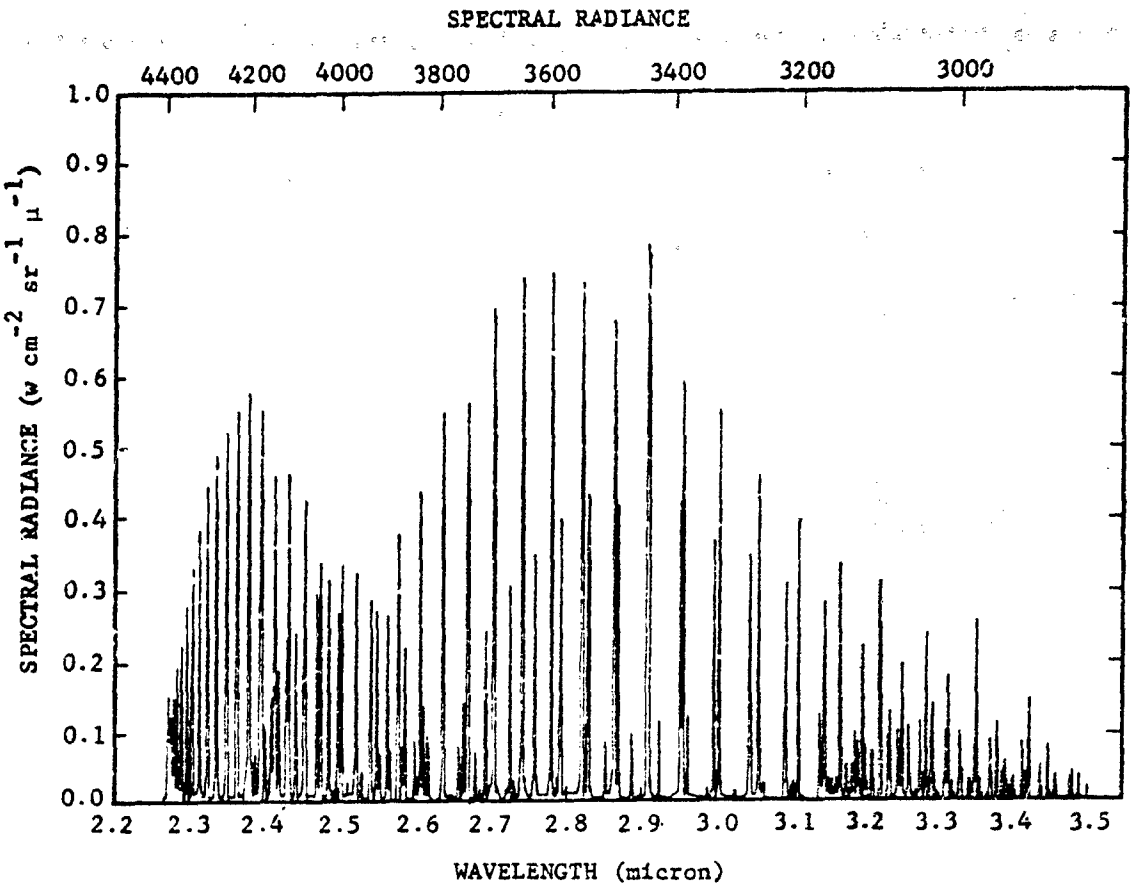


FIGURE 2-63. Emission Spectrum of HF for 0.5 cm-atm at 2640°K

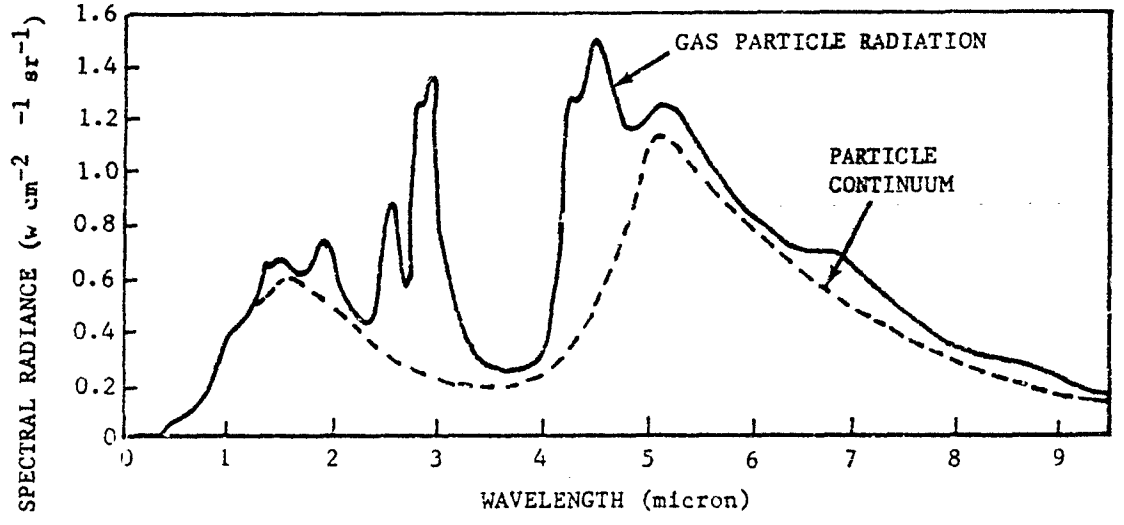


FIGURE 2-64. Theoretical Radiance Spectrum for a Solid Metallized Propellant Rocket Plume

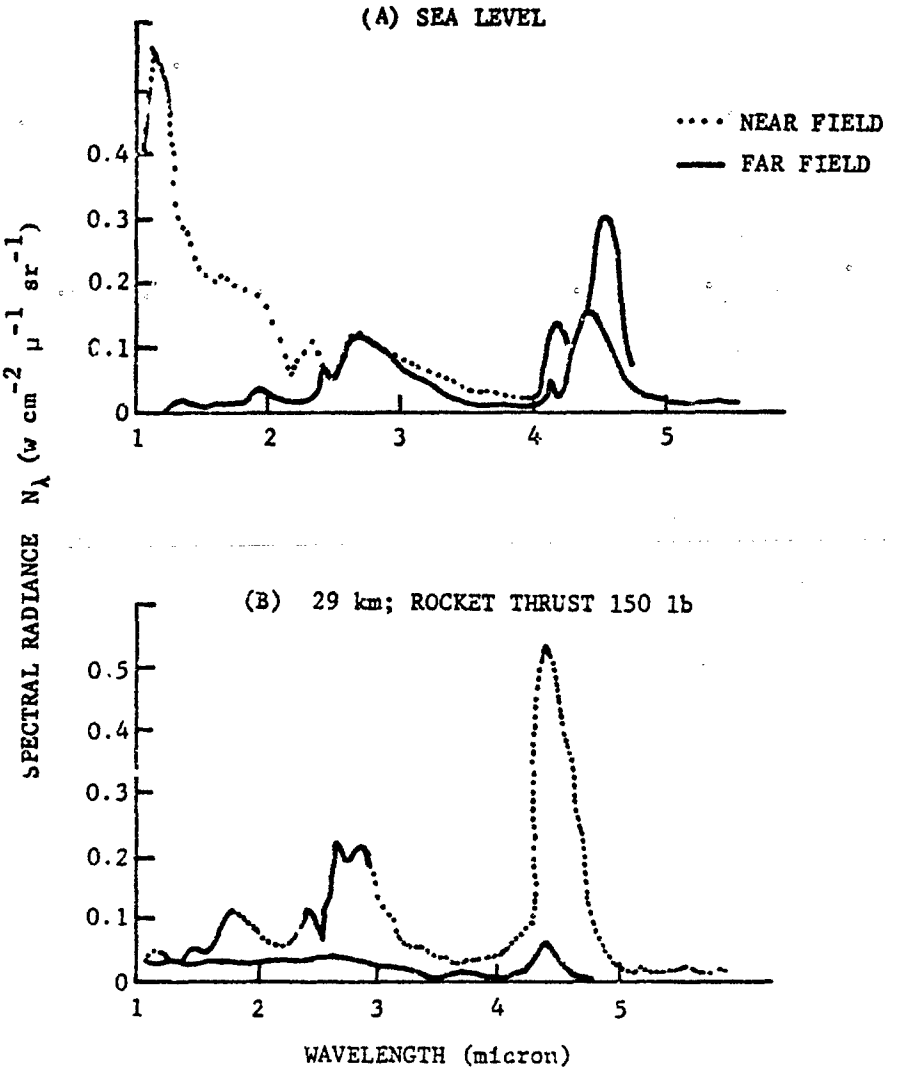


FIGURE 2-65. Spectral Radiance of RP1/O_2 Rocket Exhaust

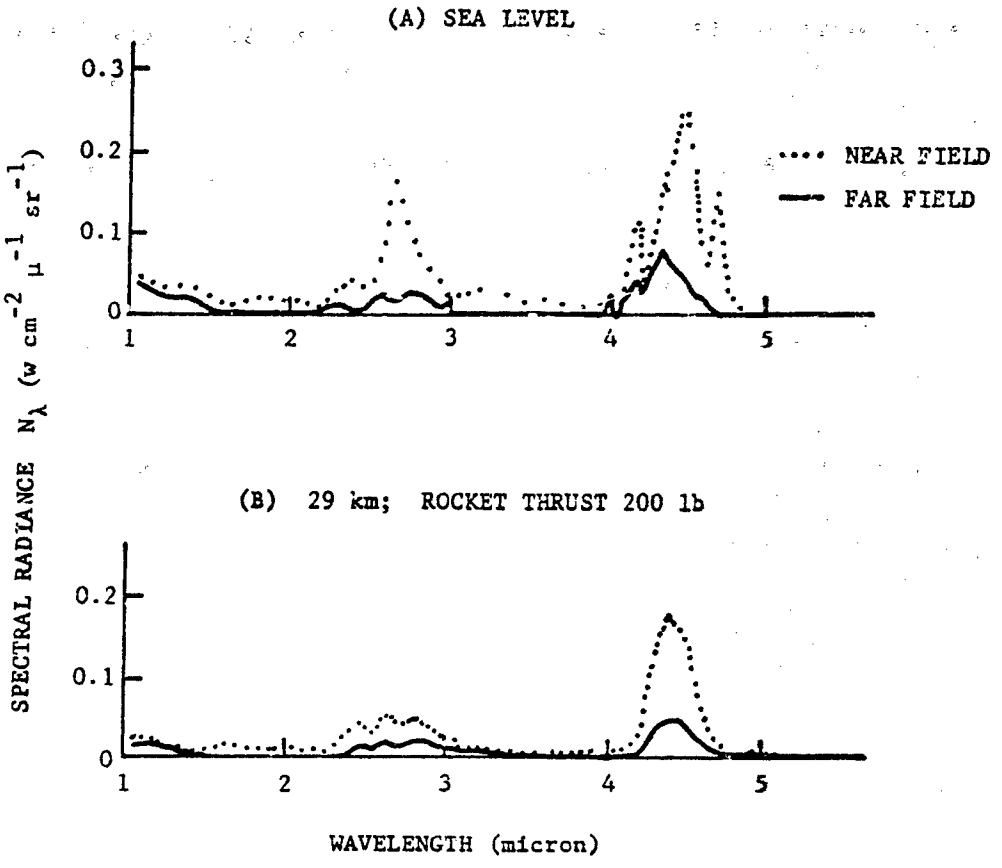


FIGURE 2-66. Spectral Radiance of Solid Propellant Rocket

2-6.2.3 Re-entry Vehicles

There are always four possible sources of radiation from a body re-entering the atmosphere: (1) the shock-heated air in front of the body, (2) the heated body surface, (3) ablation products injected into the boundary layer around the body, and (4) the wake behind the body. In the IR region the dominant radiation is almost always from the second and third sources; only in the visible and ultraviolet is the contribution from the heated air significant. The surface usually reaches temperatures in excess of 2500°K and solid particles (mostly carbon) may be sheared off into the boundary layer which can be still hotter.

The radiation from the ablated particles cannot be accurately calculated because neither the rate of particle formation during ablation nor the particle size distribution can be adequately determined from laboratory simulations. However, from the chemical composition of the heat shield, it can be determined whether or not enough particulate matter will be produced during ablation to dominate over the radiation from the hot surface. It has been found that only certain purely organic heat shields produce a char of insufficient strength to withstand the severe shear stress imposed during re-entry. All other materials (phenolic silicates comprising the largest family) retain their outermost layers. Consequently, they introduce primarily gases into the boundary layer. These gases produce intense ratiation which, like that due to air, is observed in the visible and ultraviolet[108]. CO, CO_2, and H_2O are indeed gaseous products from ablation, just as they are from missile fuel combustion (par. 2-6.2.2), but the amounts formed are orders of magnitude less.

For calculations of the aerodynamic flow around a hypersonic body the reader is referred to Refs. 109 or 110. The essential steps are outlined in the paragraphs which follow.

It is first necessary to determine whether most of the body surface will be covered with normally or obliquely shocked air. The body shape determines this, as illustrated in Fig. 2-67. If the body is blunt, the boundary layer is fed by air traversing the zone A_1 where the shock is nearly normal to the body surface and, therefore, strongest. For slender bodies, A_1 is negligible in comparison to A_2, a zone where the shock is much weaker, due to its oblique angle with respect to the body surface. Air from A_1 is very hot (several thousand degrees Kelvin) and relatively slow moving; the opposite is true for A_2. Although techniques have been developed for determining the thermodynamic properties of air under either condition, different methods must be used for different altitude regimes.

Once the properties of the boundary layer flow are determined, the heat must be transferred to the body surface via convection, radiation, and conduction. Mathematical relationships for each case are available. Essential inputs include the density, thermal conductivity, and heat capacity of the heat shield—properties which will vary during the course of re-entry. The heat transfer calculations provide a thermal map of the surface. Only in rare cases will the surface be isothermal, although this is often assumed at the altitude of peak heating.

In order to convert the thermal map of the body surface to observed radiation, the map must be divided into a number of isothermal zones. By use of the blackbody slide rule, radiation from each zone in the direction of the observer can be calculated. An emittance of 0.9 is usually satisfactory in the IR unless more accurate data are available[111]. In order to determine the angular distribution of radiation, the surface is usually assumed to be Lambertian in the absence of better information.

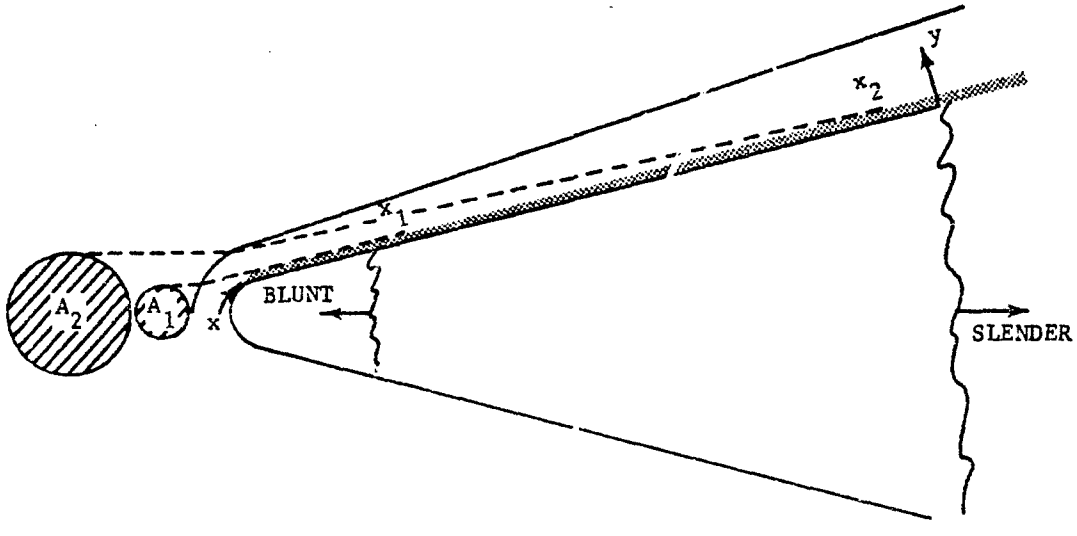


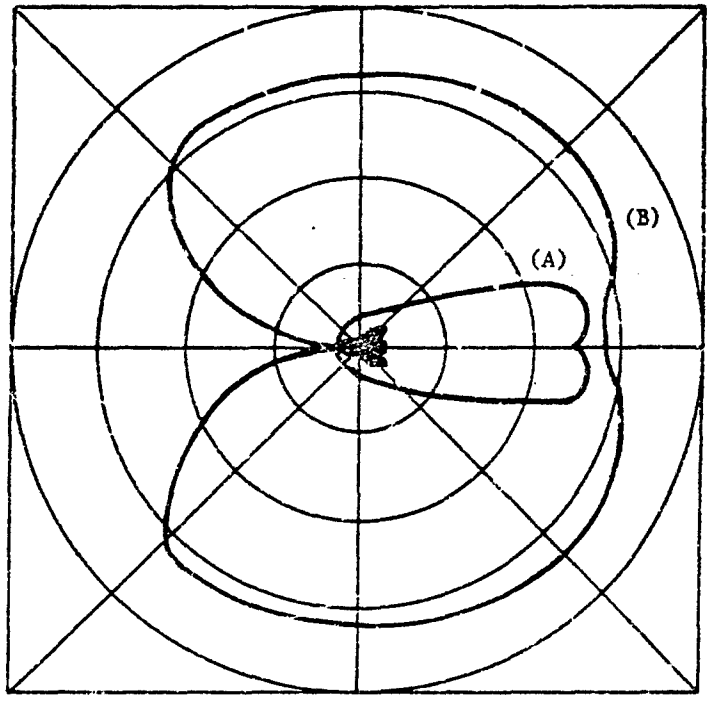
FIGURE 2-67. Flow Field Around Blunt and Slender Vehicles

2-3.2.4 Aircraft

The four sources of IR radiation from a jet aircraft are:

1. Hot metal parts of the turbine section and exhaust nozzle
2. Exhaust gases
3. Aerodynamically heated surfaces
4. Reflected sunlight

The two modes of engine operation, without afterburner (military power) and with afterburner, should be considered separately. Under military power, the most significant source of radiation is the hot metal parts inside the tailpipe. This source generates high emissivity graybody radiation. However, being inside the tailpipe, these parts are visible only from the rear. As a side aspect viewing angle is approached, aircraft radiation drops off rapidly as hot parts become blocked from view. Thus, from the side and front, aircraft radiation is low and due mainly to exhaust gases in the plume. Most radiation in the plume is due to H_2O and CO_2 molecules, the major combustion products of kerosene fuels. The main emission thus occurs in the 2.7-micron CO_2 and H_2O bands and the 4.3-micron CO_2 band. Under military power, exhaust gases are the result of lean fuel; therefore, very little, if any, afterburning takes place in the plume. At the relatively low temperatures involved, the gases are poor radiators. A typical plume spectrum, under military power, would be similar to that shown in Fig. 2-68 but with the 4.3-micron CO_2 band much more intense than the 2.7-micron H_2O band. When observed from a distance, this radiation will be strongly absorbed by atmospheric CO_2 and H_2O molecules. Thus for all practical purposes the effective plume radiance is in the wings of the bands.



(A) UNDER MILITARY POWER, (B) WITH AFTERBURNER

FIGURE 2-68. Typical Radiant Signature of a Jet Aircraft

Aerodynamic heating is not significant at the low speeds involved under military power, except perhaps at nose aspect, when most of the plume is obscured. The resulting aircraft radiation signature due to these sources is shown in Fig. 2-68(A).

In the afterburning case, the plume becomes the most significant contributor of radiation at all aspects. Radiation will peak in the side aspect, where the apparent area of the plume is greatest. At nose aspect, most of the plume is blocked by the aircraft, and radiation will be at a minimum. At these higher plume temperatures the 2.7-micron band will increase in importance relative to the 4.3-micron band. In fact, for any plume radiation, the 2.7-micron band will be much more sensitive to temperature variations than the 4.3-micron band due to the nature of blackbody spectral variation with temperatures. Furthermore, the higher plume temperature broadens the emission bands, which is especially significant since the widths of atmospheric absorption bands do not vary. Fig. 2-68(B) illustrates the signature that results from the afterburning plume. At high Mach numbers, aerodynamic heating becomes a higher significant source of radiation and exhibits graybody spectral distribution.

Determining the solar reflection from an aircraft is a highly complex problem. The spectral distribution of reflection will resemble that of the solar 6000°K blackbody radiation modified by atmospheric transmission. The magnitude of such reflection depends on: (1) the angle between the sun, aircraft, and observers, (2) the shape of the reflecting surfaces, (3) the type of reflecting surface, i.e., diffuse, and/or specular, and (4) the reflectivity of the surface.

Helicopter radiation is very similar to that from aircraft operating in the military power mode. However, the radiation signature may differ since the engine does not necessarily exhaust to the rear.

2-6.2.5 Clear Air Turbulence

Advance knowledge of atmospheric turbulence can be extremely useful to a pilot. Clear air turbulence may be detected in advance, using remote infrared sensing devices by virtue of the fact that it is accompanied by sharp temperature gradients[112]. Observation of sharp fluctuations of the infrared radiance of the atmosphere ahead of an aircraft indicates the presence of such turbulence and, if the spectral bandwidth accepted by the instrument is carefully chosen, the detection ranges may be of the order of tens of kilometers. The spectral region chosen is commonly around 13 to 14 microns, the high-frequency wing of the 15-micron CO_2 band. In the wings of a band the emissivity, though small, depends strongly on the temperature, and there will be a significant change in emissivity even for the small temperature changes of the order of 5°C associated with turbulence. At the same time the atmospheric absorption in this region is low enough that the variation in radiance can be seen over long path lengths.

2-6.2.6 Ground Targets

Infrared radiation from ground targets, under passive surveillance, is due to thermal emission and reflected solar energy. Most surface targets will be opaque or nearly so, and many objects of interest—such as roads and bridges having no internal heat supply—will be close to the ambient temperature. For these, the radiation is exactly as described in par. 2-6.2 and the emission will be different from that of the background by virtue of different emissivity or small differences in temperature. For example, since small surface area per unit mass tends to reduce temperature fluctuations, concrete roads may be expected to be cooler by day and warmer by night than gravel or dirt roads. The recent presence of stationary vehicles or camps may be detected by its "shadow", where the ground temperature may remain different from the surroundings for some time after the vehicle or camp itself has been removed. In real time reconnaissance, the recognition and identification of such targets must depend, to a great extent, upon evaluation of other factors such as shape, size, and context. Near ambient temperature (300°K) blackbody emission is a maximum around 9.5 microns and the maximum variation of spectral radiance with temperature is in the region of 8 microns. For detection of targets of reasonably high emissance, it is therefore advantageous, given suitable detectors, to use the 8- to 14-micron atmospheric window. In this spectral range the emissivity of most unpolished objects is quite high. For example, dirt has an emissivity of 90 to 95% while O.D. paint and oxidized metals have emissivities of 75 to 85%.

Few objects are truly gray and multispectral remote sensing—i.e., simultaneous measurements of albedo in several spectral channels—shows considerable promise as a reconnaissance technique and has been applied in photographic IR for some time. By measuring reflectance or emissance as a function of wavelength, to obtain a "spectral signature", less reliance need be placed on the existence of temperature differences, and definition of shape and context. This technique presently shows particular promise as a means of remote determination of ground conditions—such as type of vegetation, type of soil or rock, and degree of wetness (e. g., Refs. 113 and 114)—which affect the trafficability of terrain. A compilation of a large number of spectral signatures may be found in Ref. 115 for both natural and man-made objects.

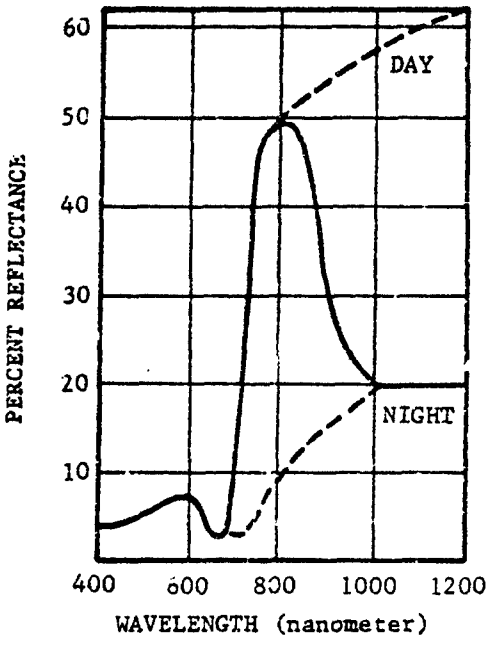


FIGURE 2-69. Idealized Reflectance Curve of a Uniform Fabric to Afford Camouflage Protection Over a Spectral Range from 0.4 to 1.2 Microns

Many ground targets—such as buildings, vehicles, and personnel—have an internal heat supply and may be much warmer than their surroundings; also, parts of a truck or tank will remain hotter than its surroundings for several hours after use. A factory chimney stack or cooling tower may be a very strong infrared source. In all cases, however, the emissivity must be considered.

Special low-emissivity paints can greatly reduce the radiation from a hot surface of a vehicle or building. However, simultaneous camouflage against both visual and infrared detection throughout the spectrum is extremely difficult. This applies not only to structures and vehicles but to personnel as well. Fig. 2-69 shows an idealized reflectance signature for a uniform fabric that provides protection against visual observations and near infrared photographic detection by day, and sniperscope detection by night[116]. Comparison with the reflectance signature of a typical Army uniform cloth shown in Fig. 2-70 is instructive. Compare also the reflectance characteristics of human skin in Fig. 2-71.

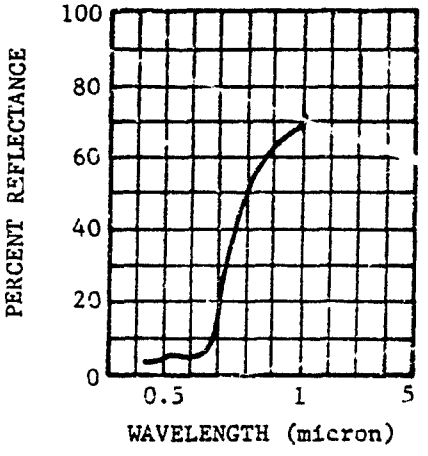


FIGURE 2-70. Reflectance from Typical Uniform Cloth

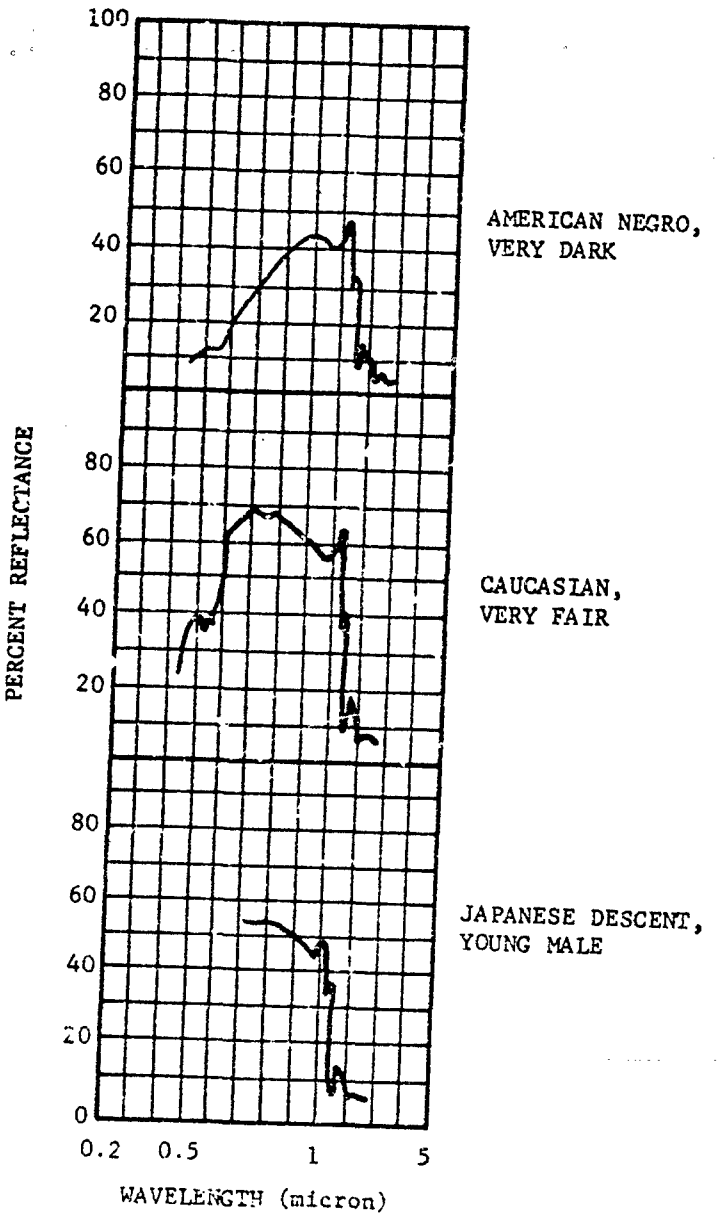


FIGURE 2-71. Reflectance from Human Skins

REFERENCES

1. *Handbook of Military Infrared Technology*, Office of Naval Research, Department of Navy, Washington, D. C., 1965.
2. E. R. Cohen and J. W. M. DuMond, "Our Knowledge of the Fundamental Constants of Physics and Chemistry", J. Rev. Mod. Physics 37, 537-594 (1965).
3. M. M. Fulk, M. M. Reynolds and O. E. Park, Proc. Cryogenic Engineering Conference, N. B. S. Report 3517, 151, (1954).
4. W. H. McAdams, *Heat Transmission*, McGraw-Hill Book Co., N. Y., 1938.
5. K. G. Ramanathan, Proc. Phys. Soc. (London) A65, 532 (1952).
6. M. B. Blackman, A. Egerton, and E. V. Truter, Proc. Roy. Soc. (London) A194, 147 (1948).
7. W. T. Ziegler and H. Cheung, Proc. Cryogenic Engineering Conference. N. B. S., Boulder, Colo. (1957).
8. M. Planck, "Regarding the Articles on Energy in the Normal Spectrum", Ann. Physik 4, 553 (1901).
9. H. L. Hackforth., *Infrared Radiation*, McGraw-Hill Book Co., N. Y., 1960.
10. P. J. Wyatt. V. R. Stull, and G. N. Plass, "Quasi-Random Model of Band Absorption", J. Opt. Soc. Am. 52,1209 (1962).
11. P. J. Wyatt, V. R. Stull, and G. N. Plass, *The Infrared Absorption of Water Vapor (Final Report)*, Report No. SSD-TDR-62-127, Vol. II, Aeronutronics Div. of Ford Motor Co., Newport Beach, California, 1962.
12. V. R. Stull, P. J. Wyatt, G. N. Plass, *The Infrared Absorption of Carbon Dioxide (Final Report)*, Report No. SSD-TDR-62-127, Vol. III, Aeronutronics Div. of Ford Motor Co., Newport Beach, California, 1962.
13. R. M. Goody, *Atmospheric Radiation*, Clarendon Press, Chapter 4 (1964).
14. D. Anding, *Band Model Methods for Computing Atmospheric Slant Path Molecular Absorption*, IRIA State of the Art Report-NAVSO P 1499-1, Willow Run Labs, University of Michigan, Ann Arbor, Michigan, 1967.
15. C. E. Junge, "Atmospheric Chemistry", Advances in Geophysics 4, Pergamon Press (1958).
16. G. M. Dobson, *Exploring the Upper Atmosphere*, Clarendon Press, Chapter 6. 1963.
17. R. N. Kulkarni, "Comparison of Ozone Variations and of its Distribution With Height Over Middle Latitudes of the Two Hemispheres", Quart. J. Roy. Meteorol. Soc. 88, 522 (1962).
18. M. Gutnick, "Mean Moisture Profiles to 31 km for Middle Latitudes", Interim Notes on Atmospheric Properties 22, Geophysics Research Directorate AFCPC (1962).
19. J. A. Jamieson, et al., *Infrared Physics and Engineering*, McGraw-Hill Book Co., N. Y., 1963.
20. R. Ladenburg and F. Reiche, "Selective Absorption by Coloured Flames". Ann. Physik 42, 181 (1913).
21. S. N. Plass, "Models for Spectral Band Absorption". J. Opt. Soc. Am. 48, 690 (1958).
22. W. Elsasser, *Harvard Meteor. Studies No. 6*, Harvard Univ., Cambridge, Mass., 1942.
23. G. Herzberg, *Molecular Spectra and Molecular Structure, Part II, Infrared and Raman Spectra*, Van Nostrand, New York, 1945.
24. *Handbook of Mathematical Functions*, U. S. Dept. of Commerce, 1964.
25. *Handbook of Physics and Chemistry*, Chemical Rubber Company, Cleveland, Ohio, 1965-66.
26. R. M. Goody, "A Note on the Regular Model of an Absorption Band", Quart. J. Roy. Meteorol. Soc. 87, 428 (1961).

27. S. N. Plass, "A Method for the Determination of Atmospheric Transmission Functions from Laboratory Absorption Measurements", J. Opt. Soc. Am. 42, 677 (1952).

28. G. Weyl, *Infrared Absorption by the Atmosphere of Flame Radiation*, Aerojet-General TM-527-61-13-658, Azusa, Cal., 17 Nov 1961.

29. E. E. Burch and D. Williams, *Infrared Absorption by Minor Atmospheric Constituents*, Ohio State University Research Foundation, Report No. 1 AFCRL Report No. TN-60-674, Cambridge, Mass., 1960.

30. J. Howard, D. Burch, and D. Williams, *Near Infrared Transmission Through Synthetic Atmospheres*, Geophysics Research Paper No. 40, AFCRC Report TR-55-213, Geophysics Research Directorate, Cambridge, Mass., 1955.

31. D. Burch, E. Singleton, W. France, and D. Williams, *Infrared Absorption by Minor Atmospheric Constituents*, Ohio State University Foundation, Final Report, AFCRL Report No. 412, Cambridge, Mass. 1960.

32. C. Walshaw, "Integrated Absorption by the 9.6μ Band of Ozone", Quart. J. Roy. Meteorol. Soc. 83, 315 (1957).

33. J. Strong, "Study of Atmospheric Absorption and Emission in the Infrared Spectrum", J. Franklin Inst. 232, 1 (1941).

34. J. Taylor, H. Yates, "Atmospheric Transmission in the Infrared", J. Opt. Soc. Am. 47, 223 (1957).

35. J. Taylor, H. Yates, *Infrared Transmission of the Atmosphere*, NRL Report 5453, U. S. Naval Res. Lab., Washington, D. C. 1960.

35A. F. R. Stauff and T. E. Walsh, "Transmittance of Water Vapor—14 to 20 Microns", J. Opt. Soc. Am. 56, 401 (1966).

36. U. Bradford, T. McCormick, and J. Selby, *Predicting the Molecular Absorption of Infrared Radiation Over Atmospheric Paths*, Report No. DMP 1422, EMI Electronics, Haves, Middlesex, England, 1963.

37. L. Abels, *A Study of the Total Absorption Near 4.5 Microns by Two Samples of N_2O as Their Total Pressure and N_2O Concentrations Were Independently Varied*, Sci. Rpt. No. 3, AFCRL-62-236, Ohio State University, Columbus, Ohio, 1962.

38. G. Morton and G. Weyl, *Water Vapor and Carbon Dioxide Absorption in the Spectral Region Around 2.7 Microns*, Aerojet-General Special Report No. 2075, Azusa, Cal., 1961.

39. T. Elder and J. Strong, "The Infrared Transmission of Atmospheric Windows", J. Franklin Inst. 255, 189 (1953).

39A. J. L. Streete, "Infrared Measurements of Atmospheric Transmission at Sea Level", Appl. Opt. 8, 1595, (1968).

40. K. Bignell, F. Saiedy, and P. A. Sheppard, "On the Atmospheric Infrared Continuum", J. Opt. Soc. Am. 53, 466 (1963).

41. W. Roach, and R. Goody, Quart. J. Roy. Meterol. Soc. 84, 313 (1958).

42. G. Weyl, *Atmospheric Transmission in the 8 to 14 Micron Band*, Aerojet-General TM-3533:68-6-835, Azusa, Cal., 1968.

43. W. Godson, "The Computation of Infrared Transmission by Atmospheric Water Vapor", J. Meteorol. 12, 272 (1955).

44. G. Morton, T. M. Tant, *A Computer Routine for the Calculation of 2.7 Micron Region Slant Path Transmission*, Aerojet-General TM 527:62-7-674, Azusa, Cal., 1962.

45. G. Morton, M. Bressler, G. Weyl, and George Greenstein, *Background and Atmospheric Transmission for Ballistic Missiles Detection Systems*, Aerojet-General Special Report No. 1971, Avionics Division, Azusa, Cal., 1961.

46. Van de Hulst, *Light Scattering by Small Particles*, John Wiley and Sons, N. Y., 1964.

47. L. Elterman, *A Model of a Clear Standard Atmosphere for Attenuation in the Visible Region and Infrared Windows*, AFCRL-63-675, Office of Aerospace Research, USAF, Hanscom Field, Mass., 1963.

48. J. A. Curcio, G. L. Knestrick, and T. H. Cosden, *Atmospheric Scattering in the Visible and Infrared*, NRL Report 5567, U. S. Naval Research Laboratory, Washington, D. C., 1961.

49. L. Elterman, *Atmospheric Attenuation Model, 1964, in the Ultraviolet, Visible, and Infrared Regions for Altitudes to 50 km*, AFCRL-64-740, Office of Aerospace Research, USAF, L. G. Hanscom Field, Mass., 1964.

50. D. Deirmendjian, "Scattering and Polarization Properties of Water Clouds and Hazes in the Visible and Infrared", Appl. Opt. 3, 187-196 (1964).

51. C. E. Junge, *Atmospheric Composition, Handbook of Geophysics*, The Macmillan Co., N. Y., 1960.

52. L. Elterman, *An Atlas of Aerosol Attenuation and Extinction Profiles for the Troposphere and Stratosphere*, AFCRL-66-828, Office of Aerospace Research, USAF, Hanscom Field, Mass., 1966.

53. R. Penndorf, "Tables of the Refractive Index for Standard Air and the Rayleigh Scattering Coefficient for the Spectral Region Between 0.2 and 20.0 Microns and Their Application to Atmospheric Optics", J. Opt. Soc. Am. 47, 176 (1957).

54. *Restoration of Atmospherically Degraded Images*, Woods Hole Summer Study, National Academy of Sciences, July 1966.

55. *Atmospheric Turbulence and the Scintillation of Starlight*, Rand Corporation Report R-406-PR, Sept. 1962.

56. V. I. Tatarski, *Wave Propagation in a Turbulent Medium*, McGraw-Hill Book Co., N. Y., 1961.

57. J. I. Davies, "Consideration of Atmospheric Turbulence in Laser Systems Design", Appl. Opt. 5, 139-146 (1966).

58. J. R. Whittier, "Q-Switched Laser Beam Propagation Over a Ten-Mile Path", Proc. IEEE 53-5, 736 (1965).

59. W. R. Hinchman, "Fluctuations in a Laser Beam Over 9- and 90-Mile Paths", Proc. IEEE 52-3, 305-306 (1964).

60. D. L. Fried, "Optical Heterodyne Detection of an Atmospherically Distorted Signal Wave Front", Proc. IEEE 55-1, 57-67 (1967).

61. C. B. Emmanual, et al., *Optical Scintillation; A Survey of the Literature*, N.B.S. Tech. Note, 255, Washington, D.C., 5 April 1965.

62. J. D. Jackson, *Classical Electrodynamics*, John Wiley and Sons, Inc., N. Y., 1962.

63. S. Ramo and J. R. Whinnery, *Fields and Waves in Modern Radio*, John Wiley and Sons, Inc., N. Y., 1953.

64. J. A. Stratton, *Electromagnetic Theory*, McGraw-Hill Book Co., N.Y., 1941.

65. Mark Kronstein, *Research Studies and Investigation on Spectral Reflectance and Absorption Characteristics of Camouflage Paints, Materials and Natural Objects*, Final Report, New York Univ., AD-100 058, March 1955.

66. H. H. Blau, J. Miles, and L. Ashman, *Thermal Characteristics of Solid Materials, A Review*, Arthur D. Little, Inc., Scientific Report No. 1. AFCRC TN 58-132, AD-146-883, 1958.

67. W. Starr, E. Streed, and A. Funai, *Principles of Camouflage for Low Temperature Targets*, U. S. Naval Civil Engineering Research and Evaluation Laboratory, Tech. Note N148, Port Hueneme, Calif., 1953.

68. G. G. Gubareff, J. E. Janssen, and R. H. Torborg, *Thermal Radiation Properties Survey*, Honeywell Research Center, Minneapolis-Honeywell Regulator Co., Minneapolis, Minn., 1960.

69. *Infrared Intelligence Analysis Study, Vol. 1 Theoretical Analysis*, H. R. Singer, Inc., Contract AF 30(602)-2683, Technical Documentary Report RADC-TDR-63-537, AD-350 048, 1963.

70. L. D. Miller and R. Horvath, *Diurnal and Seasonal Variation in Radiation of Objects and Backgrounds 4.5—5.5 Micron Spectral Region* (Project Michigan), Infrared and Optical Sensor Lab., University of Michigan, Ann Arbor,Mich., 1965.

71. E. Bell, et al., *Infrared Techniques and Measurements, Final Engineering Report*, R. F. Project 659, Contract No. AF 33(616)-3312, Ohio State University, Columbus, Ohio, AD-151 221, 1957.

72. W. R. Fredrickson, N. Ginsburg, and R. Paulson, *Infrared Spectral Emissivity of Terrain*, Final Report, Int. Dev. Report No. 2, AF 33(616)-5034, Syracuse Research Institute, Syracuse, N. Y., 1957.

73. R. H. Wunderley and M. E. Compton, Proc. IRIS 10, 1, 83 (1965).

74. C. D. Miller, J. O. Morgan, Proc. IRIS 7, 2, 231 (1962).

75. *The Tactical Interpretation of Selected Terrain Features on An/UAS-4 Infrared Imagers*, Intelligence Material Development Office, U. S. Army Electronics Laboratories, U. S. Army Electronics Command, Ft. Monmouth, N. J., Contract No. DA-18-109-AD-667, AD-366 134 (Confidential).

76. *An IR Reconnaissance Simulation Study*, AF Avionics Laboratory, Research and Technology Division, Air Force Systems Command, Wright Patterson Air Force Base, Ohio, Report No. ASD-TDR-63-880, AD-346 609, 1964 (Confidential).

77. *Final Engineering Report on Mercury Doped Germanium Detector Evaluation*, Report C1-61147-1, Texas Instruments Incorporation, Texas, Contract No. DA-36-039-SC-88977, 1961 (Confidential).

78. *Detection of Surface and Subsurface Feature in Greenland by Visual, Thermal and Radar Imagery*, Project Michigan, Report No. 2900-241-7, U. S. Army CRREL Research Report 91, Infrared Laboratory, Institute of Science and Technology, Univ. of Michigan, Ann Arbor, Mich., AD-325 637, 1961 (Secret).

79. *IR Intelligence Analysis Study*, Vol. 1, Report No. RADC-TDR-63-537, HRB Singler Incorporation, State College, Pennsylvania, Contract AF 30(602)-2683, AD-350 048, 1963 (Confidential).

80. H. McMahon, "Thermal Radiation from Partially Transparent Reflective Bodies", J. Opt. Soc. Am 40, 376 (1950).

81. A. Bondein, et al., *Infrared and Reflected Solar Measurements from the Tiros III Meteorological Satellite*, NASA Technical Note D-10967, 1961.

82. R. Vickers and R. Lyon, "Infrared Sensing from Spacecraft-A Geological Interpretation", AIAA 67, 284 (1967).

83. W. Nordberg, et al., *Preliminary Results and Radiation Measurements From the Tiros III Meteorological Satellite*, NASA Technical Note D-1338, 1962.

84. C. Cox and W. Munk, *Bull. Scripps Inst. Oceanog.*, University of California, 6, 401 (1956).

85. P. N. Tverskoi, *Physics of the Atmosphere*, Leningrad 1962, NASA Translation TTF-288.

86. H. R. Byers and R. R. Braham, *The Thunderstorm*, U. S. Government Printing Office, Washington, D.C., 1949.

87. J. H. Conover, *Cloud and Terrestrial Albedo Determinations from Tiros Satellite Pictures*, AFCRL-65-675, Sept. 1965.

88. H. Blau and R. Espinola, *Infrared Properties of High Altitude Clouds*, A. D. Little Final Report, Cambridge, Mass., Contract NONR 3556(00), 1965.

89. C. D. Kern, *Evaluation of Infrared Emission of Clouds and Ground as Measured by Weather Satellites*, AFCRL-65-840, Nov. 1965.

90. F. F. Hall, *The Effect of Cirrus Clouds on Infrared Sky Radiance*, Douglas Report DAC-61306 AD-820 503.

91. E. Bell, *Atlas of Sky and Terrain Infrared Measurements Program, Colorado Springs Region 1956, Interim Engineering Report 659-5*, Ohio State Univ., Columbus, Ohio, Contract No. AF 33(616)-3312, AD-145 464, 1957.

92. I. A. Lund, *Methods for Estimating the Probability of Clear Lines-of-Sight, or Sunshine, Through the Atmosphere*, AFCRL 66-839, December 1966.

93. E. A. Bertoni, *Clear Lines-of-Sight From Aircraft*, AFCRL 67-0435, August 1967.

94. R. C. Ramsey, "Spectral Irradiance from Stars and Planets Above the Atmosphere, from 0.1 to 100.0 Microns", Appl. Opt. 1 (1962).

95. G. Neugebauer and Robert Leighton, *Scientific American*, 51 (August 1968).

96. H. L. Johnson, F. J. Low, and D. Steinmetz, *Lunar & Planetary Laboratory Communications*, "No. 55 Infrared Observations of the Neugebauer-Martz-Leighton, 'Infrared Star' in Cygnus".

97. T. Overall, *Stellar Radiation and Distribution*, Aerojet-General TM-522-61-6-624, Azusa, Cal., 1961.

98. Barnhart and Mitchell, *Contributions from the Perkins Observatory*, Ohio Wesleyan University and Ohio State University, 1966.

99. J. M. Saari, et al., *Review of Lunar Infrared Observations*, Boeing Scientific Research Laboratory, Seattle, Washington, AD-645 548, Dec. 1966.

100. F. S. Johnson, "The Solar Constant", J. Meteorol. 11, 431 (1954).

101. F. P. Boynton and J. P. Neu, *Rocket Plume Radiance V Calculation of Adiabatic Flame Temperatures of Afterburning Rocket Exhaust*, Report ERR-AN-011, Convair (Astronautics Division), 1960.

102. R. J. Thomson, "The Chemistry of Propellants," *AGARD Symposium* (Eds. S. S. Penner and J. Decarme), Pergamon Press, 1960.

103. S. A. Golden, *Identification of Rocket Propellants by Optical Methods, Volume 1*, Report No. AFAL-TR-67-61, Air Force Avionics Lab., WPAFB, Ohio, May 1967 (Secret).

104. Various Papers in *Proc. of Specialist Conference on Molecular Radiation and its Application to Diagnostic Techniques* at Marshall Space Flight Center (Oct. 1967), NASA TM-X-53711.

105. F. S. Simmons, *Studies of Infrared Radiative Transfer in Hot Gases II: Formulation of Band Models for Nonisothermal Paths*, Willow Run Labs, Ann Arbor, Mich., Report 4613-93-T, Feb. 1965.

106. S. L. Tuttle, *Launch Phase Summary*; Vol. 1, The B.A.M.I.R.A.C. Models, Report of BAMIRAC, University of Michigan, Ann Arbor, Mich.

107. R. Zirkind, *Radiation from Rocket Exhaust Plumes*, PIBAL Report No. 984, AD-641 612, Aug. 1966.

108. W. A. Page, *A Survey of Thermal Radiation Studies of Ablating Bodies in the Ballistic Range*, NASA TN D-3741.

109. J. J. Martin, *Atmospheric Re-entry*, Prentice-Hall, Englewood Cliffs, N. J., 1966.

110. *Seventh Annual Report of BAMIRAC*, Willow Run Laboratories, Ann Arbor, Mich. (Unclassified section of SECRET report), 1966.

111. W. M. Hamilton, *Spectral Emission of Heat Shield Materials*, Proceedings AMRAC, Ann Arbor, Mich. (Unclassified paper in a SECRET symposium report), 1968.

112. G. K. Mather and A. D. Wood, *Flight Evaluation of an Infrared Spectrometer as a Clear Air Turbulence Detector*, National Research Council of Canada Report NAE-LR 477, May 1967. (FAA-RD-66-70)

113. Various Papers in *Proceedings of the Fourth Symposium on Remote Sensing of Environment*, at the University of Michigan, Ann Arbor, Mich., April 1966. AD-638 919.

114. F. C. Polcyn, *Investigation of Spectrum Matching Sensing in Agriculture VI*, Final Report Michigan University, NASA-CR-91522, November 1967.

115. D. G. Eurring and J. A. Smith, *Target Signatures Analysis Center: Data Compilation*, AFAL (AVPT) WPAFB, Ohio, July 1966. (Unclassified section NFN) (Supplement AD-819 M12).

116. A. O. Ramsby, J. T. Walwood, *Feasibility of Using Fluorescent Colorants to Afford Camouflage Protection Against Photographic Detection*, U. S. Army Natick Labs, Mass., Clothing and Organic Materials Division, RC/OM TS 145, U. S. A. National Labs TR 67-36-CM, Oct. 1966.

CHAPTER 3

IR SYSTEM COMPONENTS*

3-1 INTRODUCTION

The various major components and techniques which enable active as well as passive IR systems to emit, focus, detect, and process IR signals for any of a number of end uses are discussed here. The interrelated functions of the individual components when joined to form a variety of standard as well as specialized IR systems are then analyzed in Chapters 4 and 5.

A thorough understanding of the function of these components is essential in order to make possible the economical design of an operationally optimum IR system. The functional relationship of the major disciplines of IR technology and the associated design specialties are indicated in Fig. 3-1. The essential elements used in IR systems are identified and discussed in the following paragraphs within this chapter:

3-2 Optics
3-3 Emitters and Illuminators
3-4 Detectors
3-5 Cooling Systems
3-6 Signal Processing
3-7 Data Display and Recording
3-8 Testing IR and Associated Equipment
3-9 Ancillary IR Components and EMI Rejection Techniques

An attempt is made in this chapter to update information previously published regarding these components and techniques. Specifically, recent technological advancements in the following areas are thought to merit special attention:

a. Long wavelength IR (LWIR) multi-sensor arrays and associated electronics attained the level of development which now makes possible thermal mapping of terrestrial expanses in the 8- to 14-micron region. The use of LWIR arrays in space for object detection and in satellite communication systems represents a major achievement in space technology.

b. Cryogenic electronics have been developed for use in processing the output of high-sensivity LWIR sensors. This electronic advancement is treated here in the context of the components discussion and in AMCP 706-128 in terms of systems application.

c. Other components such as lasers, illuminators, and displays are discussed in fairly detailed form because of their novelty and the increasing interest in their application. The advent of the lasers and, especially, the recent development of high-power IR lasers has opened new frontiers in optical range finding and detection. The relative novelty of lasers, compared to other IR components, is responsible for the lack of detailed published data on the subject; consequently, a comprehensive discussion is included here as the preface for the design criteria discussion on laser systems. The remaining material included in the lasers chapter appears in current IR publications; however, an attempt is made here to update that material by including any new significant developments.

d. The requirement for real-time observation of high data-rate IR information in performing surveillance and tactical tasks has prompted detailed discussions of display systems, man-machine interface considerations, and human engineering techniques. These are discussed partly in this chapter and partly in Chapter 5.

*Written by K. Seyrafi.

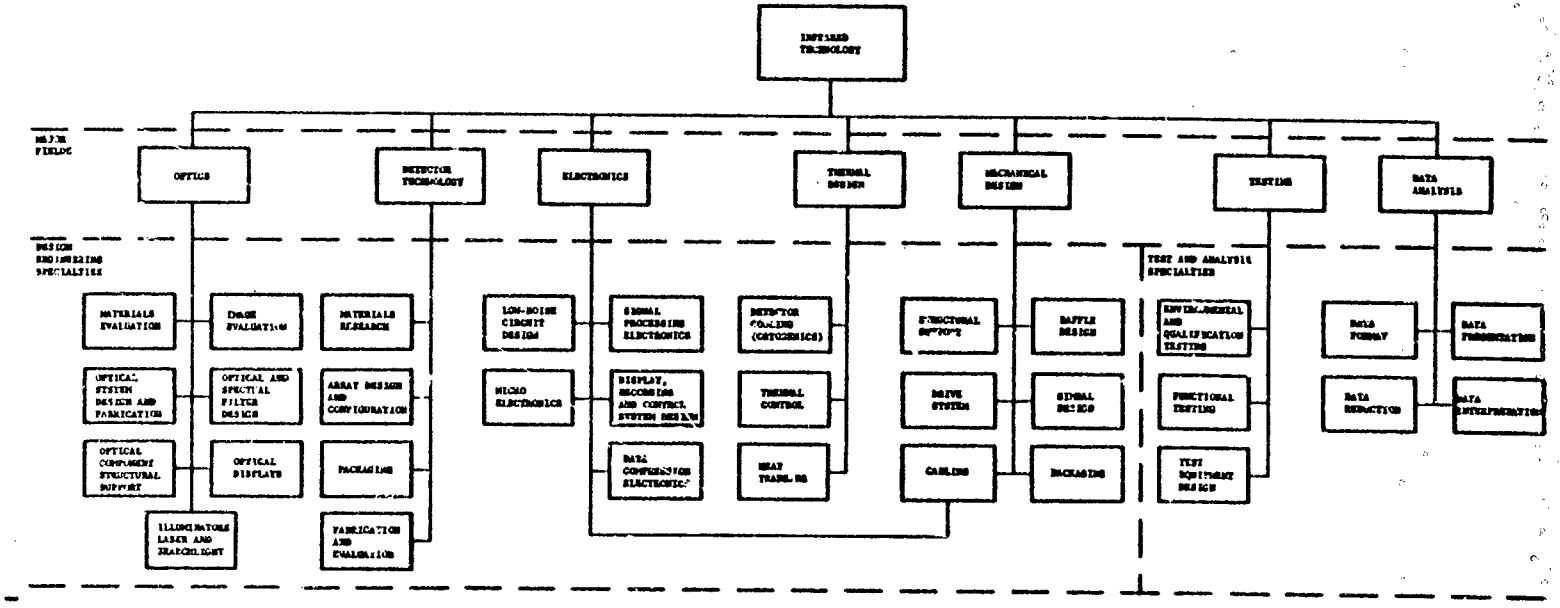


FIGURE 3-1. Functional Relationship of the Major Disciplines and Associated Design Specialties

3-2 OPTICS

3-2.1 OPTICAL MATERIALS

3-2.1.1 Material Types

3-2.1.1.1 Refractive Materials

Infrared transmitting materials include various types of glass, crystals, and plastics. In the glass category are the oxide (regular optical, high silica and fused quartz, and special oxide glasses) and nonoxide materials. Although physically suited for IR applications, the transmission range of oxide glasses is limited to wavelengths of 6 microns or less. Oxide glasses usually exhibit absorption in the 2.7- to 3.3-micron region due to the water content in the material; however, some "water free" types have been developed. Nonoxide glasses which transmit to longer wavelengths (some to 20 microns) tend to be softer and have lower softening points. Significant improvements are being made in the nonoxide chalcogenide (sulfur, selenium, and tellurium) glasses[1]. Results of investigations of many combinations and variations of glasses are reported in some of the open literature listed in the bibliography.

Crystalline materials can be either single crystalline or polycrystalline. The single crystal materials have been used primarily at the longer wavelengths, but available size, cost, and unsuitable physical properties have been limiting factors in many engineering applications. The number and size of polycrystalline materials have increased significantly during recent years. Among these are chiefly IRTRAN* materials made by Eastman Kodak and KRS-5 materials.

Plastic materials, in general, are considered unsuitable for use as optical components in IR instruments—mainly because of poor mechanical properties, low scratch resistance, temperature instability, instability due to water absorption and difficult surface finishing characteristics. The following exceptions should be noted:

1. Metal-backed epoxy mirrors can be successfully fabricated by a molding (replication) technique. Aspheric surfaces can be readily produced, thereby, resulting in components which are generally adequate for condenser systems.

2. Polystyrene has a large number of stable narrow absorption bands in the IR region. This characteristic makes it an excellent standard in the form of thin sheet for wavelength calibration of IR spectrometers and related instruments.

The index of refraction for all commercially available plastics ranges between the limits 1.49 for Lucite or Plexiglas to 1.59 for polystyrene.

3-2.1.1.2 Reflective Materials

The limited selection of suitable IR transmitting materials along with limiting size considerations have promoted the use of reflective elements in IR systems. Substrate materials used include fused quartz, Pyrex, low-expansion fused silica, glass-ceramics and metals (especially beryllium and aluminum). Vacuum-deposited aluminum is most frequently used as a reflective coating, although silver, gold, copper, and rhodium are also effective. Protective coatings of silicon monoxide or magnesium fluoride, which may also increase the reflectivity of the primary coating, are usually applied.

Beryllium, stainless steel, aluminum, and fused quartz are used in fabricating mirror substrates for cold optical devices which must be cooled to cryogenic temperatures to avoid becoming performance-limited due to background photon fluctuations. From a thermal standpoint, the mirror substrate material must have high thermal conductivity and thermal diffusivity characteristics and, if possible, low thermal capacity. These combined characteristics determine the ease with which the optical element can be cooled to the required operating temperature and maintained in a thermal gradient-free condition.

*ITRAN is a proprietary name of Eastman Kodak Co. for optical materials which transmit in the 3-5 and 2-14μ bands.

TABLE 3-1. SALIENT CHARACTERISTICS OF OPTICAL IR MATERIALS

	MATERIAL	WAVELENGTH (MICRONS) 0.5 1.0 2 5 10 20 30 50	% REFL LOSS	REFRA INDEX n	WATER SOLUB g/100g	HDNS KNOOP	EXP COEF 10^{-6}/°C	SP GR
1	POTASSIUM DIHYDROGEN PHOSPHATE (KDP)	1.65 mm	7	1.49-1.51	33			2.34
2	AMMONIUM DIHYDROGEN PHOSPHATE (ADP)	1.55 mm	8	1.47-1.50	36.8			1.80
3	GALLIUM PHOSPHIDE (GaP)	1 mm	40	3.0 at 2μ			5.3	
4	OPTICAL GLASS		7-20	1.45-1.98	INSOL	300-600	4-10	2.5-36
5	SODIUM NITRATE ($NaNO_3$)	Details not available	4	1.334 at 0.67μ	88	19.2	12	2.26
6	GALLIUM ANTIMONIDE (GaSb)	0.7 mm	51	3.789 at 2.0μ	INSOL		6.9	
7	QUARTZ, FUSED (SiO_2)	10 mm "water free"	6	1.438 at 2.0μ	INSOL	1600	0.55	2.20
8	Kel-F	Not available 9.5 mm						
9	CALCIUM CARBONATE ($CaCO_3$)	33 mm	11	1.624 at 2.0μ	.0014	3 Moh	25	2.71
10	CALCIUM ALUMINATE GLASS	2 mm	11	1.63 at 2μ		600	8.3	3.07

TABLE 3-1. SALIENT CHARACTERISTICS OF OPTICAL IR MATERIALS (Continued)

	MATERIAL	WAVELENGTH (MICRONS) 0.5 1.0 2 5 10 20 30 50	% REFL LOSS	REFRA INDEX n	WATER SOLUB g/100g	HDNS KNOOP	EXP COEF 10^{-6}/°C	SP GR
11	SPECIAL OXIDE GLASSES	1 mm, 10 mm	15	1.8 at 2μ	INSOL	≥400	8-10	3.0-6.0
12	SPINEL ($MgO \cdot 2.5Al_2O_3$)	5.4 mm	13	1.724 at 0.66μ	INSOL	1140	5.9	3.61
13	TITANIUM DIOXIDE (TiO_2)(Rutile)	6 mm	29	2.399 at 2.0μ	INSOL	890	7-9	4.26
14	SAPPHIRE (Al_2O_3)	2.5 mm	13	1.73 at 2.2μ	INSOL	~1700	5.8-7.7	3.98
15	STRONTIUM TITANATE ($SrTiO_3$)	1 mm	25	2.21 at 3.4μ		595	9.4	5.12
16	INDIUM ARSENIDE (InAs)		46	3.4 at 6μ	INSOL		5.3	
17	LEAD SULFIDE (PbS)(film)		54	4.10 at 3.0μ	8.6×10^{-5}			7.5
18	LEAD SELENIDE (PbSe)(film)	Not Av.	58	4.5 at 2.2μ	INSOL		18.4	8.1
19	LEAD TELLURIDE (PbTe)(film)	11 mm, N. Av.	64	5.35 at 3μ				8.16
20	LANTHANUM FLUORIDE (LaF_3)	Details not available	16	1.82 at 0.58μ	INSOL	MED HARD		5.94

TABLE 3-1. SALIENT CHARACTERISTICS OF OPTICAL IR MATERIALS (Continued)

	MATERIAL	WAVELENGTH (MICRONS) 0.5 1.0 2 5 10 20 30 50	% REFL LOSS	REFRA INDEX n	WATER SOLUB g/100g	HDNS KNOOP	EXP COEF 10^{-6}/°C	SP GR
21	IRTRAN 5 (Polycrys. MgO)	1 mm	7	1.482 at 8μ	INSOL	640	11.5	3.58
22	TELLURIUM (Te)	0.85 mm; N Av.	60,69	4.83, 6.29 at 6.0μ	INSOL	2.3 Moh	16.75	6.24
23	MAGNESIUM FLUORIDE (MgF_2)	2 mm; 30 mm	5	1.38 at 0.70μ	.0076	415	16.6	3.17
24	BARIUM TITANATE ($BaTiO_3$)	Not Av.; 0.094 mm	29	2.40			400	5.9
25	MAGNESIUM OXIDE (MgO)	N. Av.; 0.7 mm; 5.4 mm	12	1.68 at 3.3μ	INSOL	690	13.8	3.57
26	LITHIUM FLUORIDE (LiF)	1 mm; 12 mm	4	1.35 at 4.0μ	0.3	110	3.7	2.64
27	CADMIUM FLUORIDE (CdF_2)	Details not avrilable for this region.; 5 mm	10	1.576 at 0.59μ	4.4	3 Moh	22.6	6.34
28	MANGANESE FLUORIDE (MnF_2)	Details not available for this region.; 10 mm	5	1.4 at 6μ	.0076	MED HARD		3.98
29	T-12	5 mm	6	1.41 at 3.3μ	0.20	4.5 Moh	20.6	4.35
30	CALCIUM FLUORIDE (CaF_2)	1 mm; 10 mm	5	1.40 at 5.0μ	.0013	160	22	3.18

TABLE 3-1. SALIENT CHARACTERISTICS OF OPTICAL IR MATERIALS (Continued)

	MATERIAL	WAVELENGTH (MICRONS) 0.5 1.0 2 5 10 20 30 50	% REFL LOSS	REFRA INDEX n	WATER SOLUB g/100g	HDNS KNOOP	EXP COEF 10^{-6}/°C	SP GR
31	STRONTIUM FLUORIDE (SrF_2)	N. Av. 3 mm 27 mm	6	1.439 at 0.58μ	.0117	4 Mok		4.28
32	CESIUM FLUORIDE (CsF)	Details not available	7	1.478 at 0.59μ	.367 at 18°			3.59
33	IRTRAN 3 (Polycrys. CaF_2)	1 mm 6.4 mm	3	1.300 at 10μ	.0017	200	24	3.18
34	IRTRAN 2 (Polycrys. ZnS)	2 mm	24	2.151 at 13μ	INSOL	354	6.8	4.09
35	INDIUM PHOSPHIDE (InP)		40	3.0 at 10μ				
36	ZINC SULFIDE (ZnS)(film)	Details not available						4.1
37	INDIUM ANTIMONIDE (InSb)	0.15 mm N. Av.	52	3.95 at 10μ			5.5	
38	CADMIUM TELLURIDE (CdTe)	Details not available	32	2.56 at 10μ			4.5	6.2
39	POTASSIUM FLUORIDE (KF)	Details not available	5	1.361 at 0.57μ	92.3 at 18°			2.48
40	IRTRAN 1 (Polycrys. MgF_2)	1 mm 2 mm	4	1.34 at 4.87μ	.0076	576	10-12	3.18

TABLE 3-1. SALIENT CHARACTERISTICS OF OPTICAL IR MATERIALS (Continued)

	MATERIAL	WAVELENGTH (MICRONS) 0.5 1.0 2 5 10 20 30 50	% REFL LOSS	REFRA INDEX n	WATER SOLUB g/100g	HDNS KNOOP	EXP COEF 10^{-6}/°C	SP GR
41	SILICON (Si)	10 mm, 25°C, 450°C, Not Av.	46	3.42 at 5.0μ	INSOL	1150	4.15	2.33
42	BARIUM FLUORIDE (BaF_2)	1 mm, 10 mm	7	1.45 at 5.2μ	0.12	82	15-18	4.83
43	GALLIUM ARSENIDE (GaAs)	0.81 mm, 1 mm	42	3.135 at 10μ	INSOL		5.7	
44	POLYETHYLENE	Details not available; 0.1 mm; Details not available						
45	POLYMETHYL METHACRYLATE	Details not available; 0.02 mm; Details not available	7	1.49 at .6μ			6.3	1.19
46	LEAD FLUORIDE (PbF_2)	Details not available; 2 mm, 10 mm	11	1.65 at 4.0μ	0.064	200		8.24
47	CADMIUM SULFIDE (CdS)		27	2.30 at 1.2μ	INSOL	122	4.2	4.82
48	ARSENIC TRISULFIDE GLASS (As_2S_3)	1 mm, 12 mm	29	2.41 at 5.0μ	INSOL	125	24.6	3.2
49	SODIUM FLUORIDE (NaF)	1 mm, 10 mm	3	1.27 at 8.0μ	4.2	60	36	2.79
50	CuCl	9 mm		1.93	6.2×10^{-3}			3.53

TABLE 3-1. SALIENT CHARACTERISTICS OF OPTICAL IR MATERIALS (Continued)

	MATERIAL	WAVELENGTH (MICRONS) 0.5 1.0 2 5 10 20 30 50	% REFL LOSS	REFRA INDEX n	WATER SOLUB g/100g	HDNS KNOOP	EXP COEF 10^{-5}/°C	SP GR
51	IRTRAN 4 (Polycrys. ZnSe)	2 mm	27	2.310 at 20μ	INSOL	150	7.5	5.27
52	SILICON MONOXIDE (film)(SiO)	125 mμ, 275 mμ, N. Av.			INSOL			
53	Cd Se	1.7 mm			INSOL			5.81
54	ARSENIC SELENIUM GLASS (Se(As))	N. A.	31	2.48 at 5.0μ	INSOL		34	
55	SODIUM BROMIDE (NaBr)	Details not available	11	1.638 at 0.61μ	91			3.20
56	SODIUM IODIDE (NaI)	Details not available	14	1.765 at 0.65μ	180			3.67
57	GERMANIUM (Ge)	not av. 2 mm 300°C 25°C	>2	4.001 at 16μ	INSOL	6.25 Moh	5.5-6.5	5.327
58	SODIUM CHLORIDE (NaCl)	1 mm, 10 mm	8	1.495 at 10.0μ	36	15.2	44	3.16
59	SILVER CHLORIDE (AgCl)	10 mm	19	1.95 at 15.0μ	INSOL	9.5	30	5.59
60	SELENIUM (AMORPHOUS) (Se)	1.7 mm	30	2.45 at 2.5μ			36.8	4.26

TABLE 3-1. SALIENT CHARACTERISTICS OF OPTICAL IR MATERIALS (Continued)

	MATERIAL	WAVELENGTH (MICRONS) 0.5 1.0 2 5 10 20 30 50	% REFL LOSS	REFRA INDEX n	WATER SOLUB g/100g	HDNS KNOOP	EXP COEF $10^{-6}/^{o}C$	SP GR
61	POTASSIUM CHLORIDE (KCl)	10 mm	5	1.363 at 23μ	34.7	7.2 9.3	36	1.99
62	IRTRAN 6 [b] (Polycrys. CdTe)	3 mm	34	2.673 at 10μ	INSOL	45	5.7	5.85
63	CESIUM CHLORIDE (CsCl)	Details not available	11	1.644 at 0.54μ	186			3.97
64	THALLIUM CHLORIDE (TlCl)	1.6 mm	24	2.193 at 10μ	0.32	12.8	53	7.02
65	KRS-6 (TlBr-TlCl)	6 mm	24	2.18 at 9.0μ	0.32	30	50	7.19
66	SILVER BROMIDE (AgBr)	5 mm 27 mm	25	2.232 at 0.67μ	INSOL		34.0	6.47
67	POTASSIUM BROMIDE (KBr)	4 mm 27 mm	8	1.524 at 11.0μ	65.2	27	43	2.75
68	KRS-5 (Tl(Br,I))	5 mm 27 mm	28	2.371 at 10μ	0.05	40.2	58	7.37
69	THALLIUM BROMIDE (TlBr)	1.6 mm 6 mm	28	2.338 at 9.98μ	0.05	11.9	51	7.45
70	POTASSIUM IODIDE (KI)	Details not available 0.83 mm	11	1.62 at 10μ	144	5	42.6	3.13

TABLE 3-1. SALIENT CHARACTERISTICS OF OPTICAL IR MATERIALS (Continued)

	MATERIAL	WAVELENGTH (MICRONS) 0.5 1.0 2 5 10 30 20 50	% REFL LOSS	REFRA INDEX n	WATER SOLUB g/100g	HDNS KNOOP	EXP COEF $10^{-6}/^{o}C$	SP GR
71	CESIUM BROMIDE (CsBr)	10 mm	11	1.64 at 20_{μ}	124.3	19.5	47.9	4.44
72	CESIUM IODIDE (CsI)	5 mm	13	1.727 at 20_{μ}	85	SOFT	50	4.53
73	DIAMOND (C)	Details not available / Not available	29	2.417 in vis.	INSOL	10 Moh	1.38	3.51

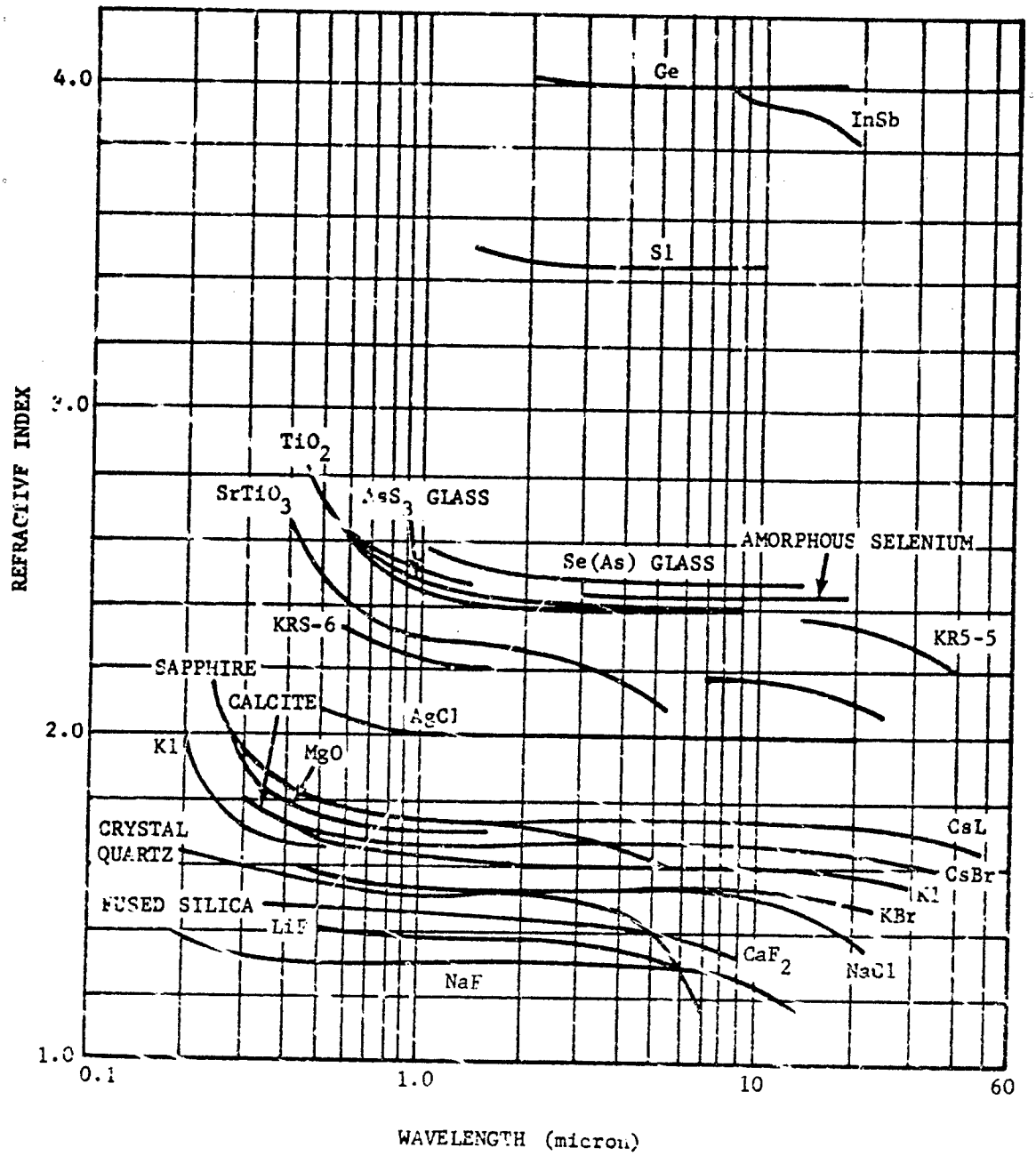


FIGURE 3-2. Refractive Index vs Wavelength for Several Optical Materials

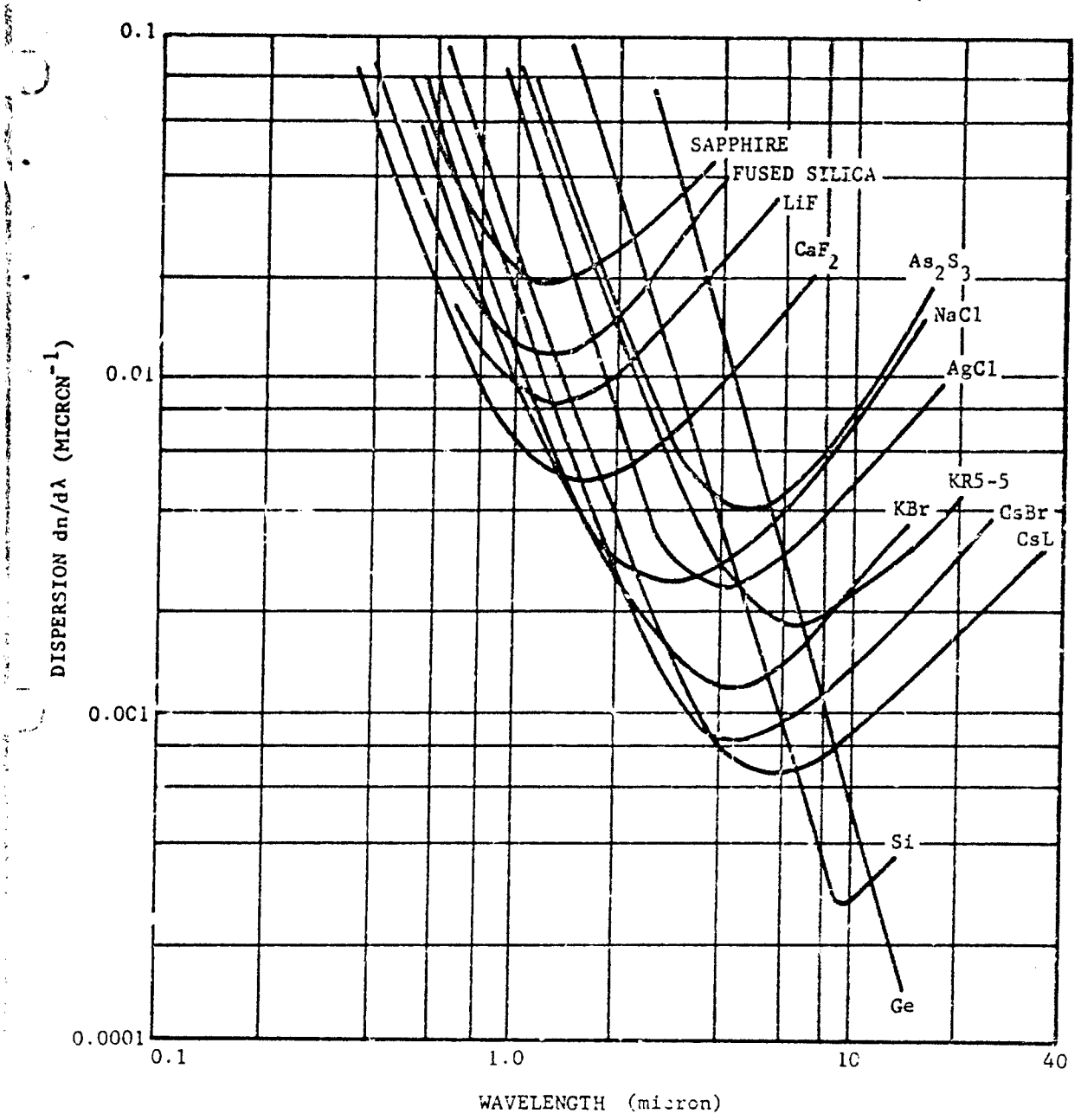
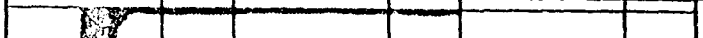
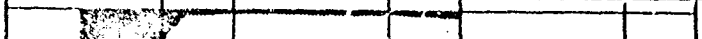


FIGURE 3-3. Dispersion vs Wavelength for Several Optical Materials

TABLE 3-2. MIRROR MATERIALS

MATERIAL	SPECIFIC GRAVITY	EXPANSION COEFFICIENT, 10^{-6}/°C	MELTING POINT, °C	YOUNG'S MODULUS, 10^6 psi	THERMAL CONDUCTIVITY, cal cm^{-1} sec^{-1} $°C^{-1}$	SPECIFIC HEAT	HARDNESS, KNOOP
Fused Quartz	2.20	0.55	1600	10.5	0.0033	0.188	460
Pyrex (7740)	2.23	3.2	820	9.5	0.0027	0.233	480
Vycor (7900 series)	2.18	0.80	1500	6.7	0.0035	0.19	530
Glass-ceramics							
Pyroceram (9606)	2.61	5.7	1350	17.3	0.0087	0.230	698
Pyroceram (9608)	2.50	0.4 - 2.0	1250	12.5	0.0047	0.235	703
CER-VIT C-101	2.50	0.15		13.4	0.0040	0.217	540
Aluminum	2.70	23.9	680	6.9	0.53	0.215	2.0 Moh
Beryllium	1.82	12.4	1300	28.0	0.38	0.516	
Invar (36% Ni)	8.0	1.30		14.8	0.026	0.095	
Magnesium	1.74	26		4.5	0.38	0.25	2.0 Moh
Titanium	4.54	8.5		11.6	0.042	0.126	

TABLE 3-3. REFLECTIVITY CHARACTERISTICS OF COMMON MIRROR COATINGS

MATERIAL	REFLECTIVITY (MICRONS) 0.2 0.3 0.5 1.0 2 3 4 5 8 10 20 30 50 80	MAX %
ALUMINUM (Al)		96
SILVER (Ag)		96
GOLD (Au)		96
COPPER (Cu)		96
RHODIUM (Rh)		96

The fact that the structure which supports the optical element is usually made of the same, or very similar, material as the mirror substrates, makes certain mechanical properties mandatory. These properties must exist at the extremely low temperature at which the device is designed to operate. The material must be rugged (insensitive to notches or stress concentrations) to avoid brittle fracture when subjected to nigh loads of high strain rates. A high strength-to-weight ratio is required of cold optical devices used in the environment of a sounding rocket, balloon, or satellite.

3-2.1.2 Material Properties

Table 3-1 lists the significant properties of the more common IR materials and illustrates graphically the transmission ranges of the materials. The transmittance curves are arranged in increasing cutoff-wavelength order. The approximate cutoff point is indicated where detailed curve data are not available. Dotted lines indicate variations in transmittance for different samples. A transmission curve only is presented for denoting the general classification of various types of glasses. Many variations of each type glass are available and no attempt has been made to give particular code names or provide curves for each variation. Fused quartz—variously called fused silica, quartz glass, or vitreous silica—can be obtained either with or without the 2.7-micron vapor absorption characteristics from several manufacturers.

The "% REFL LOSS" column lists the loss for the two surfaces at a median wavelength. This is an indication of the amount of improvement in transmittance which might be expected if an anti-reflection coating were applied to both surfaces.

The dimensions listed in Table 3-1 denote the thicknesses of the material for which the transmittance values are given. The transmittance values listed cannot be effectively extrapolated to other thicknesses since the absorption coefficients are not given (not available in most instances). Where such data are required, the manufacturer is usually the best source.

The information in the Table 3-1 should be used only as a general guide in comparing one material with another and not as an ultimate source of detailed and precise data. More complete information is contained in Refs. 2 through 17.

Curves representing the refractive index and dispersion versus wavelength characteristics of a number of the more useful IR materials are given in Figs. 3-2 and 3.3.

Data pertaining to some of the mirror substrate and support materials are listed in Table 3-2. Pyrex, Vycor, and Pyroceram are trademarks of the Corning Glass Works; CER-VIT is a registered trademark of Owens-Illinois, Inc. Table 3-3 illustrates the reflectivity characteristics of the most frequently used mirror coatings.

3-2.1.3 Anti-reflection Coatings

Transparent materials have a general property of reflecting a portion of the radiation beam as it passes through a boundary between the media of different indexes of refraction. The greater the index difference, the greater will be the reflected portion. The reflection will be 4% at a single air-glass (index 1.5) interface. Two such surfaces (as on a lens) will reflect approximately 8%. A lens made of germanium (index 4.01), which is a widely used IR material, may reflect as much as 55%, depending on the wavelength of the radiation involved.

In optical instruments of all types, surface reflection represents at least an annoyance requiring careful design to avoid serious degradation in performance. In IR systems the high percentage of reflection loss would ordinarily preclude the use of many materials having other highly desirable properties. Thus, all modern high-performance systems make use of elements which are coated with films that act to reduce the surface reflections, while producing a corresponding increase in transmission.

The principle upon which this action takes place is one of destructive interference. A film has two interfaces, one between itself and air and the other between itself and the optical element, each of which produces a reflection. If the film thickness is equal to 1/4 wavelength of the irradiance, the reflection from the second interface will arrive at the first interface 1/2 wavelength behind the incident radiation, thereby being 1/2 wavelength out of phase with that reflected at the first interface. If the amplitudes of the two reflected waves are equal they will cancel each other by destructive interference.

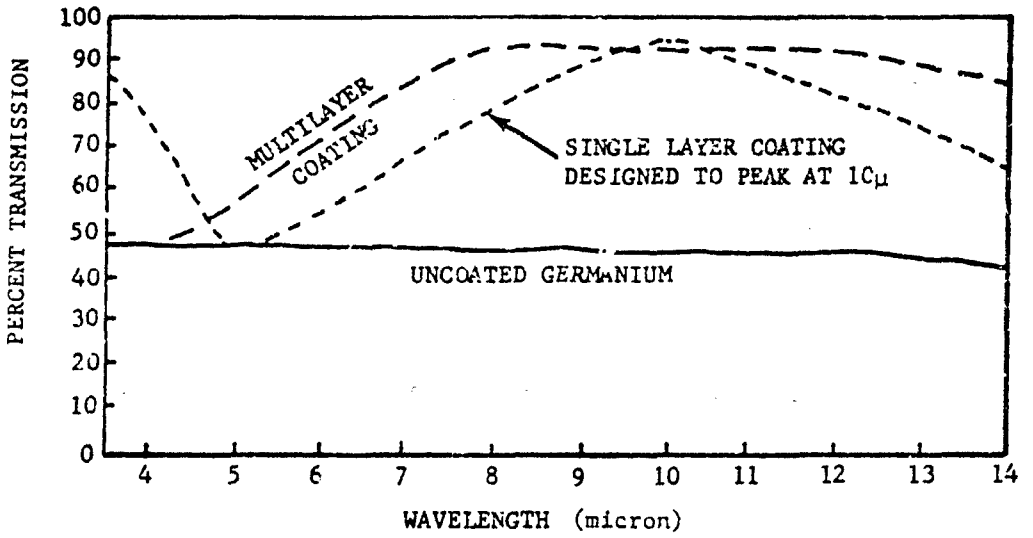


FIGURE 3-4. Effects of Anti-reflection Coatings on Ge Transmission

However, since the energy cannot be destroyed, it will appear in the transmitted beam as an increase in transmission.

In order to make the amplitudes of the two reflected waves equal, the film must have a refractive index which is the geometric mean between that of the air and the material of the optical element[18]. For air and glass (index 1.5) this is $\sqrt{1 \times 1.5} = 1.225$. For air and germanium it is $\sqrt{1 \times 4} = 2$. The 1/4-wavelength thickness can be correct for only one wavelength and for radiation which is incident on the surface at the normal. At inclined incidence the path length through the film and back to the interference point is changed because of a cos ϕ factor in the path difference equation (ϕ being the angle of the ray relative to the normal within the film). However, since the cosine of small angles does not change rapidly, the effectiveness of the coating remains high over a considerable angular range.

Some improvement in the wavelength range (bandwidth) over which a coating is effective can be made by use of multiple film layer of alternately high and low index material. The several layers are designed for peak action each at a different wavelength so that the overall effect is that of increasing the bandwidth.

In addition to the proper index, a coating material should be resistant to chemicals and abrasions and have good adhesion properties. Unfortunately, the few materials available with an index below 1.4 do not have the most desirable physical properties.

One material which has been found universally suitable as coating is magnesium fluoride. Although its index does not quite match that of the lower index component materials (1.5 to 1.75), its initial reflection losses are not very severe anyway, so that less than complete suppression of reflection is still acceptable.

For germanium and other high-index materials, the coating materials and processes have not yet been standardized. In addition, many are proprietary with the firms which specialize in film coating.

As an example of the effectiveness of anti-reflection coating in improving the transmission of IR materials, a typical set of transmission curves for germanium is presented in Fig. 3-4.

When specular reflection must be maintained in cold optical devices, the substrate is often coated with a few thousandths of an inch of Kanigen®, an electroless nickel coating developed by the General American Transportation

Corporation. This material is amorphous and polishes by material flow similarly to glass, thus obliterating the crystalline structure of the substrate below.

Unfortunately, the thermal coefficient of expansion of nickel is greatly different from most substrate materials, The bi-metallic strain produced is readily apparent optically for even modest changes in temperature when only one side of the base material is coated. Normally both sides of a mirror substrate are coated with an identical thickness of Kanigen in order to balance the bi-metallic stress.

3-2.2 OPTICAL COMPONENTS

3-2.2.1 Types of Lenses and Principal Characteristics

Lenses serve as the refracting components in optical systems. They change the convergence or divergence of the optical ray bundles as required. Convergence is the characteristic of a bundle whereby its rays are directed toward a point. Divergence means that the rays appear to originate from a point. The rate of convergence or divergence is measured by the angular size (full-cone angle) of the bundle. Positive power lenses, or simply positive lenses, will increase the rate of convergence or change a divergent bundle into a convergent bundle. Negative lenses increase the rate of divergence or render a convergent bundle divergent.

The power of a lens, or the degree to which it will change the rate of convergence or divergence, is a function of the surface curvatures and refractive index; the greater curvatures and refractive indexes produce the higher powers.

A variety of lens types along with their chief characteristics and uses are illustrated in Fig. 3-5.

3-2.2.2 Types of Mirrors

Mirrors perform the same function as do lenses (although in a different manner) in changing the rate of convergence of ray bundles. The most useful feature offered by a mirror is its freedom from chromatic aberration. Some typical mirror shapes and corresponding characteristics are illustrated in Fig. 3-6.

3-2.2.3 Filters

The purpose of a filter is to attenuate the energy of an optical beam either uniformly or in certain selected wavelength regions. A filter usually consists of thin layers of solid material deposited on a substrate or on the surface of an optical component.

Filters which attenuate uniformly over a given spectral region are called "Neutral Density Filters". These are usually produced by metallic deposits of a controlled thickness. This permits part of the beam energy to be transmitted and part reflected, while a small part is absorbed. Neutral density filters are not the only means of attenuating an electromagnetic beam. In infrared technology it is common practice to use a wire screen or a lattice of holes in an opaque plate with a set percentage of open space. However, this is only useful when diffraction effects can be ignored, which is usually not the case.

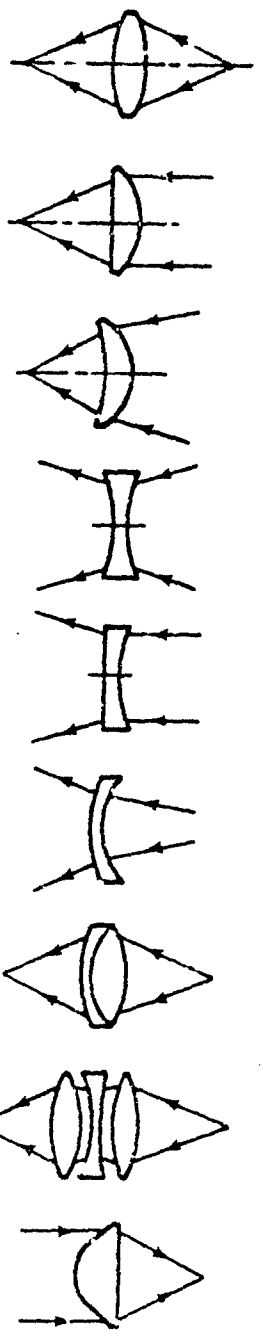
A neutral density filter, set at an angle to the beam and with both the reflected and transmitted portions being functionally useful in the optical system, acts as, and is called, a "Beam Splitter".

Spectral filtering depends on the interference effects of thin transparent coating layers which permit transmission of certain selected wavelengths while attenuating others.

Materials and processes have been developed which make possible the production of a wide variety of spectrally selective filters. These are often classified according to their characteristics as indicated in Fig. 3-7. Spike filters, Fig. 3-7(C), can be made as narrow as 2% of the wavelength at which transmission occurs for wavelengths shorter than 20 microns.

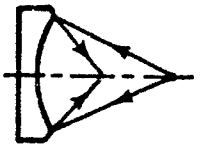
Interference filter design and fabrication is a highly specialized science and technology, much of which is proprietary with the manufacturer. An introductory treatment of the subject is given in Ref. 19, and some advanced techniques are described in Refs. 20 through 23.

Spectral as well as neutral density filters are also produced in the absorbing type. Either the substrate itself or a coating may perform the absorbing function. In this type, the spectral characteristics are dependent entirely on the selection of materials for the substrate and coating. Kodak's Wrattan filters are examples of this filter.

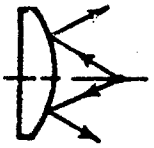


CONVEX	CONVERGING:	GENERAL USE, MAGNIFICATION
PLANO CONVEX	CONVERGING:	USED OFTEN IN OPPOSED DOUBLES TO REDUCE SPHERICAL ABERRATION
MENISCUS	CONVERGING:	REDUCED SPHERICAL ABERRATION
CONCAVE	DIVERGING:	GENERAL USE, DEMAGNIFICATION
PLANO CONCAVE	DIVERGING:	USED IN MULTI-ELEMENT COMBINATIONS
MENISCUS	DIVERGING:	REDUCED SPHERICAL ABERRATION
DOUBLET	CORRECTED FOR CHROMATIC ABERRATION	
MULTI-ELEMENT	HIGH ORDER OF ABERRATIONS CORRECTION: USED IN COMPLEX SYSTEMS	
ASPHERIC	CORRECTED FOR SPHERICAL ABERRATIONS: USED IN CONDENSER SYSTEMS	

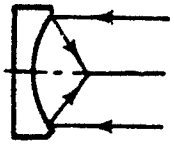
FIGURE 3-5. Basic Lens Forms



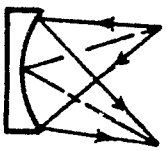
CONCAVE SPHERICAL — CONVERGING: GENERAL USE



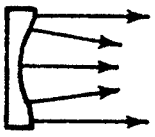
CONVEX SPHERICAL — DIVERGING: GENERAL USE



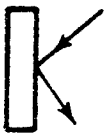
PARABOLOIDAL — ACCURATELY FOCUSES A PARALLEL BEAM OR PRODUCES A PARALLEL BEAM FROM A POINT SOURCE



CONIC: ELLIPSOIDAL — REFOCUSES A DIVERGING BUNDLE AT ANOTHER POINT (P) DISPLACED FROM THE POINT OF ORIGIN (O)



GENERAL ASPHERIC — USED MOSTLY IN COMBINATION SYSTEMS OF TWO OR MORE COMPONENTS



PLANE — CHANGE DIRECTION OF BEAM

FIGURE 3-6. Basic Mirror Forms

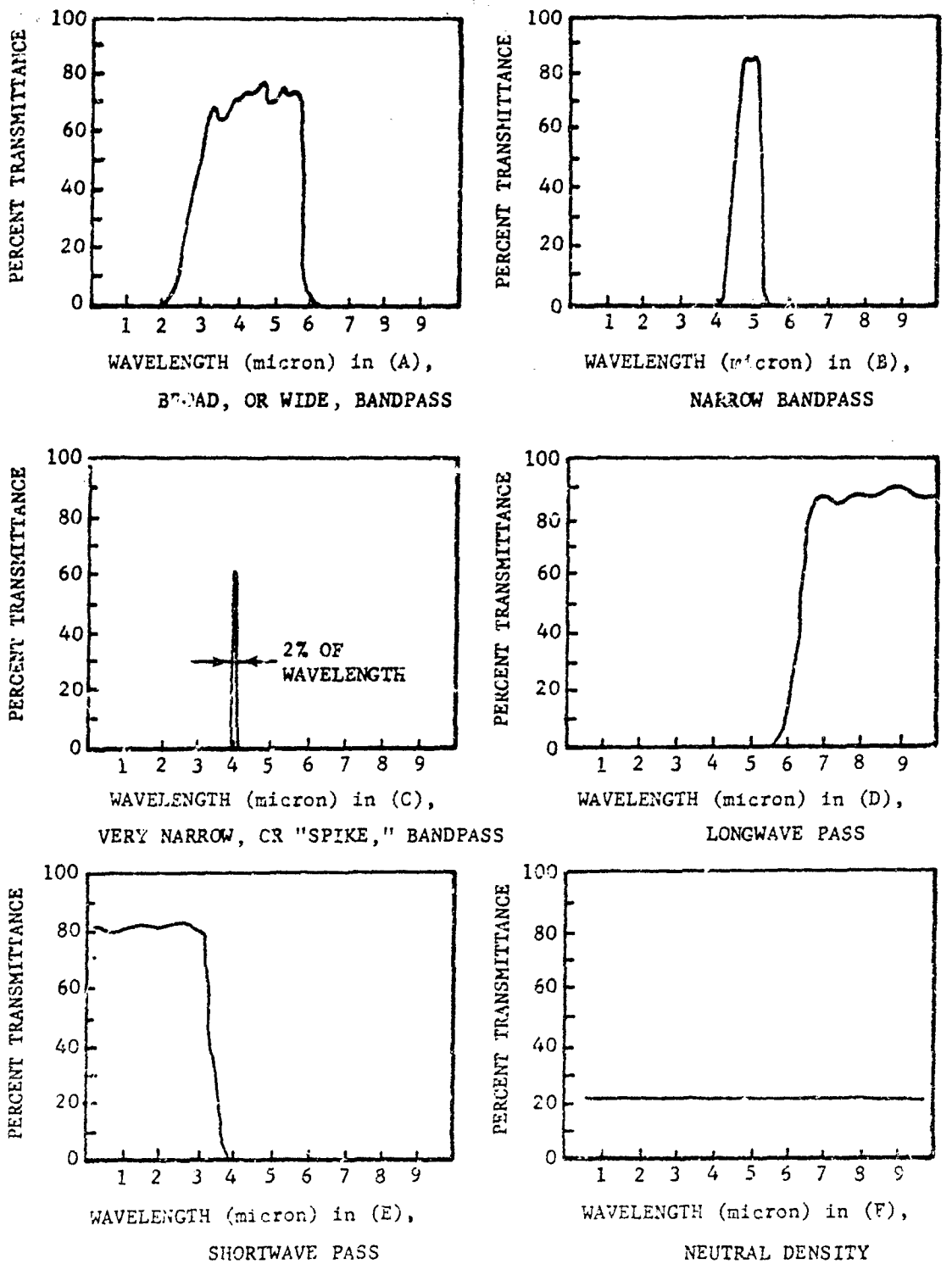


FIGURE 3-7. Typical Spectrophotometric Filter Curves

3-2.2.4 Prisms

Prisms are flat surfaced components (Fig. 3-8) used in optical systems to perform the following variety of important optical functions:

1. Dispersion. Since the refraction of electromagnetic energy rays at an interface between materials of differing densities varies with the wavelength of the radiation, the wavelength components of a composite beam passing through such an interface will be spatially separated. This action is called dispersion (Fig. 3-8 (A)). It is useful in analyzing radiances from various sources and, ultimately, in identifying and describing the sources.

2. Beam Deflection. The face formed on the hypotenuse of the 45° triangular prism (Fig. 3-8 (B)) has the property of "total internal reflection", when the material index of refraction is greater than 1/sin 45° or 1.4142. Thus, this face acts as a mirror on a beam entering through one of the other faces. The beam directed out through the third face is considered to be deflected through 90°. The advantage of this device over a plane mirror is that there is no energy loss at the reflecting surface. The entering and exiting faces can, furthermore, be coated to reduce surface reflection loss such that the net efficiency of the device can be brought close to 100%.

3. Reversion. As with a mirror, the elements of a beam in the prism of Fig. 3-8 (B) are transferred from left to right and vice versa. This action is called reversion. The image transmitted by the beam will be reverted.

4. Beam Splitting. Two prisms with their hypotenuse surfaces in contact can be made into an efficient beam splitter, Fig. 3-8 (C). In this device the total internal reflection property is suppressed, and a thin metallic film which reflects partially and transmits partially is interposed between the contacting surfaces. The advantage gained over a conventional flatplate beam splitter is the elimination of the ghost image usually formed at the second surface of such a plate.

5. Double Reversion or Inversion. An Amici or roof prism, Fig. 3-8 (D), has the property of reverting a beam and the image it carries in two directions, which is equivalent to an inversion.

6. Erection or Inversion. The Porro prism combination, Fig. 3-8 (E), has the property of inverting an image carried by a beam of radiation. An image originally formed in an inverted position by the conventional action of a telescope or binocular objective may be restored to its erect position by this Porro combination.

7. Path Length Changing. The same Porro prism, in folding the optical path back and forth several times, will reduce the total length of the instrument of which it is a part thereby making it more compact. The energy loss at the internally reflecting surfaces is nearly zero.

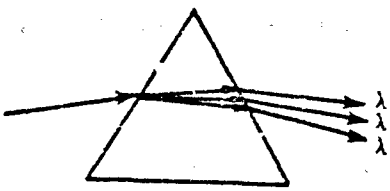
8. Constant Angle. The Penta prism, Fig. 3-8 (F), is uniquely capable of turning a beam through a fixed angle (in this case 90°) without it being precisely positioned relative to the beam. Its primary use is in rangefinders, where it is required to compare the parallelism of two spatially separated beams. A similar requirement often occurs in many testing and alignment devices.

9. Retro-reflection. An unusual type of constant deviation prism is the cube corner, Fig. 3-8 (G). This device will return a beam back on itself regardless of its orientation (within the limited range of about 15°) relative to the beam. In communication systems between an earth station and missiles or between earth and satellites, such retro-reflectors can return a maximum signal intensity from the unit (satellite or missile) being probed or interrogated.

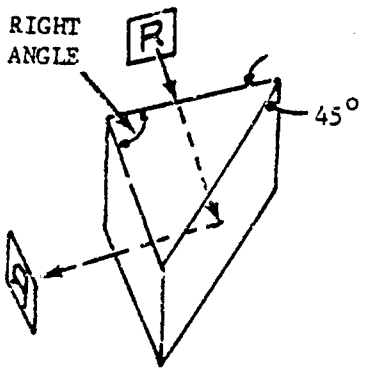
Display screens made of a multiple array of cube corner reflectors can produce a high-brightness display over a limited region.

10. Deviation. A thin prism, Fig. 3-8 (H), called a wedge, is used in devices where it is necessary that a beam or an image be laterally deviated by a small amount. Two wedges in the same beam closely spaced and arranged to rotate counter to each other, Fig. 3-8 (I), form the basic unit of an important type of infrared scanning mechanism.

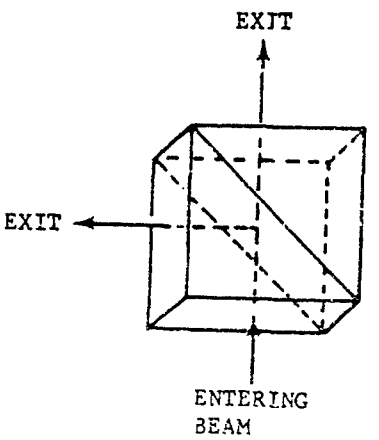
11. Parallel Displacement. If a beam is directed at an angle through a plate having parallel faces, Fig. 3-8 (J), it will be displaced but its direction will not be changed. This property is used in the design of a rotating multi-facet prism used in high-speed motion picture cameras to produce the "stop action" without mechanical interruption of the film motion.



DISPERSION
PRISM
(A)



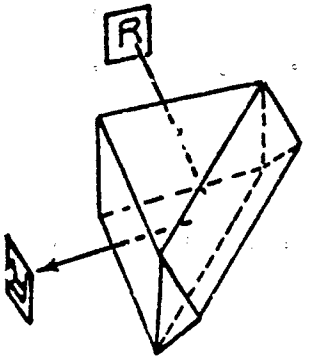
RIGHT ANGLE
PRISM
(B)



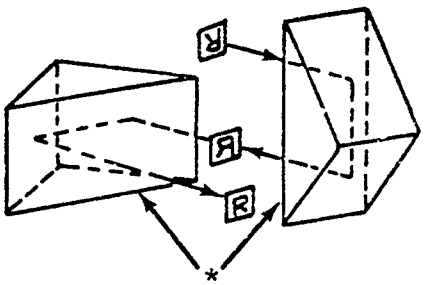
BEAM
SPLITTER
(C)

FIGURE 3-8. Prism Configurations and Functions

(Sheet 1 of 4)

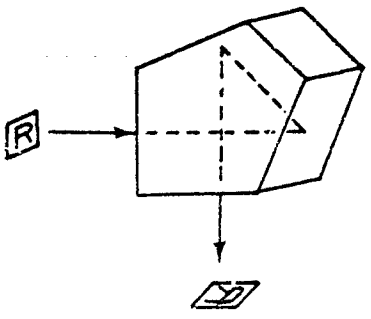


AMICI OR ROOF
PRISM
(D)



PORRO PRISM
COMBINATION (E)

*These faces are usually in contact. They have been separated for illustration only.



PENTA PRISM
(F)

FIGURE 3-8. Prism Configurations and Functions
(Sheet 2 of 4)

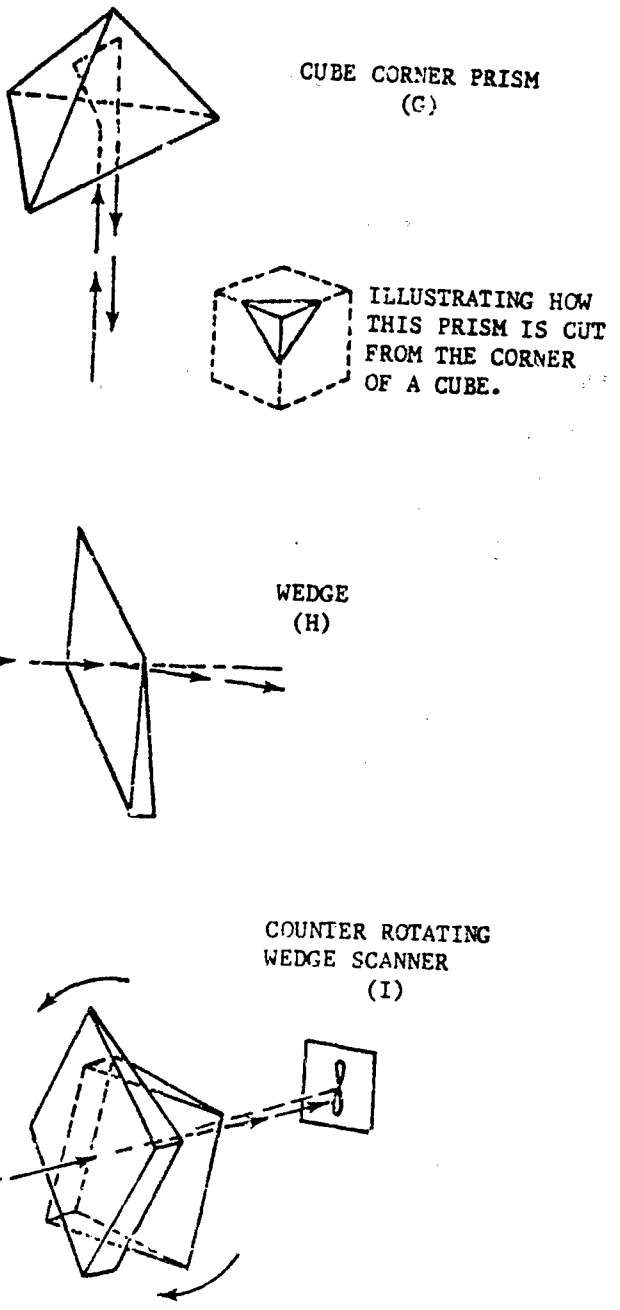
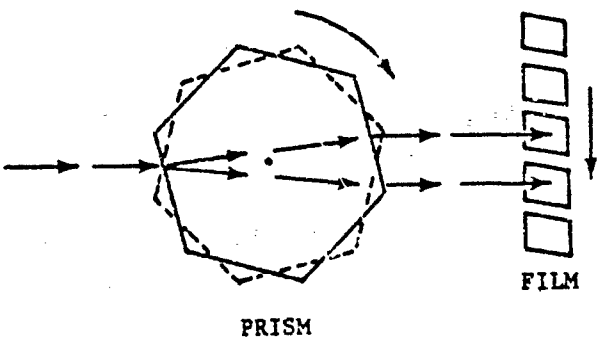
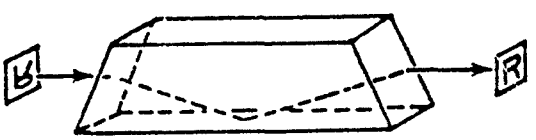


FIGURE 3-8. Prism Configurations and Functions

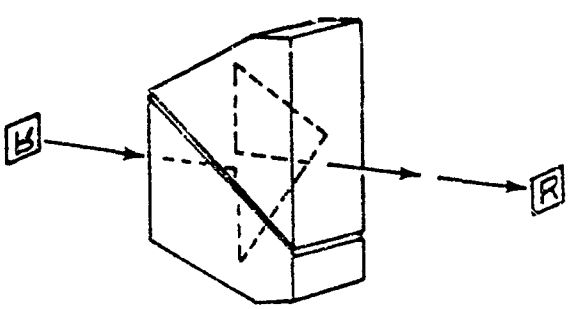
(Sheet 3 of 4)



MULTIFACET
ROTATING
PRISM
(J)



DOVE PRISM
(K)



PECHAN PRISM
(L)

FIGURE 3-8. Prism Configurations and Functions

(Sheet 4 of 4)

12. Image Rotation. There are two basic types of prisms available for producing an axial rotation of a beam or of the image it carries. Fig. 3-8 (K) shows a Dove prism and Fig. 3-8 (L) a Pechan prism. In both types the beam axis at the entrance is in line with the axis at exit. Also, rotation of the prism on the beam axis produces a double speed rotation of the image.

The Dove prism should be used only in a collimated beam of radiation since otherwise the inclined entrance and exit surfaces will introduce severe aberrations into the image. With the Pechan prism these surfaces are normal to the beam, and they affect the image to a considerably lesser degree.

The Pechan prism is made in two parts separated by an air gap of about 0.002 in. The beam is initially incident on the air gap layer at an angle steep enough to produce total internal reflection. At the second encounter, the beam is normal to the same and will be transmitted through. At the third encounter, it is again totally reflected.

Image rotator prisms are often used in dynamic display systems (panoramic scanners, periscopes, etc.) to maintain the image erect relative to a particular observer.

13. Polarization. The phenomenon of polarization is explained in another chapter of this book. In this paragraph it is intended only to make note of the use of special prisms to separate the two polarized components of a composite beam. Details of the construction and action of two types of polarizing prisms, Nicol and Rochon prisms, are found in Ref. 24, Ch. 24.

3-2.3 OPTICAL DESIGN

3-2.3.1 Definitions, Notations, and Sign Conventions

3-2.3.1.1 Optical Systems

In abstract concept, an optical system is a mathematical or geometrical construct which describes the relationship of object space to image space. Physically it is an apparatus which makes use of and controls a beam of electromagnetic radiation for the purpose of establishing and maintaining the object space to image space relationship.

3-2.3.1.2 Object

An object is the nominal source of the radiation upon which the optical system operates. That is, each ray is considered to originate at some point in the object. The ultimate source of the radiation, however, may be considerably removed from the object.

3-2.3.1.3 Image

An image is the grouping or assemblage of rays in a localized region in space. It represents the nominal terminal point of the radiation which is passed by an optical system. Unless absorbed by physical means at the image, the radiation will, nevertheless, continue beyond this terminal point.

3-2.3.1.4 Object and Image Space

The direction of rays in an optical system is from the object, through the system, to the image. All the space on the front side of the system (where the object is located) is known as object space. All the space on the side where the image is located is the image space.

3-2.3.1.5 Point Image

Most of the analytical work in optics deals with the image produced by an optical system from a point (infinitesimally small) object.

The image of a point, however, is never as small as the object due to the physical nature of the radiation that forms it (diffraction) and to imperfections in the optical system (aberrations). It is nevertheless called a "point image."

3-2.3.1.6 Extended Image

An extended image, in correspondence with an extended object, is considered to be an assemblage of an infinite number of point images. Since these images are finite in size they must necessarily be conceived as overlapping each other, and each point in the extended image is built up with contributions from many point images.

3-2.3.1.7 First-order (Gaussian) Optical Theory—Paraxial Region

The First-order Optical Theory is a logical discipline describing the relationship of an object and an image in an idealized setting. The

theory does not depend upon, and does not take into account, the physical behavior of electromagnetic radiation.

As applied to an optical system which is symmetrical about an axis, the first-order mathematical formulations describing the object-image relationship are very precise when restricted to the so-called paraxial region which is an infinitely narrow region around the axis. For practical purposes, it has been established that the same formulas are valid to a good degree of approximation when applied to a region of finite extent around the axis. This may be called the extended paraxial region.

3-2.3.1.8 Ray—Slope of a Ray

In geometrical optics, the concept of a ray has been effectively used to describe the characteristics and properties of systems, even though the fundamental nature of electromagnetic radiation does not admit to the existence of such. The basic characteristics of a ray are its infinitely narrow width and its propagation along a straight-line path in homogeneous media.

In first-order theory, a quantity u is defined as the slope of a ray relative to the optical axis, wherein the slope has the usual connotation of being the tangent of the angle which the ray makes with the axis. Since an angle and its tangent are nearly equal up to about 0.1 rad, u and tan u are often used interchangeably in formulations pertinent to the extended paraxial region.

3-2.3.1.9 Sign Conventions

In the present exposition of first-order optics, the following conventions apply and the accompanying notations and signs are used:

1. Radiation is assumed to progress generally from left to right (Fig. 3-9).

2. The direction of the axis is generally positive toward the right.

3. Radii and curvatures are positive if the center of curvature is to the right of a given surface.

4. Distances measured from left to right are positive. In diagrams, the direction of measurement is indicated with an arrow.

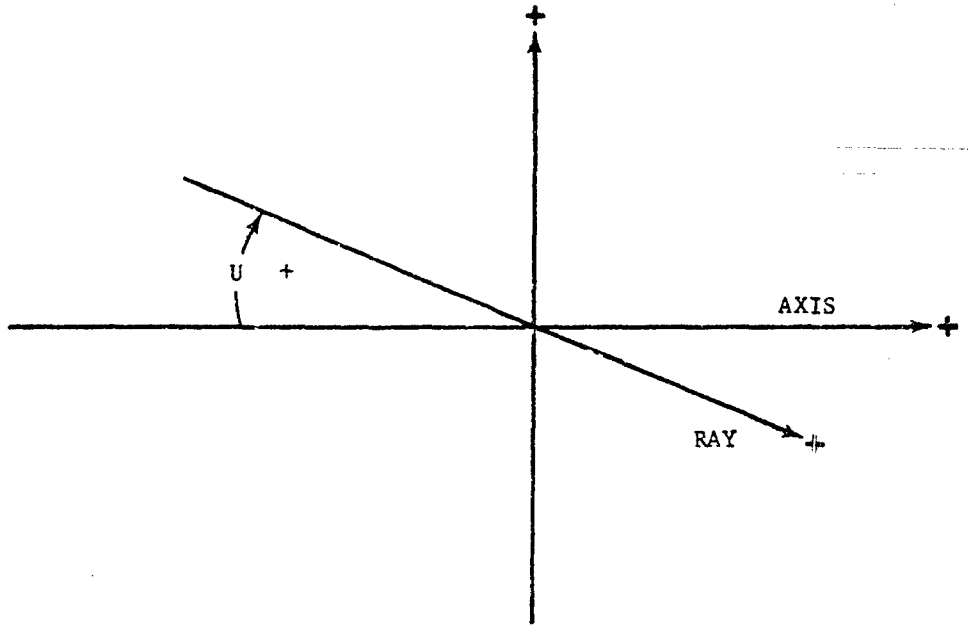


FIGURE 3-9. Conventional Geometry in First-order Optics

5. Distances measured upward from a horizontal axis are positive.

6. Each slope angle is designated by a curved arrow starting at the axis and extending to the ray. The sign of the angle is positive if the arrow shows a clockwise rotation.

7. Angles of incidence and refraction are measured from the normal to the surface. The angles are positive if the measurement is counterclockwise. Note the variance relative to Rule 6.

8. The index of refraction is a positive number greater than 1. A vacuum has an index of 1; air, slightly greater than 1. A reflective surface is considered to be equivalent to a refractive medium having an index of (-1).

3-2.3.1.10 Nomenclature

SYMBOL	DEFINITION
BFL	Back focal length
c	Surface curvature, equal to 1/r, cm^{-1}
d, *d'*	Principal plane distances from reference surfaces
EFL	Effective focal length
f_1, f_2	Focal points. An optical system characteristically acts to change a bundle of parallel rays so that they converge toward a point in space which is called a focal point.
f, f', f_a, f_b	Focal lengths or focal distances. Focal distances are the distances measured from the focal point to the principal plane.
FFL	Front focal length
h	Object or image height; radial distance from the optical axis to a point in the object or image field
i	Subscript denoting the number of a particular element or surface in a series of such, integer
k	Subscript denoting the last surface or element in a series, integer
m	Lateral magnification, dimensionless
n	Index of refraction of first medium, dimensionless
n'	Index of refraction of second medium, dimensionless
p_1, p_2	Principal planes. The intersection of the parallel rays in an idealized system with the converging rays. This is nominally the region where the optical action (refraction or reflection) takes place.
P_1, P_2	Principal points. Intercepts of the optical axis and the principal planes.
r	Radius of curvature of a surface
s, *s'*	Object-image distances from principal planes
t	Thickness of an element, i.e., the spacing between surfaces
u	Slope of a ray; angle between a ray and the optical axis
x, *x'*	Object-image distances from focal points
y	Ray height, measured from the axis to the ray at or near a refracting or reflecting surface

All mensurable quantities listed are in terms of length units, unless otherwise specified.

3-2.3.2 First-order Theory—Formulations

3-2.3.2.1 Cardinal Points of Optical Elements

The parameters which determine the image-forming characteristics of an optical element are the curvatures of the surfaces, the index of refraction of the material, and the position of the surfaces relative to the object or image. The relationship is a complicated one. However, these parameters can be converted to others, namely, the focal distances and principal plane spacings which to a first approximation define the object-image relationship in a greatly simplified manner. The focal and principal points of an element or system belong to a family called cardinal points, which include also a pair called nodal points. Nodal points have a property such

that if a lens is tilted or rotated on any axis through the same, no change in image position results if the object is fixed. For a lens in air the nodal points and principal points are coincident.

The location of focal points and principal points is computed as follows:

For a single-surface element: (Refer to Fig. 3-10)

$$f_a = \left(\frac{n'}{n' - n}\right) r \qquad (3\text{-}1)$$

where $n' > n$

and

$$f_b = -\left(\frac{n}{n' - n}\right) r \qquad (3\text{-}2)$$

Thick Lens: (Refer to Fig. 3-11)

$$FFL = f\left[1 - \left(\frac{n' - n}{nr_1}\right) t\right]; \frac{1}{r_1} = c_1 \qquad (3\text{-}3)$$

= curvature of front surface

$$BFL = f\left[1 - \left(\frac{n' - n}{n'r_2}\right) t\right]; \frac{1}{r_2} = c_2 \qquad (3\text{-}4)$$

= curvature of back surface

$$1/EFL = \frac{1}{f} = -\frac{1}{f'}$$

$$= (n' - n)\left[\left(\frac{1}{r_1}\right) - \left(\frac{1}{r_2}\right) + \left(\frac{n' - n}{n'r_1 r_2}\right) t\right] \qquad (3\text{-}5)$$

$$d = \frac{-r_1 t}{n'(r_1 - r_2) - (n' - n)t} \qquad (3\text{-}6)$$

$$d' = \frac{-r_2 t}{n'(r_1 - r_2) - (n' - n)t} \qquad (3\text{-}7)$$

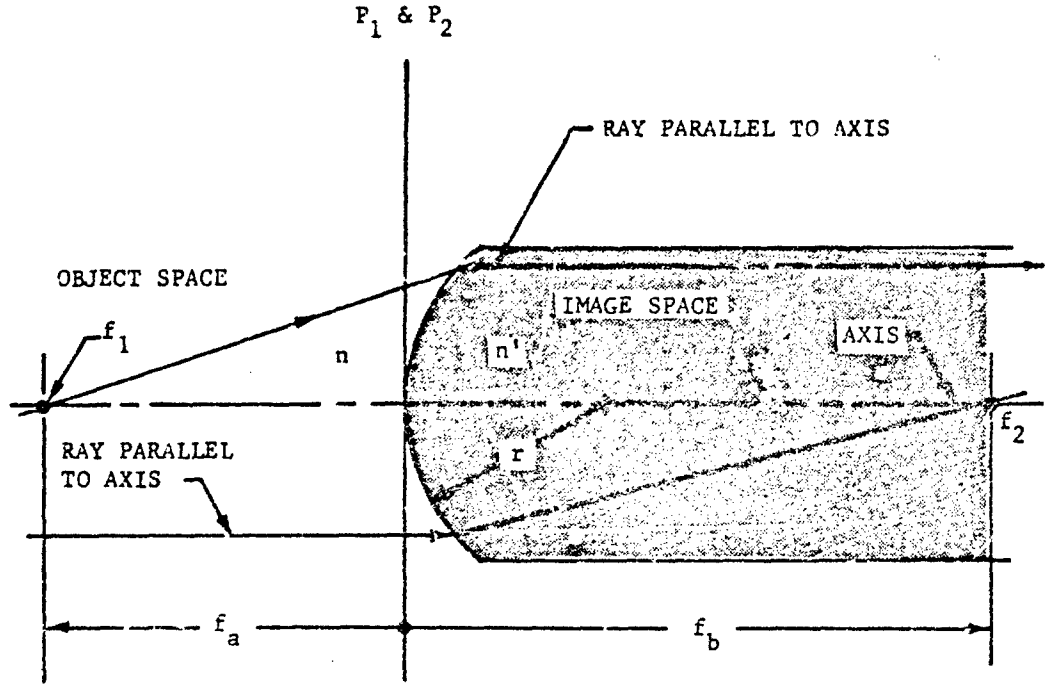


FIGURE 3-10. Cardinal Points for Single-surface Element

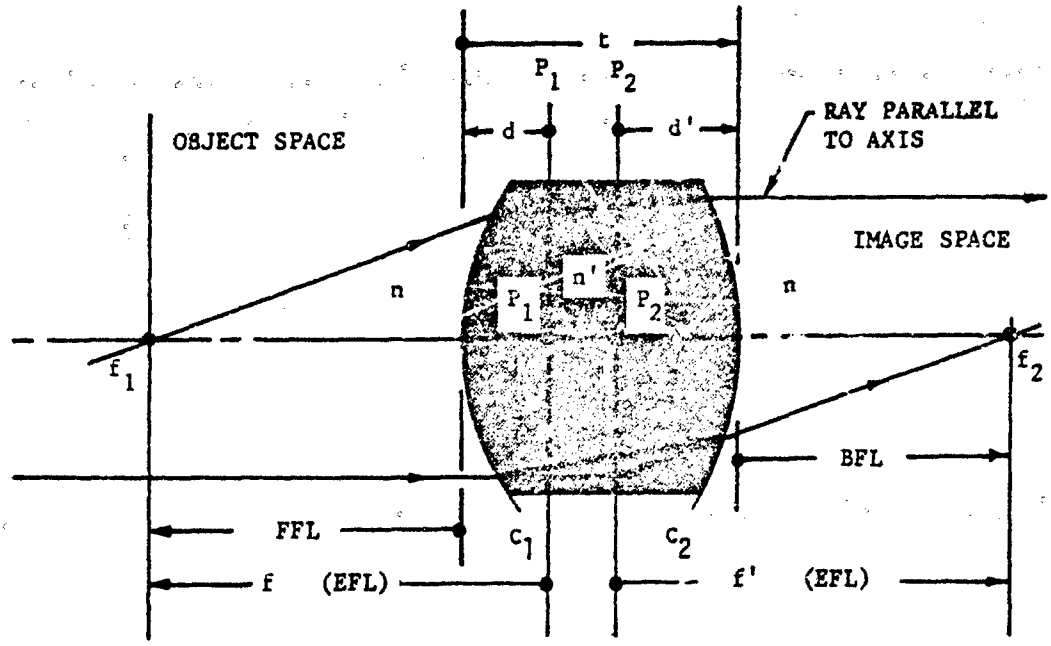


FIGURE 3-11. Cardinal Points for Double-surface Thick Lens Element

Thin Lens:

A thin lens is defined as one wherein the spacing between principal points, P_1 and P_2, is negligible compared with other parameters. Therefore, if we let $t \rightarrow 0$ in Eq. 3-5

$$1/EFL = \frac{1}{f} = (n' - n)\left(\frac{1}{r_1} - \frac{1}{r_2}\right) \qquad (3\text{-}8)$$

which is called the Lens Maker's Formula.

3-2.3.2.2 Multiple Element Systems

The location of cardinal points in multiple element systems, Fig. 3-12, is determined most easily by means of an elementary ray trace, using the pair of Eqs. 3-9 and 3-10 at each element in succession.

$$u'_i = u_i + \frac{y_i}{f_i} \qquad (3\text{-}9)$$

$$y_{i+1} = y_i - d'_i u'_i \qquad (3\text{-}10)$$

where

u_i = slope of the incident ray relative to the optical axis

y_i = height at which the ray strikes the element

(Refer to Fig. 3-12 for sample problem setup.)

For the purpose of the ray trace, assume a light beam parallel to the optical axis, i.e., let $u_1 = 0$; y = arbitrary value; $f_1, f_2, \ldots f_k$ = focal lengths of successive elements; $d'_1, d'_2, \ldots d'_k$ = spacing of elements as shown. The order of the operations is as follows:

(1) $u'_1 = u_1 + \frac{y_1}{f_1} = u_2$

(2) $y_2 = y_1 - d'_1 u'_1$

(3) $u'_2 = u_2 + \frac{y_2}{*f_2} = u_3$

(4) $y_3 = y_2 - d'_2 u'_2$

(5) $u'_3 = u_3 + \frac{y_3}{f_3}$

(6) $BFL = y_k/u'_k = y_3/u'_3$ (3-11)

(7) $EFL = y_1/u'_k = y_1/u'_3$ (3-12)

*Note: f_2 is a negative number for the negative lens.

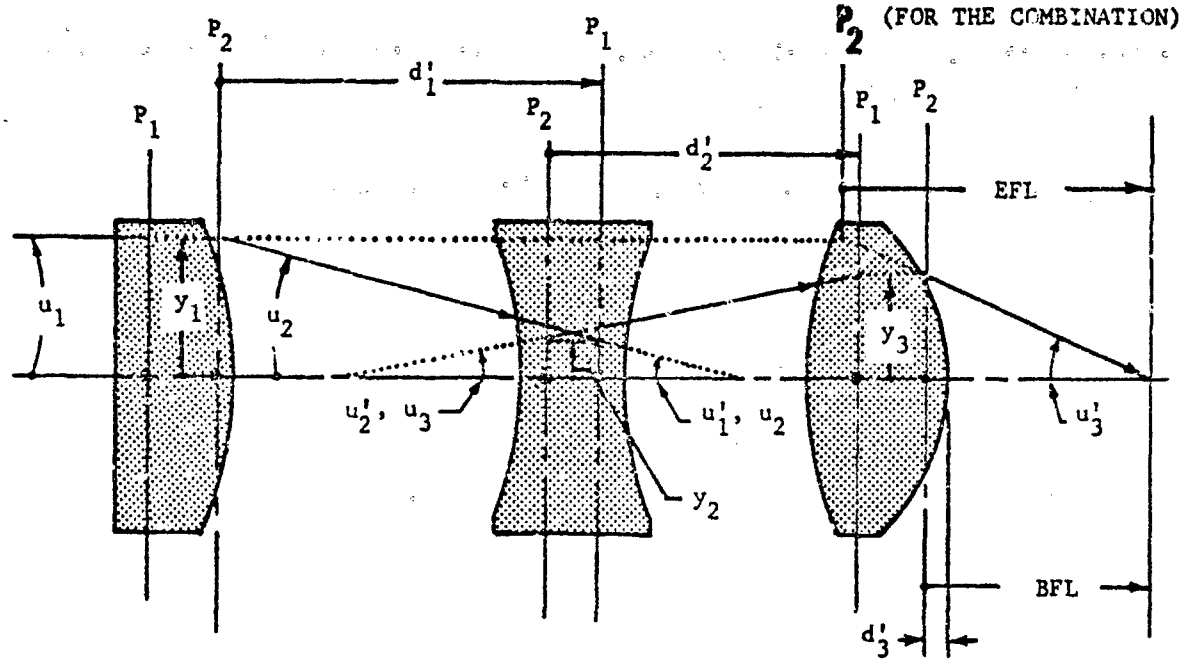


FIGURE 3-12. System Cardinal Points by Ray Tracing

Each optical system has two focal points and two principal planes. Eqs. 3-11 and 3-12 locate only one of each of these. To locate the second focal point and principal plane for the combination, trace a similar ray through the system backwards.

The *BFL* measured from the principal plane of the last element as shown is of value only to the designer. The optician who will assemble, test, and use the system needs to know the *BFL* as measured from the last surface. By treating the third element separately as a thick lens, a distance d'_3 may be calculated by subtracting the distance (Eq. 3-7) from the indicated *BFL*, thereby obtaining the required net value.

3-2.3.2.3 Image Position, Magnification, and Virtual Image

The equations for image-object and magnification relationships apply to both single element and compound systems. A single pair of focal points and principal points may be defined for a compound system to fully describe its first-order optical characteristics, regardless of the shape, index, or spacing of the individual elements.

Reference is made to Figs. 3-13, 3-14, and 3-15.

Image Position

$$\frac{1}{s'} = \frac{1}{s} + \frac{1}{f} \quad \text{Gaussian Formula} \qquad (3\text{-}13)$$

$$xx' = ff' \quad \text{Newtonian Formula} \qquad (3\text{-}14)$$

Lateral Magnification

$$m = \frac{h'}{h} = \frac{s'}{s} = -\frac{f}{x} = \frac{x'}{f'} \qquad (3\text{-}15)$$

where h and h' are the radial distances from the optical axis to a point in the object or image field, respectively.

An optical system with strong positive power will act upon the diverging ray bundles out of the object, and make them converge to form a real image as in Fig. 3-13. A weaker positive system or a negative system can only reduce or increase the divergence so that a real image cannot be formed. Instead, a virtual image is defined at the region out of which the re-directed rays appear to originate as in Fig. 3-15. The virtual image, thus, is the equivalent of a displaced object. It is to be noted that no real rays exist in the virtual image space.

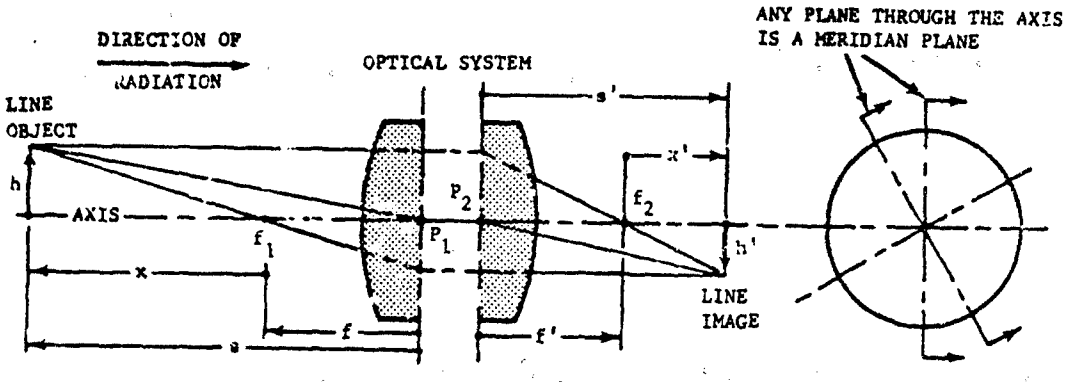
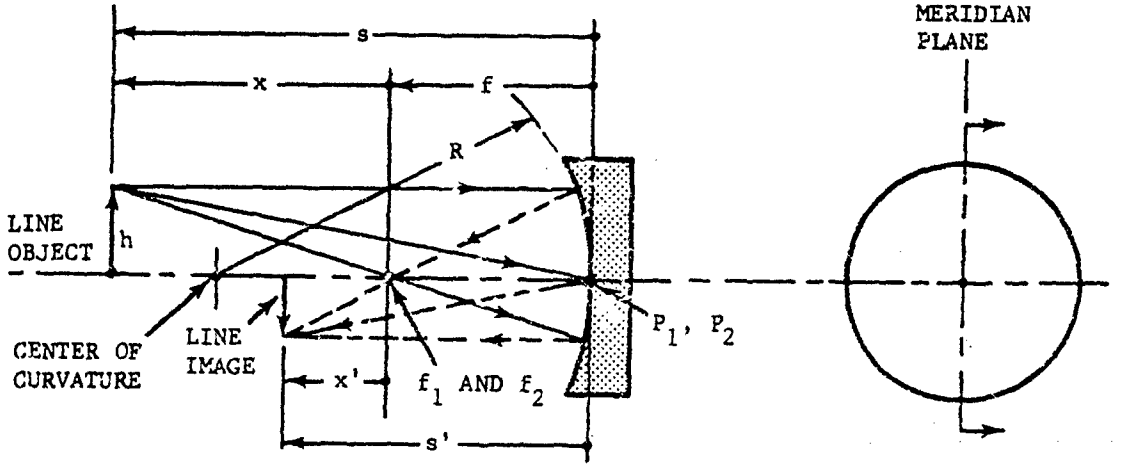


FIGURE 3-13. Meridian Plane Section of a Refractive Optical System



NOTE: Object space and image space are superimposed. Dotted lines are used to distinguish one from the other.

FIGURE 3-14. Meridian Plane Section of a Reflective System

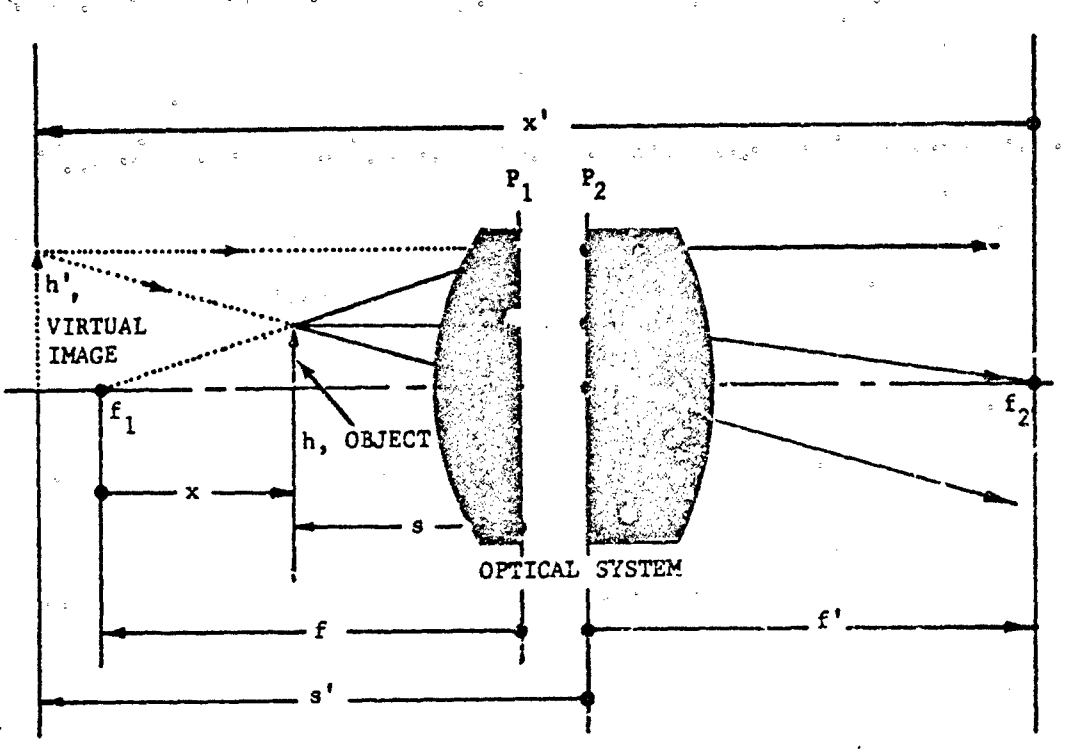


FIGURE 3-15. Virtual Image Formation by an Optical System

3-2.3.2.4 Sine Condition–Lagrange Law of Invariance

The Sine Condition of Abbe and the Lagrange Law of Invariance are highly effective tools for analyzing complex optical systems. One application of these is demonstrated in Fig. 3-16, which represents a typical optical relay system of k stages.

Abbe's Sine Condition:

$$\begin{aligned} I\,(\text{invariant}) &= n_1 h_1 \sin u_1 \\ &= n'_1 h'_1 \sin u'_1 \\ &= n_2 h_2 \sin u_2 \\ &= \ldots n'_k h'_k \sin u'_k \end{aligned} \qquad (3\text{-}16)$$

From which is obtained:

$$\begin{aligned} \frac{h'_k}{h_1} &= \text{magnification } m \\ &= \frac{(n_1) \sin u_1}{(n_k) \sin u_k} \\ &= \frac{\sin u_1}{\sin u'_k} \quad (\text{in air}) \end{aligned} \qquad (3\text{-}17)$$

This statement describes the relationship which must exist if an optical system is to produce perfect image points in the vicinity of the axis. Thus, an image to be free of aberrations requires constant magnification for all optical paths from object to image. The ratio $\sin u_1 / \sin u'_k$ must, therefore, be a constant.

The Lagrange Law of Invariance is similar to the Sine Condition with u substituted for $\sin u$. It applies in the paraxial region where $u \approx \sin u$.

If the system of Fig. 3-16 is ray-traced to obtain u'_k, then the system magnification and image height h'_k can be determined without reference to the intermediate stages using the Lagrange Law in the form,

$$\frac{h'_k}{h_1} = \frac{u_1}{u'_k} \qquad (3\text{-}18)$$

From these, the parameters of overall focal lengths, back focal lengths, and location of principal planes can be determined.

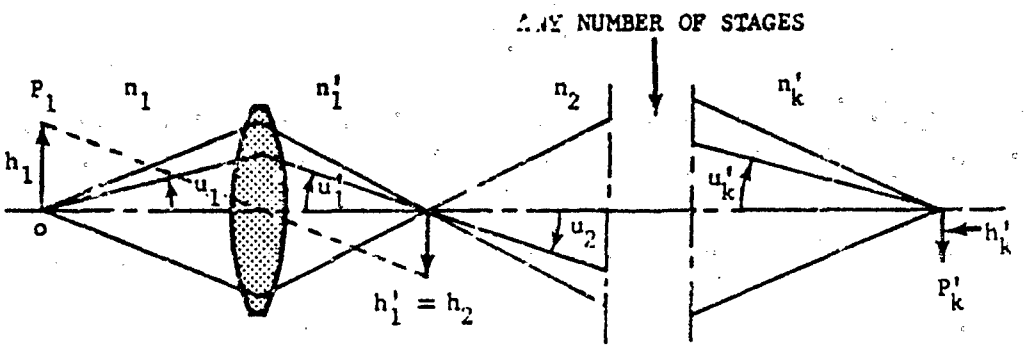


FIGURE 3-16. Application of Sine Condition

3-2.3.3 Limitation of Rays

The aperture stop of an optical system is a physical element which limits the size of the ray bundles that form each typical image point (see Figs. 3-17 and 3-18). At the position of the aperture, all the image-forming rays are completely intermingled. The principal or chief ray passes through the center of the aperture stop, hence, it is the central ray of each bundle that forms an image point. Each principal ray crosses the axis at the aperture stop and possibly at several other points. Each of these alternate crossing points locates an "image of the aperture", and each such image point could be selected as the position for the physical aperture stop, so that with suitable size adjustments, the system performance will remain unchanged.

In a compound optical system it is possible to locate an image for each lens element in the series. Some of these will be virtual images. A physical stop (called a field stop) may be placed at the position of any real image to limit or define its size. This then also defines the field angle or viewing angle of the system. The half-field angle is measured at the entrance pupil between the axis and the chief ray which intersects the outermost point (edge) of the image.

3-2.3.3.1 Entrance Pupil–Exit Pupil

In any optical system there is defined an image of the aperture stop formed by the portion of the system in front of that stop. This is the entrance pupil. Similarly, the image of the aperture stop formed by the optics behind it is the exit pupil.

As may be observed in Fig. 3-17, the entrance pupil is not always located at the first element of the assembly. In this system its size is equivalent to the clear aperture diameter of the first lens, but its position is coincident with that of the aperture stop behind the second lens. In Fig. 3-18 the entrance pupil is in front of the assembly, but its size is considerably smaller than that of the clear aperture of the objective lens.

Field angle and aperture angle (ray bundles) computation should always be based on the entrance and exit pupils as references, not on the real aperture stop.

3-2.3.3.2 Relative Aperture, Speed, f/no., and Numerical Aperture

Relative aperture, speed, and f/no. all refer to the same characteristic of an optical system, i.e., the angular size of the image-forming ray bundles. It is clear that this size determines the energy density arriving at the image. Thus, from the effect of energy density on a photographic plate, the concept of speed has been derived.

The clear aperture of a system is defined as the diameter of the entrance pupil.

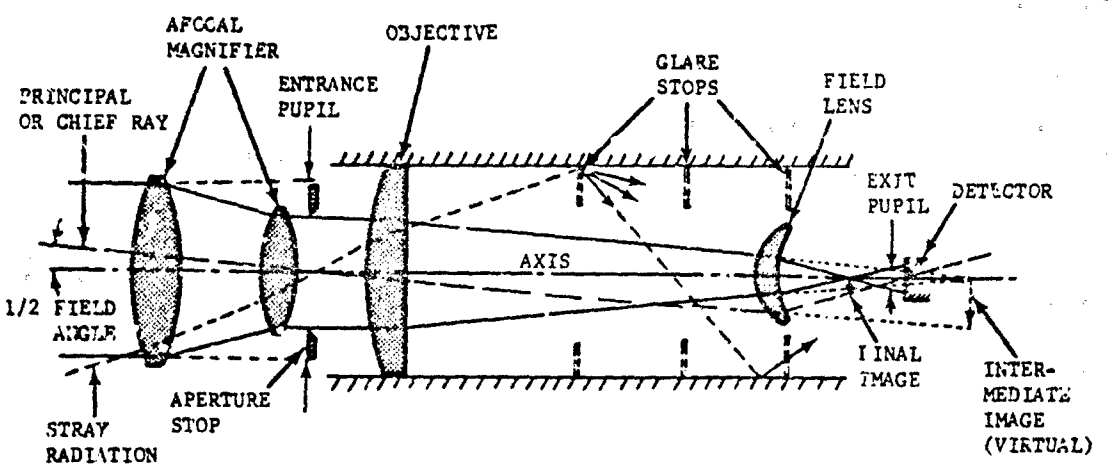


FIGURE 3-17. 19 Optical System Showing Arrangement of Stops and Baffles

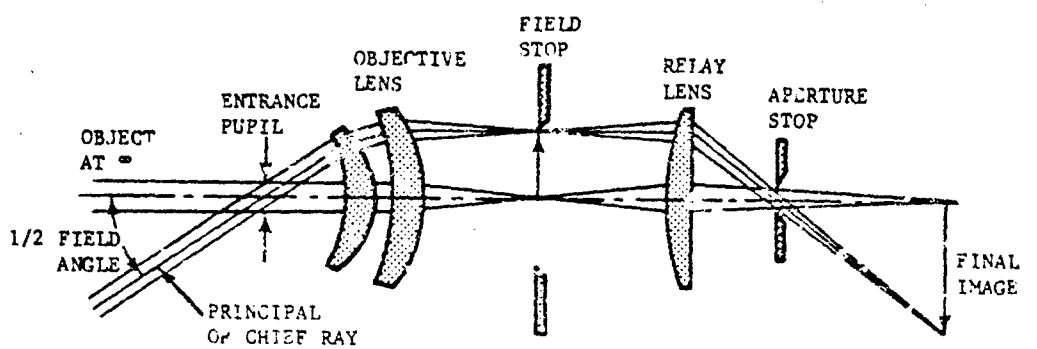


FIGURE 3-18. Display Pickup System

The f/no. of a bundle is determined from the effective focal length (EFL) and clear aperture (CA) of the system, taken as a ratio, which incidentally approximates the reciprocal of the cone angle of the bundle in radians. Thus, for a system with an object at infinity

$$\text{Relative Aperture} = \text{f/no.} = EFL/CA$$

Numerical Aperture (NA) is defined as the sine of the halfcone angle u of the image forming ray bundles multiplied by the index of refraction of the image space, thus

$$NA = n \sin u \tag{3-19}$$

This is, likewise, a measure of the same characteristic as f/no.

In a well-corrected system where the optical path length from all portions of the exit pupil is constant so that

$$\sin u = \frac{1}{2}\left(\frac{CA}{EFL}\right) \text{ (see Fig. 3-19)}$$

and if n is for air, then

$$NA = \sin u = \frac{1}{2}\left(\frac{CA}{EFL}\right) = \frac{1}{2}\,\frac{1}{(\text{f/no.})} \tag{3-20}$$

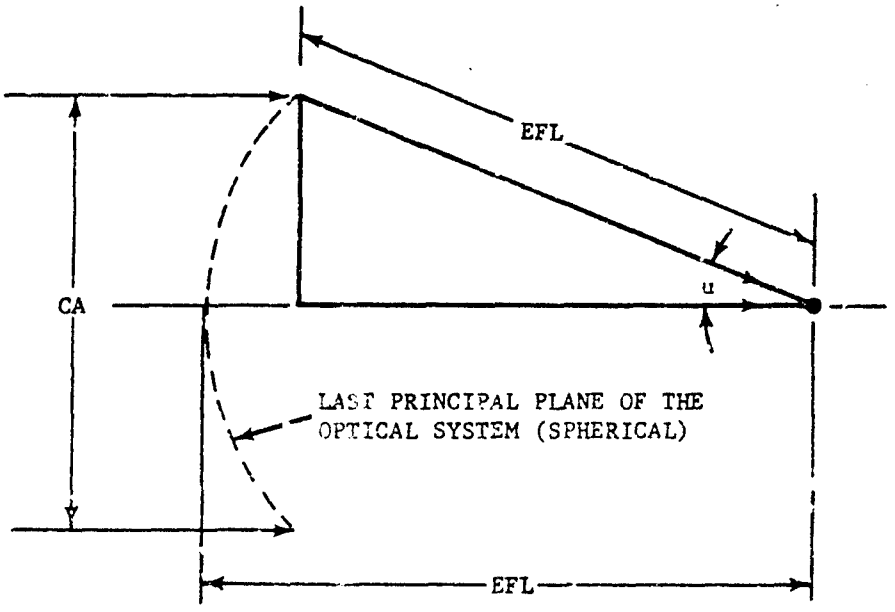


FIGURE 3-19. Relationship of Clear Aperture and Effective Focal Length

3-2.3.4 Aberrations in Third-order Optics

3-2.3.4.1 *Third-order Theory*

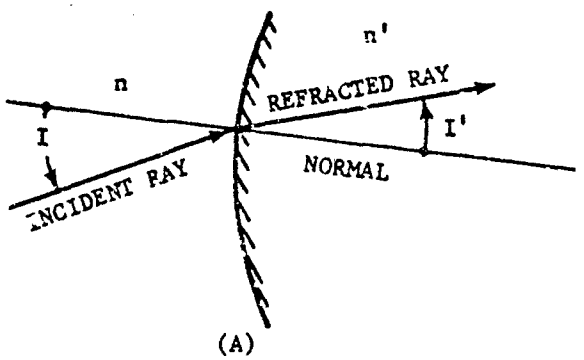
The First-order (Gaussian) Theory of optical imagery is based on a purely mathematical or geometrical structure which relates an object space to an image space. It describes, thereby, the object-image relationship of perfect or idealized optical systems. Real optical systems, however, are always imperfect and generally are afflicted with deviations from the ideal. These are called aberrations.

The exact behavior of rays in a real optical system is described by two basic laws: (1) Snell's Law of Refraction, $n' \sin I' = n \sin I$, and the Law of Reflection $I = I'$. These are illustrated in Fig. 3-20.

Snell's Law can be stated in the form:

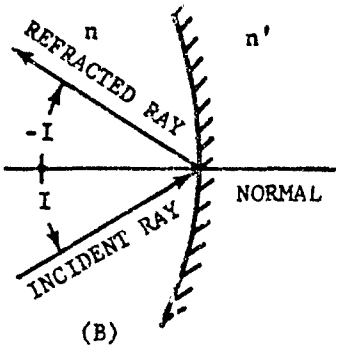
$$\sin I' = \left(\frac{n}{n'}\right) \sin I$$

from which it appears possible to have $\sin I'$ greater than 1, if n is greater than n'. This occurs when the incident ray is in the denser of the two mediums. A limiting incidence angle I exists for which I' is 90 deg, and this is called the critical angle. If the incidence angle is greater than the critical angle, the ray will be totally reflected.



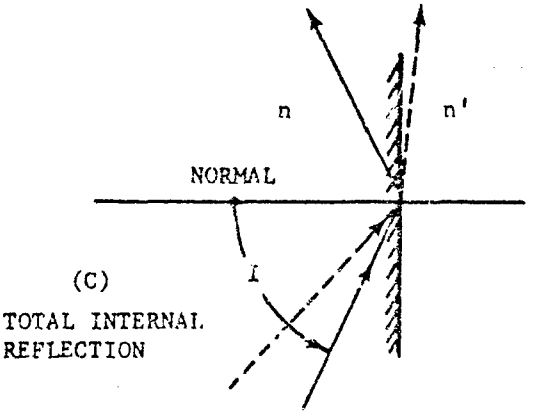
(A)

REFRACTION OF A RAY
AT AN INTERFACE BETWEEN
MEDIA OF DIFFERENT INDICES



(B)

REFLECTION OF A RAY
AT AN INTERFACE BETWEEN
MEDIA OF DIFFERENT INDICES



(C)

TOTAL INTERNAL
REFLECTION

FIGURE 3-20. Geometrical Presentation of Refracted and Reflected Rays

Using the basic laws of refraction and reflection, along with the geometry of the optical components of a system, it is possible to trace the exact path of any number of rays through the system and determine from these the performance of the system. Well-established and standardized ray tracing formulas can be found in almost any optics text book (e.g., Refs. 25 and 26). Analysis of optical systems by direct ray tracing had been a tedious and time-consuming process until the recent advent of high-speed digital computers. As a consequence, a body of theory has been developed in past years by means of which the ray tracing formulations are compared directly with the idealized formulas of the Gaussian theory so that the nature and magnitude of the aberrations may be closely examined and analyzed.

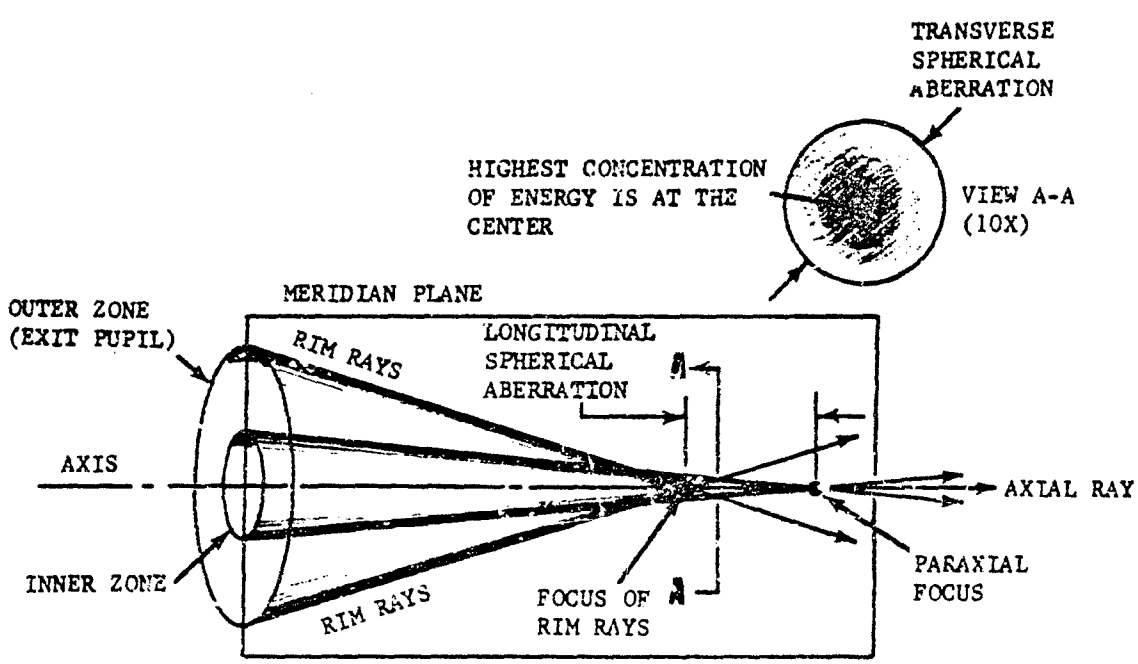
This type analysis is further aided by using, in place of certain trigonometric functions in the ray tracing formulas, their equivalent series expansions and truncating these at a low order level. It has been found that inclusion of third-order terms of the series, only, will provide a reasonably accurate accounting for the aberrations.

This technique of analysis represents the third-order theory of optical imagery.

3-2.3.4.2 Monochromatic Aberrations

In third-order theory, the deviation of a representative ray in an optical system from the path described by the Gaussian theory is expressed as a polynomial consisting of five terms. Each term describes a different type of aberration. Accordingly, any term which is equal to zero represents the absence of a particular type in the optical system. In a perfect optical system all the terms are zero.

These five Seidel aberrations (named after the investigator who first described them) include:



WHEN RIM RAYS FOCUS TO THE LEFT OF THE PARAXIAL FOCUS,
THE SPHERICAL ABERRATION IS POSITIVE (+), PER DIAGRAM.
WHEN RIM RAYS FOCUS TO THE RIGHT OF THE PARAXIAL FOCUS,
THE SPHERICAL ABERRATION IS NEGATIVE (-).

FIGURE 3-21. Optical Diagram Illustrating Spherical Aberration

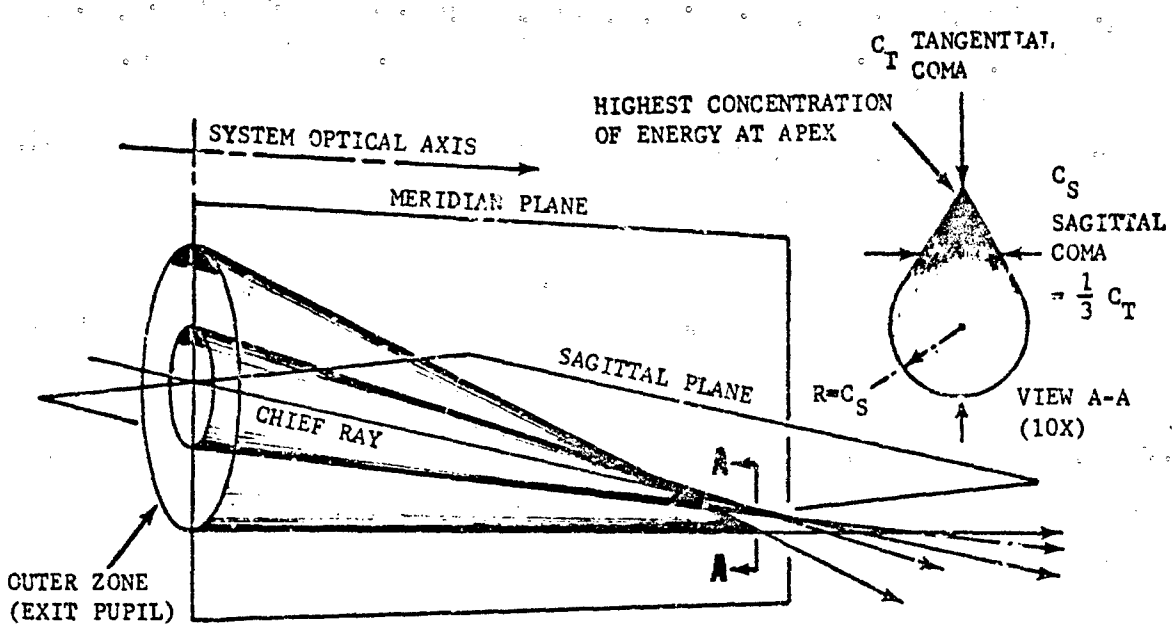


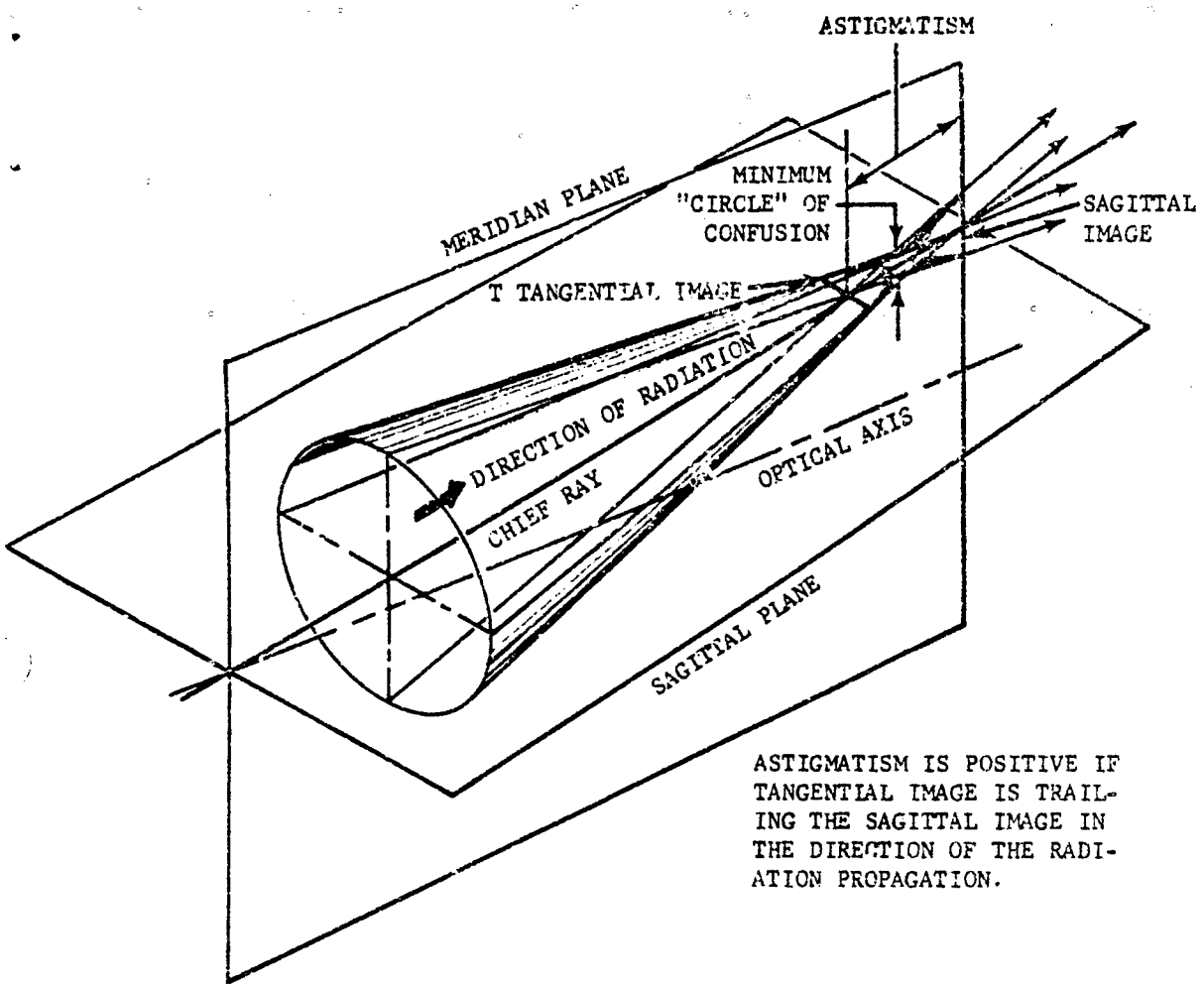
FIGURE 3-22. Optical Diagram Illustrating Coma Aberration

1. Spherical Aberration (SA). When the rays in the outer zones of a bundle come to a focus at a point axially displaced from the focus of the rays close to the axis (paraxial), the result is spherical aberration. It is measured in either the longitudinal or the transverse direction, as shown in Fig. 3-21.

2. Coma (C). When the rays in the outer zones of an off-axis bundle focus at a point which is displaced both laterally and longitudinally from the focus of the rays in the inner and central zones, coma is said to exist (Fig. 3-22). Coma is measured by the degree of deviation of the ray intercepts in the image plane from a nearly perfect image point defined by the sine condition. This is called OSC (offense against the sine condition) by Conrady[25].

3. Astigmatism (A). This aberration occurs when the rays of a meridianal fan come to a focus at a point which is longitudinally displaced from the focus of the sagittal fan of rays. Each focus is a short line, and the two lines are rotated 90 deg with respect to one another (Fig. 3-23). The linear displacement between the T and S images is a measure of the astigmatism which is present. However, in some instances it is necessary to determine the minimum circle of confusion between T and S. If the f/no. of the bundle is known, the diameter of this blur circle can be calculated as

$$\frac{1}{(\text{f/no.})} \times \frac{\text{Astigmatism}}{2} \qquad (3\text{-}21)$$



IF THE f/no. OF THE BUNDLE IS KNOWN, THEN THE DIAMETER OF MINIMUM CIRCLE OF CONFUSION IS GIVEN BY

$$\frac{1}{(f/no.)} \times \frac{ASTIG}{2}$$

FIGURE 3-23. Optical Diagram Illustrating Astigmatism

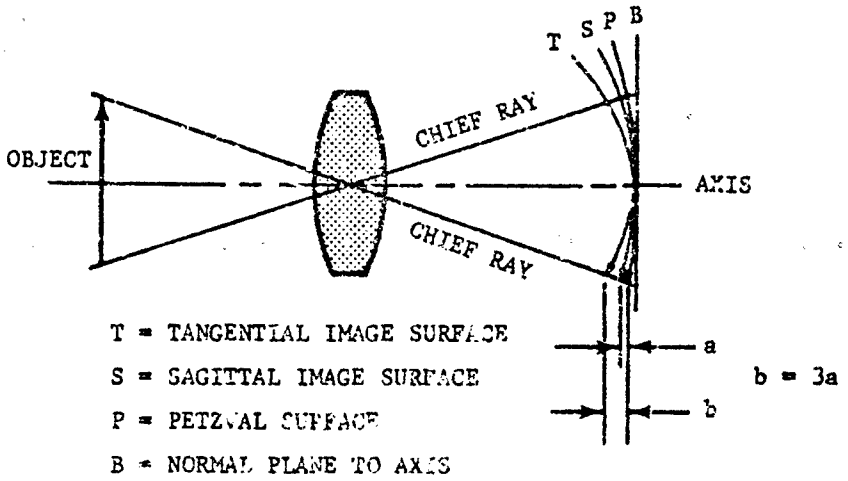


FIGURE 3-24. Field Curvature

4. Field Curvature. Field curvature is a departure of the off-axis image points from the normal plane through the paraxial focus point. Fig. 3-24 shows the loci of tangential and sagittal off-axis image points for a simple optical system along with a reference surface P called the Petzval surface. The Petzval surface is fixed by the overall parameters of an optical system and usually cannot be changed significantly by shifting or bending the lens elements. The T and S surfaces, however, can be changed by such manipulation but always in such a way that all tangential image points are three times as far from the Petzval surface as are the corresponding sagittal image points. In simple optical systems the three stated surfaces are usually paraboloids of revolution.

5. Distortion. Distortion is an aberration which affects mainly the principal rays of an optical system (Fig. 3-25). Whenever these rays do not follow the path prescribed by the Gaussian theory, the result is an image which is no longer similar in every detail to the object. Barrel distortion occurs when magnification decreases toward the outer zones of the field. Conversely, pincusnion distortion occurs when magnification increases toward the outer zones. This distortion is measured as the spatial difference between a point in a particular region of the image field and a similar point in a perfect image (an image which is exactly similar to the object). The percent distortion is the ratio of the amount of distortion to the distance of the selected image point from the optical axis. The point usually selected for distortion specification is at the extreme edge of the field. In Fig. 3-25

$$\% \text{ distortion} = 100\left(\frac{B - A}{A}\right) \qquad (3\text{-}22)$$

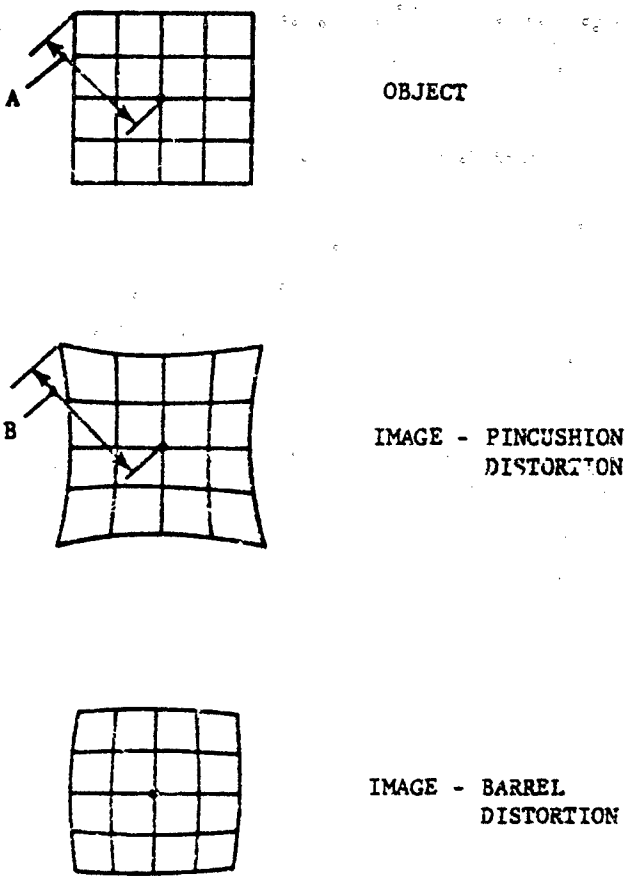


FIGURE 3-25. Examples of Pincushion and Barrel Distortion

3-2.3.4.3 Chromatic Aberration

Chromatic aberration is really a first-order effect due directly to the variation of index of refraction of the elements of the optical system. The two types of chromatic aberrations recognized are longitudinal and transverse. Longitudinal chromatic aberration is a difference in focal position between images formed by two different wavelengths of radiation. Transverse chromatic aberration at the off-axis image points is characterized by a lateral displacement between images formed by two different wavelengths of radiation.

It is conventional practice to select two wavelengths which represent the ends of the spectral band over which the optical system is expected to operate, and define the image displacements for those wavelengths as the chromatic aberration.

3-2.3.4.4 Correction of Aberrations

It should be noted that the various aberrations are characteristics of the representative image points which an optical system produces, and not of the optical system itself. They are, however, functions of the optical system parameters, and often can be explicitly stated as such. If these functional relationships are known, it becomes possible to control the aberrations by the judicious selection and adjustment of the system parameters. The more parameters available, the higher the degree of control obtainable. Thus, the more perfectly corrected systems tend to have many elements, each contributing its own set of parameters.

3-2.3.4.5 *Relation of Aberrations to Optical System Parameters*

The basic parameters of an optical system include: (1) index of refraction of the elements, (2) spacing of the elements, and/or spacing of the surfaces (thickness), (3) curvature of the surfaces, and (4) object or image distance.

In simple optical systems (one or two surfaces), the effects of these parameter variations on optical aberrations are observed directly and an immediate cause-and-effect relationship can be established.

1. Spherical Aberration (SA). In a system which contains a single refractive element (two surfaces), and with the object at infinity, the spherical aberration can be changed by "bending" the lens until minimum aberration is obtained at one particular configuration. Bending is a process of changing the curvatures of two surfaces of an element in parallel while maintaining the net curvature or power of the combined surfaces. (See Fig. 3-26.)

Also in both refractive and reflective single-element systems, the spherical aberration can be reduced to zero by making the surfaces aspheric. A paraboloidal reflective element, for example, shows no spherical aberration for an object at infinity.

2. Coma (C) and Astigmatism (A). These aberrations can be reduced by lens bending, changing the element spacing, and adjusting the stop position. Since generally, element spacing affects astigmatism much more than it does coma, an opportunity exists for differentially adjusting these aberrations.

3. Petzval curvature (PC). This condition is fixed for an element of given power and index of refraction. The PC can be approximated by

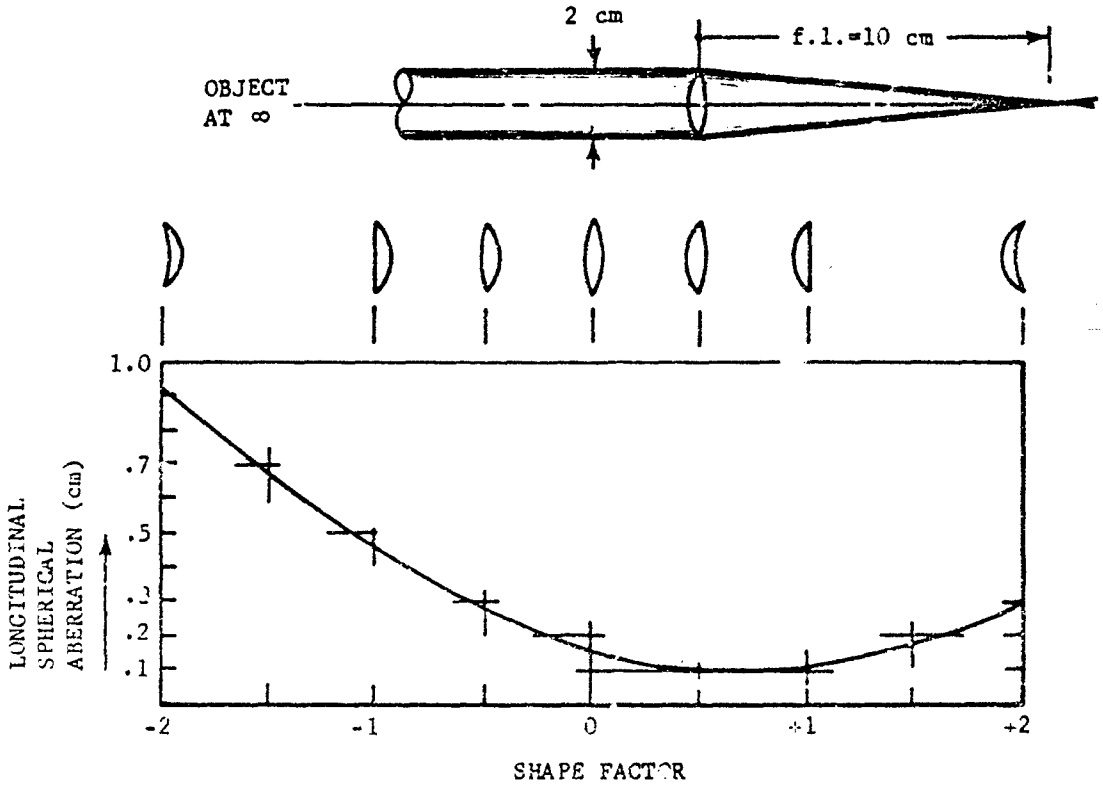


FIGURE 3-26. Lens Bending for Minimum Spherical Aberration

$$PC = \frac{1}{nf} \qquad (3\text{-}23)$$

where n is the refractive index of the material, and f is the focal length of the element.

The Petzval curvature can be changed by varying the index and adjusting the surface curvatures to maintain a fixed focal length or power.

4. Distortion. Small amounts of distortion can be controlled by locating the aperture stop at a given point in a system, and by arranging the system elements symmetrically around the stop. In wide field angle systems, where the distortion cannot be effectively controlled, it may be expedient to accept whatever degree of distortion exists after the other aberrations are balanced out.

5. Chromatic Aberration. Control of chromatic aberrations can be acheived by the combination of two or more elements; the positive power for introducing positive aberrations and negative power for introducing negative aberrations. By using suitable lens materials, the aberrations per-unit-power pertaining to the positive elements will be different from those of the negative elements, so that a combination can be designed to cancel the aberrations while maintaining a given net power.

For more complex systems, the relationship of aberrations to optical parameters has been formalized into a set of equations which show the contribution of each set of component parameters to each of the aberrations in the final image. The total (net) aberrations are the algebraic sums of the individual contributions. These are known as Seidel Sums.

The equations for the Seidel Sums (net aberrations) are given in terms other than the basic element parameters defined so far. The terms used are more immediately descriptive of the total system and particularly of the rays which traverse the system.

It is necessary to begin with an elementary ray trace of two representative rays as shown in Fig. 3-27. A new set of parameters associated with each surface, i, is thus calculated according to the listed formula. Subscripts k identify the final surface.

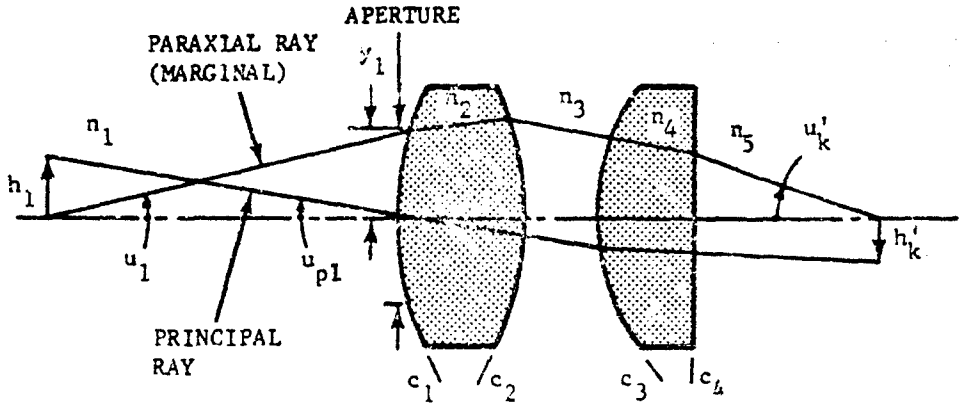


FIGURE 3-27. Elementary Ray Trace Required for Computing Seidel Aberrations

Parameter		Identity
c_i	$= 1/r_i$	Surface curvature for radius r
N_i	$= n_i/n'_i$	Index ratio for media on opposite sides of the surface
u'_i	$= c_i y_i(1 - N_i) + N_i u_i$	Slope of refracted ray
u_{i+1}	$= u'_i$	Slope of incident ray on following surface
y_{i+1}	$= y_i - t_i u'_i$	Ray height after a spacing equal to t_i between surfaces
i_i	$= c_i y_i - u_i$	Angle of incident ray at the surface
i_{pi}	$= c_i y_{pi} - u_{pi}$	Angle of incident principal ray at the surface
I	$= n(uy_p - yu_p)$	Invariant—to be computed from parameters of surface No. 1. Note: $y_{pi} = 0$, if aperture coincides with surface No. 1
h_i	—	Image height. The distance from the optical axis to the farthest point of the image
h_n	$= \dfrac{I}{n_k u_k}$	Image height at last surface or element in a series
B_i	$= n_i y_i(u'_i - i_i)(1 - N_i)/2I$	Parameter
B_{pi}	$= n_i y_{pi}(u'_{pi} - i_{pi})(1 - N_i)2I$	Parameter B associated with principal ray

The final formulas for Seidel Sums are given as:

Parameter		Identity	
ΣSC	$= B_i^2 h'_k/u'_k$	Longitudinal spherical aberration contribution	(3-24)
ΣCC	$= B_i i_p h'_k$	Coma (sagittal) contribution	(3-25)
ΣAC	$= B_{pi}^2 h'_k/u'_k$	Astigmatism contribution	(3-26)
ΣPC	$= \dfrac{I(N - 1)c}{2n}\left(\dfrac{h'_k}{u'_k}\right)$	Petzval curvature contribution	(3-27)
ΣDC	$= [B_{pi} i_p + (u'^2_p - u_p^2)/2]h'_k$	Distortion contribution	(3-28)
$\Sigma LchC$	$= yi(\Delta n - N\Delta n')/(u'_k)^2$	Longitudinal chromatic contribution	(3-29)
$\Sigma TchC$	$= yi_p(\Delta n - N\Delta n')/u'_k$	Transverse chromatic contribution	(3-30)

where $\Delta n = n_s - n_l$; and n_s, n_l are indexes for two wavelengths spanning the spectral range in which system operates.

These formulas were developed particularly for axis-centered systems of spherical surfaces, to which class the vast majority of practical optical systems belong. However, with some extensions to the set, it is possible to analyze by this technique optical systems which include aspheric surfaces of revolution (pp. 274-278, Ref. 27).

The Seidel Sums describe the aberrations as they exist at a maximum image height h_k corresponding to the intercept with the final image plane of the principal or chief ray introduced at y_{pi}. It is not valid simply to change h_k in the formulas to determine aberrations at intermediate heights. Instead, a complete new ray trace with an intermediate value of y_{pi} must be made and a complete new set of parameters determined from these.

3-2.3.5 Ray Tracing

3-2.3.5.1 Use of Computers for Ray Tracing

In modern technology, it is no longer economically sound to produce optical designs by so-called "longhand calculation", using desk calculators and slide rules. In addition to making possible the high-speed processing of established ray tracing formulas and aberration analysis, large-scale digital computers increase the scope of the analysis by multiple and iterative procedures.

For example, the precise nature and extent of aberrations can be determined by exact tracing of a large number of rays in the aperture and covering the entire field to provide a comprehensive account of the system performance. Computers will even generate a display of the data in the form of easily readable plots (spot diagrams) of the ray intercepts at the image plane.

A typical computer program will accept the basic geometrical and physical data pertinent to an optical system and produce the following performance data.

1. Detailed traces of a prescribed number of rays in the entrance pupil and in the field. These include six ray coordinates (three positional and three directional) at each surface and at the image, as well as at a number of arbitrarily designated stations.
2. Detailed listing of Seidel aberration values and sums
3. Overall cardinal points, plus effective focal length
4. Spot diagrams, as many and as detailed as required
5. Modulation transfer function data
6. Energy distribution plots

Sophisticated programs are available which permit a more rigorous analysis by automatically adjusting the input data to optimize the design. This involves the assignment and use of special merit criteria and/or logical selection programs, which may vary widely and depend heavily for their effectiveness on the ingenuity of the program designer.

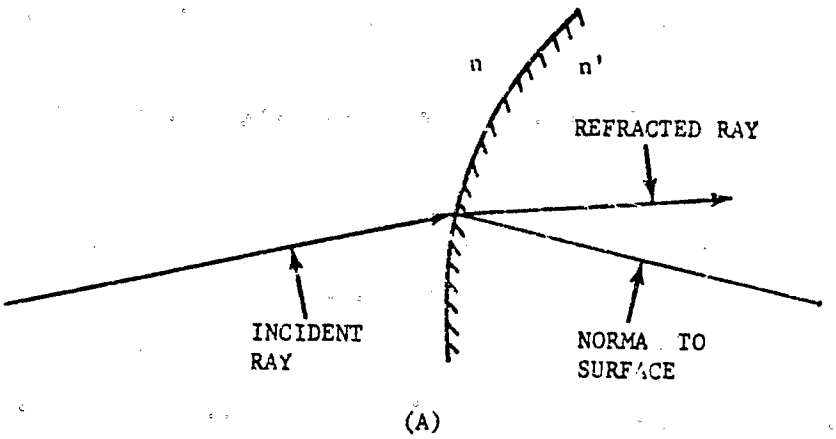
It is obvious that the optical designer no longer needs to have an understanding of detailed ray tracing techniques. Instead, he needs only to learn the relatively simple program input routine to become adequately proficient in conventional optical design.

For a detailed treatment of ray tracing, reference is made to some of the many texts available on this subject[24,25].

3-2.3.5.2 Graphical Ray Tracing

Although large-scale computers are extremely useful for analyzing optical systems which have been tentatively established, they cannot yet be used effectively in the initial synthesis of an optical design. At this stage, the designer's experience and ingenuity along with a basic understanding of the laws of optics are predominant.

System synthesis can often be expedited by graphical ray tracing techniques which help the designer visualize the projected performance of an embryonic optical system. Several such techniques are illustrated in Figs. 3-28, 3-29, and 3-30.



(A)

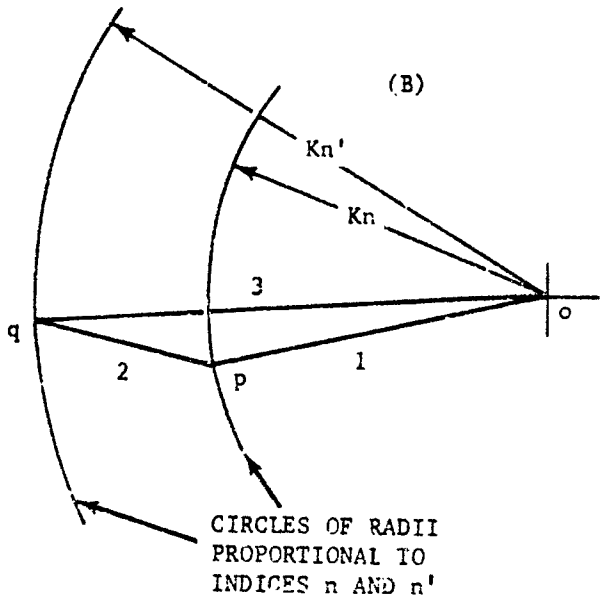
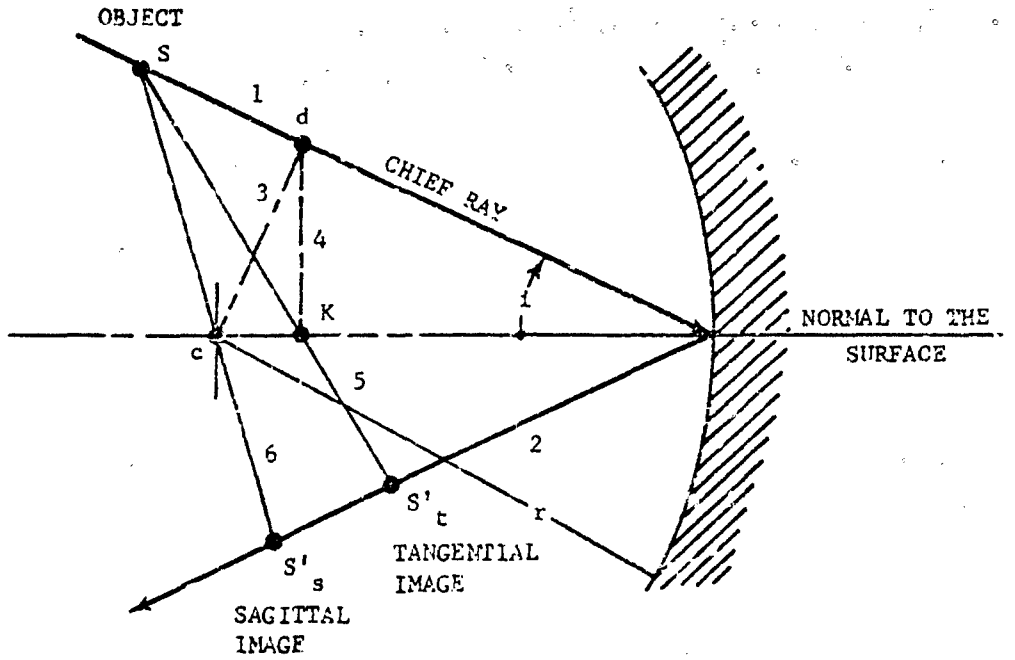


FIGURE 3-28. Graphical Ray Trace

LOCATION OF TANGENTIAL AND SAGITTAL IMAGE POINTS ALONG THE CHIEF RAY OF AN INFINITELY NARROW BUNDLE IN A REFLECTIVE SYSTEM.



PROCEDURE:

1. 2. DRAW THE CHIEF RAY ACCORDING TO THE LAW OF REFLECTION.

3. ERECT A PERPENDICULAR TO THE INCIDENT CHIEF RAY THRU c.

4. ERECT A PERPENDICULAR TO THE SURFACE NORMAL THRU d.

5. A STRAIGHT LINE THRU S AND K INTERSECTS THE REFLECTED CHIEF RAY AT S'_t.

6. A STRAIGHT LINE THRU S AND c INTERSECTS THE REFLECTED CHIEF RAY AT S'_s.

MATHEMATICALLY S'_t AND S'_s CAN BE LOCATED WITH THE FOLLOWING EQUATIONS:

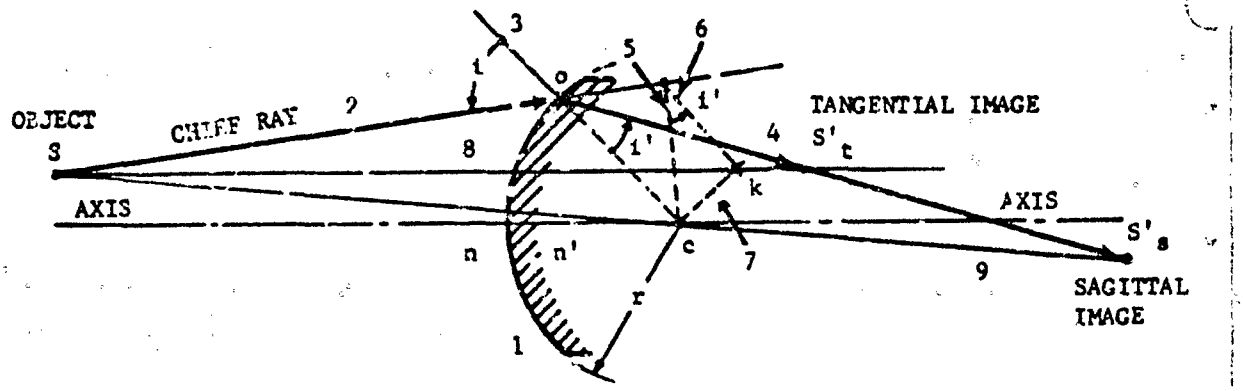
FOR TANGENTIAL IMAGE:

$$\frac{1}{S} + \frac{1}{S'_t} = \frac{2}{r \cos i}$$

FOR SAGITTAL IMAGE:

$$\frac{1}{S} + \frac{1}{S'_s} = \frac{2 \cos i}{r}$$

FIGURE 3-29. Graphical Ray Tracing. Surface May Be Convex, Flat, or Concave



CONSTRUCTION OF TANGENTIAL AND SAGITTAL IMAGE POINTS ALONG THE CHIEF RAY OF AN INFINITELY NARROW BUNDLE IN A REFRACTIVE SYSTEM.

PROCEDURE:

1 2 3 4	DRAW THE SURFACE AND TRACE THE CHIEF RAY ACCORDING TO FIG. 3-28
5	THROUGH THE CENTER OF CURVATURE c ERECT A PERPENDICULAR TO THE INCIDENT CHIEF RAY.
c	DRAW (6) AT AN ANGLE TO (5) EQUAL TO i'.
7	DRAW (7) PERPENDICULAR TO (6) THROUGH c TO FORM THE INTERSECTION k .
8	A STRAIGHT LINE THROUGH S AND k WILL INTERSECT THE REFRACTED CHIEF RAY AT S'_t.
9	A STRAIGHT LINE THROUGH S AND c WILL INTERSECT THE REFRACTED CHIEF RAY AT S'_s.

IN MULTIPLE SURFACE SYSTEMS, IF THE INTERMEDIATE S'_t AND S'_s POINTS TEND TO BECOME REMOTE AND INACCESSIBLE, THE PROCESS OF LOCATING THEM CAN BE CARRIED ON MATHEMATICALLY BY USING THE FOLLOWING EQUATIONS:

FOR THE TANGENTIAL IMAGE:

$$\frac{n' \cos^2 i'}{S'_t} - \frac{n \cos^2 i}{S} = \frac{n' \cos i' - n \cos i}{r}$$

FOR THE SAGITTAL IMAGE:

$$\frac{n'}{S'_s} - \frac{n}{S} = \frac{n' \cos i' - n \cos i}{r}$$

FIGURE 3-30. Graphical Ray Tracing. Surface May Be Concave, Flat, or Convex

Procedure:

1. Draw the normal to the surface through the point of incidence.
2. Draw an auxiliary diagram (B) consisting of two concentric circles of radii proportional to the indices n and n'.
3. Draw (1) parallel to the incident ray through the center o.
4. Draw (2) parallel to the normal through the intersection p.
5. Draw (3) through the intersection q and the center o.
6. Draw the refracted ray in (A) parallel to (3) through the point of incidence.

3-2.3.6 Image Quality in Terms of Resolution, Spot Size, and Energy Distribution

The quality of an optical system is usually a measure of its ability to produce an optimum point image from a point object, the optimum image being defined by the diffraction effect limitation which the system exhibits.

3-2.3.6.1 The Diffraction Effect–Airy Disc

The diffraction effect is an interference phenomenon which results in an energy distribution pattern at the image. The pattern consists of an intense central disc surrounded by alternate rings or zones of progressively lesser energy and zones of zero energy.

In a geometrically perfect system (no aberrations), the central disc (Airy disc) contains 84 percent of the total energy arriving at the image and its diameter is considered, therefore, as the nominal effective diameter of the image.

The size of the Airy disc depends only on the wavelength and on the diameter of the aperture and is expressed in terms of angular units (radians). The radius θ of the Airy disc is calculated using the following equation:

$$\theta = 1.22\left(\frac{\lambda}{a}\right) \qquad (3\text{-}31)$$

where λ = wavelength, a = aperture diameter; both in the same length units

Determination of the diffraction effect becomes rapidly more complicated when the system aperture is not circular. For most systems, however, the circular equivalent of the given aperture may be used to estimate first-order diffraction effects.

3-2.3.6.2 Geometrical Effects–Image Blur

The combined geometrical aberrations of a system produce a single effect termed image blur. Energy within this blur is usually distributed in a characteristic pattern consisting of a high-level concentration in a local spot with a rapid, and not necessarily uniform, fall-off in all directions. The lower level outer regions of this blur may extend a considerable distance away from the high-level region, often making it difficult to define the effective size of the blur. One arbitrary measurement assumes that the effective size is a circular area which contains 84 percent of the energy arriving at the image, similar to the measure of the diffraction effect by the size of the Airy disc, except that there is no well-defined edge to the central "hot spot".

3-2.3.6.3 Overall Effect

Mathematically, the geometrical and diffraction effects can be combined by convolution to produce an *overall* image blur. This complex mathematical operation, ordinarily done on a high-speed computer, is economically feasible only on the most precise optical system designs. For ordinary systems, it is customary to simply add the diameters of the two blur circles and assume this sum to be the nominal diameter of the overall blur circle.

The performance of an optical system is described in terms of one or more criteria based on the characteristics of the overall image blur. Most common criterion is resolution, others are energy distribution and frequency response.

3-2.3.6.4 Resolution

The resolving power of a system is defined as the smallest separation between two point-images at which the system will detect that there are two images instead of one.

A precise geometrical definition of this condition for diffraction-limited systems (systems with insignificant geometrical aberrations) is given by the so-called Rayleigh Criterion. This

criterion states that two equal diffraction-type point-images shall be considered resolved if their Airy discs overlap to the extent of not more than one disc radius. Thus, the energy peaks of the two images are separated by at least one disc radius.

The energy distribution of most point images is characteristically similar to that of the typical diffraction image. This had led to the adoption of the Rayleigh Criterion for application to all types of images, provided only that an equivalent disc diameter can be assigned thereto.

3-2.3.6.5 Geometrical Energy Distribution—Spot Diagrams

Energy distribution in a point image can be most effectively displayed by means of a spot diagram. A representative number of rays evenly spaced in the entrance pupil of the system are traced to the image plane and plotted as a grouping of dots. The density of the dot groupings will be a measure of the spatially distributed energy. A count of the dots and application of a percentage specification can provide an estimate of the nominal blur circle diameter, which can then be used as a performance criterion.

3-2.3.6.6 Energy Distribution—Spread Function

The distribution of energy in the image of a point may also be described by a "point spread function" which is a plot of the energy density versus position in the image plane. If the distribution is symmetrical, a plot of the energy density along a line through the center of the image contains all the information. If the distribution is unsymmetrical, plots along two lines at right angles to each other are used.

3-2.3.6.7 Frequency Response

The frequency response can be determined by using a series of gratings covering a range of frequencies from that at which the response (image) shows 100 percent contrast to that at which the contrast is effectively zero. This function may be used directly as a criterion of the optical system performance.

A sine wave grating consists of uniformly spaced lines and spaces with a density which is graded from one to the next according to a sine function. When such a grating is used as an object in an optical system, an image will be formed which is similar to the sine wave grating but at a reduced contrast (sine wave amplitude). The contrast may be measured with a radiometer, or with a microdensitometer if a photographic image is produced.

If a grating of higher spatial frequency (closer line spacing) is substituted, a lower contrast level will be observed.

3-2.3.6.8 Image Evaluation

When the geometrical blur of an image is small compared to that due to diffraction, an optical system is said to be diffraction limited. Such optics represent the highest quality that can be achieved.

When the geometrical and diffraction blurs are in the same dimensional range, their individual effects are often difficult to observe separately. In fact, the geometrical aberrations do not substantially increase the size of the Airy disc, instead they tend to feed energy into the outer rings of the diffraction blur. Because of this effect, an optical system in this condition is still considered to be diffraction limited. Its resolution, by definition, is unchanged, but the contrast in the image is reduced over that of a true diffraction-limited system.

A third level of image quality is that in which the geometrical aberrations predominate. In evaluating the system, it is customary to ignore the diffraction effect since the need for precision is lost among the gross system inaccuracies.

3-2.3.6.9 Depth of Focus

A perfectly corrected lens system will have a certain range called "depth of focus" throughout which the image plane may be moved without deteriorating the image beyond the Rayleigh limit.

$$\text{Depth of focus} = \pm \frac{\lambda}{n' \sin^2 u} \qquad (3\text{-}32)$$

where

λ = wavelength
u = angle of marginal ray with the axis at the image
n' = index of medium in which image exists

Example:

A camera having a 2-in. aperture objective, and focused at an object 10-ft (120-in.) distant has a depth or focus:

$$\text{Depth of Focus} = \pm \frac{0.000022}{1 \times \left(\frac{1}{120}\right)^2}$$

$$= \pm 0.32 \text{ in.}$$

Since $\lambda = 0.000022$ in. for average white light

$n' = 1$ for air

$$u = \frac{\frac{1}{2} \text{ of } 2}{120}$$

$$= \frac{1}{120} \text{ rad}$$

$$\sin u = \frac{1}{120}$$

3-2.4 OPTICAL SYSTEM DESCRIPTIONS AND ENGINEERING

3-2.4.1 Afocal Systems

An afocal optical system is one wherein the net or overall focal length is equal to infinity. Both object and image are also at infinity (see Fig. 3-31). The focal lengths of the individual components, however, are finite and the system analysis, by use of the Gaussian equations on the components, is valid.

The system has a characteristic magnification which can be derived simply by cascading the magnifications of each component system, thus

$$m = m_1 m_2 = \frac{s_1'}{s_1} \times \frac{s_2'}{s_2} \qquad (3\text{-}33)$$

but $\lim \frac{s_2'}{s_1} = 1$ as s_2' and s_1 approach infinity

$\therefore m = s_1'/s_2 = f_1/f_2$

The value of afocal optics in IR systems is in reducing the bundle size so that a large aperture may be used with relatively smaller optics for processing the beam to the detector. The reduction in bundle size, however, is achieved only at the expense of the field width.

3-2.4.2 Relay and Field Lenses

In transporting an image over a long, narrow optical path, as occurs in a periscope, it is necessary to arrange a number of short systems in series in such a way that the image of one acts as object for the next one in succession. The greater the width of bundle to be carried over a given total path, the greater will be the number of component relays needed.

A typical relay system is illustrated in Fig. 3-32. The image is reformed three times. At each intermediate position, a field lens acts to change the direction of the image forming bundles so that they will remain within the span of the next relay lens. This is the characteristic action and function of the field lens. It does not change the image itself.

A field lens is always between two relay lenses or between apertures of a system. It is, therefore, very useful to consider the field lens as a component which relays an image of the aperture from one station to the next.

In IR systems, this aperture relaying function is used to define the field and stabilize the energy level at the detector as demonstrated in Fig. 3-33. In such applications the field lens is designed so that it will also perform the demagnification function. This will result in a higher concentration of energy on the detector while permitting the use of a lower power and better corrected objective.

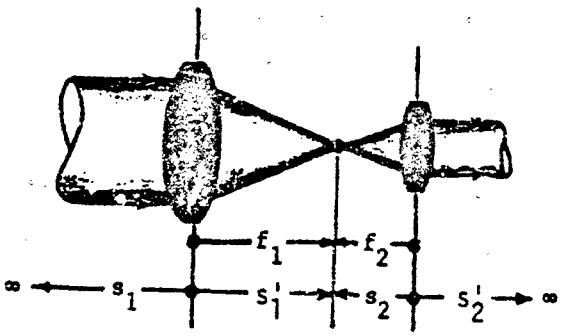
3-2.4.3 Aplanatic Systems

An optical system which produces no spherical aberration or coma in the image is said to be aplanatic. One outstanding example is a single- or double-surface refractive system wherein the object and images are as shown in Fig. 3-34. The limitations imposed by the given conjugate relationships prevent such a system from being used to form a real image from a real object, but it is very effective for increasing or decreasing powers or the field angles without adding aberrations.

3-2.4.4 Symmetrical Combinations

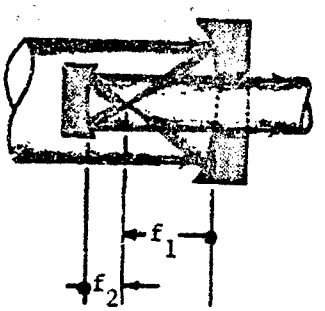
An example of a symmetrical system often used in relaying is shown in Fig. 3-35. The system is particularly useful in illustrating the techniques of adjusting the surface curvatures and stop position to minimize aberrations.

(A)

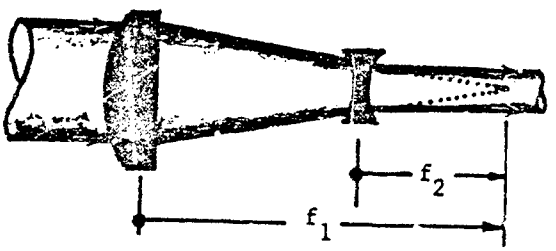


ASTRONOMICAL TELESCOPE

(C)

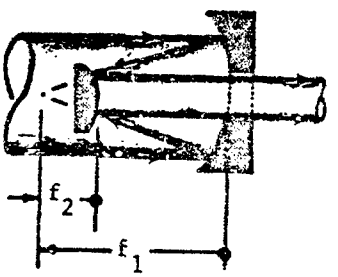


(B)



GALLILEAN TELESCOPE

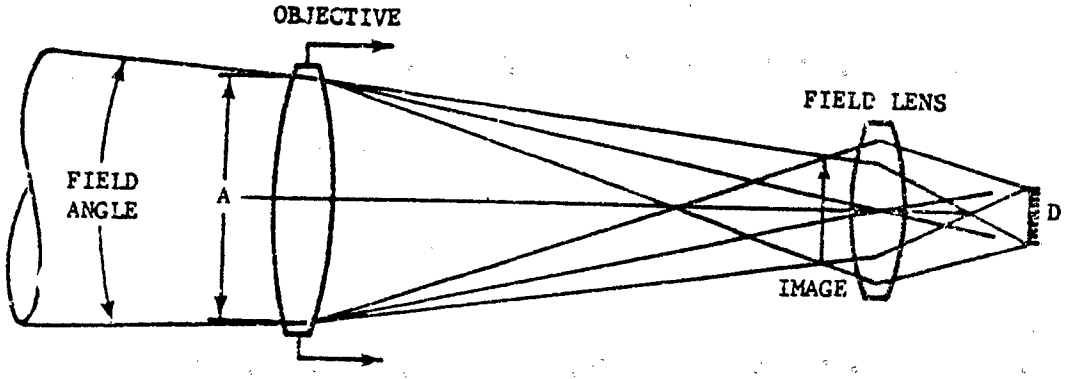
(D)



(A), (B) REFRACTIVE TYPES

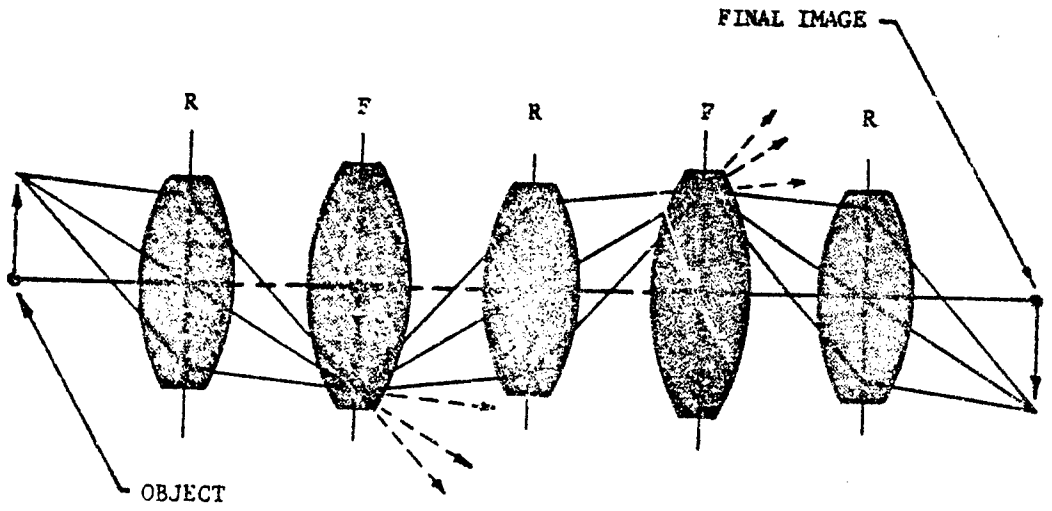
(C), (D) REFLECTIVE TYPES

FIGURE 3-31. Afocal Systems



THE OBJECTIVE LENS APERTURE (A) IS IMAGED BY THE FIELD LENS ON THE DETECTOR (D). ALL ENERGY WITHIN THE CONSTANT FIELD ANGLE IS DELIVERED TO THE DETECTOR.

FIGURE 3-32. Optical Relay System

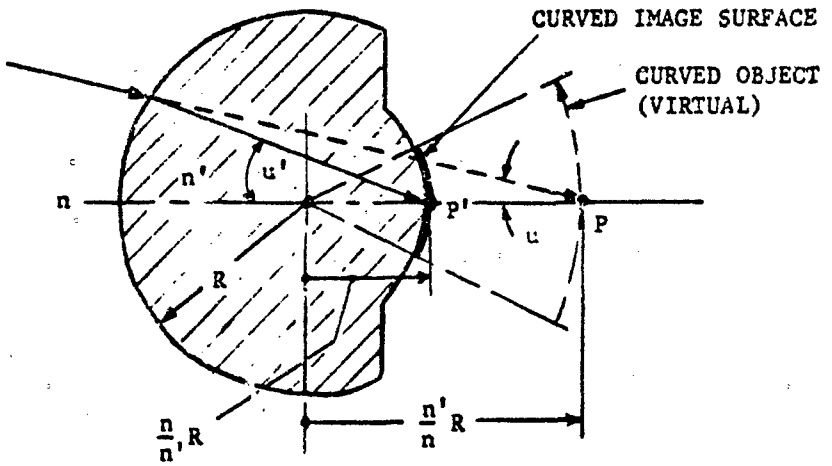


R = RELAY LENS

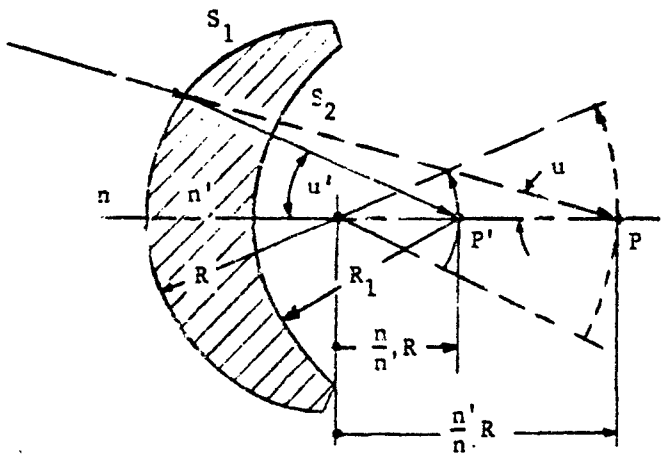
F = FIELD LENS

NOTE: Field lenses change direction of the ray bundles, but not the character of the images. Each positive relay unit inverts the image.

FIGURE 3-33. Example of Simple IR Radiometer

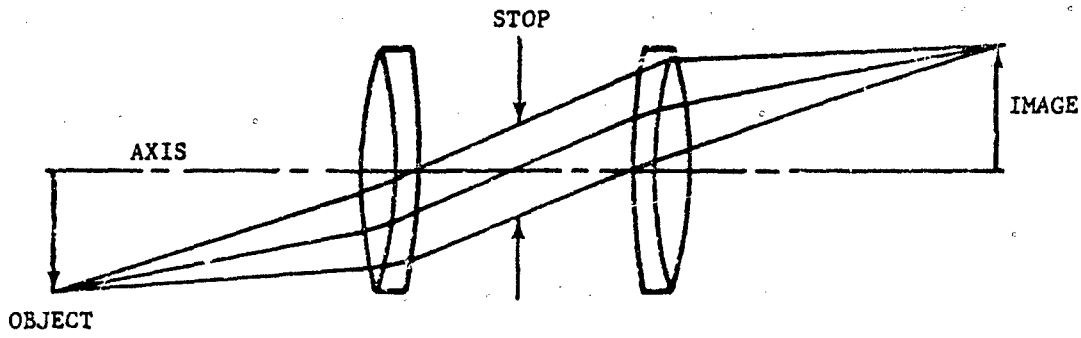


PERFECT APLANATIC SYSTEM. ZERO SPHERICAL ABERRATION. ZERO COMA. GEOMETRICALLY, THE SINE CONDITION, n sin u = n' sin u', IS SATISFIED FOR ALL RAYS THROUGH THE SYSTEM.



IMPERFECT APLANATIC SYSTEM. THE AXIAL POINT (P') IS FREE OF ABERRATION, BUT THE SECOND SURFACE S_2 INTRODUCES A SMALL AMOUNT OF COMA INTO OFF-AXIS IMAGE POINTS.

FIGURE 3-34. Geometry of an Aplanatic System



COMA, DISTORTION, AND LATERAL CHROMATIC ABERRATION, ARE CANCELLED BY SYMMETRICAL ARRANGEMENT OF ELEMENTS AROUND THE STOP

FIGURE 3-35. Symmetrical Relay System

As an initial requirement, each half of the system is designed for a minimum spherical aberration, but with a sizable coma. Ordinarily this cannot be done with a single lens, while it is feasible with a doublet or an aspheric lens. When the two halves are assembled, the lateral aberrations of coma, distortion, and lateral color are cancelled due to symmetry. The spacing of the two halves, which is equivalent to adjusting the stop position relative to each, can then be used to control astigmatism and field curvature. Spherical aberration will remain low since it is not affected by stop position.

The symmetry principle is also used quite effectively in the design of wide-angle systems.

3-2.4.5 Multiple-element Systems

A single lens element (two spherical surfaces) will always produce positive spherical aberration in the image when the object is located at infinity. By bending the lens (Fig. 3-36) so that the principal plane approaches a sphere with its center at the axial image point, the spherical aberration can be reduced to a minimum. If two or more such elements are placed in series, it might be concluded from a cursory examination of the arrangement that the positive values of spherical aberration will add to produce a greater overall effect. Quite the contrary is true. Instead, the contribution of the second and succeeding elements will drastically reduce the overall spherical aberration and, in fact, by the use of materials of suitable index, it may be entirely eliminated.

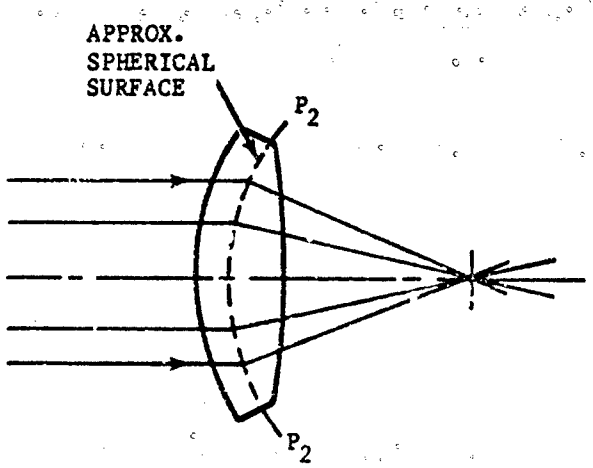
The reason for this seeming anomaly is that the spherical aberration is not a function of the optical element alone, but of the optical system, which includes object and image position. Thus, since the object which the second and succeeding elements in the series see is not at infinity, their contribution to the system aberration is drastically altered.

3-2.4.6 Reflective IR Optical Systems

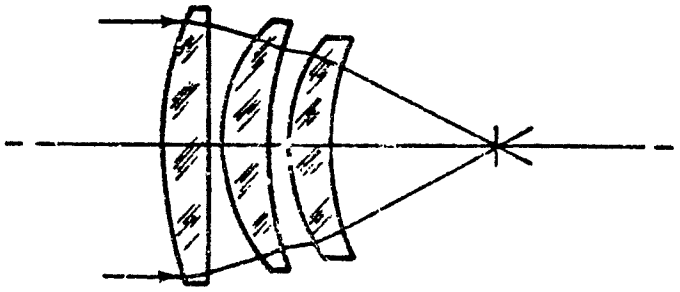
Reflecting systems are highly favored in IR technology because they perform uniformly over a wide spectral range, produce no chromatic aberrations, and are economical even in very large sizes. The most common of all-reflective optical systems configurations are shown in Fig. 3-37.

3-2.4.6.1 Simple Mirror

A simple mirror may be either spherical or paraboloidal. If the stop is placed at the center of a curvature, the sphere will introduce only spherical aberration which, nevertheless, is uniform over the curved image field. Since the image is formed within the beam, a detector array placed to sense it will produce a certain



A SINGLE REFRACTIVE ELEMENT CAN BE "BENT" FOR MINIMUM SPHERICAL ABERRATION. FOR AN OBJECT AT INFINITY, THE PRINCIPAL "PLANE" WILL APPROACH A SPHERICAL SURFACE CORRESPONDING TO THE SPHERICAL WAVE FRONT OF THE CONVERGING BEAM.



MULTIPLE REFFACTIVE ELEMENTS CAN BE COMBINED TO PRODUCE ZERO SPHERICAL ABERRATION SIMULTANEOUSLY WITH A LOW RESIDUAL COMA VALUE.

FIGURE 3-36. Reduction of Spherical Aberration in Series Lens Combinations

amount of blocking, which is detrimental only to the extent that it reduces the amount of energy which can be collected with an aperture of a given size.

A paraboloid will produce an aberrationless image point on-axis, but off-axis it introduces astigmatism and coma. For narrow field angles, these are often small enough to be tolerated.

3-2.4.5.2 Folded System

If the detector array requires substantial service support such as cooling or baffling and possibly access space, one of the folded configurations illustrated in Fig. 3-37 (B) and (D) might be suitable. The image characteristics are unchanged, but blocking of the beam is likely to be increased.

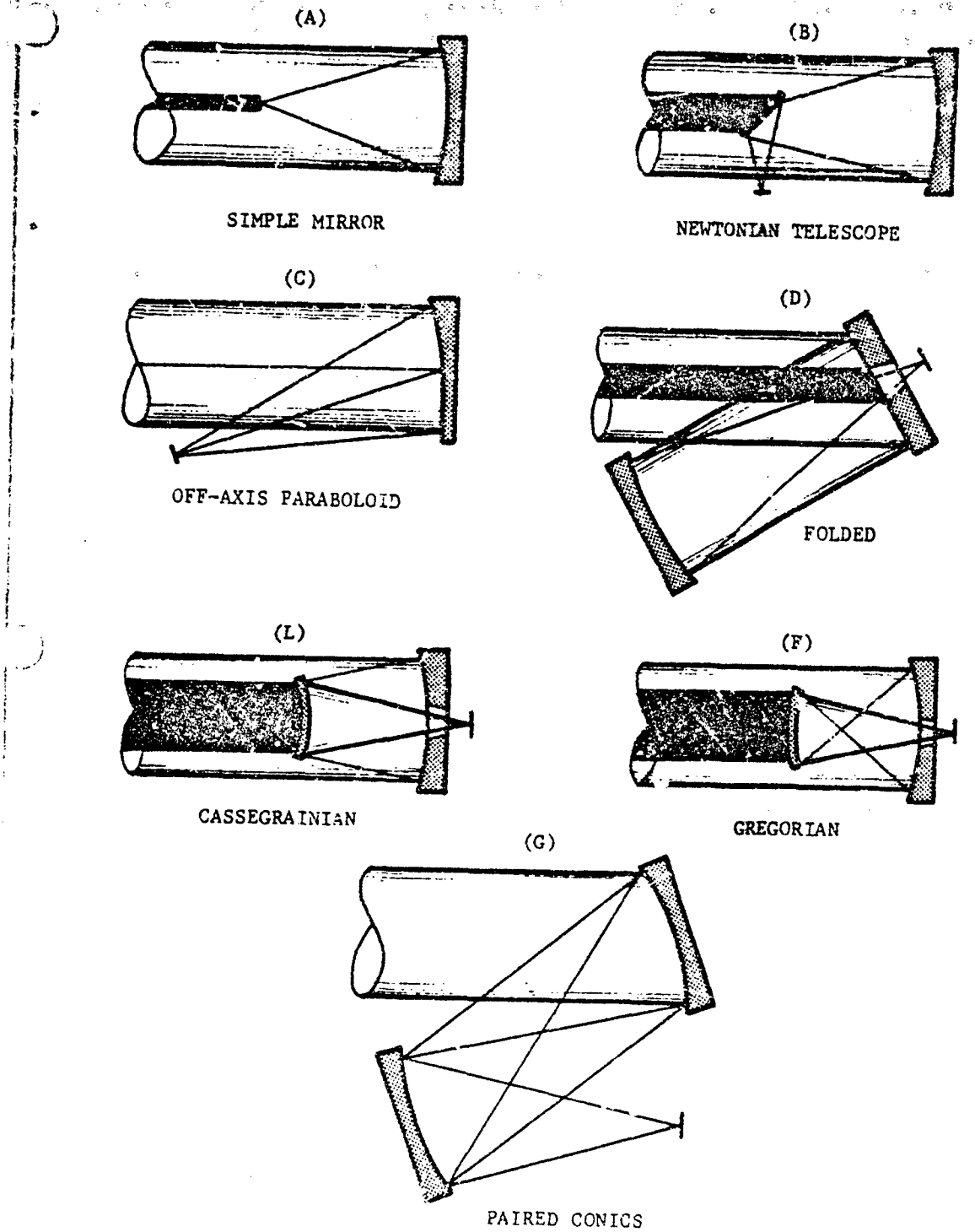


FIGURE 3-37. Reflective Systems

3-2.4.6.3 Off-axis Configuration

The off-axis paraboloid represents an elegant solution to the blocking problem, but its field is more severely limited by coma and astigmatism than it is in concentric systems. The axial image point, however, is perfect.

3-2.4.6.4 Compound Reflectors

The Cassagrainian and Gregorian systems (E) and (F) of Fig. 3-37, respectively, are essentially the same except in the placement of the secondary. Their advantages include compactness and a moderately wide field. Systems of this type result in considerable blocking of the beam.

3-2.4.6.5 Paired Conics

Systems of this type are rarely used for image-forming optics, although the dual surfaces and spacing permit the design to be optimized over a substantial field. Fabrication techniques for conic surfaces of this type have not yet been developed to the point where they might be considered economically competitive with the axis centered surfaces. This configuration is very well suited for condenser systems used for detecting very long wavelength radiation.

3-2.4.7 Catadioptric Systems

The catadioptric system (Fig. 3-38) is a combination of both reflective and refractive elements. Compared to all-reflective systems, they show superior image forming characteristics over a wide field and at high speed.

The most common types use a large aperture spherical reflector and modify its characteristics by means of one or more refractive "correcting" elements. The three distinct correcting methods which follow have been developed.

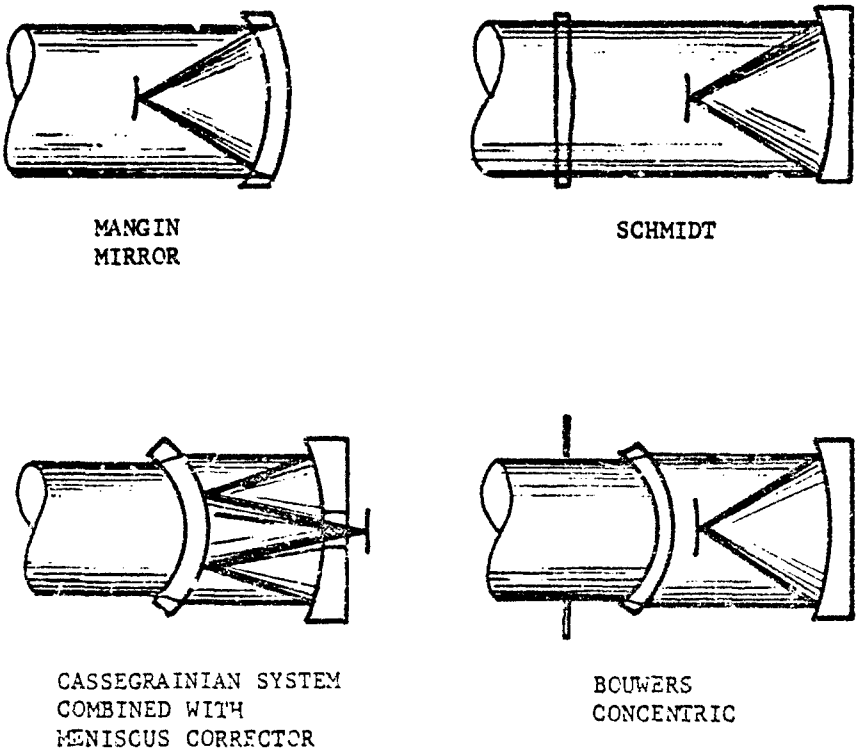


FIGURE 3-39. Catadioptric Systems

3-2.4.7.1 Mangin Mirror

A Mangin mirror has a refracting element in direct contact with the spherical mirror. In fact, the system is usually made as a single optical unit where the back (second) surface acts as the spherical reflector and the front surface of different curvature acts as the corrector. The front surface changes the ray paths twice. The radius of the front surface can be selected so that spherical aberration and coma are corrected over a small field.

3-2.4.7.2 Schmidt System

The Schmidt system makes use of an aspheric refractive corrector at the center of curvature of the spherical mirror and its stop is located at the same position.

A spherical mirror alone will produce only spherical aberration for any bundle through its center of curvature. The Schmidt corrector effectively reduces spherical aberration without introducing any sizable amounts of other aberrations.

A Schmidt system can be readily designed to perform admirably over a 25-deg wide field at a speed of f/1.0.

3-2.4.7.3 Maksutov-Bouwers System

The Maksutov system makes use of a meniscus lens of low negative power at or near the center of curvature of the spherical mirror to correct spherical aberration. It accomplishes basically the same result as the Schmidt corrector.

A special form of the Maksutov system is the Bouwers concentric, wherein the corrector is designed as two concentric surfaces, and the centers of curvature are coincident with the center of the primary spherical mirror. This makes the system uniform in its performance over an extremely wide field, although there exists a small residual spherical aberration plus a small amount of chromatic aberration.

3-2.4.7.4 Hybrid Configurations

From the three basic configurations described, a large number of more sophisticated systems can be assembled. For example:

1. A Schmidt corrector can be combined with a concentric corrector to produce a refinement of the basic Bouwers.

2. Either the concentric or Schmidt corrector can be made as a doublet to rid the system of the small chromaticism.

3. A Mangin mirror can be used as a secondary in an overall Cassegrainian system.

4. Mangin mirrors can be used as correctors in Maksutov systems.

5. A so-called Super-Schmidt has an achromatic plate corrector, a front concentric corrector, and a rear concentric corrector.

All the wide-angle systems discussed produce an image which is curved around the center of the stop. In IR systems this is of no great consequence since the detector array can be easily curved to match.

3-2.4.8 Rapid Estimation of Blur Size

All the basic optical system configurations described in the preceding paragraphs have, of course, been ray-traced many times and their characteristic aberrations have been reduced to a set of simplified rules which are often very handy for the optical designer in assembling new systems.

3-2.4.8.1 Spherical Mirror, On-axis

With the stop at the center of curvature, a spherical mirror produces only the following spherical aberration:

$$\beta_{\text{angular blur}} = \frac{0.0078}{(\text{f/no.})^3} \text{ , rad} \qquad (3\text{-}34)$$

for intermediate to high f/no.

$$\beta_{\text{angular blur}} = \frac{0.0091}{(\text{f/no.})^3} \text{ , rad}$$

for low f/no., e.g., f/1

With the stop at the mirror itself, coma is present to the extent

$$\beta_{coma_s} = \frac{0.0625\,\theta}{(\text{f/no.})^2} \text{ , rad} \qquad (3\text{-}35)$$

$$\theta = \frac{1}{2} \text{ field angle}$$

and

$$\beta_{astig} = \frac{0.5\,\theta^2}{(\text{f/no.})} \text{ , rad} \qquad (3\text{-}36)$$

A typical mirror (10-in. *fl* - f/2) exhibits a spherical aberration of $0.0078/(2)^3 = 0.001$ rad (Eq. 3-34). With the stop (4-in. CA) at the mirror, it produces (Eq. 3-35)

$$\beta_{coma_s} = 0.0625\left(\frac{\frac{10}{4}}{10}\right) / (2)^2$$

$$= 0.0039 \text{ rad at edge of field.}$$

$$\text{Astigmatism} = 0.5\left[\left(\frac{\frac{10}{4}}{10}\right)^2 / (2)^2\right]$$

$$= 0.0078 \text{ rad at edge of field (Eq. 3-36).}$$

$$\text{Total blur} = \text{sum, approximately}$$

$$= 0.001 + 0.0039 + 0.0078$$

$$= 0.0127 \text{ rad}$$

3-2.4.8.2 Spherical Mirror, Off-axis

It should be noted that the rays which define the spherical aberration blur never cross the axis. Therefore, if a concentric system is cut in half through the axis, the size of the blur is also cut in half in the direction transverse to the cut. If the half-system is again cut in half, each unit is only one-fourth the original size and the aberration is proportionally reduced. Each quarter is, by itself, an off-axis system. By utilizing this sectioning principle, an estimate of the aberrations in the off-axis system may be derived from the equations for the concentric system.

3-2.4.8.3 Paraboloidal Mirror

The paraboloidal mirror has one unique image point on-axis which is geometrically perfect.

For off-axis image points with the stop at the mirror, aberration blurs are essentially the same as for an equivalent spherical mirror, and the same equations apply.

With the stop placed at a point one focal-distance away from the mirror, the astigmatism for off-axis image points can be reduced to a relatively small value, but the coma is not changed.

3-2.4.8.4 Off-axis Paraboloidal Mirror

The aberrations of an off-axis paraboloidal mirror are difficult to estimate because of the many variables involved. The most expedient means of obtaining the desired performance data is computer ray tracing.

3-2.4.8.5 Schmidt System

The overall off-axis image blur for a Schmidt system is given by the equation

$$\beta_{Schmidt} = \frac{0.0417\,\theta^2}{(\text{f/no.})^3}, \text{ rad} \qquad (3\text{-}37)$$

$$\theta = \frac{1}{2} \text{ field angle}$$

Comparison of this equation with the equation of a spherical mirror (Eq. 3-36) will indicate that an improvement is obtained up to a width where

$$\theta^2 = \frac{0.0078}{0.0417} \text{ or}$$

$$\theta = 0.43 \text{ rad}$$

$$= 24.7 \text{ deg}$$

At $\theta = 12.5$ deg, 25 deg full field, the Schmidt system produces a blur only one-fourth that of the equivalent spherical mirror.

3-2.4.8.6 Bouwers Concentric Design

The addition of the concentric meniscus corrector to a simple spherical mirror reduces the spherical aberration of the system by about two orders of magnitude. There is some variability, depending on corrector thickness and placement.

3-2.4.8.7 Single Refractive Element

An element "bent" for minimum spherical aberration is used as a standard reference.

$$\beta_{spherical\ aberration} = \frac{K}{(\text{f/no.})^3}, \text{ rad} \qquad (3\text{-}38)$$

$$K = \begin{array}{ll} 0.067 \text{ for } n = & 1.5 \\ 0.027 & 2.0 \\ 0.0129 & 3.0 \\ 0.0087 & 4.0 \end{array}$$

$$\beta_{coma_s} = \frac{0.0625\,\theta}{(n+2)(\text{f/no.})^2}, \text{ rad} \qquad (3\text{-}39)$$

$$\theta = \frac{1}{2} \text{ field angle}$$

$$n = \text{index of refraction}$$

$$\beta_{astig} = \frac{0.5\,\theta^2}{\text{f(no.)}}, \text{ rad} \qquad (3\text{-}40)$$

$$\theta = \frac{1}{2} \text{ field angle}$$

Note the similarity of these formulas to those for spherical mirrors.

Chromatic Aberration $= \rho \equiv \frac{1}{2\nu(\text{f/no.})}$, rad (3-41)

ν = reciprocal relative to dispersion of the material.

Mangin Mirror:

Zonal Spherical $\beta = 0.00025/(\text{f/no.})^4$, rad (3-42)

Sagittal Coma $\beta = 0.031\,\theta/(\text{f/no.})^2$, rad (3-43)

Astigmatism $\beta = 0.5\,\theta^2/(\text{f/no.})$, rad (3-44)

Chromatic $\beta = \frac{1}{6\nu(\text{f/no.})}$, rad (3-45)

(All formulas for estimation of blur sizes according to Ref. 27.)

3-2.4.9 Mechanical Stability of Large Optical Systems

Weight is the greatest degrading factor of stability in large optics. Unless adequately supported, a large element will not maintain its configuration through required position changes, even from test stand to operating installation. A much-used rule of thumb for proportioning such elements dictates that the thickness shall be 1/6 to 1/4 of the diameter; even so, a support structure of at least equal rigidity must be provided.

In recent years, various kinds of lightweight structures for mirror blanks have been developed in honeycomb, egg crate, ribbing, or lightening hole configurations. The material found most adaptable to this kind of treatment is fused quartz.

Vibration also degrades the performance of large optical systems. It is often necessary to provide large seismic beds isolated from nearby foundations by means of springs or soft energy absorbing material. Many specialized vibration damping devices and mounts are available on the market.

Mechanical strain on optical elements is often avoided by the established technique of "three-point suspension", whereby any element is tied to a rigid structure by means of three localized constraints which, nevertheless, have full angular freedom (e.g., ball and socket). Another technique is the use of a sling mount which is nothing more than a strap around the bottom of a cylindrical element. Of course, this technique is not useble when the element is facing in any direction but the horizontal.

3-2.4.10 Thermal Stability

High-quality optical systems require careful attention to details in the design which involve changes due to temperature variation. Techniques used include the selection of low-coefficient materials, or materials of matched coefficients; use of massive elements; constraints with sufficient degrees of freedom for expansion; baffles or conduction paths for control of heat; and active heating or cooling of portions or all of a system. A detailed description of the various methods of achieving the required temperatures is given in par. 3-5.

3-2.4.11 Establishing Optical Tolerances

In setting up tolerance specifications for optical systems and components, guidelines are presented in the paragraphs which follow.

3-2.4.11.1 Surface Quality

Surface defects include scratches, digs, stains, grayness, or incomplete polish. The latter two can be eliminated from any particular surface simply through additional work. The first two are toleranced in terms of classification numbers such as 80-50, in which the number 80 represents a limiting width of scratch in microns, along with a limit on the total scratch lengths within a given area. The number 50 represents a limiting size of digs, pits, or bubbles in hundredths of a millimeter, along with a limit on the number of these in a given area.

Typical scratch and dig specifications of 80-50 are relatively easy to achieve; 50-30 requires a moderate amount of care in processing, and 20-10 or 10-5 represent the highest grades of surface finish. These grades are usually applied to field lenses or reticles where the surface is to be focused in the image plane.

Surface accuracy is defined as the degree of deviation of a given surface from the mathemat-

ically perfect surface specified. Spherical and flat surfaces are often measured by placing them in contact with a standard surface or test plates, and observing the interference fringes which form at the interface. Fringes are formed by wedges between the "contacting" surfaces at the rate of one fringe, bright-peak to bright-peak, per 1/2 wavelength (about 11 micro-inches) of separation.

Specification by means of interference fringes should include not only the spacing but also the uniformity of the spacing, since a given spacing represents only a wedge condition, while a variation in the spacing represents a difference in optical power of the two surfaces. A typical tolerance for a 10-in. diameter flat surface might read: "Accuracy: 4 fringes per 10 in.—fringe spacing variation 1/2 fringe".

In testing spherical surfaces, a difference in the radius between the test surface and the specimen surface will produce circular fringes. The tolerance limits on this difference may then be expressed as "(*n*) fringes of optical power". If the fringes are elliptical, the indicated astigmatism can be limited by stating tolerable optical power in two directions. Conventional shop practices can provide tolerances to about 1 fringe per inch of surface routinely, while tolerances to 1 fringe per 5 inches are not uncommon. Beyond this, great care and skill are required.

From a designer's viewpoint, the specification of surface accuracy is more rigorous and informative if given in terms of seconds of arc rather than by fringe count. In these terms, every point on the surface is specified with no lack of continuity. A specification of this type relates the direction of rays in the system to the surface normal vector, so that the effects of inaccuracies can be directly translated into image degradation. A relationship between angular measure and fringe count, therefore, deserves to be kept handy for translation of design tolerances into shop tolerances:

1 sec of arc = 1 fringe per 2 inches approximately

3-2.4.11.2 Thickness and Spacing

In most optical systems the image forming properties are relatively insensitive to element thickness and spacing. Consequently, tolerances may be relaxed to something which compares with ordinary machine operations: from ± 0.003 in. for precision work to ± 0.010 in. for routine production. With modern equipment such tolerances do not present fabrication problems.

3-2.4.11.3 Optomechanical Centering

Elements which comprise a single spherical or flat surface have no pre-established optical center or axis. Double-surface elements have an axis defined by a line which connects the centers of curvature. The objective of the centering operation is to make the mechanical axis (defined by the periphery of the element) coincide with the optical axis.

A conventional centering technique entails spinning the element on a nominal axis and directing a narrow beam through it. If the beam or the image it produces remains stationary, the spin axis and optical axis are then coincident. Only the grinding of a periphery on the element in the same setting remains to be done. Any degree of accuracy within machining tolerances of about 0.001 in. can be achieved. For extremely high accuracy a microscope may be used in observing the beam wobble. For mounting a centered element in its cell, a clearance of from 0.001 to 0.003 in. is conventional.

3-2.4.11.4 Prism Angles and Dimensions

Angular accuracy requirements for prisms vary widely according to their use, 1/2 sec of arc representing a practical limit. Tolerances on linear dimensions should be comparable to those of an equivalent machined part.

3-2.4.11.5 Materials

The material properties which most critically affect optical system performance include the transmission range, index of refraction, mechanical perfection, and stability.

On materials which have transmission bands that extend through both visible and IR regions such as quartz and special glasses, tolerance specifications on the properties indicated usually follow the standards set by MIL-G-174A. The specific tolerance values which follow have been abstracted from that document.

3-2.4.11.5.1 Index of Refraction

A standard tolerance of 0.001 is specified for values of index below 1.600 and ranging up to ± 0.002 at index 1.72 and above.

Crystalline materials have an index which is fixed by the molecular structure and, hence, need no tolerance specifications.

Although no standard tolerances have as yet been established for materials such as silicon and germanium, it should be expected that—since these are used at relatively long wavelengths—the accuracy requirement will be proportionately less.

3-2.4.11.5.2 Annealing

For time and temperature stability, large optics (usually reflective) require a high degree of freedom from residual stresses. In materials which are transparent in the visible region, this condition is usually measured in terms of "birefringence of the material", a difference in optical pathlength for light in two planes of polarization at right angles to each other.

The units of birefringence are millimicrons (nm) (of path difference) per centimeter (of pathlength). An average high quality anneal is 50 nm/cm, while the best which can be achieved (with fused quartz, for example) is about 10 nm/cm.

For the materials such as CER-VIT and large IRTRAN blanks, annealing specifications apparently do not apply.

3-2.4.11.5.3 Transmission Range

Transmission characteristics of IR materials vary widely and cannot be readily specified to tolerance. Where a specific value of transmission in a particular item is required, this might be obtained by selecting the proper material from a number of samples.

3-2.4.11.5.4 Imperfections

Bubbles, striae, sleeks, and inclusions are specified by the size and number which exist within a given volume, usually one cubic centimeter, of the material. The tolerance levels are dependent upon the intended use of the material. In windows and in components located at or near a system aperture, these imperfections will produce only a reduction in transmitted energy proportional to the percentage of the optical beam which they obstruct. As a general rule such obstruction should not exceed 1 percent and preferably should be held to 0.1 percent.

For use in field lenses, reticles, or other components near an image, the designer must decide the degree and type of obstruction that can be tolerated.

3-2.4.11.6 Summation of Tolerances

When optical components are used in series, the variance on the assembly will generally be greater than that of any individual component, but it will not be a direct arithmetic sum. The root-sum-square rule (based on statistical analysis) is used most frequently for such summation as follows

$$T = \sqrt{a_1^2 + a_2^2 + \ldots a_n^2}$$

where T is the tolerance value of the assembly and the a_i are tolerance values of the separate components.

Individual tolerances may thus be adjusted so that the total does not exceed a prescribed value.

3-2.5 TESTING OF OPTICAL SYSTEMS

Three categories of optical system and component tests are generally recognized:

1. Material Inspection Tests—Index of refraction, stress, and flaws
2. Calibration—Measurements of focal lengths, radii of curvature, flatness, magnification, reflectance, and transmittance
3. Image quality

3-2.5.1 Material Inspection Tests

3-2.5.1.1 Index of Refraction

The index of refraction of a material can be determined accurately by means of a specialized instrument, called a refractometer, designed for this purpose. Generally, a prism must be cut from the sample material and polished to a high degree of accuracy. The refractometer is then used to measure the deviation angle of a laser beam after it is passed through the specimen prism.

On large blanks, for large lenses, or windows, it is often important to determine the uniformity of the index of refraction throughout the blank. Both sides of the blank must be first ground and polished optically flat. The blank is then placed in a beam of collimated light and the emerging wave front is analyzed with an interferometer. Details on this type of measurement are to be found in Refs. 28 and 29.

For materials opaque in the visible region, a similar technique may be used by replacing the source beam with a beam of suitable wavelength and probing the interference field with a scanning detector.

3-2.5.1.2 Stresses and Strains

On blanks which are transparent in the visible region the strain may be measured with a polariscope, or alternately by placing the specimen between crossed polarizers in a beam of light and using a Babinet Compensator to measure the path difference between the two polarized components of the beam.

On opaque materials, this method of strain-stress measurement cannot be applied, of course. For these, a sample testing technique is about the only way in which the adequacy of an anneal can be determined. A small sample specimen is processed through the same fabrication and annealing process as the blank under examination. A trial flat surface is then cut and polished on the specimen which is constantly monitored during fabrication with a test flat. Undesirable movements of the surface will be observed as the polishing progresses and stresses are relieved. A piece free of stresses will not undergo such movement.

Temperature stability may be also observed by cycling the specimen through an appropriate range.

3-2.5.1.3 Imperfections

Visual examination, either direct or with the aid of mangifying devices, is usually the easiest way of determining the imperfections in a piece of transparent optical material. On large pieces, the use of a shadow graph magnifier is recommended. This consists of a point source illuminator and screen with the blank to be examined placed between them. The surfaces of the blank must be flat and polished. Imperfections appear as shadows on the screen.

On opaque materials, internal flaws can be examined by one of several established techniques using X-rays or sonic waves.

3-2.5.2 Calibration

3-2.5.2.1 Focal Length Calibration

There are several ways of calibrating to ensure the required focal length; the most direct and convenient involve the use of a calibrated collimator. An illuminated target is placed at the focal plane of the collimator and the output radiation is directed into the optical system under test. A screen used to pick up the image is moved about until the image is in sharp focus (as observed by the eye). The distance from the last optical element to the screen is the back focal length (BFL). The effective focal length (EFL) can be determined by measuring the magnification ratio. A camera back (film plate) is placed in the established position of the focal plane, and the collimator target is photographed. Its size is then compared with the target itself to determine the magnification.

The EFL of the system being tested is a function of (focal length of collimator) × (magnification). (If the focal length of the system under test is less than that of the collimator, the magnification will be a number less than unity.)

With refractive IR systems it is, of course, necessary to use a blackbody source for irradiating the collimator target. Also a detector must be used instead of the human eye for probing the image position and size.

3-2.5.2.2 Curvature Measurements

For determining the radii of a curve, a standard instrument called a spherometer is used. It permits direct measurement of the chordal height of a spherical segment of given diameter with micrometer precision.

An elegant technique exists for determining the curvature of concave spherical mirrors. A projection microscope such as illustrated in Fig. 3-39 may be used. The projection microscope is constructed so that an illuminated reticle is projected outward via a beam splitter and side source to form an image at the front focal point.

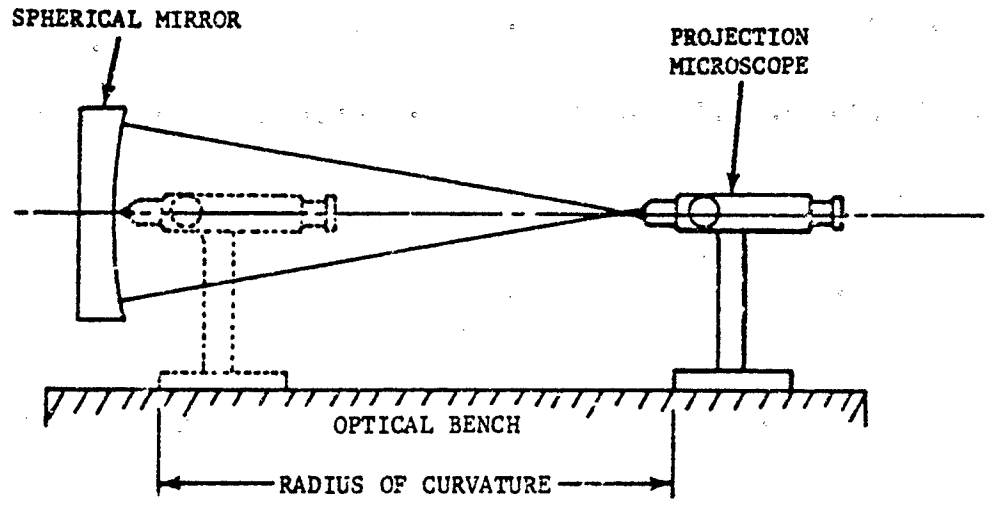


FIGURE 3-39. Setup for Determining the Curvature of Concave Spherical Mirrors

When used as in Fig. 3-39, the reticle image is re-imaged back on itself by the specimen mirror under test and can be seen through the eyepiece. When this image is sharp, the center of curvature of the specimen mirror is necessarily coincident with the microscope front focal point. The location of the microscope base can then be marked. Next, the microscope is moved close to the specimen until it is focused on the vertex point of the mirror surface. The distance from initial to final position of the microscope equals the radius of curvature.

3-2.5.2.3 Reflectance and Transmittance

The reflectance and transmittance characteristics of an optical system or of its components can only be adequately measured with specialized instrumentation. In the IR domain the instrument may be a radiometer, spectro-radiometer, reflectometer, or gonio-radiometer. Instructions for the use of these are supplied by the manufacturer.

3-2.5.3 Image Quality Measurements

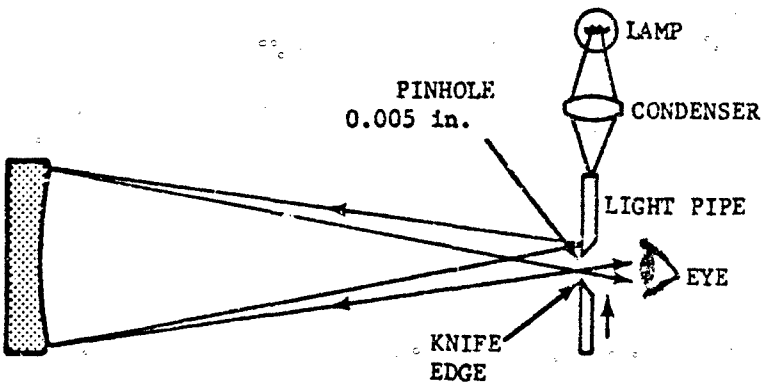
3-2.5.3.1 Knife-edge Autocollimation

The Foucault Knife-edge Test is an extraordinarily simple means for probing a point image and observing how it relates back to the optical system which produces it.

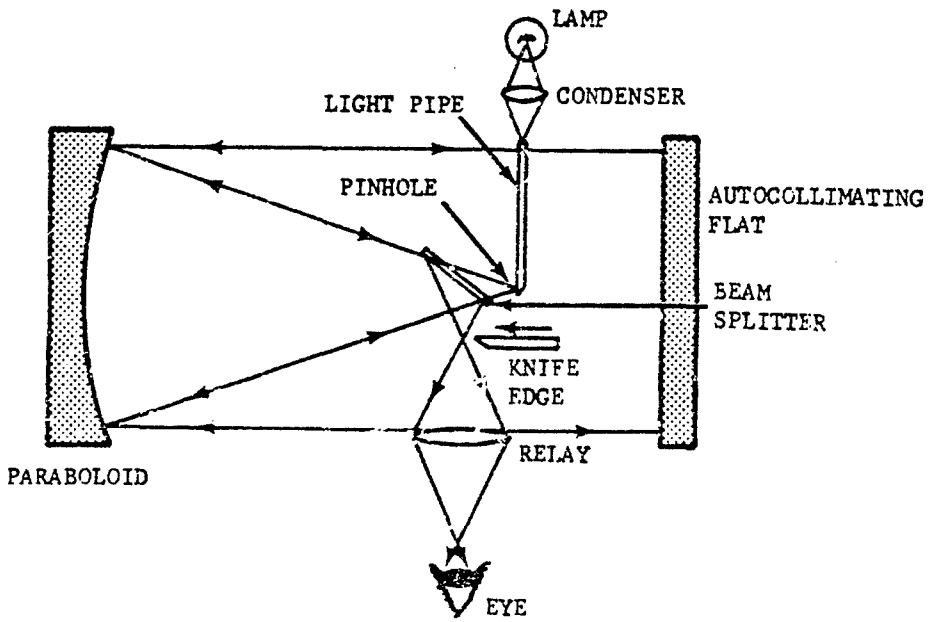
A knife-edge test setup for checking a simple spherical mirror is shown in Fig. 3-40(A). A pinhole source of light is positioned near the center of curvature of the mirror and slightly in the lateral direction. The mirror forms an image of the pinhole again near the center of curvature but an equal distance in the opposite direction. The knife edge is arranged to be moved so that it will cut across the image while the observer looks directly into it from behind. If the eye is brought sufficiently close to the image, the pupil will generally be much larger than the bundle of rays which form the image and, therefore, the entire aperture of the mirror will be observed simultaneously as a source of brightness comparable to that of the pinhole itself.

If the mirror is a perfect sphere and the diffraction effect is ignored, all the rays which form the image will converge to a point. As the knife edge is moved across this point all the rays in the bundle will be instantly cut off and the viewing field eclipsed. In a less perfect optical system the image will occupy a finite portion of space and the eclipse will be less rapid. The rate of eclipse, thus, is a measure of the perfection of the optical system.

In a diffraction-limited system, every point in the finite spot image receives energy from every point in the aperture. Consequently, the observed effect of the knife edge excursion across the image will be a gradual but uniform darkening of the entire viewing field.



(A) KNIFE-EDGE TEST SETUP
FOR SPHERICAL MIRROR



(B) KNIFE-EDGE TEST SETUP
FOR PARABOLOIDAL MIRROR

FIGURE 3-40. Knife-edge Test Setup for Spherical and Paraboloidal Mirrors

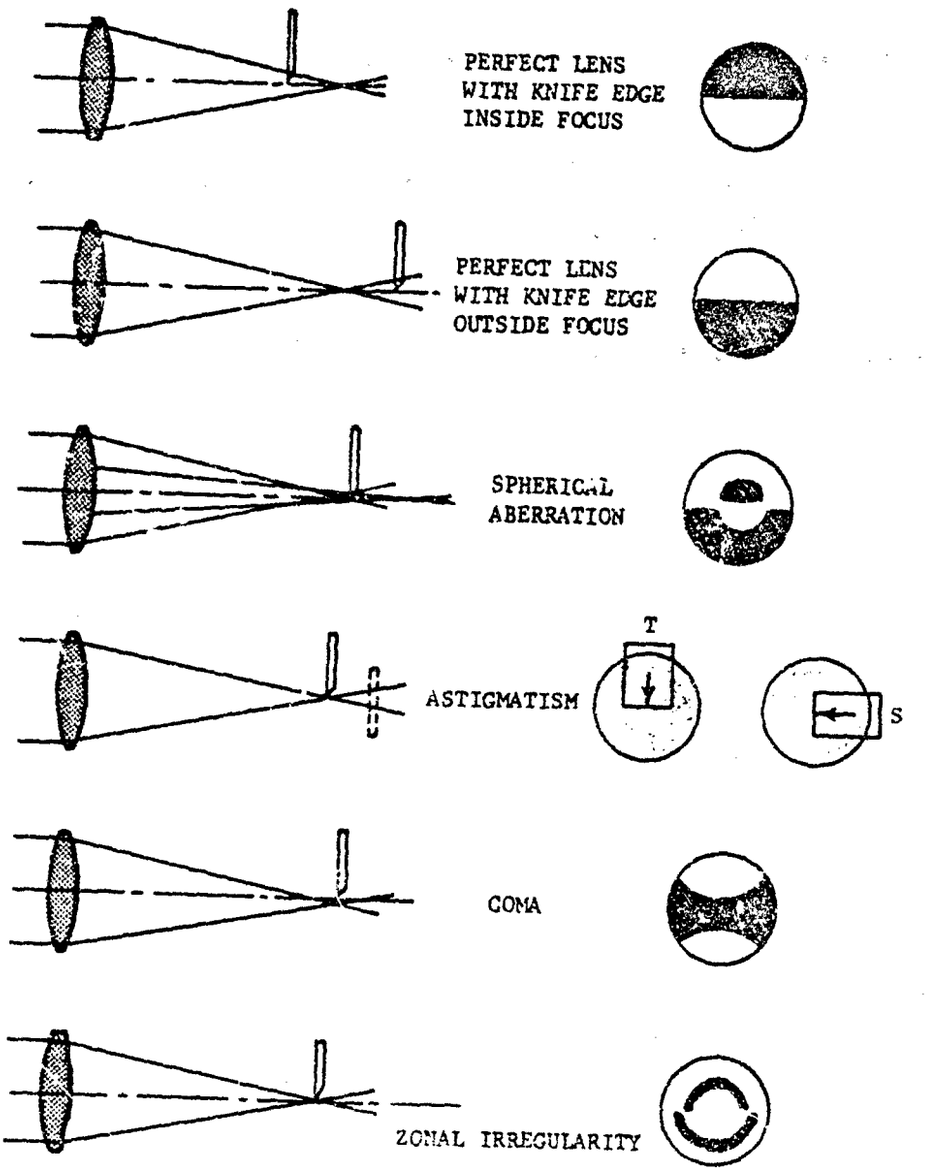


FIGURE 3-41. Knife-edge Shadow Patterns

If the system is geometrically imperfect the various portions of the image will not receive energy uniformly from all points of the aperture, and a variety of shadow effects will be observed as the knife edge progresses across the image. Note that since the diffraction image never shows any shadows, only a uniform darkening, any shadows observed are a direct and certain indication of the geometrical imperfections in the optical system.

Proper interpretation of the shadow effects can yield a great deal of information about the image being probed and about the system which produces it. All of the various aberrations, except distortion, can be readily identified and, in a limited way, measured. In particular, these shadows delineate the surface irregularities and inaccuracies in the system components so that the information derived can be used effectively to guide the figuring (final finishing) operation on such components (see Fig. 3-41).

Knife-edge probing techniques are by no means limited to spherical mirrors. All kinds of image forming systems can be tested. Fig. 3-40(B) shows an arrangement for testing a paraboloidal mirror and also demonstrates the principle of autocollimation.

Rays from the pinhole source, located at the focal point of the paraboloid, traverse the system to the autocollimating flat. From here they are reflected generally back on themselves to the paraboloid and onto an image point. This image point ordinarily coincides with the pinhole but the beam is intercepted before it reaches there by a 45-deg beam splitter so that a portion of the energy is redirected laterally to a more easily accessible region. This is the region which is to be probed with the knife edge.

In order to avoid serious blocking of the central portion of the paraboloid from view, the diameter of the beam splitter should be as small as possible. This is accomplished by locating the beam splitter close to the pinhole where the beam diameter is also small.

On wide angle systems it may then be necessary to view the image by means of a relay lens. No specific quality requirement is indicated for this lens since the knife edge has already generated an accurate shadow pattern at high magnification, which the relay lens cannot significantly alter.

3-2.5.3.2 Knife-edge Testing of Complex System

Fig. 3-42(A) shows the technique for testing an optical system of a compound form. The basic required instrumentation is a diffraction limited source of collimated energy having a focal length at least as large as that of the specimen system. Such a source system is not expensive to obtain or produce since its field angle needs to be only minutes wide.

The focal plane region of the specimen system must be accessible for placement of the knife edge. Shadow patterns due to knife-edge probing are then interpreted in the conventional manner.

Off-axis image points may be probed by simply tilting the optical specimen system relative to the collimated beam.

3-2.5.3.3 Knife-edge Testing of a Hyperbola

A setup for testing a hyperbola is shown in Fig. 3-42(B). The remote virtual focal point is simulated in real image space by retro-imaging it in a spherical mirror via a beam splitter. A real image is thus formed at the other focal point which can be probed with a knife edge.

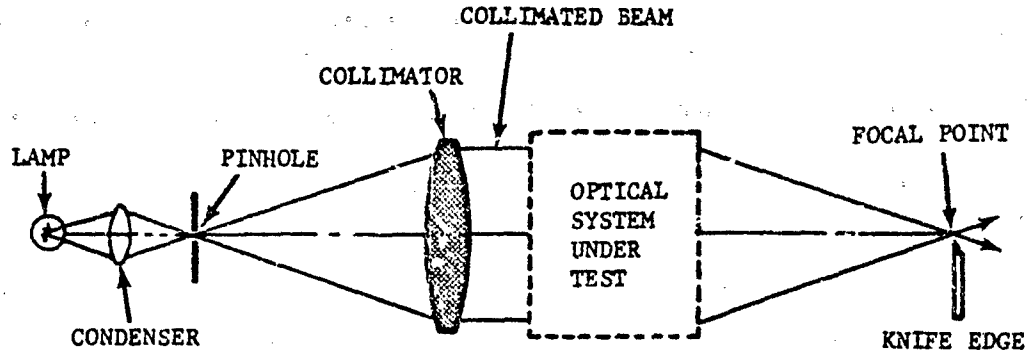
3-2.5.3.4 Knife-edge Testing of Large, Flat Mirrors

Large, flat surfaces cannot be easily tested by means of the customery test plates and interference fringe effects because of the weight of the parts which have to be moved around. Instead, a Ritchie test setup as shown in Fig. 3-43 will be most effective.

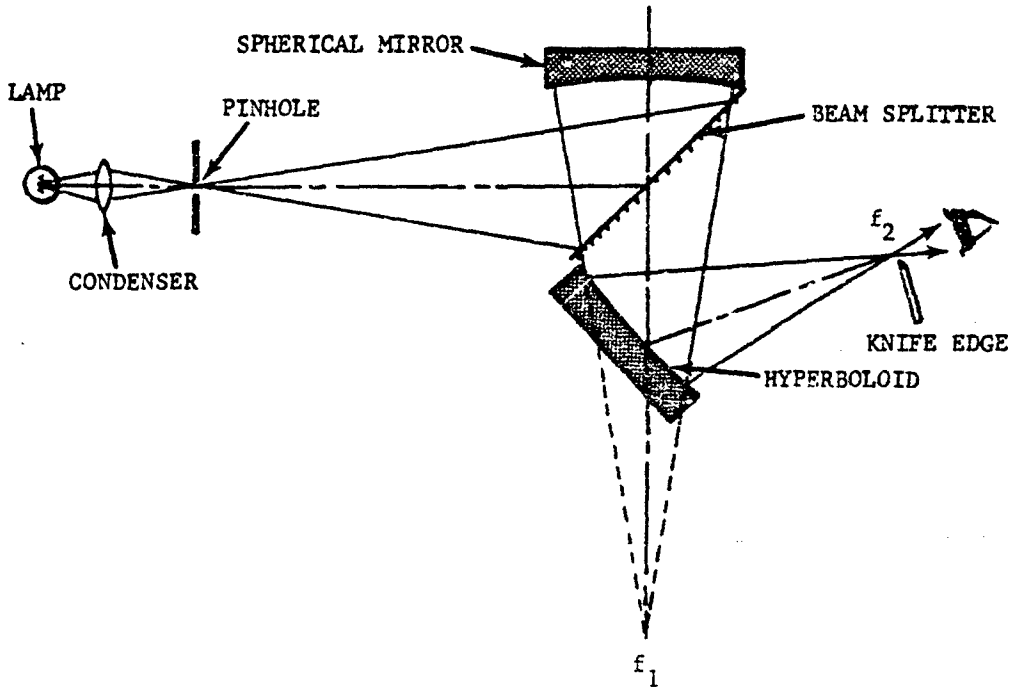
The reference spherical surface must have an accuracy of higher order than that to which the flat is to be tested. It is then assumed that all inaccuracies which the knife edge reveals are due entirely to the flat surface.

3-2.5.3.5 Ronchi Grating (Fig. 3-44)

If all rays in a bundle are converging to a *point*, it may be observed that any figure carried by the rays therein must be geometrically similar at all positions along the bundle. Thus, if a grating of parallel lines is placed at a point near the focus and viewed through the image point, an enlarged undistorted image of this will be observed at "A".



(A) KNIFE-EDGE TEST SETUP FOR COMPLEX OPTICAL SYSTEM



(B) KNIFE-EDGE TEST SETUP FOR HYPERBOLOIDAL MIRROR

FIGURE 3-42. Knife-edge Test Setup for Complex and Hyperboloidal Systems

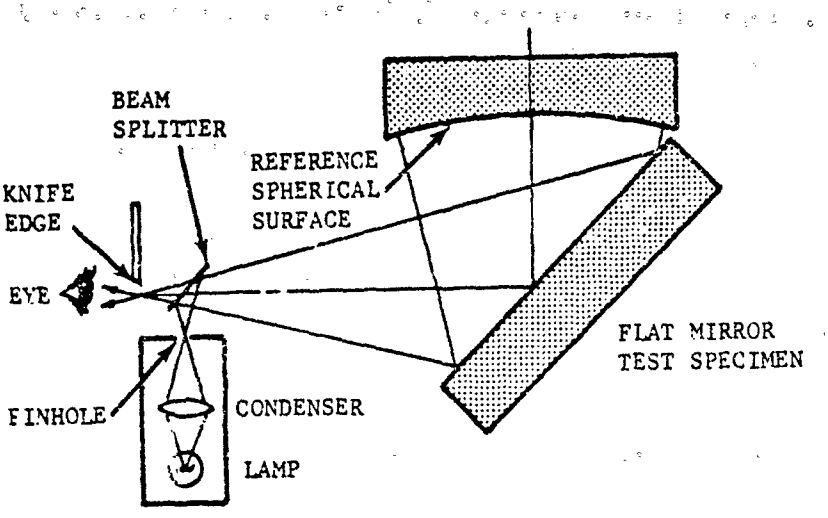


FIGURE 3-43. Ritchie Setup for Testing a Large Flat Mirror Using a Highly Accurate Spherical Mirror as Reference

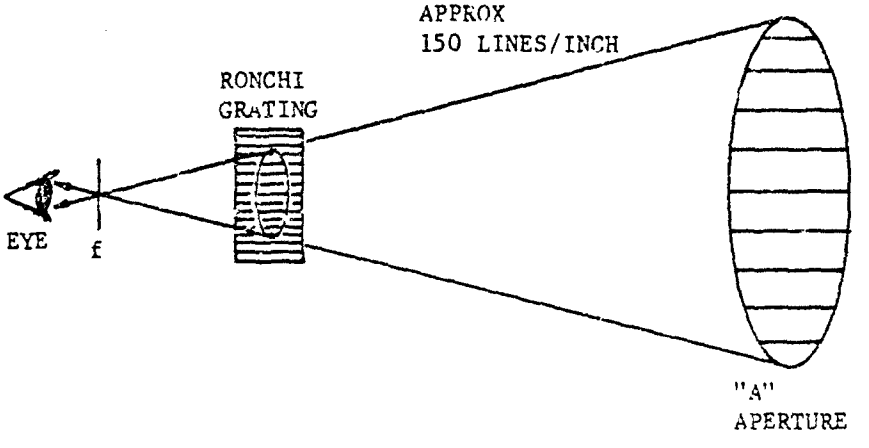


FIGURE 3-44. Ronchi Grating for Determining Duration of Rays from Perfect Focus

If the rays do not focus at a point, the lines at "A" will appear distorted to the degree which the rays deviate from a perfect focus. Furthermore, the apparent location of the distorted elements in the aperture is an indication of the location of irregularities in the system for which the aperture is a front.

Only qualitative data can be obtained from a Ronchi grating and its sensitivity is rather low.

3-2.5.4 Resolution Targets

The military services and industry have adapted a "standard" type resolution target which is widely used in the testing of camera systems operating in the visible and IR regions. The target consists of a number of groups of parallel lines and spaces, wherein each group comprises six sets of three lines and two spaces each. The lines and spaces of the several sets are graded in width from a dimension smaller than that of the least resolution element to be measured, to a nominal value greater than the largest element worth considering (see Fig. 3-45).

The target which is simply exposed to the camera under test either directly or through a collimator is photographed. Upon examination of the photograph, it is generally possible to observe one set of three lines where it is just barely possible to distinguish the three lines from each other. The spacing associated with this set is taken as the resolution limit of the camera system.

3-3 EMITTERS AND ILLUMINATORS

3-3.1 SOURCES OF ILLUMINATION

The theoretical as well as the practical aspects of incoherent and coherent sources of illumination are discussed to the extent possible without divulging classified information, specifically with respect to flares. Considerable information concerning the chemical systems and emissions of flares is classified; as a result, this Handbook deals only superficially with the design problems involved in creating effective IR emission. Additional unclassified information on these subjects is contained in Ref. 30. Persons or organizations having a need-to-know and the proper security clearance may request access to the classified literature through their Contracting Officer or the Electronic Warfare and Communications Laboratory, Wright-Patterson AFB, Ohio 45433.

A list of the symbols used throughout par. 3-3 is given in Table 3-4 along with the corresponding definitions.

3-3.1.1 Flares

The IR radiation of an object depends not only on its temperature but also on its emissivity, as discussed in par. 2-2. Consequently, high-temperature flames or incandescence are not guaranteed sources of copious amounts of IR radiation. For example, a reaction producing metal or metal oxide particles at a high temperature will not necessarily emit stronger in the IR since many of these materials have a low emissivity. Most gaseous reaction products are poor radiators even though the temperature may be quite high.

With this in mind, the flare designer must select a reaction with a high flame temperature producing primarily solid exhaust particles with a high emissivity. In addition, the chemicals in the flare must be stable at normal temperatures, relatively nonhygroscopic, and capable of being mixed and formed into a flare candle. Since gas-producing reactions suffer more degradation at reduced pressures than do solid particle reactions, this effect must be considered for flares intended for high-altitude use.

Most flare chemical systems consist of metal powders reacting with a suitable oxidizer. Emission of these flares is semi-blackbody with radiation spanning the range from visual light up to 5 or 6 microns.

Specifics of flare construction depend upon the application. Typical IR flares consist of cylindrical flare grains encased in plastic or metal housings with an end seal. Since most flares are electrically ignited, an internal squib is provided connected to an external connector in the base of the cylindrical housing. Flares may be held captive or may be assembled into a dispenser which ejects individual flares on command.

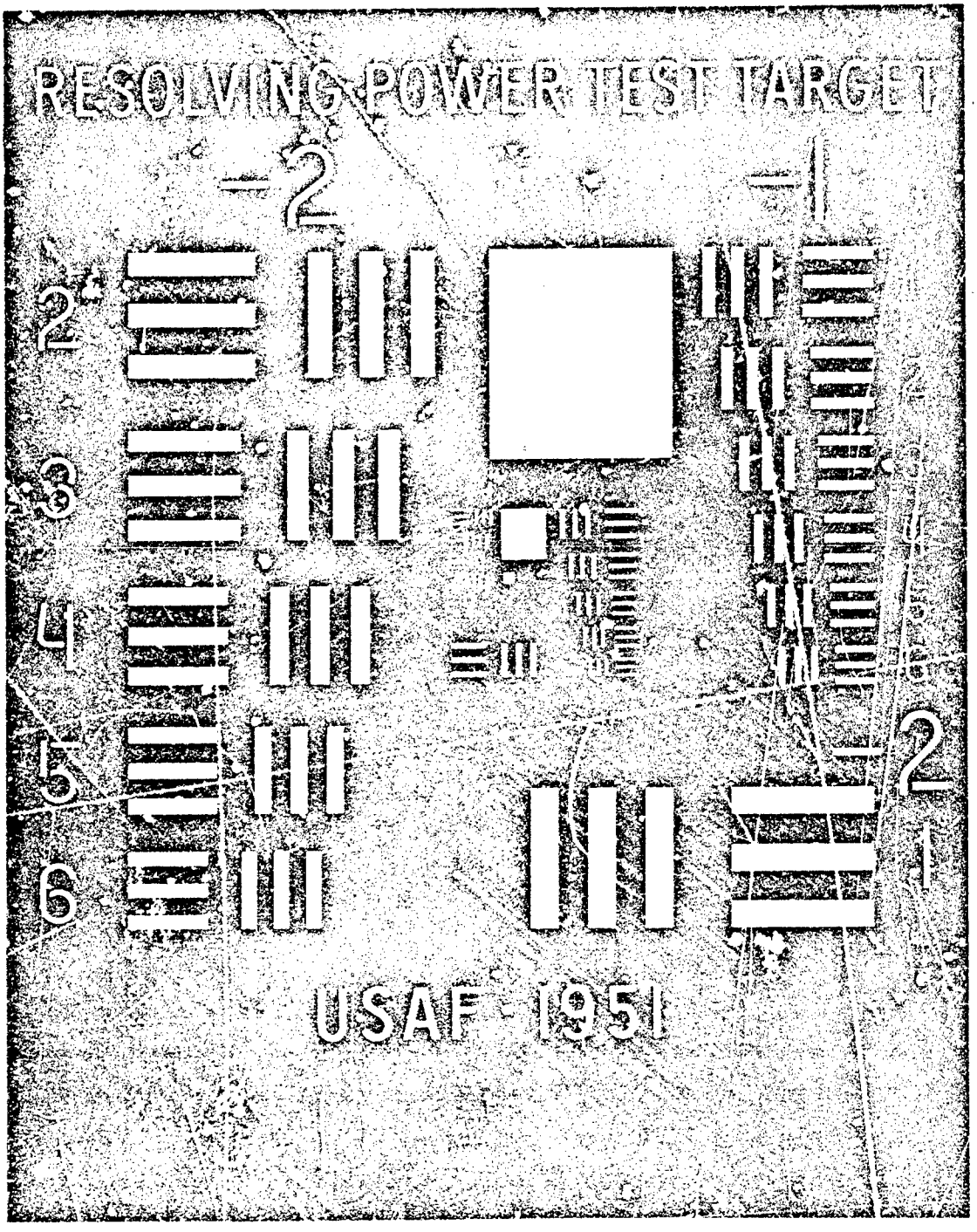


FIGURE 3-45. Resolution Chart

TABLE 3-4. STANDARD SYMBOLS FOR PAR. 3-3

SYMBOL	DEFINITION
Å	Angstrom
A_c	Area over which the phase of the wave is correlated
A_{mt}	Probability of electron falling from state (m) to lower energy state (t)
B_m	Birefringent crystals
B_{tm}	Proportionality constant
c	Speed of light
$c/2L$	Fundamental cavity frequency
d	Distance from the optics to the position Z of the point source
d_p	Separation between two prisms in frustrated internal reflected modulator
E	Energy of photon emitted by atom falling from 1st excited state to ground
E_m	Energy of photon in excited state
E_t	Energy of photon in lower energy state
f	Frequency of photon emitted by atom falling from 1st excited state to ground
f_n	Resonant frequency of optical cavity
f_{opt}	Optical frequency
Δf	Bandwidth
h	Planck's constant
L	Length of laser cavity
I_L	Incident laser power
I_{max} & I_{min}	Maximum and minimum radiation intensity, respectively
I_o	Intensity of radiation prior to passing through a given volume
I_T	Laser power transmitted through mirror
I_x	Radiation absorbed while passing through a given volume

SYMBOL	DEFINITION
L	Separation between mirrors in laser cavity; length of laser cavity
L_c	Coherence length
m	Excited state of electrons
N_m	Number of excited states per unit volume
N_t	Numer of electrons in the transition stage per unit volume
n	Wavelength integer in calculating phase reversal of wave reflected in laser cavity; index of refraction of the medium
P	Net radiated power per unit volume
P_{abs}	Power absorbed per unit volume
P_{sp}	Electron power emitted spontaneously per unit time and unit volume
P_{st}	Power emitted by stimulated emission
Q	Quality factor of an optical resonator
R	Reflectivity of both mirrors in laser cavity
S_-	A series of electro-optic switches in digital light reflectors
S_1 & S_2	Distances traveled by two radiation waves
T	Absolute temperature, °K
t	Transit time of reflection from one mirror to another
t_c	Time during which radiation may be a pure sinusoid (coherence time)
U_f	Photon density of radiation at the proper frequency
V	Visibility of the fringes
$V_{\lambda/2}$	Half-wavelength voltage
W_0	Radius of the mode at the output mirror

TABLE 3-4. STANDARD SYMBOLS FOR PAR. 3-3 (Continued)

SYMBOL	DEFINITION	SYMBOL	DEFINITION
x	Distance traveled by radiation	δ	Fraction of photons lost per transit by diffraction
Z	Point source	λ	Wavelength
Z	Distance from the collimating optics to the common focus of cavity mirrors	$\Delta\lambda$	Difference between one resonant wavelength and another
α	Optical gain coefficient	σ	Cross section for stimulated emission or absorption
$\Gamma_{1,1}(\tau)$	Autocorrelation of electric field	τ	Temporal coherence of oscillation
$\Gamma_{1,2}(\tau)$	Cross correlation of electric field	Ω	Solid angle
γ	Radiation loss in passing through cavity		

3-3.1.2 Lamps

Lamps are broadband sources which may be used for illumination in the visible and near IR regions of the spectrum. Most lamps produce continuous radiation on which line radiation may be superimposed. Filters or dichroic mirrors are used to block visible radiation in applications where only IR radiation is desired.

3-3.1.2.1 Tungsten Filament

Tungsten filament incandescent lamps, which can be operated up to 3000°K, are high-radiance sources. The hot filament radiates like a gray-body with spectral emissivity as given in Table 3-5.

The tungsten filament may be in the form of either a ribbon or wire. A ribbon-filament lamp provides a source of radiant energy having a large area of uniform radiance. Filaments as large as 4-mm wide and 40-mm long are available for use as secondary calibration standards. Most wire filaments are wound into a coil or helix in order to reduce the apparent length and increase the apparent diameter of the filament, thus providing a more compact source. In some lamps, this helix is coiled into another helix, resulting in a further increase in apparent diameter.

The filament must be contained in an evacuated or inert-gas-filled transparent envelope, usually glass or quartz. Employing an inert gas also lowers the evaporation rate of the filament. Evaporation reduces the cross section of the tungsten wire and thus increases its electrical resistance which, for a fixed lamp voltage, results in a lower power consumption and output by the lamp. The primary factor determining the life of an incandescent lamp is the evaporation rate of the filament, which is a function of the temperature. For gas-filled lamps with filament temperatures of about 3000°K, the lifetime is inversely proportional to approximately the thirty-fifth power of temperature[31]. A small percentage of the evaporated tungsten is returned to the filament by collision of the tungsten atoms with the molecules of the filling gas (argon or nitrogen) while the remainder deposits on the lamp envelope and reduces the transmission of the envelope. Collector screens are used in some lamps to collect vaporized tungsten before it reaches the envelope.

The introduction, in 1959, of the regenerative halogen cycle to tungsten lamps allows higher operating temperatures with reduced filament evaporation[32]. The most important of these lamps is the quartz-iodine lamp[33]. A small amount of iodine added to the inert filling gas in the lamp causes a chemical reaction in which the tungsten deposited on the hot bulb wall is returned to the filament. At normal bulb-wall temperatures (200° to 600°C), the iodine combines with the deposited tungsten to form the volatile compound tungsten iodide. The high

temperature in the neighborhood of the filament causes the tungsten iodide to decompose into its constituent elements with the tungsten being redeposited on the filament. These lamps are designed to operate with a bulb temperature of about 600°C and are, therefore, constructed with a high softening point material such as quartz.

A technique has been studied for reducing the input power to an incandescent lamp while still retaining its output brightness for use in projection type devices, searchlights, and other stop-limited optical systems[31]. The technique consists of placing the filament in the center of a spherical reflector which will reflect the radiated energy back to the filament and thus maintain its temperature with a smaller input power. An aperture in the reflector permits transmission of that spatial component of the radiation which can be effectively coupled into the optical system. The study showed that a power reduction of 40 percent is practical by this method.

The uses of tungsten-filament lamps as night vision illumination sources have ranged from a truck-mounted night driving IR lamp of a few watts to 18-in. tank-mounted searchlights of 2,500 w.

3-3.1.2.2 Carbon Arc

Two main types of carbon arc sources are used for illimination: low-intensity incandescent arcs and high-intensity flame arcs. In low-intensity arcs, the radiation is produced primarily by a shallow crater in the tip of the positive electrode, which is heated to its sublimation temperature by an electric arc. Additional radiation is produced by a gaseous layer immediately in front of the crater. The radiant emission from the crater appears to be close to that of a 3800°K blackbody. The ratio of measured radiance to that of a 3800°K blackbody is 95 percent in the wavelength interval of 0.63 to 4.2 microns[34].

The radial distribution of luminance across the positive electrode drops to 90 percent of its center value approximately one-half the radial distance out from the center and to 50 percent at about two-thirds the radial distance out[35].

In the high-intensity arc most of the radiant energy is produced in the arc flame with a lesser amount produced by the incandescent crater. The arc stream is made luminescent by vaporization of rare-earth salts placed in the core of the electrode. The spectral distribution of the flame radiation can be altered by changing the core material and current. The visible portion of the spectrum is generally enhanced by this method, thus obtaining color temperatures of 5000° to 9000°K.

A ballast resistor is required in the electrical power source to compensate for the negative resistance characteristic of the carbon arc.

3-3.1.2.3 Gaseous Arc

Argon, krypton, and xenon gases as well as cesium, mercury, and rubidium metal vapors have been used in arc discharge lamps in the visible and infrared spectral regions. At the present time xenon arc lamps are predominantly used as illumination sources for near-infrared night vision equipment. Input power for these lamps ranges from 150 w to 30 kw[36].

Fig. 3-46 shows the spectral distribution of a compact xenon arc lamp measured at four lamp input powers: 2.6, 5.0, 7.5, and 10 kw[37]. The resolution for these curves is 0.1 micron. It can be seen from this figure that the efficiency of generating near infrared radiation between 0.75 and 1.05 microns decreases as compared to the generation of visible radiation. Fig. 3-47 from the same reference shows the spectral distribution for a 10-kw xenon discharge with a resolution of 0.01 micron.

Fig. 3-48 shows the spectral emission from 0.35 to 1.1 microns of a 1000-w xenon lamp normalized to the input power.

Spectral distributions of xenon lamps have been measured for various current densities. As the current density is increased from a few tens of amperes per square centimeter to a few thousand amperes per square centimeter, the peak of the spectral distribution shifts toward the blue and the visible intensity increases much more rapidly than the infrared. Also, the line emission in the infrared, while strong at low current densities, is almost completely masked by continuum at high current densities.

TABLE 3-5. SPECTRAL EMISSIVITY OF TUNGSTEN AT TEMPERATURES BETWEEN 1600° AND 2800°K

	EMISSIVITY						
WAVELENGTH, μ	1600°K	1800°K	2000°K	2200°K	2400°K	2600°K	2800°K
0.25	0.449	0.442	0.437	0.430	0.424	0.416	0.410
0.30	0.482	0.478	0.474	0.470	0.465	0.461	0.456
0.35	0.489	0.486	0.483	0.480	0.477	0.474	0.471
0.40	0.481	0.477	0.475	0.471	0.469	0.464	0.461
0.50	0.469	0.465	0.462	0.458	0.455	0.451	0.448
0.60	0.455	0.451	0.448	0.444	0.441	0.438	0.434
0.70	0.444	0.440	0.436	0.432	0.428	0.423	0.419
0.80	0.421	0.426	0.420	0.414	0.409	0.404	0.400
0.90	0.413	0.407	0.401	0.396	0.390	0.386	0.383
1.0	0.392	0.388	0.382	0.378	0.375	0.371	0.368
1.1	0.367	0.364	0.361	0.358	0.355	0.353	0.352
1.2	0.344	0.343	0.342	0.341	0.340	0.339	0.338
1.3	0.325	0.324	0.323	0.323	0.323	0.322	0.322
1.4	0.313	0.311	0.310	0.308	0.306	0.302	0.300
1.5	0.302	0.299	0.296	0.292	0.288	0.284	0.281
1.6	0.292	0.288	0.283	0.278	0.273	0.268	0.264
1.8	0.275	0.268	0.262	0.255	0.247	0.241	0.234
2.0	0.259	0.251	0.243	0.235	0.227	0.219	0.210
2.2	0.245	0.236	0.228	0.218	0.210	0.201	0.190
2.4	0.233	0.224	0.215	0.206	0.196	0.187	0.176
2.6	0.224	0.214	0.205	0.194	0.185	0.175	0.164

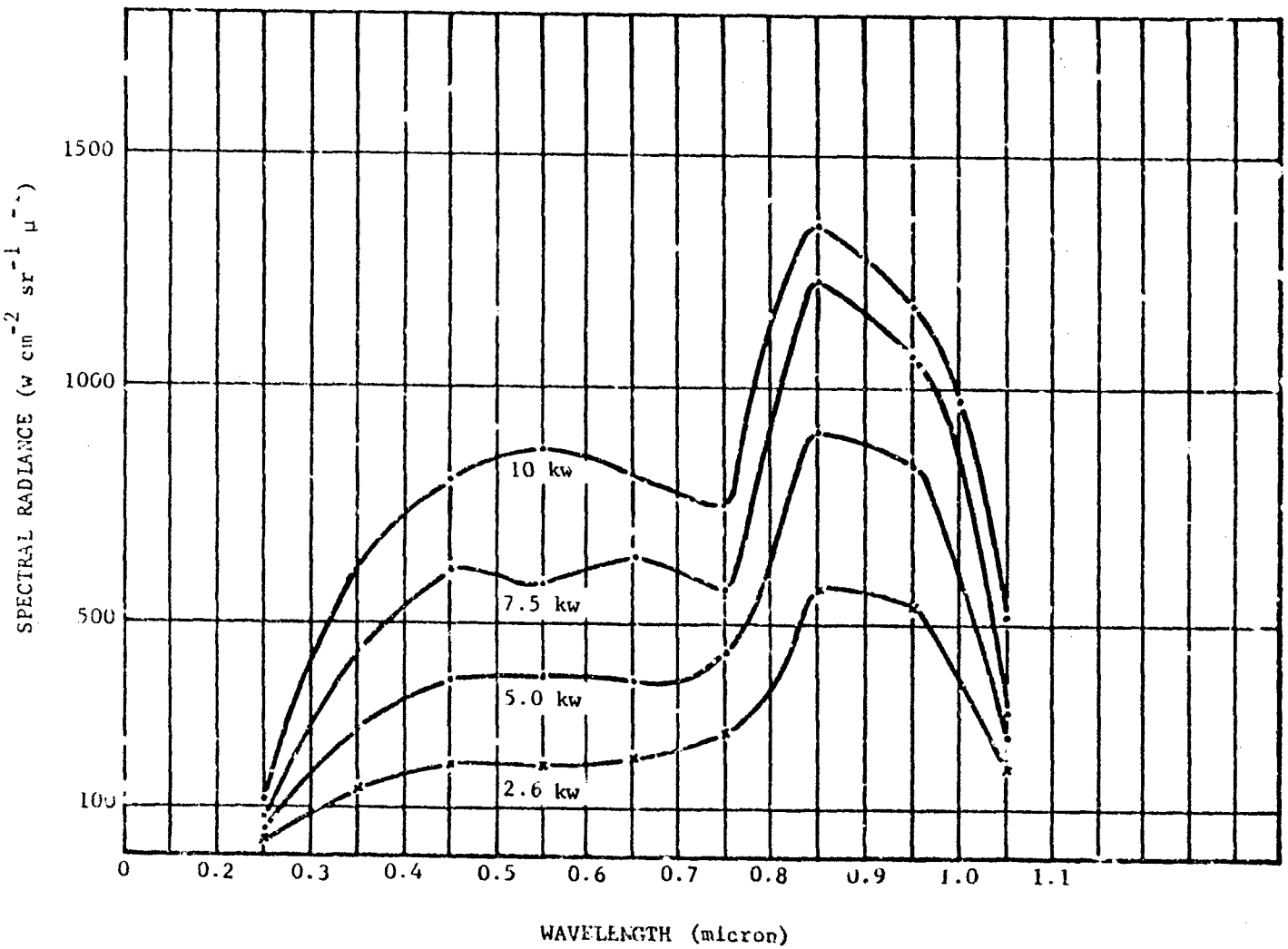


FIGURE 3-46. Spectral Distribution of Xenon Discharge With Resolution of 0.1 Micron for Selected Input Powers

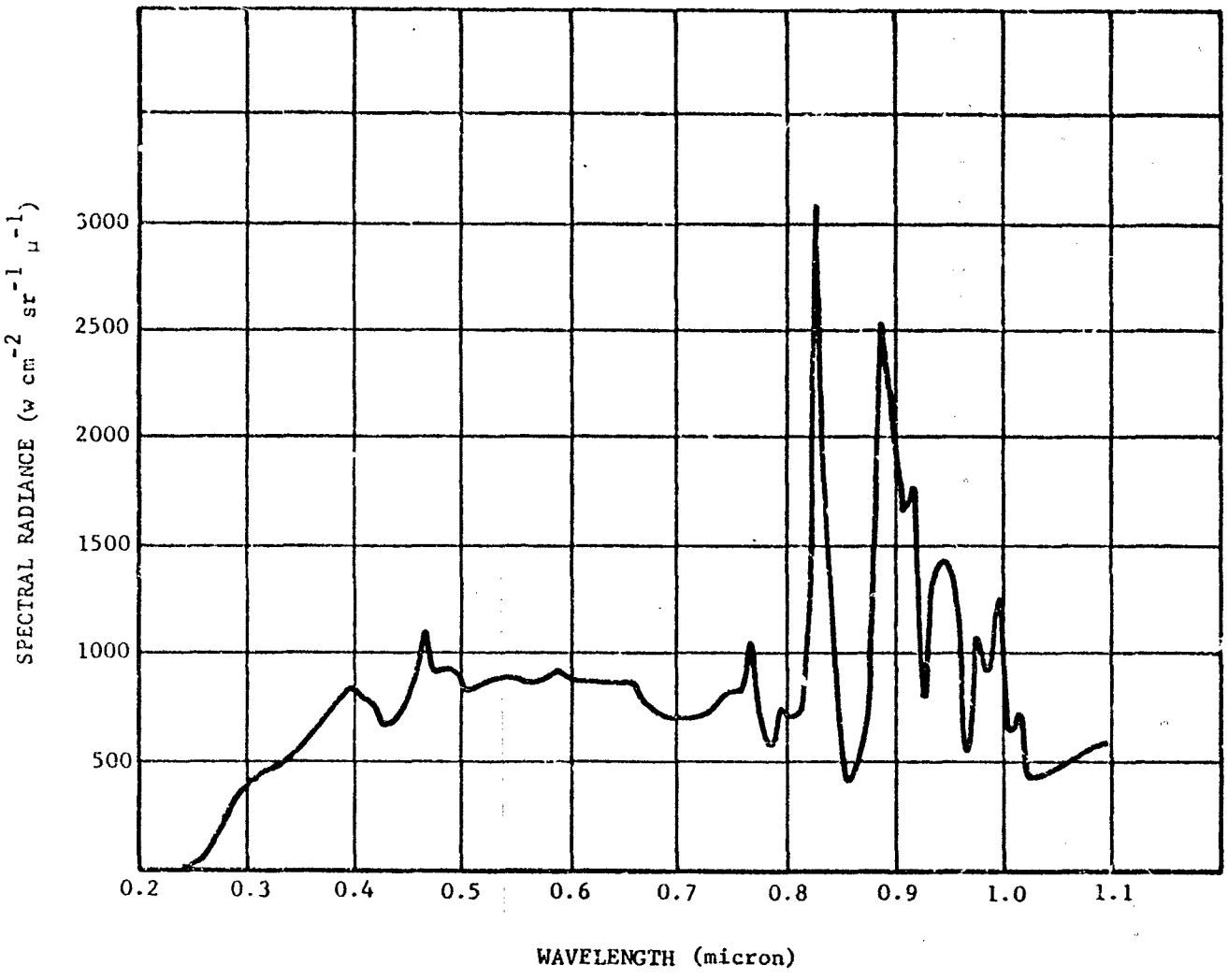


FIGURE 3-47. Spectral Distribution of 10-kilowatt Xenon Discharge With Resolution of 0.01 Micron

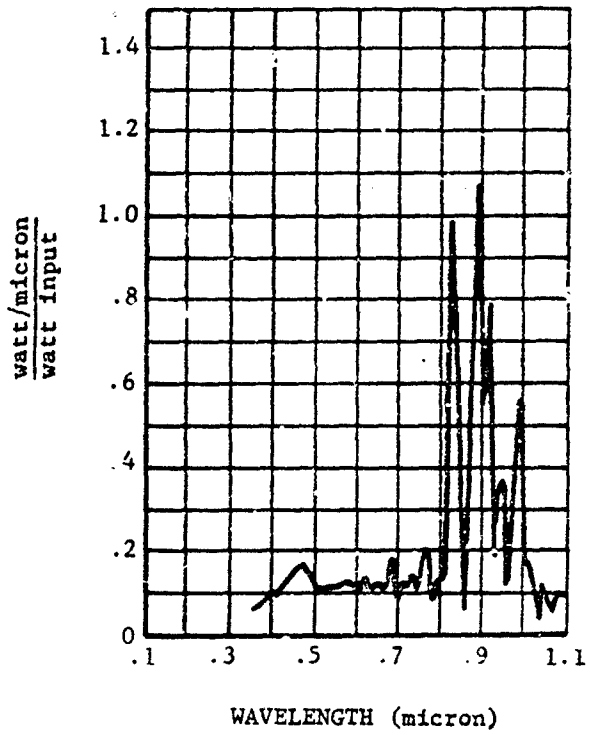


FIGURE 3-48. Spectral Radiant Power of Experimental dc Lamp (1.7-atm X_6; amp cm^{-2}) Normalized to Input Power. Spectral Resolution Equal to 0.01 Micron

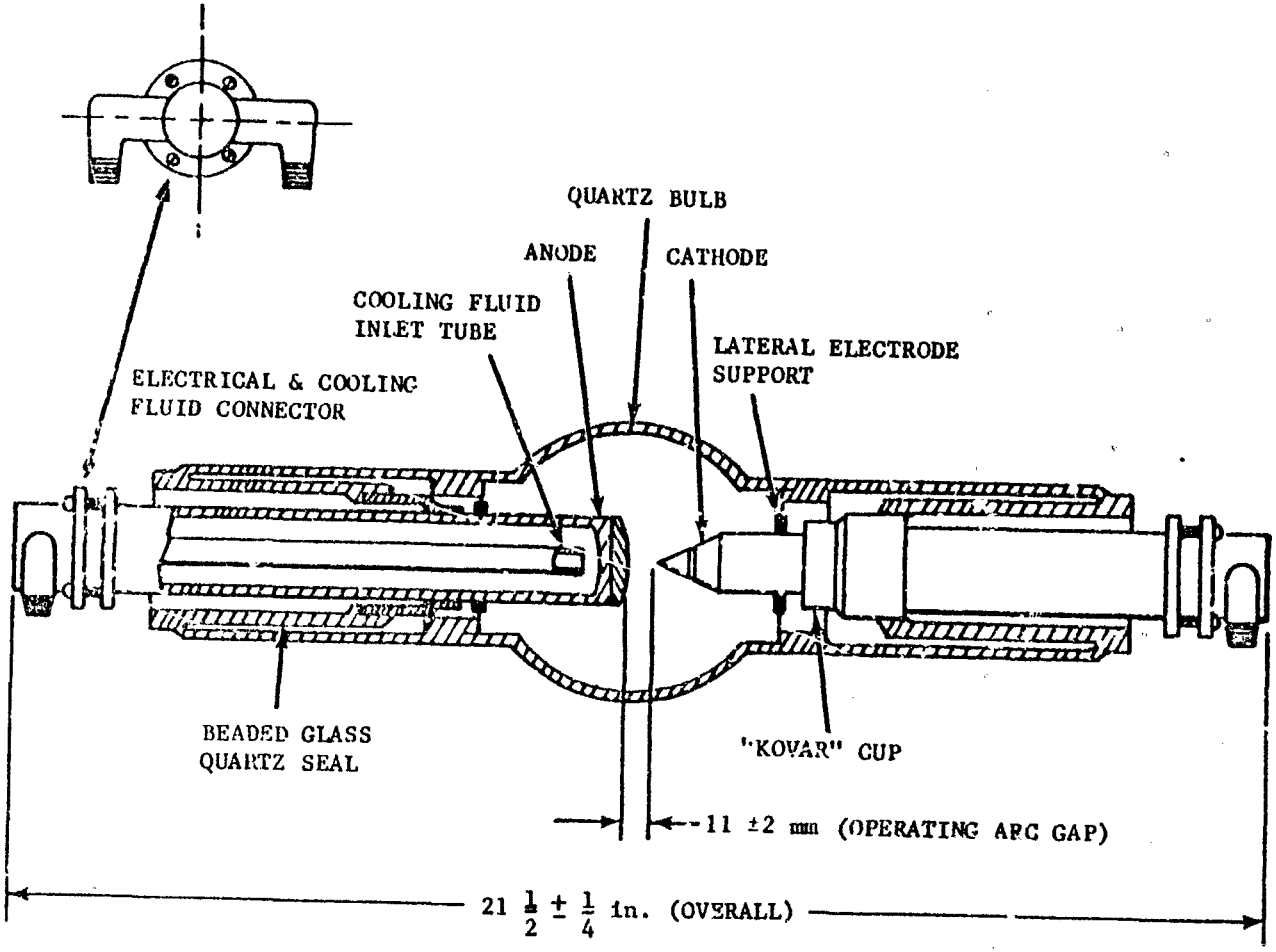


FIGURE 3-49. Diagram of the Experimental 20-kw Lamp

A 20-kw liquid-cooled, compact xenon arc lamp, shown diagramatically in Fig. 3-49, has been developed for searchlight applications[37]. The electrodes of the lamp are internally cooled by a circulating liquid, enabling the lamp to be about the same size as an air-cooled lamp of 6 kw or less.

The electrode gap is 11 ±2 mm resulting in a drop of 44.4 V across the lamp with a current of 450 A for a total power dissipation of 20 kw. A 30 to 50 kV starting pulse is required to initiate the arc discharge. The amount of lamp power dissipated by various means is as follows:

Cooling liquid	8 kw
Radiation (near IR and visible)	10 kw
Conduction and convection	2 kw

Development of arc radiation sources using mixtures of xenon with argon, neon, cesium, and rubidium has been undertaken with the goal of increasing the near infrared radiance over pure xenon lamps of equal input power[38]. The output intensities of these lamps were much less than those of the pure xenon lamps even when xenon was present in large amounts. A xenon lamp, doped with tin iodide (SnI_4), had the same intensity as a pure xenon lamp in the 0.8- to 12-micron region but was four times more intense than pure xenon in the 1.2- to 1.8-micron region[39].

Either a ballast resistor or power supply designed to have the proper volt-ampere characteristics is required for use with arc lamps to compensate for the negative resistance characteristic of arc discharge lamps.

3-3.1.2.4 *Flashlamps*

Flashlamps, which are pulsed, are rated by input and output energy rather than input and output power as for continuous lamps. As mentioned above, the spectral distribution in a xenon flashlamp is controlled by the current density during the current pulse. Fig. 3-50 shows the spectral radiance normalized to input energy of an FZ-47A flashtube manufactured by

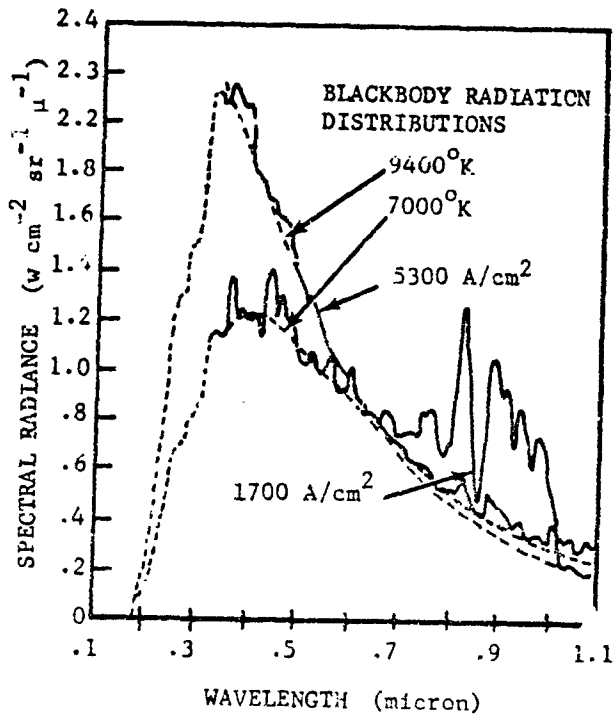


FIGURE 3-50. Spectral Radiant Energy of FZ-47A Flashtube Normalized to Input Power. Spectral Resolution Equal to 0.01 Micron

EG&G. The spectral resolution is 0.01 micron. The spectral radiance was recorded for two pulsed current densities, 1700 A cm^{-2} and 5300 A cm^{-2}. The coarse broken lines are relative spectral radiance of blackbodies at 7000° and 9400° K which determine the color temperatures of the source. The fine broken lines indicate that the measurements made in the ultraviolet are less accurate than those made in the visible and infrared.

3-3.1.3 Lasers

3-3.1.3.1 Introduction

The word *LASER* is an acronym for *L*ight *A*mplification by *S*timulated *E*mission of *R*adiation. The physical principle behind the laser was advanced as a theoretical necessity by Einstein, but it was not until after the success of the MASER (*M*icrowave *A*mplification by *S*timulated *E*mission of *R*adiation) as a low-noise microwave amplifier that the principle was applied to optical frequencies. However, for reasons that can be best understood by referring to Planck's Radiation Law (Chapter 2), the laser is not a low-noise amplifier. Its principal value is as an oscillator, i.e., as a source of coherent optical frequency radiation.

It has been increasingly clear, during the struggle of radar engineers to develop systems designed to operate at shorter and shorter wavelengths, that techniques other than wavelength-size metal cavities and waveguides would eventually be necessary. The laser represents such a technique.

Before discussing specific laser devices, a brief review of laser theory will be presented. The treatment here is simplified, and subsequent study of the references listed at the end of the chapter is recommended.

3-3.1.3.2 Laser Theory

Atoms are able to store energy in discrete amounts (quanta) in the potential and kinetic energy of their orbiting electrons. Each electron is capable of revolving around the nucleus in more than one orbit, each having a definite amount of energy. When an electron in its lowest energy state (ground state) accepts energy from some outside source, it changes its orbit from its lowest energy orbit to one of higher energy. This process is called electronic excitation, and the higher energy condition is called an excited state. Excitations can occur not only between the ground state and the lowest excited state, but between any state and a higher energy state. They may be caused by the capture of a photon having the right amount of energy to make up the difference between the energies of the two states of the electron, or excitations may be brought about by collisions that transfer energy to the atom.

Transitions also occur from higher energy states to lower energy states. In these cases, the exact amount of energy equal to the difference between the higher and lower energy states is released in some form. One way of releasing energy is the emission of a photon. The photon emitted by an atom falling from its first excited state to the ground state would be indistinguishable from the photon absorbed by the electron in its transition to the excited state. Its energy E and frequency f are related by

$$E = E_m - E_t = hf \tag{3-46}$$

where E_m and E_t are the energies of the two states, and h is Planck's constant. Rather than emitting photons, some transitions release the quantum of energy to the surroundings in the form of mechanical vibration which is then transformed into heat.

An electron may fall from an excited state (m) to a lower energy state (t) spontaneously, without any external influence. The probability of this "spontaneous" emission occurring for an atom in an excited state per unit time is defined, for a particular transition, as

$$Probability_{m\,t} = A_{mt} \tag{3-47}$$

Thus, if there are N_m of the excited states per unit volume, the power emitted spontaneously per unit time and unit volume is

$$P_{sp} = hfA_{mt}N_m \tag{3-48}$$

The transition of electrons from one state to another may also be induced by the presence of a photon having the same energy as the difference in the energies of the two states. If the direction of the induced transition is upward, the process is called "absorption". The probability of its occurrence is proportional to the

number density N_t of atoms in the energy state from which the transition is to take place. a proportionality constant B_{tm} and the photon density of radiation U_f of the proper frequency f. Thus, the power absorbed per unit volume is

$$P_{abs} = B_{tm} U_f N_t hf \tag{3-49}$$

If the induced transition is in the downward direction, resulting in the emission of a new photon, the process is called "stimulated emission". The probability per unit time and unit volume of this event is also proportional to the number density N_m of atoms in the state from which the transition is to take place a proportionality constant B_{mt}, and the photon density of radiation U_f of the frequency f. The power emitted by this process is

$$P_{st} = B_{mt} U_f N_m hf \tag{3-50}$$

In the simplest case, that of distinct energy levels,

$$B_{mt} = B_{tm} \tag{3-51}$$

The combination of stimulated emission and absorption gives a net radiated power per unit volume

$$P = hf \left(N_m - N_t \right) B_{mt} U_f \tag{3-52}$$

Adding to this power the radiation due to spontaneous emission,

$$P = hf \left[A_{mt} N_m + B_{mt} U_f (N_m - N_t) \right] \tag{3-53}$$

The proportionality constants A_{mt} and B_{mt} were first derived by Einstein, and are known as the Einstein Coefficients. They are related by

$$A_{mt} = \left(\frac{8\pi hf^3 n^3}{c^3} \right) B_{mt} \tag{3-54}$$

where c is the speed of light and n is the index of refraction of the medium.

The process that is of primary interest to a discussion of lasers is that of stimulated emission. In this process, an electron goes from a higher energy state to a lower one, the transition taking place because of the presence of a photon of the appropriate frequency. The excess energy in the transition is released in the form of a photon exactly like the one causing the transition, *even to its polarization, its optical phase, and its direction of propagation.* All laser phenomena are based upon this fact.

Similar to energy storage by atoms in electronic states, as described above, molecules store energy in mechanical vibration and rotation. Consider a molecule consisting of two atoms. The mass of the molecule is concentrated in the two atoms, and the bond between them may be thought of as a spring. The molecule is thus able to vibrate, the bond being alternately stretched and compressed, as the atoms move in and out. We know from quantum mechanics that the energy in such an osciilator cannot have any arbitrary value, but must be a multiple of some fundamental quantum of energy related to the resonant frequency f of the oscillator by Eq. 3-46. The frequencies of vibrational oscillations of molecules tend to fall in the infrared and microwave range of the frequency spectrum.

In the same way as the electrons of the atom move to different orbits by the absorption of specific amounts of energy, the molecule absorbs and stores discrete amounts of energy by going into larger amplitudes of vibration. It may release energy spontaneously, emitting photons corresponding to the energy differences between vibrational states, and it can also exhibit stimulated emission of these photons in the same way electronic states do. Since the emitted photons are coherent with those stimulating the emission, we are able to build molecular lasers, as well as those operating on electronic transitions.

We know from Boltzmann's Law (Chapter 2) that the ratio of the number densities in two energy states under conditions of thermal equilibrium is given by

$$\frac{N_t}{N_m} = \exp\left[\frac{E_m - E_t}{kT} \right] \tag{3-55}$$

where k is Boltzmann's constant and T is the absolute temperature in °K. According to this relationship, at any temperature the density of the lower state is higher than the density of the upper state. The second term of Eq. 3-49 is always negative under these conditions, representing an absorption of power. Radiation

passing through the volume under these conditions is absorbed at a rate proportional to the energy density of the radiation and proportional to the excess density of lower states. The intensity of the radiation thus decreases according to the relationship

$$I_x = I_o \exp[-\alpha x] \tag{3-56}$$

where I_o is the initial intensity, α is the decay constant, and x is the distance traveled by the radiation. The constant α is expressible in terms of the difference in densities of the two energy states as

$$\alpha = \sigma(N_t - N_m) \tag{3-57}$$

where σ is the cross section for stimulated emission or absorption. This constant is related to previously defined parameters by

$$\sigma = \frac{B_{mt}}{c} \tag{3-58}$$

Our interest in these relationships stems from the fact that they hold true whether or not the atoms are in thermal equilibrium. If, by some means, the density of atoms in the upper state can be made larger than the density of those in the lower state, Eq. 3-52 tells us that there is more stimulated emission than absorption, and Eqs. 3-56 and 3-57 tell us that the intensity of radiation passing through the volume grows exponentially rather than decaying. Since the photons generated by stimulated emission are in phase and colinear with the radiation passing through, the volume serves as a *coherent* amplifier for the radiation. (See par. 3-3.2.1.)

3-3.1.3.2.1 *Pumping*

The nonequilibrium condition in which the density (population) of atoms in an energy state is larger than the density of atoms in a lower state, is called a "population inversion", since it is the reverse of the usual distribution of populations. The generation of a population inversion in a medium is known as pumping.

Most excited states of atoms decay by spontaneous emission almost as fast as they can be pumped, and no substantial degree of population inversion can be achieved. There are states, however, from which spontaneous emission tends not to take place, because of the quantum mechanical rules governing transitions. Such transitions are called "forbidden transitions", and the states are known as "metastable states", because they are stable for relatively long periods of time before decaying by spontaneous emission. It is only in these metastable states that a large fraction of the atoms in a volume can be concentrated by pumping.

Atoms are seldom pumped directly into metastable states, as illustrated by simplified energy level diagrams in Fig. 3-51. Usually, the energy accepted by each atom puts it in one of a large number of higher energy states, from which it decays rapidly by transferring energy away nonradiatively until it comes down to the metastable state, from which all transitions are inhibited. To have a large population inversion between the metastable level and some lower state, we now need only make sure that the lower level is not highly populated.

Fig. 3-51 shows pumping from the ground state to the pumping levels, radiationless transition to the metastable level, and transition by stimulated emission through the laser transition to the terminal level. The difference between Fig. 3-51(A) and (B) is that in diagram (B), the terminal level is higher than the ground state, decaying to the ground state through a radiationless transition. Lasers of both types exist. Diagram (A) represents a three-level laser, and (B) shows a four-level laser. The difference may be appreciated when it is recalled that most atoms are ordinarily in the ground state. To obtain a population inversion between the m level and the ground state, it is necessary to put more than half of the total number of atoms into the metastable state since, otherwise, there would be more in the terminal level than in the metastable level. In the four-level laser, on the other hand, all that is required is to put more atoms into the metastable level than there are in the terminal level. If the terminal level decays rapidly to the ground state, an inversion may be obtained with only a small percentage of the total number of atoms excited.

An additional advantage of the four-level laser is that it does not absorb strongly at the laser wavelength when it is not pumped. This fact may be readily understood by referring back to Eq. 3-49, in which the power absorbed by a given electronic transition was given as a func-

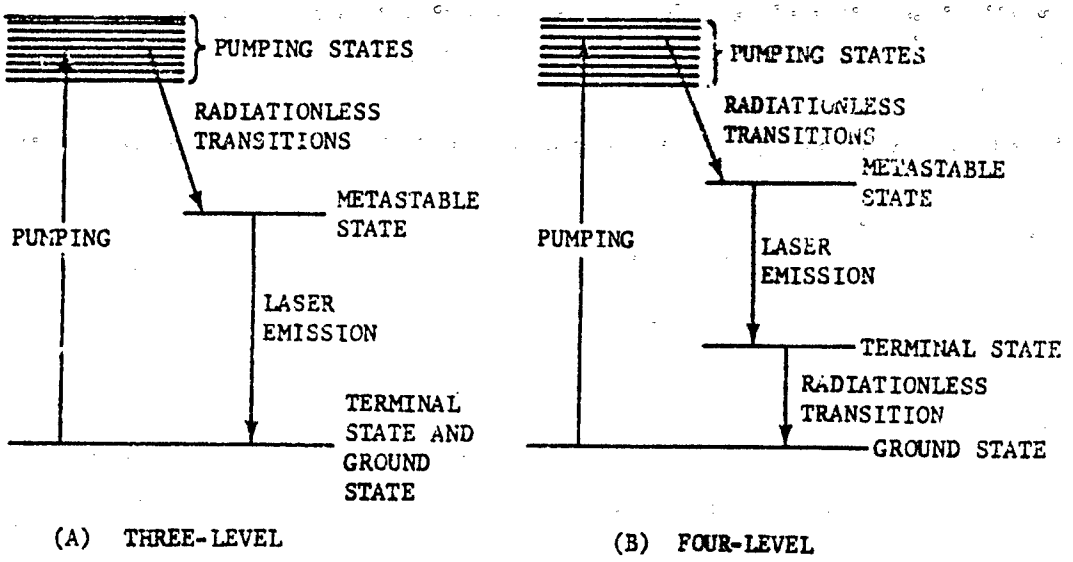


FIGURE 3-51. Simplified Laser Energy-level Diagram

tion of the density of atoms in the lower state of the transition. The three-level laser medium in an unpumped condition has most of its atoms in the ground level which is also the terminal level. It is thus in a position to absorb strongly at the laser wavelength. The four-level laser in the unpumped condition also has most of its atoms in the ground state, but few in the terminal level, so it absorbs very little at the laser wavelength. There is a significant practical difference in the care that must be taken to insure uniform pumping. In the three-level laser, any unpumped or poorly pumped portions of the medium may cause large losses in optical gain. The four-level laser requires much less care in providing for uniform pumping since, although poorly pumped portions of the medium may not contribute as much as they might, they do not create losses.

3-3.1.3.2.2 Optical Cavities

The discussion so far has described how a suitable collection of atoms may be pumped to form a coherent optical amplifier. The most useful laser devices, however, are not amplifiers but oscillators.

One makes an optical amplifier into an oscillator in the same way as any other amplifier is made into an oscillator, i.e., by providing appropriate feedback. If the amplified radiation is returned to the volume for additional amplification, and the fraction fed back is large enough to overcome the losses in the loop, the arrangement becomes an oscillator. As a rule, the devices known as "lasers" are optical oscillators.

Optical feedback is most conveniently provided by means of mirrors. A mirror having the desired reflectivity at the frequency of the radiation is placed at each end of the volume having optical gain. The mirrors are aligned so that the radiation makes reflections back and forth between them, being amplified by the active medium each time

A pair of mirrors arranged so that radiation reflects back and forth between them is a resonant cavity, similar in many respect to a microwave cavity. The principal difference is that optical wavelengths are smaller than the cavity dimensions by several orders of magnitude, while microwave cavities are typically of the order of a wavelength in size. One consequence of this difference is that diffraction plays a much smaller role in optical cavities, and they may be open, rather than completely closed as microwave cavities are. In both cases, the radiation fields within the cavity must be a

solution of Maxwell's Equation[40]. A consequence of this is that a laser oscillator can oscillate only at discrete frequencies set by the size and shape of its optical cavity. Corresponding to each resonant frequency is a spatial distribution of the electromagnetic fields in the cavity. The resonant frequencies and the spatial distributions are both spoken of as "modes of the cavity".

A convenient way of thinking about the modes of a cavity is that any solution to Maxwell's Equations for the boundary conditions represented by the cavity must be "self-reproducing". That is, if a wave of radiation of a particular frequency and a particular spatial amplitude distribution is launched in a given direction in the cavity, the frequency and spatial distribution are those of a cavity mode only if, at a later time, the wave returns to the same position with the same phase and the same amplitude distribution, differing at the most from the original wave by a constant multiplier.

Consider a wave launched parallel to the longest dimension of the cavity, which for optical cavities tends to be much larger than the other dimensions. The wave propagates to one of the mirrors of the cavity, is reflected back along the axis, reflected again from the second mirror, and finally reaches the starting point. If we assume an exact phase reversal at each mirror, the requirement that the phase be the same as the starting phase may be expressed

$$n\lambda = 2L \tag{3-59}$$

where n is an integer and L is the separation between the mirrors. Stated in words, the round-trip distance between the mirrors must be an integral number of wavelengths.

For a fixed mirror separation, each value of n corresponds to a different wavelength. The difference between one resonant wavelength and the neighboring one is given by

$$\Delta\lambda = 2L\left(\frac{1}{n} - \frac{1}{n+1}\right) \approx \frac{2L}{n^2} \tag{3-60}$$

Substituting from Eq. 3-59

$$\frac{\Delta\lambda}{\lambda} \approx \frac{1}{n} \tag{3-61}$$

Optical cavities are typically 10^4 to 10^5 wavelengths long, making n a large number, so the resonant wavelengths are closely spaced. An optical cavity thus has many resonant frequencies, given by

$$f_n = \frac{nc}{2L} \tag{3-62}$$

These frequencies are multiples of a fundamental cavity frequency, $c/(2L)$.

There are, of course, many more resonant frequencies corresponding to off-axis modes, but if the axial dimension of the cavity is much larger than the others, these frequencies are only slightly different from those of the axial modes. Any or all of these modes are capable of oscillation if the laser medium has sufficient optical gain at the corresponding frequencies. It has been seen, however, that optical gain is a rare phenomenon, and can be made to exist only at discrete frequencies. How, then, can the match between cavity mode frequencies and laser transition frequencies be achieved?

Although electronic transitions are narrow by standards applied to conventional light sources, they do have a finite width. Even in the absence of external influences, an electron in transition to a lower energy state emits a photon with a finite spectral width, related to the lifetime of the excited state. When excited atoms are subjected to outside influences, such as collisions with other atoms, the emitted photons have much larger spectral widths. If atoms are imbedded in a solid matrix such as a crystal or glass, the fields due to neighboring atoms may widen the emission line considerably by creating a distribution of photon frequencies. If the atoms are in motion, as they are in a gas laser, the motion causes random Doppler shifts in the apparent frequency emitted by individual atoms, giving an effective increase in line width. Because of these mechanisms by which laser transition lines are broadened, many cavity mode frequencies typically fall within the line width, and simultaneous oscillation at many frequencies is the rule rather than the exception. Careful design of the optical cavity is required to restrict a laser to a single operating frequency when that is needed for particular applications.

3.3.1.3.2.3 Condition for Threshold of Oscillation

To achieve a steady-state oscillation of the laser, it is necessary to feed back enough of the emitted power to make up for the losses of the system. Rewrite Eq. 3-56 with a positive sign so that α is an optical gain coefficient rather than an absorption coefficient.

$$I_x = I_o \exp[\alpha x] \tag{3-63}$$

For a laser cavity of length L with incident laser power I_L at the mirror, the total optical gain per pass is

$$\frac{I_L}{I_o} = \exp[\alpha L] \tag{3-64}$$

For simplicity, we assume that the only losses in the cavity are those due to the imperfect reflectivity of the two mirrors and both mirrors have a reflectivity R. The radiation passing through these mirrors is the useful output of the laser. The intensity of the optical wave after reflection from one of the mirrors is

$$I_R = I_L R \tag{3-65}$$

For the case in which the gain during one pass through the cavity just compensates for the loss I_R during reflection from one of the mirrors, we have

$$I_R = I_o \tag{3-66}$$

which gives

$$R = \exp[-\alpha L] \tag{3-67}$$

We may put this in more convenient form by defining a loss factor γ as

$$\gamma = -\ln R \tag{3-68}$$

Substituting in Eq. 3-67

$$\gamma = \alpha L \tag{3-69}$$

This relationship is the threshold condition. It represents the minimum feedback necessary to allow the laser to oscillate. It is clear from Eqs. 3-67 to 3-69 that threshold may be reached at a lower value of α — i.e., for a smaller population inversion, and thus a lower pumping power — if the reflectivity is high. One should not conclude from this that the reflectivity of the mirrors should be as high as possible. The useful power from the laser is the radiation that passes through the mirrors and, if absorption and scattering in the mirrors are neglected, this is given by

$$I_T = I_L(1 - R) \tag{3-70}$$

where I_T is the laser power transmitted through the mirror. To make the output power large for a fixed flux density in the cavity, one should use a small reflectance. The optimum reflectance is a compromise between having a low threshold and coupling out a large fraction of the flux in the cavity. In general, the higher the gain of the laser, the lower the optimum reflectivity of the mirrors.

It is frequently more convenient to have the output of a laser in a single beam instead of having output from both ends. This may be done by making one of the mirrors as close to a perfect reflector as possible and reducing the reflectivity of the other. Eqs. 3-65, 3-66, and 3-67 are still valid for this case of R is taken to be the geometric mean of the two reflectivities. Eq. 3-68 may be generalized by adding terms for absorption, scattering, and diffraction losses.

3.3.1.3.2.4 Heat Dissipation

The efficiency of lasers is generally poor, and most of the energy supplied for pumping is converted to heat. It is necessary to remove this heat because the range of operating temperature for a laser is limited. The gain of a laser decreases with increasing temperature because the width of the atomic transition line increases, and the gain at the peak goes down. The temperature of four-level lasers in particular must be kept low so that thermal population of the terminal laser level does not occur, reducing the population inversion.

The methods for cooling depend greatly upon the type of laser involved, and they will be discussed in the paragraphs dealing with specific lasers.

3.3.1.3.2.5 Cavity Q and Width of Resonance[41]

The Q, or quality factor, of an optical resonator is given approximately, for small losses, by

$$Q = \frac{2\pi f_{opt} t}{\delta} \tag{3-71}$$

where f_{opt} is the optical frequency; t is the transit time from one mirror to the other; and δ

is the fraction of the photons that are lost per transit by diffraction, absorption in the medium filling the cavity, and by transmission through the mirrors.

The Q is related to the sharpness of resonance in the same way as in resonant circuits and microwave cavities, i.e.,

$$\frac{\Delta\lambda}{\lambda} = \frac{1}{Q} \tag{3-72}$$

Thus, for the sharpness of resonance of an optical cavity, we have — using the relationship $f_{opt}t = L/\lambda$, and Eqs. 3-71 and 3-72 —

$$\Delta\lambda = \frac{\lambda^2\delta}{2\pi L} \tag{3-73}$$

where L is the length of the cavity. For high-reflectance mirrors, the width of an optical cavity resonance may be of the order of megahertz.

If the medium inside an optical cavity has gain at a resonant frequency of the cavity, the effect is to compensate for the losses of the cavity. The fraction δ of the photons that are lost per transit is replaced by stimulated emission, and this factor goes to zero. The Q of the system increases greatly, and the width of the resonance decreases correspondingly, perhaps to a few hertz. Under these conditions, Eqs. 3-71 and 3-73 no longer apply since the noise of spontaneous emission plays a dominant role in determining the width of the resonance.

3-3.1.3.2.6 Q-switching

In continuous steady-state oscillation, a laser finds a balance between the rate at which photons leave the cavity and the rate at which they are generated by stimulated emission. The density of photons in the cavity builds up until the depletion of the metastable states of the atoms and the filling up of the terminal states reduce the gain so that Eq. 3-69 is satisfied.

There is an alternate method of operating a laser by which much higher peak power may be generated. This technique, known as Q-switching (or Q-spoiling), consists of temporarily interrupting the optical path inside the laser cavity while the laser is pumped. Several specific methods for doing this will be discussed later. Under these conditions, the population inversion is not depleted by stimulated emission, and the gain may build up to a value limited only by the pumping rate and the rate of spontaneous emission. If the optical path is then suddenly restored, the flux density in the cavity builds up very rapidly. Since the rate of stimulation emission is proportional to the flux density and the emission adds to the flux, a regenerative buildup of flux occurs and the energy stored in the metastable states during the pumping process is released in a very short time, usually 10 to 50 nanosec. Higher peak power is obtainable from a laser by this means than any other, although the efficiency is not as high as in conventional operation.

3-3.1.3.3 Laser Types

The most convenient classification of lasers is: (1) solid-state lasers, (2) gas lasers, (3) injection lasers, and (4) liquid lasers. The term "solid-state" includes all of the crystal and glass host lasers, but does not include injection lasers, which are also made from solids, or liquids that are cooled until they are glass-like. Injection lasers are also called semiconductor lasers. Liquid lasers, usually employing dye solutions, are of relatively little importance—being used principally as tunable sources for absorption spectroscopy. They will not be discussed further.

3-3.1.3.3.1 Solid-state Lasers[42-47]

Solid-state lasers consist of ions embedded in a crystal lattice or glass. The active ions are usually either transition metals or rare earths, and the crystal lattices are usually Al_2O_3, similar oxides, or fluorides. The most common examples are Cr^{+3} in Al_2O_3 (Ruby) and Nd^{+3} in $Y_3Al_5O_{12}$ (YAG) or in glass.

Conventional solid-state lasers are optically pumped. The host material takes no part in the lasing transition although the local fields do perturb the energy levels of the metal ion. Crystal lattices have the advantage of a relatively high thermal conductivity, thus easing the problem of heat dissipation and even allowing continuous wave operation in some cases. One problem is the difficulty of preparing well-ordered, uniformly doped lattices. Glass lasers have the advantage that they can be made into shapes ranging from short, thin fibers a few microns thick to large rods two meters long.

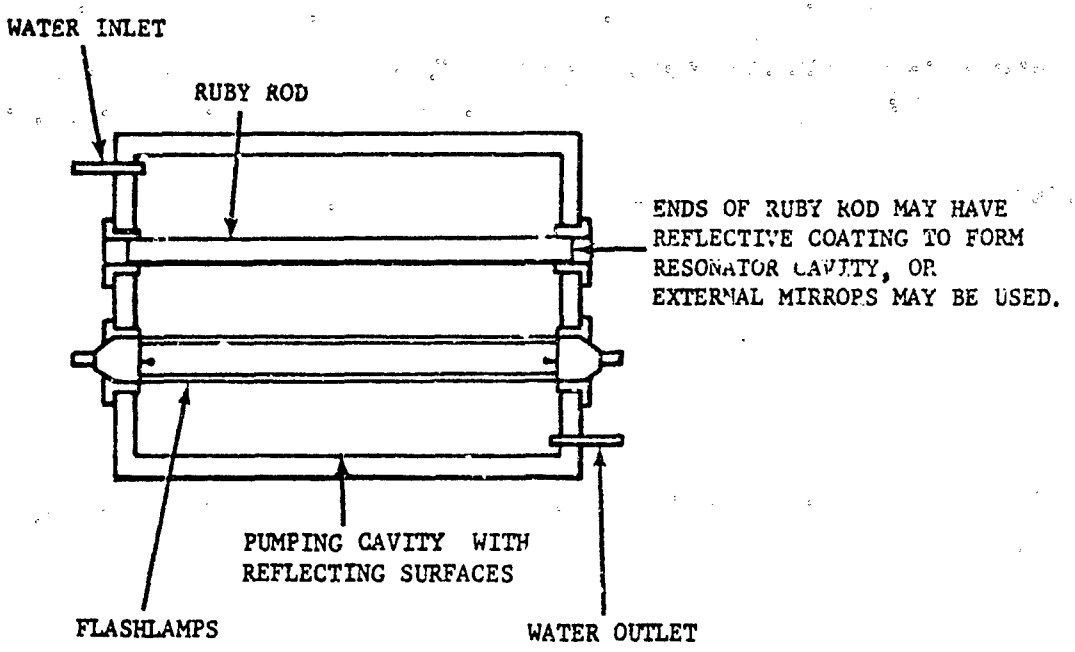


FIGURE 3-52. Outline of Solid-state Laser

Because the lifetimes of metastable states of ions embedded in lattices are relatively long (~1 msec), Q-switching techniques are successfully employed with solid-state lasers.

3-3.1.3.3.1.1 Ruby Lasers

Fig. 3-52 shows the construction of a simple ruby laser. The location of the components and shape of the pumping cavity are such that most of the optical energy from the lamp or lamps is focused onto the laser rod. High-energy linear or helical flash lamps provide intense visible radiation, certain spectral bands of which are absorbed by the active ions. Typical outputs from non-Q-switched ruby lasers are random 1 microsec pulses occurring in bursts from 0.5 to 1.5 msec in duration with total energies between 0.1 to 1000 J. Q-switched output occurs in single pulses lasting from 5 to 50 nanosec and having peak powers from 1 mw to 10 Gw.

The characteristic radiation wavelength at room temperature for ruby is 6943 Å. At liquid nitrogen temperature it shifts to 6934 Å. The measured spectral line width is about 10 picometers (0.1 Å). At any given instant, it is less than this, but local heating during a pulse shifts the spectrum. These thermal effects are due to changes in refractive index and linear expansion of the rod. Ruby is a three-level laser, and will operate continuously only at cryogenic temperatures.

3-3.1.3.3.1.2 Neodymium Ion Lasers

Neodymium ion, in its many hosts, is one of the most versatile for laser construction. The emission wavelength is 1.06 microns in the near infrared. It is a four-level laser.

Neodymium-doped glass has the largest energy storage density of any known laser material. In many respects, it is similar in operation to ruby; in fact, ruby and neodymium-doped glass may be used interchangeably in the same pumping cavity with the same lamps and driving circuit. The fact that glass rods may be made as large as desired, in contrast to the problems involved in growing ruby crystals of large size, has made neodymium the choice for applications requiring very large energy.

In yttrium aluminum garnet (YAG) host, the

neodymium ion allows cw laser action at room temperature. Multimode output up to 50 w at room temperature has been obtained with practical pump lamps, and up to 200 w with exotic pumping arrangements such as a vortex-stabilized plasma arc.

3.3.1.3.3.1.3 Other Ions and Hosts

Table 3-6 provides a summary of some of the more successful ion/host combinations, together with the wavelength type of operation and highest operating temperature achieved.

3.3.1.3.3.2 Gas Lasers[41,42,43,45,48]

Gas lasers are the most useful lasers for laboratory work because they are the most convenient and coherent. Their chief limitation is their inability to produce high peak power.

The attributes of gas lasers are derived from the fact that they are essentially sources of atomic and molecular line spectra, the transitions of which are well known and explicable by quantum mechanical analysis. Since gases are nearly optically homogeneous, theoretical resonator-mode analysis gives a very satisfactory prediction of what is actually observed.

Because of the relatively low density of available atoms within the cavity, as well as the absence of wide absorption bands, gas lasers are rarely optically pumped. The most common form of excitation is collision with electrons in an electric discharge.

Almost any gas has laser potential, and the available wavelengths are numerous. Table 3-7 lists parameters of the more commonly used gas lasers. For the purposes of describing the characteristics and mechanisms of gas lasers, it is convenient to divide them into three categories: neutral atom, ion, and molecular lasers. Representatives of each of these kinds of gas lasers will be discussed.

TABLE 3-6. SELECTED SOLID LASERS

ION	HOST	λ, μ	TYPE OF OPERATION (Pulsed or Continuous Wave)	T, °K
Ho^{+3}	CaF_2	0.55	p	77
Pr^{+3}	LaF_3	0.60	p	77
Eu^{+3}	Y_2O_3	0.61	p	220
Cr^{+3}	Al_2O_3	0.69	p, cw	350
Sm^{+2}	CaF_2	0.71	p	20
Nd^{+3}	$CaWC_4$	0.91	p	77
Yb^{+3}	$Y_3Al_5O_{12}$	1.03	p	77
Nd^{+3}	$Y_3Al_5O_{12}$	1.06	p	440
			cw	360
Tm^{+2}	CaF_2	1.12	p	77
Nd^{+3}	$CaWO_4$	1.34	p	300
Ni^{+2}	MgF_2	1.62	p	77
Er^{+3}	$Y_3Al_5O_{12}$	1.66	p	77
Co^{+2}	MgF_2	1.75	p	77
Tm^{+3}	$Y_3Al_5O_{12}$	2.01	p	300
			cw	77
Ho^{+3}	$Y_3Al_5O_{12}$	2.10	p	300
			cw	77
Dy^{+2}	CaF_2	2.36	p	145
			cw	77
U^{+3}	CaF_2	2.61	p	300
			cw	77

TABLE 3-7. COMMONLY USED GAS LASERS

TYPE	λ, Å	TYPE OF OPERATION (Pulsed or Continuous Wave)	MAXIMUM OUTPUT
Ne	3,324	p, cw	10 mw cw
N_2	3,324	p, cw	200 kw in 10 nanosec
Ar	4,579	p, cw	10 w cw
	4,658		
	4,765		
	4,880		
	5,017		
	5,145		
Kr	4,762	p, cw	5 w cw
	5,208		
	5,682		
	6,471		
	6,474		
Xe	5,419	p, cw	1 w cw
	5,971		
	6,271		
He-Ne	6,328	cw	100 mw cw
	11,523		
	33,912		
He-N_2-CO_2	10,600	cw	5 kw cw

3-3.1.3.3.2.1 Neutral Atom Laser (He-Ne)

Nearly all neutral atom gas lasers operate in the infrared region. The notable exception is the He-Ne red (6328 Å) laser, the most useful laser for work not requiring high power. Lines at 1.15 and 3.39 microns are also available from He-Ne but the visible line, because of its general utility, is usually preferred. The 3.39-micron line competes with the 6328 Å line and must be suppressed to obtain the red emission.

Neutral atom lasers are generally low power devices, a continuous wave output of 100 mw being relatively high. An He-Ne laser with that power is about two meters long and is usually air-cooled.

In the He-Ne laser, collisions with electrons produce excited He metastables, which, in turn, transfer their energy to the upper lasing state of Ne because of a near resonance in energy levels. Collisions, however, also tend to depopulate excited states, so a delicate balance is set up between opposing processes, placing narrow limits on pressures and current densities. The kinetic processes make it impossible to increase the total power output by increasing the tube diameter, or by more energetic pumping.

3-3.1.3.3.2.2 Ion Laser (Ionized Argon)

Laser transitions of the noble gas ions—including argon, krypton, and xenon—fall chiefly in the visible and ultraviolet. The Ar^+ laser is the most useful because a continuous output of several blue-green lines greater than 10 w can be obtained

The active transition of an ion laser is between two energy levels of the ionized atom. The atoms of gas must be first ionized, then excited. A large amount of power is required to maintain the ionization of the gas, so the electrical excitation must be more energetic than for a neutral gas laser. The Ar^+ laser is excited by a high-voltage, high-current discharge, and is usually water-cooled.

A consequence of the relatively high energy level of the terminal state of the laser is that a large fraction of the energy stored in the upper level is necessarily wasted and, because of this fact, the efficiency of this type of laser can

never be high. Nevertheless, ion lasers find many uses in applications not requiring high efficiency.

The principal advantage of the ion laser is the range of wavelengths available. Neutral atom and molecular lasers tend to operate only in the infrared, but ion lasers offer a wide variety of visible lines at substantial power levels. Krypton, in particular, oscillates in so many visible spectral lines simultaneously that it can produce a beam that appears to be white light.

Ionized argon lasers may be operated in a pulsed mode at peak power levels higher than the limit for continuous operation. The lifetime of the upper laser level is not long enough, however, to make Q-switching feasible.

3-3.1.3.3.2.3 Molecular Laser (Carbon Dioxide)

In a molecular laser, stimulated emission takes place between vibrational excitation levels of the molecule. In the carbon dioxide laser, the laser transition is between two vibrational modes of the carbon dioxide molecule, both states lying at relatively low energy levels. The consequence of this fact is that the efficiency is high compared to other lasers. Carbon dioxide lasers can operate at high-power levels with efficiencies of up to 30 percent. Output powers of the order of a kilowatt have been achieved, and commercial units with over 100 w continuous output are available. As a rule, 60 to 80 w are produced for every meter of length.

The carbon dioxide laser is pumped by an electrical discharge in a mixture of gases—usually carbon dioxide, nitrogen, and helium. The electrical discharge excites the nitrogen, which happens to have a metastable level very closely coinciding with a metastable level of the carbon dioxide molecule. An efficient energy transfer occurs between the nitrogen molecule and the carbon dioxide molecule, thereby pumping the upper laser level of carbon dioxide. As the upper level is depleted by stimulated emission, the energy stored in the nitrogen is transferred to it. Helium atoms serve to cool undesired rotational excitation and to depopulate the lower level of carbon dioxide. Water molecules are capable of performing the same function.

Because the lifetime of the metastable state is relatively long (more than a millisecond), the carbon dioxide laser can be Q-switched. Peak power levels one or two orders of magnitude higher than the continuous power level can be achieved.

3-3.1.3.3.3 Injection Lasers[42,43,45]

Injection lasers operate by the injection of electrons into the crystal lattice producing local concentrations of holes and electrons, which, upon recombination, emit radiation. The most common method of injection is to place a forward bias across a p-n junction; for this reason, the term *diode* laser is also often used.

In the injection laser, illustrated schematically in Fig. 3-53, a semiconductor is usually sandwiched between two metal contacts which also serve as heat sinks. One contact lies on the p-type surface, the other on the n-type. The wafer linear dimensions are typically about a millimeter. When a voltage is applied across the p-n junction is a forward direction, excess electrons that move into the p-type region and excess holes that move into the n-type region recombine in the narrow junction region, only a few microns across. The wafer surfaces normal to the junction plane are polished and serve as an optical cavity.

Small size and weight are the chief advantages of diode lasers. In addition, the injection process is relatively efficient. Typical efficiencies of 15 to 20 percent have been achieved with liquid helium and nitrogen cooling. Modulation can be achieved by varying the driving current as well as direct modulation of the output beam. Lifetimes are short, of the order of a nanosecond, so Q-switching is not possible.

The spatial coherence and beam collimation of injection lasers are relatively poor. Typical beam spreads are about 2 deg in the junction plane by 10 deg perpendicular to it.

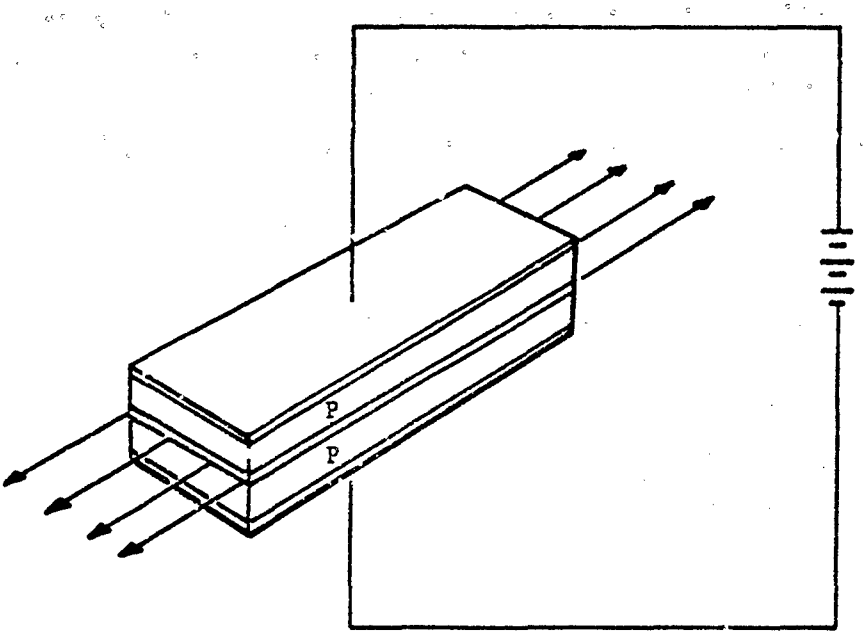


FIGURE 3-53. Injection Laser

Direct laser action has been achieved with the following binary compounds: ZnS (0.33 μ), ZnO (0.37 μ), CdS (0.49 μ), CdSe (0.68 μ), CdTe (0.78 μ), InP (0.91 μ), GaSb (1.6 μ), InAs (3.1 μ), PbS (4.3 μ), InSb (5.2 μ), PbTe (6.5 μ), and PbSe (8.5 μ).

In many cases, two compounds containing a common element can be alloyed to produce a mixed crystal whose radiating wavelength lies in between those of the two pure binary compounds. In this way it is possible to create injection lasers operating at any desired wavelength between the ultraviolet and far infrared. Combinations currently under investigation are: $GaAs_xP_y$ (0.65 to 0.85 μ). Ga_xIn_yAs (0.85 to 3.1 μ), $InAs_xP_y$ (0.91 to 3.1 μ), Ga_xIn_ySb (1.6 to 5.2 μ), PbS_xSe_y (4.3 to 8.5 μ), Zn_xCd_yTe (0.59 to 0.83 μ), Cd_xZn_yTe (0.56 to 0.66 μ), $Cd_{2x}Zn_ySe$ (0.4 to 0.63 μ), and Pb_xSn_yTe (6.5 to 20 μ). The latter family has been made to emit at 14.9 to 15.9 μ, cooled to 12°K. HgCdTe systems also show promise in the far infrared.

3-3.1.3.4 Beam Control Devices

Many laser applications — such as communication, radar, or display — require a means of controlling the light beam by modulating its amplitude, frequency, or phase, or by changing its direction of propagation.

3-3.1.3.4.1 Modulators[43]

A light modulator is a device that controls the amplitude, frequency, or phase of a light beam. There are several types. Some of the principles that will now be discussed as methods of modulation are also applicable to Q-switches and beam deflectors.

3-3.1.3.4.1.1 Mechanical Modulator

The simplest way of modulating a light beam is to interrupt all or a portion of it with a mechanical shutter. Because of the inertia of such a device, however, they are usable only at low frequencies.

3-3.1.3.4.1.2 Frustrated-internal-reflection Modulator

A frustrated-internal-reflection modulator is composed of two prisms in close proximity as shown in Fig. 3-54. When the separation d_p between the prisms is greater than about 1.5λ, Prism (_) acts as a totally internal-reflecting prism. As the surfaces of the two prisms are

brought closer together, the light begins to pass through the junction. If the surfaces of the two prisms were made perfectly flat and brought into complete contact, all of the light would be transmitted through the junction. The surfaces of the prisms cannot be made perfectly flat, of course, and they must also be coated to prevent the surfaces from adhering to each other when brought into contact. As a result, the transmission through the interface never reaches 100 percent.

Piezoelectric transducers, such as barium titanate, are used to control the separation of the prisms. The mechanical inertia of the prisms and drivers limit the response of this type of modulator to low frequencies.

3-3.1.3.4.1.3 Acoustic Modulator

Acoustic waves with frequencies up to hundreds of megahertz have been generated in liquids and solids. The waves are longitudinal (compressional), causing local variations in the index of refraction. The pattern of varying index of refraction acts as a phase grating, diffracting light out of a beam passing through the medium. By varying the power to the acoustic driver, it is possible to change the amount of light diffracted, and thus amplitude-modulate the beam. The frequency response of the modulator is determined by the time required for the acoustic waves to travel across the light beam, and is of the order of 1 megahertz.

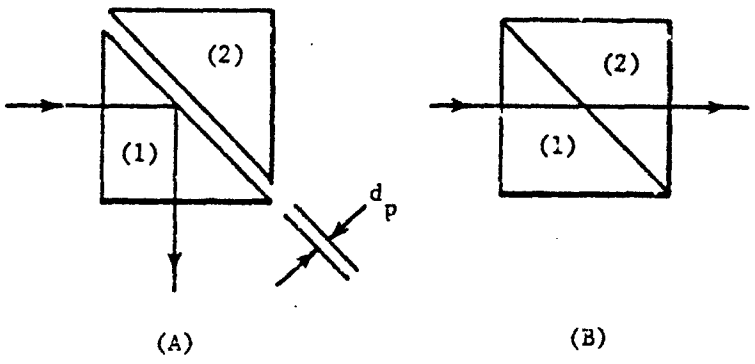


FIGURE 3-54. Frustrated-Internal-Reflection Modulator

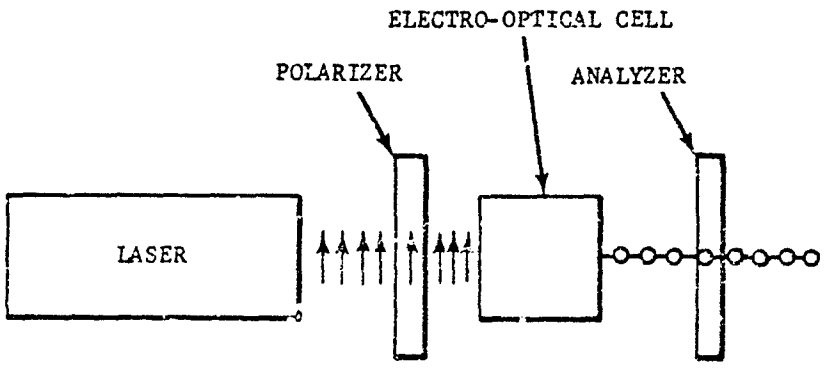


FIGURE 3-55. Electro-optical Modulator

3-3.1.3.4.1.4 Electro-optical Modulator

An electro-optical modulator consists of a polarizer, an electro-optical cell, and an analyzer, as shown in Fig. 3-55. With the polarizer and analyzer in parallel alignment and zero voltage applied to the cell, plane polarized light is transmitted through both. When an electric field is applied to the cell, however, it exhibits birefringence, i.e., the index of refraction in the medium is different for light with one polarization than it is for the other. As a result, a phase shift between the components occurs, equivalent for the proper voltage to a 90-deg rotation of the plane of polarization. At this voltage, therefore, the light is blocked by the analyzer. A polarizer is not required at the input to the modulator if the light from the laser is polarized.

The electro-optical effect works by changing the natural birefringence or by inducing birefringence in certain isotropic solids and liquids in response to an electric field, A number of crystalline solids exhibit an electro-optical effect directly proportional to the electric field. This is called the Pockels effect, or linear electro-optical effect. Crystals having a large Pockels effect are KH_2PO_4 (KDP), CuCl, GaAs, and ZnS.

Other materials, especially liquids, exhibit an electro-optical effect proportional to the square of the applied electric field. This is known as the Kerr electro-optical effect or quadratic electro-optical effect. Examples of useful materials are potassium tantalate niobate ($KTa_{.65}Nb_{.35}O_3$), strontium or barium titanate ($SrTiO$, $BaTiO_3$), and nitrobenzene ($C_6H_5NO_2$).

The electro-optical effect is relatively fast compared to modulators requiring mechanical motion or involving transit times since the effect is electronic or molecular in nature. Modulation frequencies of more than 1 GHz may be achieved.

3-3.1.3.4.1.5 Magneto-optical Modulator

The plane of polarization of light is rotated when traveling through certain materials, such as gallium-doped yttrium iron garnet, placed in magnetic fields. The amount of light passing through an analyzer then depends upon the magnetic field applied. The heavy weight of the magnetic coil and relatively high driving power requirements limit the usefulness of magneto-optical modulators. Modulation frequencies up to 1 MHz can be achieved.

3-3.1.3.4.1.6 Cavity Length Modulator

Frequency modulation of many continuous wave lasers may be accomplished by mounting one of the cavity reflectors on a piezoelectric driver and modulating the length of the cavity. Frequency shifts of tens of megahertz may be obtained at moderate drive voltages.

3-3.1.3.4.2 Q-switches

The principle of Q-switching was discussed in par. 3-3.1.3.2.6. Refer to Eq. 3-69; the equation for threshold of oscillation, a laser may temporarily be prevented from oscillating by decreasing α the gain of the medium or by increasing γ the loss factor. Both methods are used. To be most effective, a Q-switch must be timed so that the laser is allowed to oscillate at the peak of the population inversion.

3-3.1.3.4.2.1 Mechanical Shutter Q-switch

The edge of a rotating chopper wheel placed in the optical cavity of a laser acts as a simple mechanical shutter. The opaque portion of the wheel makes the cavity losses very large. As the open sector moves into the cavity, the losses are reduced to a minimum, permitting a large pulse to be produced.

3-3.1.3.4.2.2 Rotating Reflector Q-switch

An effective Q-switch may be made by replacing one of the cavity reflectors with a total-internal-reflecting roof prism mounted on a rotating shaft, the axis of which is perpendicular to the roof line. The effective reflectivity of the prism is maximum when the prism roof is perpendicular to the optical axis and very small otherwise. When peak population inversion and prism alignment occur simultaneously, maximum pulse power from the laser is generated. A roof prism is used so that precise alignment of the axis of rotation with the optical axis of the cavity is not required.

3-3.1.3.4.2.3 Electro-optical and Magneto-optical Shutter Q-switch

The electro-optical cells and, occasionally, the magneto-optical cells discussed as modulators

are used also as Q-switches.

With the correct voltage applied, the cells, in combination with the polarizer and analyzer, are effectively opaque within the laser cavity. Pumping of the laser rod during this time will produce the maximum population inversion. When the voltage applied to the crystal is suddenly removed, the Q-switch becomes transparent, and the energy stored within the laser rod is emitted as a large pulse of light. Switching of an electro-optical cell takes less than 5 nanosec.

3-3.1.3.4.2.4 Passive Q-switching

Certain opaque materials, upon absorbing a sufficient number of photons, become transparent. If this material is placed inside a laser cavity, it inhibits laser action while the rod is being pumped to a high-population inversion. The growing spontaneous emission of the laser medium then begins to bleach the material. The incipient laser action regeneratively saturates the material, rendering it transparent. The energy stored in the laser is emitted in a single 10- to 15-nanosec pulse.

Passive Q-switches do not require external control or synchronization. The saturation level of the material can be varied by changing its thickness or changing the concentration of absorbers. Therefore, a proper match of pump energy and Q-switch absorption results in threshold at peak inversion.

Passive Q-switch materials include organic solutions and doped glass, which recover quickly, and thin organic-dye films, which do not recover and must be replaced after each use.

3-3.1.3.4.3 Beam Deflection

Laser applications, such as visual display, require a means of controlling and rapidly changing the direction of the output beam. The principal means of doing this are mechanical, acoustical, and electro-optical beam deflectors.

Mechanical deflection may be accomplished by reflecting the beam from a rotating or oscillating mirror or prism. Because of the inertia of the moving parts, these devices cannot provide rapid, arbitrary positioning of a light beam. They can be used in applications where repetitive scans are needed at limited speed.

Acoustic beam deflectors are similar in operation to the acoustic modulators described in par. 3-3.1.3.4.1.3. A train of compression waves is established in a liquid or solid by means of an electro-acoustic transducer. The patterns of varying index of refraction act as a phase grating, diffracting light out of a beam passing through the medium. The concern now is not with the fraction of the light that is diffracted out of the beam, but the direction in which it is diffracted. The direction depends upon the spacing of the grating, which in this case is determined by the frequency at which the cell is driven. One can thus sweep the diffracted beam in angle by sweeping the drive voltage in frequency. The rate at which the angle can be changed is limited by the transit time.

Electro-optic beam deflectors can be made that are capable of high deflection rates, high resolution, and relatively large deflection angles. They can also give immediate access to any display position and can be built with rugged, solid-state components. These devices can be divided into two classes: those which give an analog presentation and those which give a digital presentation.

Analog deflectors depend upon the fact that a voltage applied across the face of certain crystals will produce a change in their index of refraction. The simplest of the analog deflectors is a prism with the electrodes on the triangular faces. At room temperatures potassium tantalate niobate ($KTa_{.65}Nb_{.35}O_3$), which is a mixture of potassium tantalate ($KTaO_3$) and potassium niobate ($KNbO_3$), produces one of the greatest changes in the index of refraction for a given voltage of any material known. Dielectric breakdown in this material limits the deflection to 1 deg and the number of resolvable positions of the beam to 500.

Analog light deflectors have also been made using stacks of prisms of deuterated potassium dihydrogen phosphate (KDP), a material having a large Pockels effect. For small angles, the deflection is linearly proportional to the applied electric field and to the overall length of the structure. Theoretical analysis shows that the number of resolution elements obtainable is directly proportional to the electric field strength, the length of the structure, and an electro-optical constant of the crystal. It is inversely proportional to the wavelength of the

light and the difference between the operating temperature and the Curie temperature of the crystal.

Digital deflectors may be built utilizing a combination of two effects. One is the phenomenon of birefringence, in which an unpolarized collimated light beam passing through a crystal such as calcite, splits into an ordinary and extra-ordinary ray. These rays are linearly polarized with their planes of polarization perpendicular to each other. At normal incidence, the ordinary ray passes straight through the crystal, while the extraordinary ray is laterally displaced by a distance proportional to the length of the path within the crystal. The second effect is the longitudinal electro-optical Pockels effect discussed earlier.

The basic principle of operation of a digital light deflector is shown in Fig. 3-56. A small, linearly polarized beam of light is incident on a series of electro-optical switches S_n and birefringent crystals B_m. The initial polarization is normal to the plane of the paper. If all switches are left in the open position, the light beam (as an ordinary ray) traverses all electro-optic switches and birefringent crystals undeflected and emerges at Point P_1. If, on the other hand, the half wavelength voltage $V_\lambda/2$ is applied to A_2, the polarization of the beam is rotated 90 deg and it passes through crystal B_2 as an extraordinary ray. The beam is, therefore, displaced laterally by an amount proportional to the width of B_2. As there is no voltage applied to A_3, the beam passes unchanged. Consequently, it also passes through crystal B_3 as an extraordinary ray where it is further deflected, arriving at P_2. In the digital deflector, the thickness of the crystals increases in a binary sequence from stage to stage. In this manner, a maximum of different output positions can be switched by a minimum of switches; that is 2^n positions can be controlled by n switches.

To obtain a two-dimensional scan it is necessary to place a second set of switches and crystals in the light path such that the direction of deflection is 90 deg from the first. A polarization correction half-wave plate is inserted between the x and y blanks to make the x and y binary positions independent of each other.

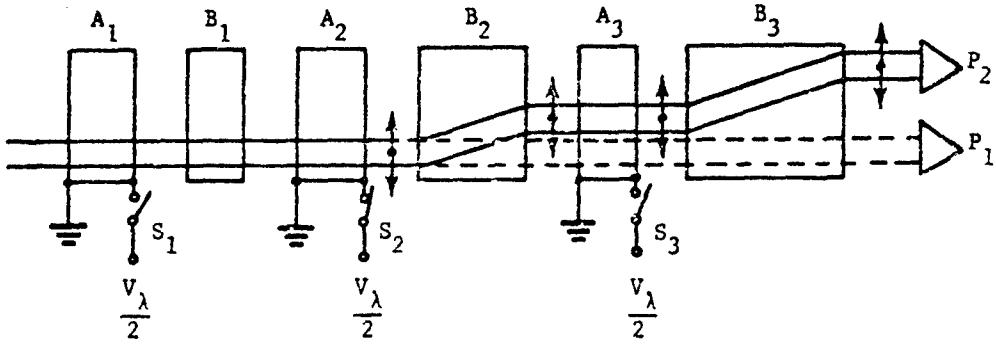


FIGURE 3-56. Digital Deflector

3-3.1.3.5 Nonlinear Optics[42,43]

3-3.1.3.5.1 Harmonic Conversion

When light propagates through materials, its oscillating electric and magnetic fields interact with the electronic charges in the atoms of the material. The electronic charges oscillate in response to the electric field and act as radiating dipoles. At low electric field strengths as encountered in normal incoherent polychromatic light, this effect is linear and is the cause of Rayleigh scattering. At the very high field strengths encountered in laser beams, the coupling of the fields to the electronic charges becomes nonlinear, causing the dipoles to have radiation components at harmonics of the original light frequency. The efficiencies of conversion of light to its harmonics is a function of the field strength. Efficiencies achievable at present are 20 percent and 1 percent, respectively, for second and third harmonic generation from laser beams of 500 Mw power level. It is also possible to mix the outputs of two or more lasers using these nonlinear effects, generating sum and difference frequencies.

3-3.1.3.5.2 Raman Conversion

Another nonlinear phenomenon used to change the wavelength of laser light is the Raman Effect. The vibrational energy of a medium, usually a liquid, is coupled with the laser output to produce new frequencies. These frequencies are displaced from the primary-beam frequency by the Raman frequencies of the medium. Conversion efficiencies of 10 percent are obtainable. If the Raman cell is placed inside the cavity of a Q-switched laser, conversion efficiencies of 50 percent can be obtained. This type of laser is called a Raman Laser.

3-3.2 DESIGN CONSIDERATIONS

3-3.2.1 Coherence

Infrared radiation may be characterized by the degree of orderliness or "coherence" of the fields. If the radiation arriving at a point has a narrow spectrum of width Δf centered at a frequency f, the oscillation may be considered to be sinusoidal, but with random fluctuations in amplitude and phase, the rapidity of which is a function of the width of the spectrum. The narrower the spectrum, the longer the average period of time between the fluctuations. If the spectrum of the radiation is limited to Δf, the radiation may be considered to be a pure sinusoid for a period of time

$$t_c = \frac{1}{\Delta f} \tag{3-74}$$

This time is known as the "coherence time". Within this time, the oscillations of the field may be assumed to be perfectly correlated.

If the phase of radiation passing a point is correlated for a period of time t_c, it is also correlated over the distance the radiation travels during that time. This distance, called the "coherence length", is given by

$$L_c = ct_c \tag{3-75}$$

This distance would be equal to the length of each photon or wave train.

Considering the fluctuation of the radiation at a point to be a stationary random process, one may ask how well the amplitude and phase of the oscillation at one time correlates to the amplitude and phase at other times. One may form, for example, the autocorrelation function of the electric field,

$$\Gamma_{1,1}(\tau) = \langle E_1(t+\tau)\, E_1^*(t) \rangle \tag{3-76}$$

where E_1^* is the complex conjugate of E_1, and the sharp brackets denote a time average. This quantity is also called the "self-coherence" of the field at the point and its value is a function of τ, a measure of the temporal coherence of the oscillation.

Carrying the analysis one step further, one may compute the cross-correlation function between the electric fields in two different radiation fields.

$$\Gamma_{1,2}(\tau) = \langle E_1(t+\tau)\, E_2^*(t) \rangle \tag{3-77}$$

This function is known as the "mutual coherence function" of the radiation in the two waves. When this function is normalized by dividing it by the geometric mean of the two intensities, it is called the "complex degree of coherence" of the radiation of the two waves.

$$\gamma_{1,2}(\tau) = \frac{\Gamma_{1,2}(\tau)}{\sqrt{I_1 I_2}} \tag{3-78}$$

It is the real part of this complex quantity, $\gamma_{1,2}(\tau)$, that determines the nature of the interference at a point.

Suppose the two radiation waves have come from the same source, starting off in phase with each other, but have traveled different distances, S_1 and S_2. The intensity of the light as a function of the difference in path length may be written in the form

$$I(x) = I_1(x) + I_2(x) + 2\sqrt{I_1(x)I_2(x)}\ \gamma_{1,2}(\tau)\left(\frac{S_2 - S_1}{c}\right) \qquad (3\text{-}79)$$

The difference in path length, $S_2 - S_1$, divided by the speed of light, is equivalent to τ in Eq. 3-78.

The degree of coherence is a number that varies between plus 1 and minus 1. If the two waves are exactly in phase and are nearly monochromatic, the value is plus 1. If the two waves are exactly out of phase, the value of the degree of coherence is minus 1. If two nearly monochromatic waves of the same frequency and polarization and equal intensity meet at a point, Eq. 3-79 says that the resultant intensity is between zero and four times the intensity of one of the waves alone. The maximum is four times the intensity of one alone, because the resultant intensity is proportional to the square of the resultant electric field, and the resultant electric field is twice as large as for only one wave present. The minimum intensity is zero, because the electric field vectors of the two waves may exactly cancel.

If two waves meeting at a point are completely uncorrelated, the value of the degree of coherence is zero, and Eq. 3-79 says that the resultant intensity is equal to the sum of the intensities of the two waves.

If two monochromatic beams of equal amplitude and frequency are brought together at an angle, the degree of coherence between the beams will vary between minus 1 and plus 1 throughout the volume in which they are superimposed. If the two beams are from the same source, the difference in phase of the two beams as a function of position may be due entirely to the difference in path length from the source to each point in the volume along the two paths. The differences in phase may, however, be due to differences in refractive index of the media in

the two paths. In either case, there is a spatial pattern of regions in which the electric vectors cancel (destructive interference). The loci of constructive and destructive interference are known as interference fringes. In the most general case, the fringes consist of surfaces on which the electric vectors of the two radiation fields add in phase or out of phase. If the radiation impinges on a surface, such as a screen, the intersection of the interference surfaces with the screen causes the appearance of bands of high and low intensity on the screen. The term "fringes" is most commonly applied to these bands. Because the electric and magnetic fields that make up infrared radiation are not directly observable, all that is known about them is deduced from the behavior of fringe patterns in various optical experiments.

Since the intensity variations are directly related to the degree of coherence, they may be used to measure it. The ratio of the amplitude of the modulation of the intensity to the average intensity along a path cutting through the fringes is defined as the "visibility" of the fringes

$$V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \qquad (3\text{-}80)$$

If the amplitudes of two interfering waves are equal, the visibility of the fringes is equal to the degree of coherence of the two waves.

The property of radiation discussed above is temporal coherence. It is a measure of the degree of correlation between the phase of a wave at one time and the phase at a later time, or the correlation between the phases of two waves that have traveled different paths since they were in phase. Also, consider the correlation between the phase of one point in a wave and another point of the wave in the same transverse plane. It is often important to know over what area in the transverse plane the phase of a wave is strongly correlated.

The area A_c over which the phase of a wave is correlated is simply related to the solid angle Ω through which the wave arrived at the area. The relation

$$A_c = \frac{\lambda^2}{\Omega} \qquad (3\text{-}81)$$

may be made plausible by considering two plane waves of the same frequency superimposed on a plane. The phase difference between the two waves, at a point, is a function of position on the plane. If a distance is chosen on the plane within which the phase difference is relatively constant, the length of the distance is inversely proportional to the angle at which the two beams meet. If this argument is extended to light that is convergent in two dimensions, Eq. 3-81 results. Or—expressing Eq. 3-81 differently—if two waves strike a surface at different angles; the smaller the angular divergence between these waves, the larger is the area over which the phase of these waves will be correlated.

The term that denotes the degree of correlation between the phases of radiation at different points in the transverse plane is "spatial coherence". Spatial coherence of radiation depends upon the size of the source and distance from the source to the point of observation of the radiation.

Radiation falling on an area A is spatially coherent if the source is small enough or distant enough to appear to be a point source. A test of this condition is whether a telescope at the position of area A and having an objective lens the size of area A could resolve the source. Naturally, the larger the area in question, the more severe would this test be

Although spatial and temporal coherence may be separated for purposes of discussion, both are required for the production of interference fringes. (It was tacitly assumed in the discussion of the visibility of fringes that the two waves that were interfering had uniform phase in their transverse planes.) The visibility of the fringes is determined jointly by the spatial and temporal coherence, and the degree of coherence must include both properties of the waves.

Stimulated emission of radiation, the mechanism responsible for the generation of laser radiation, results in the generation of new photons with the same direction of propagation and the same phase as the photons inducing the emission. It is clear from the discussion of temporal and spatial coherence that the new photons are coherent with the original ones. The laser is thus adapted to the generation of beams of radiation with unusual ability to make interference fringes. Because of the narrowness of the

spectrum emitted by a laser, high-contrast fringes may be observed over very large distances compared to fringes generated by conventional sources. Because the radiation is generated with a very small divergence angle, it is equivalent to a very distant point source and the corresponding area of correlation in the transverse plane may be very large compared to that of conventional sources.

The limitations on the achievable temperatures of thermal sources, the broad spectrum typical of thermal sources at high temperatures, and the diffuse radiation pattern make it impossible for thermal sources to generate an intense beam of high spatial and temporal coherence, no matter what sort of spectral filters or optical system is used.

3-3.2.2 Optical Design for Incoherent and Coherent Sources

Incoherent sources typically radiate in all directions and have a radiance limited by the temperature of the source. The problem of maximizing the amount of radiation power that can be coupled out of the source for a specific application involves collecting the radiation in the largest possible solid angle, measured from the source, and designing for optimum utilization of the radiation collected. The optics are usually designed to form an image of the source at the plane to be illuminated.

For a fixed-source temperature, the total radiated power is proportional to the source area. A large source area tends to produce a divergent beam, unless a correspondingly large focal length is employed. Unless the aperture of the collimating telescope is also large, the advantage of a larger source area will not be realized. Thus, if the source radiance is fixed, increased beam power come only through scaling of the entire system to larger size.

Optical design for laser systems is quite different. Because of the spatial coherence of laser emission, one must focus on the apparent source of the radiation, not on the laser aperture.

For example, consider a point source radiating in all directions. The surfaces of constant phase from this source are spherical, with the center of the spheres at the location of the source. As the spheres become very large, the curvature of the spherical surface can be distinguished from a plane only if measurements are made over a large area.

Suppose a plane wave could be generated, i.e., an optical wave whose surface of constant phase is a plane. If this wave were perceived by the eye, it would appear to be a very distant point of light. If the radiation were collected and passed through any optical system, it could be treated exactly as if the source of the radiation were a distant point. If a spherical wave were generated, it would appear that the radiation comes from a point located at the center of the sphere. In other words, the apparent source of the radiation is determined by the curvature of the wavefronts and may not be related at all to the physical size or location of the actual generator of the radiation.

The uniqueness of lasers is that they are capable of generating plane or spherical waves that are uniform in phase over the entire aperture. Not all laser radiation has complete spatial coherence, however. Therefore, two different kinds of design problems may be distinguished; namely, the single-transverse mode, or diffraction-limited laser and the multimode laser.

In the single-mode laser, the radiation has a constant phase over the entire aperture of the laser; consequently, it is equivalent to an ideal point source, and the radiation may be collimated by one lens if the apparent location of the point source is placed at the focus of the lens.

The apparent position of the point source depends upon the resonator configuration and the distance between the collimating optics and the resonator. The apparent source for a laser operating in a single transverse mode in a Fabry-Perot cavity (parallel flat mirrors) is a point at infinity if the collimating optics are close to the laser. As the optics are moved away from the laser, diffraction—due to the finite size of the aperture—begins to have an effect, and the apparent position of the point source moves according to the relation

$$d = Z\left[1 + \left(\frac{\pi W_0}{\lambda Z}\right)^2\right] \qquad (3\text{-}82)$$

where d is the distance from the optics to the position of the point source, Z is the distance from the optics to the output mirror of the laser, and W_0 is the radius of the mode at the output mirror.

For a hemispherical cavity, the apparent source is a point at the center of curvature of the spherical mirror, regardless of the position of the collimating optics.

For a spherical or convex-concave cavity, the apparent source is a point at the center of curvature of the mirrors, regardless of the position of the observer.

For a confocal cavity, the apparent position of the point source for optics close to the laser is the center of curvature of the closer mirror. As the collimating optics are moved away from the laser, the position of the apparent source moves according to Eq. 3-82, where Z is now the distance from the collimating optics to the common focus of the cavity mirrors, and W_0 is the mode radius at the focus.

The effective diameter of the source is zero, so the collimation that can be achieved or the diameter of the beam that it can be reduced to, or the size of the spot it can be focused to, is limited only by diffraction and the quality of the optics.

It is in dealing with such a source that the most stringent requirements on the quality of optical components arise. Very few optical components come close to being diffraction-limited. In many cases, the operation of a system may depend upon maintaining uniform phase across the optical wavefront. In these applications, the usual errors in a lens—rather than just increasing the diameter of the circle of confusion slightly—may make the lens unusable. Bubbles and inclusions in the glass that would be unnoticed with incoherent light may cause unacceptable diffraction effects such as concentric rings and fringes.

If we assume diffraction-limited optics, a beam may be formed with any desired degree of collimation. The area A of the required lens or mirror is given by the spatial coherence relation:

$$A = \frac{\lambda^2}{\Omega} \qquad (3\text{-}83)$$

where Ω is the solid angle to be occupied by the beam and λ is the wavelength.

The multi-mode case requires an altogether different approach. The beam divergence in this case is not due to diffraction alone, but is due to the presence of off-axis modes in the output beam. Off-axis modes, each mode having a small divergence, may diverge from each other much more rapidly. A single lens could collimate all of the individual modes, but each mode would then be collimated in a different direction.

One approach is to consider the collection of point sources corresponding to the off-axis modes as the "source" and design the system accordingly. The divergence of the beam may be reduced to any desired extent by the use of a Gallilean telescope with the beam entering at the concave lens and leaving at the convex lens. The divergence of the beam is reduced by a factor equal to the magnification of the telescope, the diameter of the beam at the objective lens being increased by the same factor.

The design may be strongly affected by the radiation pattern from each of these points, however. They do not radiate isotropically; in fact, the beam may be limited to a very small angle compared to conventional light sources.

Some characteristics of the laser tend to make optical design easier, compensating to some extent for the other problems. The spectrum of a laser is typically extremely narrow and optical components need not be achromatized. Components may also be more effectively anti-reflection coated for a single frequency and, as a result, lens designs involving more surfaces may be tolerated. These two factors may allow more advantageous balancing of the aberrations of a lens for laser system use.

3-4 DETECTORS

3-4.1 DETECTOR TERMINOLOGY

An infrared detector, in the most general sense, is any device which indicates the presence of incident infrared radiation. The detectors which are commonly used in electro-optical infrared systems generally yield an electrical signal to indicate the presence of infrared radiation. The terminology which applies to infrared detectors is defined and explained in Table 3 8.

TABLE 3-8. IR DETECTOR TERMINOLOGY FOR PAR. 3-4

TERM	SYMBOL	PREFERRED UNIT	DEFINITION	RELATIONSHIP
Detector Area	A_D	cm^2	For most detectors, the area between the electrodes. For detectors using integrating chambers, the entrance aperture area is usually the effective area.	
Detector Temperature	T_D	°K	Temperature at which detector operates	
Field of View	θ	deg	For conical fields of view, with the cone centered on the normal to the detector face, θ is the full cone apex angle	
	Ω	sr	Solid angular field-of-view, usually an effective solid angle which is weighted by Lambertian or other distributions, rather than the actual solid angle. When the field-of-view has circular symmetry, $\Omega = \pi \sin^2\left(\frac{\theta}{2}\right)$	
Background Temperature	T_B	°K	Effective temperature of the radiative background of the detector	
*Noise Equivalent Bandwidth	Δf	Hz	The bandwidth of white noise whose integrated power within that band is equal to the integrated power of the actual noise. (A white noise is defined as a noise whose power spectra is constant in frequency.)	

* For a white noise and an ideal bandpass filter, the filter bandwidth and noise equivalent bandwidth are the same. However for non-white noise or practical filter, the filter bandwidth and noise equivalent bandwidth are different.
The noise spectrum of most IR detectors and systems are white or close to white (with the exception of the low frequency band below 10 or 20 Hz). The filter bandwidth and noise equivalent bandwidth have, therefore, almost the same values.

TABLE 3-8. IR DETECTOR TERMINOLOGY FOR PAR. 3-4 (Continued)

TERM	SYMBOL	PREFERRED UNIT	DEFINITION	RELATIONSHIP
Resistance, Static	$\bar{R}$	ohm	Ratio of dc voltage across detector to dc current through it	$\bar{R} = E_{dc}/I_{dc}$
Resistance, Dynamic	$\bar{R}_d$	ohm	The slope of the voltage-current curve at bias voltage E_1	$\bar{R}_d = \left(\frac{dE}{dI}\right) F_1$
Blackbody Area	A_{BB}	cm^2	Emitting area of blackbody source	
Blackbody Temperature	T_{BB}	°K	Temperature of blackbody source	
RMS Signal Voltage	$V_{s,\,rms}$	V	Root-mean-square value of signal voltage at detector terminals, resulting from infrared signal power	
RMS Noise Voltage	$V_{n,\,rms}$	V	Root-mean-square value of noise voltage at detector terminals in bandwidth Δf	
Spectral Responsivity	$\mathcal{R}_\lambda$	$\frac{V}{w}$	Ratio of rms signal voltage at detector leads to rms monochromatic infrared signal power at wavelength λ	$\mathcal{R}_\lambda = \frac{V_{s,\,rms}}{H_{rms}A_D}$
Blackbody Responsivity	$\mathcal{R}_{BB}$	$\frac{V}{w}$	Same as above, except for use of blackbody irradiance	$\mathcal{R}_{BB} = \frac{V_{s,\,rms}}{H_{BB,rms}A_D}$
Irradiance	H	$w\,cm^{-2}$	Radiant power density incident upon a receiving plane	
Spectral Noise-equivalent Power	NEP_λ	w	The amount of rms monochromatic signal power incident upon a detector which produces a signal-to-noise ratio of unity in a detector. The values of f, Δf, λ, A_D, and photon background conditions must be specified.	$NER_\lambda = \frac{H_{rms} \times A_D}{V_{s,\,rms}/V_{n,\,rms}}$

TABLE 3-8. IR DETECTOR TERMINOLOGY FOR PAR. 3-4 (Continued)

TERM	SYMBOL	PREFERRED UNIT	DEFINITION	RELATIONSHIP
Blackbody Noise-equivalent Power	NEP_{BB}	w	The amount of rms blackbody signal power incident upon a detector which produces a signal-to-noise ratio of unity in the detector. The values of f, Δf, T_{BB}, A_D and photon background conditions must be specified.	$NEP_{BB} = \frac{H_{BB,rms} \times A_D}{V_{s,rms}/V_{n,rms}}$
Spectral Detectivity	D_λ	w^{-1}	Reciprocal of NEP. Same conditions as specified in NEP_λ	$D_\lambda = \frac{1}{NEP_\lambda}$
Blackbody Detectivity	D_{BB}	w^{-1}	Reciprocal of NEP_{BB}. Same conditions as specified in NEP_{BB}	$D_{BB} = \frac{1}{NEP_{BB}}$
Spectral D-star	$D^*(\lambda, f)$	cm $Hz^{1/2}$ w^{-1}	A normalization of spectral detectivity with respect to Δf and A_D. Background conditions must be specified for background-noise limited detectors.	$D^*(\lambda,f) = D_\lambda \sqrt{A_D \Delta f}$
Blackbody D-star	$D^*(T_{BB}, f)$	cm $Hz^{1/2} w^{-1}$	A normalization of D_{BB} with respect to Δf and A_D. Background conditions must be specified for background-noise limited detectors.	$D^*(T_{BB},f) = D_{BB}\sqrt{A_D \Delta f}$
Spectral D-double star	$D^{**}(\lambda,f)$	cm $Hz^{1/2} w^{-1}$	Only used for background-noise limited detectors. Normalizes D^* to 180° field-of-view.	$D^{**}(\lambda,f) = D^*(\lambda,f) \sin\frac{\theta}{2}$ For circularly symmetric conical fields-of-view, centered on the normal to the detector face. θ = full cone angle.
Peak Wavelength	λ_p	micron	Wavelength of maximum D^*	

TABLE 3-8. IR DETECTOR TERMINOLOGY FOR PAR. 3-4 (Continued)

TERM	SYMBOL	PREFERRED UNIT	DEFINITION	RELATIONSHIP
Cut-off Wavelength	λ_c	micron	Wavelength at which D^* has degraded to 50% of peak value.	
Responsive Quantum Efficiency	η_R	dimensionless	Ratio of the number of countable output electrons to number of incident photons	
Detective Quantum Efficiency	η_D	dimensionless	The square of the ratio of measured D^* to theoretical limit D^*	$\eta_D = \left(\frac{D^* measure}{D^* theoretical}\right)^2$
Noise Spectrum			Plot of $V_{n,\,rms}$ versus frequency	See par. 3-4.3.6 for noise spectrum discussion.
Modulation Factor	MF	dimensionless	Ratio of rms value of fundamental component of modulated irradiance to the peak-to-peak value of the modulated irradiance	For sine wave chopping, $MF = 0.35$ For square wave chopping, $MF = 0.45$

3-4.2 DETECTOR CLASSIFICATIONS

Detectors can be divided into two basic categories: quantum detectors and thermal detectors. In an ideal quantum detector, each individual incident photon that is absorbed produces one or more countable electrons in the external measuring circuit. In a thermal detector, the energy of an incident, absorbed photon is used to heat the entire detector and thus indirectly cause an externally-measurable effect such as a change in resistance, voltage, or volumetric length in liquid thermometers.

Since thermal detectors are sensitive to the amount of incident *energy* absorbed, their spectral detectivity is relatively flat. On the contrary, the detectivity of quantum detectors, being a function of the incident *number* of photons absorbed, increases as the energy per photon decreases. Finally at the long-wavelength limit, the photon has too little energy to cause any excitation at all.

3-4.2.1 Quantum Detectors

3-4.2.1.1 Photoemissive Detectors

a. Vacuum Photodiode. A vacuum photodiode (Fig. 3-57(A)) is a vacuum-tube diode which contains a thin wire anode and a specially coated cold photocathode. Incident radiation on the photocathode excites electrons into the vacuum (Fig. 3-58). The electrons are accelerated to the anode by the potential difference (or electrical field) across the two electrodes.

The photocathode is coated with a special (*low-work function*) material to facilitate electron emission. A coating material designated S-1 is useful in the near infrared region. The response time of vacuum photodiodes, being essentially the transit time, is extremely fast (a few nanoseconds)[49].

b. Photomultiplier. A photomultiplier is a photodiode which has a built-in electron amplifier. Electrons released at the photocathode of a vacuum photodiode are attracted to a dynode. A dynode is a specially coated anode which emits more than one electron into the vacuum per each bombarding electron. These electrons, in turn, are accelerated to a second dynode, etc. A gain of 10^6 is possible using eight consecutive dynodes[49]. The dynode gains are achieved under conditions of high input impedance with far less noise than in a conventional preamplifier. Photomultipliers also have fast response times (a few nanoseconds) being essentially the transit time.

c. Channeltron. A channeltron is a narrow, evacuated glass tube with a photocathode at one end and an anode at the other. Electrons emitted into the vacuum are accelerated down the tube. Due to lack of collimation, the electrons bounce off the wall of the tube. Since the wall is coated with a dynode coating, more electrons are emitted from the wall than bounce into it, thus producing gain as in a photomultiplier. Gains as high as 10^4 are practicable. Bundles of hollow-glass-fiber channeltrons can be used to make an *image* photomultiplier.

3-4.2.1.2 Photoconductive Detectors

Photoconductive detectors are made from semiconductor materials. The resistance of the semiconductor decreases due to electrons and/or holes liberated by absorption of incident photons (Fig. 3-57(B)).

At absolute zero in an intrinsic semiconductor all the electron states in the valence band are filled and all the electron states in the conduction band are empty (Fig. 3-59). The lowest conduction band energy level and the highest valence band energy level are separated by a forbidden energy band called the band gap E_g. In this condition, the semiconductor cannot conduct.

At temperatures above absolute zero, some of the valence band electrons are thermally excited into the conduction band leaving "holes" behind in the valence band (Fig. 3-60). Both the conduction band electrons and the valence band holes are free to conduct.

An optical excitation occurs when the energy from an incident photon is absorbed by an electron. In an intrinsic semiconductor this produces a free electron and a free hole (Fig. 3-61). For efficient detection it is required that the rate of thermal excitations be small compared to the rate of optical excitations. (This is true for *all* quantum detectors.) The optical excitation of an electron into the conduction band where it can conduct freely under the influence of an electric field (Fig. 3-62) makes the semiconductor an analog of the vacuum photodiode. Note, however, that the holes also can conduct.

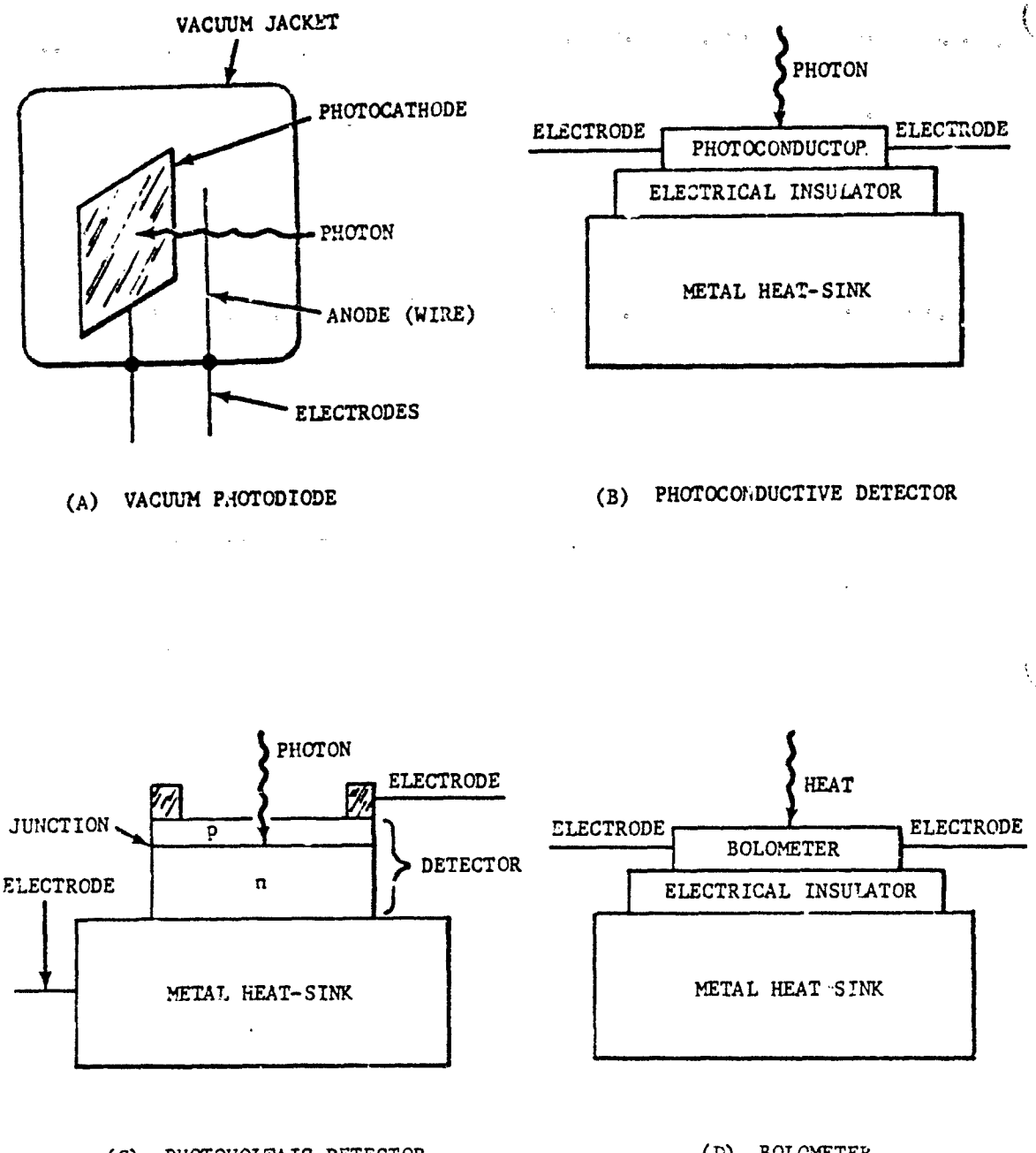


FIGURE 3-57. Schematic Diagram Showing Basic Types of Detectors

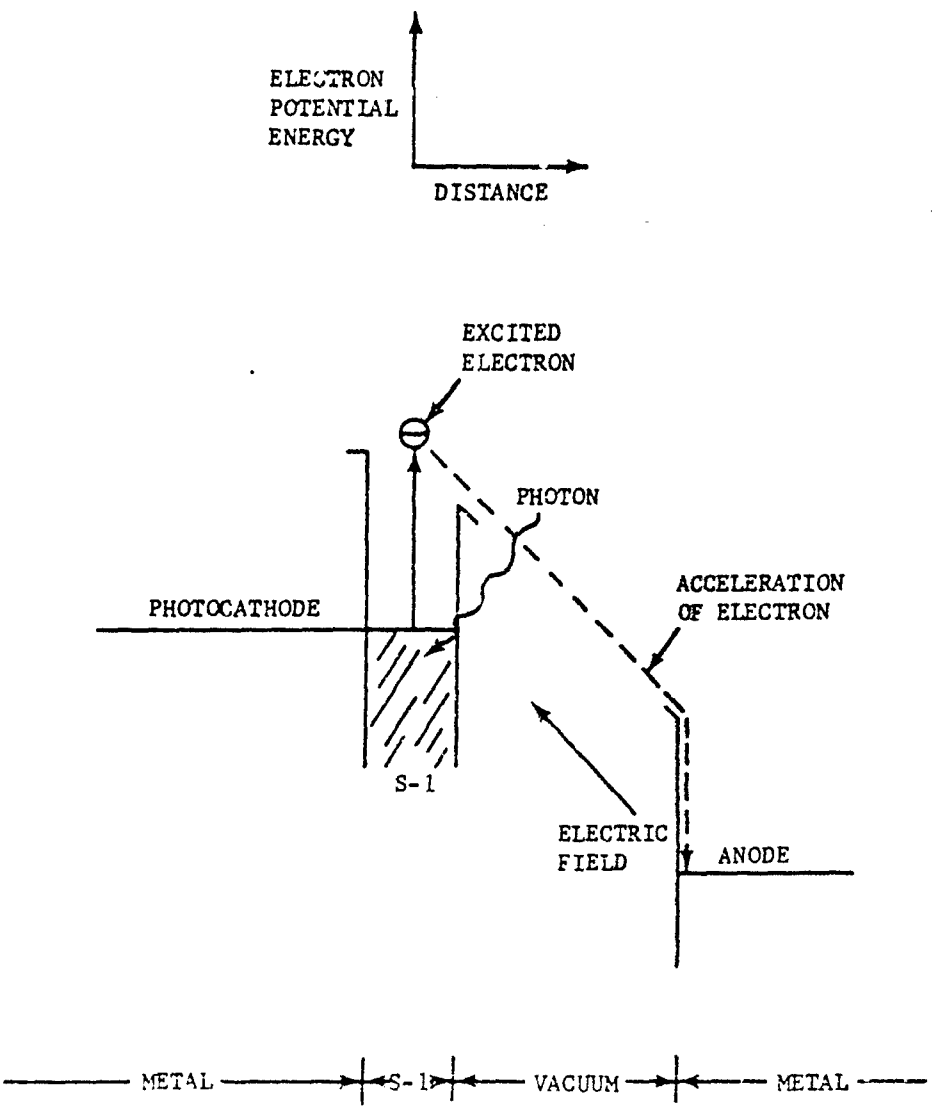


FIGURE 3-58. Electron Excited Into Vacuum by IR Radiation Impinging on Photocathode

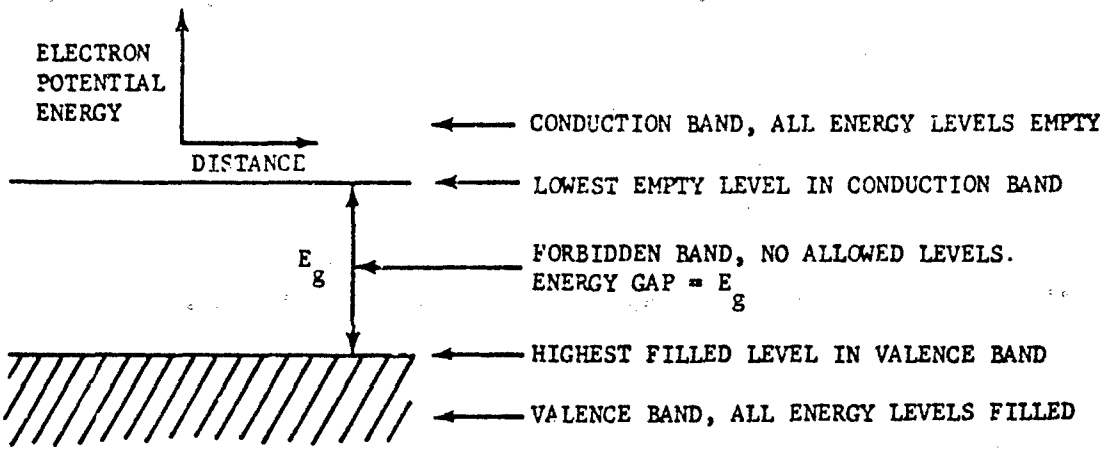


FIGURE 3-59. Semiconductor Energy Levels at Absolute Zero

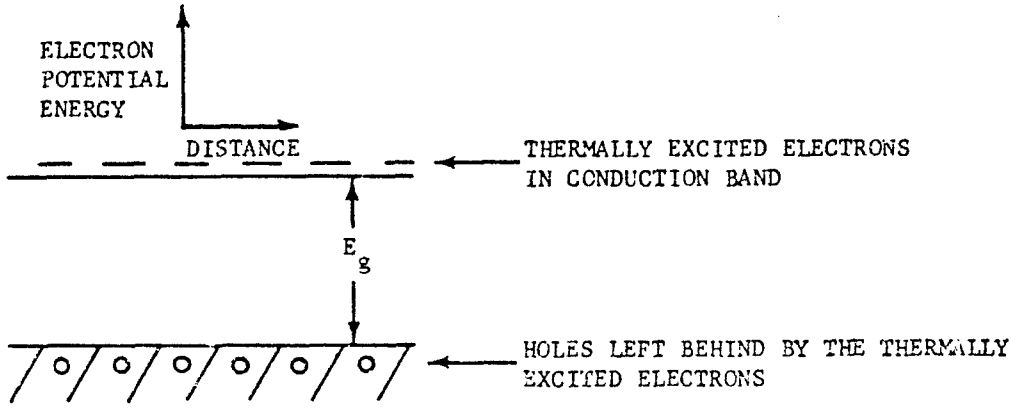


FIGURE 3-60. Semiconductor With Thermally Excited Electrons

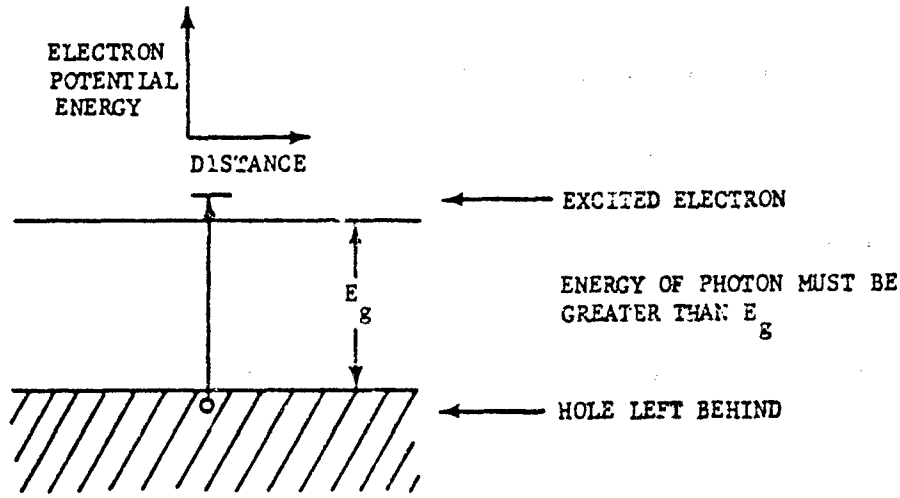


FIGURE 3-61. Electron Excited by a Photon

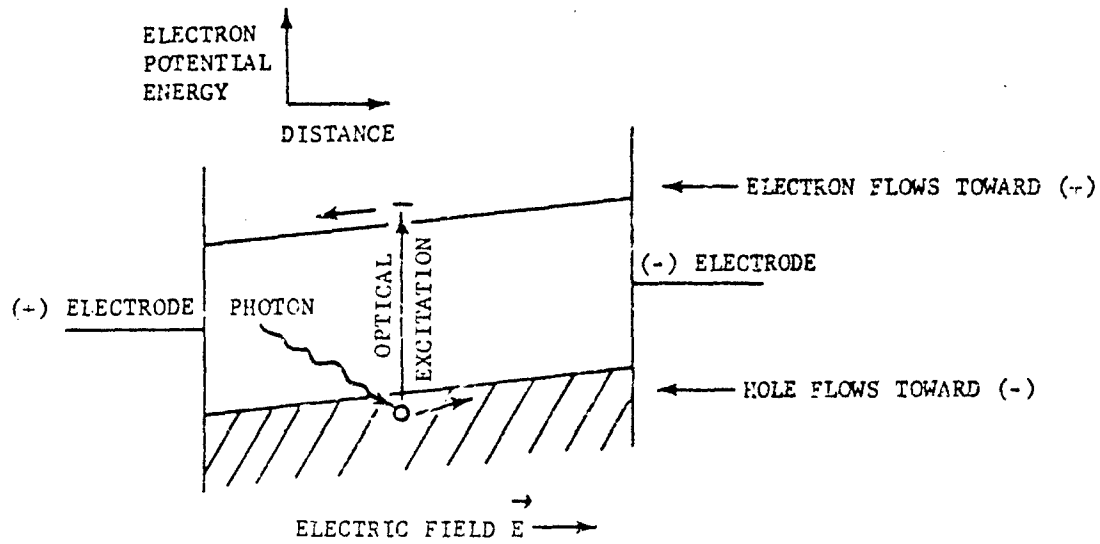


FIGURE 3-62. Semiconductor (Intrinsic Detector) With Electric Field $\vec{E}$

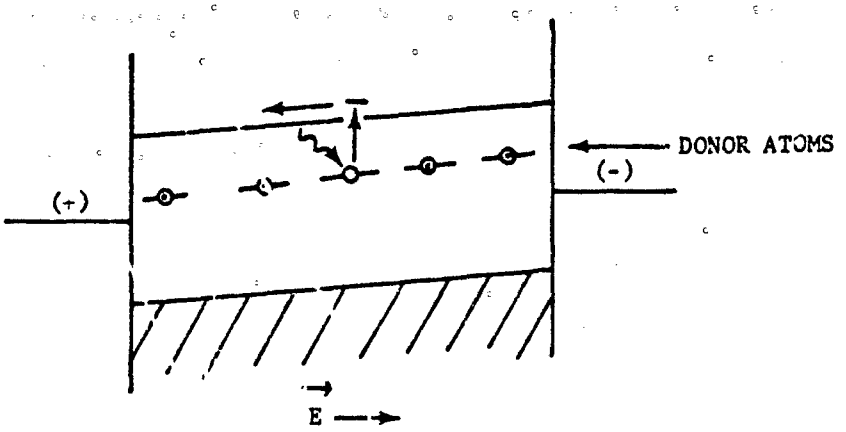


FIGURE 3-63. Semiconductor (N-type Extrinsic Detector) With Donors

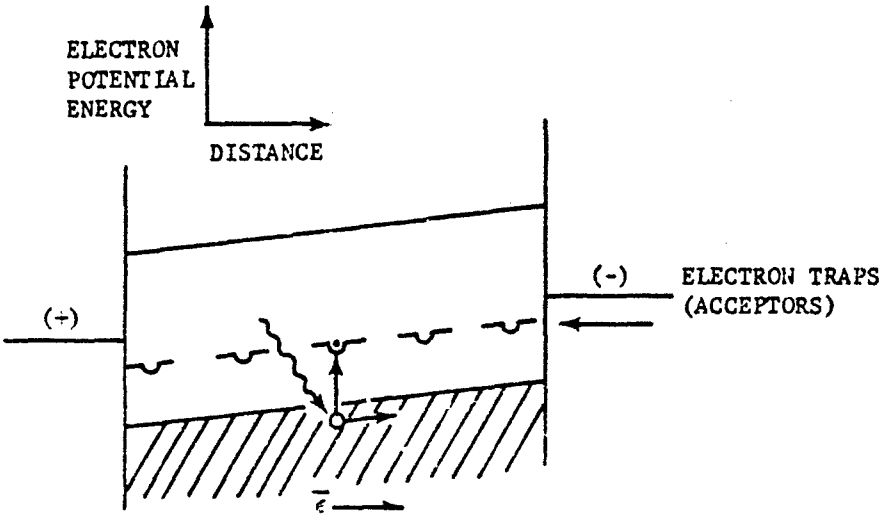


FIGURE 3-64. Semiconductor (P-type Extrinsic Detector) With Acceptors

Certain impurities (Cu, Hg, Zn, Au, etc.) can cause traps within the forbidden gap. An electron caught in one of these traps cannot conduct since the impurity is not free to move. Optical excitations of trapped electrons produce free conduction electrons (Fig. 3-63). Optical excitations of valence electrons into the traps produce free holes (Fig. 3-64).

The following is a description of various types of photoconductive detectors:

a. Intrinsic Elements. Absorption is caused by excitation of a valence band electron into the conduction band. Both the electron and the resultant hole are free to conduct. Examples of intrinsic elements are Si, Se, and Te.

b. Extrinsic Elements. Absorption is caused either by excitation of a valence band electron to an acceptor leaving a free hole or by excitation of a donor electron to the conduction band making a free electron. Examples of

extrinsic elements are Ge:Hg, Ge:Cu, and Ge:Zn.

c. III-V Compounds. A III-V compound is a binary compound having one element from column III and one element from column V of the periodic chart of the elements per molecule of compound. Semiconducting III-V compounds can be used as intrinsic photoconductive detectors. InSb is an example of a III-V photoconductive detector.

d. Lead Salts (PbS, PbSe, PbTe). Lead salts are intrinsic detectors but their exact operation is complicated. Either excess lead or excess chalcogen (S, Se, Te) can cause effective impurities. Generally, the material is made p-type and then electron-trapping impurities are introduced until stoichiometry[50] is achieved at the desired operating temperature.

e. Pseudo-binary Compounds. A pseudo-binary compound is an (A-B)C compound. For each atom of Type C there is one atom of A or B per molecule. Thus C is one half of the binary compound and the A-B system is the other half. Hence the name pseudo-binary compound. Mercury-cadmium-telluride is an example.

Mercury-cadmium-telluride can be used as an intrinsic photoconductive detector. The material is an alloy of mercury-telluride and cadmium-telluride. The band gap is a function of the mercury-to-cadmium ratio. Either excess tellurium or excess mercury-plus-cadmium causes effective impurities. Other typical (A-B)C compounds include (Pb-Sn)Te and (Pb-Sn)Se.

3-4.2.1.3 *Photovoltaic Detectors*

When optical excitation occurs in the region of a p-n junction (Fig. 3-57(C)), a hole and an electron are set free (Fig. 3-65). Due to the built-in potential gradient, they separate; the hole going to the p side, the electron to the n side. The external measuring circuit can measure this either as current or as voltage. A photovoltaic detector operates with little or no bias. Examples of photovoltaic detectors are Si photodiode, GaAs, InAs, and InSb.

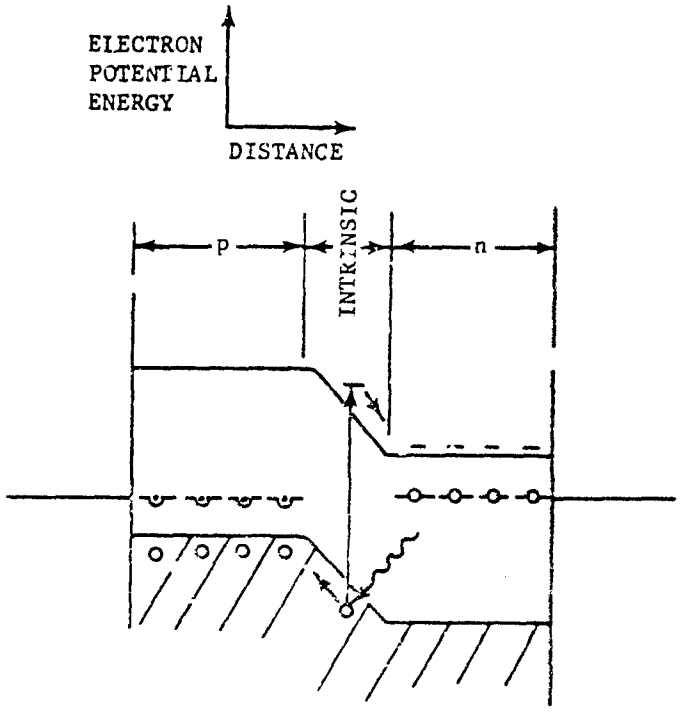


FIGURE 3-65. P-n Junction (Photovoltaic Detector)

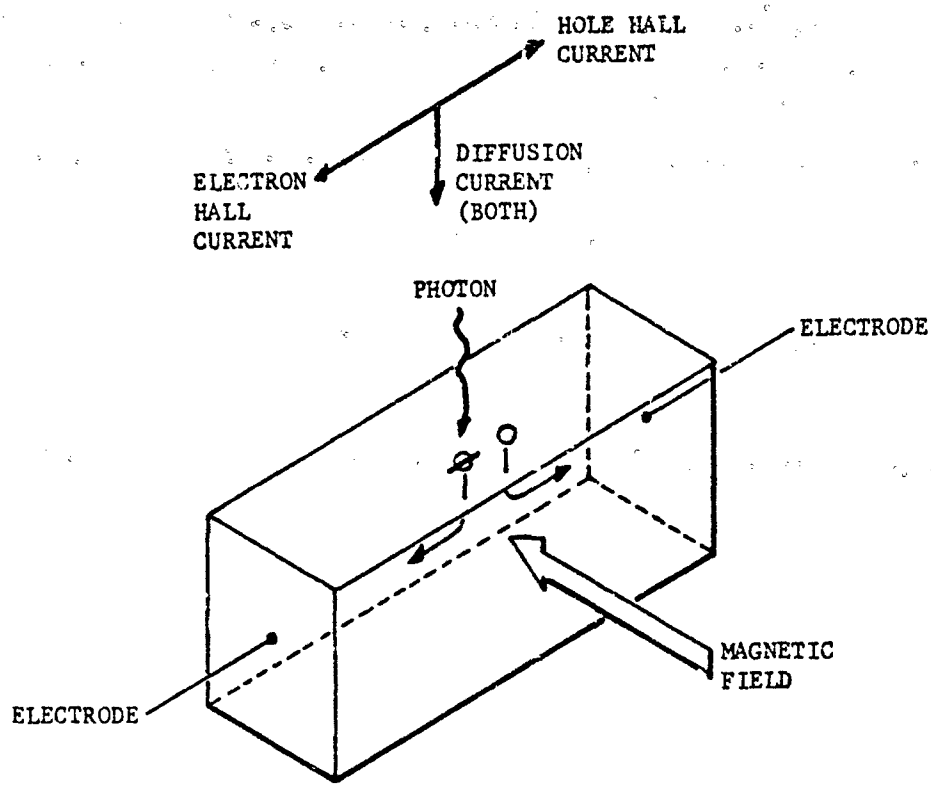


FIGURE 3-66. Photo-induced Hall Effect (Photoelectromagnetic Detector)

3-4.2.1.4 Photoelectromagnetic Detectors

A photoelectromagnetic detector operates in accordance with the Hall effect using photo-induced carriers (Ref. 51 and Fig. 3-66). Incident radiation generates electrons and/or holes mostly at the front surface of the detector due to strong absorption. The electrons and holes, therefore, diffuse away from the surface. A magnetic field is applied parallel to the absorbing surface, deflecting the carriers and causing a Hall-type current which is the signal current. InSb can be used as this kind of detector.

3-4.2.2 Thermal Detectors

3-4.2.2.1 Liquid Thermometer

Incident radiation is absorbed by a liquid reservoir causing an increase in temperature which causes an increase in volume. The volumetric change of the liquid is converted into a change in length by a capillary tube. Liquid thermometers are relatively insensitive due to the high heat capacity of the reservoir. Examples are mercury- and alcohol-in-glass thermometers. In 1800, when Herschel discovered infrared radiation in the solar spectrum, his detector was a mercury-in-glass thermometer.

3-4.2.2.2 Golay Cell[52]

A Golay cell is a gas thermometer. Incident radiation is absorbed by a wall of a gas-chamber which heats the gas causing increases in both pressure and volume of the chamber, thus deforming a membrane which is another wall of the chamber. This change is measured by a sensitive optical system using a reticle that is either in-focus or out-of-focus depending on the membrane's position. The Golay cell is sensitive to mechanical shock and sudden changes in

ambient pressure. Hence, its use in limited to spectrographs in general.

3-4.2.2.3 Calorimeter

A calorimeter measures total heat introduced into the calorimeter (in calories, for instance). Calorimeters, although relatively insensitive detectors, are suitable for use in calibrating lasers (such as CO_2 at 10.6 microns). The power input is proportional to the time rate of change of temperature. A beaker of water, a mercury-in-glass thermometer, and a stopwatch are sometimes all that are needed for performing an approximate power output calibration.

3-4.2.2.4 Thermocouple

A thermocouple is a bi-metallic junction. The voltage generated across the junction of the two dissimilar metals is a function of the junction temperature. For small temperature changes (the usual case in infrared radiation measurements) the voltage is proportional to the temperature change. Normally the voltage measured is the difference between two such junctions, one held at a fixed temperature. The response time is on the order of milliseconds. An example is the Sb-Bi junction.

3-4.2.2.5 Thermopile

A thermopile consists of a series of similar thermocouples with alternate ones illuminated; the others are held at a fixed temperature. The numerical value of the measured voltage is thus increased. However, the responsivity and detectivity are the same as for a single thermocouple. The response time is usually slower than the single thermocouple.

3-4.2.2.6 Bolometer

A bolometer is a device with a temperature-sensitive passive electrical bulk property (Fig. 3-57(D)). Bolometer responsivity $\mathcal{R}$ is defined as

$$\mathcal{R} = \frac{1}{R}\left(\frac{dR}{dT}\right) \qquad (3\text{-}84)$$

where the bulk property R is usually resistance.

a. Metal Bolometer. A thin metal wire ribbon, or film is the sensing material in a metal bolometer. Typical metals are Ag, Al, Ni, and Pt. Since metals reflect infrared radiation, the metal is coated with an absorbent film. The resistance of metals is nearly linear with temperature. At room temperature, $\mathcal{R}$ is approximately equal to T^{-1}. For room temperature metal bolometers, $\mathcal{R} = 3 \times 10^{-3}\,^{\circ}K^{-1}$. Bolometers with response times as short as 1 microsec have been made but are very fragile[53].

b. Semiconductor Bolometer. A semiconductor is the sensing material of a semiconductor bolometer. A semiconductor's resistance decreases exponentially with temperature as

$$R = R_o e^{A'/T}. \qquad (3\text{-}85)$$

Thus

$$\mathcal{R} = -A'T^{-2} \qquad (3\text{-}86)$$

where A' defines the change of resistivity as a function of temperature change. For germanium, $A' = 3000^{\circ}K$. Therefore, $\mathcal{R} = -3 \times 10^{2}\,^{\circ}K^{-1}$ (Ref. 49).

Cooling increases the magnitude of $\mathcal{R}$, but the resistance quickly becomes too high to measure. One solution is to use impurity-doped semiconductors, lowering A' but allowing much lower T. Thus $\mathcal{R} = 6$ in Ge:Ga at $4.2^{\circ}K$.

c. Carbon Resistor Bolometer. At low temperatures, the carbon rod in an ordinary carbon resistor shows an exponential dependence of resistance on temperature[54]. This is because graphite is a semiconductor with a very small band gap. The carbon resistor bolometer, as a result, is a special case of the semiconductor bolometer.

d. Thermistor. A semiconductor bolometer made of sintered semiconducting oxides is called a thermistor.

e. Phase Transition Bolometer. A phase transition bolometer uses a material with a phase transition which occurs at the bolometer's operating temperature T_o. The phase change is not instantaneously abrupt but occurs over a range $\pm\Delta T$ about T_o. If one phase is high resistivity and the other low, then the change in resistance with temperature about T can be very high. A typical semiconductor example is vanadium dioxide with $T_o = 68^{\circ}C$ and $\mathcal{R} = 1$ (Ref.55).

f. Superconducting Bolometer[56]. A phase transition bolometer using the superconducting phase transition is called a superconducting

bolometer. For superconducting tin, the transition temperature T_o is 3.7°K and $R = 3.5$. Special equipment is required to maintain the transition temperature, hence superconducting bolometers are used only as a laboratory instrument.

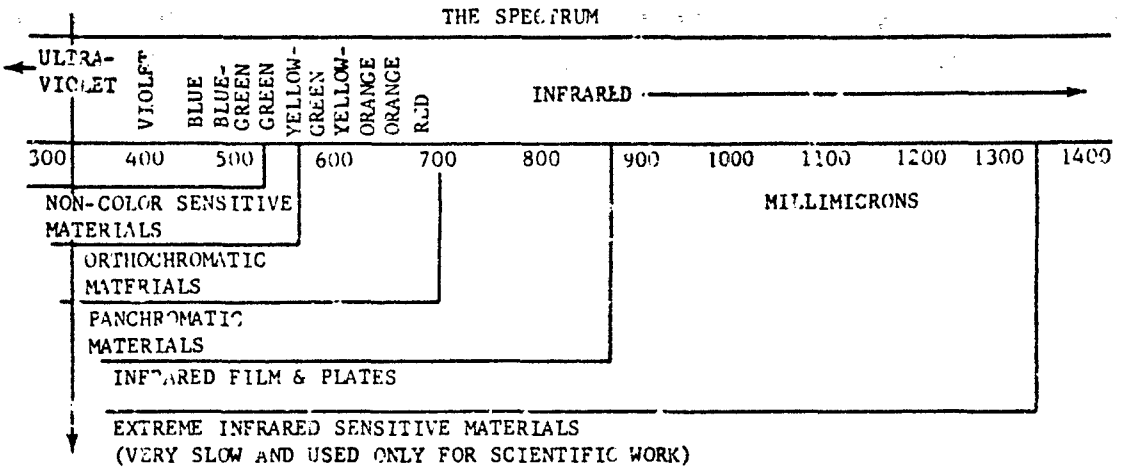
3-4.2.3 Imaging Devices

Imaging devices may be either thermal or quantum in nature, or both.

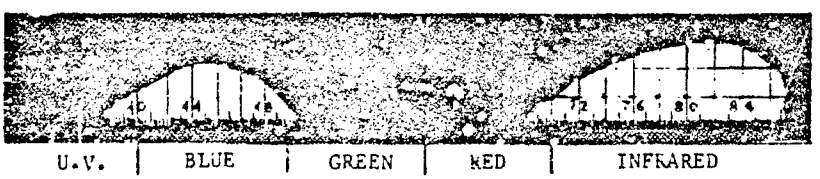
3-4.2.3.1 Photographic Film

Through treatment with special dyes, photographic emulsions may be rendered infrared sensitive. These emulsions can serve as both infrared detectors and display devices. By proper optical design, a camera can be constructed for use in the infrared and emulsions, sensitive as far out as 1.350 microns are available. Most infrared photography is performed in the 0.70- to 0.86-micron region (Fig. 3-67(A)). Since longer wavelength radiation is scattered to a lesser degree than visible light by atmospheric particles, infrared photography produces much clearer pictures than ordinary photography when the scene is obscured by haze.

Currently available infrared films and plates



(A) THE SPECTRAL RANGE OF PHOTOGRAPHIC SENSITIVITY



(B) SPECTRAL SENSITIVITY OF KODAK INFRARED MATERIALS, NAMELY, KODAK INFRARED FILM (IN SHEETS) AND KODAK INFRARED FILM (IN MAGAZINES)

FIGURE 3-67. Spectral Range and Sensitivity of IR Photographic Material

may be used in a manner not unlike that of panchromatic materials. Exposures, however, must be made with a higher degree of precision to obtain satisfactory results. The materials should be handled in a dark room with no white light present. Where a safelight is required, the Kodak Safelight Filter, Wratten Series 7, is recommended for films and plates. With high-speed infrared film, however, no safelight can be used. Some plastics, black paper, and hard rubber are not opaque to infrared radiation, therefore, care must be taken to use a suitable holder for the film if fogging is to be avoided. Metal foil is adequate for this purpose.

TABLE 3-9. FILTERS FOR INFRARED PHOTOGRAPHY

MANUFACTURER	NAME & NUMBER OF FILTER		WAVELENGTH BEYOND WHICH FILTER TRANSMITS, micron
Eastman Kodak Co.	Wratten	23A	0.560
	"	25	0.600
	"	29(F)	0.620
	"	70	0.670
	"	89	0.680
	"	89A	0.700
	"	88	0.720
	"	88A	0.740
	"	87	0.770
		87(C)	
Agfa	Agfa	42	0.600
	"	83	0.725
	"	84	0.750
	"	85	0.830
	"	87	0.840
	"	89	0.930
Ilford		201	0.560
		202	0.580
		204	0.600
		205	0.630
		206	0.660
		207	0.760
		207 & 813	0.800
Corning Glass Works		246	0.580
		245	0.600
		244	0.610
		243	0.620
		242	0.630
		241	0.640
		244 & 555	0.690
		244 & 585	0.710
		254	0.780
		255	0.760

Infrared film is available in sheets of all common sizes. Either normal or high contrast is available by proper adjustment of the development process. The film is also available in 35-mm rolls. High-speed infrared film is available in both 16-mm and 35-mm rolls for use in motion picture applications. As shown in Fig. 3-67 (B), infrared materials are sensitive also to the red and blue regions of the spectrum. It is necessary, therefore, to use a red filter to eliminate the blue portion if only infrared sensitivity is desired. The best filter is the Kodak Wratten Filter No. 25(A). Other filters that may be employed with satisfactory results are listed in Table 3-9.

Because of their long wavelength, infrared rays do not come to focus in the same plane as visible rays, therefore the lens-to-film distance must be increased for infrared photography. The correction should neither exceed 0.5 percent for most lenses nor 0.35 percent for good lenses. For best definition, the smallest lens opening allowable by the conditions should be used. Special infrared optics should be used since ordinary optics are not corrected for aberrations in the infrared region and moving the lens corrects only the longitudinal color aberrations. Reducing the aperture helps to limit the magnitude of the off-axis aberrations. Recommended exposure settings are listed in Table 3-10.

The smaller the specimen to be photographed, the further the lens must be extended from the film plane. The effective aperture may be much less than the nominal aperture setting. The effective f/no. for close-up work is computed by using the following relation:

$$\text{Effective f/no.} = \frac{(\text{nominal f/no.}) \times (\text{lens-to-film distance})}{\text{focal length}} \qquad (3\text{-}87)$$

The lens-to-film distance is the focal length plus the lens extension from its position for infinity focus.

Infrared-sensitive film may be kept at room temperature, but should be refrigerated if it is to be stored for several months. High-speed infrared film should not be stored at temperatures above 50° F.

The speed of infrared film and plates can be almost doubled by hypersensitizing immediately prior to use. The hypersensitizing solution is made by diluting 4 parts of 28 percent ammonia

TABLE 3-10. EXPOSURES OF INFRARED PHOTOGRAPHY ASSUMING OPEN LANDSCAPE, SUMMER SUN, CLOSEUPS, AND BRIGHT SUN

MATERIAL	FILTER	EXPOSURE
Kodak Infrared Film	Wratten No. 25, 29, or 70	1/25 sec, f/8
	Wratten No. 88, 89, or 89A	1/25 sec, f/6.3
	Wratten No. 87 or 88A	1/25 sec, f/5.6
	none	1/100 sec, f/11
Kodak Infrared Sheet	Wratten No. 25, 29, or 70	1/25 sec, f/8
	Wratten No. 87 or 88A	1/25 sec, f/5.6
	none	1/50 sec, f/16
Kodak Infrared Sensitive Plates	Wratten No. 25, 29, or 70	1/25 sec, f/6.3
	Wratten No. 88, 89, or 89A	1/25 sec, f/5.6
	Wratten No. 87 or 88A	1/25 sec, f/4.5
	none	1/100 sec, f/11

with 100 parts of water. The plates should be bathed for about 3 minutes in such a solution at a temperature below 50°F; then rapidly dried, using a blower if possible, to prevent formation of streaks. This must be done in total darkness. Development and fixing are the same as for ordinary films.

When an emulsion is exposed and developed, the silver halide crystals affected by the exposure are converted to particles of metallic silver, while those not exposed are removed in the fixing bath. The photographic image consists of large numbers of minute silver grains. The density D of grains determines the opacity O of the negative and is defined as the logarithm of the opacity.

$$D = \log O \tag{3-88}$$

where

$O = I/I'$
I = light incident on film
I' = light transmitted by film

The density E of the image increases with exposure. It is defined by the relationship

$$E = It \tag{3-89}$$

where

I = light incident on film
t = time during which light is incident, i.e., exposure time

A curve of density vs logarithm of exposure, plotted for an emulsion, is known as its characteristic curve. This curve usually contains a straight-line portion, the slope of which, called the gamma (γ) of the film, is a measure of its contrast properties. The higher the gamma, the greater the contrast ratio for a given difference in exposure. Some typical characteristic curves are presented in Figs. 3-68, 3-69, and 3-70 for infrared emulsions. Time-gamma curves, giving the gamma value to be obtained by a given process time in the particular developing agents are also shown. Characteristic curves give the density of the negative as a function of exposure for the particular gamma obtained in the development process.

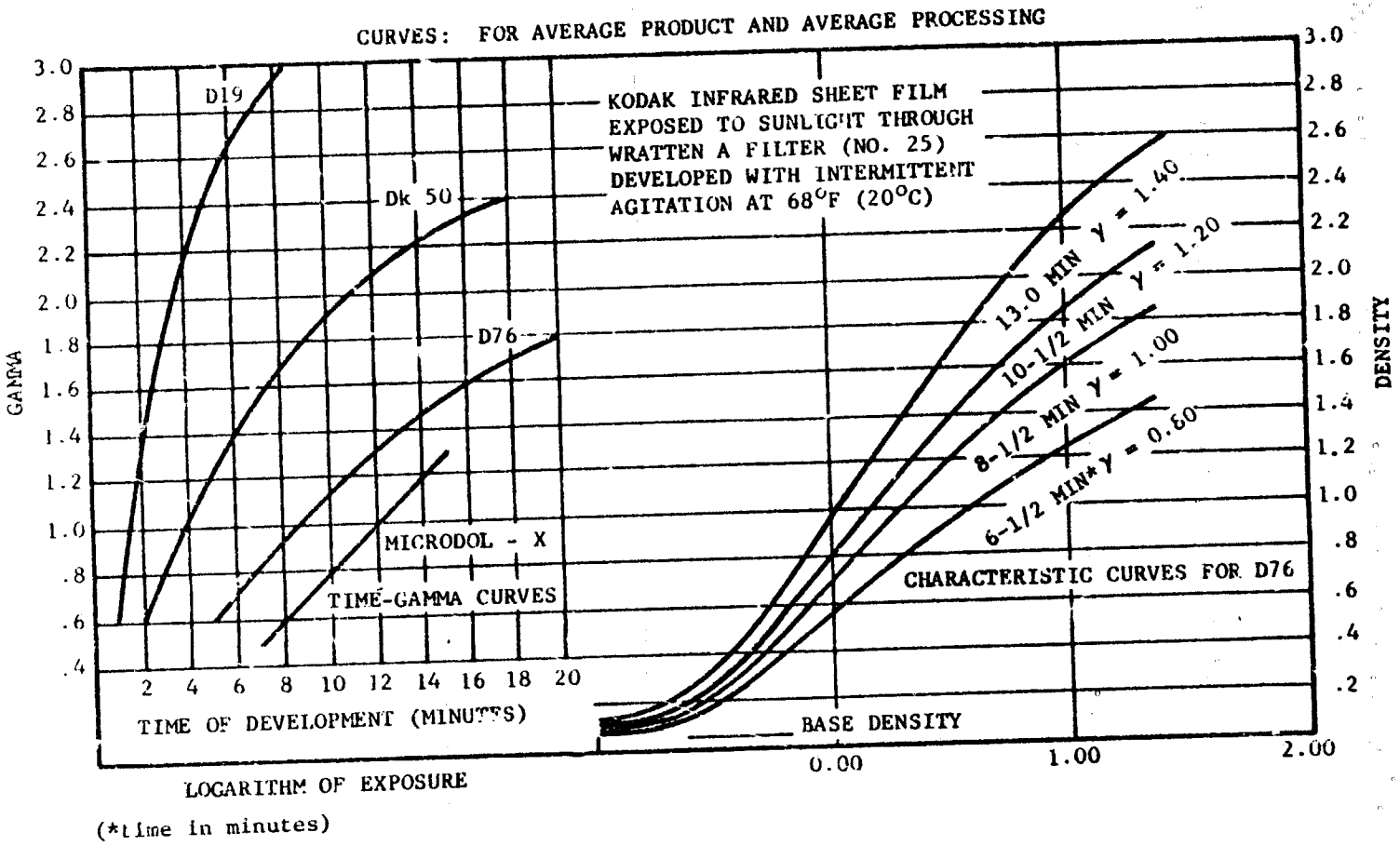
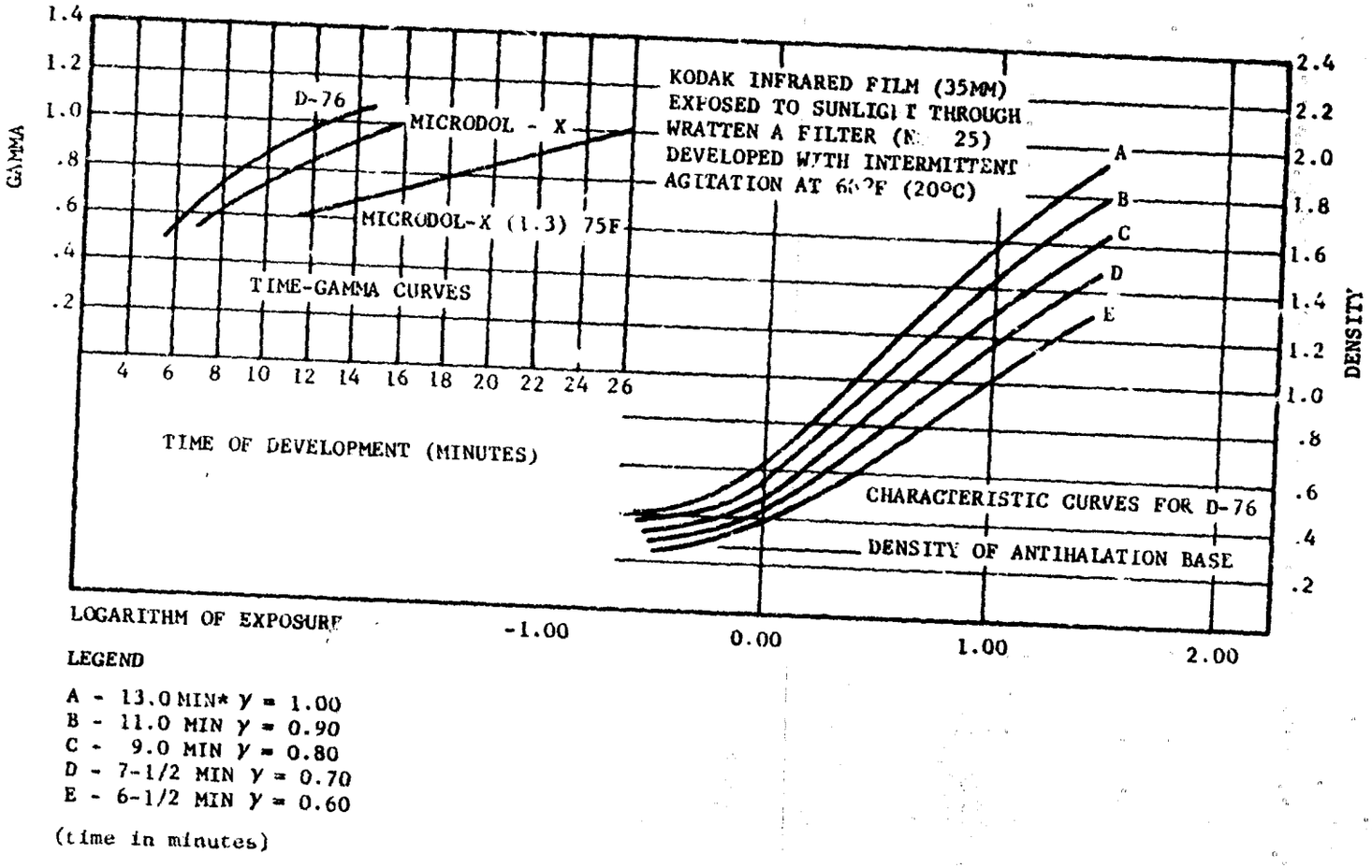


FIGURE 3-68. Density vs Exposure for Various IR Film Sheets



LEGEND

A - 13.0 MIN* γ = 1.00
B - 11.0 MIN γ = 0.90
C - 9.0 MIN γ = 0.80
D - 7-1/2 MIN γ = 0.70
E - 6-1/2 MIN γ = 0.60

(time in minutes)

FIGURE 3-69. Density vs Exposure for Various Types of 35-mm IR Film

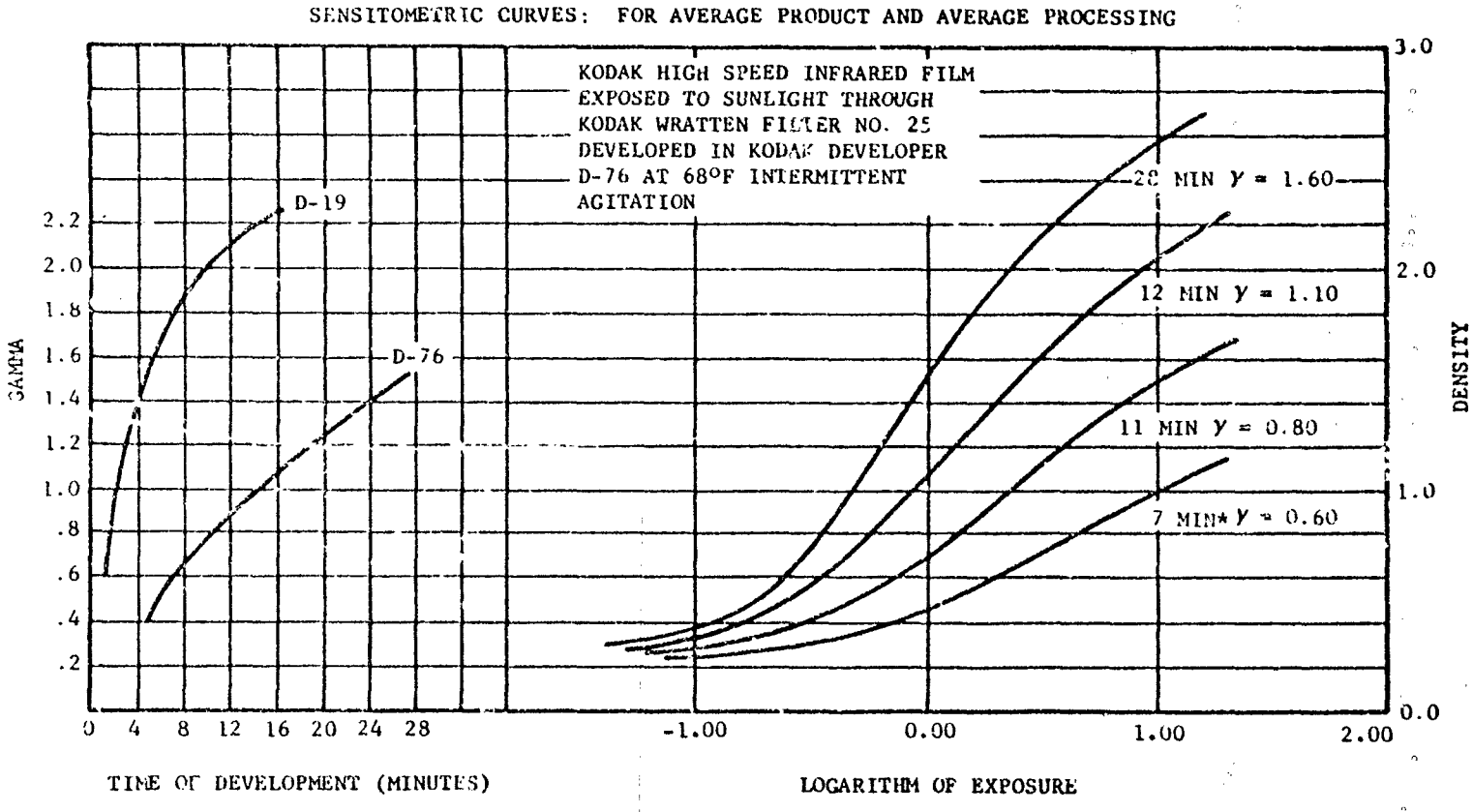


FIGURE 3-70. Density vs Exposure for Various Types of High-speed IR Film

3-4.2.3.2 IR Vidicon

A semiconductor film when cooled until there are almost no thermally generated carriers becomes a high-resistivity dielectric. Incident radiation absorbed by the film causes regions of lowered resistivity. If the film were the dielectric of a capacitor, then the lowered resistivity would allow leakage between the plates, thus lowering the potential difference across the capacitor. An infrared vidicon operates on this principle.

The generally high resistance of a cold semiconductor film effectively isolates each discrete point on the film as a separate capacitor. When an infrared image is focused on the film, the voltage across the capacitor is a function of the infrared intensity of the image at each discrete point. The voltage is sampled by scanning the film with a low-voltage electron beam.

Either the beam current is measured directly or the beam current is kept constant and the reflected beam is measured by a series of dynodes. The latter system is more sensitive. For infrared work, the semiconductors used include doped germanium, doped silicon, and lead sulfide.

3-4.2.3.3 Image Converter

An infrared image is focused on a photoemissive cathode; emitted electrons are then accelerated and focused to a fluorescent screen. Focusing can be accomplished by close proximity, magnetic coils, or electrostatic plates. The latter is preferred for best imaging. The 1P28 image tube is an example of an image converter.

3-4.3 DETECTOR PERFORMANCE AND TEST PROCEDURES

3-4.3.1 Detector Figures of Merit

3-4.3.1.1 Noise Equivalent Power (NEP)

The *NEP* of a detector is defined as the amount of incident rms infrared signal power which causes an output signal-to-noise ratio of unity. Thus, the smaller the numerical value of *NEP*, the better the detector.

3-4.3.1.2 Detectivity (D)

Detectivity D is defined as the reciprocal of *NEP*[57].

$$D = \frac{1}{NEP} \tag{3-90}$$

3-4.3.1.3 Specific Detectivity (D^*)

NEP and D are two detector parameters which vary as a function of detector area A_D and noise bandwidth Δf, and as such, A_D and Δf must be specified to make *NEP* or D meaningful.

A figure of merit D^* has been normalized for detector area and noise bandwidth, based upon the fact that *NEP* of detectors is proportional to the square root of the detector area A_D and the square root of the noise equivalent bandwidth Δf (Ref. 57). It is related to *NEP* and D as indicated by the following equations:

$$D^* = \frac{\sqrt{A_D \Delta f}}{NEP} \tag{3-91}$$

$$D^* = \sqrt{A_D \Delta f}\ D$$

For detectors which behave according to these normalizing conditions, the concept of D^* is valid.

The formula for calculating D^* from experimental data is

$$D^* = \frac{\left(\frac{V_{s,rms}}{V_{n,rms}}\right)\left(\sqrt{\Delta f}\right)}{(H_{rms})\left(\sqrt{A_D}\right)}, \text{ cm Hz}^{1/2}\ \text{w}^{-1} \tag{3-92}$$

The numerical value of D^* is the normalized rms signal-to-noise ratio which would be obtained from a 1 cm^2 detector with 1 w cm^{-2} incident rms irradiance, measured in a 1 Hz bandwidth.

Two basic types of D^* may be specified for a detector; blackbody D^* (D^*_{BB}) or spectral D^* (D^*_λ). The blackbody D^* is obtained by using the total irradiance from a certain temperature blackbody in the calculation. The spectral D^* is obtained by using a monochromatic irradiance at a certain wavelength in the calculation. The spectral peak D^* i.e.,(D^*_{peak}), is a special case of D^*_λ, where λ is the wavelength of maximum response for the detector.

The conditions under which the D^* value is specified follow the D^* notation in parentheses as follows

$$D^*(T_{BB} \text{ or } \lambda, f_o) \qquad (3\text{-}33)$$

Therefore, D^* (500°K, 1000) means that the D^* is measured using a 500°K blackbody source and a modulation frequency of 1000 Hz. The D^* (10μ, 500) value specifies a spectral D^* at 10 microns and at a 500-Hz modulation frequency. The noise is either measured in a 1-Hz bandwidth or normalized to a 1-Hz bandwidth.

3-4.3.2 Theoretical Background Limited Detectors

When the inherent, internally generated noise of a detector is negligible compared to the noise caused by random arrival of background photons, the detector is said to be background limited. This is optimum sensitivity for a detector because both the signal and the noise of the detector are established by the photon environment, and the detector itself adds no appreciable noise to degrade the signal-to-noise ratio. A detector with such properties is said to be a BLIP (Background-limited Infrared Photodetector)[58] detector. This background noise, in semiconductor detectors, exhibits itself in the form of generation-recombination noise due to photon generated carriers.

This BLIP condition is an ultimate limit to D^*. The theoretical maximum D^*_λ for a photoconductive detector[59] is

$$D^*(\lambda_p, f) = \frac{\lambda_c}{hc}\sqrt{\frac{\eta}{4Q_{(0-\lambda_c)}}} \qquad (3\text{-}94)$$

where

$D^*(\lambda_p, f)$ = D^* at peak wavelength, cm $Hz^{1/2}$ w^{-1}
λ_c = long wavelength cutoff of detector, cm
h = Planck's constant, J-sec
c = speed of light, cm sec^{-1}
η = quantum efficiency of detector
$Q_{(0-\lambda_c)}$ = background photon flux density from 0 to λ_c wavelength, photons cm^{-2} sec^{-1}

The BLIP D^* for a photovoltaic detector is higher by a factor of $\sqrt{2}$ due to the absence of recombination noise[60]. In a photovoltaic detector the carriers are collected at the electrodes before they recombine. Thus, the BLIP D^* limit for a photovoltaic detector is expressed as

$$D^*(\lambda_p, f) = \frac{\lambda}{hc}\sqrt{\frac{\eta}{2Q_{(0-\lambda_c)}}} \qquad (3\text{-}95)$$

As can be seen from this equation, the BLIP D^* is inversely proportional to the square root of the background photon flux.

One of the most common means of reducing the background level of a detector to increase its D^* is by "cold-stopping" the angular field-of-view of the detector to the extent that the detector views only the solid angle of interest. When used in a telescope, the detector should ideally be "cold-stopped" so that it views only the active optical elements, and not the telescope walls. "Cold-stopping" is generally accomplished with an aperture plate which is cooled to the extent that its photon emission is negligible compared to the photon flux through the aperture. When a BLIP detector is "cold-stopped" to a conical solid angle, centered on the normal to the detector face, D^* is related to the cone apex angle θ as follows:

$$\frac{D^*(\theta_1)}{D^*(\theta_2)} = \frac{\sin\left(\frac{\theta_2}{2}\right)}{\sin\left(\frac{\theta_1}{2}\right)} \qquad (3\text{-}96)$$

where θ_1 and θ_2 are full-cone angles. This relationship is true only for uniformly radiant backgrounds.

Based upon the above angular dependence for BLIP D^*, the figure of merit D^{**} was introduced by R. C. Jones[61] as a means of normalizing D^* to a 180-degree (hemispherical) field-of-view as follows:

$$D^{**} = D^*_{(\theta = 180^\circ)}$$

Therefore,

$$D^{**} = D^*_\theta \sin\left(\frac{\theta}{2}\right) \qquad (3\text{-}97)$$

Other methods for reducing the background photon flux include:

1. Cold spectral filters which pass only the wavelengths of interest

2. Cold optical elements and telescopes (to reduce telescope emission to an amount which is small compared to external background)

Most commercially available photoconductive and photovoltaic infrared detectors exhibit background dependence over a limited range of backgrounds. The impurity-activated germanium detectors, such as Ge:Hg and Ge:Cu, are the most nearly ideal BLIP detectors presently available, being BLIP over a large range of background photon flux densities. The D^* of these detectors under low photon-flux densities is classified and is discussed in AMCP 706-128.

3-4.3.3 Measurement of D^*

To determine the D^* of a detector and specify it meaningfully, it is necessary to measure the ratio of the rms signal-voltage to rms noise-voltage under a known set of conditions which include:

1. H_{rms}: Infrared signal irradiance at detector, rms value of modulated component. This irradiance may either be blackbody or monochromatic for determining D^*_{bb} or D^*_λ respectively.

2. A_D: Detector area. This may be either the actual area or, in some cases, an effective area such as with detectors in integrating cavities where the entrance aperture acts as A_D.

3. Δf: Noise equivalent bandwidth used for noise voltage measurement

4. f_o: Signal chopping frequency, fundamental component

5. T_D: Detector operating temperature

6. Background conditions: Either the photon flux density incident upon detector at all wavelengths from 0 to long wavelength cutoff; or the detector field-of-view and effective background temperature

A typical D^* measurement set-up is shown schematically in Fig. 3-71.

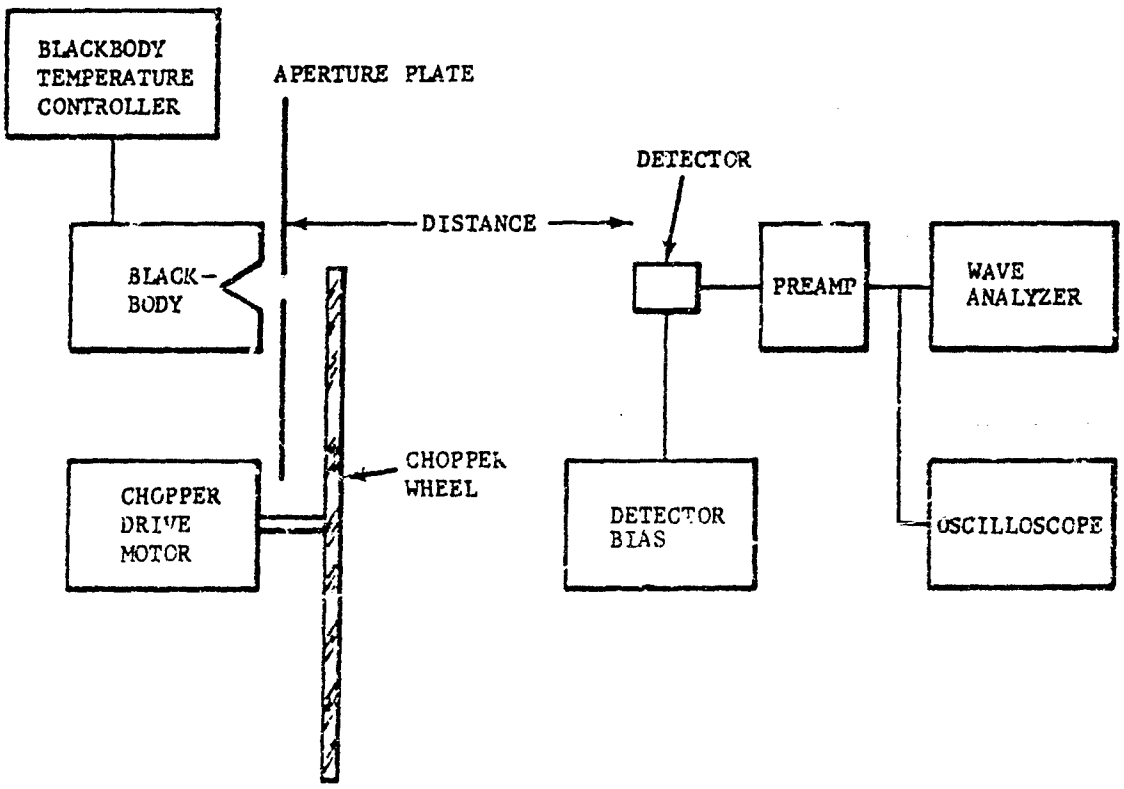


FIGURE 3-71. Block Diagram of D^* Measurement Apparatus

In this setup, a calibrated blackbody of known temperature serves as the infrared signal source. Its radiating area A_{BB} is limited by an aperture plate mounted in close proximity. To modulate the infrared signal, a chopper wheel with alternate teeth and spaces rotates in front of the blackbody aperture. The detector, placed at a known distance d from the blackbody stopping aperture, receives this chopped infrared radiation as a signal. The $H_{rms,BB}$ at the detector can be determined in the manner which follows.

For a Lambertian source, when d is much greater than the linear dimensions of A_{BB} and A_D and detector is centered on line normal to A_{BB},

$$H_{BB} = \frac{W_{BP} A_{BB}}{\pi d^2} \tag{3-98}$$

where

$$W_{BB} = \epsilon \sigma T^4 .$$

Therefore, under these conditions

$$H_{BB} = \frac{\epsilon \sigma T^4 A_{BB}}{\pi d^2} \tag{3-99}$$

In order to determine $H_{rms,BB}$, which is the rms value of the fundamental component of the chopped radiation, the modulation factor must be known and used as follows:

$$H_{rms,BB} = \frac{\epsilon \sigma T^4 A_{BB}}{\pi d^2} \times \begin{pmatrix}\text{modulation} \\ \text{factor}\end{pmatrix} \tag{3-100}$$

where the modulation factor for sine wave chopping is 0.35 and for square wave chopping is 0.45 (Ref. 62). (See Fig. 3-72.) With the rms irradiance thus established, the detector is coupled through appropriate biasing and preamplifying circuits to display instruments for reading signal and noise voltages. The most commonly used instrument for this purpose is a wave analyzer which is basically a variable frequency, narrow band voltmeter. A typical wave analyzer can be tuned to center frequencies f_o, from 20 Hz to 50 kHz, with a fixed Δf of about 7 Hz. The display must indicate rms voltages, preferably using a true rms meter. However, for the narrow Δf of 7 Hz, a peak reading meter which is calibrated for sine wave rms values is adequate with negligible error.

With the above apparatus operating and the wave analyzer tuned to the chopper frequency, the rms signal voltage ($V_{s,rms}$) can be measured. The blackbody aperture is then closed and the rms noise voltage ($V_{n,rms}$) is measured under the same conditions with no incident signal irradiance.

Since the signal-to-noise ratio sometimes varies with bias voltage, several bias points should be tried to maximize the signal to noise ratio. The bias for maximum signal-to-noise ratio is called "optimum bias".

In performing D^* measurements it is important that all noise sources other than the detector be negligibly small compared to the detector noise, otherwise the D^* would be limited by measurement system noise and would not be true measure of the detector.

With the measured information, the D^* is calculated and specified as

$$D^*(T_{BB}, f) = \frac{\left(\frac{V_{s,rms}}{V_{n,rms}}\right)\sqrt{\Delta f}}{(H_{rms,BB})\sqrt{A_D}} \tag{3-101}$$

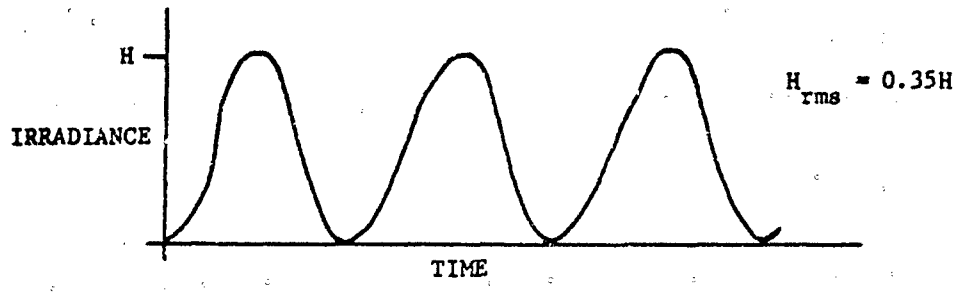
The proper units for D^* are cm $Hz^{1/2}$ w^{-1}.

This experimental setup can also be used to measure D^*_λ by inserting a calibrated spike-filter in front of the blackbody aperture. This filter transmits nearly monochomatic radiation at wavelength λ. This causes W_{BB} to be reduced to W_{filter}, which is the total power density (w cm^{-2}) transmitted by the filter.

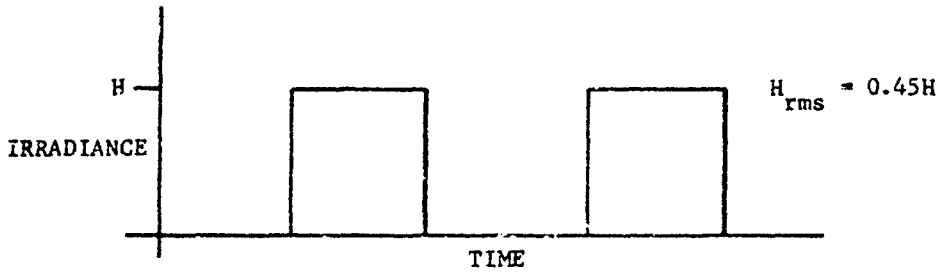
As a result, $H_{rms,BB}$ at the detector is reduced by the same amount to $H_{rms,filter}$. This filtered irradiance is then used in the D^* formula as

$$D^*(\lambda, f) = \frac{\left(\frac{V_{s,rms}}{V_{n,rms}}\right)\sqrt{\Delta f}}{(H_{rms,filter})\sqrt{A_D}} \tag{3-102}$$

In order to make D^* values meaningful, especially with background-limited detectors, the photon background conditions should be stated along with the D^* value. The preferred method is to state the total incident background photon flux density at all wavelengths from zero to the long wavelength cutoff of the detector. An alternate method is to state the field-of-view and effective background temperature.



(A) SINE WAVE CHOPPING



(B) SQUARE WAVE CHOPPING

FIGURE 3-72. Modulation Factor for Sine and Square Wave Chopping

3-4.3.4 Spectral Responsivity

The responsivity $\mathcal{R}$ of a detector is defined as the ratio of rms signal voltage, measured at the detector, to the incident rms infrared signal power. The common units are volts per watt.

For blackbody irradiance

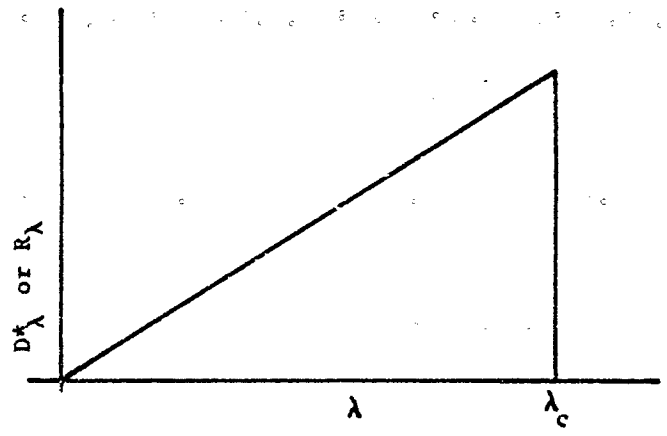
$$\mathcal{R}_{BB} = \frac{V_{s,rms}}{(H_{BH,rms})A_D} \qquad (3\text{-}103)$$

For spectral irradiance

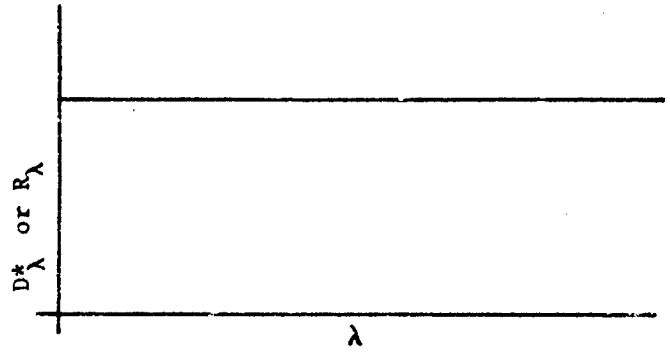
$$\mathcal{R}_{\lambda} = \frac{V_{s,rms}}{(H_{\lambda,rms})A_D} \qquad (3\text{-}104)$$

A plot of spectral responsivity as a function of wavelength is called a spectral response curve. This curve may represent either relative spectral response, in which case the curve is normalized to its own peak value as the 100 percent point, or the curve may carry an absolute calibration of responsivity vs wavelength.

Since the curve represents the variation of V_s with λ, and V_n is constant because the only variable parameter is signal irradiance, the curve also represents $\frac{V_s}{V_n}$ vs λ. Since $\frac{V_s}{V_n}$ is proportional to D^*, the spectral response curve also represents the variation D^*_λ vs λ.



(A) SPECTRAL RESPONSE OF IDEAL PHOTON DETECTOR



(B) SPECTRAL RESPONSE OF IDEAL THERMAL DETECTOR

FIGURE 3-72. Idealized Spectral Response of Photon and Thermal Detectors

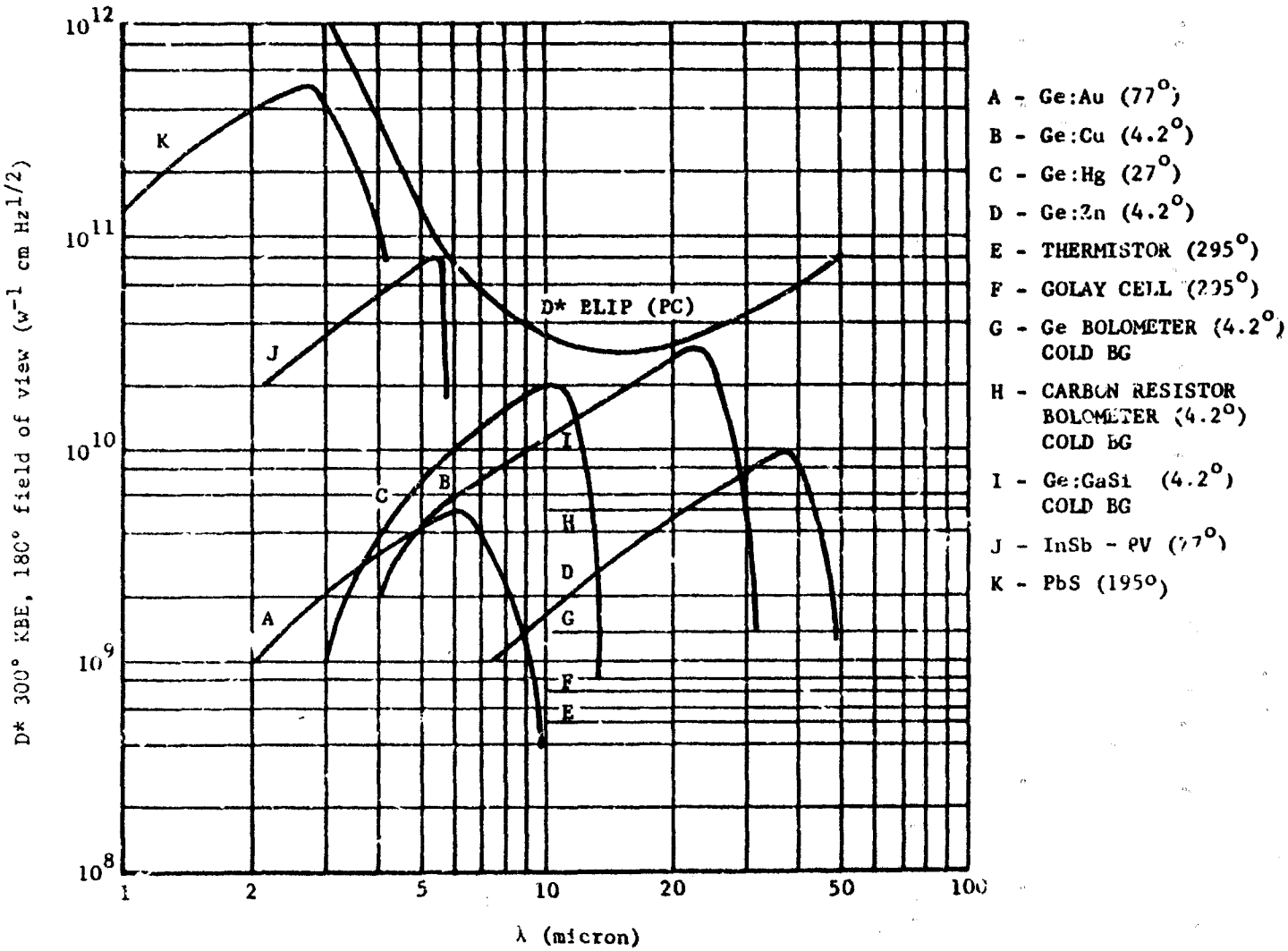


FIGURE 3-74. Detectivity of Long-wavelength Detectors

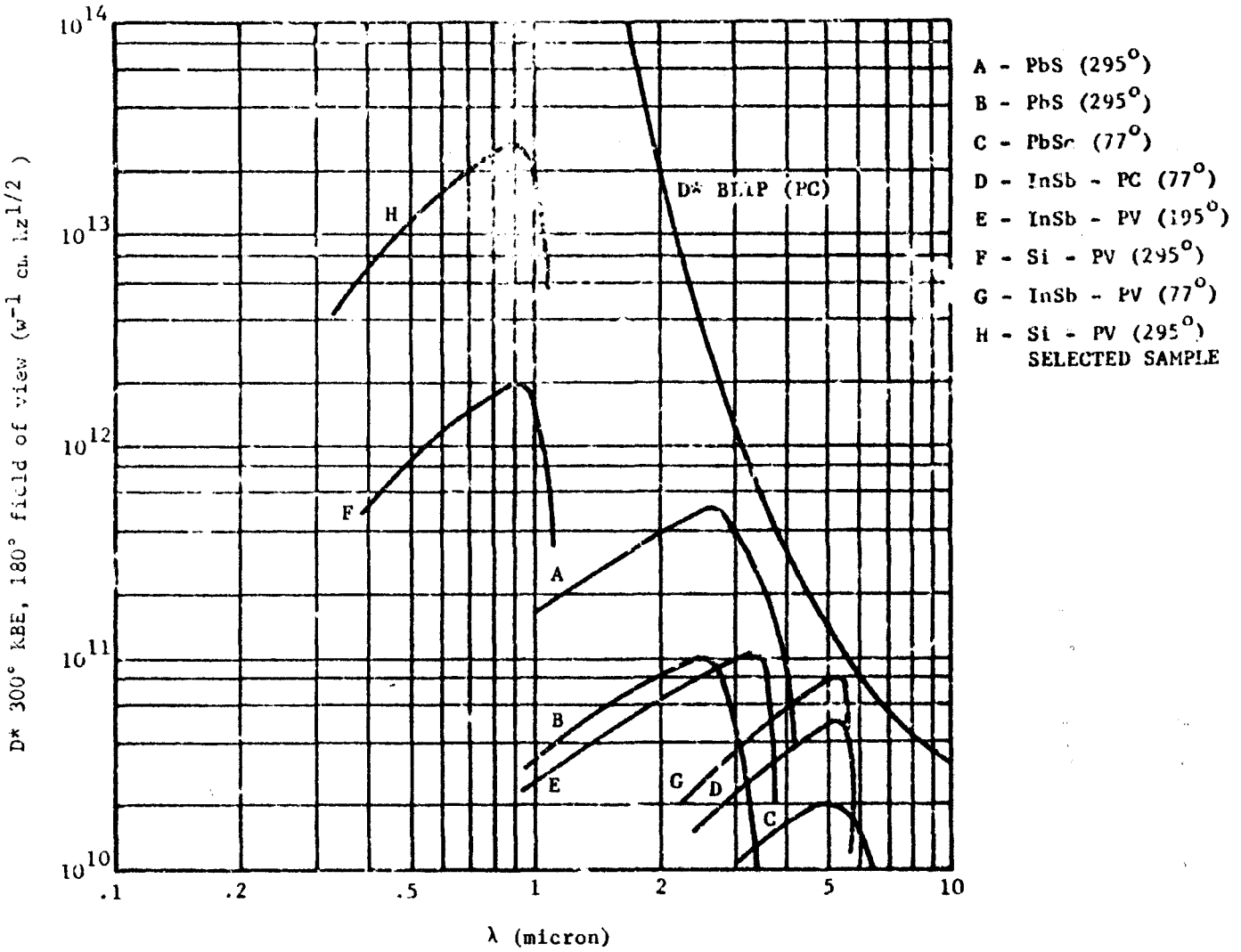


FIGURE 3-75. Detectivity of Short-wavelength Detectors

Spectral response curves are basically different for thermal and photon detectors. This is explained in the paragraphs which follow.

1. Photon Detector Spectral Response

As discussed in par. 3-4.2, a photon detector operates by direct absorption of photons. The energy of incident photons must be equal to or greater than a certain threshold amount in order to be absorbed. Photons possessing less energy (long wavelength) will not be absorbed. Thus a long wavelength cutoff for the detector's spectral response is exhibited. Photon energy is related to wavelength as follows[43]:

$$E = \frac{1.24}{\lambda} \qquad (3\text{-}105)$$

where

E = photon energy, eV (electron volt)
λ = wavelength, micron

Therefore, a photoconductor with a bandgap energy of 0.1 eV would respond only to wavelengths shorter than 12.4 microns.

Photon detectors have nearly constant quantum efficiency at wavelengths less than the long wavelength cutoff, so that the V_s for equal numbers of signal photons at various wavelengths would be nearly constant. However, the spectral responsivity (V/w) decreases as the wavelength shortens because the energy per photon is increasing without causing an increase in V_s.

Fig. 3-73(A) illustrates the idealized spectral response of a photon detector.

Spectral response curves for actual detectors are shown in Figs. 3-74 and 3-75.

2. Thermal Detector Spectral Response

The response of a thermal detector is proportional to the total photon energy absorbed. The response is limited only by the detector's ability to absorb photons of various wavelengths. The incident surface of thermal detectors is generally coated with materials which are highly absorptive over large wavelength regions. Therefore, the spectral responsivity of thermal detectors tends to be constant as shown in Fig. 3-73(B).

The spectral response curve of a detector may be determined experimentally by measuring its relative signal response in the output beam of a calibrated monochrometer as a function of λ.

An alternate approach to determine the approximate spectral response is to measure the $\mathcal{R}_\lambda$ of the detector at various discrete wavelengths, using a series of calibrated narrow-band spectral filters in front of a blackbody source.

3-4.3.5 Input-output Relationship

The relationship between detector output signal voltage and the input radiation can be described as a function of frequency (frequency response) or time (impulse response).

3-4.3.5.1 Frequency Response

In this method the detector is exposed to a constant amplitude, variable frequency, sinusoidally modulated infrared signal. As the signal chopping frequency is varied through the frequency range of interest, V_s of the detector is monitored and recorded. The results are then plotted as V_s vs f. This plot, when normalized to its peak value is called the frequency response or amplitude response.

3-4.3.5.2 Time Response

In this method the detector is exposed to pulses of infrared signal whose intensity vs time varies as a square wave function. The rise and decay times of the radiation pulses are made small compared to the expected time response of the detector. The output signal voltage of the detector is amplified and displayed on an oscilloscope. The rise and decay times of the output signal voltage pulses can be measured on this oscilloscope, and represent the rise and decay times of the detector.

Some of the most important processes governing detector time response include:

1. Free Carrier Lifetime. The carrier lifetime in semiconductor photon detectors is a measure of the average time that an excess electron or hole exists before recombination. Free carrier lifetime is the major factor governing decay time of most semiconductor photo detectors.

2. Thermal Time Constant. In thermal detectors the major process governing the time response is the rate at which its temperature can rise in response to signal power, and the rate at which its temperature can drop back to ambient

when the signal is removed. This time is a function of specific heats of the materials, thermal impedance to the heat sink, mass, and geometry. Thermal time constants are generally fairly slow, in the milliseconds to seconds range.

3. Resistance-capacitance (RC) Time Constant. A detector can often be limited in its response by the shunting effect of the capacitance of the detector and its surroundings (Fig. 3-76).

The breakpoint frequency f_b for a simple parallel RC time constant is expressed as

$$f_b = \frac{1}{2\pi \bar{R} C} \tag{3-106}$$

where

$\bar{R}$ = total parallel resistance, ohm
C = total parallel capacitance, farad

RC time constants can be minimized by the use of low capacitance cables and connectors.

The time constant of photon detectors is generally much faster than thermal detectors because thermal detectors work primarily on the basis of heat transfer which is a relatively slow process.

Many infrared detectors have a simple exponential time response which can be described with a single time constant. However, some detectors may have multiple processes simultaneously determining a complex time response which cannot be described in terms of a single time constant. In these cases, a frequency response curve or an impulse response plot is required to describe the detector behavior in a meaningful manner.

The frequency response of a detector can be measured in the D^* measurement setup illustrated in Fig. 3-71 simply by varying the rotational rate of the chopper and monitoring V_s on the wave analyzer. Due to the practical problems involved in rotating choppers at high speeds, this method is limited to about 100 kHz or less.

Higher frequencies can be measured using forward biased semiconductor diodes such as InAs and GaAs, which emit photons due to the direct recombination of minority carriers. The amount of radiant power emitted by these diodes is proportional to the bias current, which may be modulated at high frequencies, resulting in high-frequency modulated infrared signals. In this manner signals in the megahertz range can be generated to extend the frequency response measurement range. One drawback of the diode emitter is that it emits nearly monochromatic radiation, which may or may not be usable on the detector of interest.

Fast pulses of infrared power can be generated by:

1. Properly configured chopper wheels
2. Square wave biased emitter diodes
3. Rotating mirrors which scan a beam of infrared power across a detector rapidly
4. Pulsed lasers

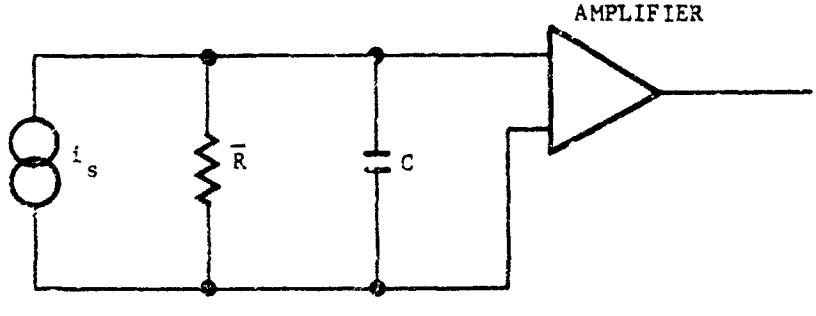


FIGURE 3-75. Equivalent Circuit Illustrating Detector-associated Capacitance

3-4.3.6 Detector Noise

There are a number of basic noise-generating mechanisms in detectors. The most important of these sources are explained in the paragraphs which follow.

3-4.3.6.1 Johnson Noise (also Nyquist or Thermal)

This noise, present in all resistors, is caused by the random motion of charge carriers in a resistor at thermal equilibrium. The rms value of this noise may be calculated as follows[63]:

$$Johnson\ V_{n,rms} = (4kT\bar{R}\Delta f)^{1/2} \qquad (3\text{-}107)$$

where

k = Boltzmann's constant

The spectrum of this noise is flat up to very high frequencies (microwave frequencies).

3-4.3.6.2 Temperature Noise

Temperature fluctuations in the detector, caused either by the radiative exchange with the environment or conductive exchange with the heat sink, can cause output voltage fluctuations. In thermal detectors, if the temperature noise is caused by radiative exchange with the environment, the detector is said to be at its theoretical limit.

The following relationship expresses the temperature noise variations for a thermal detector[49]:

$$\overline{\Delta T^2} = \frac{4kT^2K\Delta f}{K^2 + 4\pi^2 f^2 C^2} \qquad (3\text{-}108)$$

where

$\overline{\Delta T^2}$ = mean square temperature fluctuation
k = Boltzmann's constant, w sec $^\circ K^{-1}$
T = detector temperature, $^\circ K$
K = thermal conductance, w $^\circ K^{-1}$
C = heat capacity of the detector, w sec $^\circ K^{-1}$
Δf = noise equivalent bandwidth (must be small compared to frequency f), Hz

The relationship between $\overline{\Delta T^2}$ and $V_{n,rms}$ must be determined for each specific detector.

3-4.3.6.3 Generation-recombination Noise

The statistical fluctuation in the carrier generation and recombination rates in semiconductor detectors is called generation-recombination (G-R) noise. If the generation of carriers is due primarily to the temperature of the detector, the noise is termed thermal G-R noise. If the generation of carriers is due primarily to direct photon absorption, the noise is termed photon G-R noise. If the photon induced G-R noise is the predominant noise from the detector, it is said to be photon or background-noise limited. G-R noise has the same frequency dependence as signal.

The G-R noise for a photoconductor is expressed as follows[64]:

$$V_{n,rms} = \bar{R}I_{DC}\left[\frac{2\tau(\Delta f)}{n(1 + 4\pi^2 f^2 \tau^2)}\right]^{1/2} V \qquad (3\text{-}109)$$

where

$\bar{R}$ = detector static resistance, ohm
τ = carrier lifetime, sec
n = charge carrier density, dimensionless
Δf = noise equivalent bandwidth (must be small compared to frequency f) mHz
I_{DC} = dc current, A

For photovoltaic detectors, $V_{n,rms}$ is smaller by a factor of $\sqrt{2}$ due to the absence of recombination noise.

3-4.3.6.4 Shot Noise

Although shot noise—caused by the random fluctuations in the number of electrons emitted by a cathode—occurs in most detectors, it is sometimes the dominant noise in photovoltaic and photoemissive detectors. The rms value of shot noise in Δf Hz bandwidth may be calculated as follows[65]:

$$V_{n,rms} = \bar{R}(2qI_{DC}\Delta f)^{1/2} \qquad (3\text{-}110)$$

where

q = charge on an electron, 1.59×10^{-19} C
Δf = filter bandwidth, Hz

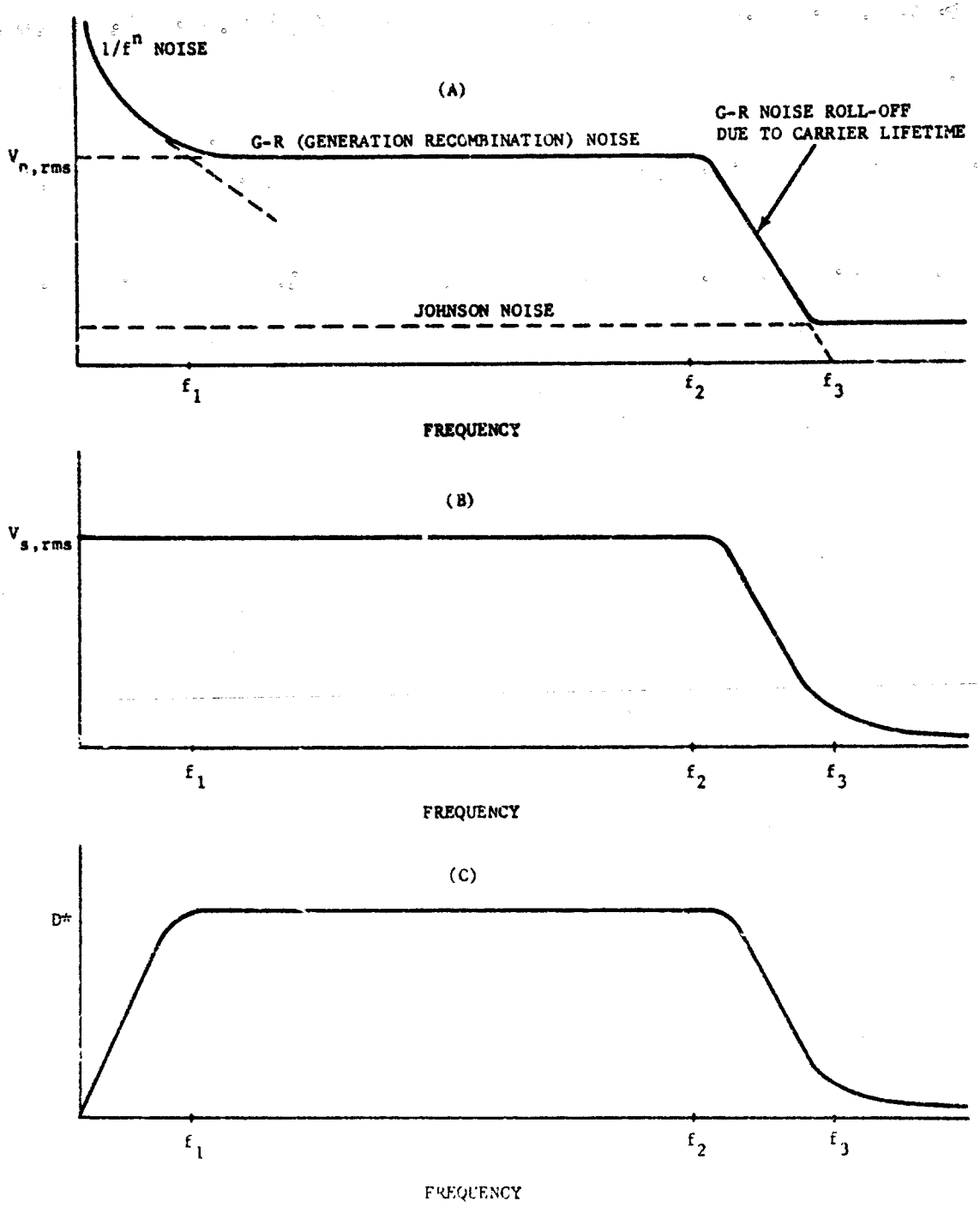


FIGURE 3-77. Typical Noise Voltage Spectrum and Signal Response of a Photoconductor

3-4.3.6.5 Current Noise

This noise is characterized by a $1/f^n$ spectrum[63], usually called $1/f$ noise. The mechanisms causing this noise are not well understood and probably encompass many causes. Some of the factors influencing $1/f$ noise are contacts, surface quality, crystalline dislocation, and bias current. This noise source is generally not dominant beyond frequencies of a few hundred Hz.

Since several of these noise mechanisms usually demonstrate themselves simultaneously at various frequencies in a detector, it is desirable to plot a noise voltage spectrum for the detector. This is a plot of absolute or relative $V_{n,\,rms}$ versus frequency.

The typical noise voltage spectrum and signal frequency response of a photoconductor are illustrated in Fig. 3-77(A) and (B). The signal-to-noise ratio is represented by Fig. 3-77(C).

The noise voltage spectrum of a detector can be determined by measuring $V_{n,\,rms}$ in narrow Δf increments over the frequency range of interest. This can be accomplished with a wave analyzer (narrow Δf, tunable voltmeter).

3-4.4 DETECTOR FABRICATION

3-4.4.1 General

Since the types of infrared detectors are so diverse, a thorough discussion of fabrication methods would be prohibitively lengthy for this handbook. Therefore, only some of the most important electro-optical detectors will be discussed.

3-4.4.2 Photon Detectors

Most infrared photon detectors are made from semiconductor materials, both intrinsic (such as Si, InSb, HgCdTe, etc.) and extrinsic (such as Ge:Hg, Ge:Cu, etc.). In some cases the semiconductor material is grown in large crystals; then sawed, lapped, and chemically etched into detector elements. In some other cases, a thin film of the semiconductor material is applied to a substrate by vapor-deposition, chemical precipitation, or epitaxial deposition.

A photoconductive detector, Fig. 3-57(B), is made by applying a properly prepared semiconductor element to a heat sink, and attaching bias electrodes and wires.

A photovoltaic detector is made by diffusing a p-n junction into a semiconductor element, mounting it to a heat sink, then electroding and wiring it, Fig. 3-57(C).

If the detector requires cryogenic cooling, it is attached to a cold finger in a Dewar housing which is equipped with a window which will admit IR signals.

Some of the critical parameters involved in making high-quality IR detectors include:

1. The purity, stoichiometry, and crystalline perfection of the semiconductor material

2. Closely controlled amounts of the proper activation impurities in extrinsic photoconductors and in photovoltaic detectors

3. Careful surface treatments such as lapping and etching to minimize surface noise

4. Application of low-noise electrodes to the element

5. Optimum thickness for maximum photon absorption

Arrays of detectors can be made by mounting several individual detectors onto a heat-sink, which is practicable method if only a small number of fairly large detectors is involved. A method which is particularly suited for fabricating arrays consisting of large numbers of small detector elements involves forming the array from a single monolithic crystal using a multi-blade saw. High-density arrays can also be produced by photo-etching or scribing techniques. All of these methods have been used to form detectors as small as a few thousands of a square inch in arrays consisting of hundreds of elements[66].

3-4.4.3 Thermal Detectors

The most common thermal detectors are thermocouples and bolometers. Radiation thermocouples are composed of a thin, blackened, gold foil behind which is attached a thermoelectric junction of dissimilar metals. The incident radiant power is absorbed by the blackgold foil and a temperature rise results which is sensed by the thermocouple. Low-mass foils and junctions are required for fast time response.

A thermistor bolometer contains a sensitive element which is a mixture of manganese, nickel, and cobalt oxides. This element, which is a temperature-sensitive semiconductor, is mounted to a heat-sink and electrodes and wires are attached. The surface of the sensitive element is blackened for maximum infrared absorption. A thermistor bolometer generally consists of two sensitive elements mounted to the same heat-sink, with only one of them exposed to the signal radiation. The other element acts as temperature-compensating load resistor.

In general, thermal detectors are designed with (1) low mass and good thermal conduction to the heat sink for the fastest time response, and (2) thermal isolation from the heat sink for maximum sensitivity.

3-4.5 COMPARATIVE PERFORMANCE OF INFRARED DETECTORS

The comparative performance of infrared detectors is presented in Table 3-11.

TABLE 3-11. COMPARATIVE PERFORMANCE OF INFRARED DETECTORS

MATERIAL	MODE	NOMINAL OPERATING TEMPERATURE, °K	λ_{max}, micron	$D^*_{500^\circ K_{BB}, 300^\circ K_{BG}}$, w^{-1} cm $Hz^{1/2}$	$D^*_{\lambda max, 300^\circ K_{BG}}$, w^{-1} cm $Hz^{1/2}$	$D^*_{\partial P}$, w^{-1} cm $Hz^{1/2}$	$\tau_{300^\circ K_{BG}}$, sec
1. PbS	PC	295	2.5	1×10^9	1×10^{11}	$> 2 \times 10^{11}$	3×10^{-4}
2. PbS	PC	195	2.6	1×10^{10}	5×10^{11}	$> 2 \times 10^{12}$	2×10^{-3}
3. PbS	PC	77	3.2	1×10^{10}	2×10^{11}	$> 2 \times 10^{13}$	2×10^{-1}
4. PbSe	PC	295	4.0	3×10^8	3×10^9	$> 1 \times 10^{10}$	2×10^{-6}
5. PbSe	PC	195	4.6	5×10^9	3×10^{10}		3×10^{-5}
6. PbSe	PC	77	5.0	5×10^9	2×10^{10}		4×10^{-5}
7. PbTe	PC	77	4.0	4×10^8	3×10^9		3×10^{-5}
8. Ge:Au	PC	77	6.0	2×10^9	5×10^9	$> 3 \times 10^{10}$	$< 1 \times 10^{-6}$
9. Ge:Zn	PC	4.2	36	4×10^9	1×10^{10}	$> 3 \times 10^{10}$	$< 1 \times 10^{-8}$
10. Ge:Hg	PC	27	10	1×10^{10}	2×10^{10}	$> 1 \times 10^{11}$	3×10^{-7}
11. Ge:Cu	PC	4.2	23	1×10^{10}	3×10^{10}	$> 1 \times 10^{11}$	$< 1 \times 10^{-6}$
12. Ge:Cd	PC	4.2	16	7×10^9	2×10^{10}		$< 1 \times 10^{-6}$
13. Ge-Si:Au	PC	50	7.3	3×10^9	7×10^9		1×10^{-7}
16. InSb	PC	77	5.4	1×10^{10}	5×10^{10}	$> 2 \times 10^{12}$	2×10^{-6}
17. InSb	PV	77	5.3	2×10^{10}	8×10^{10}	$> 8 \times 10^{11}$	$< 1 \times 10^{-6}$
18. InSb	PEM	295	6.2	1×10^8	3×10^8	$> 3 \times 10^8$	2×10^{-7}
19. InAs	PC	295	3.6	2×10^7	2×10^8	$> 2 \times 10^8$	2×10^{-7}
20. InAs	PV	295	3.5	3×10^8	3×10^9	$> 5 \times 10^9$	$< 2 \times 10^{-7}$
20a. InAs	PV	195	3.5	3×10^9	1×10^{11}	$> 2 \times 10^{11}$	$< 2 \times 10^{-6}$
20b. InAs	PV	77	3.1	3×10^9	2×10^{11}		$< 1 \times 10^{-5}$

TABLE 3-11. COMPARATIVE PERFORMANCE OF INFRARED DETECTORS (Continued)

MATERIAL	MODE	NOMINAL OPERATING TEMPERATURE, °K	λ_{max}, micron	$D^*_{500^\circ K_{BB}, 300^\circ K_{BG}}$, w^{-1} cm $Hz^{1/2}$	$D^*_{\lambda max, 300^\circ K_{BG}}$, w^{-1} cm $Hz^{1/2}$	D^*_{BP}, w^{-1} cm $Hz^{1/2}$	$\tau_{300^\circ K_{BG}}$, sec
21. InAs	PEM	295	2.5	2×10^7	2×10^8	$> 2 \times 10^8$	2×10^{-7}
22. Te	PC	17	3.5	4×10^9	6×10^{10}	$> 1 \times 10^{11}$	6×10^{-5}
23. Thermistor	SB	295	Flat	2×10^8	2×10^8	5×10^8	2×10^{-3}
24. Thermo-couple	TE	295	Flat	2×10^9	2×10^9	2×10^9	4×10^{-2}
25. Golay	GT	295	Flat	2×10^9	2×10^9	2×10^9	2×10^{-2}
26. NbN	SPTB	15	Flat	5×10^9	5×10^9	5×10^9	5×10^{-4}
27. Carbon Resistor	SB	4.2	Flat	5×10^9	5×10^9	5×10^9	1×10^{-2}
28. CdSe	PC	295	0.7		2×10^{11}	$> 2 \times 10^{11}$	1×10^{-2}
29. GaAs	PV	295	0.8		5×10^{11}	$> 5 \times 10^{11}$	1×10^{-3}
31. S-1	PE	295	0.8		3×10^{12}	$> 1 \times 10^{13}$	$< 1 \times 10^{-6}$
32. Si	PC	295	0.8		8×10^{12}	$> 8 \times 10^{12}$	4×10^{-6}
33. Si	PV	295	0.9		2×10^{13}	$> 2 \times 10^{13}$	$< 1 \times 10^{-6}$
34. Ge:Ga	SB	4.2	Flat	1×10^{10}	1×10^{10}	1×10^{10}	1×10^{-2}
35. (Hg-Cd)Te	PC	77	10		1×10^9	1×10^9	$\sim 1 \times 10^{-7}$
36. (Hg-Cd)Te	PV	77	10		1×10^9	$> 1 \times 10^9$	$\sim 1 \times 10^{-6}$

BB — Blackbody
BG — Background
BP — Best Published
GT — Gas Thermometer
PEM — Photoelectromagnetic
PC — Photoconductive
PV — Photovoltaic
SB — Semiconductive Bolometer
SPTB — Superconducting Phase Transition Bolometer
TE — Thermoelectric
PE — Photoemissive

(More detailed information concerning the performance of these detectors is contained in AMCP 706-128.)

3-5 COOLING SYSTEMS

3-5.1 REQUIREMENTS

Optimum performance in the intermediate- and far-IR wavelength regions requires that solid-state detectors and optical components such as telescope barrels, baffles, and mirrors be operated at very low (cryogenic) temperatures. Stable cryogenic temperature levels are achieved in the laboratories with relative ease by using gases such as helium, hydrogen, nitrogen, and oxygen whose boiling point temperatures when liquefied are 4°, 20°, 77°, and 83°K, respectively. The problem is to achieve and maintain the required cryogenic temperatures in the field with equipment generally encumbered by space, weight, power, and other limitations.

In addition to hardware limitations are performance requirements such as the temperature variation tolerances which generally range from 10° to 15°K. There is also the need to ensure structural integrity and optical alignment of the components and to minimize background noise during operation. Fulfilling these general requirements involves performing detailed thermal analyses; and evaluating methods of cooling, transient flow characteristics of heat, and the efficiency of selected insulating methods.

3-5.1.1 Detector Cooling

The data plotted in Fig. 3-78 illustrate the sensing capabilities versus operating temperature ranges of various detectors. Note that all of the detectors achieve maximum performance in the cryogenic temperature range below 200°K, and that half have operating temperatures below 77°K.

Achieving the required temperatures within the framework of minimum weight and volume, reliability, and simple operation is the task which confronts the designer.

3-5.1.2 Cooling of IR Optics

Although the cooling of optical components in an IR system does not involve the low temperatures required for detectors, the cooling load is generally much higher because of the greater mass and area involved. As mentioned previously, optical barrels, baffles, and mirrors must be maintained at certain temperatures for satisfactory operation of the IR system. These temperatures are a necessary factor in achieving the required low-background noise levels for detectors. Temperature gradients in optical components are extremely critical since excessive gradients can result in the distortion of optical or baffle surfaces, thereby, changing the optical characteristics of the system. Optics used in space vehicles can be cooled passively by radiators on the vehicle.

3-5.2 TYPES OF COOLING SYSTEMS

3-5.2.1 Direct Contact

Direct-contact cooling methods which apply the principles of heat of fusion or heat of vaporization of a material, or of radiation from a surface, are reliable and relatively simple to analyze. Selection of solidified or liquefied coolant and of the heat transfer mode is dictated to a major extent by the required temperature. Dual cryogenic systems are sometimes considered for long-life, minimum-weight applications in which one cryogen serves as a protective jacket for the other. Constant temperature can be maintained by melting or vaporizing the coolant through phase-change cooling methods. Conduction and radiation cooling will provide stable temperature levels as long as the radiator thermal environment is constant.

3-5.2.1.1 Liquid Coolant Systems

a. Description

Cooling by means of a "liquid cooling" system entails maintaining thermal contact between the coolant and the device to be cooled. Thermal contact can be achieved through the high-conductivity walls of the storage vessel or by means of a heat-conduction rod (cold finger) immersed in the fluid. Fig. 3-79 shows a diagram of a typical liquid coolant system.

b. Thermodynamics

The thermodynamics involved in the performance analysis of liquid coolant systems are relatively simple. The amount of liquid coolant required can be calculated from the heat balance equation

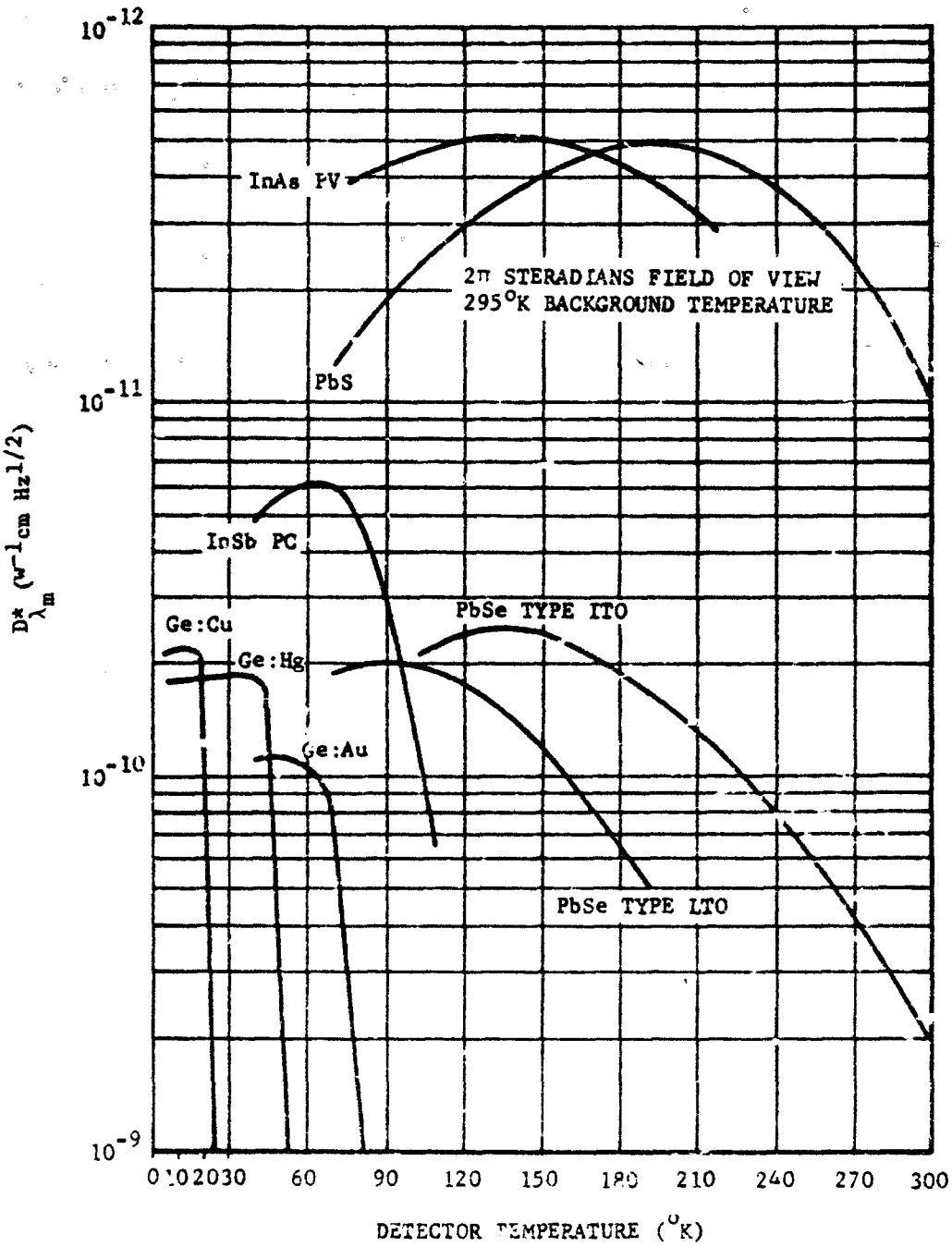


FIGURE 3-78. Detector Performance vs Temperature

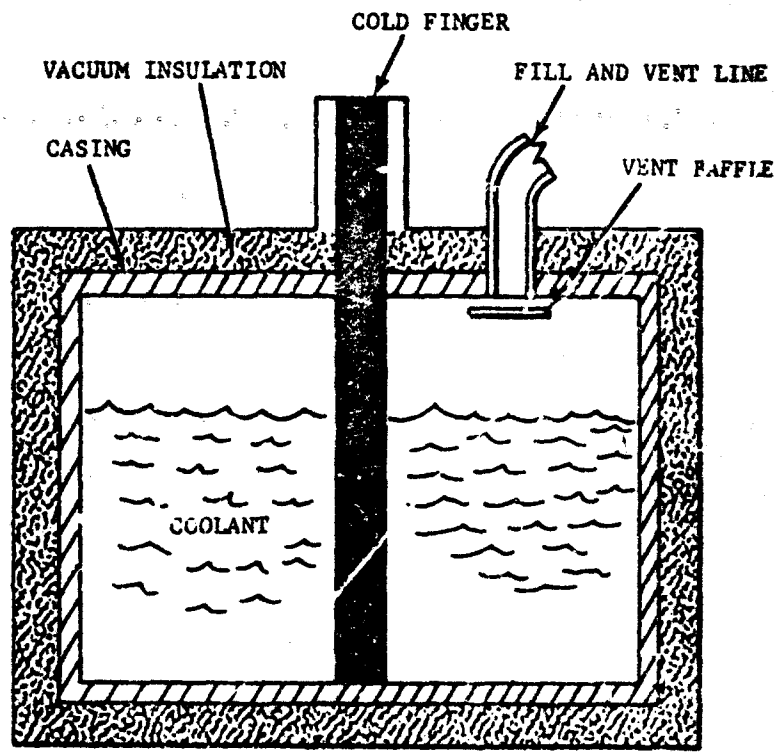


FIGURE 3-79. Liquid Cryogen Cooling System

$$W = \frac{3.41Qt}{h_v}, \text{ lb} \qquad (3\text{-}111)$$

where

W = weight of coolant required for the time of operation, lb
Q = total heat rate to the coolant (including that of the detectors, insulation, supports, and fill- and vent-ports), w
t = time, hr
h_v = latent heat of vaporization of the coolant, Btu/lb

3.41 Btu/hr = 1 w

(These as well as additional symbols used throughout this paragraph on cooling systems are listed in Table 3-12.)

The temperature level of an evaporative coolant can be varied to some extent by controlling the environment pressure in the coolant-supply tank. Vapor pressure curves vs temperature characteristics of the most common coolants are plotted in Fig. 3-80.

c. Range of Application

The principal limitation to the use of the liquid evaporative cooling system is in providing a sufficiently long operating life with a given heat load. In general, for flight-type IR cooling applications, this cooling system can be used only when the heat load is low or when the operating time is short.

d. Existing Problems

The behavior of two-phase fluids in a vessel that is subject to reorientation and to various acceleration forces is very complex. In addition, the vessel design and fluid properties (viscosity, surface tension) are extremely critical. If the thermal efficiency of the system is closely dependent on the liquid cryogen interface being adjacent to the device being cooled, positive methods of controlling the liquid (such as a capillary structure for holding the liquid for supercritical operation) must then be applied.

TABLE 3-12. STANDARD SYMBOLS FOR FAR. 3-5

SYMBOL	DEFINITION	TYPICAL UNITS
A	Heat flow area	cm^2
a	Reflected solar energy from planetary bodies (albedo)	
C	Thermal capacitance of a body	w-hr/°K
COP	Carnot coefficient of performance	
ΔT	Temperature difference	°K
e	Planetary emission	
g	Gold	
h_s	Heat of sublimation	Btu/lb
h_v	Latent heat of vaporization	Btu/lb
I	Current	A
K_t	Thermal conductivity	mw/cm°K
k	Kovar	
L	Heat flow-path length	cm
$\dot{m}$	Gas flow rate	g/sec
M	Molecular weight	g/mol
N_c	Compressor efficiency	
p	Body internal power generation	w
P	Pressure	arbitrary units
q	Heat flow	mw
Q	Total heat rate	w
Q_a	Absorbed incident solar reflectance from planetary bodies	
Q_e	Absorbed incident planetary emission	w
Q_r	Body emitted energy to space	w
Q_s	Absorbed incident solar energy	w
R	Universal gas constant	cal/g mol-°K
R_e	Electrical resistance	ohm
R_t	Thermal resistance	°K/mw
r	Radiation	
S	Seebeck coefficient	V/°K
T	Temperature	°K
tot	Total	
t	Time	hr
T_c	Cold-face temperature	°K
T_h	Hot-face temperature	°K
U	Thermal conductance	w/°K
W	Coolant required in weight for a given time	lb
W_c	Compressor power	w

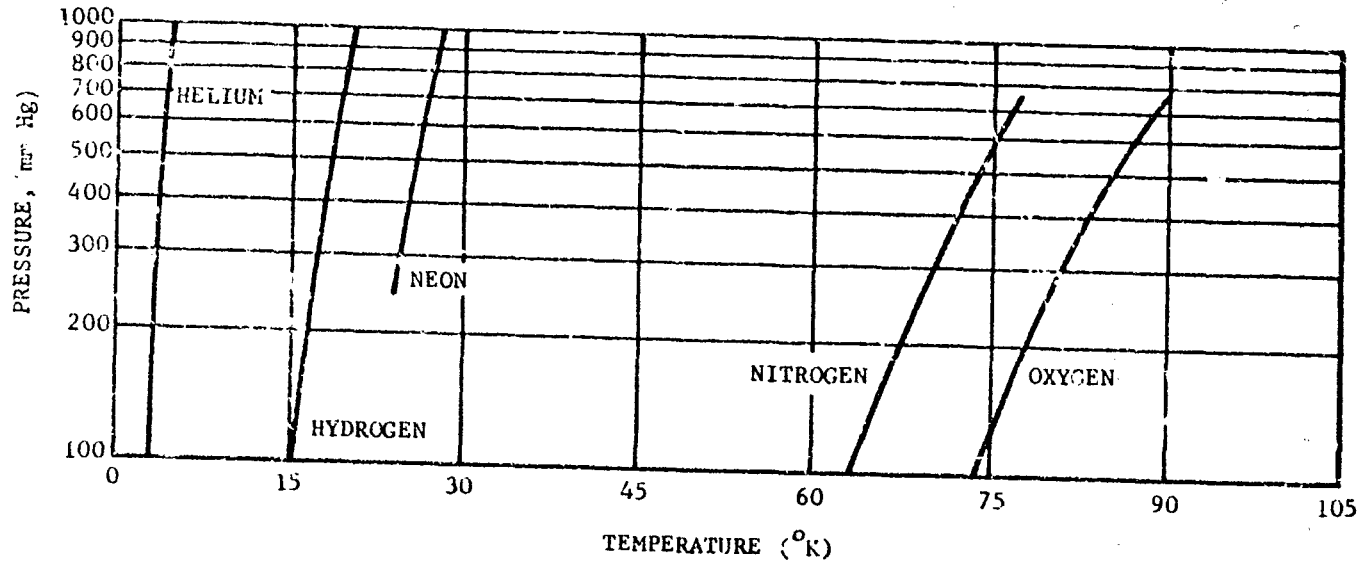


FIGURE 3-80. Vapor Pressure of Liquid Cryogens

3-5.2.1.2 Solid Coolant Systems

a. Description

Solid coolant system operation is based on the absorption of thermal energy through sublimation of a solidified coolant. A typical "solid cooler" arrangement designed for IR detector cooling is illustrated in Fig. 3-81. The temperature level of operation is dependent on the vapor pressure of the coolant. Vapor pressure curves of some potential solid coolants are given in Fig. 3-82. The problems due to the behavior of liquid coolant in a vessel during reorientation or to the variation in the acceleration forces on the cooler are avoided by the use of solidified coolants. Additional advantages of solidified coolants include lower temperatures than liquids, more coolant per unit volume, and additional cooling capability per unit mass.

b. Thermodynamics

The overall thermodynamic performance of solid coolers is developed from the following simple heat balance relationship based on the heat of sublimation equation:

$$W = \frac{3.41Qt}{h_s}, \text{ lb} \qquad (3\text{-}112)$$

where

W = weight of coolant required for the time of operation, lb
Q = total heat rate to the coolant (including detectors, insulation, supports, and fill- and vent-ports), w
t = time of operation, hr
h_s = heat of sublimation, Btu/lb

3.41 Btu/hr = 1 w

c. Range of Application

Solid coolers are most promising for applications requiring very low temperatures. Limitations include the size and weight of the cooler as dictated by the heat of sublimation and density of the selected coolant, the heat load, and the required length of operation. The requirement for a continuous vacuum in this cooling system may restrict its use in some applications.

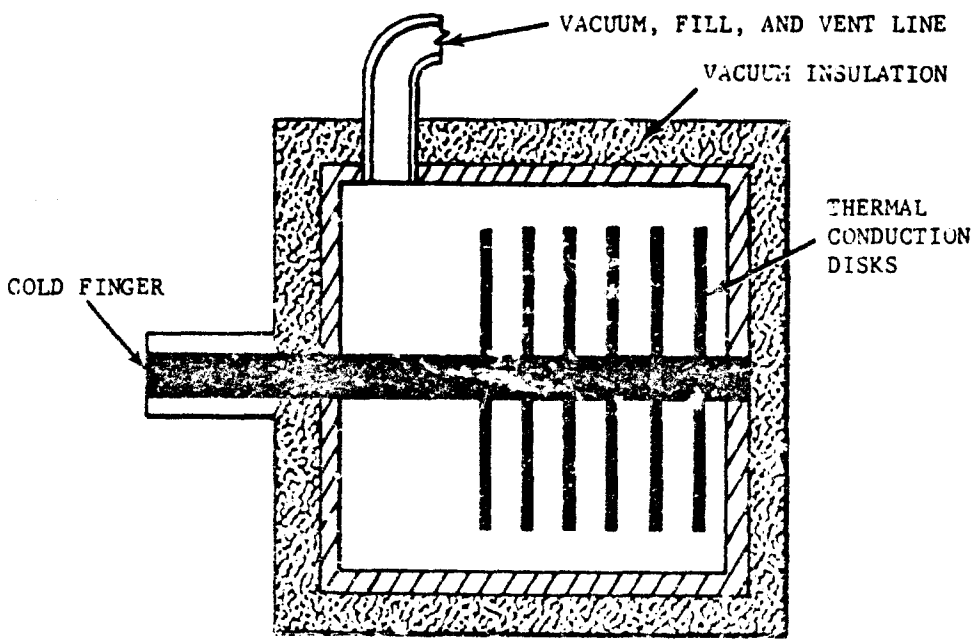


FIGURE 3-81. Solid Cryogen Cooling System

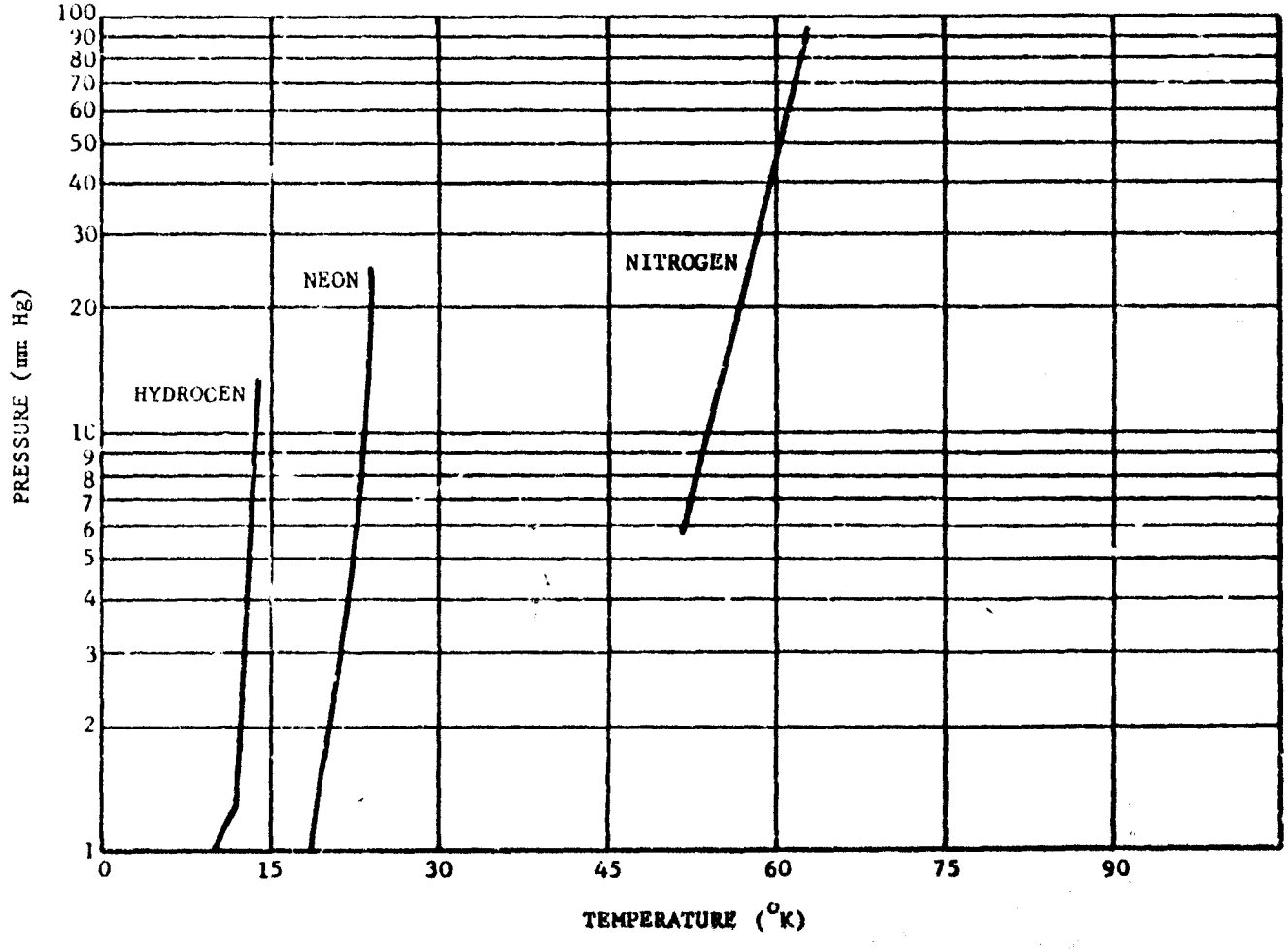


FIGURE 3-82. Vapor Pressure of Solid Cryogens

3-5.2.1.3 Passive Cooling

Passive, IR detector temperature-control techniques can be applied in an environment where advantage can be taken of the low, effective, space temperature. The technique entails controlling the balance of the absorbed external heat loads and the emitted energy to space for achieving the desired operating temperature.

The heat balance equation for a body in space is as follows:

$$C\frac{dT}{dt} = Q_s + Q_a + Q_e + p - Q_r \qquad (3\text{-}113)$$

where

C = thermal capacitance of the body, w-hr/°K
T = body temperature, °K
Q_s = absorbed incident solar energy, w
Q_a = absorbed incident solar reflectance from planetary bodies, w
Q_e = absorbed incident planetary emission, w
p = body internal power generation, w
Q_r = body emitted energy to space, w
t = time, hr

One of the most important factors influencing the stability of passive, temperature-control systems intended for long-term operation in space is the degradation of external thermal control coatings. Sources of degradation are: (1) ultraviolet component of solar radiation, (2) high-vacuum environment, (3) high-energy radiation and particles from the Van Allen belt and the sun, (4) atmospheric ions, protons, and electrons, and (5) micrometeorite erosion.

Passive techniques for achieving detector temperature control are limited, in general, to temperatures greater than -148°F (-100°C) and, thus, are quite restrictive with respect to the type of detector being cooled.

Since it is not possible to adjust the operating temperature of passive cooling systems in orbit, a great deal of care and skill must be applied during the design analyses and testing phases of the cooler to assure the verification of the required performance characteristics.

3-5.2.2 Joule-Thomson

The Joule-Thomson principle of obtaining low temperatures is commonly used in the industrial manufacturing of cryogenic fluids. It is the basis of many closed- and open-cycle cooling systems used for detector cooling at intermediate cryogenic temperature levels. The application of the Joule-Thomson principle to low cryogenic temperatures for detector cooling is being studied.

3-5.2.2.1 Closed Cycle

a. Description. The Joule-Thomson closed-cycle cooling system is shown schematically in Fig. 3-83. High-pressure gas from the compressor passes through regenerative heat exchangers to the throttling orifice where, by expanding to a lower pressure, part of the gas becomes liquefied. The cold unliquefied vapors and the vapors from evaporation of the liquid during heat addition return to the compressor after flowing through the regenerative heat exchangers.

b. Thermodynamics. A temperature-entropy diagram of the Joule-Thomson closed cycle is shown in Fig. 3-84. The cycle involves isothermal compression from 1 to 2, constant-pressure cooling process through the regenerative heat exchanger from 2 to 3, and adiabatic expansion from 3 to 4 through an orifice, constant temperature evaporation of the condensed liquid from 4 to 5, and constant pressure heating in the regenerative heat exchanger from 5 to 1.

The Carnot coefficient of performance COP is expressed as

$$COP = \frac{T_{cold}}{T_{ambient} - T_{cold}} \qquad (3\text{-}114)$$

where both temperatures are absolute.

The power required by the compressor can be determined from Eq. 3-115

$$W_c = \dot{m}\left(\frac{RT}{MN_c}\right) ln\left(\frac{P_{high}}{P_{low}}\right), \text{ w} \qquad (3\text{-}115)$$

where

W_c = compressor power, w
$\dot{m}$ = gas flow rate, g/sec
R = universal gas constant, cal/g mol-°K
T = temperature, °K
M = molecular weight, g/mol
N_c = compressor efficiency
P = pressure, arbitrary units

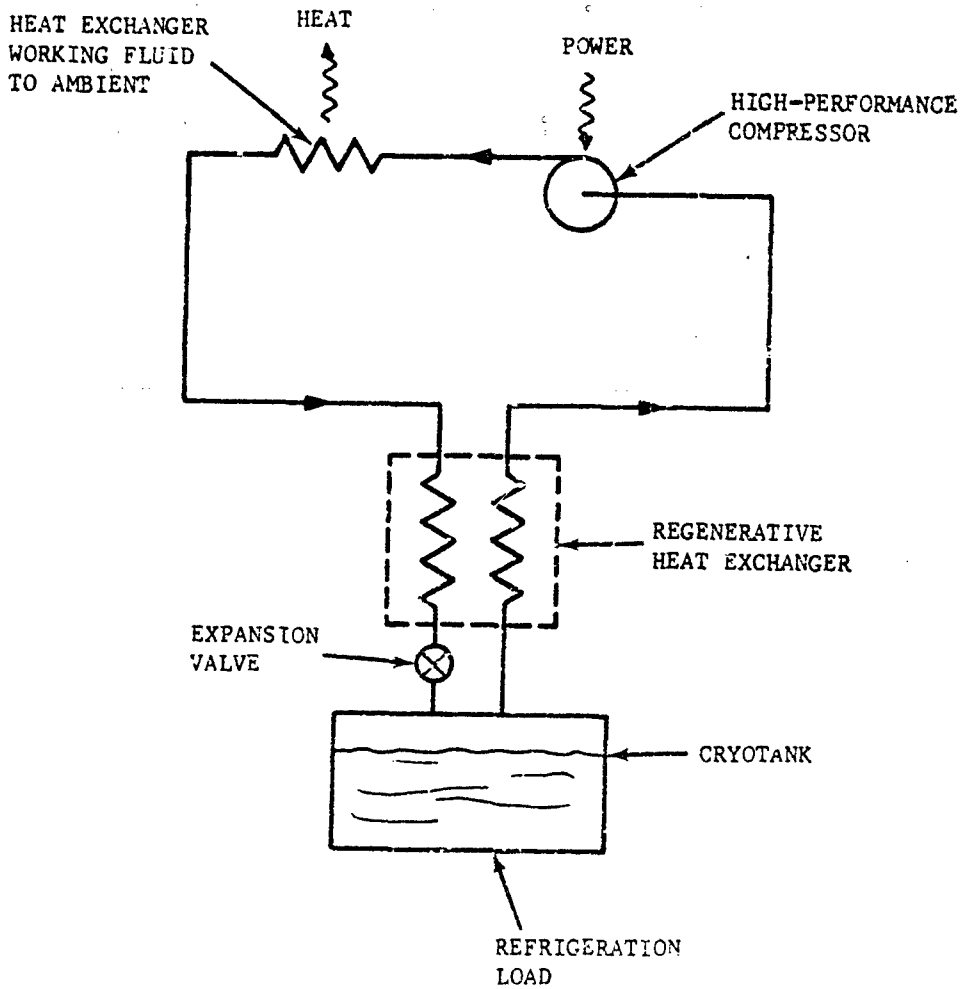


FIGURE 3-83. Closed Cycle Joule-Thomson System

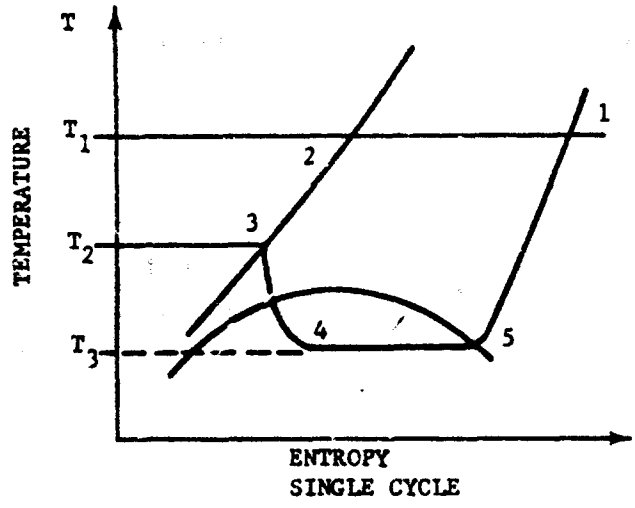


FIGURE 3-84. Joule-Thomson Closed Cycle System Temperature-entropy Diagram

3-5.2.2.2 Open Cycle

Open-cycle Joule-Thomson systems differ from closed-cycle systems in that gas is supplied from a high-pressure cylinder instead of a compressor, and expanded gas is not recovered.

3-5.2.3 Expansion Engine

Expansion of a gas in an engine where the gas performs work in the expansion process lowers the temperature of the gas. The cooling cycles described in the subsequent paragraphs are based on this principle.

3-5.2.3.1 Stirling Cycle

The Stirling Cycle cools by isothermal expansion of a gas in a smaller chamber called the cold head. For increased coefficient of performance, it makes use of a regenerator which is a thermal capacitor which alternately absorbs heat from, and rejects heat to, the gas stream and operates in four steps as illustrated in Fig. 3-85. The first step, insets (1) and (2) of Fig. 3-85, consists of compressing the gas at room temperature with heat removal. The gas is then transferred to the cold chamber through the regenerator, insets (2) and (3). The gas expands due to the piston and displacer retracting, inset (4), thus cooling the gas. The gas is finally transferred back through the regenerator to the warm chamber, insets (5) and (6). The effectiveness of the Stirling Cycle method of cryogenic cooling is illustrated in Fig. 3-86, where it is compared with three other cooling methods.

3-5.2.3.2 Claude Cycle (Reversed Brayton)

The Claude Cycle, Fig. 3-87, cools by means of gas expanded through a turbo-alternator from which useful electrical energy can be obtained. This cycle makes use of a staged compressor and an expander with a regenerative heat exchanger. The compressor is separate from the expander and can be located remotely.

3-5.2.3.3 Solvay Cycle (Gifford-McMahon)

Solvay Cycle operation is based on gas compression and expansion processes similar to those of the Stirling Cycle, but the compressor is located remotely and the cyclic process is achieved through valves located at the compressor. The system illustrated in Fig. 3-88 is relatively simple, and the weight of the refrigerating part of the system is competitive with other cooling systems.

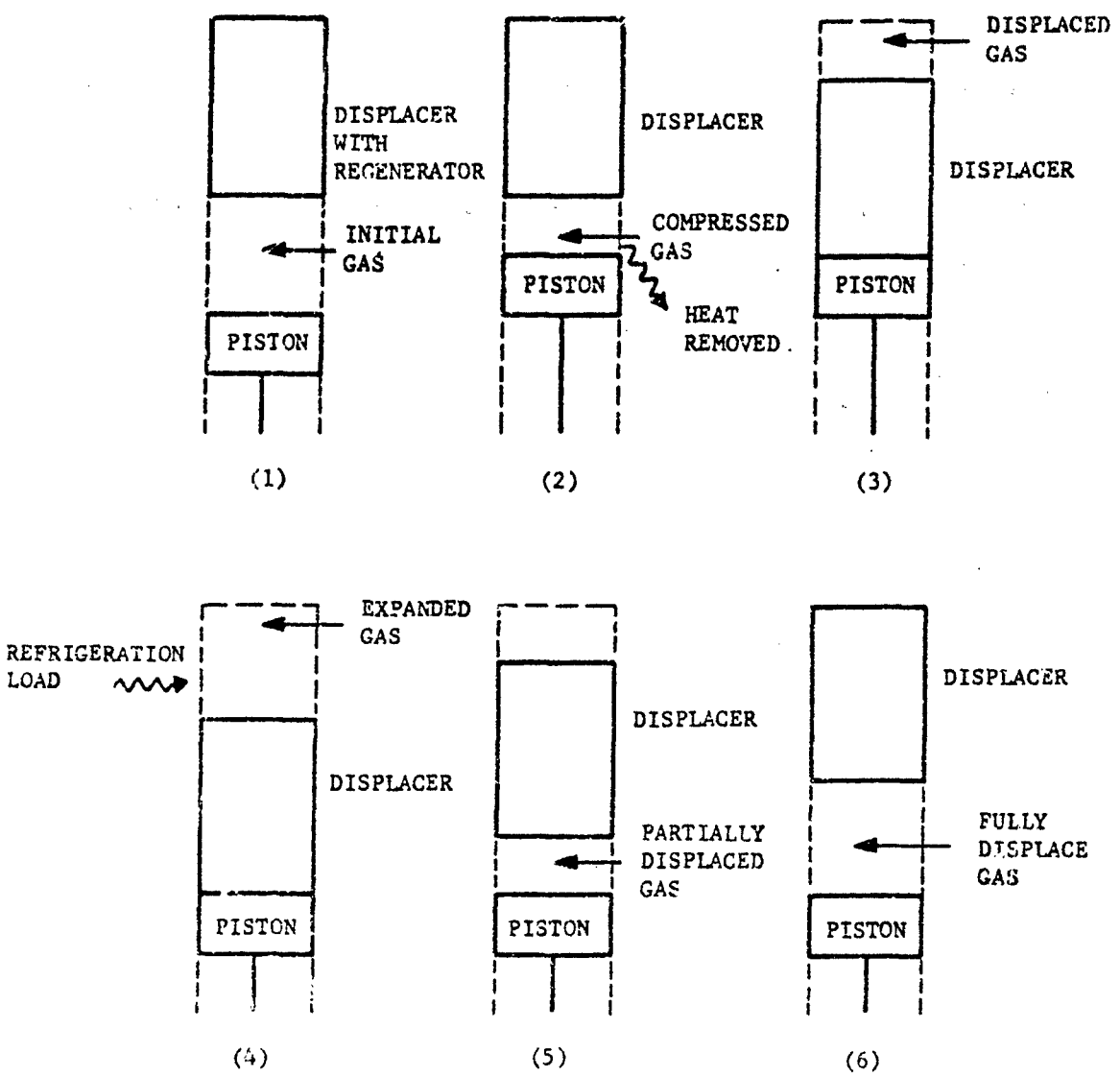


FIGURE 3-85. Closed Cycle Stirling System

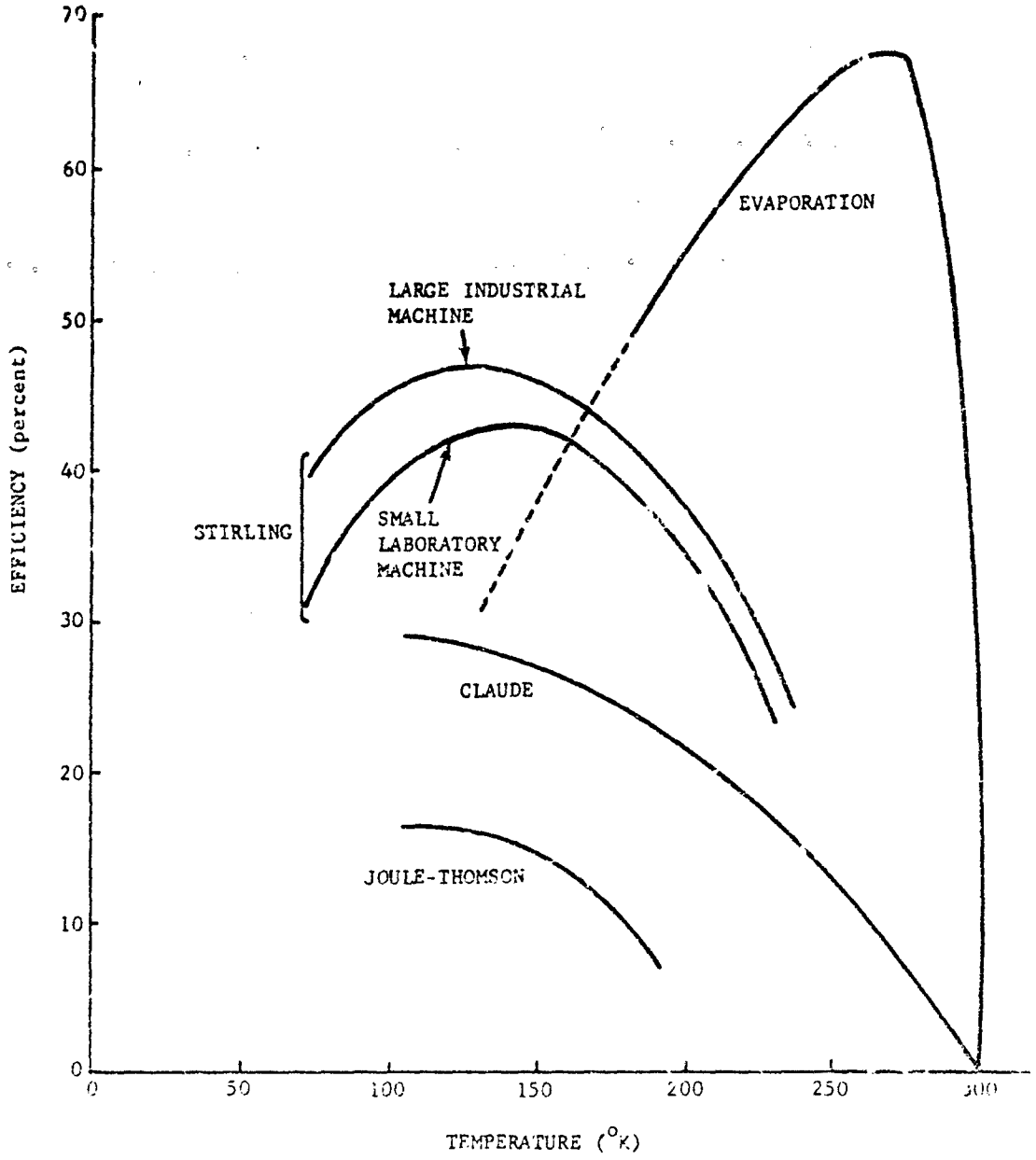


FIGURE 3-86. Refrigeration Efficiency vs Temperature

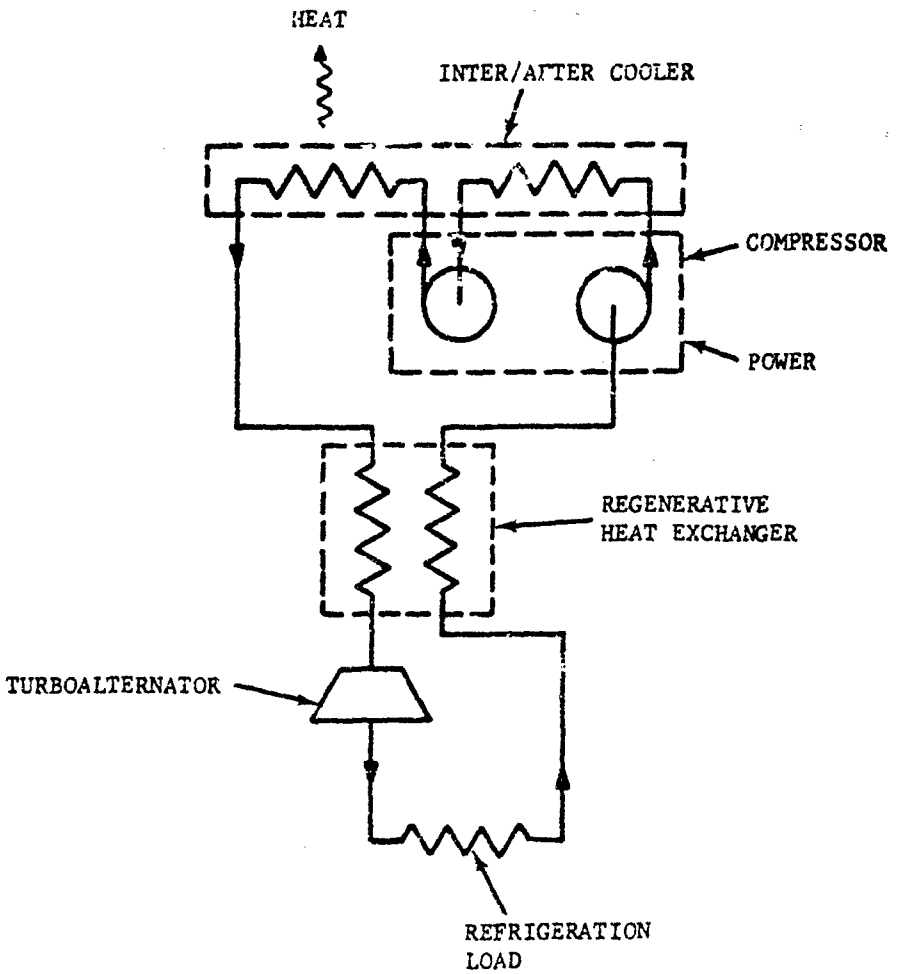


FIGURE 3-67. Closed Cycle Claude (Reversed Brayton) System

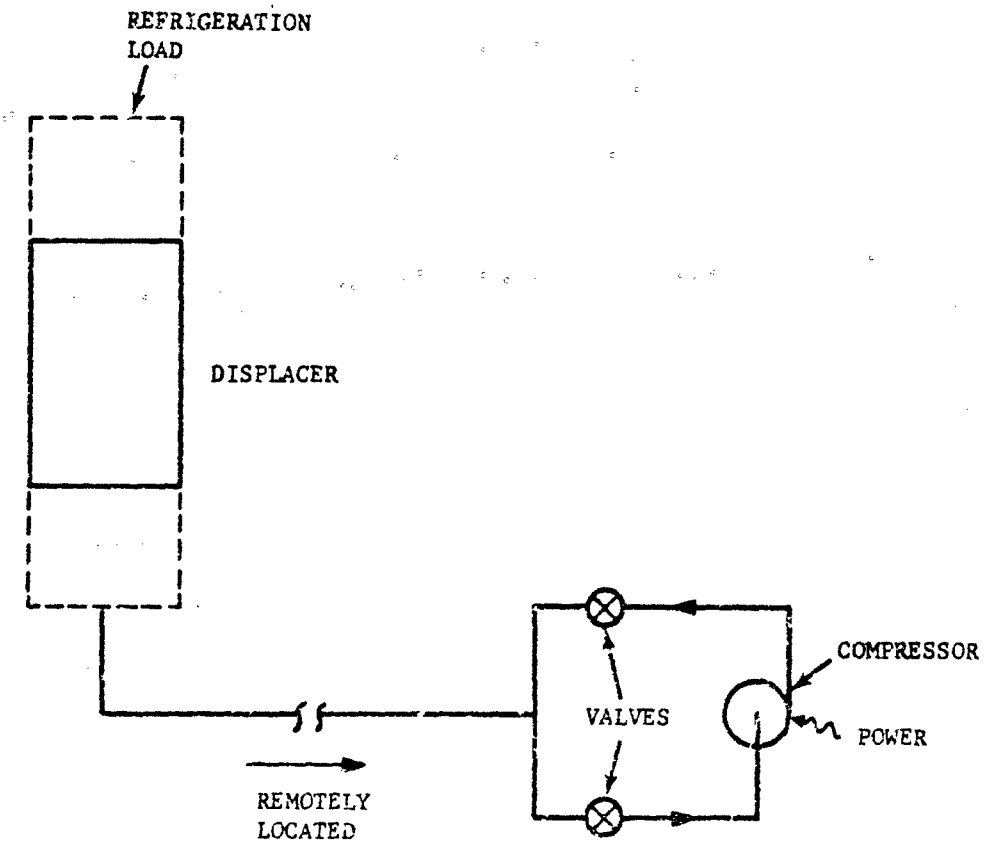


FIGURE 3.98. Closed Cycle Solvay (Gifford-McMahon) System

3.5.2.3.4 Vuilleumier Cycle

This heat-driven system has only recently been considered seriously for IR detector cooling application. The addition of heat replaces the conventional compressor and can take the form of electrical heat; isotope heat is also possible for long-life space applications. The Vuilleumier Cycle makes use of regenerators and displacers in much the same way as does the Stirling Cycle. A schematic of the system is shown in Fig. 3-89.

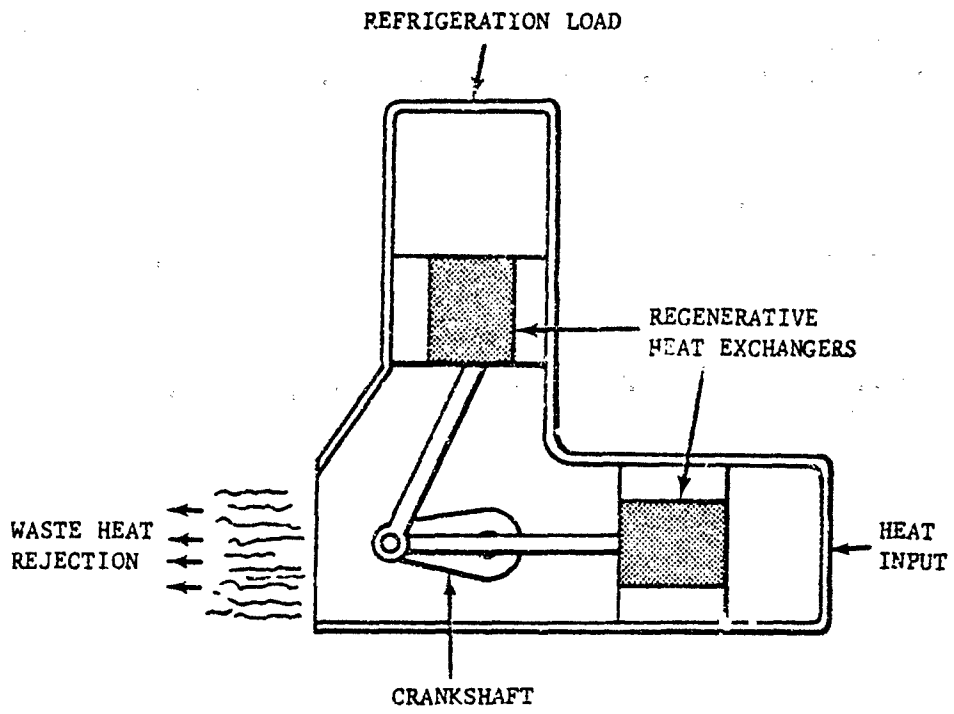


FIGURE 3-89. Vuilleumier Cycle Cooling System

3-5.2.4 Thermoelectric

Thermoelectric cooling is based on the absorption of heat which results when an electric current flows through a junction of two dissimilar metals. This is called the Peltier Effect whereby the current creates a temperature differential. Heat is absorbed at the low-temperature junction and is rejected at the high-temperature junction.

In the performance analysis of thermoelectric elements, the cooling capacity of the element is expressed as

$$Q = (S)IT_c - \frac{1}{2}(I^2R_e) - U(T_h - T_c) \quad (3\text{-}116)$$

where

Q = heat pumped, w
S = Seebeck coefficient, V/°K
I = current, A
T_c = cold-face temperature, °K
R_e = electrical resistance, ohm
U = thermal conductance, w/°K
T_h = hot-face temperature, °K

The power required to obtain the above cooling capacity is calculated as

$$Power\ Required = I^2R_e + (S)(T_h - T_c)I, \text{ w} \quad (3\text{-}117)$$

Thermoelectric cooling systems at present achieve temperatures down to 200°K from an ambient-temperature environment. Below this temperature they are not competitive with other systems in terms of power requirements and weight, using existing materials.

3-5.2.5 Other Cooling Systems

The various other types of cooling systems described in the paragraphs which follow are potentially suited for use in IR cooling applications.

3-5.2.5.1 Pulse Tube

In the pulse tube cooler the expansion of a gas from a dead-ended tube is used to reduce the temperature. The idealized performance of the pulse tube cannot be attained due to various

TABLE 3-13. COMPARISON OF COOLING TECHNIQUES

	Size	Weight	Power, w	Life	Temp, °K	Cooling Capacity, w	Heat Rejection	Working Fluid	Cooldown Time	Remarks
Liquid	1 ft³	10 lb	None	20 min	10	4	Vaporization	Helium	12 hr	Highly variable, depending on mission
Solid	12 in. D × 18 in.	22 lb	None	1 yr	65	40×10^{-3} 76×10^{-3}	Sublimation	Methane Carbon Dioxide	24 hr	Highly variable, depending on mission, Cryogenic Technology Jan/Feb '68
Passive	1 ft²/w for radiator	10 lb/w	None	1 yr	50	10	Radiation to space	None	1 to 30 days	Highly variable, depending on mission
Joule-Thomson	6.5 in. D × 12 in.	18 lb	650	500 hr maintenance	77	5	Air cooled	Nitrogen	3 min	Cyrogenic Engineering News 6/67
	10 in. × 12 in. × 13 in.	50 lb	1,050	500 hr maintenance	30	0.35	Air cooled	Nitrogen Neon	28 min	Air Products Corp. Data
Stirling	5 in. D × 14 in.	15 lb	375	500 hr maintenance	29	1	Air cooled	Helium	10 min	Cyrogenic Engineering News 6/67
	3 in. D × 12 in.	5 lb	30	500 hr maintenance	77	1	Air cooled	Helium	8 min	Malaker Corp. Data
Claude	55 ft³ 524 ft³	965 lb 19,000 lb	3,400 75,000		4.5 4.3	2.1 200				Advances in Cyrogenic Engineering, Vol. 12
Solvay	7 ft³	500 lb	5,000	500 hr maintenance	4.2	5	Air cooled	Helium	30 min	In development stage—Advances in Cyrogenic Engineering, Vol. 12
Vuilleumier	8 in. D × 30 in.	10 lb		Indefinitely	15		Liquid nitrogen cooled	Helium		In development stage—Advances in Cyrogenic Engineering, Vol. 9
Thermoelectric	¾ in. × 1 in. × 1 in.	2 lb for power supply	3	Indefinitely	240	0.3	Forced convection	None	5 min	Highly variable, depending on mission, Borg-Warner Handbook
Pulse Tube	¼ in. D × 4 in.	20 lb	335	Indefinitely	170	1.75 @ 220°R	Water cooled	Freon		In development stage—Advances in Cryogenic Engineering, Vol. 12
Vortex Tube				Indefinitely						In development stage
Adsorption				Minutes	4	1				In development stage

losses in the system, including heat conduction from the hot end of the cylinder to the cold end. However, the process has been demonstrated, and the simplicity of the system warrants further study.

3-5.2.5.2 Vortex Tube

The Vortex Tube, sometimes called the Ranke-Hilsh Tube, is a steady-state operating device which operates on the same basic physical principle as does the pulse tube. Its application is restricted to fulfilling the higher temperature requirements associated with cooling of optics rather than detectors. In its present stage of development, the Vortex Tube is not competitive in performance with other more conventional cooling methods. Its efficiency is low, and high pressure losses occur.

3-5.2.5.3 Adsorption

Coolants adsorbed on a surface will decrease in temperature when the adsorbent surface is exposed to lower pressures. This technique is of interest for providing the low temperatures required for cooling IR detectors. It is simple and has good standby characteristics.

3-5.2.6 Comparison of Cooling Systems

Table 3-13 lists the characteristics of the various cooling systems discussed for comparison. Many of the systems listed are highly variable in size, weight, power consumption, life, minimum obtainable temperature, and cooling capacity. The entries in the table represent typical systems. For the most part, data were obtained from available literature.

Although the liquid coolant system listed in the table is a short-life device, it is possible to produce liquid cooling systems which have lifetime ratings greater than one year. Cool-down time for the system shown is highly dependent on the minimum temperature. Helium which is used has the lowest temperature of all liquids, therefore it requires the longest time for initial cooling. The choice of coolant depends, of course, on the temperature desired.

Coolants different from those listed for liquid, solid, Joule-Thomson, and pulse-tube systems can be chosen, depending upon mission requirements. Temperatures will vary for the different coolants, and life will also vary as a function of the size of the system.

3-5.3 COOLING SYSTEM SELECTION

In order to select the proper cooling system for a particular application, it is necessary to:

1. Analyze cooling requirements

2. Compare performance characteristics of the candidate cooling systems with application requirements

3. Select systems which appear to meet the requirements

4. Develop design data (weight, volume, power requirements, etc.) pertaining to selected systems

5. Rate cooling systems in terms of application potential

6. Select system which provides desired characteristics and requires fewest trade-offs

Various factors involved in selecting IR cooling systems are as described in the example which follows.

Example

Consider as a typical example, a cryogenic cooler intended for use with an air-to-ground missile system. The first step in the design of such a cooling system is the calculation of heat loads. Assume a detector temperature of 27°K is required with ambient temperature of 300°K. The eight detectors comprising the IR sensor are to be mounted in a glass Dewar which has an access channel 0.440 in. in diameter, 0.040 in. thick, and 3.0 in. long. Each of the eight detectors requires an electrical lead and a common ground wire. These 0.040-in. diameter Kovar leads are attached to the detectors by means of 0.002-in. diameter gold paper wires each 0.25 in. long.

Notation for Example

A = heat flow area, cm^2
L = heat flow path length, cm
K_{total} = thermal conductance, mw $°K^{-1}$
K_t = thermal conductivity, mw $cm^{-1}\,°K^{-1}$
Q = heat flow, mw
R_t = thermal resistance, $°K\ mw^{-1}$
T = temperature, °K
T_{avg} = average temperature, °K
ΔT = temperature difference, °K
σ = Stefan-Boltzmann constant, 5.67×10^{-12} w $cm^{-2}\,°K^{-4}$

Subscripts

a = gold
g = glass
k = Kovar
l = leads
r = radiation

Conduction Along Glass

$$T_{avg} = (300 + 27)/2 \approx 165°K$$

Thermal conductivity K_{tg} of glass at $T = 165°K$ is:

$$K_{tg} = 8.10 \text{ mw } cm^{-1}\,°K^{-1}$$

The glass Dewar has an access channel of 0.440 in. in diameter and 0.040 in. thick. This gives a cross-sectional area A approximately equal to:

$$A_g \approx \pi(D_{inside} + 0.040) \times 0.040 \approx 0.0602 \text{ in}^2$$

The thermal conductance K_{total} along the glass Dewar is therefore:

$$K_{total} = K_t \times \frac{(area)}{length}$$

$$= \frac{8.1(0.0602 \times 2.54)^2}{3 \times 2.54}$$

$$= 0.41 \text{ mw } °K^{-1}$$

The thermal resistance R_{tg} of the glass Dewar is:

$$R_{tg} = \frac{1}{K_{total}} = \frac{1}{0.41} = 2.42\ °K\ mw^{-1}$$

$$\Delta T = 300 - 27$$

$$= 273°K$$

The total heat rate Q_g is therefore:

$$Q_g = K_{total} \times \Delta T$$

$$= \frac{\Delta T}{R_{tg}}$$

$$= \frac{273}{2.42}$$

$$= 112 \text{ mw}$$

Conduction Along Leads

Conduction along the leads follows a similar analysis. The results of the analysis follow:

K_{tk} = 120 mw/cm°K (Kovar at $T = 165°K$)
$A_k = \pi(0.040)^2/4 = 0.00125 \text{ in}^2$
L_k = 3.0 in.
H = 9 (number of leads including 8 detector leads and 1 common ground)
$R_{tk} = (3)/[(120)(0.00125)(2.54)(9)]$
$= 0.877°K\ mw^{-1}$
K_{ta} = 4300 mw $cm^{-1}\,°K^{-1}$ (gold at $T = 27°K$)
$A_a = \pi(0.002)^2/4 = 3.14 \times 10^{-6} \text{ in}^2$.
L_a = 0.25 in.
$R_{ta} = 0.25/[(4300)(3.14 \times 10^{-6})(2.54)(9)]$
$= 0.811°K\ mw^{-1}$
ΔT = 273°K
$Q_l = \Delta T/(R_{tk} + R_{ta})$
$= 273/(0.877 + 0.811)$
$= 161$ mw

Radiation to Detectors

$$A = \pi[0.44 + 2(0.04)]^2/4 = 0.212 \text{ in}^2$$

$$(Q_r)_{max} = A\sigma(\Delta T)^4$$
$$= (5.67 \times 10^{-12})(0.212)(2.54)^2[(3 \times 10^2)^4 - (0.27 \times 10^2)^4]$$
$$= 63 \times 10^{-3} \text{ w}$$
$$= 63 \text{ mw}$$

Total Heat Load

$$Q_{total} = Q_g + Q_l + (Q_r)_{max}$$
$$= 112 + 161 + 63$$
$$= 336 \text{ mw}$$

Cooler Selection

A closed cycle Stirling refrigerator would be recommended for this application. It would weight 16 lb, measure 5 in. in diameter and 13 in. in length, and require 400 w of power. Justification for this choice is:

1. Closed cycle coolers provide the most trouble-free operation of all coolers, requiring only power and periodic maintenance.

2. Stirling refrigerators are the lightest, most compact, and most efficient coolers made to provide the required temperature of 27°K.

Alternative Cooling Systems

Other methods for cooling detectors to 27°K include two-stage Joule-Thomson expansion refrigerators and Gifford-McMahon coolers. Both were rejected for this particular application in which the missile is not released from the plane but held captive for a series of tests.

Joule-Thomson refrigerators were rejected because they are less compact and less efficient than Stirling Cycle coolers and present unique contamination problems that require greater care in handling and operating.

Gifford-McMahon coolers were rejected because they are also less compact than are the Stirling coolers; their principal advantage being a cold head assembly which can be remotely located from the compressor. For this application, such a feature is not an advantage.

3-5.3.1 Design Criteria

It is important that the design criteria for cooling systems be established early in the evaluation study in order to ensure that the design effort is directed to the best suited system. Because the large number of methods available for cooling IR systems offer such a diversity of external requirements, design features, and performance characteristics; many can be eliminated early in the application study if realistic design criteria for the application have been established.

The criteria include independent parameters such as temperature and total cooling load, while others are dependent parameters such as weight and volume. Independent parameters are set by the application and, therefore, must be met by any system that is considered in the evaluation. Dependent variables are flexible, depending on the designer's selection of materials and arrangement.

3-5.3.1.1 Weight

For most airborne and space cooling systems, weight is important. Preliminary comparison of cooling systems for a given application will generally be made on the basis of weight. The lightest system would be selected from those that meet the general requirements.

3-5.3.1.2 Volume

The volume of a cooling system for IR applications is frequently a major factor in the selection of a system. The compactness and the design arrangement of the IR system itself is of primary importance, and the cooling system is generally a secondary factor in the overall system design. In some installations the space left for the cooling system dictates that the volume be limited, and that it be installed in a particular arrangement.

3-5.3.1.3 Durability

Durability becomes a factor when the system is subjected to long-term operation [illegible] conditions. For long-term applications, the simple evaporative or sublimative cooling systems are not suitable because of excessive volume and weight, therefore, a mechanical system would be selected. Durability of mechanical systems is generally a function of the mechanical design of the moving components in which wear can occur.

3-5.3.1.4 Logistics

Dependence of a system on ground support is a major factor where readiness and remoteness are involved. This becomes a major factor where long standby conditions are required. The use of liquid neon or liquid helium would seriously handicap operation under these requirements unless portable liquefiers were available. Where stringent logistic requirements exist, use of mechanical cooling systems which only require electrical power is most desirable. Logistic requirements are independent factors dictated by the application. Little can be done by the coolant system engineer to improve these requirements.

3-5.3.2 Reliability

The simplicity of the evaporative or sublimative coolant systems implies the probability of highly reliable performance once the physical operation is verified under actual or simulated operating conditions. The reliability of a mechanical system having reciprocating or rotating parts will be a function of its mechanical design. Operating times of 300 to 500 hr without maintenance are now being achieved in the field with closed-cycle systems. At present, reliability of cooling systems has been obtained by overdesign (large weight and volume). The continuing need to provide lower temperatures along with the more stringent weight and volume requirements for IR cooling systems dictates the need for engineering advances which will result in a better compromise between these factors.

3-5.3.2.1 Mechanical Design

Mechanical design of the cooling system must consider the reliability of the design concepts used. Mechanical reliability of solid and liquid coolant systems with no moving parts will depend on the design and operation of valves used for filling and venting of the coolant. Systems having reciprocating components must be designed for high degree of freedom from failure caused by thermal and mechanical shock. This is a particularly critical factor for very low-temperature operation where the ductility of most materials is poor.

3-5.3.2.2 *Structural*

Structural reliability of cooling systems is of major importance where high-pressure operation is encountered. Factors such as thermal stresses, thermal shock, and fatigue must be considered in the structural design of each system. The method of supporting coolant systems against "g" loads must be done with consideration for minimum heat-leak and minimum weight.

3-5.3.2.3 *Vibrational*

IR cooling systems are often subjected to vibrational loads when installed in a vehicle. Mechanical cooling systems generate vibration due to the reciprocating and rotary motion of the equipment. It is essential that the cooling systems be capable of withstanding the external vibration loads imposed on them. It is also important that the cooling system frequencies not be harmonics of external frequencies which would allow excessive loads to be developed. Vibrational reliability of a cooling system can only be developed by means of environmental tests which simulate the conditions expected in the application under consideration.

3-5.4 INSULATION

Since IR cooling systems operate in the cryogenic temperature range, thermal isolation of low temperature components from the higher temperature environment is a prerequisite for efficient operation. The lower the operating temperature, the more critical this insulation problem becomes. For flight and space vehicles this thermal isolation must be achieved with minimum weight.

A combination of several insulating methods which, individually, provide a thermal barrier for one mode of heat transfer is being considered for insulation of cryogenic systems. Vacuum barriers are used to reduce convective and conduction heat transfer between surfaces. Radiation transfer is reduced by using multiple radiation shields between two radiating surfaces or between a surface and space. Low thermal-conductivity materials are used to reduce heat conduction in areas where physical contact must be maintained to transmit structural loads.

3-5.4.1 Vacuum Insulation

Use of vacuum is prevalent for cryogenic systems, but vacuum by itself does not prevent radiation heat transfer. Therefore, vacuum insulating systems are generally used in conjunction with radiation barriers.

3-5.4.2 Radiation Shields

Radiation shielding has been a primary method of reducing radiation heat transfer for vacuum-jacketed cryogenic systems. The technology of fabricating and using radiation shields has accelerated in the past decade and, at present, many materials are used for radiation shields depending on the application requirements. Available materials for radiation shielding include, from polished aluminum to aluminum- or gold-coated Mylar as well as other very thin, lightweight plastics. Typical shield thickness is less than 1 mil, and blankets of radiation shields with separators such as silk or glass fiber may be 1 or 2 in. thick, typically.

3-5.4.3 Superinsulation Materials

Superinsulation is made up of a large number of radiation shields, sometimes as many as 100, arranged in the form of a blanket which is then wrapped around the surface to be insulated. High performance and lightweight characteristics of superinsulations are a result of using thin, lightweight shields with high reflectivity in conjunction with thin, low-thermal-conducting spacers. Effective thermal conductivities for superinsulations vary from 0.35 $\mu w\ cm^{-1}°K^{-1}$ to ten times this value, depending upon the materials used and the methods of applying the insulation to the surface being insulated.

The insulating characteristics of superinsulating blankets vary as a function of pressure. If the thin radiation shields are arranged very close together, they may contact each other under pressure thus providing a thermal-conducting path.

3-5.4.4 Support of Insulating Materials

Severe degradation in the overall performance of superinsulating blankets can occur if judicious design of the insulation supports is not performed. High thermal conductivity through the coolant container supposts can nullify the required insulation performance. Successful support design requires the design of individual supports to provide low heat leak and a container arrangement whereby a minimum number of

supports are required. Total heat leak to an insulated surface can easily be increased by an order of magnitude if care is not exercised in insulation support design.

3-6 SIGNAL PROCESSING

3-6.1 INTRODUCTION

To maximize the signal amplitude at the output of an optical sensor with respect to the amplitude of background clutter and system noise, it is necessary to take advantage of the following distinctions in the target and background properties:

1. Spectral radiance
2. Spatial (geometrical) qualities
3. Signal frequency spectrum vs noise power spectrum

The spectral radiance of targets and backgrounds is discussed in Chapter 2. The criteria for design of an optimum spectral filter is described in Chapter 5. Par. 3-6 deals with techniques for target signal enhancement and background discrimination on the basis of their spatial and temporal variations. The first technique, spatial filtering, is discussed in par. 3-6.2 and the latter, temporal filtering, is discussed in par. 3-6.3.

Spatial filtering involves modification of the signal by the telescope spread functions and by the detector geometry. Temporal filtering is a result of the frequency response of the detectors and electronics and, finally, of the detection or display circuits. Consequently, the output signal is affected by the transfer functions of all of these elements. Since the transfer function of most IR components is linear, the total system transfer function is obtained from the transfer function product of the individual elements. These functions include:

1. Spatial transfer function of the telescope
2. Spatial and temporal transfer function of the detector
3. Transfer function of the electronics
4. Transfer function of the detection or display circuits

The conventional technique for describing the transfer function of each element entails the use of the Fourier transform. This transform describes the input-output relationship in terms of the system impulse response which is the transform of an input δ-function (i.e., the response to a point source input for spatial function and a time impulse for temporal function). As mentioned previously, the overall system transfer function is simply the product of the transfer function of each independent element (optics, detector, filter amplifier, displays, etc.) as

$$T_s(\omega) = \prod_{i=1}^{n} T_i(\omega) \qquad (3\text{-}118)$$

where

$T_s(\omega)$ = overall system transfer function
$T_i(\omega)$ = transfer function of each ith element
ω = frequency variable
n = number of elements arranged in tandem

Since IR optical components are described in terms of their spatial frequency response and the electronic components by their temporal frequency response, the combined optical-spatial frequency is multiplied by the scanning rate of the IR system in order to determine the temporal frequency. For example, if the scanning rate is 100 rad per sec and a target of interest is one cycle per rad, then the corresponding temporal frequency is calculated as

$$(1 \text{ cycle/rad}) \times (100 \text{ rad/sec}) = 100 \text{ cycles/sec or } 100 \text{ Hz} \qquad (3\text{-}119)$$

3-6.2 SPATIAL FILTERING

Spatial filtering is one of the more novel and promising techniques for detecting objects of interest in the presence of unwanted background radiation. As the name implies, spatial filtering involves detection and discrimination on the basis of shape and size of an object, and of the surrounding background. By dividing the field of view of the IR instrument into areas that correspond to the size of the predicted target, most background (extended source) objects are rejected since they are considerably larger than the areas of the field occupied by the (point source) target.

Spatial filtering may also involve signal conditioning whereby the frequency of the irradiance upon the detector may, if necessary, be converted to whatever frequency can be detected and processed with optimum results. This is done by chopping (interrupting) the radiation by means of a reticle thereby shifting the frequency, typically from dc to a few hundred Hz. In the follow-up discussion, the transfer function of the optics is described first followed by the detector transform and its spatial filtering effect.

3-6.2.1 Optical Transfer Function (OTF)

The optical transfer function (OTF) is the two-dimensional Fourier transform of the point spread function. Thus, if $O(x,y)$ were the object function, $I(x',y')$ the image function, and $S(x,y)$ the point spread function; the non-coherent image formation would be determined by the convolution integral

$$I(x',y') = \iint S(x'-x; y'-y)O(x,y)dx\,dy \qquad (3\text{-}120)$$

Eq. 3-120 can be expressed in the transform domain as

$$I(\omega_x,\omega_y) = \text{two-dimensional transform of image intensity function } I(x',y') \qquad (3\text{-}121)$$

where

$T(\omega_x,\omega_y)$ = two-dimensional transform of the point spread function $S(x,y)$ or transfer function of the system

$O(\omega_x,\omega_y)$ = two-dimensional transform of object intensity function $O(x,y)$

ω_x,ω_y = spatial frequencies in x- and y-directions, respectively, rad

In general, $T(\omega_x,\omega_y)$ is a complex quantity

$$T(\omega_x,\omega_y) = |T(\omega_x,\omega_y)| \exp[\phi(\omega_x,\omega_y)] \qquad (3\text{-}122)$$

where

$|T(\omega_x,\omega_y)|$ = amplitude function

and

$\phi(\omega_x,\omega_y)$ = phase function.

In the special cases of an aberration-free system or for on-axis symmetrical aberrations, the transfer function has no phase shift and is a real function. Asymmetric aberrations lead to complex-valued transfer functions.

The OTF is the result of the diffraction effect and the aberration effect. The diffraction effect is a function of the aperture size and the wavelength of light that forms the image. The wavelength of light λ and the smallest dimension of an aperture determines the lens cutoff spatial frequency (the lens transfer function falls to zero). The cutoff frequency, which is independent of the aperture shape, is defined as D/λ radians where D is the diameter of a circular aperture or the smallest dimension of a slit or other shape aperture, and λ is the wavelength of the image forming light. The terms D and λ are in the same units. For hetero-chromatic images which form light, the longest wavelength gives the largest or worst-case value for the diffraction limit.

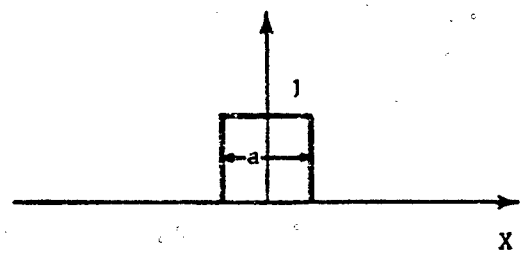
The aberration effect is the result of imperfections in the design and in the fabrication of the optical elements.

3-6.2.1.1 Point Spread Function (PSF)

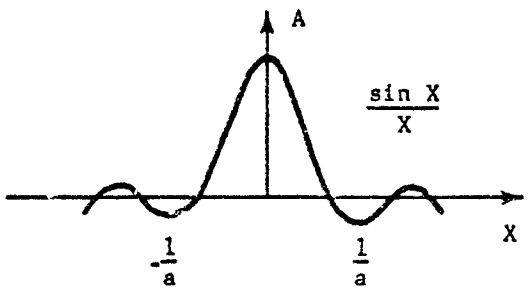
The point spread function (PSF) is the two-dimensional Fraunhofer image of a point source, normalized to unit volume. It is a mathematical description of the illuminance distribution of the point-source image in the x- and y-dimensions at the image plane which is normal to the optical axis of the lens systems.

3-6.2.1.1.1 Modulation Transfer Function (MTF)

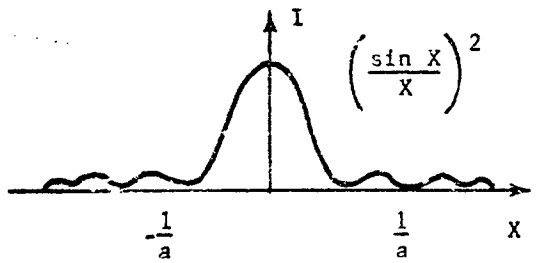
Modulation transfer function (MTF) is the amplitude spectrum of the Fourier transform of a normalized point spread function. It is the ratio of the amplitude at the output of an optical system to the amplitude at the input as a function of spatial frequency. The PSF and MTF of a diffraction limited optical slit are illustrated in Fig. 3-90. Fig. 3-90(A) represents the transmission of the slit aperture. The Fraunhofer image of a point source viewed through the slit aperture is shown in Fig. 3-90(B). The intensity distribution of this image, or point spread function, is illustrated in Fig. 3-90(C). The Fourier transform of Fig. 3-90(C) is the optical transfer function whose amplitude function or MTF is plotted in Fig. 3-90(D).



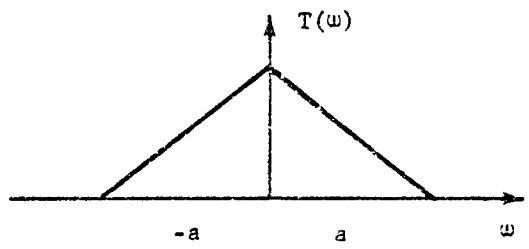
(A) TRANSMISSION OF A RECTANGULAR SLIT



(B) FRAUNHOFER IMAGE OF A RECTANGULAR SLIT



(C) POINT SPREAD FUNCTION OF A RECTANGULAR SLIT



(D) MODULATION TRANSFER FUNCTION OF A RECTANGULAR SLIT

FIGURE 3-50. Point Spread Function and Modulation Transfer Function of a Rectangular Slit

3-6.2.1.1.2 One-dimensional Image Analysis

In the case of systems that scan in only one direction, and for purposes of simplified computation and practical measurements, the one-dimensional image function is used.

The PSF can only be expressed in two-dimensional space, consequently, an analogous function, line spread function (LSF), is required for one-dimensional image analysis.

3-6.2.1.2 Line Spread Function (LSF)

The LSF is a mathematical description of the illuminance distribution of the image of a line source in the x-dimension at the image plane.

The LSF, $L(x')$ in direction x, is an integration of the total radiance of a section through the point spread function defined as

$$L(x') = \int_{-\infty}^{\infty} S(x',y')dy' \qquad (3\text{-}123)$$

The Fourier transform of the LSF is a complex quantity consisting of a modulus (or amplitude) and a phase. The phase function is a measure of the lateral displacement of the image from the geometric center of the optical system as a function of spatial frequencies. The modulus of the transfer function is a measure of the ratio of the amplitude of spatial sine wave output to the amplitude of the spatial sine wave input which is termed the modulation transfer function (MTF). The maximum value of the transfer function amplitude $T(\omega_x)$, which occurs at $\omega_x = 0$, is assigned a normalized value of 1 which denotes 100 percent transmission at zero spatial frequency.

3-6.2.1.3 Resolution Criteria and MTF

The angular resolution for a diffraction-limited system is the angular subtense of the radius of the Airy disc. For a circular aperture, the Rayleigh criteria of resolution in radians is expressed as

$$\theta_c = \frac{3.83\lambda}{\pi D} \approx \frac{1.22\lambda}{D} \qquad (3\text{-}124)$$

where 3.83 is the first zero of the Bessel function of the first order in radians, λ is the wavelength of light, and D the diameter of the optical aperture. The radius of the Airy disc in inches is given as

$$X_r = \frac{1.22\lambda}{D}(f\ell) \qquad (3\text{-}125)$$

$$= 1.22\lambda \ (\text{f/no.}), \text{ in.}$$

The diameter of the Airy disc, also in inches, is given as

$$X_d = \frac{2.44\lambda}{D}(f\ell) \qquad (3\text{-}126)$$

$$= 2.44\lambda \ (\text{f/no.}), \text{ in.}$$

where

$f\ell$ = optical focal length

For a slit aperture

$$\theta_s = \frac{\lambda}{D}, \text{ rad} \qquad (3\text{-}127)$$

$$X_r = \frac{\lambda}{D}, (f\ell), \text{ in.} \qquad (3\text{-}128)$$

3-6.2.2 Detector Spatial Filtering

Spatial filtering can also be achieved by means of spot scanning devices, reticles, arrays of detectors, and combinations thereof.

The mathematical representation of filtering effect of a detector is complicated by the one-dimensional (time) characteristic of the detector output having to describe the three-dimensional spatial-temporal scene radiance. The detector output is represented mathematically as

$$D(t) = \iint S(x,y,t)R(t)dx\,dy \qquad (3\text{-}129)$$

where the scene radiance is represented by $S(x,y,t)$ and the spatial filter is described by a weighting function of space and time $R(x,y,t)$.

The methods of analysis are described in the subsequent paragraphs where examples are given.

3-6.2.2.1 Analyses

Spatial filtering for detection purposes can be analyzed in the time domain as well as in the frequency domain.

3-6.2.2.2 Analysis in the Time Domain

Any known $S(x,y,t)$ intensity function and $R(x,y,t)$ weighting function can be analyzed simply by simulating Eq. 3-129 on a computer. The values of parameters affecting $R(x,y,t)$ such as scan rates, slit widths, and detector arrays may be changed as required to determine their effect, thereby, permitting the selection of

optimum values for the parameters. In a similar manner the intensity function, $S(x,y,t)$, can be varied to demonstrate the effects of input conditions, such as background or transmission characteristics. In addition, the simulation can include the downstream processing through electronic filters, transmission, display, logic processing, and decision making. This time-domain analysis by simulation is widely used today on digital, analog, or hybrid computers.

3-6.2.2.3 Analysis in the Frequency Domain

Detector spatial filters may also be analyzed in what is called the frequency domain. This consists basically of taking the Fourier transform of both sides of Eq. 3-129 with respect to x, y, and t. This transforms the independent variables from x,y, and t to k_x,k_y, and f, where k_x and k_y represent frequencies of sinusoidal signals in x- and y-direction, and f refers to sinusoidal signals in time. The advantage of this method is that Eq. 3-129 is changed from an integration to a product of the transformed quantities, which is easier to analyze than an integration. An additional advantage of the frequency or transformed domain analysis is that both the input, $S(x,y,t)$, and the output, $D(t)$, are often described in the frequency domain. The optics can be described more simply in terms of transformed quantities such as the optical transfer or modulation transfer function. The transform of $S(x,y,t)$ is obtained by transforming the source or object intensity times the optical transfer function. Also, the downstream processing (usually electronic) is often described in terms of its frequency response or transfer function. Thus, the output can be described by the transform of $D(t)$ times the electronic transfer function. The entire system from input source intensity to processed output can be treated in the frequency domain, which utilizes the analysis techniques developed in electronic signal-processing applications.

Another advantage of frequency domain analysis is that the signals in the transformed domain contain the same information as is in the space and time domain; therefore, it may be transmitted and the original scene reconstructed by a second transformation.

The analysis of general problems involved in the frequency domain method is quite complex. Pars. 3-6.2.2.4 and 3-6.2.2.5 describe two cases which are simple, yet practical. Excellent, thorough, and extensive treatment of spatial filter analysis by the frequency domain method is contained in Refs. 67 and 68.

3-6.2.2.4 Spatial Filtering Neglecting the Time-effect

Assuming the scene radiance $S(x,y,t)$ is independent of time, i.e., $S(x,y)$, only; the Fourier transform $\underline{S}$ of $S(x,y)$ would then be

$$\underline{S}(k_x,k_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(e^{-j2\pi k_x x})(e^{-j2\pi k_y y})S(x,y)dxdy \qquad (3\text{-}130)$$

This transform describes the intensity pattern in terms of the spatial frequencies k_x, k_y, cycles/unit length.

The Fourier transform $\underline{R}$ of the spatial filter $R(x,y)$ is

$$\underline{R}(k_x,k_y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(e^{-j2\pi k_x x})(e^{-j2\pi k_y y})R(x,y)dxdy \qquad (3\text{-}131)$$

This transform describes the spatial filter in terms of spatial frequencies, k_x and k_y. Both $\underline{R}(k_x,k_y)$ and $\underline{S}(k_x,k_y)$ are complex quantities containing amplitude and phase descriptions.

The detector input is considered as the sum of all spatial frequencies due to the integral of the SR product. It must be noted that the detector input due to any k_x,k_y spatial frequency cannot be considered as simply the product of $\underline{S}\,\underline{R}$, since the phasing may not be zero. However, the contribution to the detector cannot be any greater than the product of the amplitudes. This establishes an upper limit so that any frequency component of $\underline{S}$ will not contribute more than $\underline{S}\,\underline{R}$ to the detector. Thus if $R(k_x,k_y)$ is zero at any spatial frequency k_x,k_y; the detector output would produce a zero amplitude at that frequency. This attenuation of spatial frequency signals in S by $\underline{R}(k_x,k_y)$ is similar to filtering in electronic circuits. The paragraphs which follow describe some of the most commonly used spatial filters in IR systems. Refs. 67, 68, and 69 give detailed descriptions and transfer functions of various spatial filters.

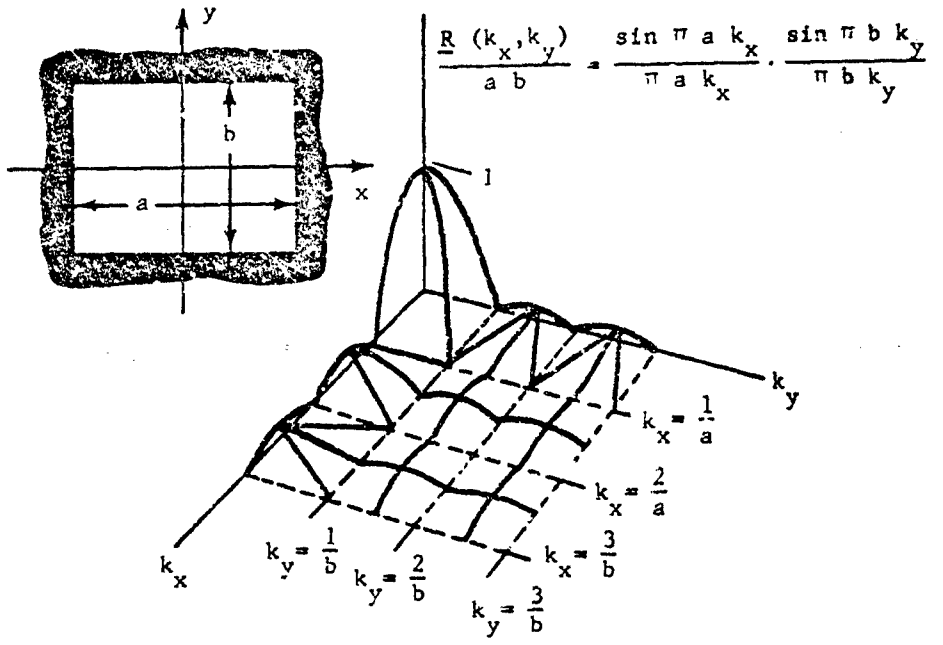


FIGURE 3-91. Spatial Filter Amplitude Response of Rectangular Detector

3-6.2.2.5 Spatial Filter of a Detector Scanning in One Dimension Only

Consider a simple spatial filter scanning in the x-direction at a constant velocity v, neglecting y variations in $S(x,y)$. The output of the detecto can now be written as

$$D(t) = \int S(x)R(x - vt)dx \qquad (3\text{-}132)$$

Note that the detector output in the frequency domain is simply the transform of $D(t)$. The transform of the right hand side of Eq. 3-132 is shown[68] as

$$D(f) = \frac{1}{v}\underline{S}\left(\frac{f}{v}\right)\underline{R}^*\left(\frac{f}{v}\right) \qquad (3\text{-}133)$$

where f is frequency in Hz; $\underline{S}$ and $\underline{R}^*$ are the one-dimensional transforms of S and R, respectively, with f/v substituted for k_x; and $\underline{R}^*$ is the complex conjugate of $\underline{R}$.

Thus for scanning system, the detector output in the frequency domain is the product of the transform of the intensity and the spatial filter transform. Since many systems scan with the spatial filter at a constant velocity, Eq. 3-133 has wide application.

3-6.2.2.6 Rectangular Detector of Width a and Length b

The Fourier transform of a detector of width a and length b is

$$\underline{R}(k_x,k_y) = \int_{-a/2}^{a/2}\int_{-b/2}^{b/2}\left(e^{-j2\pi(k_x x + k_y y)}\right)dx\,dy$$

$$= \int_{-a/2}^{a/2}(e^{-j2\pi k_x x})dx \int_{-b/2}^{b/2}(e^{-j2\pi k_y y})dy$$

$$= \left(\frac{\sin \pi k_x a}{\pi k_x}\right)\cdot\left(\frac{\sin \pi k_y b}{\pi k_y}\right) \qquad (3\text{-}134)$$

or

$$\frac{R(k_x,k_y)}{ab} = \left(\frac{\sin \pi a k_x}{\pi a k_x}\right)\cdot\left(\frac{\sin \pi b k_y}{\pi b k_y}\right)$$

where x and y are coordinates in the width and and length direction, respectively; k_x and k_y are spatial frequencies in cycles/unit in width and length directions, respectively. This transfer function means that sinusoidal components of the intensity function in the x-direction of frequency k_x will be attenuated by $\frac{\sin \pi k_x a}{\pi k_x}$ and similarly for sine waves in the y-direction. (See Fig. 3-91.)

Note that the transfer function has zeros for $k_x a = n$ and $k_y b = m$ where n and m are integers.

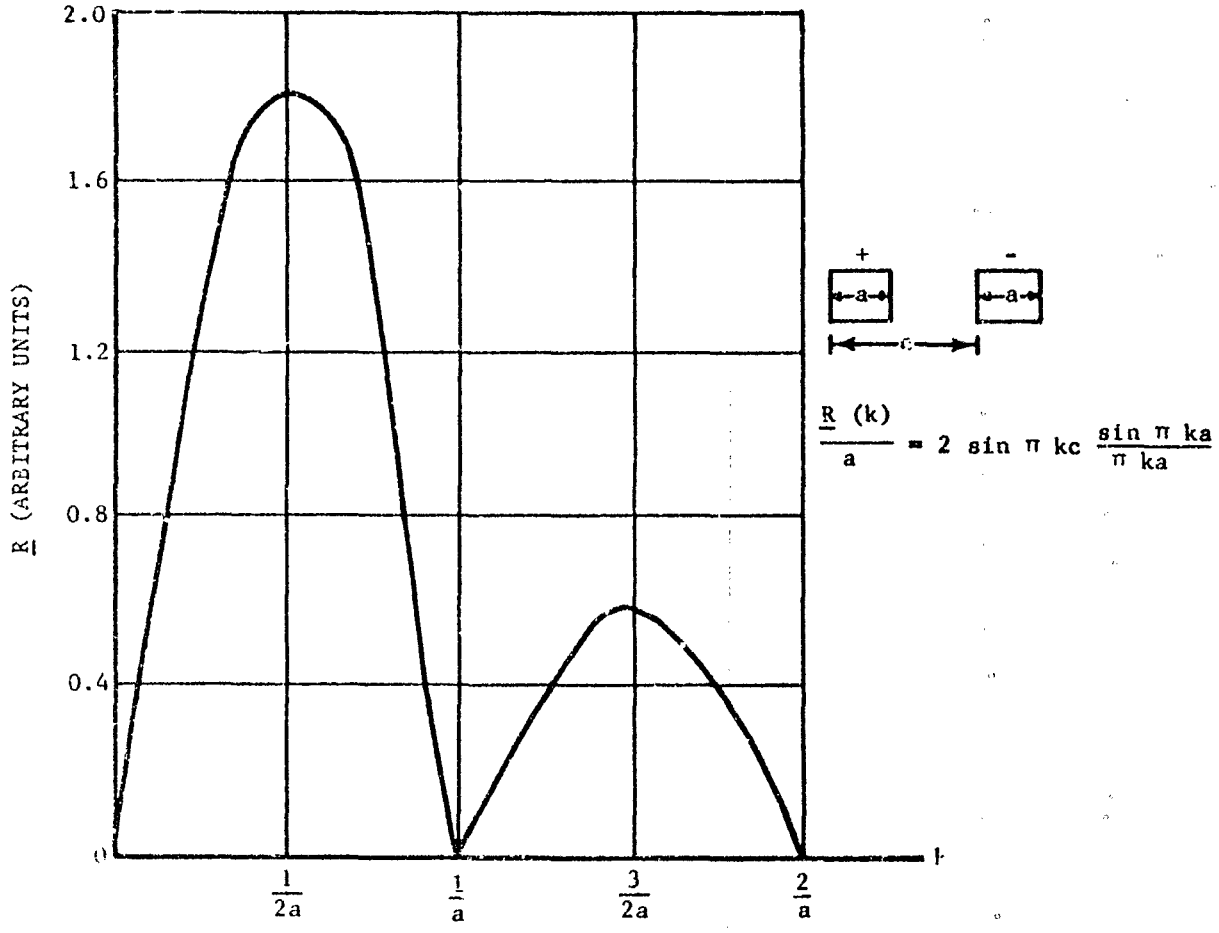


FIGURE 3-92. Spatial Frequency Amplitude Response (R) of a Plus-minus Detector Spaced One Detector Width Apart (c=2a)

3-6.2.2.7 Plus-minus Detectors

In many detection systems, the spatial discrimination technique can be employed due to the use of multi-detector arrays. For detection of a point source target, for example, it is desirable to process the information through a band-pass filter since the background-clutter-noise dominates in lower frequency. One method of eliminating the low-frequency signals is to use two detectors separated by a distance c between centers and by connecting the detectors in opposite polarity. The transfer function of the second detector is the same as the first detector multiplied by $e^{-j2\pi k_x c}$ due to the delay c. Thus,

$$\underline{R}(k_x,k_y) = (1 - e^{-j2\pi k_x c})\left(\frac{\sin \pi k_x a}{\pi k_y}\right)\left(\frac{\sin \pi k_y b}{\pi k_y}\right) \tag{3-135}$$

$$= e^{-j\pi k_x c}(e^{j\pi k_x c} - e^{-j\pi k_x c})\left(\frac{\sin \pi k_x a}{\pi k_x}\right)\left(\frac{\sin \pi k_y b}{\pi k_y}\right)$$

$$= 2je^{-j\pi k_x c}(\sin \pi k_x c)\left(\frac{\sin \pi k_x a}{\pi k_x}\right)\left(\frac{\sin \pi k_y b}{\pi k_y}\right)$$

The term $je^{-j\pi k_x c}$, being simply a phase function, may be dropped. The filter amplitude response is, therefore,

$$\underline{R}(k_x,k_y) = 2(\sin \pi k_x c)\left(\frac{\sin \pi k_x a}{\pi k_x}\right)\left(\frac{\sin \pi k_y b}{\pi k_y}\right) \tag{3-136}$$

Fig. 3-92 shows the bandpass characteristic of this type of spatial filter.

3-6.2.2.8 Circular Detectors

Circular detectors have the same characteristic in all directions. The filter amplitude response can be shown as

$$\underline{R}(k) = \frac{rJ_1(2\pi rk)}{k} \tag{3-137}$$

where

r = radius of the circle
k = spatial frequency in radial direction, cycles/unit length.
J_1 = Bessel function of the first kind and first order.

Fig. 3-93 illustrates the normalized amplitude response of such a circular detector.

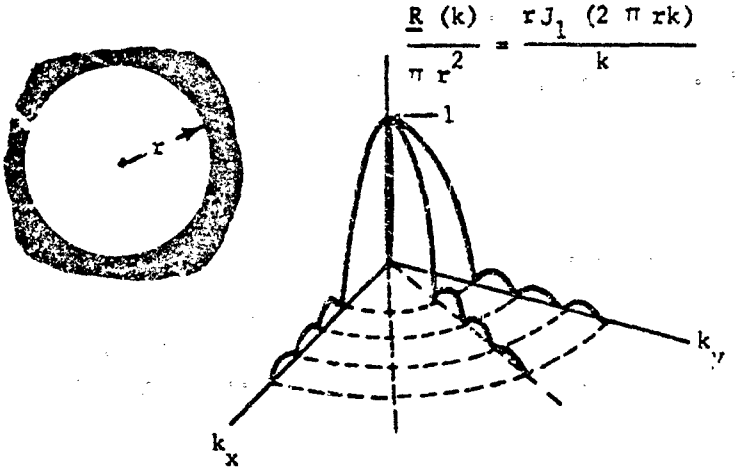


FIGURE 3-93. Spatial Amplitude Response of Circular Detector

3-6.3 ELECTRONIC SIGNAL PROCESSING

Requirements for greater IR system sensitivity have led to the development of highly sensitive (low-noise) detector performance. Indeed, present-day IR detectors are often background-limited, such that all but a negligible amount of noise is the result of photon-induced generation-recombination. Consequently, if a system is to remain background-limited, the noise of the processing electronics must be negligible with respect to detector noise.

Quite often the electronic signal at the output of the detector is in the order of microvolts. In order to preserve the potential sensitivity of the detector, this low-level signal must be amplified by whatever means will result in minimum noise and distortions. Fundamental principles involved in mechanization of this amplification function are discussed in par. 3-6.3.1. Constraints such as detector and circuit-limiting time constants relegate most IR preamplification requirements to the 1- to 20-kHz band. Consequently, amplification devices which operate at higher frequencies are not discussed here.

IR systems designed to "detect" the presence of an object within a given field of view are primarily concerned with the strength of the signal from the object to not only internally generated noise but also to external noise due to fluctuations of the scene radiance. Usually the characteristics and particularly the form of the expected "signal" are known. Thus, the signal frequency spectrum is predeterminable. Signal-to-noise ratio and detection probability are directly related in these cases and can be vastly enhanced by relatively simple electronic filtering devices. The various signal processing electronics applicable to these cases are discussed in the paragraphs which follow.

Reproduction of an IR detected scene on a cathode ray display tube (CRT) requires that the output of each detector be sampled periodically and presented to the CRT intensity control in a time sequence correlated with the scanning beam. A typical example is the IR surveillance system which frequently incorporates arrays, comprised of literally hundreds of individual detectors. Reproduction of the scene in these cases requires the application of time-division multiplexing techniques discussed in par. 3-6.3.3.

3-6.3.1 Low-noise Amplification

This paragraph is addressed to the problems confronted in preserving the signal-to-noise ratio which exists at the output terminals of IR detectors, while providing sufficient amplification to minimize the noise contributions of subsequent signal-processing stages.

Amplification of low-noise signals is achieved almost exclusively by means of solid-state devices since low noise is associated with the lower operating current of solid-state devices. In addition, stringent present-day packaging re-

quirements dictate the performance of numerous electronic functions with hardware that requires minimum space, low power consumption, and provides a high degree of reliability. Although the theoretical advantages of such techniques as transformer coupling are mentioned, solutions more amenable to microminiaturization are emphasized.

3-6.3.1.1 Thermal Noise

The makeup of thermal noise in electronic components including resistive elements and semiconductor devices is similar to that of detectors as described in par. 3-4.3.6. In Fig. 3-94, a noisy resistor is shown as a noiseless resistor plus a voltage (A) or current (B) generator.

In junction transistors, thermal noise is primarily associated with the transverse current in the base region, i.e., the ohmic base resistance r_b. Noise associated with ohmic resistance in the emitter and collector regions is almost always negligible. In field-effect transistors (FET's), thermal noise is generated by the channel conductance.

3-6.3.1.2 Shot Noise

The p-n junction of a transistor is a source of "shot" noise[76]. Both the emitter and collector p-n junctions are sources of this type of noise. Transistor shot noise is similar to that produced in a vacuum tube diode. In both cases, the mean-squared noise current i^2_{sh} is proportional to the direct current I_{dc} through the diode, the

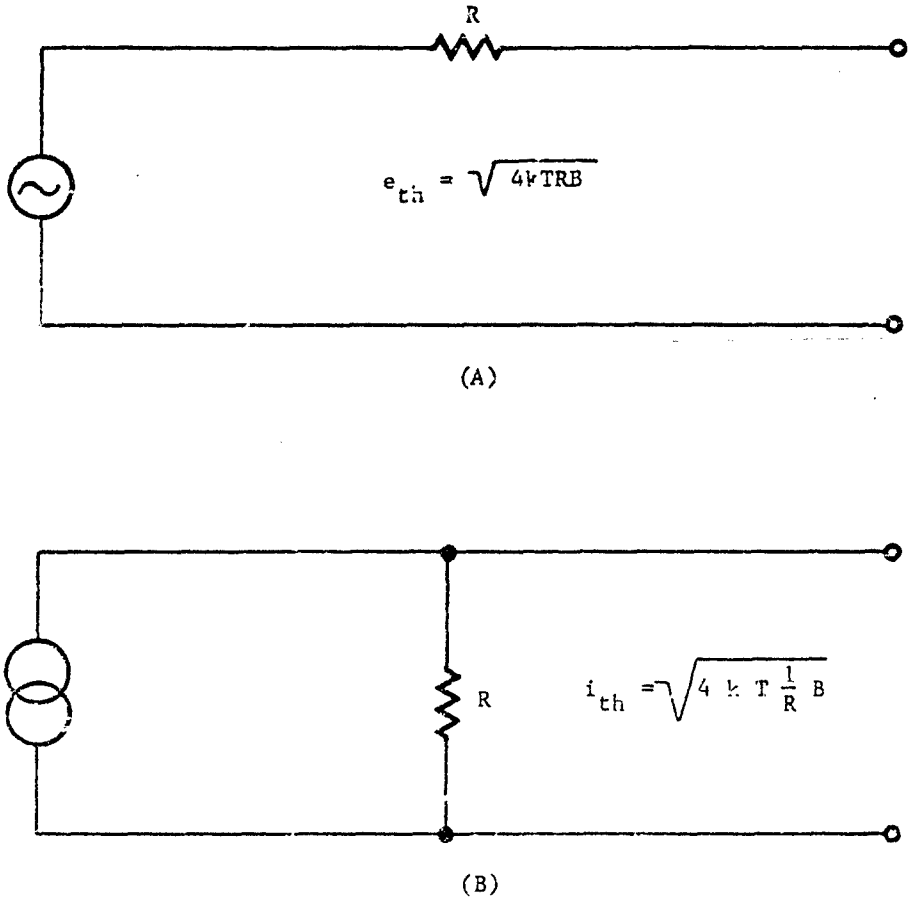


FIGURE 3-94. Characterization of Thermal Noise from (A) Voltage Source, and (B) Current Source

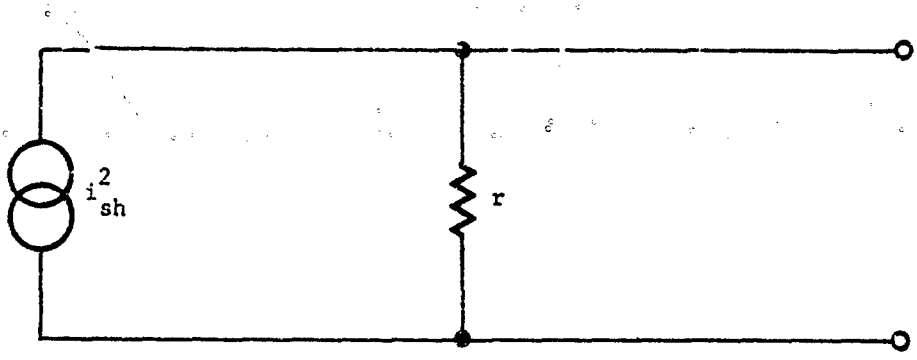
electron charge q, and the noise equivalent bandwidth Δf. Stated mathematically

$$i_{sh}^2 \approx 2qI_{dc}\Delta f \qquad (3\text{-}138)$$

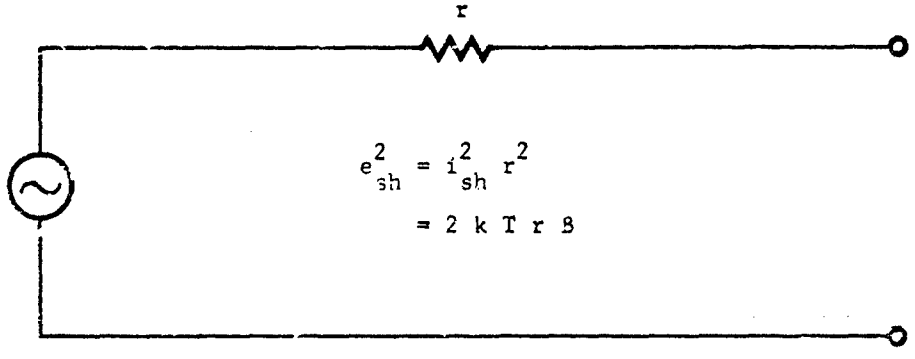
where

q = charge on electron
= 1.59×10^{-19} C

The noise junction is represented in Fig. 3-95 as a resistance which is equal to the dynamic resistance of the junction in parallel with a noise-current generator. The value of the shot-noise current, although approximate, is sufficiently accurate for most design purposes. A more rigorous representation is provided by van der Ziel[71,72].



(A)



(B)

$$r = \frac{n\ k\ T}{q\ I_{dc}} \simeq \frac{25}{I_{dc}}, \text{ ohm where } I_{dc} \text{ is measured in mA}$$

n = a constant $\simeq$ 1 for a transistor

FIGURE 3-95. Characterization of Shot-noise from (A) Current Source, and (B) Voltage Source

3-6.3.1.3 1/f Noise

At low frequencies (usually below 1 kHz) increases in noise power follow the $1/f$ law, i.e., the noise-power spectrum decreases with increasing frequency. This noise is sometimes referred to as excess noise, surface-effect noise, or simply as $1/f$ (one over f) noise. It occurs in most solid-state devices and is generally attributed to inhomogeneities or traps either in the semiconductor material or on the surface of the transistor. Since this type of noise varies with individual devices, it is difficult to specify an internal noise generator for use in an equivalent circuit. In practice, this parameter normally is established by direct measurement.

3-6.3.1.4 Performance Characterization

The best insight into the selection and comparison of amplification devices and circuits for low-noise applications is provided by their "noise figure" characteristics. The noise figure (or noise factor) NF is a direct measure of the degree to which the amplification process preserves the signal-to-noise ratio at the output terminals of the detector before connection of the amplification circuitry, thus

$$NF = \frac{S_o/N_o}{S_i/N_i} \tag{3-139}$$

where S_o/N_o is the signal-to-noise-power ratio at the output and S_i/N_i is the corresponding quantity at the input. An alternate form for NF is given by

$$NF = \frac{N_o}{GN_i} \tag{3-140}$$

where G is the gain S_o/S_i.

There are two approaches for determining the NF in general use. The classic approach entails establishing an equivalent circuit composed of a sufficient number of known noise generators such as those discussed previously under noise sources. This approach requires that many of the basic semiconductor parameters be available from data sheets (or from measurements).

A more recent approach which is gaining wide acceptance involves only two parameters for determining the noise figure, i.e., equivalent input noise voltage and current discussed in par. 3-6.3.1.4.3.

3-6.3.1.4.1 Classic Approach for Conventional Transistors

Internal noise in a transistor consists of mainly:

1. Shot noise in the emitter-base junction
2. Shot noise in the collector base junction
3. Thermal noise in the base resistance

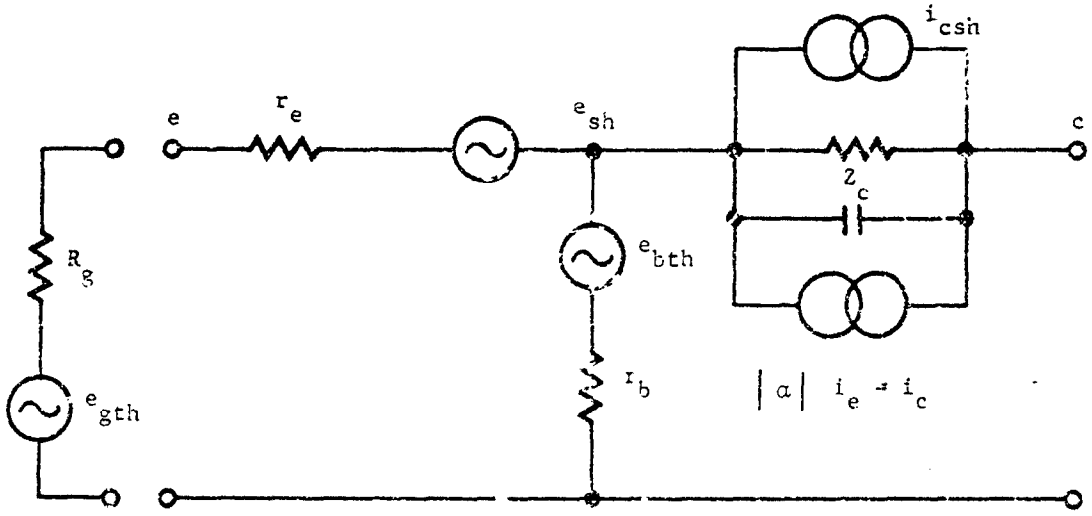


FIGURE 3-96. Classic Conventional Transistor Noise-equivalent Circuit

Fig. 3-96 illustrates the equivalent common-base circuit which includes these noise sources where

e_{sh} = emitter shot-noise voltage
$= \sqrt{2kTr_e\Delta f}$

i_{csh} = collector shot-noise current
$= \sqrt{2qI_E\Delta f}$

e_{gth} = generator thermal noise voltage
$= \sqrt{4kTR_g\Delta f}$

e_{bth} = base resistance (r'_b) thermal voltage
$= \sqrt{4kTr_b\Delta f}$

i_c = collector current (at test frequency)

α = common-base current gain (at test frequency)

r_e = emitter diode resistance $\approx 25/I_E$ (low frequency) where I_E is measured in mA

R_g = source resistance

Δf = noise equivalent bandwidth, Hz

By use of the second form of the definition of noise figure Eq. 3-140, the noise figure can be derived for the equivalent circuit[73,74].

$$NF = 1 + \frac{r'_b}{R_g} + \frac{r_e}{2R_g} + \frac{(R_g + r_e + r'_b)^2}{2\alpha_o R_g r_e}\left[\frac{1}{h_{FE}} + \frac{I_{CO}}{I_E} + \left(\frac{f}{f_\alpha}\right)^2\right] \qquad (3\text{-}141)$$

where

I_E = dc emitter current

I_{CO} = dc collector cutoff current

h_{FE} = common emitter dc current gain

r'_b = base resistance

R_g = source resistance

α_o = low frequency alpha (not the dc)

f = frequency of measurement

f_α = frequency at which alpha has decreased to $\alpha_o/\sqrt{2}$

By differentiating Eq. 3-141 with respect to the variable R_g and setting the result equal to zero, an optimum value for R_g is obtained as given below.

$$R_g(opt) = \left[(r_e + r'_b)^2 + \frac{\alpha_o r_e(2r'_b + r_e)}{\frac{1}{h_{FE}} + \frac{f^2}{f_\alpha} + \frac{I_{CO}}{I_E}}\right]^{1/2} \qquad (3\text{-}142)$$

Examination of this equation reveals $R_g(opt)$ to be frequency dependent. For most IR systems, f/f_α is small and since r_e is normally small with respect to r'_b, $R_g(opt)$ can be simplified to

$$R_g(opt) = \left[(r'_b)^2 + \frac{\alpha_o 2 r'_b r_e}{1 \quad \alpha_o} \right]^{1/2} \qquad (3\text{-}143)$$

As indicated in Fig. 3-102, the value for $R_g(opt)$ in typical devices is not critical (i.e., a reasonably broad range of values does not significantly affect the noise figure).

It is also of interest to note that the expression for the noise figure is identical for the common emitter configuration. Also since $R_g(opt)$ is close to the common emitter input impedance, higher gain is possible in the common emitter configuration.

As indicated earlier, the $1/f$ noise characteristics are not completely understood and the low-frequency corner (Fig. 3-97) must be determined by test.

3.6.3.1.4.2 Classic Approach for Field-effect Transistors

Field-effect Transistors (FET's) exhibit excellent low-noise characteristics. According to van der Ziel[75], the equivalent noise resistance of FET's is "about a factor of four better than the shot-noise resistance of a vacuum tube having comparable transconductance characterisitics".

The principal noise sources in FET's are:

a. Shot noise due to gate leakage current expressed as

$$i^2_{sh} = 2qI_{gt}\Delta f \qquad (3\text{-}144)$$

where q is the electronic charge, I_{gt} is the gate leakage current, and Δf is the noise equivalent bandwidth.

b. Thermal noise in the conducting channel given by

$$e^2_{th} = 4kT\Delta f/g_m \qquad (3\text{-}145)$$

where g_m is the transconductance and k is the Boltzmann constant.

c. Generation-recombination (G-R) noise in the space charge layer. The G-R noise gives rise to the $1/f$ noise power spectral characteristic previously noted for conventional transistors and is expressed by

$$e^2_f = \left(e^2_{th}\, 1 + \frac{f_L}{f} \right) \qquad (3\text{-}146)$$

In the normal reverse bias mode of operation, the gate leakage current for junction FET's range from 10^{-11} to 10^{-8} A. For insulated gate FET's (e.g., MOSFET's), leakages are normally less than 10^{-15} A and, therefore, shot noise is not significant.

A noise model of the FET is given in Fig. 3-98.

The noise figure of the FET[76] can be reduced to

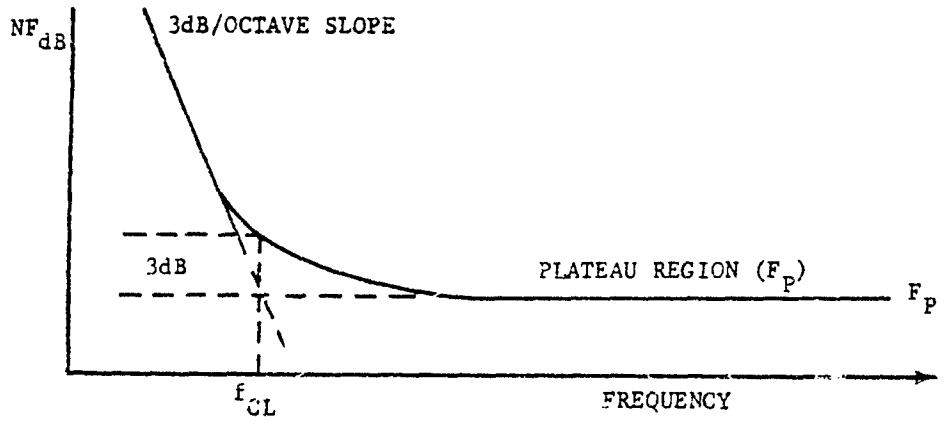


FIGURE 3-97. Low-frequency (1/f) Noise Characteristics

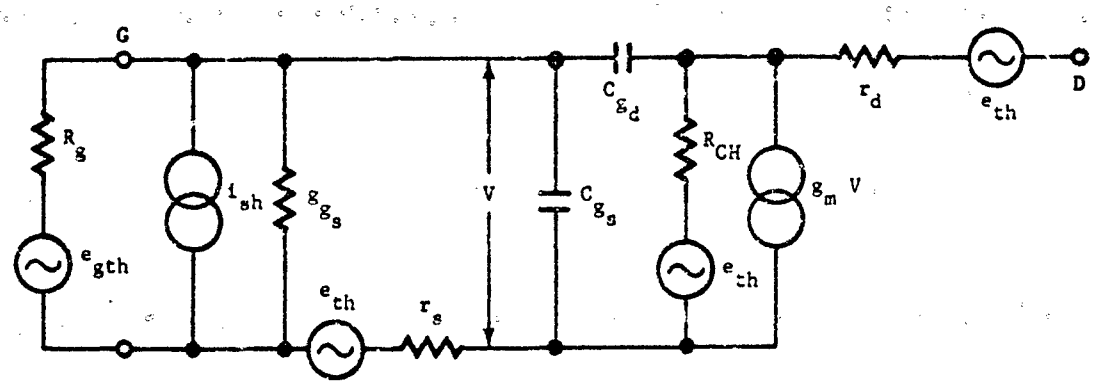


FIGURE 3-93. FET Noise Equivalent Circuit

$$NF = 1 + \frac{1}{g_m}\omega(C_{g_s} + C_{g_d}) + \frac{R_g}{2g_m}\left[\frac{1}{R_g} + \omega(C_{g_s} + C_{g_d})\right]^2 \qquad (3\text{-}147)$$

where C_{g_s} and C_{g_d} are interelectrode capacitors between gate and source, and gate and drain, respectively; and ω is the electrical frequency in rad/sec.

The optimum value of source resistance is

$$R_g(opt) = \frac{1}{\omega(C_{g_s} + C_{g_d})} \qquad (3\text{-}148)$$

For low frequencies (excluding the 1/*f* region), the noise figure reduces to

$$NF_{lf} = 1 + \frac{1}{2g_m R_g} \qquad (3\text{-}149)$$

3-6.3.1.4.3 $\bar{e}_n$, $\bar{i}_n$ Approach for Conventional Transistors

Noise performance can be disassociated from the circuit parameters (i.e., source resistance) by dealing in terms of noise parameters. In general, four such parameters are required. These are generally taken to be the amplitudes of two noise generators plus a complex correlation coefficient. Although numerous combinations of parameters can be used, the most general combination is illustrated in Fig. 3-99.

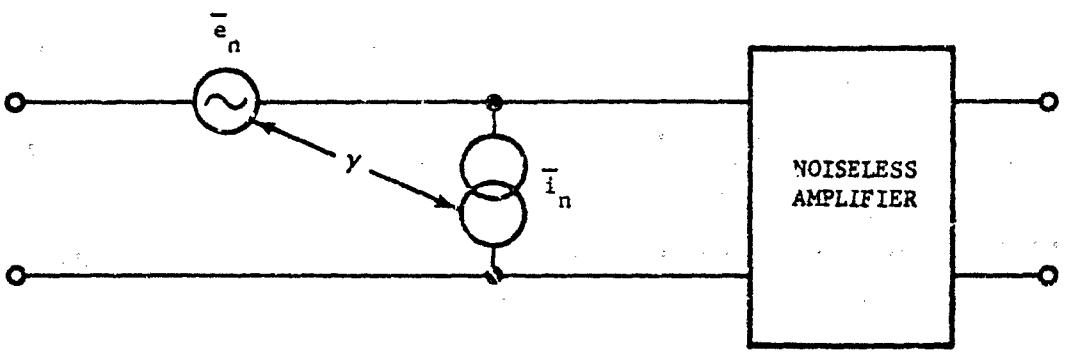


FIGURE 3-99. $\bar{e}_n$, $\bar{i}_n$ Noise Equivalent Circuit

The "noisy" amplifier is represented by a voltage source $\bar{e}_n$ in series with the input of a noiseless amplifier plus a current source $\bar{i}_n$ in parallel with the input. The parameters $\bar{e}_n$ and $\bar{i}_n$ are measured under input open circuit and input short circuited conditions, respectively. The total noise outputs under these conditions are assumed to be due to $\bar{e}_n$ and $\bar{i}_n$.

Preamplifiers for IR applications are usually relatively low-frequency devices. Experimental evidence seems to indicate that, in the low-frequency region only, the correlation coefficient between the input voltage and current noise generators is small for most types of junction transistors. This is particularly fortuitous in the $\bar{e}_n$, $\bar{i}_n$ approach.

Fig. 3-100 shows a small-signal amplifier which consists of a generator having a signal voltage e_g and an internal resistance R_g. A corresponding noise voltage e_{gn} is connected to the input of a noiseless amplifier together with the two input noise generators $\bar{e}_n$ and $\bar{i}_n$, respectively, and with a low admittance Y_l.

The noise figure is given by[76]

$$NF = 1 + \frac{1}{4kT\Delta f}\left[\frac{\bar{e}_n^2}{R_g} + \bar{i}_n^2\frac{|Z_g|^2}{R_g} + 2\left(\bar{e}_n^2\bar{i}_n^2\right)^{1/2}\left(\gamma_r - \gamma_i\frac{X_g}{R_g}\right)\right] \qquad (3\text{-}150)$$

where

R_g = source resistance

Z_g = source impedance

X_g = generator reactance

γ_r, γ_i = correlation coefficient of $\overline{e_n^2}$ and $\overline{i_n^2}$, respectively

For a resistive input impedance, Eq. 3-149 reduces to

$$NF = 1 + \frac{1}{4kT\Delta f}\left[\frac{\overline{e_n^2}}{R_g} + \overline{i_n^2}R_g + 2\overline{e_n}\,\overline{i_n}\gamma_r\right] \tag{3-151}$$

The low-frequency correlation coefficient γ also reduces to a value near zero for a resistive input impedance. Therefore, by differentiating this expression with respect to R_g and setting the result equal to zero, the minimum attainable noise figure becomes

$$NF_{(min)} = 1 + \frac{\overline{e_n}\,\overline{i_n}}{2kT} \tag{3-152}$$

$$R_g\,(opt) = \frac{\overline{e_n}}{\overline{i_n}} \tag{3-153}$$

The e_n, i_n approach for the FET is identical to that presented for conventional transistors. In the FET case, the measurement of $\overline{e_n}$ is straightforward, but that of $\overline{i_n}$ is difficult. Fortunately, it is sufficient in most cases to measure the gate leakage current and calculate $\overline{i_n}$ by

$$\overline{i_n^2} = 2qI_{gl} \tag{3-154}$$

where I_{gl} is the gate leakage current.

Also since the correlation coefficient for FET is small, the noise figure is given by[77]

$$NF = 1 + \frac{1}{4kT\Delta f}\left(\frac{\overline{e_n^2}}{R_g} + \overline{i_n^2}R_g\right) \tag{3-155}$$

3-6.3.1.5 IR Application Problems

3-6.3.1.5.1 Field-effect vs Conventional Transistor Characteristics

The curves of Figs. 3-101 and 3-102 provide a gross indication of the regions of applicability of the two devices relative to frequency spectrum and source impedance.

Generally, FET's can provide a lower noise figure in applications where the lower frequencies (1 to 20 Hz) are important. Also, in general, optimum source resistance for bipolar transistors is in the kilohm region whereas optimum for FET's tends toward the megohm region. Internal impedances of IR detectors for various applications cover the very wide range from a few ohms to hundreds of megohms.

Although optimum source resistance is not greatly significant where transformer coupling is possible, a signal processing system with virtually hundreds of parallel identical amplification channels, as in many present-day military

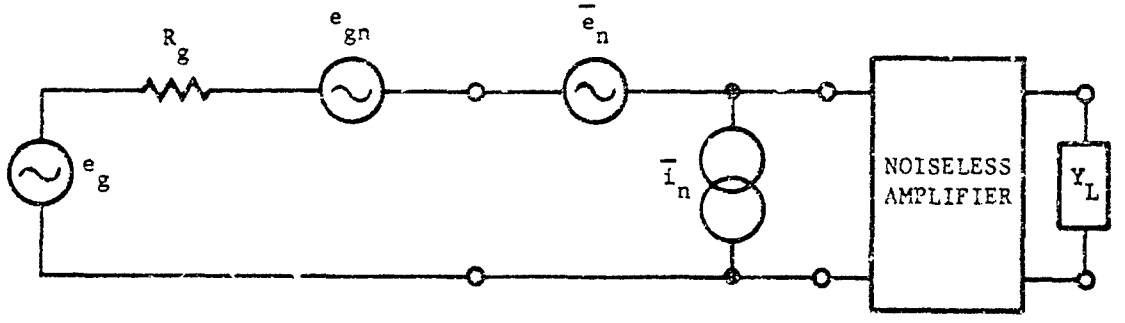


FIGURE 3-100. $\overline{e_n}$, $\overline{i_n}$ Approach Noise Figure Computational Model

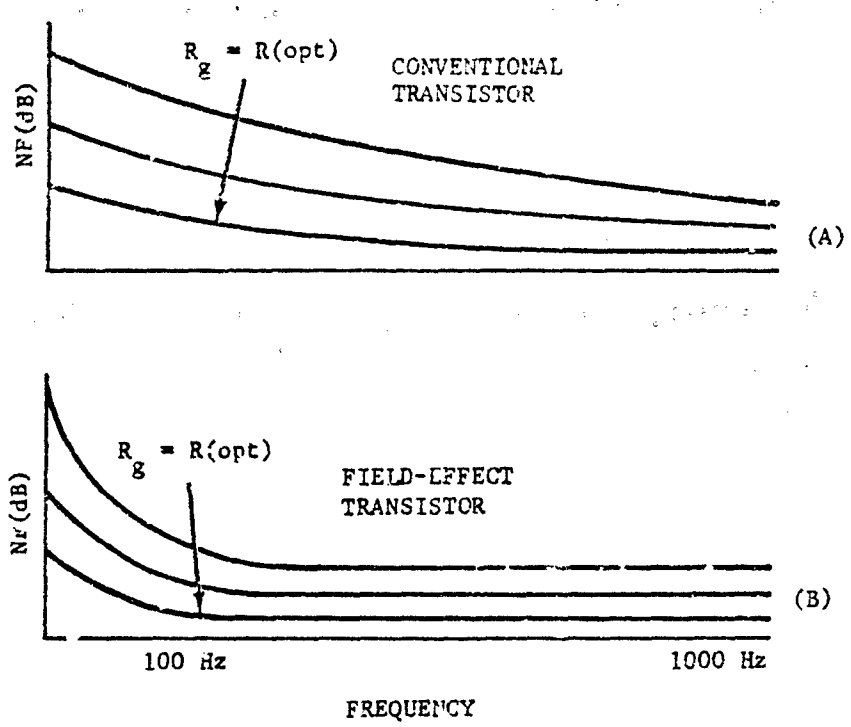


FIGURE 3-101. Noise Figure As a Function of Frequency from (A) Transistor, and (B) FET

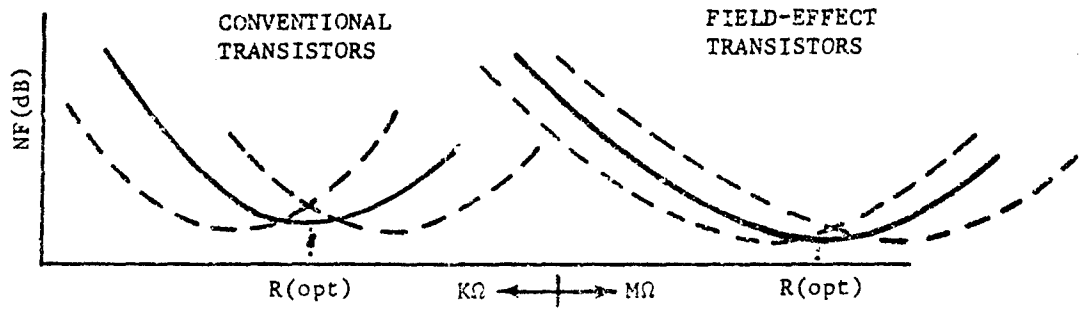


FIGURE 3-102. Gross Comparison of Transistor and FET Noise Figures vs Source Resistance

IR systems, limits the use of transformers. The transformer does not lend itself to the microelectronic configurations thus demanded. It should also be noted that, in general, transformers for low-noise and low-frequency applications present very special design problems.

3-6.3.1.5.2 True Noise-figure

Amplifier-noise-figure data are based on the assumption that the source resistance generates true Johnson (thermal) noise at a level caused by a temperature of 290°K (see definition of noise figure in par. 3-6.3.1.4). If the only objective is the comparison of amplifier performance, this is a reasonable assumption. However, if it is necessary to determine the degree of perfection of the amplification process relative to the detector capabilities, the actual type of noise generated by the detector (source resistance) and its actual temperature must then be considered. At room temperature, detectors will generate noise greatly in excess of Johnson noise. To avoid overdesign, i.e., specifying an unnecessarily expensive low-noise design, the actual conditions should be examined.

An amplifier having a published noise figure of 2 (3 dB) might have an effective noise figure of 1.1 (0.04 dB). On the other hand, many present-day IR detectors are cooled to very low temperatures whereby normal noise figure data would not indicate the extent of degradation caused by the amplification process. An analysis of the actual situation might reveal the wisdom of specifying a more sophisticated low-noise design.

Reference to the performance characterizations (par. 3-6.3.1.4) will show that the noise figure NF is of the form

$$NF = 1 + \frac{A}{T_o} \qquad (3\text{-}156)$$

where T_o is normally specified as room temperature, or 300°K.

If the "actual" noise figure is defined as

$$NF_a = 1 + \frac{A}{T_g} \qquad (3\text{-}157)$$

where T_g is the operating temperature of the detector, then combining Eqs. 3-156 and 3-157 yields

$$NF_a = 1 + (NF - 1)\frac{T_o}{T_g} \qquad (3\text{-}158)$$

which provides a simple method of converting from room temperature to actual operating conditions.

3-6.3.1.6 Cryogenic Amplifiers

As described in par. 3-4, long-wavelength extrinsic photoconductors are operated at cryogenic temperatures. At the desired operating conditions, the detector impedance is often extremely high. Coupling of the detector to an amplifier generally introduces finite capacitance (input capacitance as well as distributed load capacitance) which results in a low cut-off frequency. This, in turn, reduces the signal output of the detector. In systems that are preamplifier-noise limited, this capacitive loading may be detrimental to the signal-to-noise ratio. In order to reduce the capacitance associated with detector wiring, it is often advantageous to utilize preamplifier circuits operating at the detector temperature and located adjacent to the detector on the same cold sink. This paragraph describes some of the considerations necessary to achieve functional cryogenic electronic circuits for use with infrared detectors.

The development of preamplifiers which operate at temperatures between 4.2° and 300°K has resulted from recent improvements in FET technology. Silicon junction FET's operate at temperatures as low as 77°K, with thermal de-ionization reducing the channel carrier concentration at lower temperatures. Metal-oxide-semiconductor FET's (MOSFET's) do not depend on thermal ionization for carriers in the channel; hence, these devices provide relatively consistent performance over the temperature range 4.2° to 300°K. In addition, the high-input resistance and low-input capacitance make these devices suitable for use with high-impedance infrared detectors. Preamplifiers using MOSFET's have been successfully used at 4.2°K [78].

Junction FET's exhibit significant variations with temperature between 300° and 77°K. The behavior of the transconductance g_m and drain saturation current I_{DSS} for a typical device is

illustrated in Fig. 3-103. Both parameters are controlled by thermal scattering up to 300°K and, therefore, are approximately proportional to $T^{-3/2}$. Below 200°K, thermal deionization of impurities causes a sharp decrease in performances.

The noise of junction FET's exhibits a peculiar behavior at low temperatures. As shown in Fig. 3-104, successive maxima and minima occur as the temperature is varied. This effect is due to carrier generation at Shockley-Read-Hall centers in the gate-channel depletion region[79]. Individual FET's behave differently; however, n-channel devices consistently exhibit a noise peak near 100°K. Lower-noise performance near this temperature is usually provided by p-channel devices.

The performance of suitable MOSFET's is relatively constant from 4.2° to 300°K. The drain current and transconductance usually increase slightly from 300° to 77°K, and remain constant or decrease at lower temperatures. The noise in these devices is due to fluctuations at surface interface states and does not change significantly over this temperature range. The short-circuit noise voltage spectrum of a typical MOSFET can be compared in Fig. 3-105 with the noise spectra of an n-channel silicon junction FET. Notice that the MOSFET noise is approximately one order of magnitude greater than the junction FET noise at low temperatures.

Some MOSFET's exhibit current-dependent noise oscillations which can be detrimental to low-noise performance in the 77° to 4.2°K region. This mode of operation results from a space-charge-limited conduction characteristic near 4.2°K.

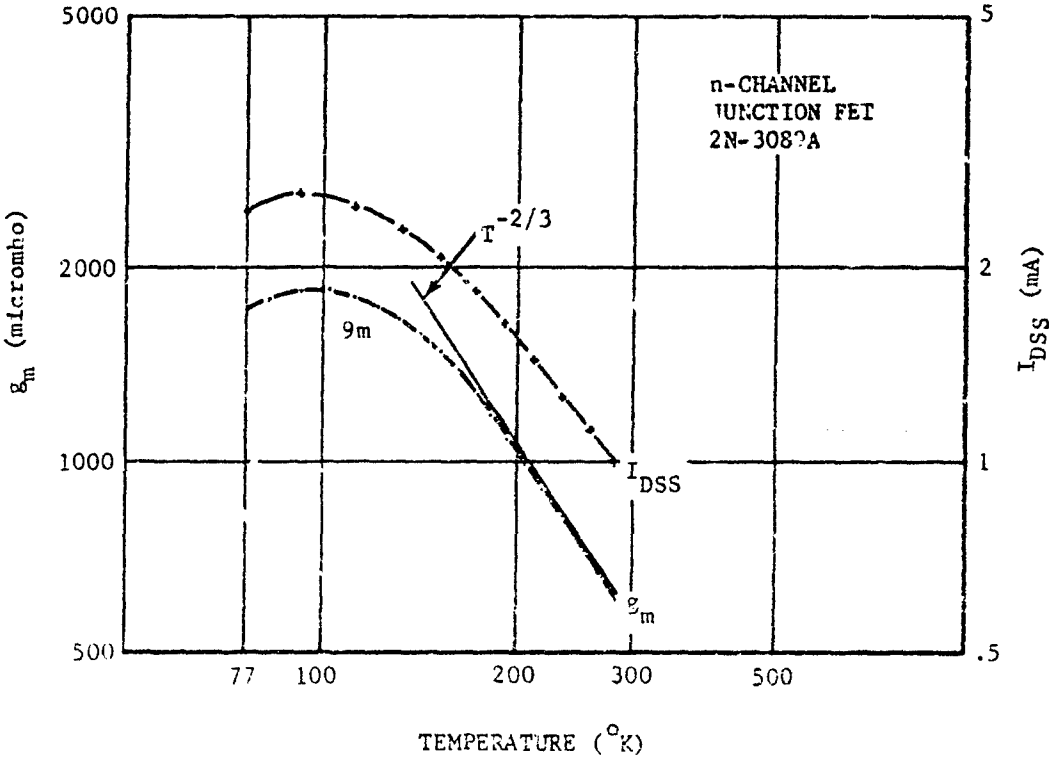


FIGURE 3-103. Transconductance and Drain Current as a Function of Temperature

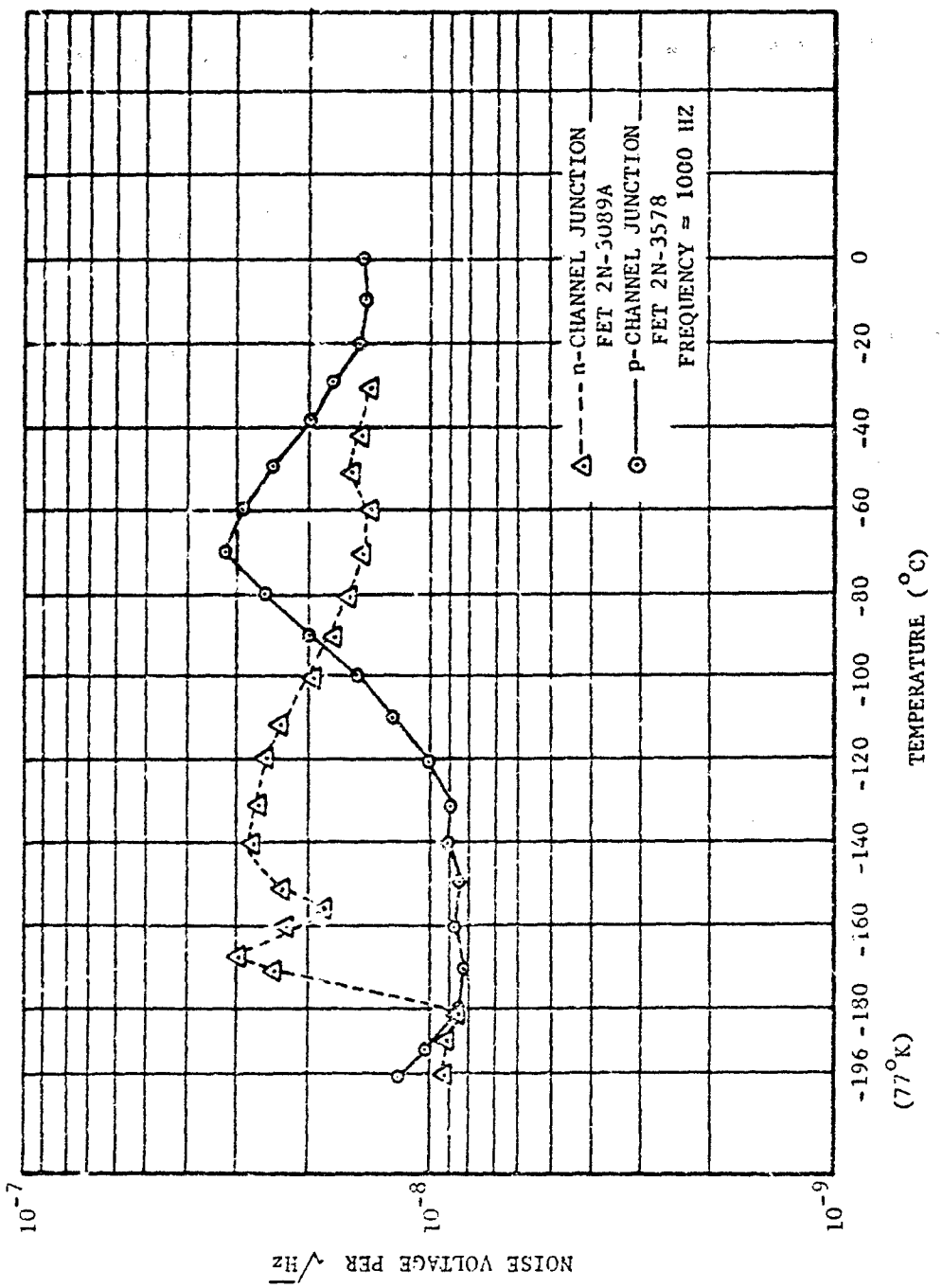


FIGURE 3-104. Junction FET Noise As a Function of Temperature

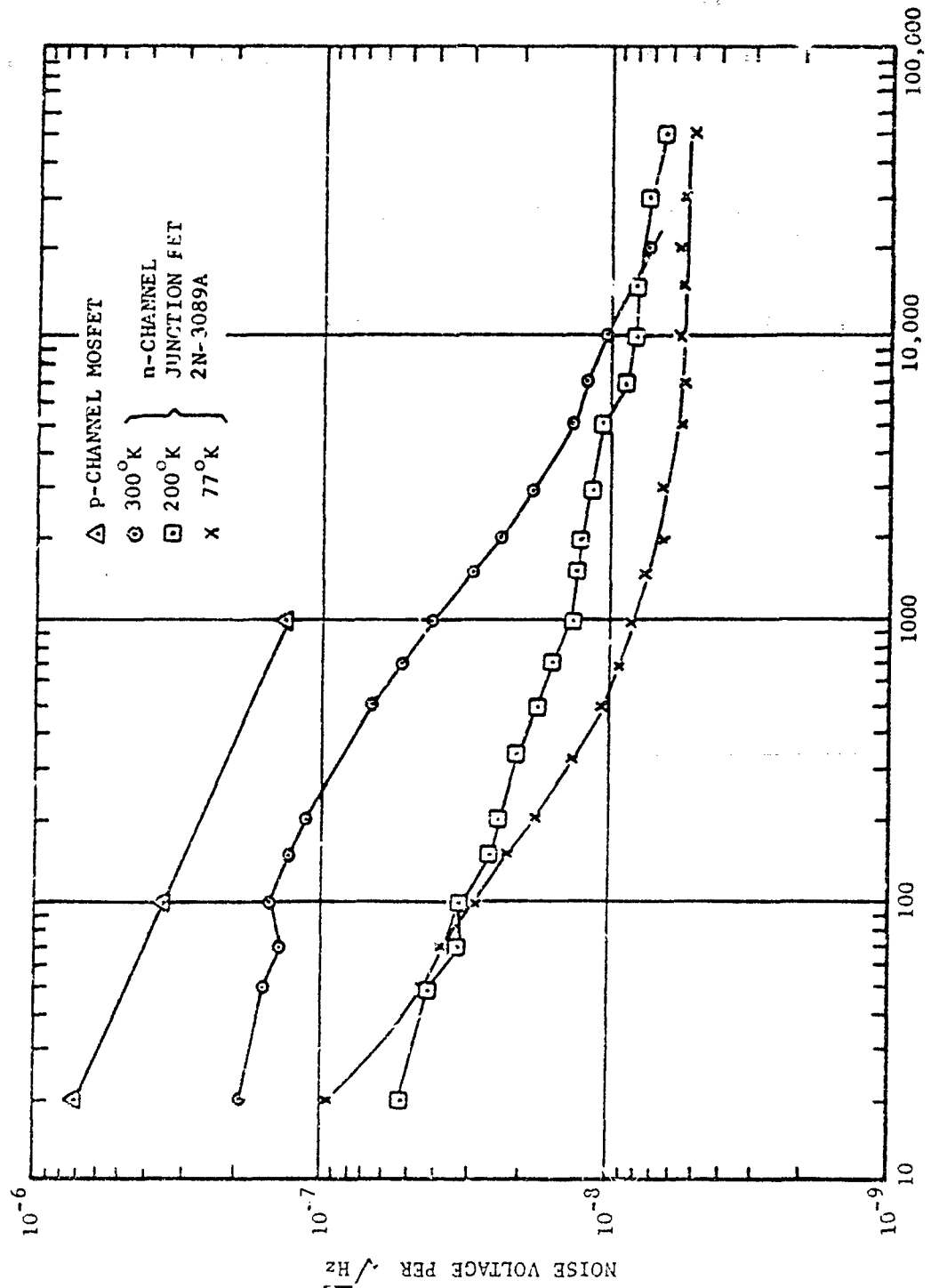


FIGURE 3-105. FET Noise As a Function of Frequency

The noise of an FET is usually represented by the equivalent noise voltage and current ($\overline{e_n}$, $\overline{i_n}$) generator method, discussed previously. In junction FET's, $\overline{i_n}$ decreases to negligible values at low temperatures, leaving $\overline{e_n}$ as the dominant noise source for impedances as high as 10^8 ohms. In MOSFET's the noise source $\overline{e_n}$ is dominant at all temperatures and source impedances. Thermal noise due to resistive components in the preamplifier input circuit may be significant even at low temperatures and may be controlled by careful selection of component values.

Since these active devices are not capable of temperature-independent performance in this cryogenic region, a source-follower circuit configuration is normally used to eliminate variations in voltage gain due to changes in transconductance. Such a circuit is best suited for use with high-impedance photodetectors.

In designing preamplifiers for cryogenic applications, excessive power dissipation within the active devices and passive components must be carefully avoided. This is a particularly acute problem in applications where large numbers of active elements are used. Active cryogenic devices commonly dissipate approximately 5 mw, due to the high-current density required in the FET channel.

3-6.3.2 Electronic Filtering

3-6.3.2.1 Application and Response Analysis

As indicated previously, electronic filtering plays an important role in determining the effectiveness of the signal detection process. Filtering may be defined, in general, as the process of removing or attenuating certain frequency components from a given composite input signal. The performance criteria for a given task will, in general, determine the specific type of filtering required (e.g., maximization of peak signal-to-rms noise, faithful signal reproduction, pulse time determination, etc.). The types of filters required in IR systems range from the simplest RC low-pass filters to the most sophisticated state-of-the-art designs.

The design of conventional filters is treated thoroughly in existing literature[80,81] and, therefore, is not discussed here. The discussion is also restricted to the area of linear time-invariant filters. The most important "language" used to describe such filters is the Laplace-Transform Theory; several texts on this theory[82,83,84] are available. The paragraphs which follow give some of its applications:

Two important factors in the design of filters entail determining the output response of the filter to a signal and the effect of the filter on noise. A distinction between response to signals and noise is made here since noise by definition is nondeterminable and must be treated somewhat differently than signals. Fig. 3-106 shows the filter response problem posed here. The object is to determine the output signal $e_o(t)$, and noise $n_o(t)$ if the input signal $e_i(t)$, noise $n_i(t)$, and the filter transfer junction are given.

From the transform theory, the Laplace Transform of the output is related to the input signal transform by

$$E_o(s) = H(s)E_i(s) \qquad (3\text{-}159)$$

where

$E_o(s)$ = Laplace Transform of $e_o(t)$

$H(s)$ = transfer function or impulse response of the filter

$E_i(s)$ = Laplace Transform of $e_i(t)$

s = Laplace notation for complex frequency

The time response of the output is given by the relation referred to as the convolution integral

$$e_o(t) = \int_{-\infty}^{\infty} e_i(\tau)h(t-\tau)d\tau \qquad (3\text{-}160)$$

which may be solved graphically by techniques given in the literature[82,83,84]. The output may also be determined by means of the Laplace Transform tables.

The noise at the output is determined differently. The input noise n_i is given in terms of its power-density spectrum $|N_i(f)|^2$, the output noise power density spectrum $|N_o(f)|^2$ is given by

$$|N_o(f)|^2 = |N_i(f)|^2 \times |H(f)|^2 \qquad (3\text{-}161)$$

where $H(f)$ is the transfer function of the network.

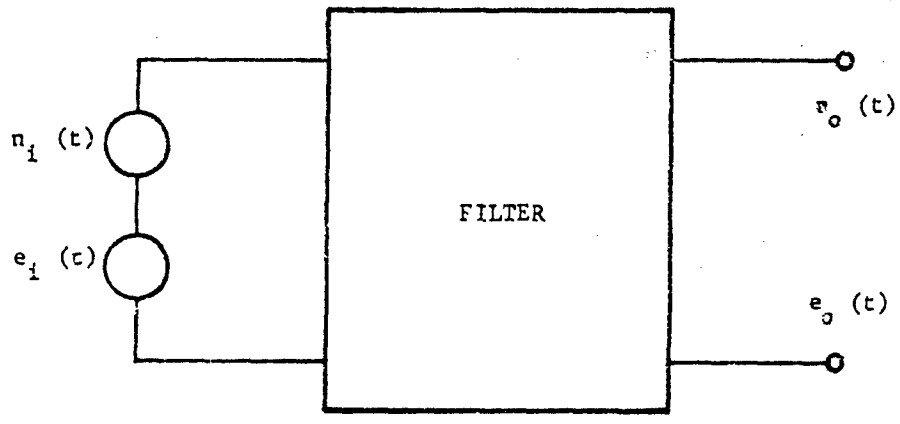


FIGURE 3-106. Model of Filter Relating Output Response $e_o(t)$ to Input $e_i(t)$ and Output Noise n_o to Input Noise n_i

3-6.3.2.2 Detection Filters

One of the most important and most frequently encountered requirements in IR systems is that of filtering for detection. The filter of concern here would be required to maximize the peak signal-to-rms-noise ratio with regard to waveform distortion. The solution to this problem entails using a matched filter. The derivation of the matched filter is lucidly treated in the referenced literature[69,85]. The matched filter solution to the maximum signal-to-rms-noise problem is represented mathematically as

$$H(f) = K \frac{X^*(f)}{|N_x(f)|^2} \exp[-j2\pi ft] \qquad (3\text{-}162)$$

where

K = constant

$H(f)$ = the frequency transfer function of the matched filter

$X^*(f)$ = the conjugate of the signal transform of the input signal

$|N_x(f)|^2$ = the input noise power density spectrum

In many cases, such as thermal noise, the spectrum of $N_x(f)$ is a constant (white noise).

The matched filtering concept is important in terms of measuring performance rather than as a basis for an actual filter design. Practical limitations often dictate the use of a relatively simple filter. The problem then becomes one of determining how well a simpler filter approximates the performance of a matched filter.

The examples which follow assume the presence of white noise.

In the first case[85], the input signal is shaped as a rectangular pulse of width τ and the filter is a simple RC stage as shown in Fig. 3-107. The results under these conditions are illustrated in Fig. 3-108. As shown, the single-stage RC filter is only 1 dB worse than the matched filter. The responses of a Gaussian or multi-stage (two or more stages) RC filter and an ideal low-pass filter are also shown.

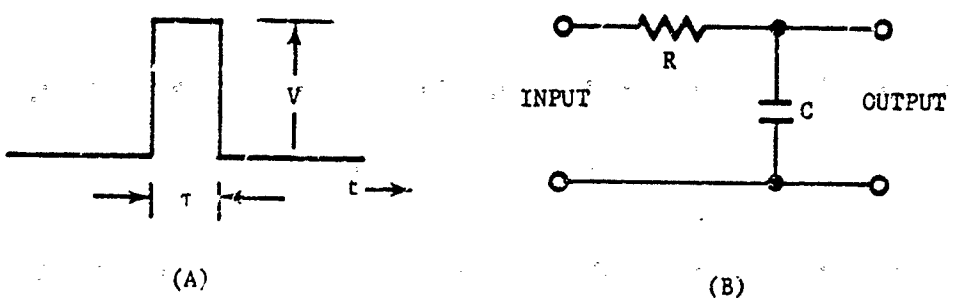
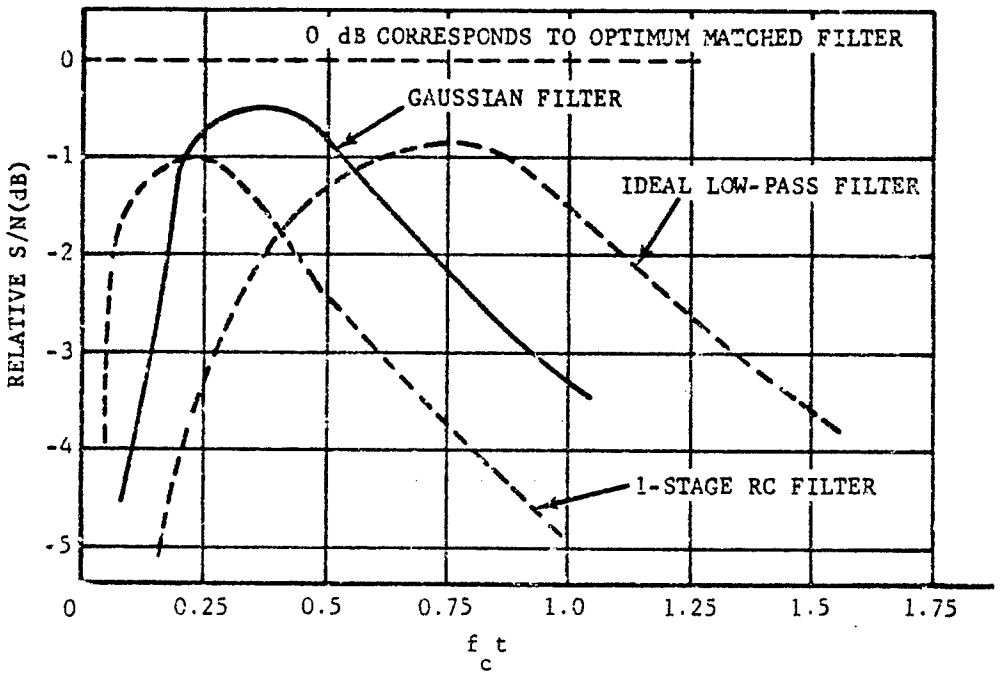


FIGURE 3-107. Rectangular Pulse Input and Simple RC Filter



where f_c = cutoff frequency
t = input pulse width

FIGURE 3-108. Peak S/N for Various Filters, Compared With Matched Filter (Rectangular Pulse Input)

For the second case[69], the input is a sinusoidal pulse train and the filter is a simple, second-order filter as illustrated in Fig. 3-109. The results shown in Fig. 3-110 also indicate reasonably good performance with this simple, tuned filter.

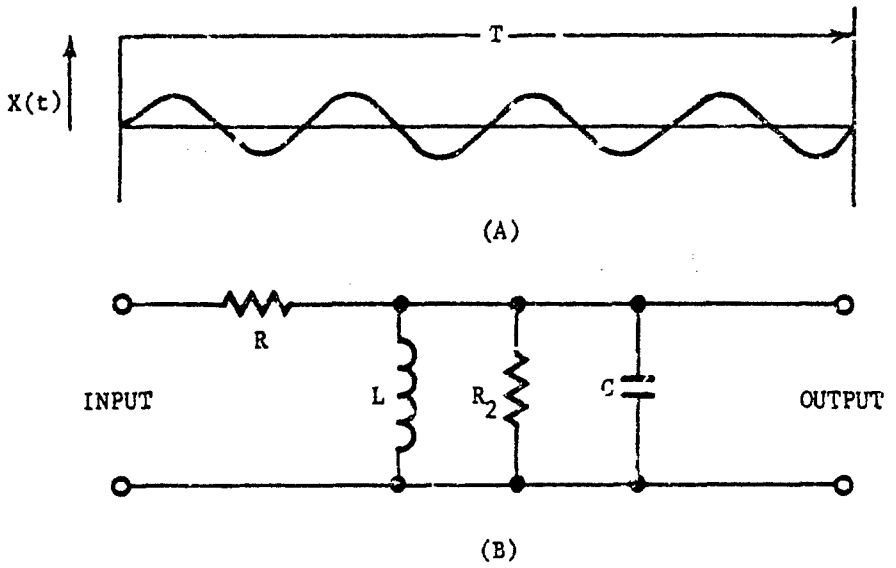


FIGURE 3-109. Input Sinusoidal Pulse Train and Second-order Filter

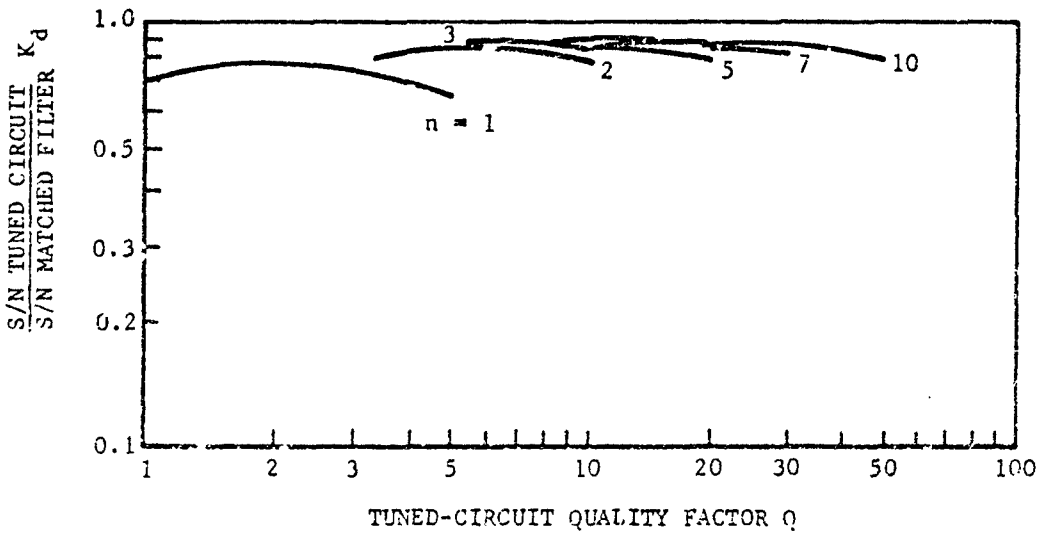


FIGURE 3-110. Signal-to-Noise Performance for a Single Tuned Circuit of Quality Q-filtering a Train of n-sinusoidal Pulses

3-6.3.3 Electronic Multiplexers

As pointed out by Black[86], although not formulated with mathematical rigor, some of the fundamental concepts of interest here are almost as old as the electrical transmission of telegraph messages. For example, multiplexing by time-division was conceived by an American, Moses B. Farmer, in 1852. The patent awarded in 1853 describes the transmission of a plurality of telegraph messages over a single line by the process of allocating independent intervals of time to the transmission of each message. To accomplish this, Farmer utilized two rapidly revolving synchronous commutations, one at each end of the line. This is the essence of time-division multiplexing.

A fundamental issue immediately apparent is the question of the required speed of sampling. The answer is found in the "Sampling Principle" usually stated [86] as follows:

"If a signal that is a magnitude-time function is sampled instantaneously at regular intervals and at a rate slightly higher than twice the highest significant signal frequency, then the samples contain all of the information of the original signal."

It follows that, in general, all the information contained in a multiplicity of separate amplitude-time function channels can be collected and transmitted through a single channel in real time by taking advantage of the unused portions of the sampling intervals.

Where a system makes use of a device such as the IR-sensitive image orthicon tube, multiplexing is achieved by physically sweeping the electron beam of the tube. Where greater sensitivity is required, individual detector elements are used in a linear array configuration. A scene can be scanned with the array or the scene can be made to sweep across the array. In order to reproduce the scene on a device such as a CRT, the amplitude data from all resolution elements of the scene must be collected and time-sequenced.

The example which follows provides an insight into the switching speeds required. Assume a matrix consisting of 400 vertical by 800 horizontal resolution elements. The linear array of detectors is vertically oriented, thereby, resulting in 400 detector elements with their associated individual amplifiers and multiplexing switches. The frame rate is set at 30 Hz to prevent visual "flicker" of the scene. Furthermore, assume that the scene consists of alternately bright and dark bands whose width is equal to the width of the resolution element. Fig. 3-111 illustrates the signal output of an individual detector as the scene is scanned. From this illustration, it is apparent that the linear array of 400 elements must be sampled at least once during the time period t and in so doing the sampling principle for this digital waveform will be satisfied. The time t' available for an individual sample is

$$t' = \frac{1}{(FR) \times M \times N}, \text{ sec} \qquad (3\text{-}163)$$

where

FR = frame rate, Hz

M = number of vertical elements

N = number of horizontal elements in matrix

$$t' \leq \frac{1}{30 \times 800 \times 400}, \text{ sec}$$

The clock rate f_c therefore, must be *at least*

$$f_c = \frac{1}{t'} = 9.6 \text{ MHz} \qquad (3\text{-}164)$$

Multiplexers capable of performing efficiently and reliably at such speeds have only recently become feasible. This can be attributed to the development of more efficient, high-speed switching transistors, and to progress in microelectronic fabrication techniques. These factors have been combined to achieve high-speed multiplexing with the smallest inter-component spacing.

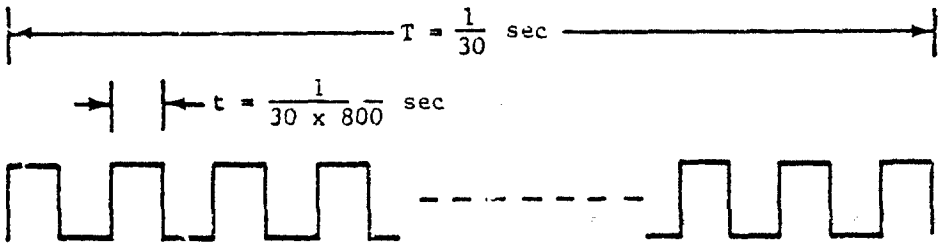


FIGURE 3-111. Detector Output for One Frame

3-7 DATA DISPLAY AND RECORDING

3-7.1 INTRODUCTION

A display may be defined in a general sense as a device for presenting visual information to a human observer. There are two basic formats for displaying information: (1) alpha-numeric and (2) pictorial. The first category consists of numerals, letters, symbols, and event indicators (binary code). The second display format recreates an external scene in as full detail as desirable. In both cases, the information must be presented in a manner which can be viewed comfortably, readily understood, and not easily misinterpreted by the observer.

Par. 3-7 provides a brief look at some of the display techniques and recording equipment most frequently used in conjunction with infrared systems and a short discussion of the design considerations regarding display equipment. The following devices are described:

1. Cathode Ray Tubes
2. Storage Tubes
3. Character Generation Tubes
4. Image Converter Tubes
5. Tape Recording Techniques
6. Photographs
7. Holograms
8. Projection Cathode Ray Tubes
9. Multi-gun Cathode Ray Tubes
10. Electroluminescent Panels
11. Photoemitter Diode Arrays

3-7.2 TERMINOLOGY

Symbols used in par. 3-7 on displays are summarized in Table 3-14. A detailed definition of those terms most frequently employed in a discussion of display techniques is given by the following:

a. *Brightness (or Luminance) B.* A photometric measure of the luminous intensity (emitted by an extended source) per unit area of the luminous flux per steradian. It is expressed by the relation,

$$B = \frac{I_L}{A} \tag{3-165}$$

where I_L is the luminous intensity, and A is the area. The units most often employed for B are the candle/ft^2 and the footlambert which is $\frac{1}{\pi}$ candle/ft^2. Other units in which it may be expressed include: stilbs (candle/m^2), apostilbs (10^{-4} lambert), and millilamberts (10^{-3} lambert) which is approximately 1 footlambert.

b. *Contrast C.* The contrast of an image is defined as

$$C = \frac{B_i - B_b}{B_b} \tag{3-166}$$

where B_i is the image brightness, and B_b is the average brightness in the image plane. When $B_i - B_b = \Delta B$ is small, it is found that the same visual sensation is produced for a constant value of $\frac{\Delta B}{B_b}$ for a wide range of B_b values. This is called Fechner's Law. It gives the value of ΔB

required to produce equal contrast increments. Over the useful range from a few footlamberts to a few thousand footlamberts, the quantity $\frac{\Delta B}{B_b}$ has a threshold value of about 0.02.

c. *Contrast Ratio* C_R. This quantity is a measure of the brightness range of the display surface and is given by the relation

$$C_R = \frac{B_{max}}{B_{min}} \tag{3-167}$$

where B_{max} is the maximum brightness, and B_{min} is the minimum brightness of the display. When including the brightness effects of ambient lighting $B_{ambient}$ the contrast ratio is expressed as

$$C_R = \frac{B_{max} + B_{ambient}}{B_{min} + B_{ambient}} \tag{3-168}$$

d. *Field.* A single scan through the display area once in the chosen line pattern. It is one of the two, or more, equal parts into which a frame is divided in the process of interlaced scanning.

e. *Frame.* A single scan covering all the resolution points of the entire display area. When using an interlaced scan, a frame is composed of two fields.

f. *Flicker.* The fluttering sensation caused by the eye's response to a periodic visual stimulation.

g. *Gray Level.* A unit employed to express contrast which is taken as a factor $\sqrt{2}$ difference in brightness between two sources. Thus, if B_o and B are two brightness levels which differ by a factor $\sqrt{2}$, i.e.,

$$\frac{B_o}{B} = \sqrt{2} \tag{3-169}$$

then the brightness step from B_o to B is one gray level.

h. *Illuminance* E. The incident luminous flux per unit area in lumen/ft^2 (foot candle), lumen/m^2 (lux or meter candle), or lumen/cm^2 (phot,

$$E = \frac{F}{A} \tag{3-170}$$

where

A = area

F = luminous flux, lumen

i. *Interlaced Scanning.* The process in which scanned lines are spaced an integral number of line widths and every other line is scanned during one field while the missed lines are scanned in the succeeding field.

j. *Luminance.* (See *Brightness.*) A photometric measure of the luminous intensity (emitted by an extended source)/unit projected area. It is expressed in either candle/ft^2 or candle/cm^2.

k. *Luminous Equivalent.* The ratio of luminous flux to radiant flux expressed in terms of lumen/w of radiant flux.

$$\text{Luminous Equivalent} = \frac{680 \int_{\lambda_1}^{\lambda_2} V_\lambda P_\lambda \, d\lambda}{\int_{\lambda_1}^{\lambda_2} P_\lambda \, d\lambda} \tag{3-171}$$

where λ_1, λ_2 is the wavelength interval of the radiation, V_λ is the standard luminosity function and P_λ is the radiant flux/unit of wavelength. The factor 680 lumen/w is the maximum response of the average eye where $V_\lambda = 1$.

l. *Luminous Flux* F. The quantity of visible radiation passing/unit time defined in terms of the standard luminosity function V_λ such that

$$F = \int_{\lambda_1}^{\lambda_2} V_\lambda P_\lambda \, d\lambda, \text{ lumen} \tag{3-172}$$

It is expressed in lumens where 1 lumen is the flux through a unit solid angle (1 steradian) from a point source of 1 candle.

m. *Luminous Intensity* I_l. The luminous flux F of a point source/unit solid angle Ω subtended at the source is called the luminous intensity and is expressed as

$$I_l = \frac{F}{\Omega}, \text{ lumen/sr} \tag{3-173}$$

It includes the power radiated over the entire spectral range of the source and is expressed in candles.

n. *Radiant Flux* P. The radiant flux is the time rate of flow of energy expressed in radiated watts. It is sometimes specified as a function of wavelength P_λ.

o. *Raster.* A raster is a predetermined pattern of scanning lines which provides substantially uniform coverage of the display area.

TABLE 3-14. STANDARD SYMBOLS FOR PAR. 3-7

SYMBOL	DEFINITION	TYPICAL UNITS
A	Area	ft^2
a	Aspect or width-to-height ratio	
B	Brightness	candle/ft^2, footlambert
B'	Brightness of phosphor screen	lumen
B_a	Screen brightness	
B_i	Image brightness	
B_b	Average brightness in the image plane	
B_{max}	Maximum brightness	
B_{min}	Minimum brightness	
$B_{ambient}$	Ambient lighting	
B_s	Separation between plates	m
B_m	Magnetic field intensity	Wb/m^2
C	Contrast	%
C_R	Contrast ratio	%
CRT	Cathode ray tube	
D	Film density	
D_s	Screen diameter	
D_s	Deflection of the spot on the screen	m
d_c	Mean diameter of the coil	
E	Illuminance of the screen	lumen/ft^2, footcandle
E	Exposure	
E	Illuminance	
EL	Electroluminescent	
e/m	Electron mass ratio	
F	Luminous flux	lumen
F	Fraction of the scan line used for information presentation	
f	Frequency, character regeneration frequency	Hz
f_c	Focal length of the coil	
$f/no.$	Aperture (focal length/diameter)	
G	Radiant power gain	w/w
H	Horizontal resolution	TV lines/in.
I	Electrical current	A
I	Intensity of incident light	
I'	Intensity of light transmitted	
I_a	Screen current	
I_ϱ	Luminous intensity	lumen/sr
IN	Number of ampere turns	
L	Distance from the center of the plates to screen	m
$L.E.$	Luminous efficiency	lumen/μA
ϱ	Length of the deflection plates	m
m	Magnification	
NLR	Nonlinear resistor	
n	Number of lines, number of symbols to be displayed	

TABLE 3-14. STANDARD SYMBOLS FOR PAR. 3-7 (Continued)

SYBMOL	DEFINITION	TYPICAL UNITS
$\frac{1}{n}$	Width of each line is $\frac{1}{n}$ times the picture height	
O	Opacity	
P	Radiant power	w
P'	Fraction of time the beam is on	
PC	Photoconductor element	
P_λ	Radiant flux/function of wave length	
$P.F.$	Power factor	lumen output/w input
PPI	Plan-position indication	
P_f	Fraction of time available for character generation	
R	Screen radius of curvature	
R	Repetition rate	Hz
RF	Radio frequency	
R_r	Character writing rate	
S	Electrostatic-deflection sensitivity of a cathode ray tube	m/V
SCR	Silicon control rectifier	
T	Character writing time	μsec
$1/T$	Basic generation frequency required of the sinusoidal signal	MHz
V	RMS applied voltage	
V_a	Anode voltage, accelerating voltage	kV,V
V_s	Screen voltage	V
V_d	Deflection plate voltage	V
V_λ	Standard luminosity function	
γ	Slope of characteristic curve for an emulsion	
ϵ_a	Transmission of the optical system	%
η	Phosphor screen radiant efficiency	w/w
λ	Wavelength	
σ_s	Photocathode radiant sensitivity	A/w
τ	Transmittance	%
Ω	Solid angle	

p. *Resolution.* The ability to delineate picture detail is expressed in the number of lines resolvable on a test chart. For a number of lines n alternately black and white, the width of each line is $1/n$ times the picture height. According to television standards, a TV line is taken to be either a black or white bar. The optical standard defines a black-white bar pair as a line. We shall employ the television standard unless stated otherwise.

q. *Scanning.* The process of analyzing the light values of a picture element consisting of the total picture area according to a predetermined method.

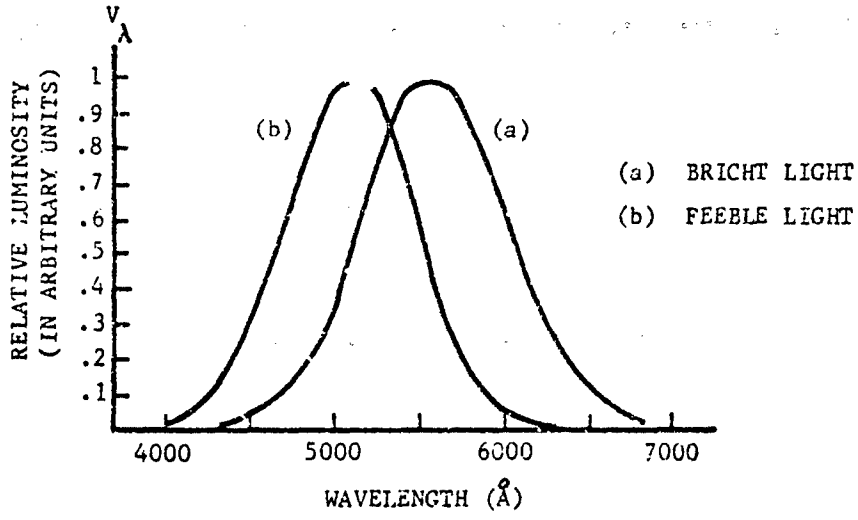


FIGURE 3-112. Standard Luminocity Curve for the Average Human Eye

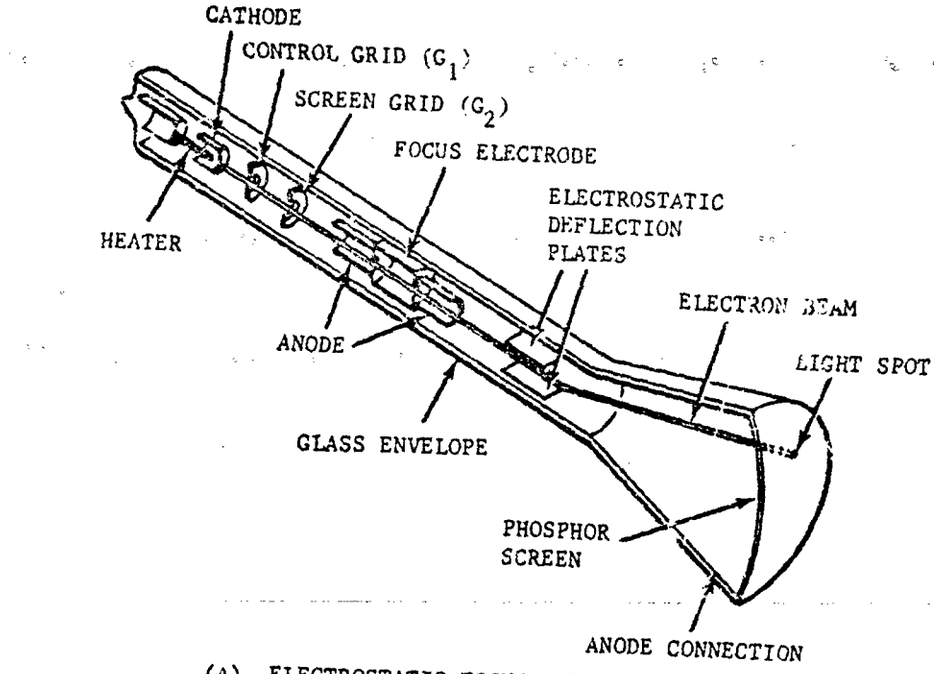
r. *Standard Luminosity Function* V_λ. Also termed "the relative visibility curve", the standard luminosity function represents the ratio of the power at the wavelength of the eye's greatest sensitivity to produce a given brightness sensation to the power at the chosen wavelength necessary to produce the same brightness sensation. The function is plotted in Fig. 3-112. The shift of the wavelength of maximum response toward the blue in feeble light is termed the Purkinje shift.

3-7.3 TYPES OF DISPLAYS

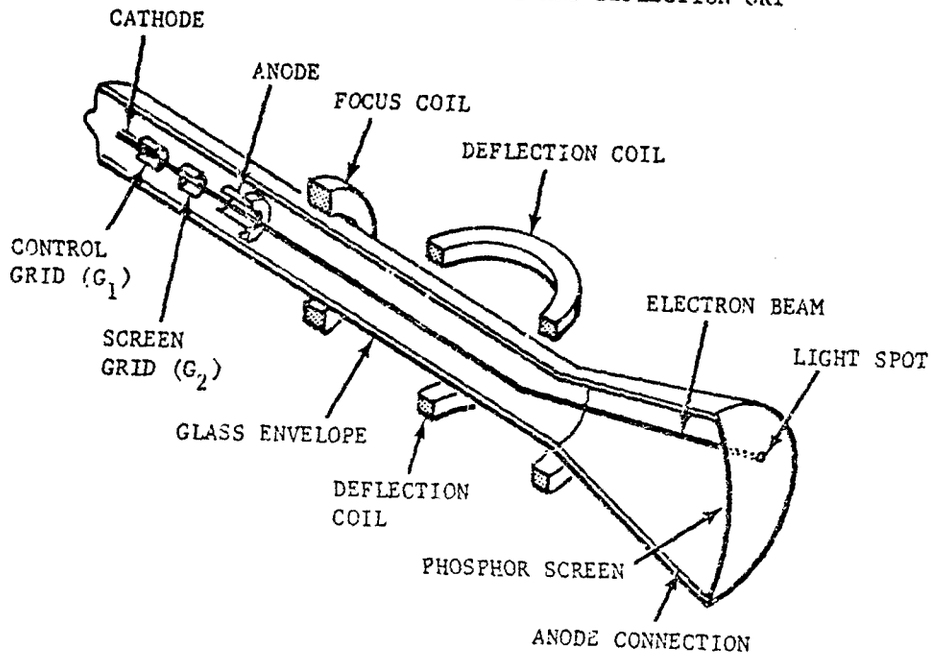
3-7.3.1 Cathode Ray Tube (CRT)

The CRT is a thermionic vacuum tube containing an electron-beam source, focusing system, deflection system, and phosphor screen. The conventional focus and deflection systems used are of the magnetic and electrostatic type shown in Fig. 3-113. Common to both types is the glass envelope (evacuated to 10^{-6} mm of Hg pressure or lower), the electron-beam source (gun), and the phosphor screen. The electrons are boiled-off from a cathode by electrical heating in typical usage and accelerated toward the phosphor screen by a large potential difference (from 6 to 18 kV typically) between the screen and the cathode. The cathode consists of a mixture of barium and strontium oxides. In traversing the length of the tube, the electron beam passes through a small circular aperture in a disk-shaped metal electrode called the control grid (G_1). This grid operates at about −60 or −70 V dc with respect to cathode potential and serves as a control to increase or decrease beam intensity. Next, the beam passes through a circular aperture in the disk called the screen grid G_2, which is at several hundred volts positive compared with the cathode and serves to accelerate the beam initially. After these, either an electrostatic lens or a magnetic coil is used to focus the electrons into a narrow beam with a focal point on the phosphor screen. Varying the potential of the electrostatic lens or the field strength in the coils allows the beam to be brought to a focus at different points along the tube axis. The beam is deflected in directions normal to the tube axis by use of electrostatic plates or magnetic coils. A final accelerating voltage of 6 to 18 kV is provided by a cylindrically shaped electrode and a conductive interior bulb coating.

Several basic factors which should be considered when evaluating the performance of a CRT are resolution, brightness, contrast ratio, focus and deflection sensitivity.



(A) ELECTROSTATIC FOCUS AND DEFLECTION CRT



(B) MAGNETIC FOCUS AND DEFLECTION CRT

FIGURE 3-113. Two Basic Types of Cathode Ray Tubes

3-7.3.1.1 Resolution

About 500-1000 TV lines are considered to be a commercially satisfactory tube. High resolution tubes employ about 4500 lines. The resolution is determined primarily by the size of the spot produced on the phosphor screen by the electron beam. Twenty- to fifty-mil spot sizes are typical in standard CRT's. Size depends upon phosphor thickness, grid voltage, and beam current. The thickness of the phosphor screen is held to a minimum (about 2 microns per kilovolt of anode supply during common usage) since the electron beam spreads during its penetration into the screen in proportion to its thickness. This spreading is the enlargement of the beam diameter as it penetrates into the phosphor layer. Spot size varies with the screen grid potential Vg_2 (with screen grid in a magnetically focused tube) in a manner shown by the curves in Fig. 3-114. The variation of spot size with beam current is shown in Fig. 3-115 for a typical high resolution tube. In a conventional tube, a change of from 50 to 300 μA would increase the spot size by a factor of 1.8. Under normal operating conditions the anode voltage will have little effect, if any, upon the spot size. Increases in anode potential will increase brightness which allows the beam current to be reduced to maintain the same brightness. So, for constant brightness, the spot size may be reduced by increasing the anode potential since the spot size varies directly with the beam current as mentioned above. Deflection defocusing is reduced with an increase in anode potential. For an electrostatically focused tube an additional increase in spot size is seen as the beam current is varied. This is corrected with dynamic focus control. Due to difficulties in focusing very small beams, we generally do not find tubes larger than 5 in. in diameter with spot sizes below 1 mil. Beam currents must usually be kept below 10 to 25 μA to achieve a small spot size. In extreme cases a spot size of 0.0005 in. can be attained but the useful screen area is small due to defocusing of the spot when it is moved a short distance off axis.

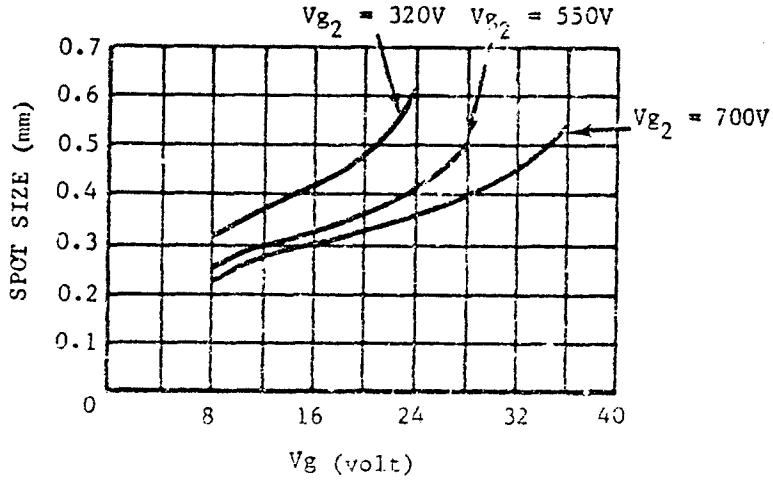


FIGURE 3-114. Change in Spot Size With Vg_2, Spot Size vs Vg_1; Anode Voltage V_a = 7000 V

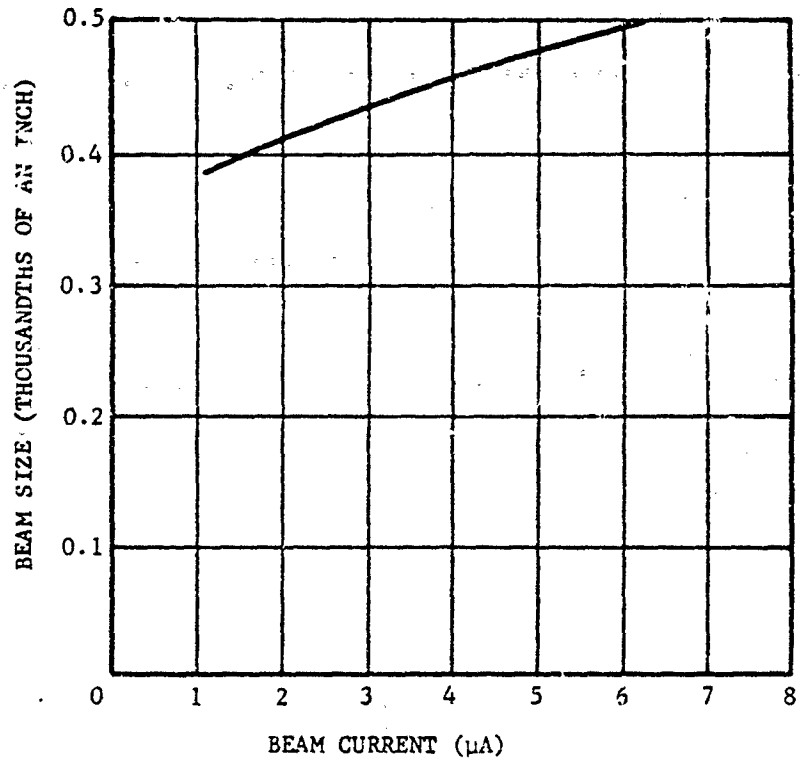


FIGURE 3-115. Spot Size for Several Beam Currents

3-7.3.1.2 Brightness

When the electron beam strikes the phosphor layer, it produces an excitation of the phosphors resulting in the emission of light referred to as cathodluminescence. More commonly, engineers refer to the brightness of the CRT screen. Screen brightness may be determined by using the relation,

$$B_a = (P.F.) \times I_a \times V_a \qquad (3\text{-}174)$$

where B_a is the screen brightness, $P.F.$ is the power factor, I is the screen current, and V_a is the screen voltage. (It should be noted that screen voltage and current differ from anode voltage and current due to effects of backscattering, secondary emission, transmission losses in penetrating the aluminized layer, thermal radiation, and other causes.)

To obtain the power factor (lumen output per watt input) multiply the power efficiency of the phosphor (radiated power in watts per electrical power input in watts) by its luminous equivalent. These quantities are listed in Table 3-15 for some commercially available phosphors.

The luminous intensity varies with the applied potential for constant beam current density as shown by the curves in Fig. 3-116. With constant applied potential there is a variation of luminous intensity with beam current density as shown in Fig. 3-117. For voltages above 6 to 8 kV, the free surface of the phosphor is coated with a thin aluminum film to prevent screen charging and to enhance reflection of light in the forward direction. The relative outputs of aluminized and nonaluminized screens are presented in Fig. 3-118. For voltages below the cross-over point (~ 6 kV) of the two curves, it is best to use an unaluminized screen; while for higher voltages an aluminized backing is recommended. The frequently used phosphors include zinc sulphides, zinc silicate, tungstates, and fluorides with the appropriate impurity activators. Table 3-16 lists the commercially available phosphor screens.

TABLE 3-15. PHOSPHOR DATA

PHOSPHOR	LUMINOUS EQUIVALENT, lumen/radiated watt	POWER EFFICIENCY, w/w	POWER FACTOR, lumen/w
P-1	521	0.06	31.3
P-2	462	0.07	32.3
P-3	381	0.041	15.6
P-4	291	0.016	4.7
P-5	88	0.025	2.2
P-11	137	0.101	13.8
P-15	253	0.051	12.9
P-16	25	0.049	1.2
P-19	387	0.0002	0.8
P-20	482	0.135	65.1
P-22B	44	0.15	6.6
P-22G	526	0.06	31.6
P-22R	104	0.043	4.5
P-24	363	0.026	9.4
P-25	319	0.0125	4.0
P-28	500	0.08	40
P-31	230	0.22	50.6

(For 10 kV potential and average beam currents)

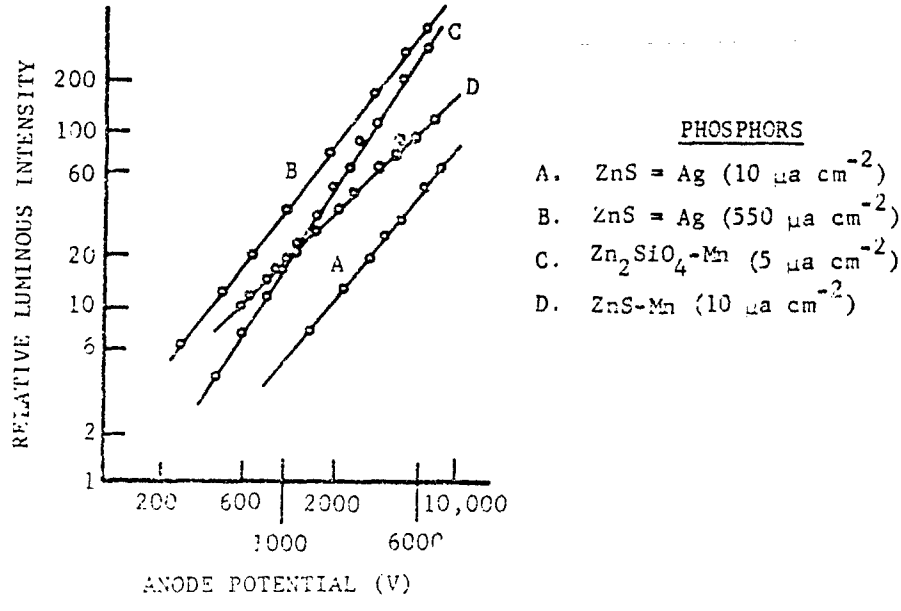


FIGURE 3-116. Variation of Luminous Intensity With Anode Potential

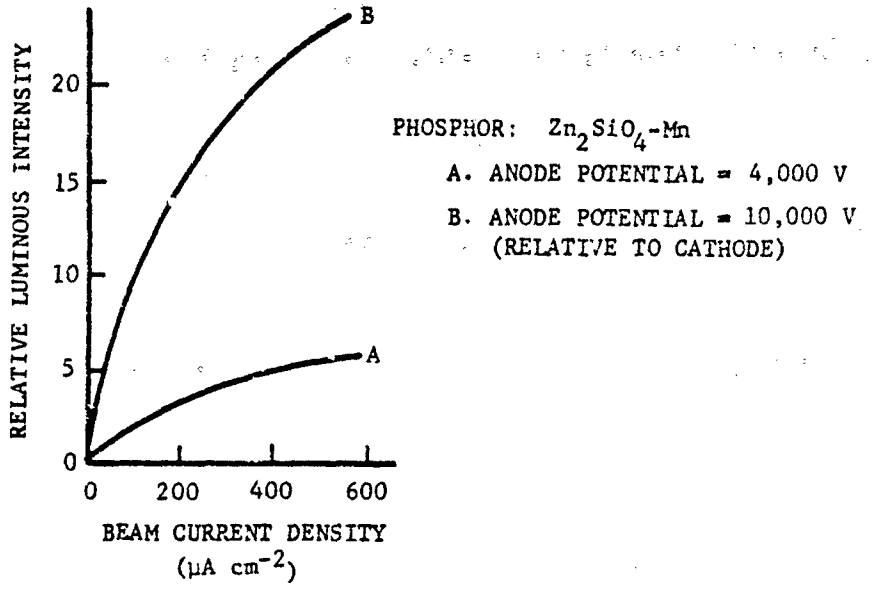


FIGURE 3-117. Variation of Luminous Intensity With Electron Beam Current Density

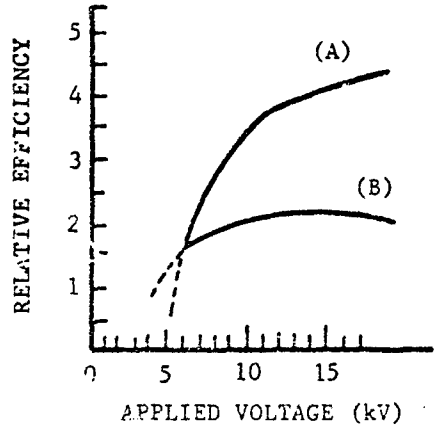


FIGURE 3-118. Efficiency of (A) Aluminized, and (B) Unaluminized Screens

TABLE 3-16. PHOSPHOR SCREEN CHART

REGISTERED PHOSPHOR TYPE	COLOR* FLUOR-ESCENT	COLOR* PHOSPHOR-ESCENT	PERSISTENCE**	GENERAL USE
P-1	YG	YG	M	oscillography, radar
P-2	YG	YG	M	oscillography
P-3	YO	YO	M	projection TV (with blue)
P-4	W	W	M-S	direct view TV
P-4	W	W	M to M-S (Sulfide Silicate Type)	theater projection TV
P-5	B	B	M-S	photographic applications
P-6	W	W	S	direct view TV
P-7	W	Y	M-S (blue) L (yellow)	radar and oscillography
P-10		(scotophor)		radar
P-11	B	B	M-S	photographic applications
P-12	O	O	L	radar
P-13	R O	R-O	M	radar
P-14	P-B	YO	M-S (blue) M (yellow orange)	radar and oscillography
P-15	UV and G	G	V-S (UV) S (Green)	flying spot scanning and photography
P-16	B-P	B-P	V-S	flying spot scanning and photography
P-17	W	Y	S (blue) L (yellow)	radar and oscillography
P-18	W	W	M to M-S	projection TV
P-19	O	O	L	radar
P-20	YG	YG	M to M-S	oscillography
P-21	RO	RO	M	radar
P-22	W	W	M-S (three phosphors R,B, and G)	color TV
P-23	W	W	M-S	direct view TV
P-24	G-W	G-W	S	flying spot scanning
P-25	O	O	M	radar
P-26	O	O	V-L	radar
P-27	RO	RO	M	color TV monitors
P-28	YG	YG	L	radar
P-29	W	YW	M(P-2+P-25)	radar and oscillography
P-31	G	G	M-S	oscillography
P-32	PB	YG	L	radar
P-33	O	O	V-L	radar

TABLE 3-16. PHOSPHOR SCREEN CHART (Continued)

REGISTERED PHOSPHOR TYPE	COLOR* FLUOR-ESCENT	COLOR* PHOSPHOR-ESCENT	PERSISTENCE**	GENERAL USE
P-34	BG	YG	V-L	oscillography, radar, virual information storage
P-35	B	B	M-S	photographic applications

*COLOR CODE

UV	Ultra Violet
P	Purple
B	Blue
G	Green
Y	Yellow
O	Orange
R	Red
W	White

**PERSISTENCE CODE

V-L	Very long	1 sec or over
L	Long	100 msec to 1 sec
M	Medium	1 msec to 100 msec
M-S	Medium short	10 μsec - 1 msec
S	Short	1 μsec - 10 μsec
V-S	Very short	less than 1 μsec

In addition, the brightness per unit area of the CRT display surface is proportional to the electron beam dwell time (i.e., the time the electron beam spends in one position on the screen), regeneration rate of the display, and persistence of the phosphor.

The desired brightness levels for most displays should be anywhere from 20 to 70 footlamberts for normal ambient brightness. For daylight TV one might reasonably require in excess of 100 footlamberts, while 1 to 10 footlamberts will suffice for night operation.

3-7.3.1.3 Contrast

The maximum contrast obtainable in a CRT is determined by reflection of light, screen curvature, and scattered electrons. The reflection of light from both surfaces of the faceplate and the walls of the tube causes a contrast loss of about 9 to 10 percent. The curvature of the screen enables light scattered from the point of impact of the electron beam to reach other parts of the screen and thus places a limit on the contrast ratio obtainable. An empirical relation for this is given as[87]

$$C_R(limit) = 1250\left(\frac{D_d}{R}\right)^2 \qquad (3\text{-}175)$$

where R is the screen radius of curvature, and D_d is the screen diameter. The contrast loss due to scattered electrons can usually be eliminated in the design of the tube. Typical contrast ratios for a CRT are about 20 to 1, which provides a contrast range containing about 8 to 9 gray levels. Based upon Fechner's Law, an observer should theoretically be able to distinguish 400 brightness increments over a 20 to 1 contrast range. Thus, limitations are in the quality of gray levels provided by the CRT.

3-7.3.1.4 Deflections

The electrostatic-deflection sensitivity S of a cathode-ray tube is defined as the deflection (in meters) on the screen per volt of the deflecting deflecting potential[88]

$$S = \frac{D_s}{V_d} = \frac{\ell L}{2B_s V_a}, \text{ m/V} \qquad (3\text{-}176)$$

where

ℓ = length of deflection plates, m
L = distance from center of plates to screen along axis of tube, m
B_s = separation between plates, m
V_a = anode potential, V
V_d = deflection plate potential, V
D_s = deflection of spot on screen, m

Sensitivity of the CRT tube is seen to be independent of both V_d and e/m (the electron charge to mass ratio) and depends solely upon the geometry of the tube (which is a constant) and the accelerating potential. Typical values of tube sensitivities are 1.0 to 0.1 mm/V, corresponding to deflection factors of approximately 25 to 250 V/in. These values are determined experimentally and given in the specification sheets.

In the case of magnetic deflection we find the deflection per unit magnetic field intensity is given by

$$(D_s/B_m) = \frac{\ell L}{3.4 \times 10^{-6}} \left(\frac{1}{V_a}\right)^{1/2}, \text{ m/(Wb/m}^2) \qquad (3\text{-}177)$$

where ℓ, L, V_a are as defined in Eq. 3-176, B_m is the magnetic field intensity in Wb/m^2. Since ℓ and L are geometrical constants we see that the sensitivity varies inversely with the square root of the anode potential.

When operation over a wide range of frequencies is desired, an electric deflection CRT is preferred. The upper frequency limit (about 10 MHz) is determined by the capacitance between the deflecting plates (a few picofarads) and the capacitance of the amplifier and lead wires.

Magnetic deflection has a much smaller bandwidth (about 10 kHz) due to the inductance of the coils. The same input voltage at different frequencies will result in different coil currents and hence different deflections. This limits magnetic deflection to applications involving a constant sweep frequency.

Most displays require a fairly bright screen (about 100 footlamberts) which demands high beam currents (25μA) at large values (12 kV) of anode potential V_a. Since electric deflection varies inversely with V_a and magnetic deflection varies inversely as $V_a^{1/2}$, as shown in Eq. 3-177, we see that it becomes more economical to obtain large deflections with a magnetic-deflection tube. The time required for a full screen magnetic deflection is from 15 to 30 microsec due to the setting time of the yoke. The standard deflection rates are 60 Hz for vertical and 15,750 Hz for horizontal.

3-7.3.1.5 Focus

Electric focusing is always used in electrically-deflected tubes. This type of focusing requires little power and is insensitive to power supply variations since the focusing electrode voltage is obtained from a bleeder across the accelerating voltage supply. Thus, these two voltages will vary in proportion and the spot will remain focused for small variations in the dc supply. Magnetic focusing is obtained at more than one value of the current in the coil. The best focus is at the minimum value. For an average coil, use the relation

$$IN = 220(V_a d_c/f_c)^{1/2} \qquad (3\text{-}178)$$

where

IN = number of ampere turns
V_a = anode voltage, kV
d_c = coil mean diameter } same units
f_c = coil focal length }

3-7.3.1.6 Bandwidth

The video bandwidth requirement may be calculated from the following general relationship

$$Bandwidth = \frac{1}{2}\left(\frac{aH}{F_r}\right) nR \qquad (3\text{-}179)$$

where

a = aspect ratio (width-to-height ratio)
H = horizontal resolution, TV lines/in.
n = number of scanning lines in field
R = repetition rate, Hz
F_r = fraction of scan line used for information presentation

Frequently used values are a = 4:3, H = 60 TV lines per inch, n = 500 lines, R = 60 Hz, F_r = 0.84. The usual bandwidth requirements are from 1 to 8 MHz.

3-7.3.2 Storage Type Cathode Ray Tubes

These devices are special types of cathode ray tubes which have the capability to store information on the display screen and to erase it when desired.

This type of tube is particularly useful for display of transient or temporary information and narrow band TV pictures for extended intervals or at controlled persistence rates, and with higher light output than is available from long persistence phosphors. This makes the display on a direct view storage tube extremely useful in infrared systems as a means of presenting the information to the observer. These tubes are found in sizes from 3 to 21 in. with display storage time of from 30 sec to 3 or more min. Various signal levels can be stored so that a full-tone picture can be presented.

As shown in Fig. 3-119 the basic components are a write gun, a flood gun, a deflection system, a collimating system, a storage screen, and a display screen. By secondary emission charging, a moderately energetic writing gun deposits a positive charge pattern on the insulating target. The secondary electrons are collected at the collector. The charge pattern controls the transmission of a broad, well-collimated beam of electrons which floods the target. Where the charge pattern is positive, the low-energy flood electrons pass through the holes in the screen and are accelerated to the phosphor screen which is operated at high potential to provide a bright display. The flood beam has low energy, but high current. The storage grid is comprised of a thin deposit of insulating material covering the side of the fine metal mesh facing the guns. The collimating system makes all electrons approach perpendicular to the display screen to avoid shading.

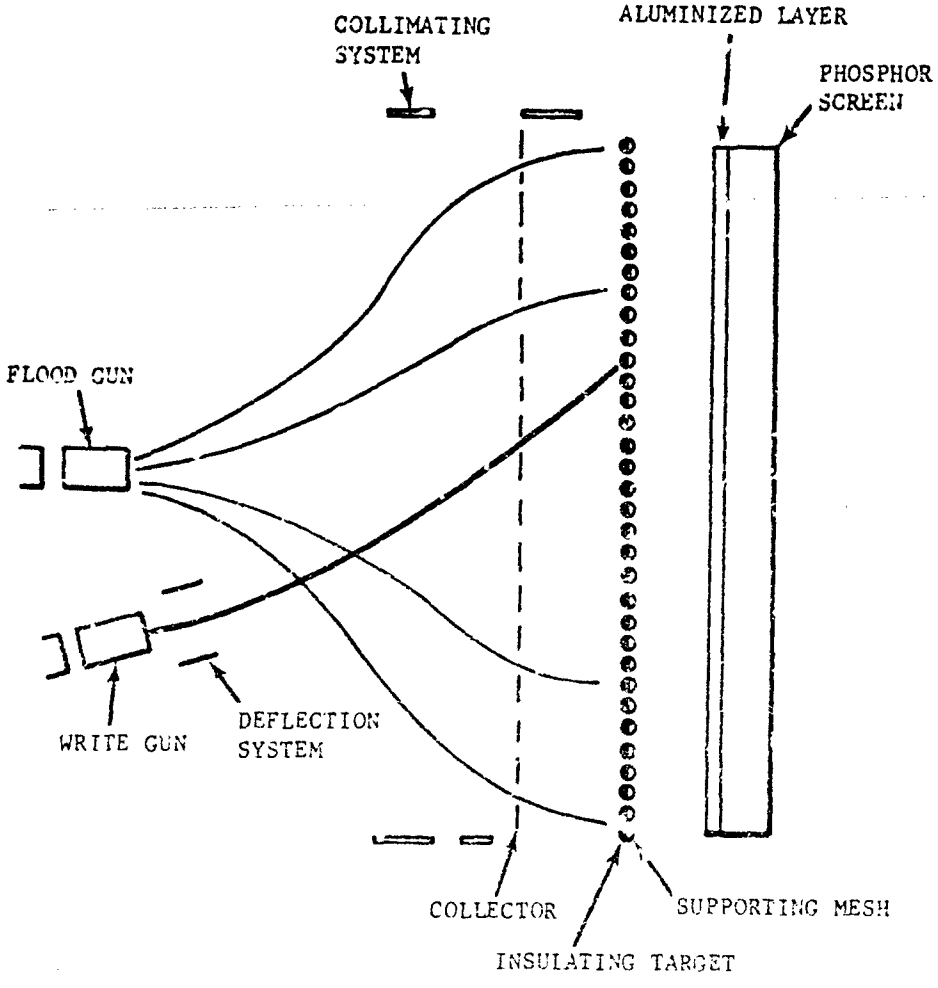


FIGURE 3-119. Display Storage Tube

The flood electrons do not alter the charge distribution on the storage screen, however, they do produce positive ions which do alter the charge on this grid. For this reason, the storage time is limited to only a few minutes. Since the production of positive ions and brightness of display image are both proportional to the flood beam current, a trade-off must be made between storage time and screen brightness. Erasure of the stored charge pattern is accomplished simply by applying a 5-V negative pulse to the storage surface through the metal mesh. This lowers the potential of the storage grid below the cutoff, so the secondary electron emission is less than unity. Erasure can be achieved by application of one long pulse of several milliseconds duration or by a train of narrow, high-frequency pulses to obtain a viewing time of variable duration without the flicker which accompanies a low-frequency erase.

Typical output brightness ranges from 1250 to 4000 footlamberts. A special projection model is available with brightness near 20,000 footlamberts. At 10 percent saturation brightness, the resolution may be about 90 to 100 lines per in., but at 90 percent saturation brightness it reduces to near 35 to 40 lines/in. Under general operation conditions, resolutions are from 50 to 70 lines/in. for these tubes. Writing speeds at 50 percent of saturation brightness range from 10,000 to [illegible],000 in./sec. Viewing screen currents are hi[illegible] 500 μA at 10 kV. Several stable high[illegible]plies are required, for the guns and [illegible] must be coupled into high-voltage circuits.

There are three types of display storage tube–the halftone, bistable, and selective erasure tubes–which operate in this same manner with minor alterations. Some halftone tubes are the Iatron, Storatron, and Tonotron with Memotron and Typotron being the best available types. These are discussed in more detail by such manufacturers as ITT Industrial Laboratories and Hughes Aircraft Company.

3-7.3.3 Electrical-readout Storage Tubes

This device is an electrical input-electrical output storage tube in which information can be stored for hours, erased in a second, or erased gradually at a controlled rate and superseded with new information. Write in and read-out of information can be performed simultaneously or sequentially.

The device can be used to convert from one scan frequency or scan mode to another as in wide-to-narrow bandwidth transmission, and plan-position indication (PPI) to TV scan; and for data transmission and storage with random, simultaneous, and repetitive access. Halftone operation enables transmission of video, multilevel, and time dimension information. Integrating properties of these tubes allow for the buildup of repetitive signals submerged in noise and identification of superimposed information. Thus, weak signals may be enhanced by a time exposure process to yield a higher signal-to-noise ratio. Also, a constantly decaying picture can be converted into one of uniform brightness and may be studied for several minutes. These features make this an extremely useful device for presentation of infrared data where signals are typically weak. These tubes can also be operated to indicate only moving targets by electrically cancelling repetitive signals and retaining only new signals.

The tube has two high-resolution, opposed, on-axis electron guns for the write-in and read-out functions. These use either the electrostatic or magnetic types of focus and deflection, depending upon the model. The two beams are collimated to approach the storage surface normally at uniform velocities to prevent shading. The storage surface has a grid which is charged to a level corresponding to the input modulation on the write-in gun and thus controls the transmission of the reading beam to the collector grid. A voltage variation corresponding to the original input signal is generated across a load impedance in the collector grid circuit. RF separation or video cancellation is required to eliminate the write signal from the output when simultaneous writing and reading is performed. Stored information can be caused to decay at a controlled rate by varying the collector to storage dielectric potential. Increasing the positive potential on the collector increases the electron collection efficiency and the storage time is long.

The storage grid is a thin dielectric layer evaporated upon a thin metal film, supported by a fine mesh screen (see Fig. 3-120). The focused, low-energy read-out beam scans the storage grid with a constant current beam to maintain it at a

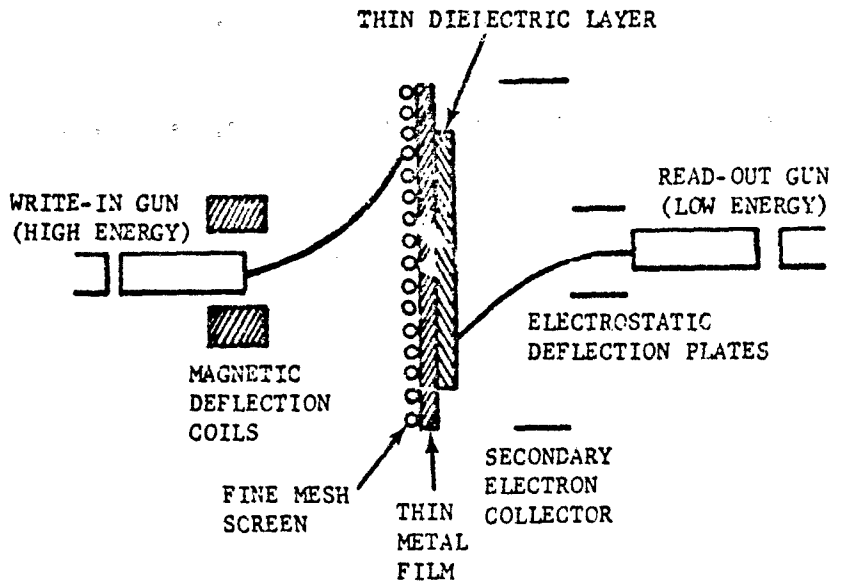


FIGURE 3-120. Typical Scan-conversion Tube

fixed potential approximately equal to the potential of the collector and about 20 V above read gun cathode potential. Electrons penetrating the storage grid, as determined by the charge levels established on the dielectric surface by the write-in beam, are accelerated to the positive collector grid where they are collected, generating an output signal across the impedance in the output grid circuit. The read-out beam does not disturb the charge pattern on the negatively charged storage grid. The write-in beam bombards the opposite side with high energy electrons which pass through the metal film and induce a volume conductivity in the target dielectric partially discharging the uniform gradient established by the secondary beam. The change in potential is proportional to the charge supplied by the write-in beam. Intensity modulation of this beam allows a charge pattern to be placed on the target. The highest secondary emission ratio, and therefore the fastest writing speed, is attained typically at a cathode to dielectric potential difference of about 350 V. Any area of the target can be erased by operating the write-in gun above the secondary emission crossover, and written by operating the write-in gun below the crossover (the crossover is the point at which the ratio of secondary to primary electrons is equal to unity). These tubes are capable of resolution near 1000 TV lines. Manufacturers of electrical-readout storage tubes should be contacted for additional data.

3-7.3.4 Character Generation Tubes

The devices most frequently employed in systems where it is desired to display alphanumeric information are described in the paragraphs which follow.

3-7.3.4.1 Matricon

The matricon is an electron beam device which generates alphanumeric characters in the form of electronic video signals for display on a cathode ray tube. Its basic components are an electron gun, horizontal and vertical electrostatic deflection plates, a collector, a shield plate, and a target as presented in Fig. 3-121. The target area consists of a matrix pattern of electrically isolated conducting elements. There are about 50 separate metallic elements which are independently switch-controlled. The electron beam is made to raster scan over the target matrix. Application of from 0 to +5 V on these

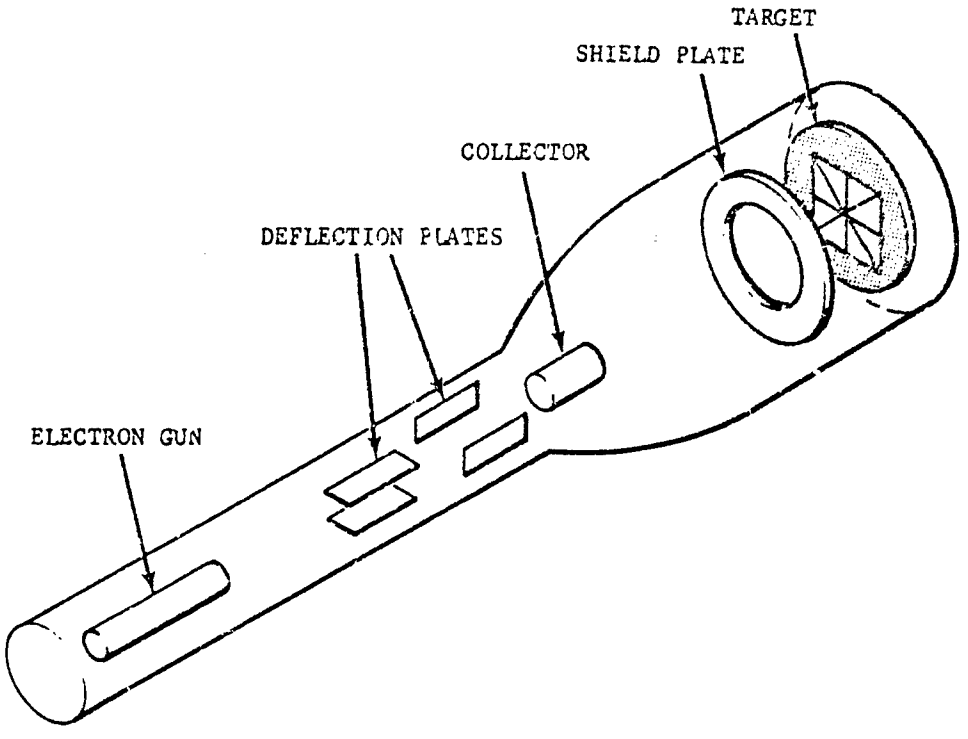


FIGURE 3-121. Matricon Tube

target elements varies their secondary emission. The secondary electrons are collected at the collector and form a video signal. By selecting the appropriate matrix elements, an alphanumeric character can be produced as a video signal and displayed on another CRT which is raster scanning in synchronization with the matricon beam. The decelerator mesh and shield plate serve to control the electron optics of the tube. The matricon character writing speed is about 50 microsec; it is produced by the Sperry Rand Corporation.

3-7.3.4.2 Monoscope

Another character generating tube frequently employed is the monoscope shown in Fig. 3-122. This tube also operates on the principle of secondary emission and generates an electronic video signal for display on a cathode ray tube. The electron beam in the tube raster scans the selected target element, which is one of 64 alphanumeric characters of aluminum on a carbon-ink background. Due to the difference in the secondary emission ratio between the aluminum and the ink, a variation is produced in the current collected by the collector. This current variation is transformed into a standard video signal. This video signal can be displayed on a standard CRT which is scanned synchronously with the monoscope electron beam. Standard monoscopes generate average quality characters with writing times of approximately 33 microsec, which is slightly slower than shaped-beam tube rates. For more detailed information, the reader should contact the A. B. Dick Company.

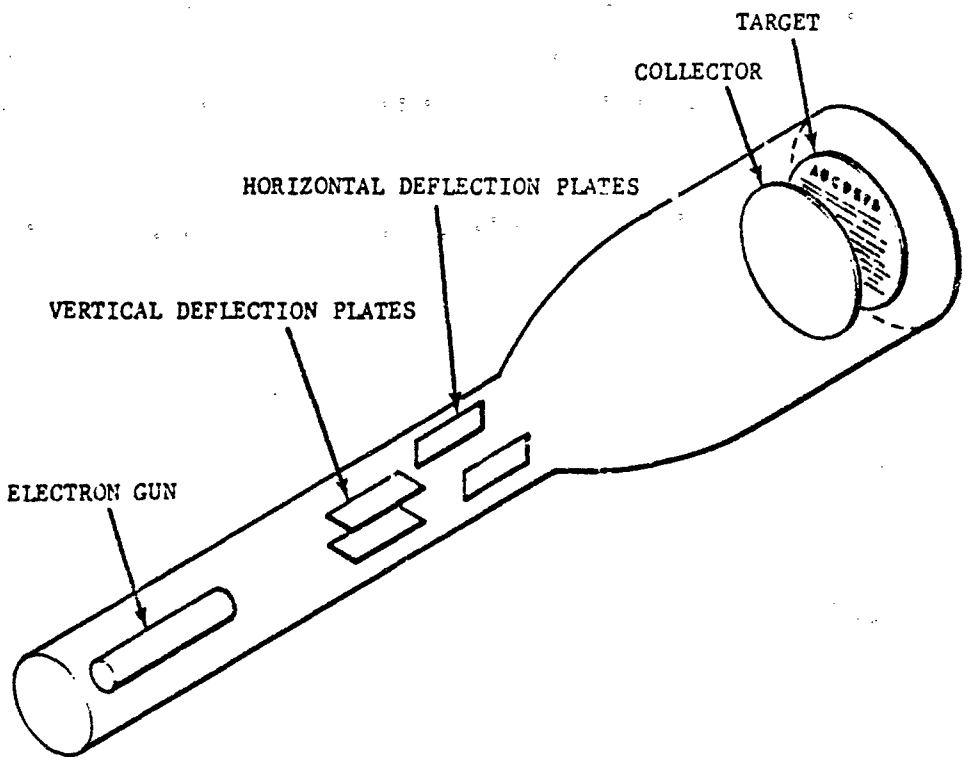


FIGURE 3-122. Monoscope Tube

3-7.3.4.3 Shaped-beam Tube

The shaped (or extruded) electron beam tube produces alphanumeric characters on a phosphor screen for direct viewing. The tube consists of an electron gun, character selection deflection plates, a character mask, a convergence coil, compensation plates, and electrostatic deflection plates as shown in Fig. 3-123. The mask has 64 character-shaped apertures in it. The electron beam is directed through the selected aperture to form it into the shape of the desired character which it then generates on the phosphor screen. The characters can be made from 0.1 to 0.8 in. high. These are very legible with constant brightness from character to character. Deflection of the shaped beam to the desired position on the tube face is accomplished by electrostatic or electromagnetic deflection in a step-like manner. Selection, reference, and deflection voltages are established to display the desired character at the specified position on the screen; the beam is unblanked and then a new set of values established. The ultimate speed of this device is determined by the choice of components. With magnetic deflection, one can achieve from 10,000 to 40,000 characters per sec. If electrostatic deflection is used, the deflection voltages can be established in 2-3 microsec. With a display time of 8 microsec per character, a display rate of about 100,000 characters per sec is reasonably achieved. The tube is generally very long, about 40 in. for a 19-in. screen, due to the many components involved. They are found with screen sizes from 7 to 21 in. In addition, a tube may be obtained having an 8.5 × 0.5 in. fiber-optic faceplate. This type of tube provides page-width printing. These tubes tend to be very expensive.

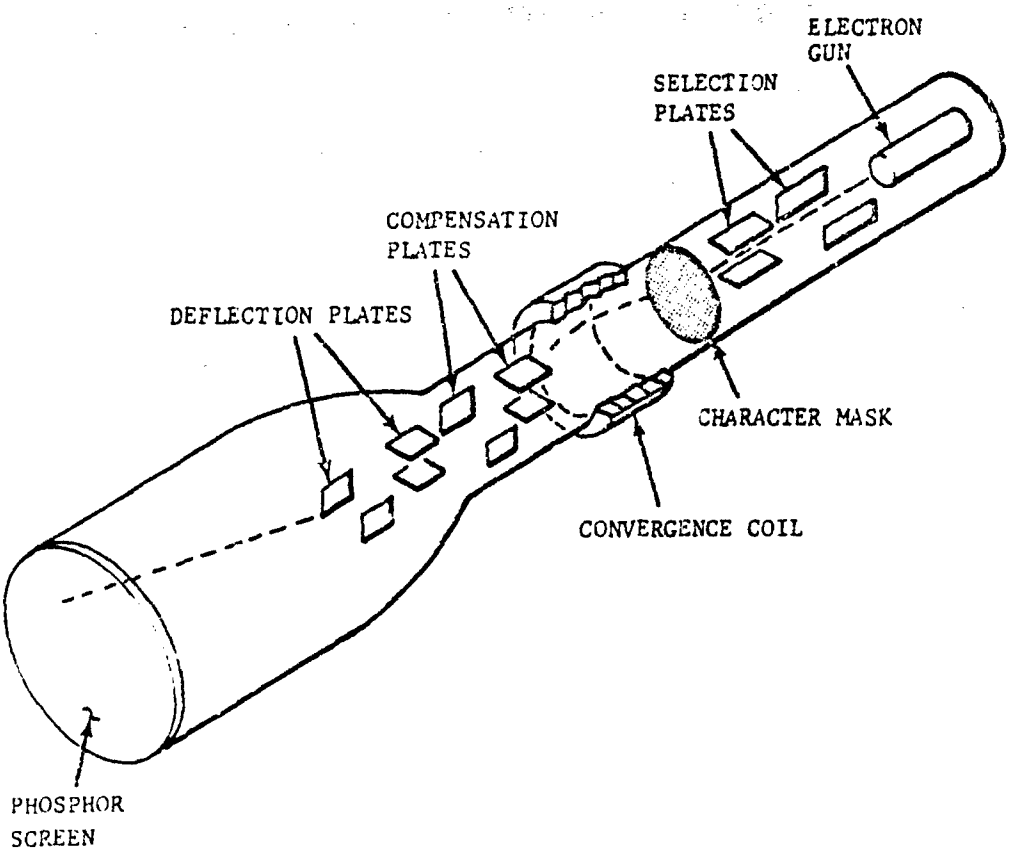


FIGURE 3-123. Shaped-beam Tube

3-7.3.4.4 Stroke Technique

Another technique for the generation of characters is the stroke method in which the characters are formed by a sequence of line segments, called strokes, produced on the screen of a cathode ray tube by the movement of the electron beam. Generally, from 9 to 16 consecutive strokes must be used to produce a character. Each stroke is produced by changing the voltages applied to the deflection plates. A resistor matrix is constructed for each required voltage change and a character generator clock selects the matrix point whose voltage is to appear at the deflection plate or coil, at each instant of time. Also, a Z-axis modulation is performed synchronously. A separate resistor matrix exists for each character. Stroke techniques have the highest character writing speeds, about 5 microsec per character.

3-7.3.4.5 Lissajous Techniques

Characters are also generated in a cathode ray tube by the Lissajous technique. A Fourier series of sine and cosine terms is used to construct a voltage waveshape which is applied to the deflection plates for the generation of the character. The specific character displayed depends upon the particular waveshape. For example, to produce the letter "O", a sine and cosine wave of the same frequency are employed. Usually, the first five harmonics of this frequency will suffice to generate good quality characters. If T is the character writing time, the $\frac{1}{T}$ is the basic generation frequency required of the sinusoidal signals. For the same bandwidth, this technique produces higher quality characters than the stroke technique, but costs much more. The Lissajous technique is capable of

writing characters at the rate of 53 microsec each.

3-7.3.4.6 *Alphanumeric Indicator Tubes*

These tubes are capable of displaying letters, symbols, and numbers, but only one character at a time per tube. By operating several tubes simultaneously a multiple character format may be displayed. The most popular of these devices is the NIXIE tube manufactured by Burroughs Corporation. This indicator is an inexpensive, gas-filled, cold-cathode, glow-discharge tube having a common anode and ten individual metallic cathodes which are formed in the shape of numerals, alphabetic characters, or special symbols. When a negative voltage, with respect to anode potential, is applied to the selected character it becomes the cathode for a gas discharge diode. The cathode will be surrounded by the glow of the ionized gas and only that character will be visible. A minimum cathode current is required for a complete glow. Current values do not usually exceed from 2 mA to 6 mA (depending upon tube size) when used at the rated voltage of 170 V dc. Brightness in the order of 200 footlamberts may be expected for each character. Character heights range from 0.3 in., visible at a distance of 11 to 14 ft, to 2.0 in. which is visible at a distance of 90 to 100 ft. Image contrast will depend upon ambient lighting conditions. Characters may be generated at rates in excess of 1000 per sec and as high as 20,000 per sec. Tube life is better than 20,000 hr.

3-7.3.4.7 *Comparison of Character Generating Devices*

The factors to be weighed when comparing the various character generating devices should include writing rates, brightness, bandwidth, cost, size, quality and type of characters, and versatility. The writing rate is found by the relationship

$$R_r = fnP_f \qquad (3\text{-}180)$$

where R_r is the character writing rate, f is the character regeneration frequency, n is the number of symbols to be displayed, and P_f is the fraction of time available for character generation. In general, the character generation frequency need not be greater than 60 Hz, but should not fall below 15 Hz at the risk of being disturbing to the receiver of a CRT display. The greater the writing frequency the lower will be the brightness of the characters. The brightness in footlamberts is given by the relationship

$$B = \frac{B'fP'}{R_rA} \qquad (3\text{-}181)$$

where B' is the brightness in lumens of the phosphor screen as determined by beam current, anode voltage, and phosphor type; f is the regeneration frequency in Hz; P' is the fraction of time the beam is on; R_r is the character writing rate; and A is the character area in ft^2. With the highest speed character generators and regeneration frequency of 30 Hz, one can reasonably expect character brightness in the order of 20 footlamberts.

Bandwidth for a raster scan system should be 150 to 200 times the character writing rate. For a stroke character generating technique, a bandwidth of 50 to 100 times the writing rate will suffice. In the case of Lissajous character generators, a bandwidth of 10 times the writing rate will suffice.

3-7.3.5 Image-converter Tubes

This device, illustrated in Fig. 3-124, converts an image formed in one wavelength of radiation into an image in a different wavelength for viewing. The tube contains both the sensor and display in one unit. The sensor is the photocathode at the front end of the tube which is held at ground potential. Light incident upon the photocathode causes the emission of electrons which travel down the tube following the potential gradient and striking the phosphor screen at the rear of the tube which is held several (typically 5 to 15) kilovolts above cathode potential. The phosphor screen then emits visible radiation. So, by projecting an image in one wavelength on the photocathode surface, that image is reproduced in a different wavelength on the phosphor screen. Some image converters employ proximity focusing of the electron image onto the phosphor screen; most, however, use electrostatic focusing by addition of an annular focusing ring to the tube. Dark current is usually below 0.1 μA. The resolving power of image converter tubes ranges from 10 to 80 line pairs/mm depending upon the tube type and quality. A magnifying eyepiece is

normally employed due to the small screen size. The eyepiece should be a good quality achromat with a field slightly in excess of the useful phosphor screen area. A 7-power Hastings triplet will suffice for most instances.

For a visual display in which the human eye is to observe the screen directly, a P-20 phosphor is most suitable since its spectral emissive curve most closely matches the response curve of the eye (standard luminosity curve). The brightness of the screen may be calculated by using the relation,

$$F = (L.E.) \times I \qquad (3\text{-}182)$$

where

F = luminous flux, lumen

$L.E.$ = luminous efficiency, lumen/μA

I = current, μA

For a screen voltage of 6 kV, the luminous efficiency should be about 0.19 lumen/μA. A curve of luminous efficiency versus screen voltage is given in Fig. 3-125.

The total power radiated by the phosphor screen is found by

$$P = \eta V_a I \qquad (3\text{-}183)$$

where

P = radiant power, w

η = phosphor screen radiant efficiency, w/w

V_a = accelerating voltage, V

I = current, A

At 6 kV, the radiant efficiency of P-20 is about 0.066 w/w.

In addition to the above two measures of output brightness, there is another significant quantity to be considered when evaluating an image tube for display purposes. This is the radiant power gain which is defined as the ratio of total exit radiant flux in watts to total monochromatic incident radiant flux in watts and rates the image brightness intensifying ability of the tube. We may write,

$$G = \sigma_s V_a \eta \qquad (3\text{-}184)$$

where

G = radiant power gain, w/w

σ_s = photocathode radiant sensitivity, A/w

V_a = accelerating voltage, V

η = phosphor screen efficiency, w/w

The quantity G will vary with wavelength as does the photocathode sensitivities shown in Fig. 3-126.

For an S-1 photocathode and P-20 phosphor combination at screen potential of 16 kV, the radiant power gain at 0.8 micron wavelength is 3.8 w/w.

The standard image converter tube has a magnification factor less than unity, usually about 0.75 or less, which means the image is demagnified electron-optically. The electron density in the image increases inversely as the square of the magnification, so an image reduction of say 0.20 gives an electron density and resultant brightness gain of 25 times. This additional brightness gain, due to demagnification, must be multiplied by the radiant power gain G, previously discussed, to give the net radiant power gain of the tube. It should be noted that these gain factors are idealized in that they do not take into account any losses in the tube.

When evaluating the quality of the image produced, its distortion is measured. The distortion of the tube is expressed as a percentage departure of the radial distance of an image point on the screen from the product of the radial distance of the corresponding object point on the photocathode and the paraxial magnification. The percentage of distortion is given as a curve in the image characteristics of the tube. Typical curves are shown in Fig. 3-127. As an example of their utility, an illuminated point 5 mm from the center of the photocathode would theoretically yield a point 3.3 mm from the center of the screen, but the 4.2 percent distortion at this radial distance causes the point to appear at a radial distance 3.44 mm–3.3 + 0.042 (3.3) = 3.44–from the center. The curve for resolving power shows that it falls off with increased radial distance from the center. For further discussion of the image tube with concern to its ability as a sensor, refer to par. 3-4.

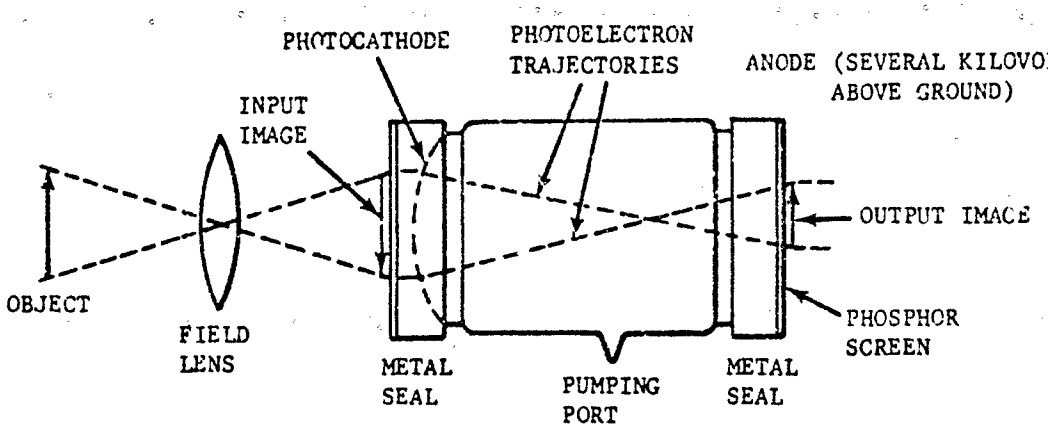


FIGURE 3-124. Typical Image-converter Tube in Use

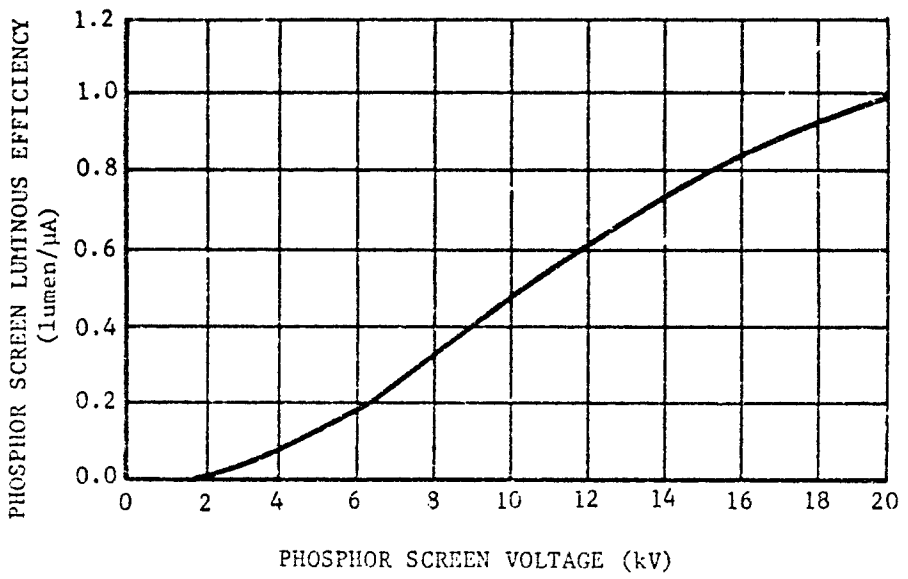


FIGURE 3-125. Screen Luminous Efficiency

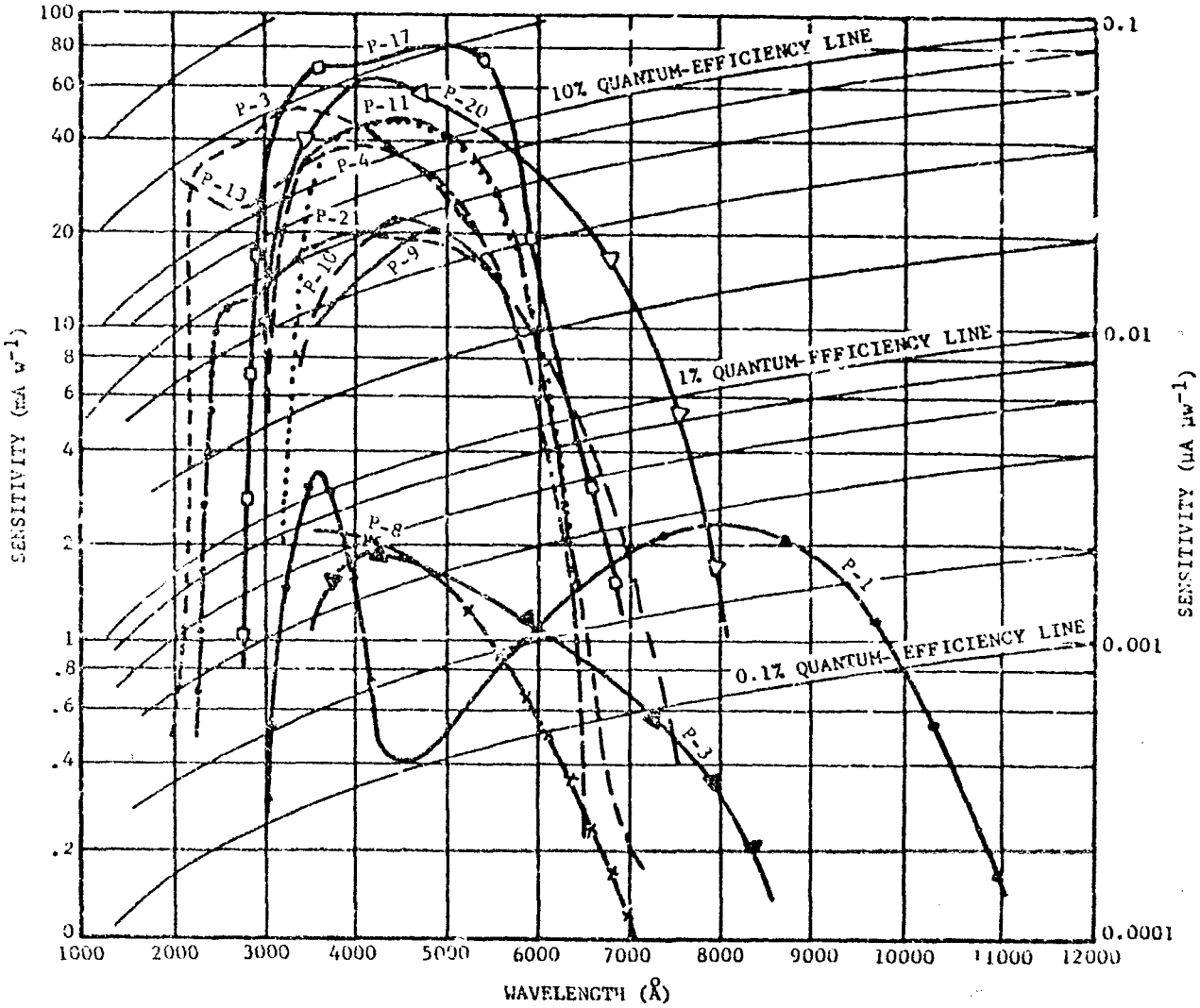


FIGURE 3-126. Photocathode Sensitivities

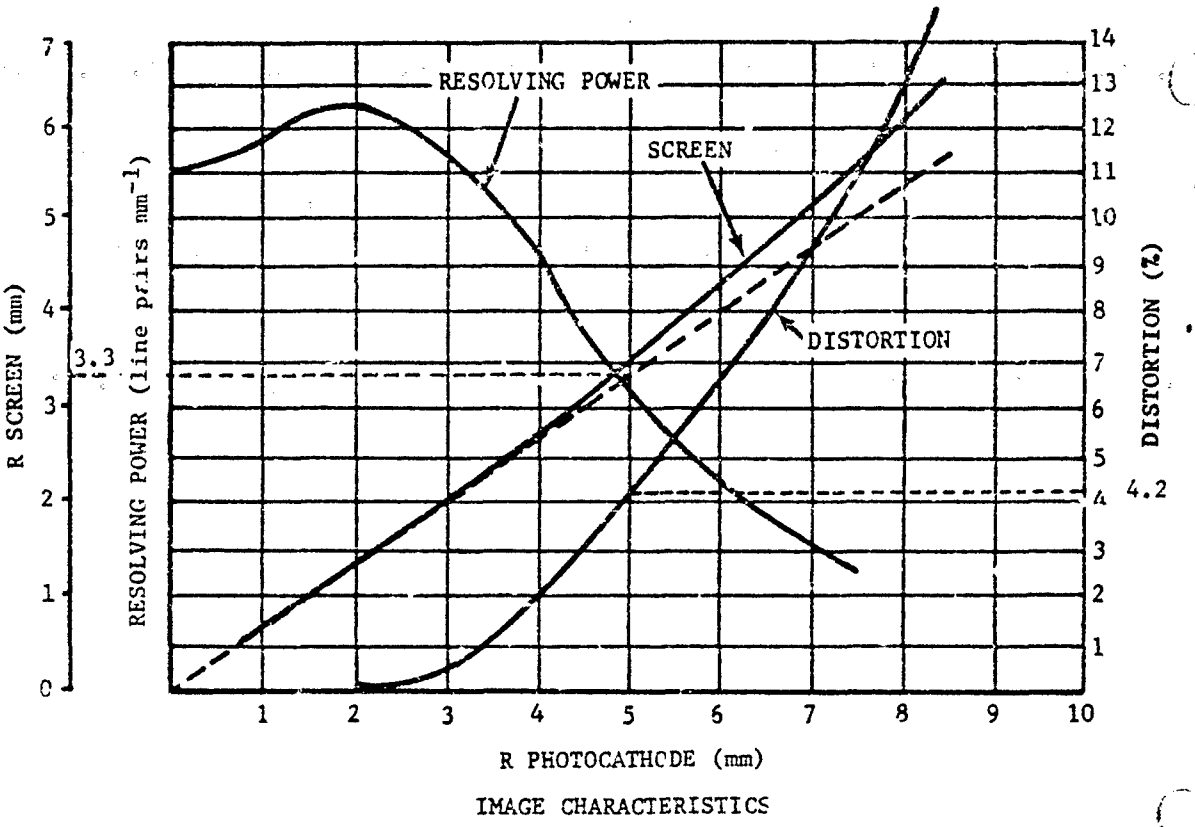


FIGURE 3-127. Image Quality of Image Converter Tube in Terms of Distortion

3-7.3.6 Tape Recording

Display systems[39] utilize tape recording as a method of storing and displaying video information. The two most widely used methods of recording on tape are magnetic and thermoplastic recording.

3-7.3.6.1 Magnetic Tape

Magnetic or signal-waveform recorders utilize magnetic-tape apparatus and are known as television tape recorders. Recorded tapes are reproduced on the same type of equipment; no special reproducer is required.

The recording of a video waveform on magnetic tape has the advantages of immediate reproduction of a recording without processing, fidelity of picture reproduction, erasability, and reusability of the recording medium.

The range of frequencies for video recording is approximately from 5 MHz down to 15 Hz, about 18 octaves at a tape speed of 15 in./sec. A conventional magnetic recording system is limited in bandwidth to about 10 octaves with a practical signal-to-noise ratio. This limitation is imposed at high frequencies by the impossibility of satisfactory reproduction of recorded wavelengths shorter than the head gap width. Lower frequencies on reproduction suffer a 6-dB loss in signal level per octave because of the reduced rate of change of flux across the gap. The signal level 10 octaves below a given reference frequency thus has a 60-dB lower signal-to-noise ratio.

By utilizing the lower sidebands of an FM system, the video spectrum is contained within a range of three octaves, consistent with a high degree of signal transfer linearity. The undeviated carrier frequency is typically between 5 and 5.5 MHz, corresponding to the blanking level of

the waveform. A peak white signal deviates the carrier to approximately 7, and at the synchronizing pulse dips to 4.3 MHz. Picture resolution is in the range 350-450 lines, and signal-to-noise ratios of 42 to 48 dB are obtained. During the reproduction process, the FM signals picked up from the tape are preamplified, sequentially switched, limited, and demodulated.

A high degree of mechanical accuracy is required of the recording head from angular position. An error of 1 degree represents a timing error of about 12 microsec (approximately 1/5 of a TV line).

In digital magnetic tape recording, standards are about 1000 bits/in. for parallel recording. This can be increased to approximately 2000 bits/in. under laboratory-controlled conditions. For serial recording, packing densities of 4000 bits/in. may be achieved with a conventional nonreturn to zero technique. If ultra-high density methods, employing modulation techniques and bias are used, 10,000 to 20,000 bits/in. may be achieved; thus, in the laboratory, rates of 20 megabits/sec represent the state-of-the-art.

3-7.3.6.2 Thermoplastic Film

The thermoplastic recording is made in the form of small ripples on the surface of a plastic film. A scanning electron beam produces the ripples. The electrical input is similar to the magnetic case and the image is like that on photographic film. The film consists of a thin thermoplastic material coating on top of a transparent conducting layer, which in turn is adhered to a base like that of a standard motion picture film base. The surface of the thermoplastic is charged by an electron beam in a pattern corresponding to the ripple pattern which will form the image. The film is heated as it moves past an RF heating element causing the thermoplastic to melt. The attraction of the charges to the conducting base depresses the thermoplastic's surface to form the ripples which are frozen into place by cooling the film.

The film can be used again by melting it to a high enough temperature for the charges to leak away and the surface will smooth back to its original state. The process must be carried out in vacuum since an electron beam is involved. The ripple pattern is illuminated by a series of line sources imaged on a set of bars in front of the projection lens. Ripples on the surface of the film scatter light through the bar system. Light passing through the projection lens images the ripple as a white spot on the viewing screen. The film generally used is the size of 16-mm photographic film. The film runs at 10 in./sec (25.4 cm/sec) and large size recordings use half the film width; the images are about 5 mm wide.

Commercial thermoplastic systems are available at present with resolution of 40 optical lines per mm. Others have been demonstrated in the laboratory with resolution near 600 optical lines per mm. Theoretically the resolution obtainable is near 2500 optical lines per mm. The General Electric Company has already produced a commercial thermoplastic system.

3-7.3.7 Photography

Photography deals with the permanent recording of an object by a light-sensitive material so that it may be examined at a later time. The light-sensitive material employed is a support substance coated with a colloidal suspension of silver halide microcrystals immersed in a gelatinous medium. This medium is a gelatin emulsion, and its thickness ranges from 0.0002 to 0.0015 in. Support is usually provided by glass, paper, or film.

When extreme dimensional accuracy in the photograph is required, the glass support material is chosen. This material has the disadvantages of fragility, excessive weight, and inflexibility. The glass plates range in thickness from 0.060 to 3/16 in.

Paper is used as a support material for photographic prints and applications requiring an opaque and highly reflective base. It is strong, economical, and resistant to photographic solutions. Paper used in making prints comes in three weights referred to as lightweight (less than 0.006 in. thickness), single-weight (0.006 to 0.010 in. thickness), and double-weight (greater than 0.010 in. thickness).

By far the most widely used emulsion base is plastic film in sheet or roll form. This substance provides a tough, strong, transparent base with good dimensional stability and relative inertness toward common solvents and chemicals. Sheet film is found in a variety of sizes from 2¼ by 3¼ in. up to 30 by 40 in., with 4 by 5 in. being the most popular. Roll film is available in 70-, 35-,

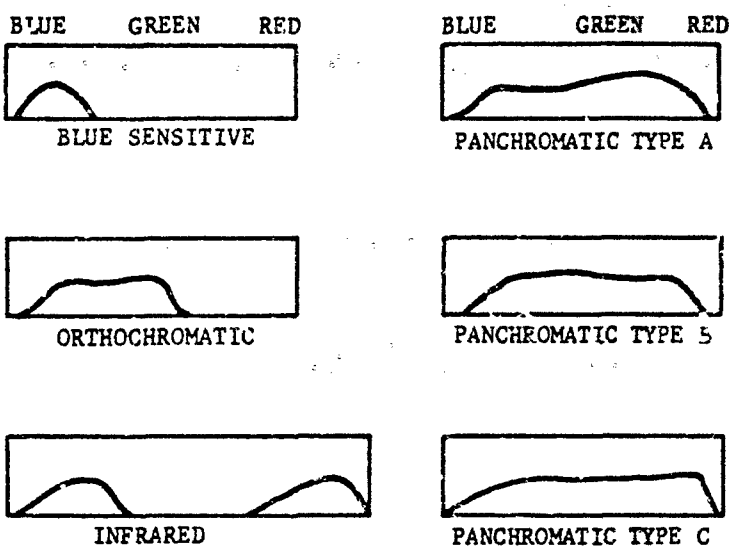


FIGURE 3-128. Spectral Sensitivity of Various Types of Film

and 16-mm widths. Most 8-mm cameras use 16-mm film and expose it one-half width at a time.

Photographic emulsions in wide use are divided into six general divisions:

1. Ordinary blue sensitive
2. Orthochromatic
3. Panchromatic type A
4. Panchromatic type B
5. Panchromatic type C
6. Infrared

The spectral regions of sensitivity of these emulsions are shown in Fig. 3-128. These emulsions will reproduce the object in black, white, and intermediate shades of gray in proportion to the local area brightness on the object. In addition to these, color emulsions are also available which reproduce the object in the colors which it appeared to the eye. The most widely used color films are based on the tripack subtractive processes produced by Eastman Kodak Company and Ansco. Trade names are Anscochrome, Kodacolor, Ektachrome, and Kodachrome.

To be useful as a display device the photographic reproduction must provide an accurate representation of form, and a faithful rendition of tone. The dimensions of the image are determined by the optics employed in the recording process. For a discussion of this problem refer to Chapter 2 in this handbook on optical design and to Ref. 90. The tonal qualities of the image result from exposure and are related to the photo-sensitivity of the emulsion. Perfect reproduction may be defined as one in which the amount of light transmitted or reflected by its various areas is directly proportional to the brightness of the corresponding areas of the subject. If a negative reproduction is used, the definition is one in which the amount of light transmitted by the various areas of the negative image is inversely proportional to the brightness of the corresponding areas of the subject. To understand more about the quality of the photographic image we must consider the sensitometric characteristics of emulsions.

When an exposed film is developed, the silver halide crystals affected by the exposure are converted into particles of metallic silver, while those not affected are removed in the fixing bath. The residual silver grains compose the photographic image by providing a certain opacity to light. This is determined by the relation [91]

$$O = \frac{I}{I'} \tag{3-185}$$

where O is the opacity, I is the intensity of light incident upon the emulsion, and I' is the intensity of light transmitted by it. It is customary to speak of density rather than opacity. The density D is given as[91]

$$D = \log_{10} O = \log_{10}\left(\frac{I}{I'}\right) \tag{3-186}$$

An increase in exposure of the film increases the number of residual silver grains formed, and hence increases the density. The relationship between density and exposure for a particular film is given by the "Characteristic Curve" of the film (also called the H and D curve after Hurter and Driffield). A typical curve is shown in Fig. 3-129 where the density is plotted versus the logarithm (to the base 10) of the exposure[91]. The lower segment (AB) of the curve, the "toe", is the region of underexposure. The middle section (BC) is the linear part of the curve. The uppermost segment (CD), the "shoulder", is the area of overexposure.

The BC portion of the curve indicates a linear relationship between optical density and log exposure. The range of exposure corresponding to this segment of the curve represents the region of correct exposure. As long as the darkest areas of the subject do not fall below B and the brightest areas do not reach beyond C, all gradations of tone within the subject will be faithfully reproduced as proportional density variations in the negative. The slope of this straight-line portion of the curve is called the gamma (γ). If $\gamma = 1$, then opacity is directly proportional to exposure. Tonal gradation in the negative is determined both by the shape of the characteristic curve and the level of exposure used to expose the negative. If a level of exposure that places the entire brightness range of the subject on either the toe or shoulder portions of the characteristic curve is chosen, the corresponding density gradations within the negative will be compressed, resulting in either under- or over-exposure. Thus, a correctly exposed negative results when the tonal range of the subject falls on the linear region of the curve.

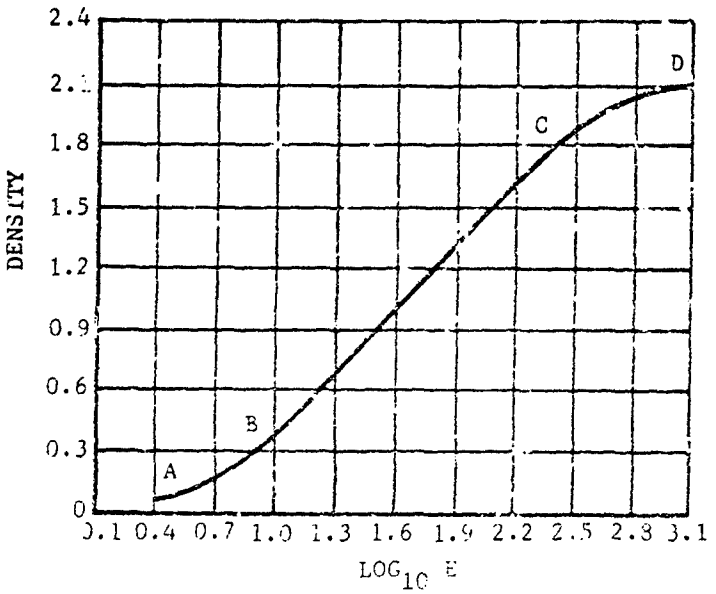


FIGURE 3-129. Typical Characteristic Curve of a Photographic Emulsion

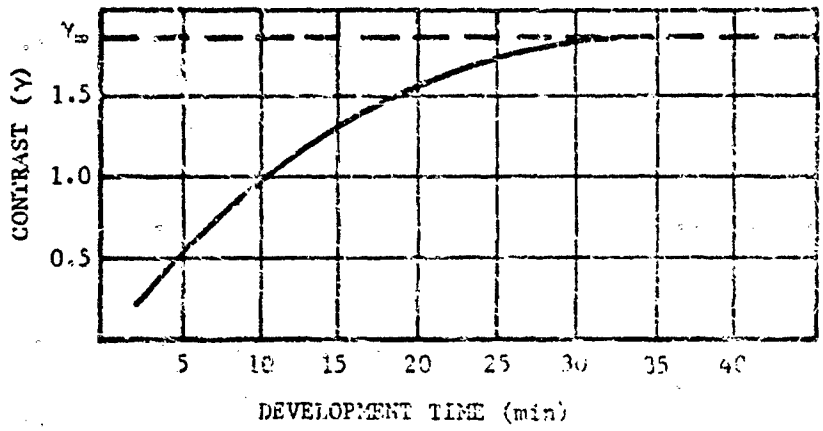


FIGURE 3-130. Typical Variation of Gamma With Development Time

Contrast in a negative is the difference in densities corresponding to the lighter and darker portions of the negative. Contrast in the reproduction is determined by both the subject contrast and the γ of the characteristic curve. If the tonal range of the negative is restricted to the linear region of the curve, the contrast of the subject and the reproduction will be the same for $\gamma = 1$. When the gamma exceeds unity, the negative contrast will be greater than the subject contrast, and vice versa.

The gamma of the curve is found to increase directly as a function of development time as shown by the plot in Fig. 3-130. A low gamma results from underdevelopment and a high gamma from overdevelopment. Extended development results in the buildup of an overall fog that reduces contrast and limits the maximum contrast obtainable to a value defined as gamma infinity (γ_∞).

With the preceding discussion of the sensitometry of emulsions understood, we may discuss the physical characteristics of the developed photographic image. The important factors to be considered are graininess, turbidity, sharpness, resolving power, and range of contrast.

When the developed image is examined under low-power magnification, it exhibits granularity. Although the silver grains are microscopic in size, the inhomogeneous appearance or graininess is noticeable at low magnification due to the clumping together of individual grains. As a general rule, graininess increases with increasing sensitivity of the emulsion, with greater gamma, and with the density of the image. It is most evident within large areas of uniform exposure at density levels of about 0.8. Graininess limits the degree of enlargement which can be attained without disturbing the uniform appearance of the image and preventing the visual recognition of small, unsharp, low contrast detail. This effect can be reduced by using slower fine-grain emulsions in combination with low-energy developers containing silver halide solvents.

When a knife edge is held in contact with an emulsion and the emulsion is exposed to light, the developed image does not end precisely at the knife's edge but extends into the shaded area. Light reaches the shielded silver halide grains by the physical processes of diffraction, reflection, and refraction in the emulsion. This property of the emulsion is called its turbidity and is to be minimized where possible. This quantity may be expressed by the relation[92]

$$Turbidity = \frac{dx}{d(\log_{10} E)} \qquad (3\text{-}187)$$

where x is the distance into the shaded area beyond the knife edge, and E is exposure.

The sharpness of an image is defined by the expression[93]

$$\text{Sharpness} = \frac{dD}{dx} \qquad (3\text{-}188)$$

where dD/dx is the rate of change of image density D with distance x into the shaded region behind a knife edge held in contact with the emulsion surface during exposure. A rough product between the sharpness and the turbidity will equal the gamma of the film[94]

$$\frac{dD}{dx}\left[\frac{dx}{d(\log_{10} E)}\right] = \gamma \qquad (3\text{-}189)$$

The ability of an emulsion to record fine detail is called the resolving power or resolution of the film. The resolution of a typical high-speed emulsion is in the order of 40 to 50 lines/mm while positive types may resolve 80 to 100 lines/mm. Special emulsions can resolve up to 2000 lines/mm. Resolution increases with sharpness; inversely with graininess and turbidity; inversely with the wavelength of the exposing light; directly with contrast of both the emulsion and the object; and, finally, passes through a maximum as exposure increases.

The range of contrasts attainable by a film is given by its latitude which is the length of the straight line portion of the characteristic curve. If the difference between B and C in Fig. 3-129 on the $\log_{10} E$ scale is 2, then from Eq. 3-186 an intensity range of 100:1 in the subject may be represented by the film in the reproduction. In practice, contrast ranges of 150:1 or even 200:1 are recordable.

3-7.3.8 Holography*

A hologram is a recording of the diffraction or scattering pattern of an object. When light reflects from, or is scattered by an object, information as to the location, size, shape, and texture of the object is contained in the complex waves reflected from the object.

By means of conventional optics, such as a lens, it is possible to collect a portion of these waves and refract them in such a way that the spherical waves diverging from each point of the object are made into converging spherical waves. The collection of converging spherical waves then come to a focus, forming an image of the object. By means of holography, it is possible to record, in a permanent form, the complex waves themselves. At any later time, it is then possible to pass light through the record, or hologram, and obtain waves exactly like those that came from the original object (limited, of course, by noise and nonlinearities in the process). If one examines the waves by eye, he "sees" the original object in its original position with respect to the hologram plane, in three dimensions, with all spatial relationships preserved. This is to be expected, if the original waves coming from the object have been reproduced in detail. The process of recreating an image from a hologram is called "reconstruction".

Information is contained in both the amplitude and phase functions of a complex optical wave. Recording the amplitude function alone is not sufficient to allow reconstruction of the wave. It is necessary to record both amplitude and phase. The most widely used medium for recording holograms is photographic film, which is known not to be sensitive to phase. It responds only to the intensity, or square, of the amplitude. Phase information may be recorded, however, if a reference beam is superimposed on the wave to be recorded, so that the relative phase between the two beams gives rise to fringe patterns. The amplitude and phase distribution of the wave to be recorded is then represented by the amplitude and position of the fringes.

The way in which the amplitude and phase of a complex wave can be recorded by a detector sensitive only to intensity, and the manner in which the wave can be reconstructed from such a record will now be outlined. Let the electric field vector of the wave from the object in the plane of the plate be represented by a complex variable $E_o(x,y)$. Let the reference beam be represented by the complex function $E_r(x,y)$. When these two waves are superimposed on the photographic plate, the resultant electric field is the sum of the two functions, $E_o + E_r$, where the dependence on x and y is implicit. The plate, however, responds to the intensity of the resultant wave, which is given (except for a constant factor) by

* The definition of holography includes sound waves and the whole electromagnetic spectrum, including microwaves, X rays, ultraviolet, and infrared. From the practical view of display, however, the visible region is of primary interest.

$$(E_s + E_r)(E_s + E_r)^* = E_sE_s^* + E_sE_r^* + E_rE_r^* + E_sE_r^* + E_s^*E_r \qquad (3\text{-}190)$$

where * denotes the complex conjugate.

The film is exposed in this way and then processed. Assuming, for simplicity, that the amplitude transmittance of the developed film is a linear function of exposure energy density, the transmittance τ as a function of x and y is

$$\tau = A - B(E_sE_s^* + E_rE_r^* + E_sE_r^* + E_s^*E_r) \qquad (3\text{-}191)$$

where A and B are constants for a given film and development.

Now suppose the processed plate is illuminated with a beam having the same amplitude and phase distribution as the reference beam. The distribution of the resulting wave after it passes through the plate (hologram) is

$$E_r\tau = E_rA - B[E_rE_sE_s^* + E_r(E_r^*E_r) + E_r(E_r^*E_s) + E_s^*E_rE_r] \qquad (3\text{-}192)$$

If the reference beam has nearly constant amplitude across the plate, the product $E_r^*E_r$ may be considered a scalar constant. With this in mind, the terms of Eq. 3-192 may be examined. The first and third terms represent a constant times the reference beam, and are equivalent to a zero-order beam, if the hologram is considered as a diffraction grating. The second term represents noise due to interaction of the signal with itself. The fourth term is the original signal wave multiplied by a constant; this term gives the true image. The last term contains the conjugate of the original signal wave, and gives rise to a false image called the conjugate image. The goal of holography is to control the process of making holograms and reconstructing the images so that linearity is maintained, and the light beams corresponding to various terms of Eq. 3-192 are spatially separated so that they do not overlap and degrade the desired image.

A typical optical arrangement for making holograms is illustrated in Fig. 3-131. Coherent, monochromatic light is divided into two beams by a beam splitter. One of the beams is spread by a lens so that it illuminates the object. Light reflected and scattered from the object falls on the photographic plate. At the same time, the

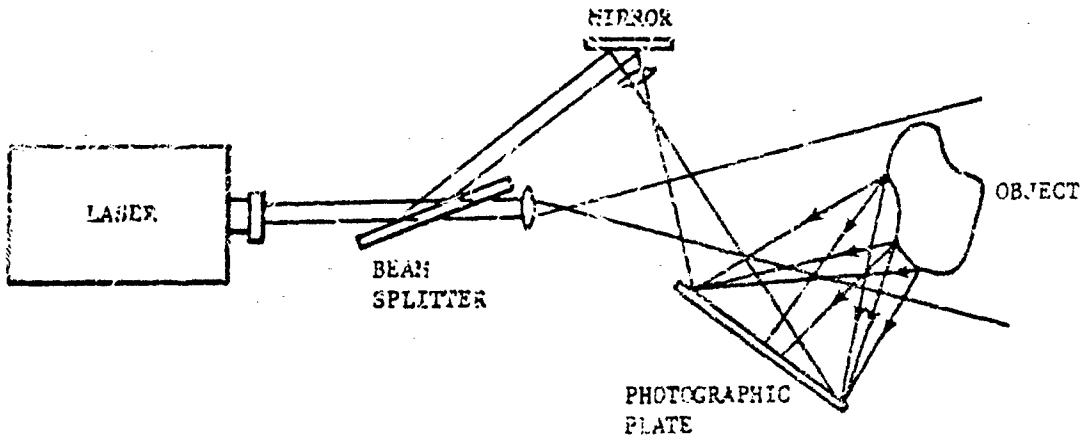


FIGURE 3-131. Typical Arrangement for Making Holograms

other beam is also diverged and made to fall on the plate, providing the reference beam. It is essential for the plate to be exposed to both beams simultaneously, so that the plate can record the interference fringes.

If the plate is processed and replaced in the optical system, with the object removed, the light diffracted by the hologram into the viewer's eye placed at the position shown in Fig. 3-132, seems to come from a three-dimensional object exactly like the original object. The reconstruction of the size and shape of the object is so exact that if the original object is not removed but left in place, and illuminated as before, optical interference can be observed between the reconstructed wave and the waves scattered from the object. Small movements or distortion of the object can be measured to fractions of a wavelength by counting the fringes due to this interference.

It should be noted that the viewer does not look along the direction of the illuminating beam. The light in that direction is the zero-order beam, which is shown by Eq. 3-192 to contain no information about the object. *(One should also avoid looking into a laser beam for safety reasons.)* The object information is contained in the first-order diffraction from the hologram. There are two first-order diffraction beams, one on each side of the zero-order beam. The beam represented by the fourth term of Eq. 3-192 forms a true virtual image that can be viewed from the position indicated in Fig. 3-132. The other first-order diffraction beam, described by the fifth term of the equation, forms a real image on the other side of the zero-order beam, but the image is "pseudoscopic", i.e., it is reversed, or turned inside out in a peculiar way because it is formed by the complex conjugate of the object wave.

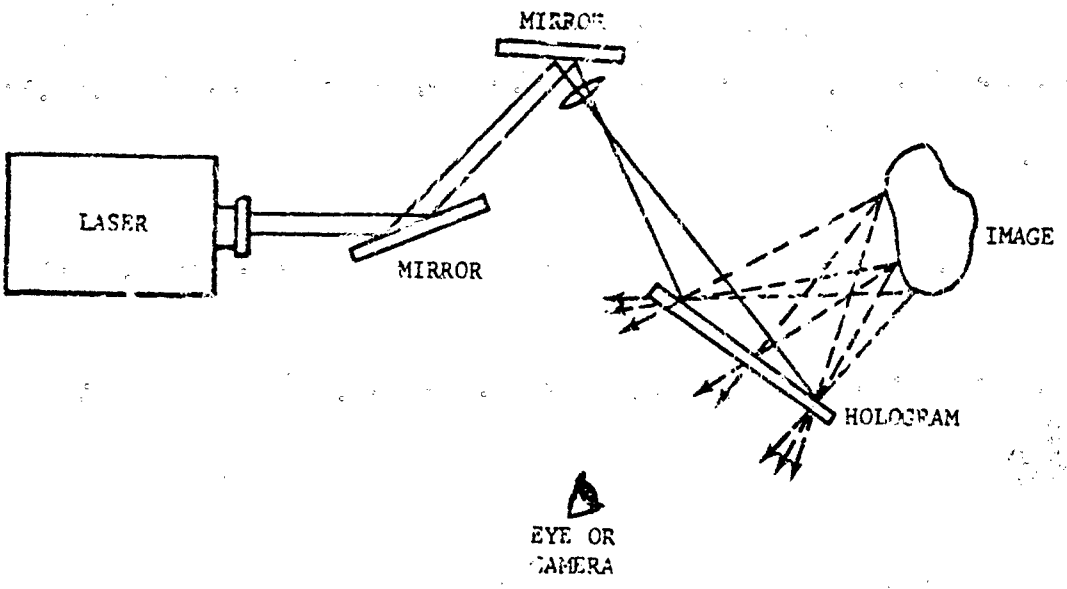


FIGURE 3-132. Reconstruction of a Hologram

The spatial detail in the fringes that must be recorded in a hologram is extremely fine, approaching the resolution limit of the finest-grained spectroscopic plates. Consequently, small movements of the plate or the object during the exposure can completely wash out the fringe pattern. In such cases, one does not get a blurred image upon reconstruction, one gets no image at all. The requirement for extreme rigidity in a holographic recording may by relieved considerably by the use of a pulsed laser for illumination. Even the shortest available laser pulses, however, do not allow reconstruction of rapidly moving objects such as bullets.*

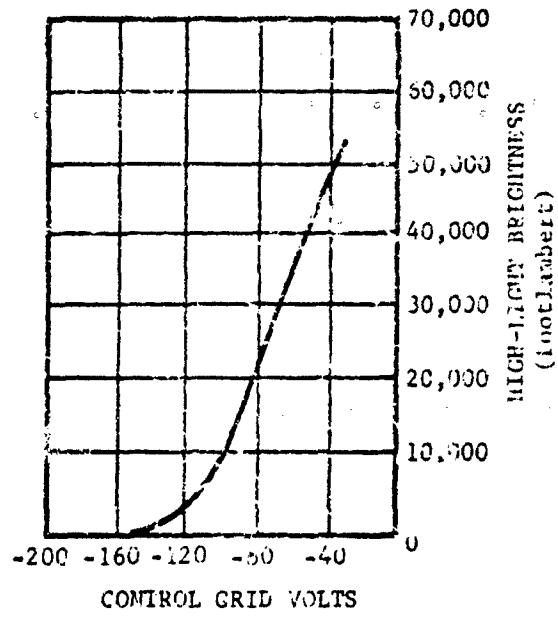
Another restriction on the present usefulness of holography is the requirement for coherence of the illuminating light throughout the volume occupied by the object. (Refer to par. 3-8 for a discussion of coherence.) This requirement may be relieved to some extent by tricks, such as the use of multiple beam splitters to illuminate long objects, or careful matching of path lengths but, in general, the depth of field of a hologram is limited to the coherence length of the laser. For typical He-Ne lasers, used for most holography to date, the length is about 20 cm. There is little doubt that improved coherence of lasers will greatly broaden the applications of holography.

3-7.3.8 Special Purpose Displays

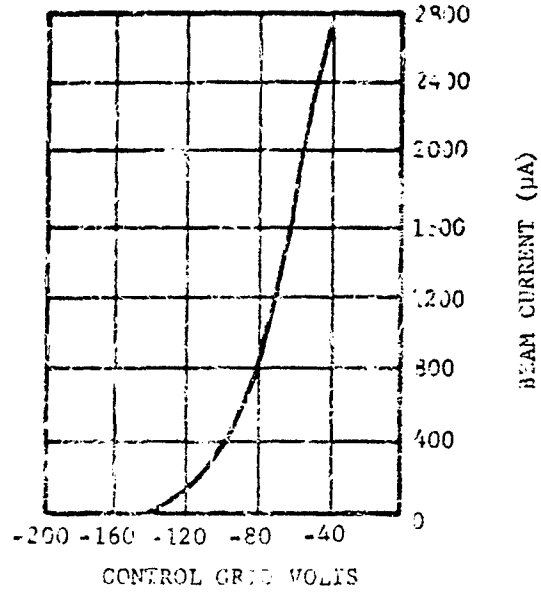
3-7.3.8.1 Projection CRT

The projection tube is a special version of a CRT and operates on the same principles, but at greater values of beam current and voltage than the standard, to produce higher output brightness. The tube operates at anode voltages of 40 kV, beam currents of 500 μA with brightness of 20,000 footlamberts under typical conditions. Brightness levels approaching 40,000 footlamberts are possible with some of these tubes. Magnetic deflection and focus are employed because of the high beam currents. Maximum deflections of the beam are 38 deg.

* Silhouette holograms have been made of bullets in flight. This fact has caused some confusion since bullets travel many wavelengths during an exposure. Light reflected from the moving object washes out completely, and does not contribute to the image. However, in these pictures, only the diffuse background actually reconstructs and the hologram of the silhouette of a bullet appears in the picture.



(A) BRIGHTNESS AS A FUNCTION OF CONTROL GRID VOLTAGE



(B) BEAM CURRENT AS A FUNCTION OF CONTROL GRID VOLTAGE

FIGURE 3-133. Projection CRT Characteristics

Typical projection CRT characteristic curves are presented in Fig. 3-133. A special glass is used for the faceplate to prevent discoloration under such intense electron-beam bombardment. The faceplate is optically polished to have a curvature matching the optical system and is usually about 7 in. in diameter. Assuming a 4:3 aspect ratio, the light is concentrated into an area about 5 in. × 4 in. If the tube has 30,000 footlamberts at its faceplate, then projecting it on a large screen, say 20 ft × 15 ft, gives a screen brightness of 10 footlamberts. This is considered to be satisfactory; 5 footlamberts is the minimum acceptable.

For a projected display, the illumination on the screen will be given by

$$E = \frac{\pi B \epsilon_a}{4(f/no.)^2 (1+m)^2}, \text{ footlambert} \quad (3\text{-}193)$$

where

E = illuminance of screen, footlambert

B = available source brightness, footlambert

$f/no.$ = aperture (focal length/diameter)

m = magnification

ϵ_a = transmission of optical system

The resolution of these tubes is 600 TV lines. A special shielding must be used around the tube since X-rays are produced by the high voltages in the tube. In addition, a heat exchanger (either gaseous or liquid) is required to maintain the proper operating temperature.

3-7.3.9.2 Multi-gun CRT

CRT's have been constructed with several complete but separate electron guns contained within a single tube. Each gun operates on the same principle as a single-gun CRT, having independent control over the intensity and deflection of its own electron beam. If the several beams in one tube are required to operate independently, then independent electrostatic deflection and focus are necessary. The main drawback to these tubes is a lack of precise registration of the beams; it is difficult to maintain a registration of two beams closer than about 1 percent of the tube diameter over the total useful screen area and there is crosstalk between the several deflection systems to de-

TABLE 3-17. MULTI-GUN TUBES

Faceplate dimension, in.	5-1/4	5-1/4	5-1/2	5-1/2 × 5-1/2	11-3/4 × 10-1/2
Length, in.	18-1/4	18-1/2	18-1/2	18-1/4	21-3/4
Number of guns	2	3	4 (offset)	5 (offset)	6 (with independent displays)
Faceplate radius of curvature	flat	flat	flat	40 in.	49 min
Post accel. Volt. dc	4000	12,000	8000	4000	...
Accel. Volt. dc	2000	4000	2500	2000	4000
Sens. Volt., esu	400 to 684	900 to 1300	550 to 900	450 to 650	900 to 1200
Neg. grid cutoff (Volt. dc)	45 to 75	60 to 98	55 to 100	50 to 90	70 to 120
Deflection D_1 D_2	54 to 66	120 to 160	30 to 50	68 to 84	80 to 100
(V/in.) D_3 D_4, esu	43 to 53	104 to 156	50 to 70	27 to 37	37 to 50
Useful Scan D_1 D_2 (in.)	4.0	4.25	3.5	4.5	8.5
D_3 D_4	4.0	4.25	0.84	1.5	2.5
Line Width	0.026 in. at 25μA dc	0.015 in. at 50μA dc	0.012 in. at 2μA dc	0.026 in. at 25μA dc	0.022 in. at 25μA dc

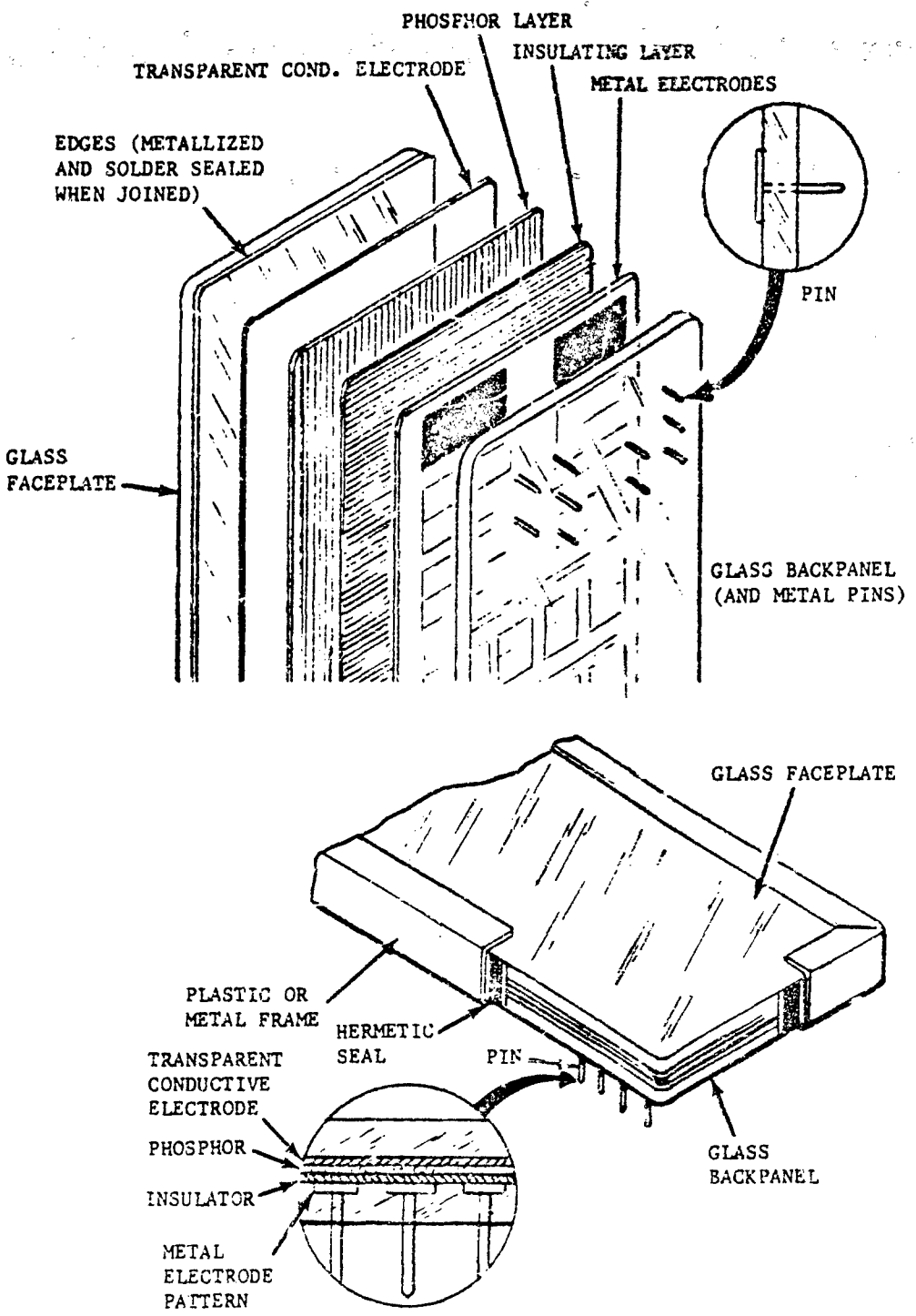


FIGURE 3-134. Exploded View of All-glass EL Readout Panel and Sectional View of Metal-glass Unit

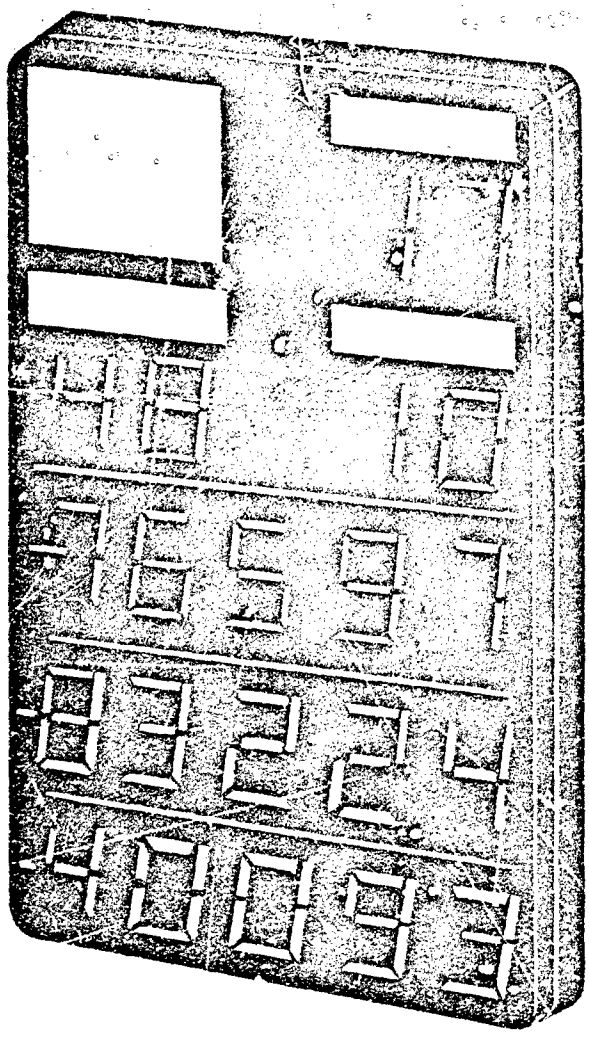


FIGURE 3-135. Example of Electroluminescent (EL) Panel

grade image quality. In addition to the three-gun color tubes in prominent use, there are CRT's available with up to ten guns. The most commonly used are from 2- to 6-gun tubes for which some typical characteristics are listed in Table 3-17. These tubes are useful for display of information received from several parallel channels simultaneously.

3-7.3.9.3 Electroluminescent (EL) Panels

An EL panel consists of a phosphor layer and an insulating layer sandwiched between a pair of electrodes. This is shown in Figs. 3-134 and 3-135 where one electrode is transparent, the other metallic, and the entire assembly is hermetically sealed between a pair of flat glass plates for protection. By applying an alternating potential to the electrodes, an oscillating electric field is set up in the phosphor dielectric layer which causes electrons to become excited into higher energy states from which they subsequently decay with the emission of visible radiation.

The most frequently employed phosphor material is ZnS with Cu or Al activator impurities added. The behavior of the EL layer will depend

very strongly upon the method of preparation; however, an empirical relation generally referred to in calculations of brightness B is given by

$$B = A_o \exp\left(-\frac{b}{V^{1/2}}\right) \qquad (3\text{-}194)$$

where A_o and b are constants, and V is the rms applied voltage. This relation attempts to represent the empirical curves for brightness as a function of voltage, as shown in Fig. 3-136. Brightness of the phosphor also varies with the frequency f at which the excitation is applied, as shown in Fig. 3-136. EL panels very rarely suffer catastrophic failure, but gradually diminish in terms of output brightness. The time required for a panel to diminish to one-half of its initial brightness level is directly proportional to the number of cycles it has been put through and independent of the frequency at which it operates. Thus, the higher the frequency of applied voltage the more cycles the panel will pass through in a given interval of time and, hence, the sooner it will reach its half-brightness point.

Commercially available EL panels provide brightness levels of up to 50 footlamberts at an applied potential of 250 V, 400 Hz. The applied voltages vary from 100 to 600 V ac, at rates of from 100 to 4000 Hz. Typical EL phosphor efficiencies are in the range of 0.5 to 2.0 lumen/w. Light is observed from a panel within 0.2 μsec after the electric field is applied, but full brightness is not attained until after about 10 cycles of operation. EL panels offer the widest viewing angle of any display device, almost 180 deg. They may be used to display numbers, letters, analog data, quantitative comparisons, and even pictorial data, although they do not perform the latter task as well as desired. Resolution attainable is about 50 lines/in. at this time. Contrast ratios are in the order of 20:1 for most available devices. It is possible to increase the contrast by changing the transmission characteristics of the faceplate resulting in a net decrease in panel brightness. By using a 50 percent transmission glass faceplate, for example, the contrast is increased significantly, but only one-half as much light output is available as

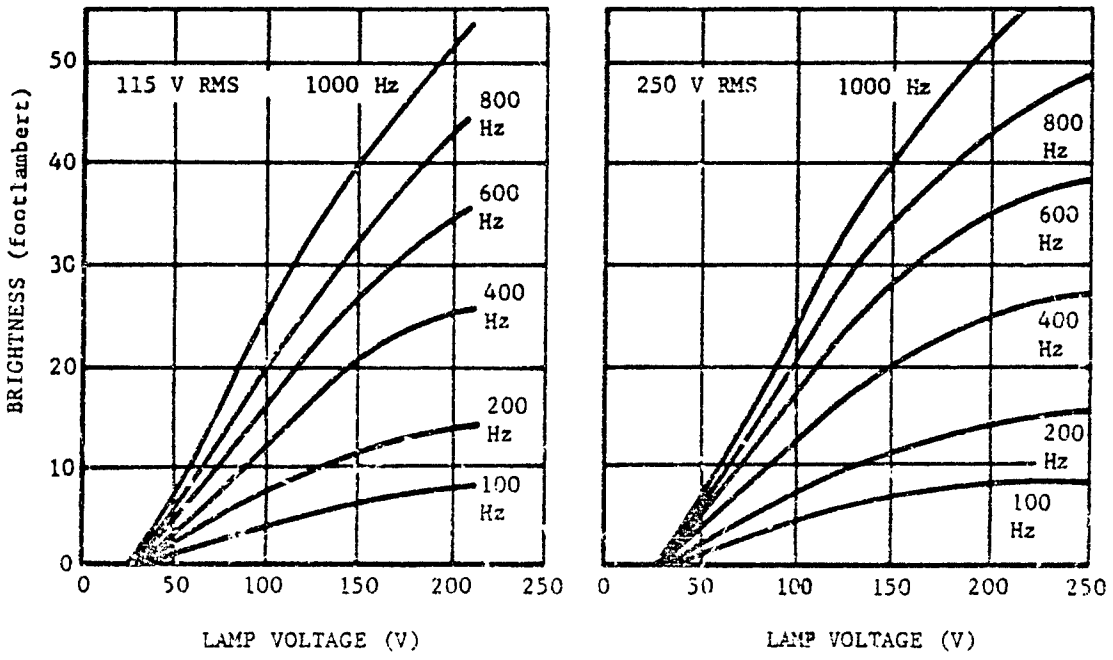


FIGURE 3-136. Brightness vs Voltage With 50% Transmission Glass for Both Metal-glass and All-glass EL Readout Panels at 115 and 250 V rms, Respectively

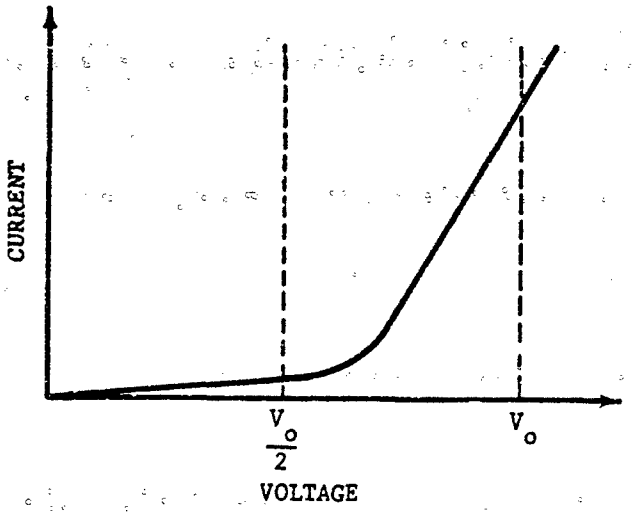


FIGURE 3-137. Voltage-current Curve for Nonlinear Resistor

with a 100 percent transmission faceplate.

EL panels may be considered in two classes: (1) monostable and (2) bistable. The monostable must be continually refreshed to maintain a presentation, while the bistable has a memory and need be addressed only once. In a typical panel the front set of n parallel transparent electrodes is aligned in the X-direction, while the rear set of n parallel electrodes is aligned with the Y-axis. The n^2 intersections of these electrodes may then be addressed in an X, Y fashion. By placing one-half the required rms excitation voltage on a given X-electrode, say X_i, and the other half on the Y-electrode, Y_j, then the position X_i, Y_j on the EL panel will luminesce.

The two prominent methods for constructing monostable EL panels employ: (1) nonlinear resistors (NLR) and (2) ferro-electric material. The NLR method consists of an EL phosphor layer in contact with an NLR layer sandwiched between two orthogonal sets of electrodes. The NLR layer provides electrical isolation between the EL layer and one set of electrodes. As the voltage applied to the electrodes is increased, the impedance of the NLR layer decreases (Fig. 3-137) until reaching a point V_o at which the EL layer has enough applied voltage across it to luminesce. In the second type of monostable panel, ferro-electric material replaces the NLR layer to provide both an electrical isolation and a partial storage capability. The cross-grid electrode arrangement is applied here, too. This type of panel suffers less from flicker effects and offers a gray-scale sufficient for pictorial presentations.

The bistable configuration is obtained with any of three techniques: (1) silicon control rectifiers (SCR); (2) cadmium selenide (CdSe); or (3) photoconductor (PC) elements. The SCR serves as a control element for the EL cell. A gate voltage is applied to the SCR element to open it, applying the excitation voltage to the EL layer. When the gate pulse is removed the SCR effectively blocks the voltage and turns the EL cell off. The CdSe is a polycrystalline material which possesses an abrupt discontinuity and hysteresis in voltage versus current (Fig. 3-138). The current rises linearly with applied voltage in the CdSe element until reaching the discontinuous value, at which point the impedance decreases abruptly by several orders of magnitude allowing the passage of current sufficient to produce luminescense in the EL layer. By virtue of its hysteresis property, the CdSe remains in this on-condition until the applied voltage is reduced sufficiently below cut-off value.

The PC technique uses a photoconductive technique in place of either the SCR of CdSe layer in contact with the EL layer. When the EL layer luminesces, some of the light produced will be scattered into the PC layer rendering it

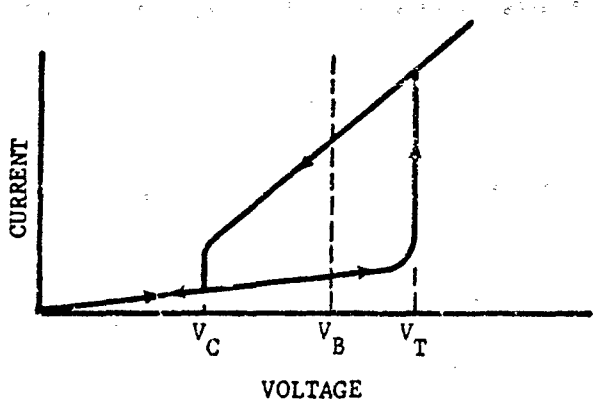


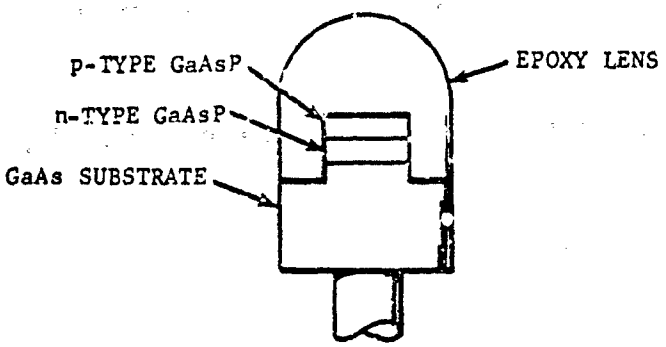
FIGURE 3-138. CdSe Hysteresis Curve

conductive at those points. This allows the applied voltage to pass easily through the PC layer at positions where there is already luminescence underway and thereby sustain it. All of the preceding bistable techniques thus possess a memory and require no updating signals.

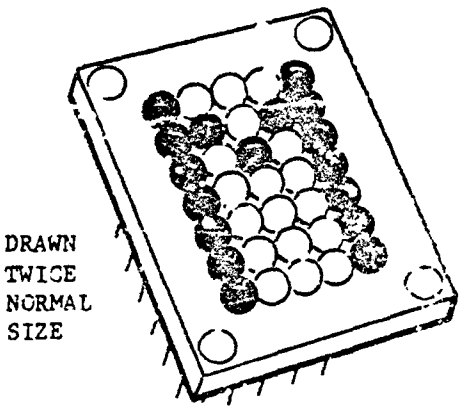
3-7.3.9.4 Photo-emitter Diodes

Discrete solid-state visible light emitters are available for use individually as indicator or event lights and in array format for presentation of alphanumeric information. These are gallium arsenide phosphide diodes of the diffused mesa structure (Fig, 3-139) which emit noncoherent light in the wavelength range from 500 to 700 millimicrons (mμ). With a forward bias voltage across the p-n junction in the mesa, electrons and holes combine with the emission of radiation. As the phophide concentration is increased, the wavelength of peak radiation emitted decreases, as does the efficiency of the device.

The radiation emitted peaks at about 665 millimicrons with a narrow spectral distribution as shown in Fig. 3-140. Typically, a forward bias of 1.6 V will produce an output brightness proportional to the current, as shown by the characteristic curves in Fig. 3-141. This means that a brightness of 300 footlamberts is obtained at 50 mA and 80 mw. The total radiated power at this level of operation is about 25 μw. The device behaves like a standard p-n junction diode, in accordance with its characteristic curve shown in Fig. 3-142. The spatial distribution of the emitted radiation is shown in Fig. 3-143. These photo-emitter diodes may be operated in either steady-state or pulsed mode as desired. The turn-on and turn-off times are 10 nanosec each. Properly prepared diodes have a very small probability of failure and appear to have lifetimes equal to, or greater than, most other small-light sources.



(A) INDIVIDUAL GaAsP Diode



(B) GaAsP EMITTER DIODE ARRAY DISPLAYING LETTER M

FIGURE 3-139. Illuminating Panel Consisting of GaAs Photo-emitter Diode Array

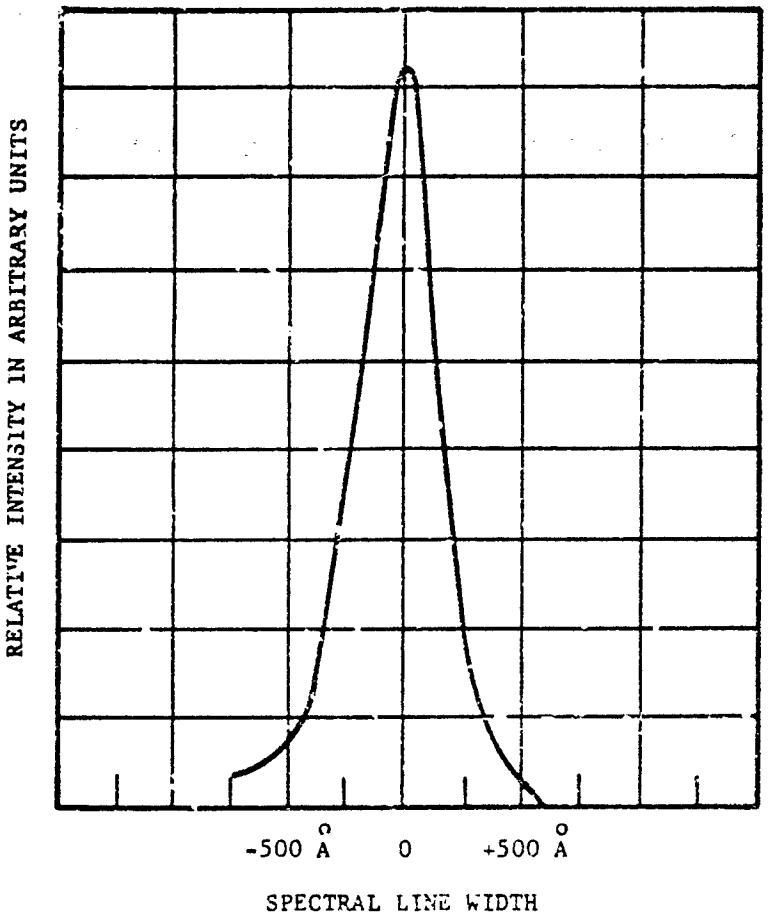


FIGURE 3-14C Spectral Line Width for GaAsP Emitter Diode

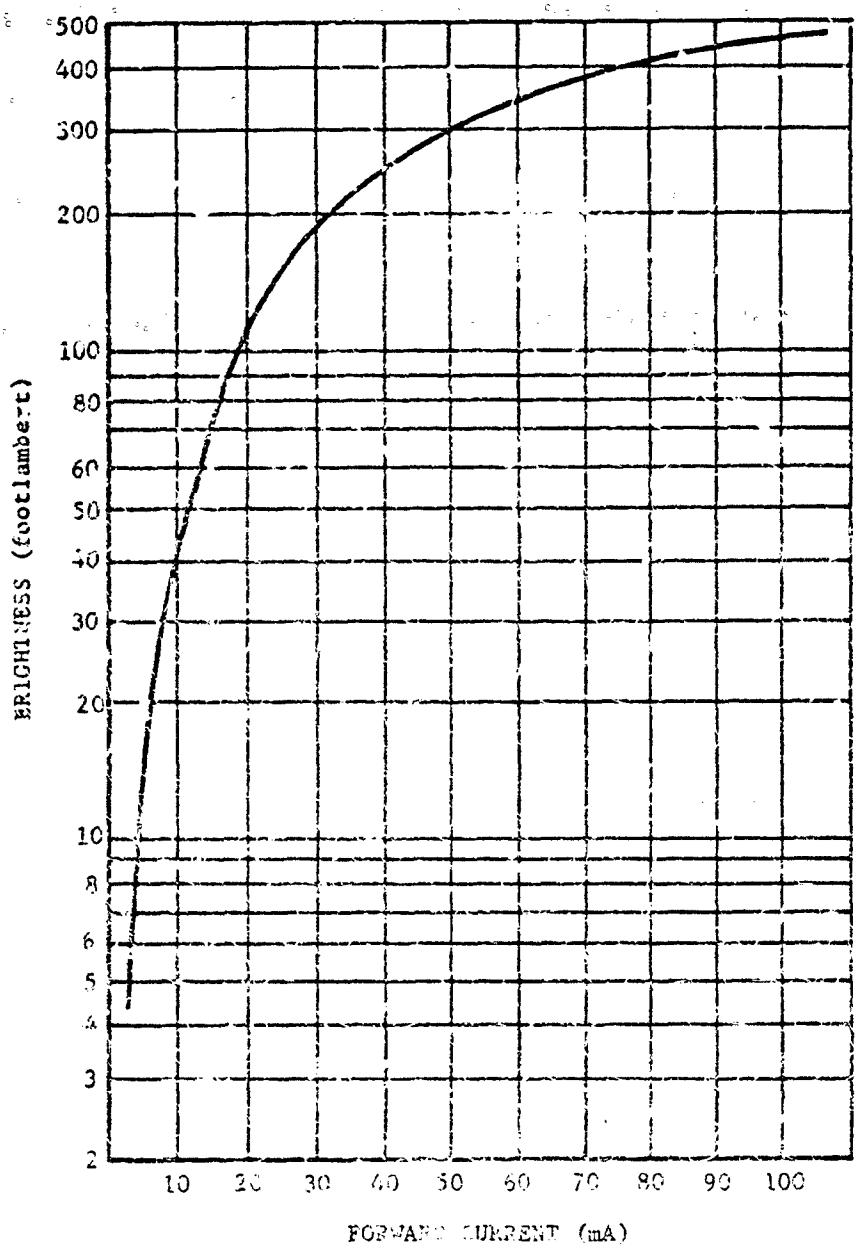


FIGURE 3-141. Characteristic Curve for Typical GaAsP Emitter Diode

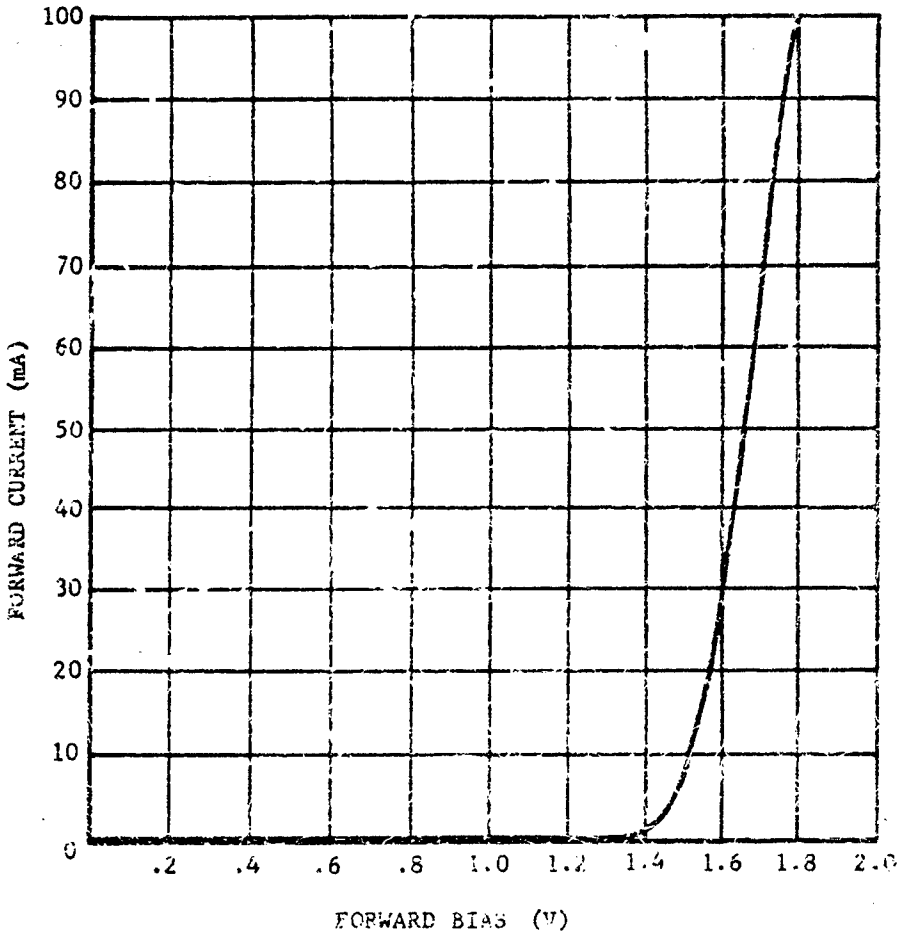


FIGURE 3-142 Forward Current vs Forward Bias Voltage of a Typical GaAsP Emitter Diode

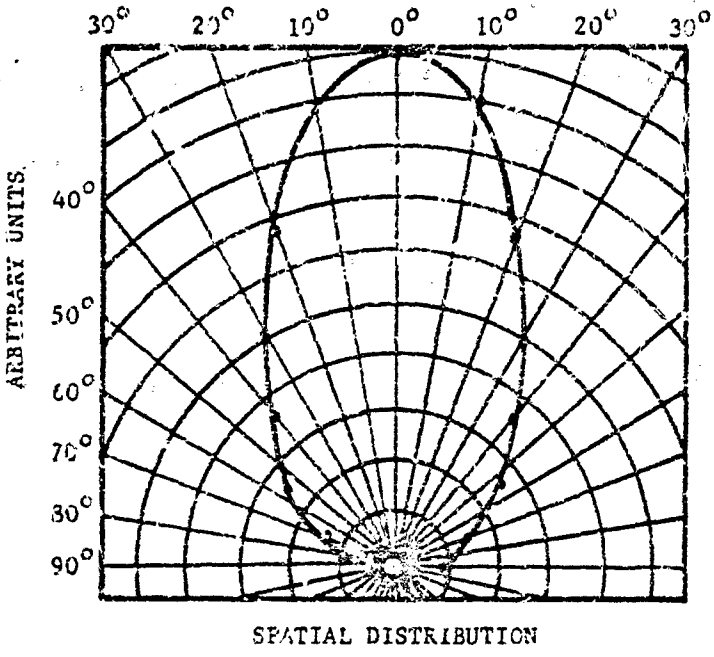


FIGURE 3-143. Spatial Distribution of Radiation Emitted by GaAsP Diode

3-7.4 DISPLAY EQUIPMENT DESIGN PARAMETERS

The design of a display should start with a complete analysis of the system in which it is to be integrated to determine the components, both man and machine, and the conditions under which they operate. In doing this, the following steps can be taken:

1. Determine the internal and external environments of the system
2. Itemize all sensing devices required for system inputs
3. List all equipment employed in processing sensor data
4. Specify human requirements (i.e., numbers and kinds of functions)
5. Specify number and kinds of displays and controls required for system output

After obtaining the information given above, it is possible to set intelligent requirements upon the display by assigning a measure of importance to each of the parameters in the following list.

Information Recognition. Decide upon the amount and density of data to be presented concerning the relative importance of the data and the observer's ability to recognize it (see Chapter 5). Data will include a consideration of resolution, contrast, brightness, symbology and display time, and rate.

Number of Inputs. The number of inputs to be presented on one display unit.

Number of Displays. The number of separate display units required.

Real Time. The time delay allowed between sensor response and data display to observer.

Versatility. The different types of data to be presented in various formats (e.g., alphanumerics, symbols, maps, pictures, etc.) on one display either simultaneously or sequentially.

Screen Size. The actual size of the display surface.

Display Size. The physical size of the display package.

Power Requirements. Be certain the power requirements of the display are compatible with the overall system design.

Image Quality. The accuracy with which the display must depict what is "seen" by the detectors, including the resolution and distortion of images.

Brightness and Contrast. The brightness required will be determined by the environment to a large extent (daylight or nighttime use) and the contrast attainable is a function of this brightness requirement.

Repetition Rate. How often the information is refreshed is determined by cost of refreshing and flicker tolerance of the observer.

Update Rate. The number of times new information is to be substituted on the display.

Reliability. It is always advisable to guard against catastrophic failure by employing a display which deteriorates gradually.

3-8 TESTING IR AND ASSOCIATED EQUIPMENT

Testing of IR equipment must be tailored to prove the performance of each specific system and entails testing electronic and mechanical as well as optical equipment. Standard symbols used in this discussion are shown in Table 3-18. A general list of test requirements is derived by considering the following operating particulars which define IR sensor systems:

1. *IR System Types by Function*
   a. Radiometers
   b. Spectrometers
   c. Target sights
   d. Target trackers
   e. Target search
   f. Search-track
   g. Mapping

2. *IR System Types by Wavelength Response*
   a. Short wavelength (0.72 - 1.2 microns)
   b. Intermediate wavelength (1.2 - 7 microns)
   c. Long wavelength (7 - 30 microns)
   d. Multiple wavelength, multiple filters
   e. Multiple wavelength monochrometers
   f. Wide bandwidth (2 - 10 microns wide)
   g. Narrow bandwidth (0.1 - 1 micron wide)
   h. Target temperature

3. *IR Systems Types by Field of View*
   a. Narrow field, single detector
   b. Wide field, single detector, field lens
   c. Wide field, single detector, scanner
   d. Wide field, line detector array, scanner
   e. Wide field, detector matrix

4. *IR System Types by Signal Processing Mode*
   a. Direct scan
   b. Interlaced scan
   c. Spatial discriminating reticles
   d. Spatial discriminating logic
   e. Photographic recording
   f. Kinescope display
   g. Alarm signal
   h. Error signals
   i. Time frequency transforms of spatial signals
   j. Time frequency transforms of wavelength signals

5. *IR System Types by Environments*
   a. Temperature
   b. Humidity
   c. Pressure
   d. Shock and vibration

TABLE 3-18. STANDARD SYMBOLS FOR PAR. 3-8

Symbol	Definition	Typical Units
A	Area of aperture; area of blackbody aperture	cm^2
D	Distance from a surface to reference blackbody	cm
D_1	Pathlength through collimator	cm
D_2	Free pathlength outside collimator	cm
D_m	Diameter of beam deviating mirror	
D_b	Diameter of collimated beam	mrad
$F(\Omega)$	Modulation transfer function amplitude	
$f/no.$	Ratio of (focal length)/(aperture diameter)	
H	Radiant flux density	$w\ cm^2$
H_o	Radiant flux density of blackbody at reference temperature	$w\ cm^2$
H_1	Radiant flux density at a finite distance from a reference blackbody	$w\ cm^2$
dH	Incremental change in flux density	$w\ cm^2$
k	Empirical constants (Elder and Strong)[93]	
$°K$	Absolute temperature	degree Kelvin
n	Integer	
Q	Figure of merit of tuned circuit	
S_b	Displacement of collimated beam	in.
S_s	Distance from sensor to beam-deviating mirror	
t	Atmospheric transmission	
t_c	Collimator transmittance	%
t_e	Empirical constants (Elder and Strong)[93]	
t_1	Transmittance over a reference distance	%
t_2	Transmittance over a measured distance	%
T	Absolute temperature	°K
dT	Incremental temperature change	°K
w	Precipitation over a given pathlength	cm
W	Total radiation	w
W_n	Width of bar	in.

GREEK LETTERS

γ	Angle between source-point direction and blackbody axis	deg
ϵ	Emissivity	ratio 0 to 1
θ_b	Angle of collimated beam deviation	deg
$\Delta\theta_m$	Angle of beam-deviating mirror rotation	deg
λ	Wavelength measurement of light	micron
σ	Stefan-Boltzmann constant	$5.67 \times 10^{-12}\ w\ cm^{-2}\ °K^{-4}$

6. *General Test Requirements*

a. Sensitivity

b. Resolution

c. Field of view

d. Field distortion

e. Spectral response

f. Axes orthogonality

g. Tracking rates

h. Positioning accuracy

i. Signal saturation

j. Background discrimination

k. False-alarm rate

l. Built-in test source calibration

m. Baffle effectiveness

n. Mechanical references

3-8.1 TEST CATEGORIES

The two general categories of tests are identified as Type 1 and Type 2 and involve testing for deviations from a specified standard.

Type 1 tests are generally conducted in a laboratory environment using general purpose equipment of high stability, accuracy, and precision. Type 2 tests are generally conducted in the field with special purpose equipment of acceptable precision and stability. The accuracy of Type 2 equipment is associated with its calibration rather than with any integral reference standard.

The reference for radiation measurements is the certified blackbody standard and calibrated temperature controller. The blackbody, in conjunction with accurate length and angle measuring instruments, is the basic tool for the calibration and alignment of all IR optical instruments.

3-8.2 LABORATORY INSTRUMENTS AND TECHNIQUES

3-8.2.1 IR Collimator

A set of guidelines can be established for the selection of a laboratory IR collimator on the basis of the information contained in par. 3-2 and from the general requirements of an IR laboratory. Par. 3-2 contains a detailed discussion of the optical properties of lenses, spectral filters, reflectance, and image quality.

The collimator is used for simulating an object at infinity (or at a lesser distance) to which the IR system under test will be exposed.

In general, a laboratory model IR collimator should have:

a. All reflective optics

b. Large diameter clear aperture relative to the system under test

c. Long focal length relative to the system under test and a relatively large f/no. (f/10)

d. Preferably, a Herschelian configuration

The image quality of a collimator is a function of surface errors, astigmatism, spherical aberrations, coma, and diffraction. In a setup where the collimator is larger in diameter than the IR system under test, resultant diffraction will be due to the diameter of the system aperture and the spherical aberrations will be greatly reduced because of the resulting higher effective f/no. (The small-diameter IR system acts as an aperture stop for the collimator.) Coma can be avoided by using an on-axis collimator setup so that the two critical specifications are surface errors and astigmatism. When it is specified that these be small (i.e., 1 to 3 sec of arc) the collimator will, for all practical purposes, be diffraction-limited when used to test IR systems whose optical diameter is smaller than that of the collimator.

A large-diameter collimator will ensure a more uniform central bundle, and also permit translation and rotation of the IR system within the collimated bundle while avoiding vignetting effects.

All reflective optics permit the widest range of uniform spectral response, as compared to catadioptric or refractive type systems.

The four basic configurations of reflective optics collimators are illustrated in par. 3-2. The three on-axis types have central obscuration due to the source having been inserted at the focal plane, while the off-axis Herschelian type does not.

Where obscuration does occur, small aperture IR systems should be placed in the clear aperture area of the collimator. Newtonian systems are such that the obscuration can be limited to 1/10 - 1/15 of the beam diameter. Cassegrainian systems having a large central obscuration are really suitable only for testing other Cassegrainian systems of equal diameter or systems very much smaller since the obscuration ranges generally from 1/2 to 1/4 of the beam diameter.

The Herschelian collimator will provide a clear aperture at the expense of a smaller effective f/no. since the aberrations are determined by the full lens diameter f/no., while the usable diameter determines the diffraction limit. The special testing problems associated with long wavelength systems are reduced by using the Herschelian collimator. Since ambient temperature objects are targets of considerable irradiance in the long wavelength spectral region (including the IR system under test) the use of a Herschelian collimator will place the system under test such that it does not re-image onto itself as would be the case with on-axis collimators. Re-imaging in an on-axis collimator occurs when a point is placed at the radius of curvature of a reflective lens thereby causing the point to be re-imaged onto itself. Thus, for a typical 100-in. focal length collimator, a system under test placed on-axis at 200 in. from the lens would be re-imaged back onto itself forming its own target. Furthermore, because of the large depth of focus, a sensible return image would result for large variations of this distance and would interfere with the test target of interest. Under test situations where the effectiveness of IR sensor optical baffles are measured by illuminating the baffles at various off-axis angles, the baffle reflections would become an on-axis source of far greater intensity than a low-level target of interest. The Herschelian, off-axis collimator serves to image the sensor onto the target where it can be properly "light-trapped" and thus not be returned as a signal.

In testing long wavelength wide field-of-view systems, the facility or enclosure housing the collimator and sensor is critical with respect to the uniformity of irradiance since part of the facility will be within the field of view of the system. The use of low-temperature shrouds and highly reflective, i.e., low emissivity, surfaces does not reduce the problem unless the shrouds are free of irradiance gradients. At a temperature of 4°K , the irradiance of a source is $\approx 10^{-9}$ w/cm^2, several orders above the IR system sensitivity. Low-emissivity surfaces must be used with care to ensure that no gradients such as the edges of the collimator structure, room corners, and the sensor under test are reflected from these surfaces as targets of interest.

The lower limit of a test background temperature can be set at 4.2°K since this is the lowest temperature background within which the IR system will operate; however, this limit need not be realized to provide an accurate test. Since IR sensor systems are spatially and temporally selective for maximum response to small size targets, discrimination against a low frequency, uniform background will occur.

A target source mounted on a two-axis cross-slide can be moved readily over a desired field of view and focus range. Since the focusing sensitivity of a collimator is the square of the ratio of the collimator focal length to the focal length of the system under test, the optimum focus position of the IR sensor can be tested by defocusing the collimator. For example, if the focal length of the collimator is 100 in. and the sensor focal length is 10 in., then $\left(\frac{100 \text{ in.}}{10 \text{ in.}}\right)^2 = (10)^2 = 100$, which means that a 1-in. change of focus in the collimator is exactly equivalent to a 0.01-in. change in focus of the sensor. Thus, when it is not possible to refocus the sensor physically, the collimator can be refocused for best spatial resolution and compared to the sensor response measured at an infinity focus position. Any response better than the collimator response at infinity focus would indicate an improperly focused sensor. The exact amount of refocusing required would be determined by the collimator focus change times the squared focal-length ratio. An increase in the collimator focal distance from its infinity focus position would indicate a shorter-than-infinity focal distance for the lens under test and vice versa. With suitable spatial resolution targets, positioning accuracy ≈ 0.001 in. is readily achievable.

Wide field-of-view systems can be tested—one channel at a time—with a narrow field collimator by rotating the IR sensor about its

entrance aperture, using a precision rotary table for angle readout. When the scanning mode of the IR sensor requires continuous rotation, a beam deviating mirror placed in the collimated bundle can be pivoted to change the direction of the collimated beam relative to the sensor. (See Fig. 3-144.) The angle of beam deviation θ_b is twice the angle of mirror rotation $\Delta\theta_m$ since beam position is a reflection of the collimator. For moderate field angles (less than 20 deg) the IR sensor can simply be rotated. This applies provided the aperture diameter of the sensor is less than two-thirds that of the collimator beam, since rotation of the beam-deviating mirror will translate the beam over the field so that the full beam diameter is only usable at one angular position. As illustrated in Fig. 3-144 the displacement of the beam is $S_d = S_s \tan\theta_b$ where S_s = distance of the sensor to the beam deviating mirror. The beam deviating mirror is, of course, larger than the collimated beam due to its inclination, therefore, its maximum dimension in the direction of rotation of the beam deviating mirror is $D_m = \frac{D_b}{\cos\theta_{m_2}}$ where D_b is the diameter of the collimated beam. This arrangement will cause the sensor to "walk across" the collimated beam due to the beam translation. Consequently, nonuniformity of the collimated beam will affect the relative response of the sensor for the same input target source.

A beam uniformity of ± 2 percent can be expected with a high image quality collimator

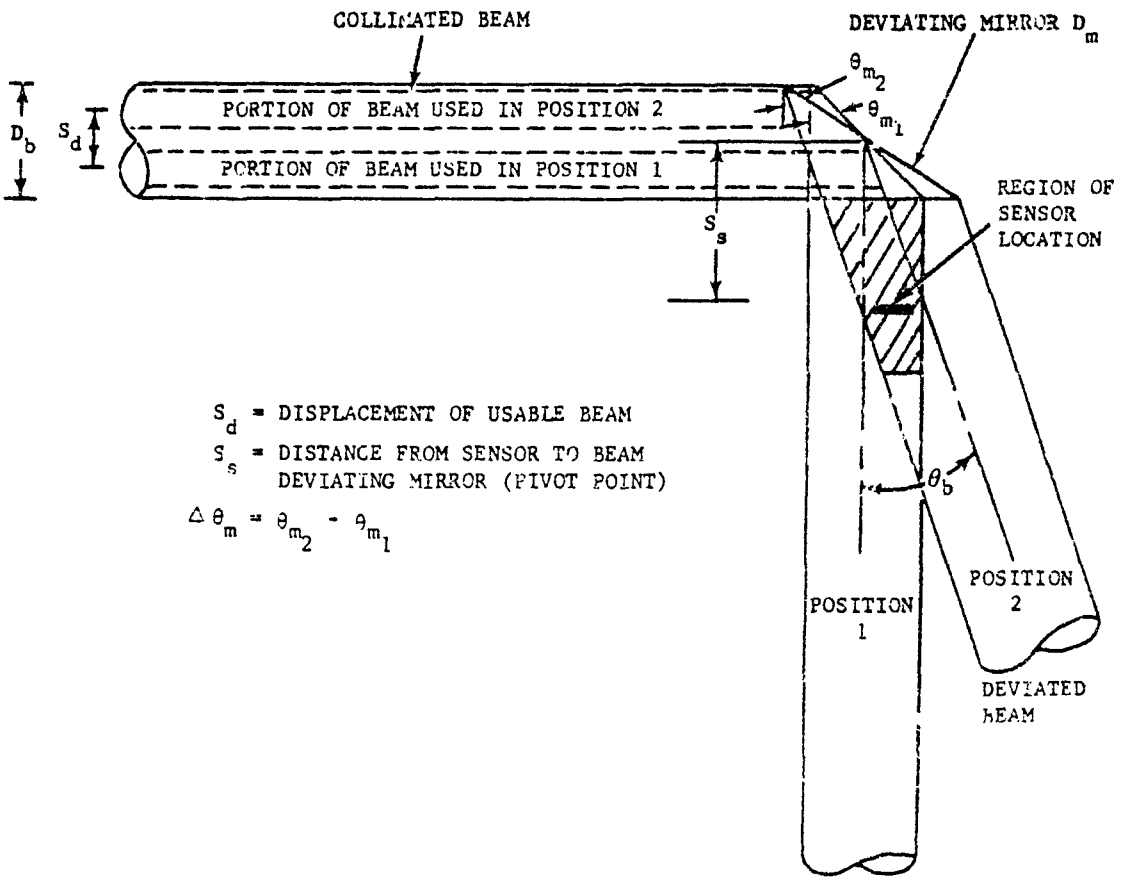


FIGURE 3-144. Beam-deviating Mirror Schematic Diagram

using a pinhole target less than 0.1 mrad in diameter. With narrow-field paraboloidal collimators, the beam uniformity is very sensitive to target diameter while for wide-field collimators, beam uniformity is a characteristic of the design. Consequently, for some collimator designs, highly accurate measurements require that the collimator transmission be determined for each size of metering aperture. A change in target diameter of 10 times, i.e., from a 0.1 to 1.0 mrad pin-hole may introduce a 20 percent relative error in collimated beam flux density. Where the wide field of view of a sensor makes it impractical to use a simple rotating beam deviator, the beam deviator can also be translated to minimize "walking across" the beam while achieving angular changes $\approx \pm 3$ deg.

3-8.2.2 Blackbody

The laboratory blackbody is the irradiance standard for IR system measurements. It consists of a heated cavity that is temperature-controlled by means of a servo loop which senses the error signal between a monitoring thermometer and a calibrated temperature control potentiometer. The cavity is a passive element which, as a consequence of its shape factor (cone or sphere), can achieve an emissivity ≈ 0.99 when the cavity walls have an emissivity as low as 0.85. The choice of shape factor is as much dependent upon the manufacturer's experience as upon personal preferences. Proportional controllers are thought to be superior to the SCR type of control because of the tighter control loop and freedom from electromagnetic interference (EMI).

Laboratory blackbodies are available over a temperature range from 20° above ambient to 3000°C and down to −10°C with the use of an auxiliary self-contained cooler. The standard cavity apertures are 1/2 and 1 in. in diameter with a Lambertian radiation over a cone angle of 30 deg.

Fig. 3-145 is a plot of the iso-flux density curves for a Lambertian radiator which is a circle of unit radius with the radiation target tangent to the circumference of this circle. In particular, it should be noted that equal distances from the blackbody, over the 30-deg field, do *not* fall on the iso-flux density curve since the blackbody exit aperture is on the circumference of the radiation circle, not at its center. A pin-hole aperture will generate a new radiation circle while, at the same time, reducing the Lambertian radiation angle according to how far the aperture is from the cavity.

Since the blackbody is most often used with metering apertures or spatial resolution targets, the cavity merely serves as an extended source for uniformly "illuminating" the clear spaces against which the pin-hole or target is a "silhouette". A larger expanse of uniform background will permit larger targets to be used without vignetting the Lambertian cone angle.

The high level radiant energy from the blackbody cavity tends to heat the pin-hole apertures to the point of sensible self-emission, therefore, precautions should be taken to cool the aperture. This may consist of circulating water, making the aperture surface highly reflective, or shuttering the cavity to block the heat during standby periods.

The f/no. of the collimator used with the blackbody will determine the minimum Lambertian cone angle required to fill the collimator aperture. For example, an f/10 collimator with 100-in. focal length and 10-in. diameter requires a 1/10 rad (6-deg) cone angle while an f/2 collimator requires a cone angle of about 30 deg. The radiant flux density H_o of the blackbody is determined by its temperature from Eq. 3-195

$$H_o = \epsilon \sigma T^4, \text{ w cm}^{-2} \tag{3-195}$$

where

ϵ = emissivity

σ = Stefan-Boltzmann constant, 5.6697×10^{-12} w cm^{-2} °K^{-4}

T = temperature, °K

Thus, for a 1°K source temperature and an emissivity of 1, the irradiance is $H_o = 5.7 \times 10^{-12}$ w cm^{-2}. When the size of a metering aperture A is included in the calculation

$$W = \epsilon \sigma A T^4, \text{ w} \tag{3-196}$$

where W is the total radiation and A is the aperture area in cm^2.

The radiant flux density H_1 at some point distant from the blackbody source is calculated as

$$H_1 = \frac{W \cos \alpha}{\pi D^2} \quad , w\ cm^{-2} \qquad (3\text{-}197)$$

where

W = total radiation, w

α = angle between the source-point direction with the blackbody axis, i.e., off-axis condition

(For on-axis condition, $\alpha = 0$, $H_1 = \frac{W}{\pi D^2}$)

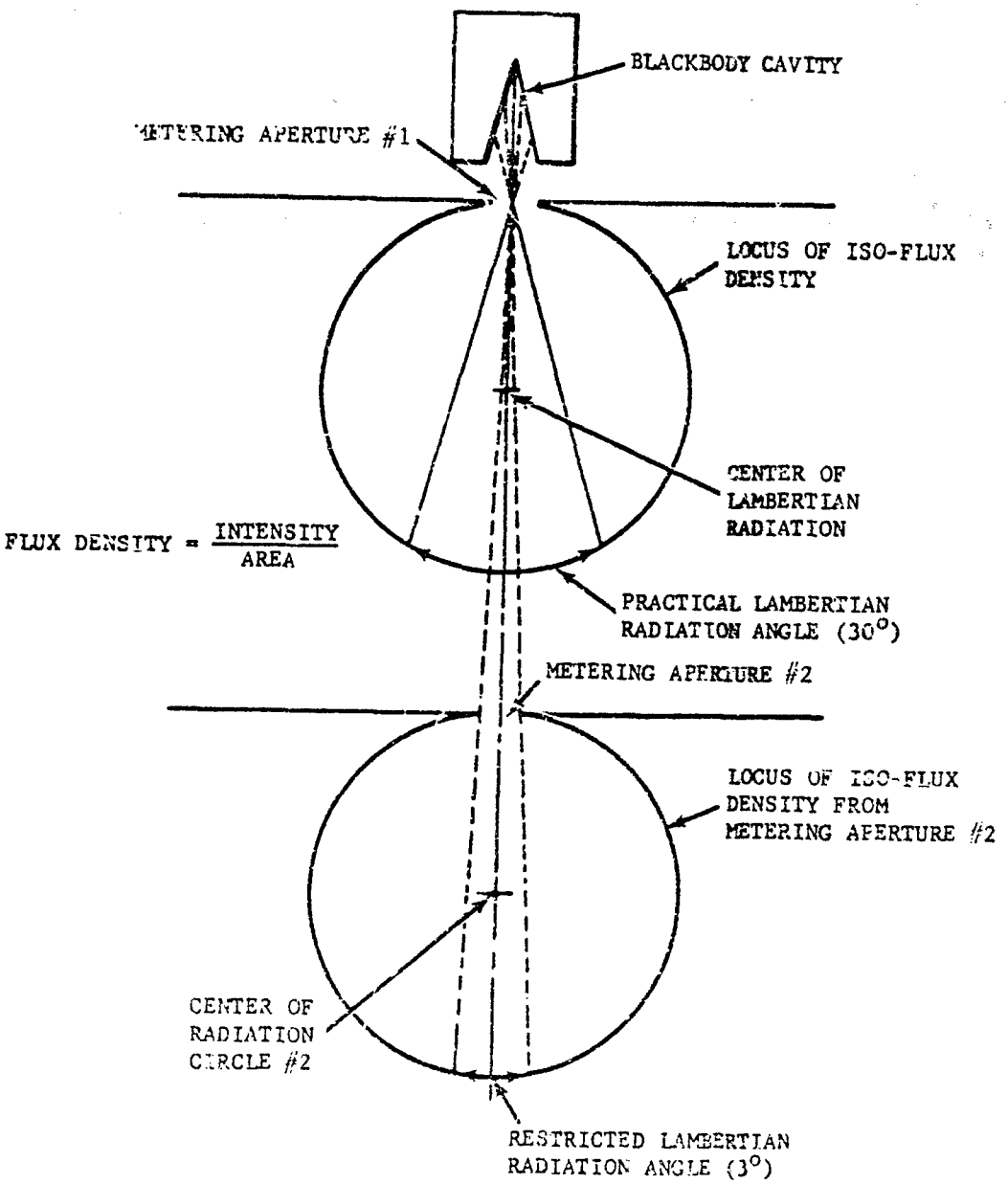


FIGURE 3-145. Lambertian Radiation

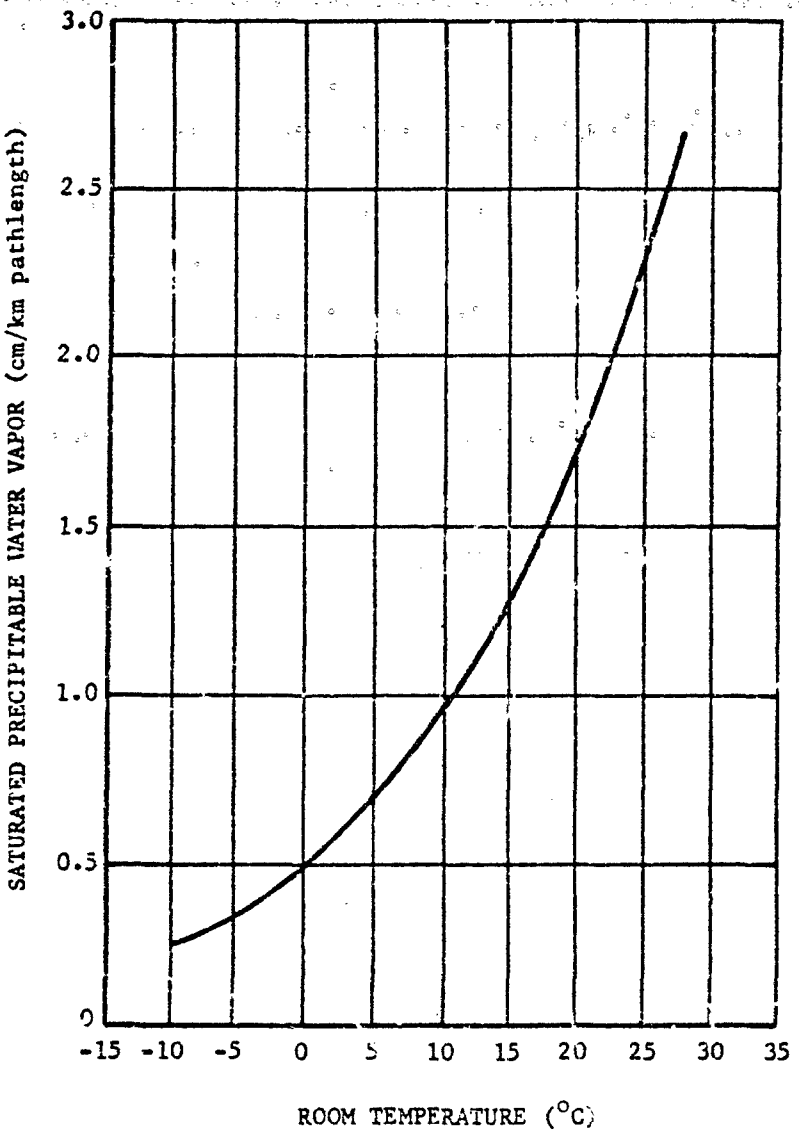


FIGURE 3-146. Density of Saturated Water Vapor (100% Relative Humidity) at 1 atm, 760 mm Hg

D = distance of the point from source, cm

As an example, for $T = 1000°K$, $A = 1\ cm^2$, $D = 1000$ cm, $\alpha = 0$, and $\epsilon = 1$:

$$H_1 = (1)(5.7 \times 10^{-12})(1)(1000^4)/[\pi(1000^2)]$$
$$= 1.8 \times 10^{-6}\ w\ cm^{-2}$$

Eq. 3-196 indicates that the metering aperture has a direct influence on the flux density according to its area or in proportion to the diameter squared. Thus, relative flux densities can be established in steps of 4X by doubling the diameter of successive metering apertures. For example, starting with a 0.001-in. diameter pin-hole aperture, the irradiance of a 0.002-in. diameter aperture would be four times greater; a 0.004-in. diameter aperture would be 16 times greater; a 0.008-in. diameter aperture would be 256 times greater, etc.

However, when used at the focal plane of a collimator, this relationship may not hold for large apertures due to the off-axis aberrations of the collimator, therefore, the relative flux density must be calibrated.

The ideas conveyed by Eqs. 3-197 and 3-198 are addition significant calibration considerations.

For small temperature differences it is sometimes convenient to calculate the change in flux density by using the implicit derivative of Eq. 3-195, $H = \sigma T^4$. Emissivity ϵ of 1.0 can be assumed.

$$dH = 4\epsilon\sigma T^3\, dT \qquad (3\text{-}198)$$

Thus, for $dT = 10°K$ at $T = 1000°K$

$$dH = 4 \times 5.7 \times 10^{-12} \times (10^3)^3 \times 10$$
$$= 0.23\ w\ cm^{-2}$$
$$dH = 2.3 \times 10^{-1}\ w\ cm^{-2}$$

The incremental flux density from an aperture A at a distance D is obtained from the derivative of Eq. 3-196 as

$$dH = 4\epsilon\sigma A\left(\frac{T^3}{\pi D^2}\right) dT \qquad (3\text{-}199)$$

which for $T = 1000°K$, $\epsilon = 1$, $A = 1\ cm^2$, $D = 1000$ cm, and $dT = 10°K$

$$dH = 4(1)(5.7 \times 10^{-12})(1)(1000^3)(10)/[\pi(1000^2)]$$
$$= 7.3 \times 10^{-8}\ w\ cm^{-2}$$

i.e., a 10°K change in blackbody temperature will result in a change of 7.3×10^{-8} w cm^{-2}.

3-8.2.3 Chopper-modulator

A mechanical chopper is normally used to modulate the blackbody energy thereby permitting ac amplification of the detector output or allowing the center frequency of the IR system under test to be matched. Placing the chopper between the blackbody and the metering aperture will:

a. Provide the proper reference for different aperture sizes by making the angle subtended by the chopper identical to the angle subtended by the aperture. Particular attention should be paid to this geometry when testing long wavelength IR systems since a chopper placed between the system and the aperture may radiate more power than that received through the aperture. Placing the chopper behind the aperture will result in a flux difference proportional only to the relative temperature and emissivity of the chopper and blackbody.

b. Prevent inadvertent changes in the intended signal waveshape since a chopper placed in front of the pin hole (located at the focal point of the collimator) may result in partial vignetting of the pin-hole image during the chopping cycle.

c. The chopper can serve to shutter and cool the aperture from the radiated heat of the blackbody cavity.

The use of a frequency pick-off on the chopper modulator in conjunction with a synchronous demodulator serves effectively to increase signal-to-noise ratios. This is a highly effective technique for rejecting transients signals such as power-line ripple. It also permits varying the bandwidth from 1000 to 0.001 Hz without changing the Q of the tuned circuit or altering the signal level. Thus, extremely low level signals which would otherwise be masked by noise may be recovered by narrowing the bandwidth.

3-8.2.4 Blackbody Usage

The standard blackbody in an IR laboratory facility serves as a standard source of radiation and as a stable uniform background for spatial frequency measurements.

As a standard source of radiation, complemented by a set of precision metering apertures, the blackbody establishes a specified flux density at a known distance. The irradiance on the optical test equipment must be corrected to compensate for the effect of the atmosphere over the pathlength.

Over the relatively short pathlength in a laboratory setup, H_2O is the principal attenuating constituent of the atmosphere. The correction factor discussed in Chapter 2 must be applied. In correcting for the pathlength attenuation by the H_2O content, the water vapor concentration is the measured variable and the transmission is determined as

$$t = -k \log_{10} w + t_o \qquad (3\text{-}200)$$

where

w = precipitable centimeters of water over the pathlength

k, t_o = empirical numerical constants that vary with the selective spectral band listed in Table 3-19.

The precipitable centimeters of water can be derived from Fig. 3-145 when the atmospheric room pressure is 1 atm and

w = (saturated precipitable water vapor cm/km) × (pathlength) × (relative humidity)

Since the equivalent pathlength is approximately inversely proportional to the square of the pressure ratio, the data provided in Fig. 3-146 can be corrected as required. The corrections are generally small for distances involved under laboratory conditions since at a 1000-ft altitude the transmittance over a 20-m pathlength, when the relative humidity is 50 percent, will be reduced by approximately 0.1 percent of the transmittance at zero elevation.

For example, determine the radiance of a ¼ cm source 600°C blackbody under the following conditions:

Pathlength:	10 m in 1.9- to 2.7-micron band
Relative humidity:	45%
Temperature	23°C
Pressure:	1 atm

a. The precipitable centimeters of water (from Fig. 3-146) is

$$2\left(\frac{10\text{m}}{1000\text{m}}\right) 0.45 = 0.009 \text{ cm}$$

b. From Table 3-19 $k = 13.1$ and $t_o = 72.5$. Therefore, Eq. 3-199 takes the value:

$$t = -13.1 \log_{10} 0.009 + 72.5$$

$$t = -13.1(-2.0458) + 72.5$$

$$t = 99.3\%$$

c. The average percentage of irradiance in the spectral band using a blackbody slide rule as described in Chapter 2 is 13.5% - 2.6% = 10.9 % of the total blackbody irradiance.

d. Thus, the combined spectral and atmospheric effects will result in an effective irradiance that is 0.109 × 0.993 = 0.108 of the total irradiance.

The transmittance obtained by these calculations is an average ratio of irradiance absorbed to the irradiance emitted over the entire spectral interval and is not uniform for each wavelength in that spectral band. (See Chapter 2 for spectral transmittance data.)

TABLE 3-19. WINDOW REGIONS IN THE INFRARED[93]

SPECTRAL BAND, μ	k	t_o
0.72 - 0.92	15.1	106.3
0.92 - 1.1	16.5	106.3
1.1 - 1.4	17.1	96.3
1.4 - 1.9	13.1	81.0
1.9 - 2.7	13.1	72.5
2.7 - 4.3	12.5	72.3
4.3 - 5.9	21.2	51.2
5.9 - 14	*	*

* In the 5.9 - 14μ band, specifically beyond 13μ, CO_2 becomes a significant absorber; water vapor has relatively little absorption from 8 - 11μ as discussed in Chapter 2.

For example, determine the irradiance of a 1 cm² 600°C blackbody source under the following conditions:

Pathlength:	20 m
Relative humidity:	45%
Temperature:	23°C
Pressure:	1 atm

a. Spectral band of interest: 1.9 to 2.7 μ

b. Cm of precipitable water: From Fig. 3-143 the precipitable water vapor for 23°C temperature is found to be 2 cm/km. For a 20 m pathlength and 45% relative humidity, the precipitable water is

$$2\left(\frac{20\text{ m}}{1000\text{ m}}\right)0.45 = 0.0108\text{ cm}$$

c. $k = 13.1$ and $t_o = 72.5$ [from Table 3-19]

d. Pathlength transmittance = 82.5% [Eq. 3-200]

e. Radiance in the band of interest = 0.111 [from radiation slide rule].

Thus, the combined spectral and atmospheric effects result in an effective irradiance of $0.111 \times 0.825 = 0.0916$ of the total radiance of the 1 cm² 600°C source.

The transmittance thus obtained represents an average over the entire spectral interval and is not uniform for each wavelength in the band, as discussed in Chapter 2.

A blackbody source located at the focal plane of a collimator will be attenuated by the reflectivity or transmissivity of the collimator optics. The extent of attenuation can be determined by direct measurement with a double-beam IR spectrophotometer or alternately by means of a stable radiometer, spectrally band-passed over the region of interest. When using a radiometer, the irradiance level in the collimated beam is first measured and recorded. Then—without changing the blackbody metering aperture, spectral filtering, or temperature—the blackbody is removed from the collimator allowing the radiometer to be placed at a free-path distance from the blackbody so that the same signal level can be measured with the radiometer.

As an example, the average transmission of a collimator is determined here, assuming the following conditions:

Collimator focal length:	300 cm
Total pathlength from blackbody to IR sensor:	2000 cm
Number of collimator reflection surfaces of unknown reflectivity:	2
Spectral band of interest:	1.9 to 2.7μ
Temperature:	23°C
Relative humidity:	45%
Pressure:	1 atm

From Fig. 3-146, Table 3-19, and Eq. 3-200, the transmittance $t_1 = 82.26\%$.

The blackbody is removed from the collimator and the radiometer distance is adjusted so that the same signal level is measured. If this free-path distance is 2150 cm, the collimator transmissivity is determined by,

$$t_{c_1} = \left(\frac{D_1}{D_2}\right)^2\left(\frac{t_1}{t_2}\right)$$

where

D_1 = pathlength through the collimator = 2000 cm

D_2 = free pathlength outside the collimator = 2150 cm

t_1 = transmittance over the D_1 distance = 82.26%

t_2 = transmittance over the D_2 distance = 81.8%

t_{c_1} = collimator transmittance

Thus

$$t_{c_1} = \left(\frac{2000}{2150}\right)^2\left(\frac{82.26}{81.8}\right) = 86.8\%$$

This transmittance of 86.8% for the collimator indicates an average reflectivity of 95.6% for each of the two collimator surfaces and a collimated blackbody irradiance of $t_1 \times t_{c_1} = 82.26 \times 86.8 = 71.3\%$ of the unattenuated available energy.

Where extreme accuracy is required in a changing atmospheric environment, it may be necessary to replace the ambient atmosphere with dry air or nitrogen to ensure stable and repeatable measurements.

3-8.2.5 Modulation Transfer Function (MTF) Measurements

In visual-region optics, MTF can be measured by using spatial sinewave radiance targets produced on a film transparency where the variability in transmission over each spatial cycle is sinusoidal. To permit the measurement of fre-

quencies approaching the grain size of the transparency, optical minification is used such that the image contains the spatial radiance amplitude of the aspect at these frequencies.

Film and substrate materials being opaque for most of the long wavelengths of interest in IR measurements, dictate that particular consideration be given to the spatial frequency targets used in the IR.

In the recording of MTF measurements, the irradiance at the image plane is incident on a photomultiplier through a narrow slit. Generally, the slit should be in the order of one-fourth the diameter of the Airy disc of the system under test. Where the required dimension of the slit is impractical, the image is magnified by microscope optics so that a practical size slit can be used.

The relatively low sensitivity of IR detectors compared to photomultipliers makes the use of ultra-narrow slits impractical, notwithstanding the practical difficulties of fabricating narrow slits and IR microscopes. Because of these constraints, clear-space bar reticles are used for infrared MTF targets. The reticles are generally etched from 1 mil-thick beryllium-copper stock.

The three basic methods of making MTF measurements in the IR entail the use of (1) a nonperiodic target (Method 1), (2) a continuously varying square-wave spatial frequency target (Method 2), or (3) a discrete square-wave spatial frequency target (Method 3).

3-8.2.5.1 Method 1

Method 1 is implemented by placing a knife-edge object in the focal plane of a collimator and wide detector at the image plane of the optics. It can be shown that the recorded raw data from a knife-edge object and a wide detector are the double integral of the optical line spread-function. Consequently, the second derivative of the raw data is required for determining the line spread function. Unfortunately, it has been found impractical to compute

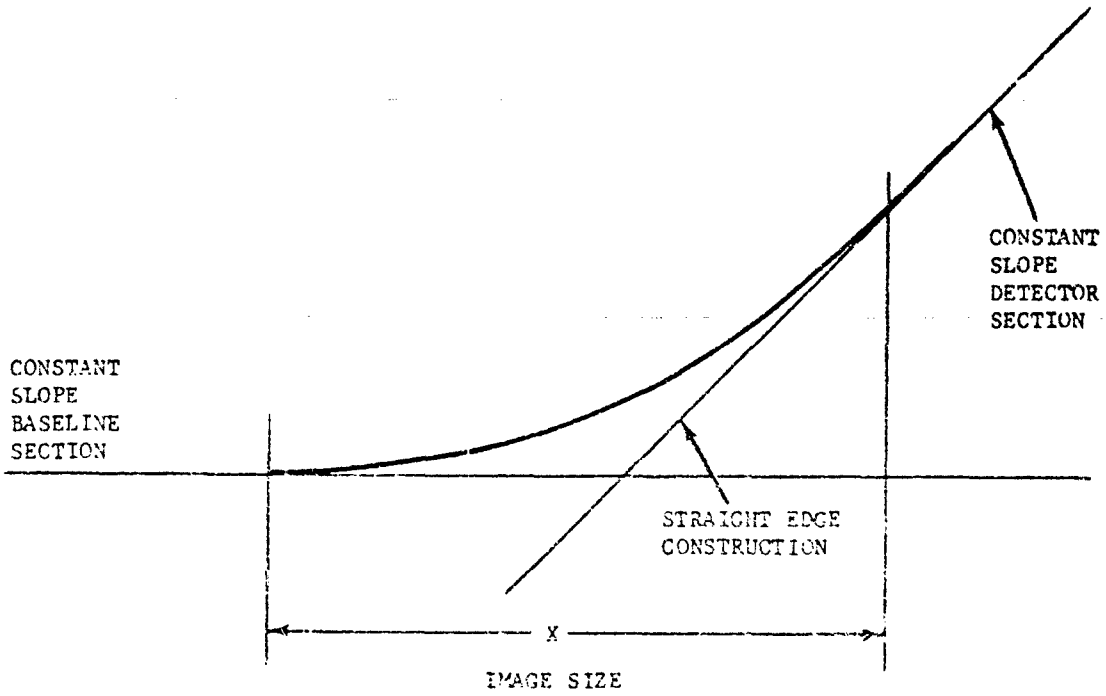


FIGURE 3-141. Section of Knife-edge Recording Characteristic Curve

these derivatives directly because of the extreme sensitivity of the second derivative to minute fluctuations in the slope of the raw data which can be easily distorted by noise. If the Fourier transforms are used however, the MTF can be computed directly by calculating the transform of the second derivative. The Fourier integral is less sensitive to noise amplitudes in the raw data than direct differentiation, thereby, providing a practical computational procedure for Method 1 measurements. When using this method it is recommended that at least 200 data points be recorded to ensure an accurate representation of the optical high-frequency response.

Fig. 3-147 is an illustration of a raw data plot using Method 1. The region of interest is the curved portion of the data (X) which represents the minimum circle of confusion of the optical system under test. The size of the image is determined by the scale of the recorded X dimension multiplied by the ratio of collimator-to-test-lens focal lengths.

Fig. 3-148 illustrates a typical set of recordings for five different focal positions at 0.002-in. focus change intervals. Curve No. 3 represents the best focus position, as indicated by having the smallest dimension in X. This curve can be used to compute the lens transfer function and line spread function.

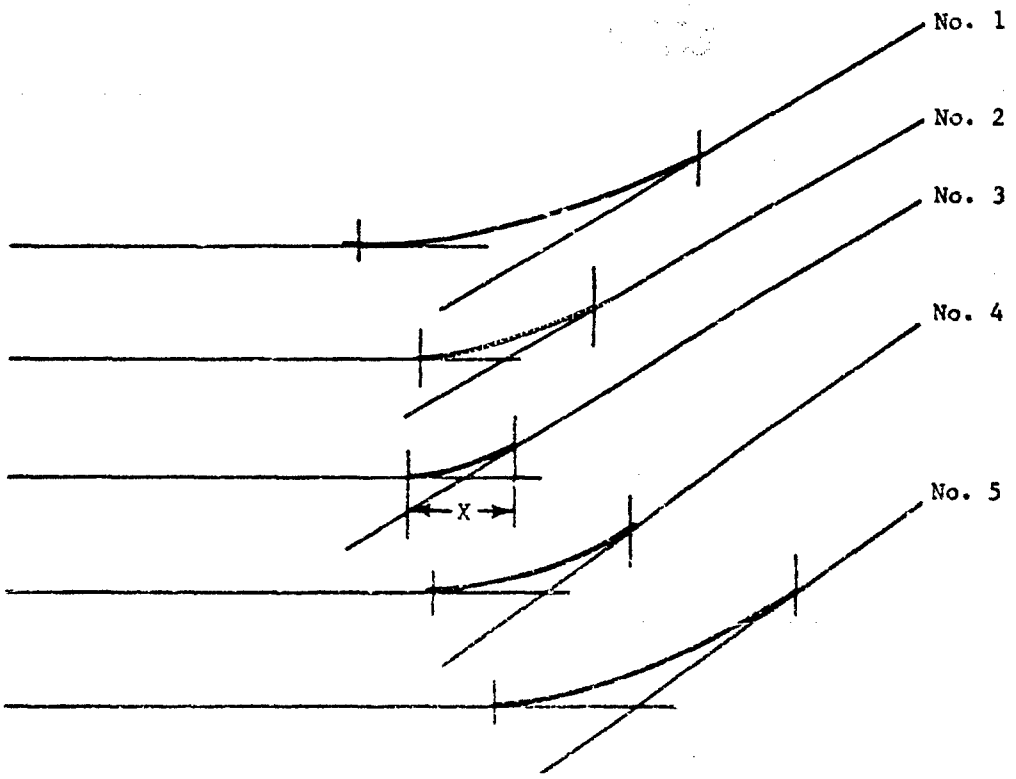


FIGURE 3-143. Typical Knife-edge Record for Different Focus Positions

3-8.2.5.2 Method 2

This method requires a continuously variable spatial frequency target scanned across the sensor field of view. The target, which is "back lighted" by a blackbody cavity, serves to modulate the irradiance of the blackbody spatially. To ensure that the sidebands generated by this modulation are below the cutoff region of the electronic amplifiers and the response of an *x*-*y*-plotting recorder, the field should be scanned by the target at a relatively constant time-rate for each spatial frequency. This implies that the physical scan rate should decrease in proportion to the increasing spatial frequency and, to accommodate and *x*-*y*-plotter of 1 Hz (full scale) response, the constant time-rate should be 1 Hz or slower.

This requirement can be met by a target whose bar and space width decrease at a constant rate, e.g.; with the rate being 5 percent and the first space 1.0 in. wide, the following equation gives the width of the successive bars:

$$W_n = \frac{1}{(1.05)^{n-1}}, \text{ in.}$$

where

W_n = width of each bar, in.

n = an integer

The second bar, for example, will be

$$W_2 = \frac{1}{1.05} = 0.9524 \text{ in.}$$

The third bar will be

$$W_3 = \frac{1}{(1.05)^2} = 0.9073 \text{ in.}$$

And so on to a practical fabrication limit or to a limit beyond the lens cutoff frequency. A target of this type can be driven by a rate servo using a linear rate potentiometer to achieve the desired constant temporal frequency.

In order for the low-frequency electrical signal to pass through the IR system amplifiers, a constant-speed chopper is required to serve as a carrier frequency for the spatial target modulation. The chopper speed which determines this carrier is set at some nominal frequency within the passband of the system electronics. The rectified carrier signal will produce a signal that is proportional to the target modulation which can be recorded on an *x*-*y*-plotter and/or a digital data acquisition system.

Since the target input signal is a train of square waves, and the optical transfer function is analyzed as a sine wave modulation transfer function (MTF), the waveform should be filtered so that only the sinewave fundamental of the modulation is recorded. This can be accomplished by a post-rectifier low-pass filter. Several off-the-shelf synchronous-demodulator amplifiers are available with a wide range of post-rectifier filter time constants that can be manually selected to perform the required filtering. The envelope of the resulting analog record of the target modulation will be the MTF of the IR system.

When it is desirable to determine the MTF of the IR optics only, the effect of the system's finite width detector must be removed from the recorded data. Since the detector function attenuates the $F(\Omega)$ amplitude of the optics MTF in a known manner, the data can be readily corrected geometrically on the analog record or by means of the computer. Since the detector represents a $\frac{\sin x}{x}$ function, the envelope must be accordingly corrected.

3-8.2.5.3 Method 3

The spatial response of an IR system can be determined simply and directly when the specified performance is expressed as a percentage of modulation at a particular spatial frequency. This method requires the use of a target having six bar-space pairs corresponding to the specified spatial frequency and a wide space corresponding to a zero spatial frequency. The wide space will produce 100 percent modulation and the specified bar space pairs will produce the performance modulation. For example, if the specifications state that 75 percent modulation is required for a 1 mrad target, then the dimensions of each bar and space should subtend 1 mrad at the collimator focal plane; while the reference wide space, i.e., 100 percent modulation, should be considerably larger. However, because in actual practice, the reference space cannot be sufficiently wide to provide 100 percent modulation, a space width at least four times greater than the specified spatial frequency can be used and the modulation level of the

reference space can be determined from the lens design MTF. If, in the above example, the lens were to have been diffraction-limited, then where 1 mrad produced 75 percent modulation, 4 mrad would have produced 95 percent modulation. As a result, the relative amplitude between the reference space and the six bar-space pairs would be 75/95 = 79 percent provided the detector width did not attenuate the $F(\Omega)$ amplitude, which is not the case for the illustrative example where the detector width is assumed to be 1 mrad.

Since $F(\Omega)$ vanishes at 1/2 mrad and at 1/4 mrad, at 1 mrad the modulation level is only 70 percent instead of 75 percent (because of the detector effect) and 93 percent instead of 95 percent at 4 mrad. Thus, with a 4-mrad space reference and 1-mrad bars and spaces, the recorded modulation would be 70/93 or 75 percent. By selecting the reference spatial dimension judiciously, the analog record can provide a direct indication of the specified modulation. It is also apparent from the example that care should be taken not to specify a modulation level for spatial frequencies close to the vanishing point of $F(\Omega)$, i.e., at even multiples of the detector width. Since the spatial resolution of an IR system is generally limited by the detector width, it is desirable to standardize the dimensions of the bar and space target to equal the detector width (i.e., a bar-space pair is twice as wide as the detector) and to specify the reference space at four times the width of the detector.

3-8.3 ON-BOARD CALIBRATION

On-board calibration makes possible the performance evaluation of an IR system during actual operation and may be qualitative or quantitative depending on the specific mission requirements of the IR system.

Quantitative calibration is used when precision channel detectivity is required. When the expected radiometric stability of the system is less than the required precision of measurement, an on-board signal of high-stability can be used to calibrate the radiometric responsivity of the system periodically.

Qualitative calibration provides end-to-end testing to ascertain whether or not all subassemblies are functioning.

There are two calibration methods: (1) radiometric source and (2) simulated electronic signal.

3-8.3.1 Radiometric Source

3-8.3.1.1 Quantitative Calibration

The principle of operation for using a radiometric source is that the target flux incident on the detectors can be compared with the incident flux from a precision source. Since the detector responsivity is measured in

$$\frac{\text{output, V}}{\text{input irradiance, w}},$$

radiation which produces the same output voltage must have the same input power. Once having determined by calibration the power and stability of a test source, detector responsivity could be ascertained at will by irradiating the detectors with the calibrated source.

3-8.3.1.2 Qualitative Calibration

Under conditions where it is necessary to ensure periodically that the IR system is functioning, e.g., when targets of interest are infrequent and short-lived, a built-in test source (BITS) can be energized periodically to simulate a target signal. This permits stand-by checkout of the system. Source radiation waveform and spectral content are not as critical, therefore, a stable source will permit a one-point quantitative calibration.

Qualitative calibration of long-wavelength systems can often be implemented by viewing portions of the sensor mount or vehicle structure and surroundings, obviating the need for a BITS.

3-8.3.2 Simulated Electronic Signal

A stable electronic signal of the proper waveform can be used to determine the amplifier gain and dynamic range for detector signals of the same waveform. From this measurement, qualitative information can be inferred about the relative detectivity and response of the detector. However, since changes in detector dark-noise cannot be accurately related to either detectivity or responsivity for small changes in noise, the method serves only to judge the relative change of one detector to another in a

multiple array and to suspect data from detectors that exhibit large noise variances. In systems where several detectors of an array scan the same real target, knowing the actual amplifier gain of each channel will permit data correction and thereby detector calibration. For example, through a cross-comparison of levels, the average level of target irradiance can be determined more accurately than from a single measurement.

3-8.3.3 Types of BITS

3-8.3.3.1 Lamps

Tungsten lamps are stable graybody sources of IR irradiance spectrally limited by the envelope material or window material of specially built lamps. The usable dynamic range of irradiance is approximately three orders of magnitude when the lamp is operated from 10 percent to 130 percent of its nominal rated voltage. Two magnitudes of the dynamic range are over the range of 10 percent to 50 percent of the nominal voltage rating. The lower limit of dynamic range is determined by the required stability since at low voltage/filament temperatures, the ambient environment will have a greater effect on the apparent irradiance. (Thus, a stable, cooled, lamp housing could achieve another magnitude of lower level irradiance.) The upper level is limited by lamp life; however, since the voltage range from 50 percent to 130 percent of the nominal rating only accounts for about one-half the magnitude of high-level irradiance, lamp life is shortened more rapidly than irradiance is increased. Thus an upper long-life limit of 90 percent of the nominal operating voltage is recommended for tungsten lamp sources.

To increase the dynamic range (at the low end of irradiance) neutral density filters, reflective diffusers, or properly designed metering apertures can be used. (Transmissive diffusers for long wavelengths are not practical.)

Thin glass envelopes do not greatly attenuate the spectral irradiance of commercial production miniature lamps; therefore, these are by far the most reliable and uniform sources available. When a special envelope is required, extensive testing is necessary to satisfy the uniformity, i.e., interchangeability, and reliability design requirements.

The physical design of the lamp filament is crucial to its use as a test source. The most stable design consists of a single coil since it does not exhibit physical displacement when heated and it survives greater shock and vibration loads than does the flat ribbon or strand filament types.

The single disadvantage of the coiled filament type lamp is that the filament must be imaged directly onto a detector or reticle which is smaller than a group of coils since very small displacements will change the incident irradiance of the lamp. Where no practical design can overcome this condition, a ribbon or strand filament must be used. In this instance, the filament should be specially designed as a single element, coiled at each end of the straight imaged section to minimize sag and provide isolation from mechanical shock and vibration.

A controlled, constant-current power supply source is required to ensure that the lamp irradiance achieves the highest potential accuracy. The irradiance intensity can be varied by adjusting a reference potentiometer in the comparator circuit of the supply. In this manner, the lamp irradiance can be made relatively independent of the environment and the precision of irradiance may be expected to be in the order of 1 percent at medium and high levels, decreasing to 10 percent at low levels. Where the environment is expected to be sufficiently stable to ensure the required calibration accuracy, the use of low resistance (large conductor cable) and a well regulated power supply will be adequate.

3-8.3.3.2 Emitters

Emitter diodes are commonly used as calibration sources. Output radiance is proportional to the driving voltage over the specific operating range, beyond which internal heating will reduce the output. By reducing the duty cycle, the output can be increased to the point where the average heat dissipation is at the operating limit. Since the frequency response of emitters is in the megahertz region, the duty cycle will be limited by the response of the IR system. With dc or sine wave ac voltage, a dynamic range of two orders of magnitude may be realized. Cooling of the diode will permit doubling of the output at the high-level end while lowering the duty cycle will permit an increase of another one to two orders of magnitude at the high

level limited by the peak breakdown voltage of the diode.

The diodes are particularly suitable where a compact modulated source is required. Since the bandwidth of the diode response is several hundred times greater than the usual electrical bandpass of IR systems, diodes can provide brightness with virtually any required modulation waveform at any desired detector amplifier center frequency.

Typical emitter cw output at 25°C ranges from 10 to 20 μw in a 900Å-wide band centered at 6650Å. Modulation rise-times are in the order of 10 to 100 nanosec. Peak power can be increased to 100 w when operated at a low duty cycle of 3×10^{-4}. Other emitters are available for operation in the 4-micron region. Because of the wide variety of emitters and the rapid advances in emitter technology, manufacturers should be consulted for assistance.

Because the emitter output is in a narrow and relatively short wavelength IR spectral band, its use as a calibration source must be by direct radiation of the detectors to be calibrated, i.e., without having to pass through any short wavelength cutoff filters.

Where the system filter also serves as a window of the cooled detector cryostat, the emitter can be mounted within the cryostat, which will then also serve as a total integrator of the emitter energy.

3-8.3.3.3 Blackbody Sources

The temperature-controlled blackbody cavity is an excellent calibration source where space and weight can be accommodated.

Miniature blackbodies approximately 2 $in.^3$ in volume are commercially available with compact controllers approximately 8 $in.^3$ in volume. Because of the small cavity size, the exit aperture is limited to 1/16 in. in diameter and operating temperature up to 600°C. Special designs can generally be implemented to accommodate special configuration and temperature requirements.

Because of their small mass, blackbody temperatures can be rapidly changed, requiring no more than 2 to 3 minutes to stabilize for temperature changes from one extreme to the other. The dynamic range of the output flux (which is approximately 30 to 1) can be increased at the low-level end by metering apertures and neutral density filters. The temperature is controlled by means of a feedback loop that senses the voltage drop across the heater element.

Recently developed miniature blackbodies are much larger than miniature lamps and diodes, have relatively more complex electronic controllers, and require several minutes to reach ambient temperature after having been turned off. They are particularly useful in providing a relatively large uniform source of blackbody radiation. This is independent of ambient temperatures lower than 10°C of the set point. Stability of 0.1°C can be readily achieved.

3-8.3.3.4 Modulation of Blackbody Radiation

Continuous sources of radiance such as lamps and blackbody cavities require mechanical interruption or modulation of their emission at a frequency that is within the electrical bandpass of the IR system. Diode emission can be interrupted electrically and, thus, presents no special mechanical consideration since the purpose of the modulation is the same for all types of BITS. The specific method used for modulating BITS must be determined by the IR system design and calibration requirements; therefore no single design can be expected to be universally applicable.

The principle of operation and possible application of several different modulation methods is described in general terms. In addition, the reader is referred to a general discussion of chopper modulators contained in par. 3-8.2.3. There are three general types of BITS choppers: (1) transmissive, (2) reflective, and (3) combined transmissive-reflective.

Typical of the *transmissive chopper* is the simple slotted, opaque, rotary disc placed between the BITS and the detector so as to alternately pass and block the energy from the BITS. The duty cycle and waveform is a function of the relative dimensions of the slots, the opaque area, and the size of the BITS. In general, a square wave is approached when the BITS is small relative to the slots, and a sine wave is approached when the slots and the BITS are circles of the same diameter. The average frequency of the chopper is the speed of disc

rotation multiplied by the number of slots. Where the IR system utilizes a chopper, a reticle might be used also to modulate the BITS provided an ancillary optical system can be devised to image the BITS onto the reticle.

The *reflective chopper* is a two-sided, mirror-surface slotted disc so placed as to alternately reflect the target signal and the BITS signal onto the detectors. The slotted segments are generally uniform in order to provide symmetry between the two signal sources. Reflective choppers are also considered when the BITS is required to scan across the detector so as to approximate the scan of an actual target. In this manner, an effort is made to provide an accurate simulation of the real target waveform at various levels of irradiance. The scanning chopper is generally a cube (or other multi-sided reflector) that sweeps a reflected image of the BITS across an imaging calibration lens, or directly across the detector. The waveform is determined by the size of the BITS aperture, the focal distance of the lens, and the line spread function of the lens. By varying these physical dimensions, a wide variety of BITS slopes can be produced.

Combined transmissive-reflective choppers consist of a mirrored-surface slotted disc arranged so that the BITS is alternately reflected onto the detectors and the target signal is passed through the open slots. Generally, to provide symmetry of the two input signals, the slotted disc will be uniformly spaced. This type of chopper can be used in null type radiometers.

The BITS signal in a null radiometer is varied to match the level of the target signal. A null will result when the average energies of the two signals are equal.

3-8.3.4 Calibration of BITS

In general, reliance on analytical methods to determine the absolute level of irradiance from each detector channel of an IR system is unsatisfactory because of the multiple variables in the path between the BITS and the detectors. Nonetheless, analytical calculations are required to establish parameters for selection of individual components, lenses, chopper-modulator, power source, heat loads and source size, and to establish parametric levels against which the physical calibration can be checked. Differences in the order of 10 to 50 percent of the calculated values may result, depending on how accurately the calibration design has been modeled and analyzed. Realistic evaluation of target parameters during laboratory experiments entails simulating the target radiance level, spatial extension, spectral distortion, etc.

The factors to be considered in the analysis depend upon the particular calibration requirement. This may involve knowing one or more of the following:

1. Absolute level of radiance in the spectral band of interest for each level of power input to the BITS

2. The spatial extension and the geometry of the source object

3. The transfer function of the source optics

4. The optics transmission

5. The transmission path losses

6. The spatial transfer function of the detector

7. The temporal transfer function of the detector amplifier chain

8. The parameters affecting the characteristics of the data gathering system transfer function

9. The environmental effects on each of these factors

The physical method of calibration is universally applicable to all IR systems. It entails the following procedure:

1. The IR system is aligned to a collimated blackbody.

2. A system output voltage is recorded for various levels of blackbody radiance at a fixed aperture size.

3. A system output voltage is recorded for various levels of BITS radiance.

4. The BITS calibration is then the known radiance of the blackbody standard for each corresponding output voltage level of the IR system.

5. Each channel of a multichannel system requires an independent calibration for errors due to:

(a) Differences in size of the target

(b) Differences in the operating environment

(c) Differences in channel response of a multichannel system

(d) BITS variability

From the sensitivity analysis, factors affecting calibration can be evaluated and an estimate of the calibration accuracy can be made.

3-9 ANCILLARY IR COMPONENTS AND EMI REJECTION TECHNIQUES

The material in par. 3-9 concerns the various electromechanical components associated with mounting and aiming IR sensors, monitoring their output, and protecting them from direct exposure to sunlight. Finally, the electromagnetic interference and its attendant

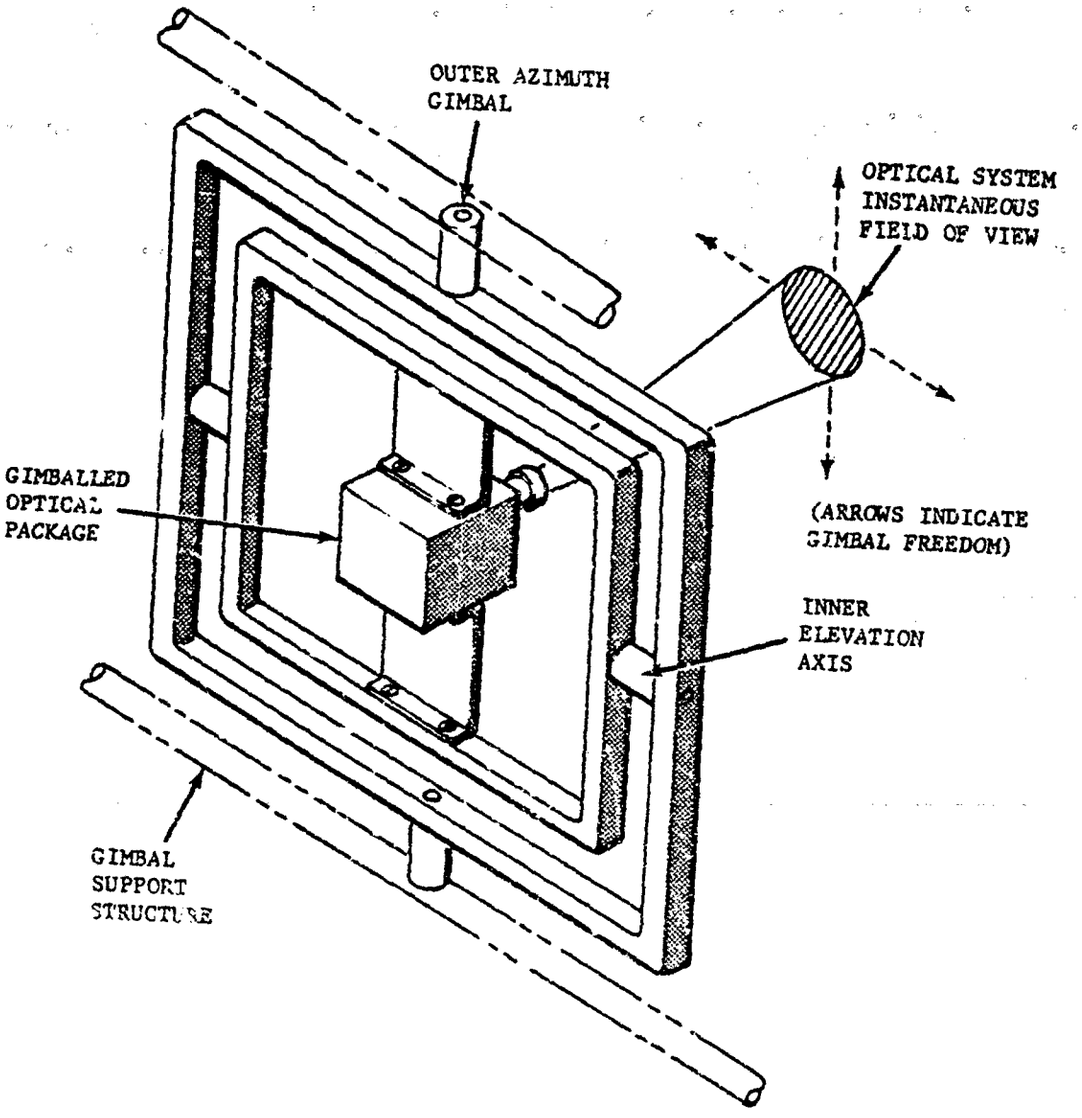


FIGURE 3-149. Conceptual Layout of Two-axis Gimbal Assembly

problems are discussed with respect to its effect on electrical and electronic systems used with IR equipment.

3-9.1 GIMBALS

Nearly all IR systems for the military are used to perform search, track, or measurement functions or combinations thereof. In each case the field of view of the system is usually small in order to achieve the high sensitivity required. Therefore, a suitable method must be used for pointing the IR sensor of such a system with the necessary accuracy. The gimbal serves this purpose by supporting the sensor on whatever platform it is mounted and controlling the angular position of its boresight axis. Some gimbals used in terminal guidance systems may be required to operate over a range of only a few degrees, while others such as those used with search sets may be required to cover a full 360-deg azimuth.

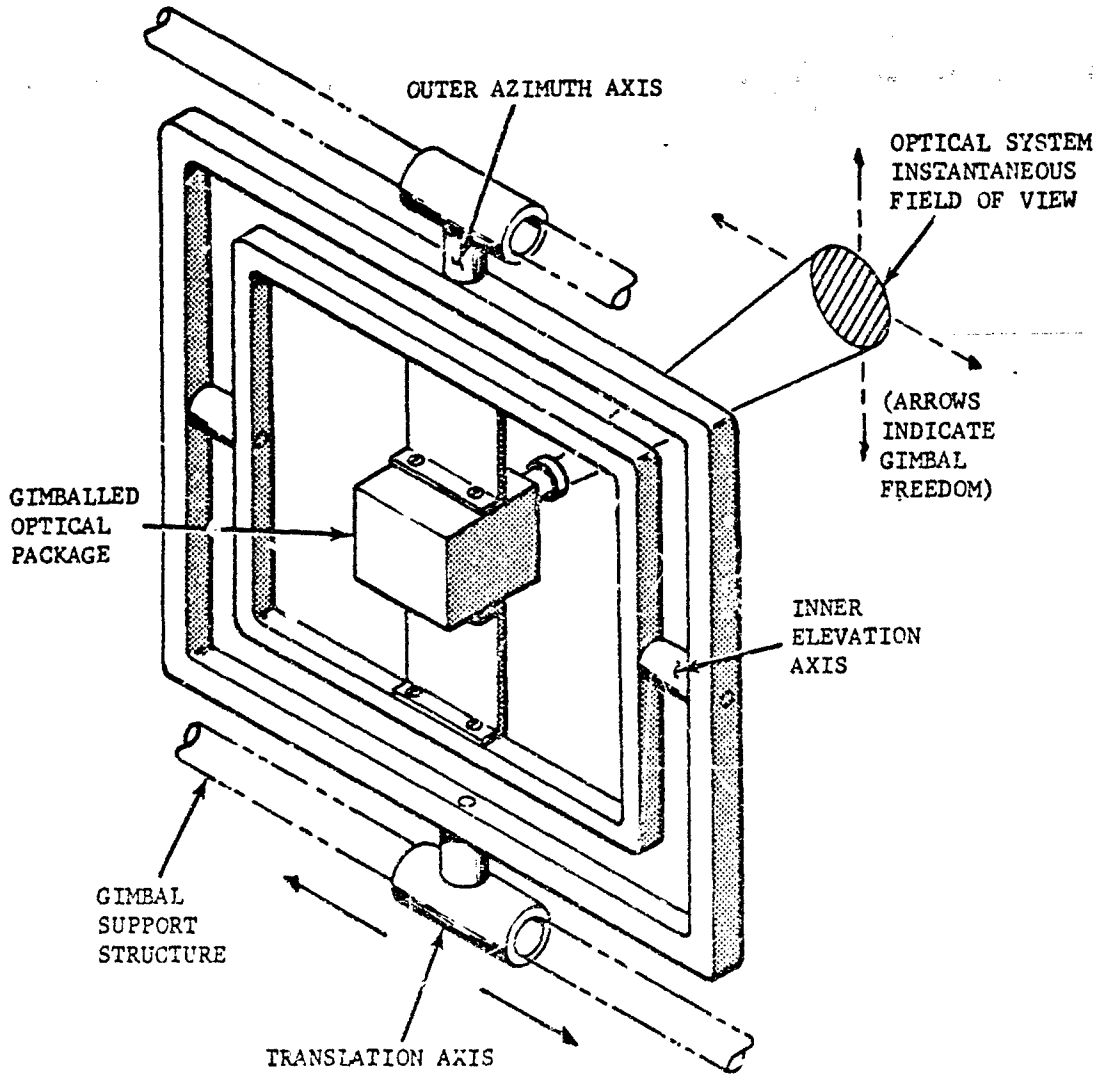
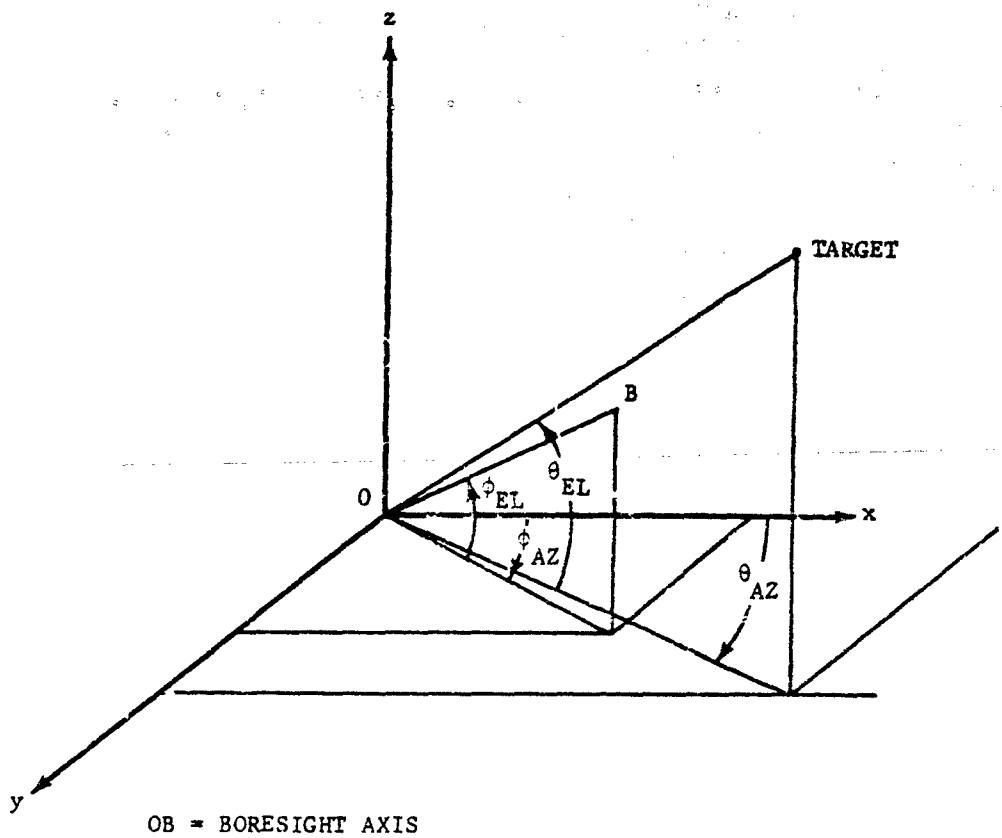


FIGURE 3-150. Conceptual Layout of Three-axis Gimbal Assembly

In the case of the two-axis gimbal (Fig. 3-149) the boresight axis of the optical package can move in either or both directions. Some special requirements may dictate the use of more than two independent axes (Fig. 3-150). For example, certain optical systems operating from inside an aircraft must track point-source targets through the aircraft windows. Consequently, when the line-of-sight to the target describes a large angle with respect to the center line of the tracker-aircraft window, there is a possibility of radiation blockage due to the finite size of the window. In order to minimize this blocking, gimbals capable of translating parallel to the window are used to provide less obstructed coverage of the field of view. Fig. 3-150 illustrates a gimbal which has one translational and two rotational axes. Systems having four degrees of freedom (two rotational and two translational) have been built[94] for the purpose of minimizing blockage.

The design of a gimbal system usually begins with the definition of the applicable coordinate system and associated terminology. The system requirements—i.e., tracking accuracy, tracking rate, and acceleration—must then be established. Next, components are selected and the



OB = BORESIGHT AXIS
Φ = BORESIGHT AXIS WITH RESPECT TO INSTRUMENT PLATFORM
θ = TARGET ANGLE WITH RESPECT TO INSTRUMENT PLATFORM
AZ = AZIMUTH
EL = ELEVATION

FIGURE 3-151. Relationship of Gimbal and Target Coordinates

system is analyzed and modified, if required, to ensure that the system meets the performance requirements.

3-9.1.1 Coordinate System

Fig. 3-151 illustrates a typical gimbal coordinate system. The coordinate system x, y, z is fixed to the platform upon which the IR system will be mounted. Line OB represents the boresight axis of the IR sensor (usually the center of the field of view). Angle ϕ denotes angles of OB with respect to the platform and θ denotes angular coordinates of the target with respect to the platform. The subscripts AZ and EL denote azimuth and elevation, respectively.

3-9.1.2 Servo Loop

Fig. 3-152 is a diagram of a typical servo loop used for positioning the gimbal used with a typical IR search or measurement system. The program source generates a signal which the boresight axis is to follow. In the case of a search system, the program signal would describe a search pattern. In a measurement system, the program signal might come from the gimbal of a tracker, thereby, slaving the measurement system to the tracker. The angle measurement circuit in the feedback loop provides the position angle of the boresight axis with respect to the platform. The program and boresight axis position signals are compared to produce an error signal. The error signal is then amplified and applied to a motor which drives the boresight axis toward the target. The load on the motor consists of the optics, detector(s), cooling system if required, and usually part of the electronics. These components are described in pars. 3-2 through 3-6. Often, the search pattern in one axis is a constant angular rate, and a rate feedback is used as illustrated in Fig. 3-153. Here, a constant angular rate ϕ is generated by a constant command input.

Fig. 3-154 is a diagram of a typical servo loop used with a tracking system. The reference input is the target angle θ and the output ϕ is still the boresight axis. The processing portion of the tracker generates signals which correspond to the position of the target in the field of view. As before, these signals are used to drive the boresight axis toward the target.

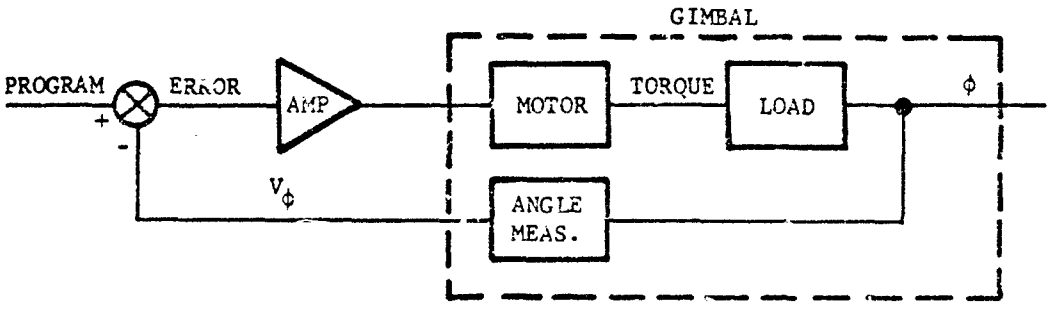


FIGURE 3-152. Servo Loop for Search or Measurement System

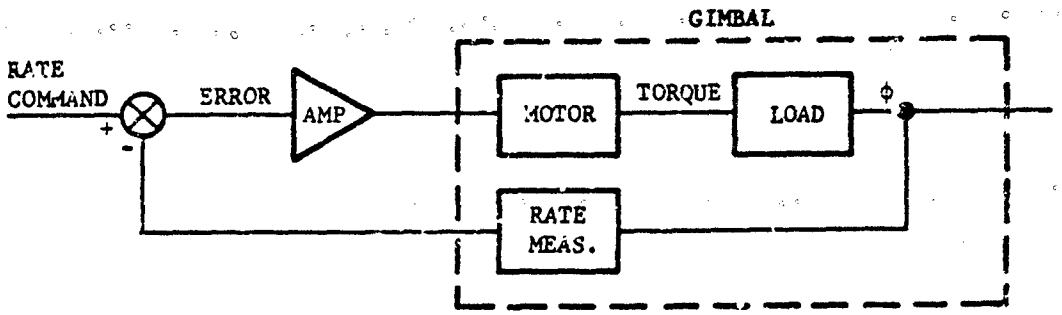


FIGURE 3-153. Servo Loop for Beta Search

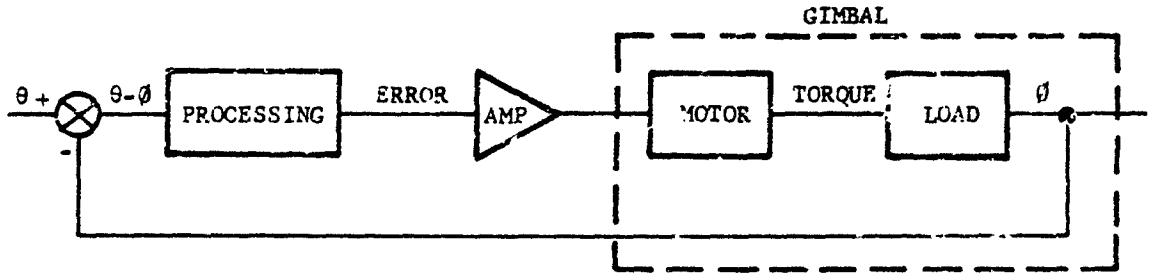


FIGURE 3-154. Servo Loop for Tracker

3-9.1.3 Specifications

The purpose of a gimbal is to cause the boresight axis of the instrument mounted thereon to conform as closely as possible to the direction provided by a command signal input. The performance criterion of the gimbal is usually specified in terms of the error between the boresight axis and the command signal. Because the input and output of the servo are time varying, the error is also a time varying quantity; consequently, the error specification usually includes a transient specification and a steady-state error limit.

3-9.1.3.1 Steady-state Errors

For the simple servo loop illustrated in Fig. 3-155, the transfer function is given as the Laplace transform of the response of the system to a unit impulse function (δ function). The following equations describe the system:

$$E(s) = R(s) - C(s) \qquad (3\text{-}201)$$

$$C(s) = E(s)G(s) \qquad (3\text{-}202)$$

where

$E(s)$ = Laplace transform of the error signal

$R(s)$ = Laplace transform of the input signal

$C(s)$ = Laplace transform of the output signal

$G(s)$ = transfer function of the system

Solving for $C(s)$

$$C(s) = \frac{G(s)R(s)}{1 + G(s)} \qquad (3\text{-}203)$$

Solving for $E(s)$

$$E(s) = \frac{R(s)}{1 + G(s)} \qquad (3\text{-}204)$$

Using the final value theorem[35]

$$\lim_{t \to \infty} e(t) = \lim_{s \to 0} sE(s) = \lim_{s \to 0} \frac{s\,R(s)}{1 + G(s)} \qquad (3\text{-}205)$$

Eq. 3-205 describes the steady-state error which remains after all the transients have become negligibly small.

The inputs $R(s)$ normally used to specify the system performance are

1. unit step $R(s) = \frac{1}{s}$
2. unit ramp $R(s) = \frac{1}{s^2}$
3. unit parabolic function $R(s) = \frac{1}{s^3}$

For most gimbals the input and output signals represent angular positions such that: the unit step input corresponds to a sudden shift in the input angular command; the unit ramp is a sudden application of a constant angular rate or velocity command; and, the unit parabolic function is the sudden application of a constant angular acceleration.

Substituting the unit step for $R(s)$ in Eq. 3-205

$$\lim_{s \to 0} sE(s) = \lim_{s \to 0} \frac{1}{1 + G(s)} = \frac{1}{1 + \lim_{s \to 0} G(s)} \qquad (3\text{-}206)$$

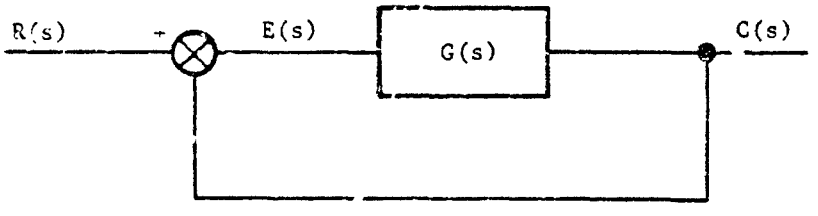


FIGURE 3-155. Servo Loop Diagram

Substituting the unit ramp

$$\lim_{s\to 0} sE(s) = \lim_{s\to 0} \frac{\frac{1}{s}}{1 + G(s)} = \frac{1}{\lim_{s\to 0} s\,G(s)} \tag{3-207}$$

Substituting the unit parabolic function

$$\lim_{s\to 0} sE(s) = \lim_{s\to 0} \frac{s^{1/2}}{1 + G(s)} = \frac{1}{\lim_{s\to 0} s^2\,G(s)} \tag{3-208}$$

The values derived in Eqs. 3-206 through 3-208 are defined as error constants:

$K_p = \lim_{s\to 0} G(s)$ = position constant

$K_v = \lim_{s\to 0} s\,G(s)$ = velocity constant

$K_a = \lim_{s\to 0} s^2\,G(s)$ = acceleration constant

The transfer function can be written in general factored form as follows:

$$G(s) = \frac{(s + Z_1)(s + Z_2)(s + Z_3)\ldots}{s^n(s + P_1)(s + P_2)(s + P_3)\ldots} \tag{3-209}$$

The value of n in Eq. 3-209 reveals a great deal of information about the steady state characteristics of the system. For example, if

$$n = 0$$

$$K_p = \lim_{s\to 0} G(s) = \frac{Z_1 Z_2 Z_3 \ldots}{P_1 P_2 P_3 \ldots} \tag{3-210}$$

and K_p has a finite non-zero value. A system in which $n = 0$ is known as a Type 0 system. For a Type 0 system the steady-state error is given by Eq. 3-206 and results in a finite non-zero value.

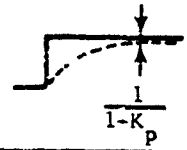
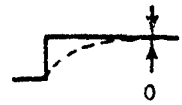
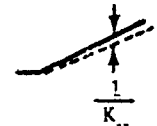
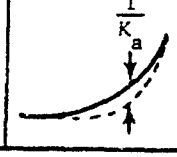
TYPE	K_p	K_v	K_a	OUTPUT FOR UNIT STEP IN	OUTPUT FOR UNIT RAMP IN	OUTPUT FOR UNIT ACCEL. IN
0	K_p	0	0			
I	∞	K_v	0			
II	∞	∞	K_a			

FIGURE 3-156. Error Constants and Responses of Type 0, I, and II Servos

Calculating K_v for a Type 0 system

$$K_v = \lim_{s \to 0} s\,G(s) = 0 \qquad (3\text{-}211)$$

For a Type 0 system, Eq. 3-207 shows that the steady-state error for a unit ramp input is infinite.

Systems whose transfer function (Eq. 3-209) show $n = 1$ and $n = 2$ are known as Type I and II systems, respectively. The error constants and steady-state errors for Type 0, I, and II systems are illustrated in Fig. 3-156 to permit direct comparison among the different types of systems for various inputs.

IR sensors used with search systems often are required to perform a search pattern at a constant rate. The search rate is usually important in determining the spectrum of the detector output which in turn determines the filtering required. Thus, the servo specification is often expressed in terms of a maximum error allowable while the gimbal moves at a constant rate. This describes the role of a Type I servo system and also determines the velocity constant required. As indicated in Fig. 3-156 the output rate is not equal to the input rate for a Type 0 system.

The specifications governing IR systems used for tracking often impose a constant position error with a constant input rate. Here again a Type I servo system would be appropriate. However, there may be additional considerations involved. For example, if the tracker has a very small field of view, the signal-to-noise ratio may be seriously degraded when the target is even

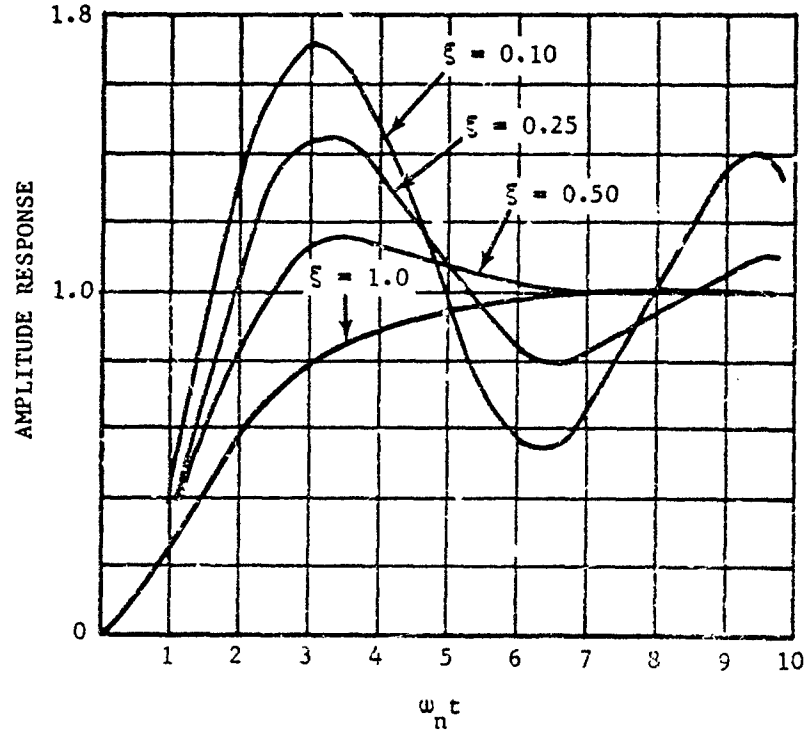


FIGURE 3-157. Servo Response of $\frac{\omega_n^2}{s^2 + 2\xi\omega_n s + \omega_n^2}$ to Unit Step Input for Various Damping Factors

slightly off center. Thus, it may be desirable to use a Type II servo in order for the tracker to operate at a constant velocity with zero steady-state-error.

IR systems used with measurement equipment may vary considerably in the performance required. For a simple IR camera, a Type 0 system may be sufficient if its function is merely to position the camera and if tracking of moving targets is not required. Many measurement systems may be slaved to a tracker which is tracking a moving target. These usually require a finite position error with an input rate. Still others which have an extremely small field of view may require a Type II servo for the same reason as in tracker operations.

3-9.1.3.2 Transient Response

Usually the response of a servo system is largely controlled by a few poles which determine the low-frequency characteristics of the transfer function. Thus, servos are often specified by natural frequency and damping ratio. The closed-loop response of the system is calculated by solving Eqs. 3-201 and 3-202:

$$\frac{C(s)}{R(s)} = \frac{G(s)}{1 + G(s)} \tag{3-212}$$

If $\frac{C(s)}{R(s)}$ is assumed to be represented by a second-order equation, it is convenient to write it in the form

$$\frac{C(s)}{R(s)} = \frac{\omega_n^2}{s^2 + 2\xi\omega_n s + \omega_n^2} \tag{3-213}$$

where

ξ = damping factor

ω_n = resonant frequency

Then the response of the system to a unit step function input is expressed as

$$C(s) = \frac{\omega_n^2}{s(s^2 + 2\xi\omega_n s + \omega_n^2)} \tag{3-214}$$

The output signal $C(t)$ is obtained by taking the inverse transform of Eq. 3-214

$$C(t) = \frac{1}{\omega_n^2}\left|1 + \frac{1}{\sqrt{1-\xi^2}}\, e^{-\xi\omega_n t} \sin\left[\omega_n\sqrt{1-\xi^2}\; t - \mathrm{Tan}^{-1}\sqrt{\frac{1-\xi}{-\xi}}\right]\right| \tag{3-215}$$

which is plotted in Fig. 3-157 for various values of the damping factor ξ.

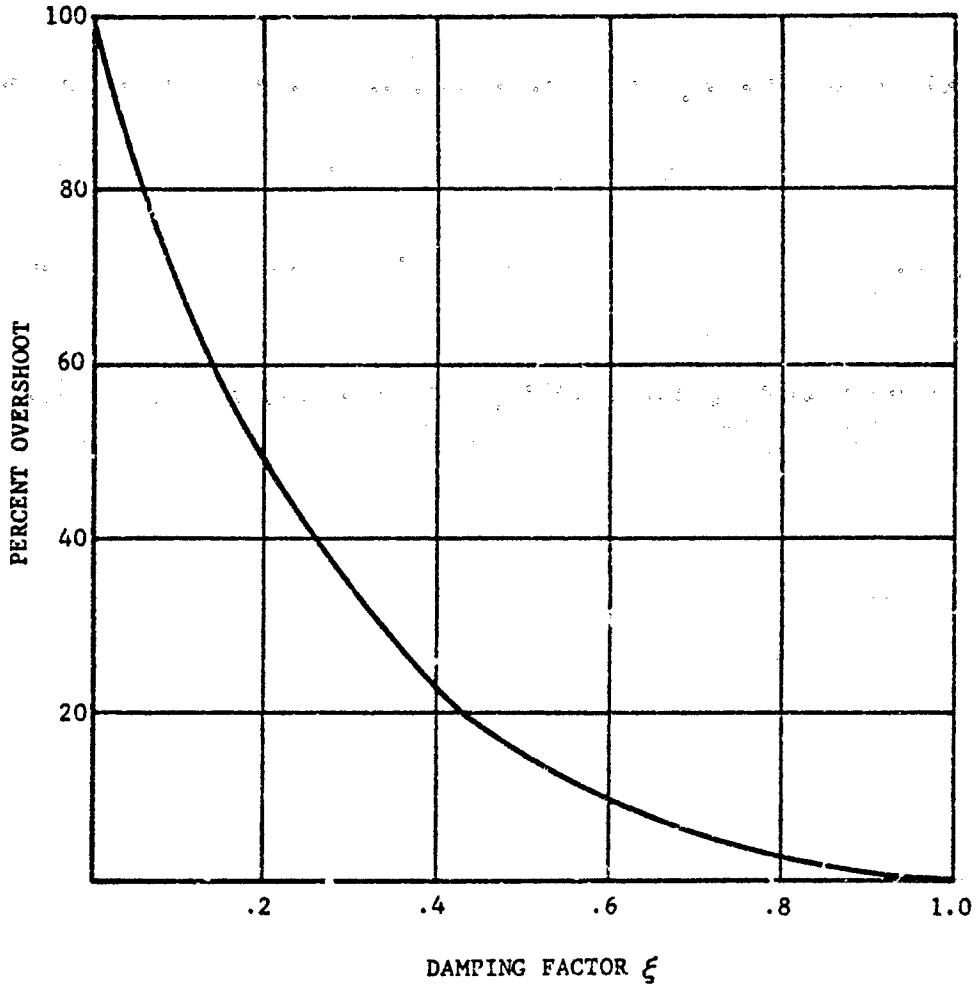


FIGURE 3-158. Overshoot vs Damping Factor for Second-order System

It can easily be seen from Fig. 3-157 that for small values of ξ the output response reaches the input value of unity much faster but it also overshoots more. The overshoot versus damping factor is plotted in Fig. 3-158. Thus, a compromise must be made between a fast initial response and more overshoot. A value of ξ greater than 0.8 usually results in a loop gain which is lower than required while values less than about 0.5 usually make a system excessively oscillatory and subject to instability. Thus a value of ξ of about 0.6 or 0.7 represents a satisfactory compromise. From Eq. 3-215 it can be seen that the first maximum in the response is at $\omega_n t = \frac{\pi}{\sqrt{1-\xi^2}}$.

A typical specification requirement for a servo system might be for a maximum of 10 percent overshoot and may also require that the output be within 10 percent of its steady-state value when a unit step-input within 1 sec is applied. Fig. 3-158 illustrates that a damping factor of $\xi = 0.6$ will meet the overshoot specification. Interpolation in Fig. 3-157 shows that for $\xi = 0.6$ the amplitude reaches 0.9 at $\omega_n t = 2.3$. Thus for $t = 1$ sec, the resonant frequency $\omega_n = 2.3$ rad/sec.

3-9.1.4 Design Considerations

Usually the servo design and the procurement of certain components must be started before the IR system design is complete. Thus, certain

estimates must be made on which to base the servo design. These estimates generally include:

1. The moment of inertia of the gimballed package
2. The maximum acceleration required

Other system performance specifications which are required in the servo design and which are usually available from the system specifications include:

1. Angular accuracy
2. Maximum angular rate
3. Accuracy at a tracking rate

Whether the IR system is to be used in performing a search, track, or measurement function the target motion usually determines these three system specifications. If the potential target is stationary, the information which satisfies specifications 1 and 2 may be sufficient. For systems which are required to track moving targets, data for all three are required.

3-9.1.4.1 Example Design Problem

In the paragraphs which follow, a typical design problem consisting of a dc torquer with tach generator feedback will be solved to illustrate the application of some of the design and analysis techniques. The specifications selected include a tracking accuracy of 2 mrad while tracking at 100 mrad/sec, and the ability to attain a rate of 100 mrad/sec in 0.05 sec.

3-9.1.4.1.1 Moment of Inertia

The moment of inertia estimate is based on a simple tracker having a 6-in. diameter aperture optics and consisting of an 8-in. diameter 1/4-in. thick, 1-ft long aluminum cylinder with 2 lb of equipment at each end. The 2-lb weight at each end is intended to keep the gimballed package balanced. Mass unbalance is so undesirable that provisions are usually made for adding small weights to the gimballed package after the construction is completed in order to achieve precise balance. Any mass unbalance will appear to the servo loop as a torque disturbance, thereby, causing errors in the loop. Furthermore, the torque disturbance input to the servo loop is directly proportional to any g loads imposed upon the system.

According to Ref. 96 the moment of inertia J_{cyl} of a thin-walled hollow cylinder is

$$J_{cyl} = m\left(\frac{r^2}{2} + \frac{\ell^2}{12}\right), \text{ slug-ft}^2 \qquad (3\text{-}216)$$

where

m = mass, slug

r = radius, ft

ℓ = length, ft

The volume of the hollow cylinder is approximately $V_{cyl} = 2\pi r t_t \ell$ where t_t is the thickness. The density of aluminum is 169 lb/ft^3 (Ref. 97). Thus, the cylinder would weigh 7.35 lb and its mass would be 0.23 slug. Substituting these values into Eq. 3-216, $J_{cyl} = 0.032$ ft-lb-sec^2. The moment of inertia of the 2-lb loads at the ends is 0.0313 ft-lb-sec^2. Thus, the total moment of inertia is 0.0633 ft-lb-sec^2. Since this is only an estimate, a value of $J = 0.1$ ft-lb-sec^2 will be used.

3-9.1.4.1.2 Acceleration

In general, the maximum acceleration must be sufficient to satisfy all of the requirements imposed on the system. The requirement of 100 mrad/sec in 0.05 sec requires an acceleration rate of 2 rad/sec^2.

3-9.1.4.1.3 Component Selection

The main consideration in the selection of a torque motor is the maximum torque required. For the system with a maximum acceleration of 2 rad/sec^2 and a moment of inertia of 0.1 ft-lb-sec^2, a maximum torque of 0.2 ft-lb is required. However, it is usually desirable to select a motor whose torque capability is several times the maximum required. One reason being that, in the early stages of design, it is very difficult to predict accurately the moment of inertia which is a factor that can easily increase substantially. Also systems with a high-velocity constant usually require wide bandwidth operation in order to achieve high gain while still maintaining the required system stability. Since bandwidth is related to the torque-to-inertia ratio, it is desirable to select a motor with the necessary high torque rating. For these reasons, a motor rated at 1 ft-lb would be used. According to Ref. 93 the specifications of a typical motor are as follows:

Maximum torque	1 ft-lb
Torque constant K_T	0.23 ft-lb/A
Back emf K_B	0.31 V/rad/sec
Winding resistance R	4.3 ohm
Winding inductance L	0.01 H (henry)
Weight	1.6 lb
Moment of inertia J_{MOT}	0.27×10^{-3} ft-lb-sec^2
Friction torque	0.02 ft-lb

The first block of the diagram shown as Fig. 3-159 represents the electrical circuit of the motor. The term R_T is the sum of the resistances of the motor and the output resistance of the amplifier. Since the output resistance of a typical amplifier is about 0.2 ohm, $R_T = 4.5$ ohms. The term L represents the motor inductance. The next block represents the conversion of motor current into torque. The summing junction allows the application of a disturbance torque. The next block represents the conversion of torque into output angle. The term in the feedback path is the application of the back emf to the voltage input to the motor.

The complete transfer function of the motor is calculated :

$$G(s) = \frac{\theta}{e} = \frac{\frac{K_T}{JL}}{s\left(s^2 + \frac{R}{L}s + \frac{K_T K_B}{JL}\right)} \qquad (3\text{-}217)$$

where

θ = output torque angle, rad

e = input voltage disturbance, V

It is convenient to factor Eq. 3-217 into the following form

$$G(s) = \frac{\frac{K_T}{JL}}{s\left(s + \frac{R}{L}\right)\left(s + \frac{K_T K_B}{JR}\right)} \qquad (3\text{-}218)$$

This is possible if $\frac{R}{L} \gg \frac{K_B K_T}{JR}$ which is true for most dc motors.

For purposes of the example problem, spatial rate is not involved; therefore, a tachometer generator will be used in the analysis. A typical tachometer generator has the following specifications:

Voltage sensitivity K_G	1 V/rad/sec
Generator resistance R_G	215 ohms
Generator inductance L_G	0.28 H
Weight	13.5 oz
Moment of inertia J_{TG}	0.028 oz-in.-sec^2
Friction	2 oz-in.

The equivalent circuit of a tachometer generator is shown in Fig. 3-160. The block diagram of Fig. 3-161 shows the tachometer generator in Laplace transform notation. The load resistance R_L presented to the tachometer generator is usually made large in order that the electrical pole $\frac{R_L + R_G}{L}$ will be very large and most of the voltage $K_G\left(\frac{d\theta}{dt}\right)$ will appear at the output. For the example, if R_L = 50 kohm, the electrical pole of the tachometer generator will be 178,000 rad/sec, which is well beyond the range of the other poles in the system.

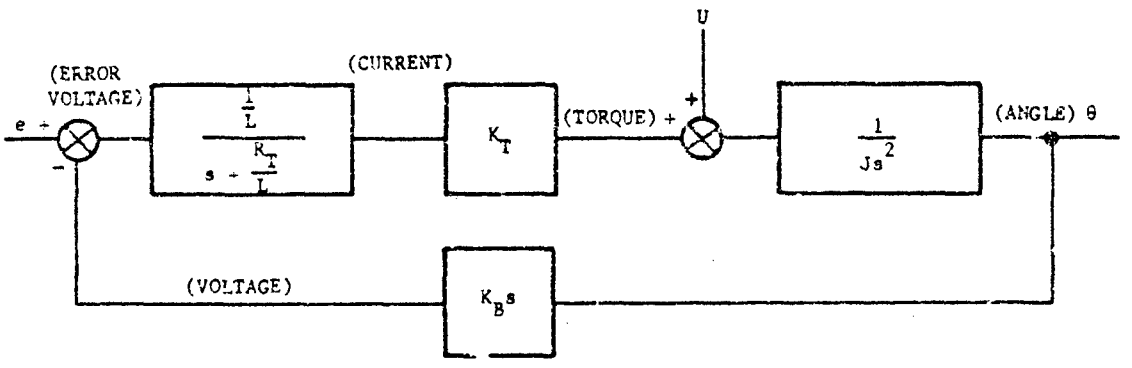


FIGURE 3-159. Block Diagram of dc Torque Motor

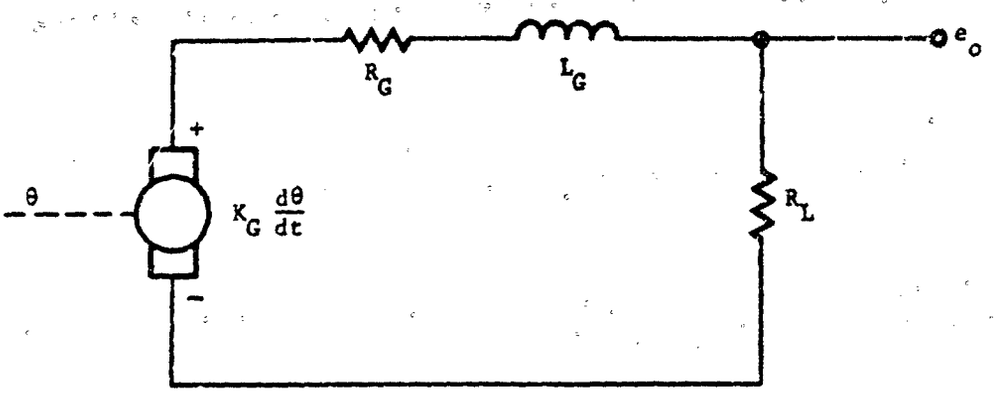


FIGURE 3-160. Equivalent Circuit of Tachometer Generator

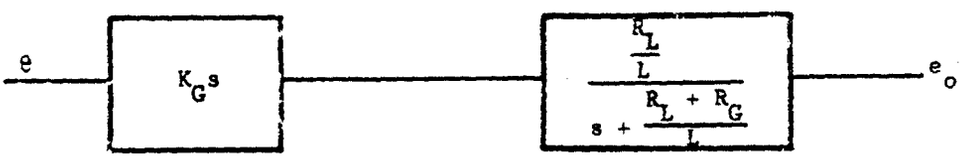


FIGURE 3-161. Block Diagram of Tachometer Generator

3-9.1.4.1.4 Analysis

In order to illustrate some of the analysis methods, the system shown in Fig. 3-162 will be analyzed, and the gains K_1 and K_2 will be set for proper operation using the root locus method.

3-9.1.4.1.4.1 Rate Loop

The system represented in Fig. 3-162 has two loops. The inner (rate) loop is such that for a constant input R_2, a constant output rate $c(t)$ will be obtained. This loop can also be used to obtain a constant-velocity search pattern. In this analysis $C(s)$ is the Laplace transform of $c(t)$, the two terms conventionally used for frequency domain and time domain, respectively, in electrical engineering.

The first step in the analysis is to set the gain K_2. The closed-loop response for the inner (rate) loop is

$$\frac{C(s)}{R_2(s)} = \frac{K_2 G(s)}{1 + K_2 G(s) K_G s} \qquad (3\text{-}219)$$

Now substituting Eq. 3-218 into Eq. 3-219, we have

$$\frac{C(s)}{R_2(s)} = \frac{K_2 K_T/(JL)}{s\{(s + R/L)[s + K_E K_T/(JR)] + K_2 K_T K_G/(JL)\}}$$

Now substitute the values, previously given, for the motor and tachometer generator, and the moment of inertia from par. 3-9.1.4.1.1

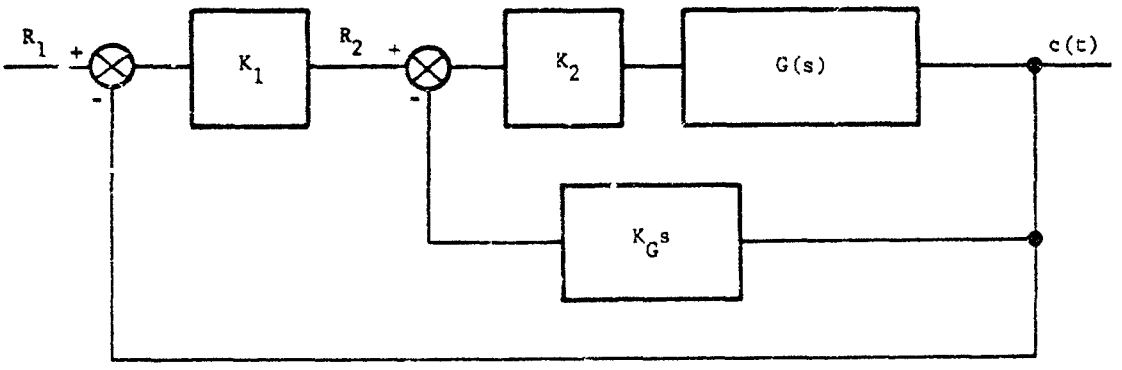


FIGURE 3-162. Block Diagram of System in Example Problem

$$\frac{C(s)}{R_2(s)} = \frac{\dfrac{0.23K_2}{0.1(0.01)}}{s\left\{\left(s + \dfrac{1.3}{0.01}\right)\left[s + \dfrac{0.31(0.23)}{0.1\ (4.3)}\right] + \dfrac{0.23(1)K_2}{0.1\ (0.01)}\right\}}$$

$$= \frac{230K_2}{s[(s + 430)(s + 0.166) + 230K_2]}$$

$$= \frac{230K_2}{s(s^2 + 430s + 71.3 + 230K_2)}$$

The root locus is shown in Fig. 3-163. If we assume a damping factor of $\xi = 0.707$, the complex conjugate poles are at $s = -215 \pm j215$. In this case the most direct way to solve for K_2 is to note that the denominator of the right-hand side of Eq. 3-219 is of the form $s[(s + \alpha)^2 + \beta^2)]$ with $\alpha = 215$, $\beta = 215$ which gives a value of $K_2 = 405$.

Thus

$$\frac{C(s)}{R_2(s)} = \frac{230(405)}{s[(s + 215)^2 + (215)^2]} \qquad (3\text{-}220)$$

$$= \frac{9.3 \times 10^4}{s[(s + 215)^2 + (215)^2]}$$

In order to get an idea of the transient response of the rate loop, let $R_2(s) = \dfrac{0.1}{s}$. This is a step function whose amplitude will command an output rate of 100 mrad/sec.

$$C(s) = \frac{(0.1)(9.3 \times 10^4)}{s^2[(s + 215)^2 + (215)^2]} \qquad (3\text{-}221)$$

It is more illustrative if the output rate $\dot{c}(t)$ rather than the output is plotted. Since $sC(s)$ is the transform of $\dot{c}(t)$ then Eq. 3-221 can be written as

$$\mathcal{L}[\dot{c}(t)] = \frac{(0.1)(9.3 \times 10^4)}{s[(s + 215)^2 + (215)^2]} \qquad (3\text{-}222)$$

Taking the inverse transform[25]

$$\dot{c}(t) = 0.1[1 + 1.41e^{-215t}\sin(215t - 135°)] \qquad (3\text{-}223)$$

Eq. 3-223 is plotted in Fig. 3-164.

It is also important to know the transient response of the motor torque. Since torque is $J\ddot{c}(t)$, therefore, from Eq. 3-222 we find

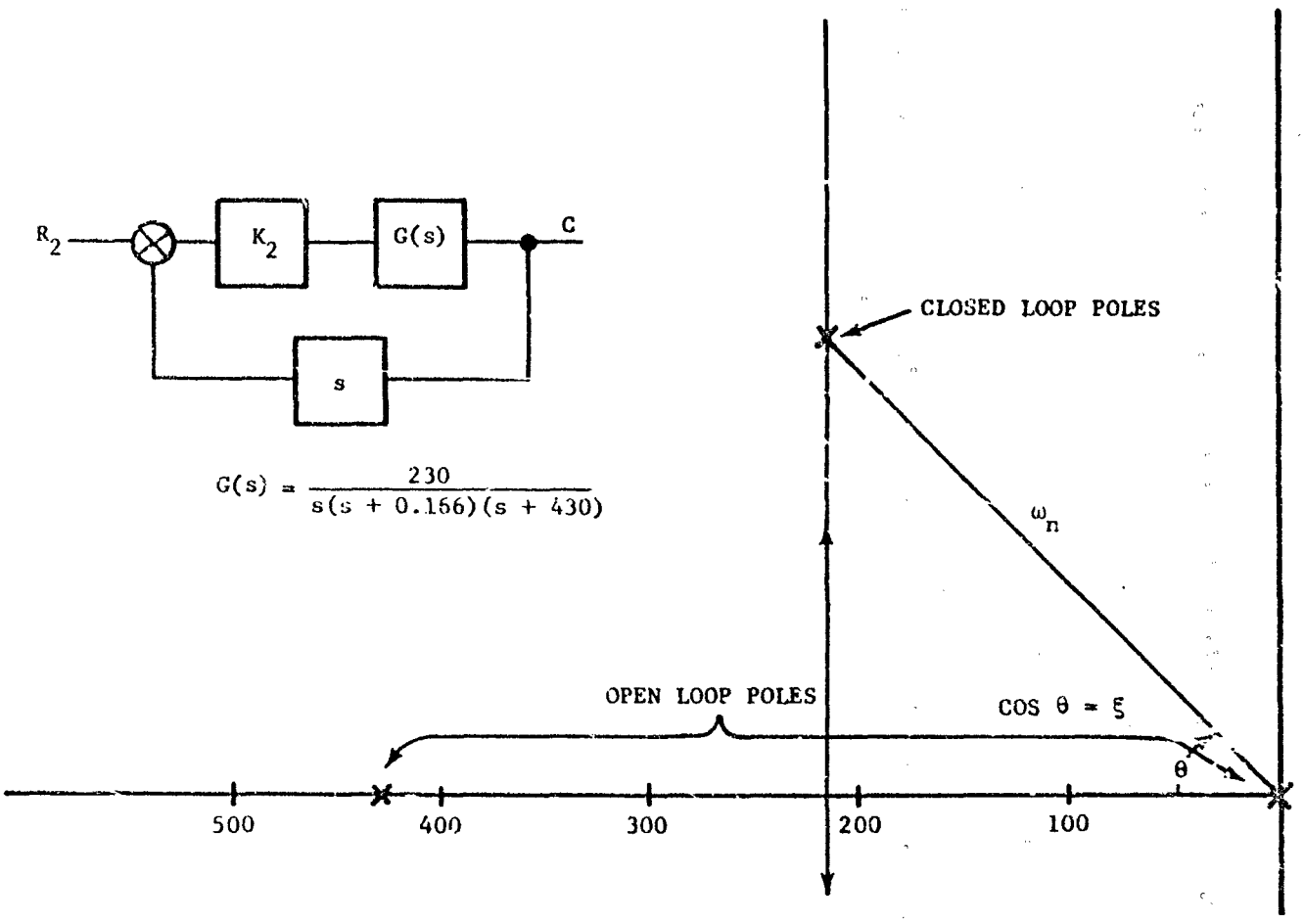


FIGURE 3-163. Root Locus for Inner (Rate) Loop

$$\mathcal{L}[T(t)] = [J\ddot{c}(t)]$$

$$= Js^2 C(s)$$

$$= \frac{(0.1)(0.1)(9.3 \times 10^4)}{(s + 215)^2 + (215)^2} \quad (3\text{-}224)$$

The inverse transform [95] is

$$T(t) = 4.32e^{-215t}\sin(215t) \quad (3\text{-}225)$$

Eq. 3-225 is plotted in Fig. 3-164. The maximum torque is 1.4 ft-lb, indicating that a step function input of 100 mrad/sec saturates the system slightly.

3-9.1.4.1.4.2 Position Loop

Next, the root locus of the outer position loop is drawn (Fig. 3-165). The poles of $\frac{C(s)}{R_2(s)}$ become the open loop poles of the outer loop.

In order to obtain reasonably high gain with reasonable stability, the value of $\xi = 0.607$ is selected, resulting in poles at $s = -130 \pm j170$. The closed-loop response for the outer loop is—using the value of $\frac{C(s)}{R_2(s)}$ from Eq. 3-220 since the rate loop has already been analyzed—

$$\frac{C(s)}{R_1(s)} = \left[\frac{\dfrac{9.3 \times 10^4 K_1}{s(s^2 + 430s + 9.3 \times 10^4)}}{1 + \dfrac{9.3 \times 10^4 K_1}{s(s^2 + 430s + 9.3 \times 10^4)}}\right]$$

$$= \frac{93000K_1}{s^3 + 430s^2 + 93000s + 93000K_1} \quad (3\text{-}226)$$

In order to obtain a reasonably high gain with reasonable stability, a value of $\xi = 0.707$ is selected which results in complex poles at $s = 130 \pm j170$; therefore, it can be written

$$(s + \alpha)[(s + 130)^2 + (170)^2] = (s + \alpha)(s^2 + 260s + 45900)$$

Then, by dividing $s^2 + 260s + 45900$ into $s^3 + 430s^2 + 93000s + 93000K_1 = s + 170$. Also, $93000K_1 = (45900)(170)$ and $K_1 = 84$. Now the total closed-loop response can be written as

$$\frac{C(s)}{R_1(s)} = \frac{7.8 \times 10^6}{(s + 170)[(s + 130)^2 + (170)^2]} \quad (3\text{-}227)$$

from which velocity constant of the system is calculated as

$$K_v = \lim_{s \to 0} sK_1 \frac{C(s)}{R_2(s)} = 84$$

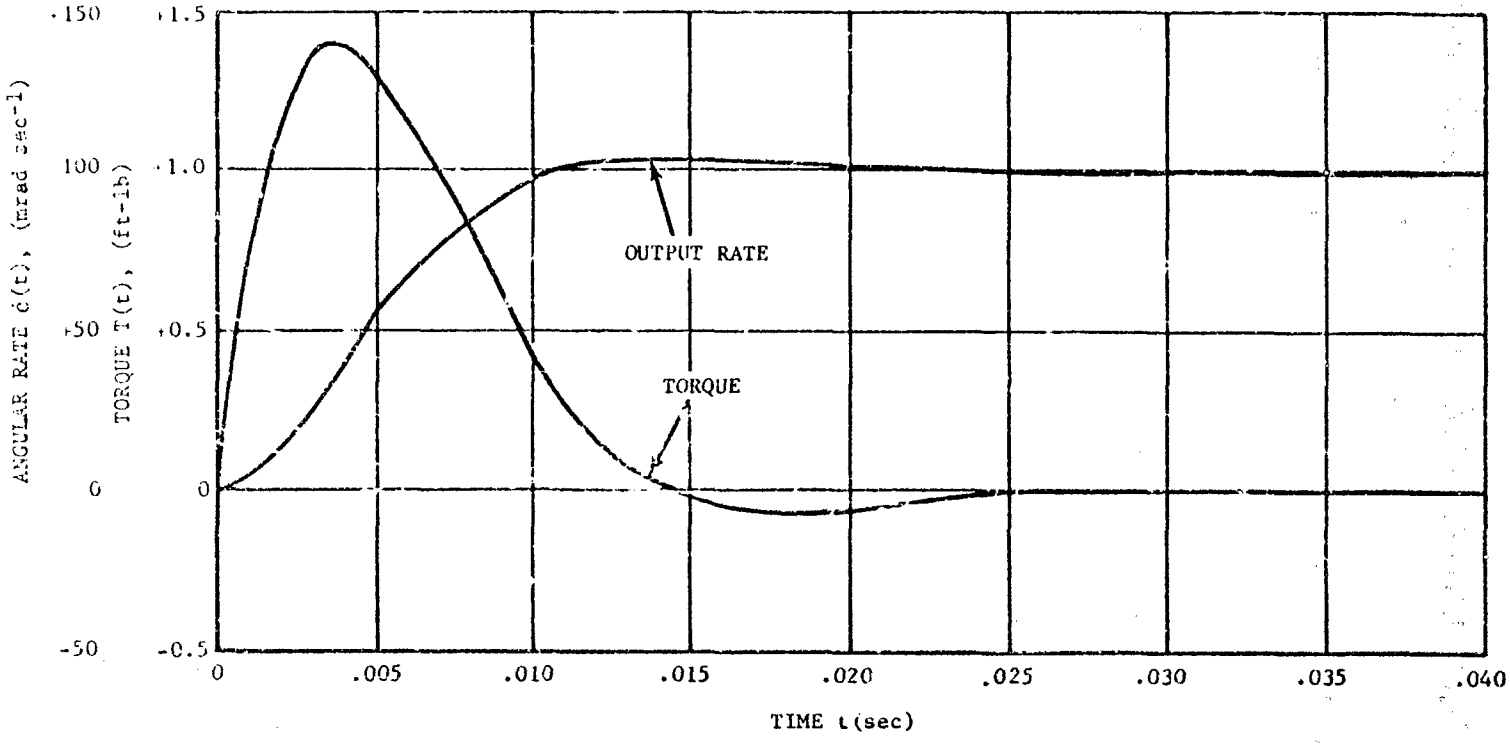


FIGURE 3-164. Output-Rate and Torque Respon.e for Step-function Input to Rate Loop

The result is better than the value of 50 required by the specification of 2 mrad accuracy while tracking at a rate of 100 mrad/sec.

Note that for a tracker, the gain component K_1 would consist of the error processing or that part of the optics, optical filtering, detectors, and electronics that converts the angular error into voltage. In general, this part of the system will include at least one pole which is likely to be closer to the origin than the complex poles of $\frac{C(s)}{R_2(s)}$. Although this could affect the design considerably, the principles would remain the same. The poles of the error processing in a tracker are usually made as close to the origin as possible to minimize the bandwidth and, therefore, maximize the signal-to-noise ratio. But for the servo design, it would be convenient to place the poles out as far possible. It is likely that for a tracker the servo design would require compensation[99-103].

For a step function input $R_1(s) = \frac{10^{-4}}{s}$ and from Eq. 3-227

$$C(s)c(t) = \mathcal{L}[c(t)] = \frac{(10^{-4})7.8 \times 10^6}{s(s + 170)[(s + 130)^2 + (170)^2]} \tag{3-228}$$

$$c(t) = 10^{-4}\left|1 - 1.5e^{-170t} + 1.23e^{-130t}\sin(170t - 204°)\right| \tag{3-229}$$

Eq. 3-229 is plotted in Fig. 3-166.

The torque response for the step input is

$$\mathcal{L}[T(t)] = \mathcal{L}[J\ddot{c}(t)] = \frac{(10^{-4})7.8 \times 10^6 s}{(s + 170)[(s + 130)^2 + (170)^2]} \tag{3-230}$$

The inverse transform is

$$T(t) = -4.48e^{-170t} + 5.71e^{-130t}\sin(170t + 51°) \tag{3-231}$$

When plotted in Fig. 3-166, Eq. 3-231 illustrates that a step-function input of 0.1 mrad is as large as the system will tolerate without becoming saturated.

3-9.1.4.1.4.3 Response to Torque Disturbance

In any servo loop, the output response to an

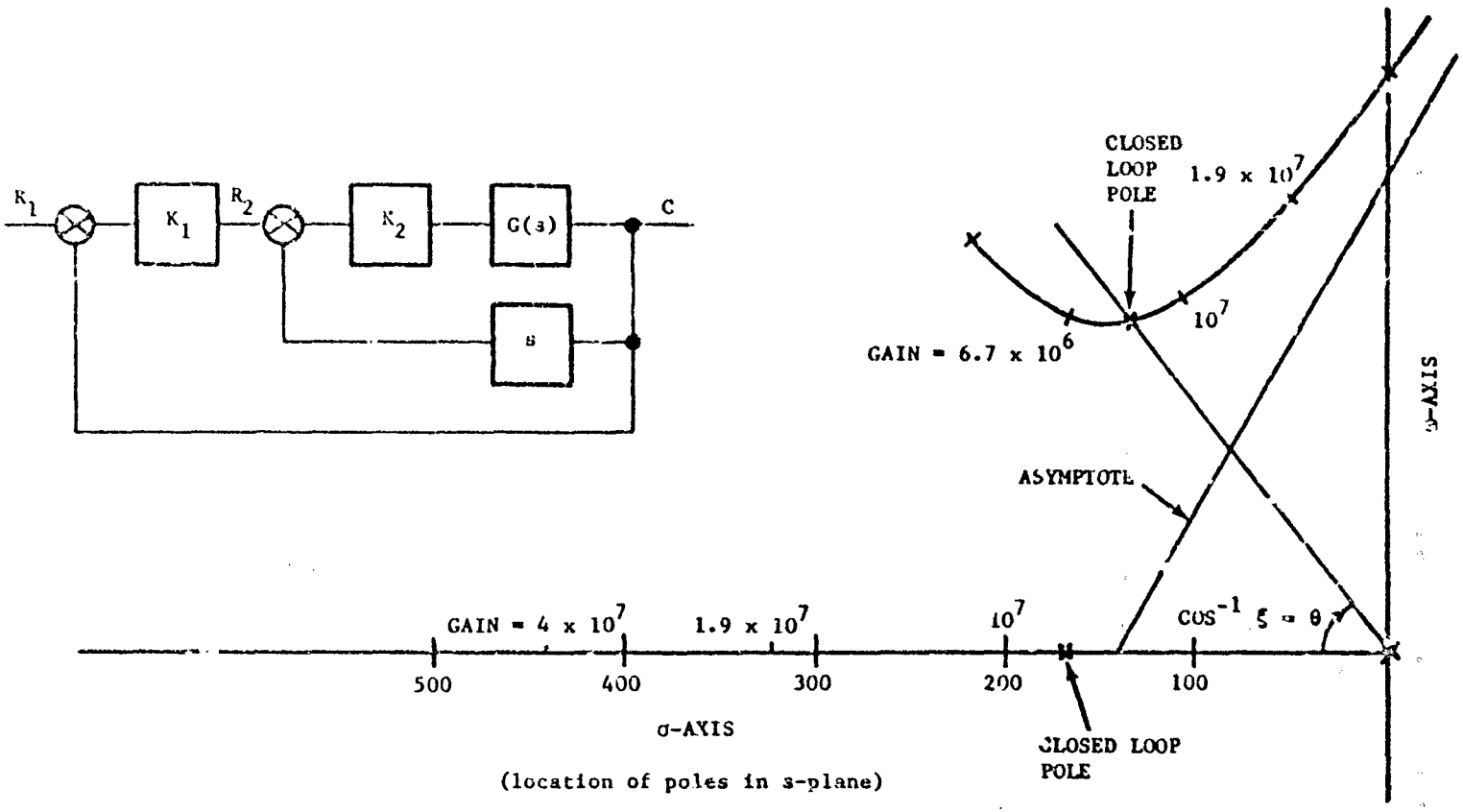


FIGURE 3-165. Root Locus for Outer Loop

input disturbance is given by the $\frac{C(s)}{R(s)}$ response divided by the forward transfer function from R to the disturbance U. From Figs. 3-159 and 3-162

$$\frac{C(s)}{U(s)} = \left[\frac{1}{\frac{(8.2)(405)(0.23)(100)}{s + 430}}\right] \frac{C(s)}{R_1(s)}$$

$$= \left(\frac{s + 430}{7.8 \times 10^3}\right) \frac{C(s)}{R_1(s)}$$

$$\frac{C(s)}{U(s)} = \frac{10(s + 430)}{(s + 170)[(s + 130)^2 + (170)^2]} \tag{3-232}$$

The output response to a torque step function of 1 ft-lb magnitude is

$$\mathcal{L}[c(t)] = \frac{10(s + 430)}{s(s + 170)[(s + 130)^2 + (170)^2]} \tag{3-233}$$

Taking the inverse transform

$$c(t) = 5.52 \times 10^{-4} - 5.1 \times 10^{-4} e^{-170t} + 5.41 \times 10^{-4} e^{-130t} \sin(170t - 174.6°) \tag{3-234}$$

According to this equation the steady-state response is defined by 5.52×10^{-4} rad, the first term in the right-hand side.

In most systems, mass unbalance can be made about as small as the friction torque. Referring to the motor and tachometer generator specifications, the sum of their friction torques is 0.04 ft-lb. Even if this were doubled to take into account electrical connections, adverse environmental conditions, and a g load of 10, the torque disturbance input would be 0.8 ft-lb. Substituting this value in Eq. 3-232 and using the final value theorem, the steady-state error in the output would be 0.44 mrad.

Several other considerations are appropriate to the complete analysis of the servo. These include an error analysis and a nonlinear analysis which are beyond the scope of this handbook but are mentioned here for the sake of completeness.

There are many components in the servo whose characteristics can vary with environmental conditions or aging, and many vary individually. An example of individual component variations is the inductance of the motor which is 0.01 H ±30%. This could cause a variation of

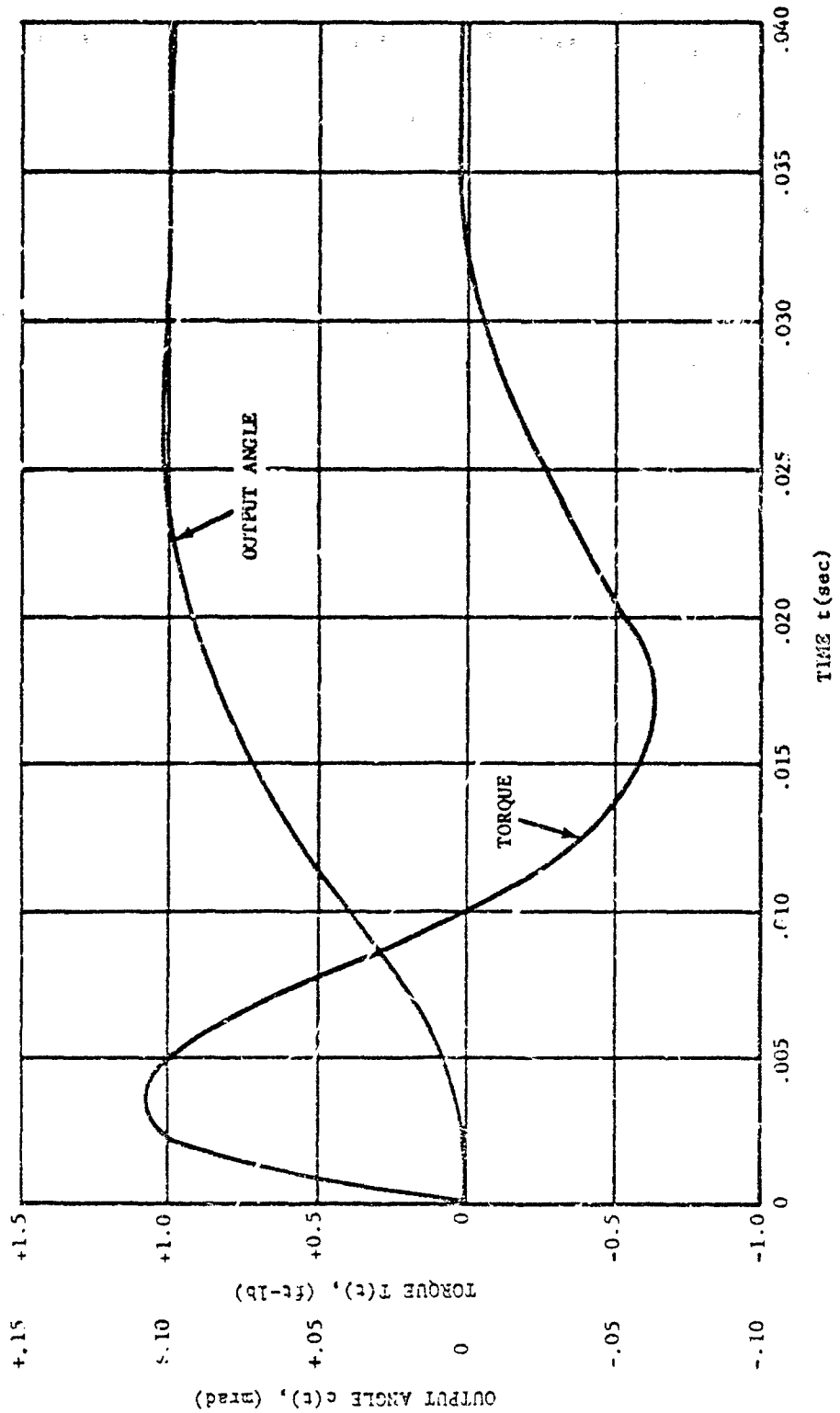


FIGURE 3-166. Output Angle and Torque Response for Step-function Input to Position Loop

approximately 30 percent in the electrical pole of the motor and in the forward gain of the $\frac{\frac{1}{L}}{s + \frac{R_T}{L}}$ block of Fig. 3-159. These variations must be taken into account and, to be safe, the system should be designed to meet the performance specifications even at the extreme limits of the component variations.

As illustrated in Fig. 3-166, a step-function input of 0.1 mrad is about as large as can be tolerated by the system without experiencing saturation. Larger inputs would cause the system to saturate. A common method of protecting all applicable components is to include a current-limiting function in the power amplifier. In the example design problem, the only significant nonlinearity is saturation. The describing function method[99] shows that the system will be stable under saturation. Also there is negative rate feedback both when the system is linear and when it is saturated. When linear, the rate feedback is provided primarily by the tachometer generator and, when saturated, by the back emf of the motor. However, saturation will vary the transient response considerably and the effect in this case is to slow the system down. A phase-plane analysis[99] or a direct-transient response analysis will provide complete information on the stability and transient response of the system.

One final servo analysis technique mentioned involves the emerging development and use of image-forming systems such as vidicons and large IR detector arrays for search and track. Sampled data techniques will probably be required[99].

3-9.1.5 Components

An important part of any servo design is the selection of suitable components. This generally starts with some knowledge of the required closed-loop response followed by a review of the components available. Some of the components commonly used in IR systems are discussed. Some of the advantages and disadvantages of various types are mentioned. Ahrendt and Savant have compiled an outstanding book on components for use in servo systems[104].

3-9.1.5.1 Bearings

Since IR systems usually are required to be highly accurate, precision ball- or roller-bearings are used. Only the pertinent facts relative to the use of bearings in servo designs are mentioned here since a vast store of information is available[105,106,107] on bearings in general.

The loads to be supported by the bearings should be analyzed to ensure reliability in the operational environment. Bearings are designed to support radial as well as axial loads, and some will support both in different relative amounts.

The bearing friction must be low. Bearing friction or viscous damping will appear in the servo loop as rate feedback. Rate feedback is often desirable, but usually friction is not desirable for two reasons. First, mechanical friction is difficult to control. Viscous components are usually sensitive to temperature, bearing load, state of lubrication, and other variables. Secondly, viscous damping requires the dissipation of energy which, in turn, imposes a load on the power supply, requires a higher capacity amplifier, and results in heated bearings.

For a complete discussion on the design, specifications, and use of bearings, refer to Refs. 104, 105, and 106.

3-9.1.5.2 Motors

The most common types of drives used with IR systems are the two-phase ac servo motor and the dc torquer.

The ac motor is usually rated for high speed and low torque. A gear train is used between the motor and load to match speed and torque characteristics. The motor has two windings, one connected to the ac line source and the other to an amplifier. About half the power required is generally supplied to each winding. Motors are commonly available for operating from 28 V rms to 130 V rms sources, and at frequencies from 60 to 400 Hz. These represent the most common voltage and frequency ranges, however, others are available. Ac servo motors are commonly available in sizes ranging from 1/2 to 100 w. Larger ac motors are usually less efficient and experience larger temperature increases.

Dc torque motors are available to satisfy most requirements imposed upon IR systems. Since dc motors apply torque directly to the load, many problems associated with gear trains (particularly backlash) are eliminated. The dc amplifier associated with this type motor must be carefully considered, as discussed in par. 3-9.1.5.4.

3-9.1.5.3 Gears

Gears are used for matching the motor speed and torque characteristics to those of the load to be driven. As with bearings, gears can cause a dead zone and viscous damping due to static and dynamic friction, respectively.

Probably the most difficult design problem involving gears is that associated with backlash. Anti-backlash gears are commercially available; however, these are generally used for instrument applications and are not suitable for transmission of high torque.

Many of the considerations associated with bearings are also applicable to gears. These include high-precision design and manufacture[108], lubrication, foreign material contamination, and reliability.

3-9.1.5.4 Amplifiers

The object of an amplifier in a servo system is to increase the power level of the error signal for application to the motor at the proper voltage. Some of the desirable characteristics of a good servo amplifier include high-gain, stability, suitable frequency response, low noise-level, and low output-impedance.

In achieving high gain, careful design is required to minimize noise. Noise appearing at the output will cause the boresight axis to "jitter" or move about in random fashion. White noise generated in the early stages of an amplifier can contribute to the noise. More commonly, several types of pickup and feedthrough can be troublesome. Spurious radiation can be coupled to low-level input wires routed close to high magnetic or electrostatic fields. Also, care must be taken to filter the ripple from signals which have been converted to dc and from power supplies. One technique, sometimes helpful here, is the use of an ac frequency which is outside the frequency response of the servo. However, even in this case, the amplifier could be saturated by the ripple, thereby causing heating of the motor and amplifier.

For maximum power transfer from the amplifier to the motor, the output impedance of the amplifier must match the input impedance of the motor. This is usually impractical because the back emf causes an apparent change in the input impedance of the motor with speed. In dc motors, the input impedance of the motor is inductive, thereby causing an electrical frequency response problem. Both effects are undesirable, adding to the complexity of the servo analysis. Thus, the output impedance of the amplifier is generally made as low as practicable.

3-9.1.6 Gimbal Associated Measurements

3-9.1.6.1 Angle Measurement

The two most common methods of obtaining angular measurements require the use of potentiometers and differential transformers. Both measure the relative angle between the platform on which the instrument is mounted and the gimbal. In certain applications, a gyro system is used on the gimbal to measure the spatial coordinates of the gimbal.

A wide variety of potentiometers is available. Some of the more important characteristics of potentiometers include friction, resolution, noise, and linearity.

In measuring very small angles, the resolution of a potentiometer can be the limiting factor. This is particularly true of wire-wound potentiometers where the resolution is generally limited by the number of turns of wire. However, wire-wound components offer advantages such as stability and accuracy of total resistance. Some manufacturers use a continuous resistance element in potentiometers for which they claim infinite resolution.

The other angular measurement component in common use is the differential transformer. Many variations of the technique (e.g., synchro, induction potentiometer, microsyn, etc.) operate on the general principle of generating an ac signal which corresponds to the relative position of one or more windings of a transformer with respect to the others. Another variation, the variable reluctance pickup, makes use of the motion in part of the magnetic circuit of a

transformer to produce an output which corresponds to the input angular position.

All of these magnetic devices require ac excitation and produce an ac output. They are usually relatively free of the friction problems associated with potentiometers; also, they offer infinite resolution and low noise, while some units can measure angles within a few seconds of arc.

3-9.1.5.2 Rate Measurement

The two most common components used for measuring angular rates are the tachometer generator and the rate gyro[109,110]. Sometimes a suitable position signal can be differentiated; however, since this method often results in excessive noise on the rate signal, the rate is usually measured directly.

Tachometer generators are commercially available in a great many varieties. They can provide either an ac or dc output, and a large range of input rates, sizes, and accuracy ratings.

The rate gyro provides the most accurate rate measurement. It measures the spatial rate of the gimbal and is capable of measuring extremely low rates. The rate gyro consists of a rotor which is capable of very high speeds up to 20,000 rpm. With an angular input rate in an axis perpendicular to the spin axis of the rotor, the rotor axis tries to precess in an axis normal to both the rotor axis and the input axis. It deflects the retaining spring an amount proportional to the input rate. This deflection is measured with an accurate position pick off. This signal becomes the output and is proportional to the input rate. The gyro is usually mounted in a liquid to make it insensitive to accelerations; the movable part is balanced to an extreme accuracy, and the temperature is often controlled[109,110].

3-9.2 IR SYSTEM MONITORS

3-9.2.1 Requirements for Monitor Devices

Various methods of monitoring IR systems are used to provide periodic indications of their operational behavior in a system. This becomes necessary in order to account for operational changes in a system and, thereby, correct the IR system's output data accordingly. Such devices are especially necessary for monitoring IR sensors used in spacecraft where components cannot be replaced readily and where degradation of component performance can occur due to rigorous environmental conditions. Undetected changes in system performance due to degradation or variation in component performance are highly undesirable. Only critical components used in IR systems installed in aircraft, helicopters, or on ships are monitored since periodic system checkout can be performed using ground equipment at scheduled maintenance intervals.

Monitoring devices include temperature sensors on critical components, and voltage and current monitors on the electrical system. Temporary devices can be used during development tests and operational checkout; conversely, monitors can be permanent components in a system required for use in space.

3-9.2.2 Thermocouples

Thermocouples are used for measuring the operating temperature of system components. The major advantage in using thermocouples is that they can be readily installed on the surface to be measured by soldering, spot welding, or peening. The thermocouple generates an emf when a standard cold-junction reference is used, thereby, providing a direct indication of the surface temperature. Direct readout temperature indicators are used for visual monitoring where the number of readings is small. An example of this application is the monitoring of component temperature levels on an aircraft-mounted IR system during operational checkout. Fig. 3-167 is a schematic of a thermocouple circuit.

The accuracy of thermocouples falls off at lower temperatures due to the weaker emf signal generated and the nonlinearity of the signal. Thermocouples should be calibrated in place if they are soldered or spot welded to account for the effect of the installation method. In addition, an accurate reference junction must be provided such as an ice bath or an electronic reference junction. The thermocouple itself requires no external power, although the weak signals are often amplified, especially when detecting low temperatures.

Thermopiles, consisting of a group of thermocouples placed in series, provide higher output signals; however, thermistors are now being used in most cases where greater signal power is required for accurate measurement.

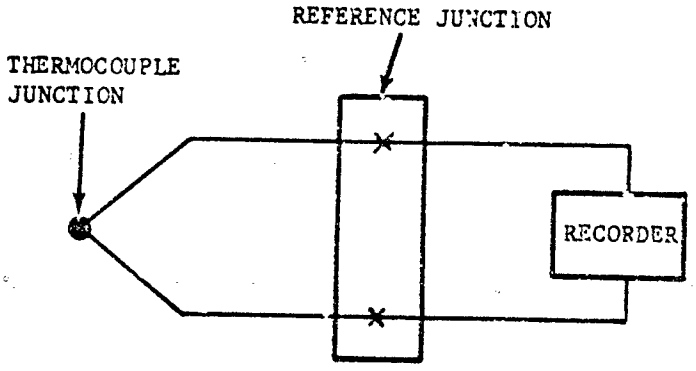


FIGURE 3-167. Thermocouple Circuit Schematic

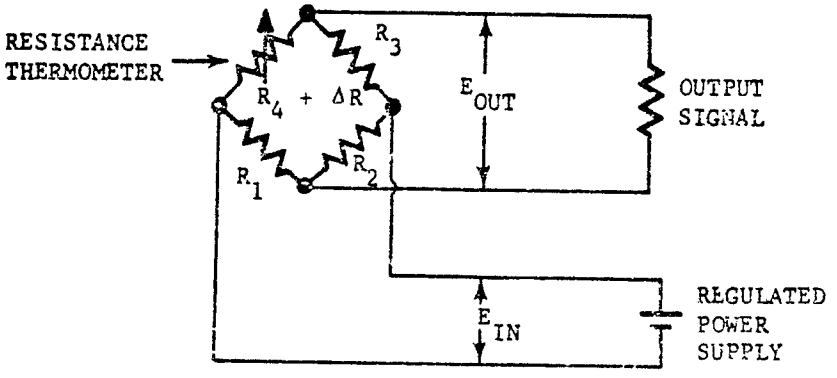


FIGURE 3-168. Resistance Thermometer Circuit

3-8.2.3 Thermistors and Resistance Thermometers

Resistance thermometers and thermistors are passive sensors which require a bridge circuit and a controlled input voltage for recording the resistance of the element. A schematic of a resistance thermometer circuit is shown in Fig. 3-168. Thermistors contain semiconductor or carbon elements while resistance thermometers have platinum, tungsten, or nickel wire elements. Resistance thermometers and thermistors are available both in cartridge and sheet form. They are secured on a surface by cementing, strapping, or they may be held by screws. Good thermal contact between the temperature-measuring sensor and the surface to be measured is required especially when transient conditions are involved. These elements have essentially a linear resistance variation with temperature. When the sensor element is uniformly heated or cooled, the resistance changes. This, in turn, causes the bridge to change the output-signal voltage. The basic equation for the platinum resistance thermometer is

$$t = 100\left(\frac{R_t - R_o}{R_x - R_o}\right) + \delta\left(\frac{t}{100} - 1\right)\frac{t}{100} \qquad (3\text{-}235)$$

where

t = temperature

R_t = resistance of sensor at temperature t

R_o = resistance of sensor at reference temperature t_o

R_x = resistance of sensor at reference temperature t_x

δ = constant of the sensor

To process the readings from a group of resistance thermometers which are monitoring an IR system, a specific voltage range is selected and the signals from the thermistors are adjusted by amplifiers or voltage dividers to fall within the selected range. The signals from all the thermistors can then be processed by multiplexing through the same system used for recording or telemetering data to earth from a satellite.

Resistance thermometers must also be calibrated periodically to ensure that accurate data are being reported. Calibration consists essentially of applying a known signal voltage to the device and comparing the expected and actual voltages.

3-9.2.4 Voltage and Current Monitors

Voltage and current are conventionally measured with somewhat identical methods. Direct-reading voltmeters and ammeters operate on the principle that a current flow will produce a voltage indication in one scheme and current in another. Conventional meters can be used for direct monitoring of ground or airborne instruments. However, for spacecraft systems, the telemetering requirement dictates that voltage and current signals be converted generally to frequency or pulse signals for transmission. This is accomplished by means of an oscillator used in the setup illustrated in Fig. 3-169. The conversion of data to this form permits multiplexing, whereby a number of monitor signals can be processed in sequence through the same transmitting channel.

3-9.2.5 Calibration of IR Sensors

A reference or calibration source can usually be designed into an IR system from which highly reliable data are required. The calibration source must provide a nearly constant level of radiation to which the detector can be exposed. A chopper or rotating mirror may be used periodically as a means of exposing the detector to the calibration source. A simple tungsten-filament lamp, provided with a controlled power input, can serve as a reference of known constant temperature. The actual temperature level in the calibration source lamp usually ranges from 1500° to 2000°K. It is extremely important to keep the temperature variation along the active filament length at a minimum.

Two typical IR sensor calibration arrangements are illustrated in Fig. 3-170. In one setup, a calibration source is exposed to a multi-element array through a single mirror which is rotated at required intervals for periodic calibration. In the other setup, a detector is exposed to selected wavelengths by use of filters located on a chopper wheel. A mirror located on the wheel reflects the beam from the calibration lamp to the detector once each revolution.

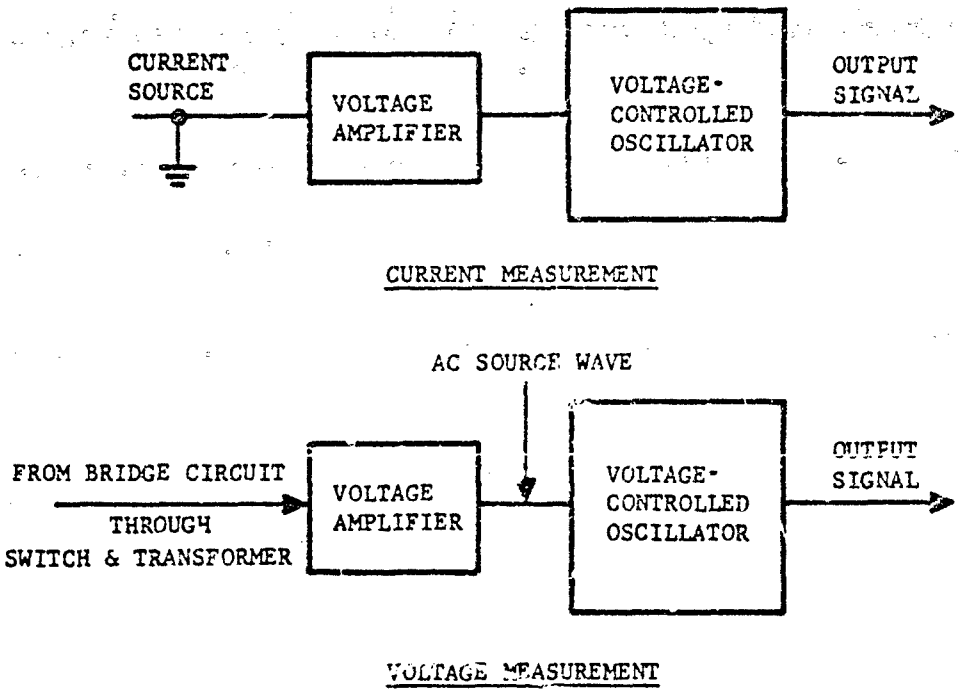
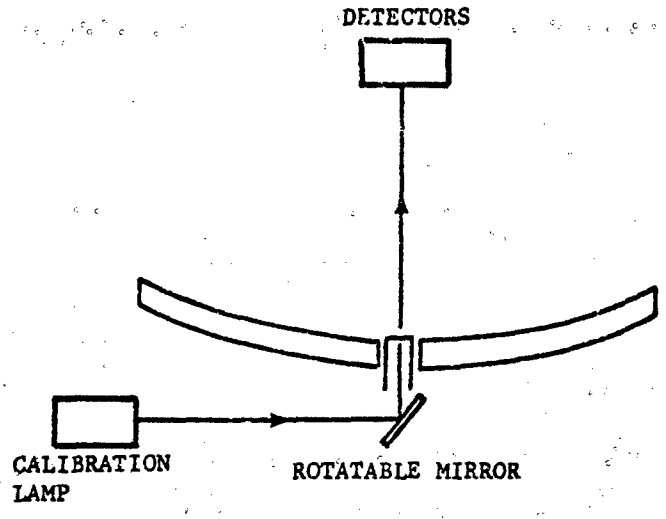


FIGURE 3-169. Block Diagram of Voltage and Current Monitors

3-9.3 SUN SHUTTERS

IR sensors, particularly those in spacecraft applications, must be protected from exposure to direct illumination from the sun. Solar energy incident on a surface one astronomical unit ($92{,}897 \times 10^3$ mi) from the sun is high, 127 w/ft^2. This intensity increases proportionately with optical magnification; consequently, some means of obscuring the sun must be provided for the protection of mirrors, lenses, and detector arrays since these are easily distorted and damaged by heat. Various means of preventing (by obscuring or interrupting) solar energy from impinging upon the detector arrays are outlined in Table 3-20 along with the estimated complexity of the mechanisms, the design problems, and actuating mechanisms. Most devices are electrically actuated upon command from a sun sensor. The listing in Table 3-20, although by no means all-inclusive, may encourage the designer to devise other improved methods of obscuring the path of the solar energy. A method that may be applicable for one design may not be suitable for another.

One of the various techniques makes use of an iris type solar shutter at the entrance to the optical system. This type, normally used with cameras, may be mechanically or electrically actuated. A mechanically actuated unit may be as shown in Fig. 3-171; its application here is for a Cassegrainian optical system. This consists of a bi-metallic actuator heated by the sun. As the sun begins to illuminate the aperture, energy is concentrated on a bi-metallic shutter, thereby closing the aperture. Extreme care must be taken to maximize the actuation speed to prevent damage to the detectors. The electrical counterpart of this method, shown in Fig. 3-172, has the advantage that the sun sensor can be placed almost anywhere and, most important, the actuation mechanism is rapid.



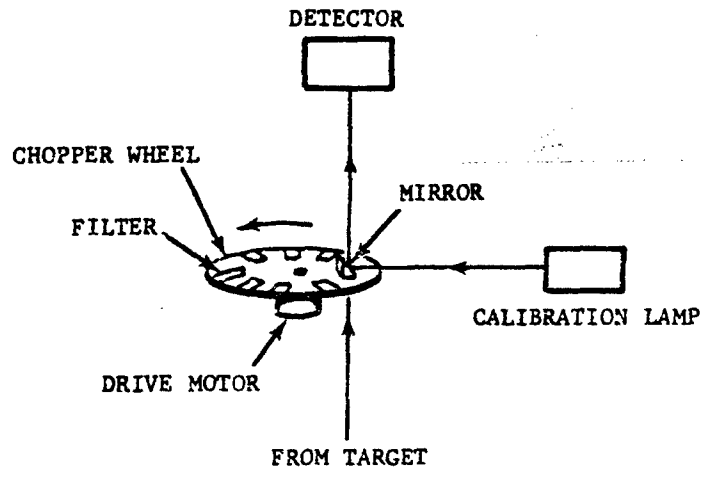


FIGURE 3-170. Diagrams of Typical IR Calibration Systems

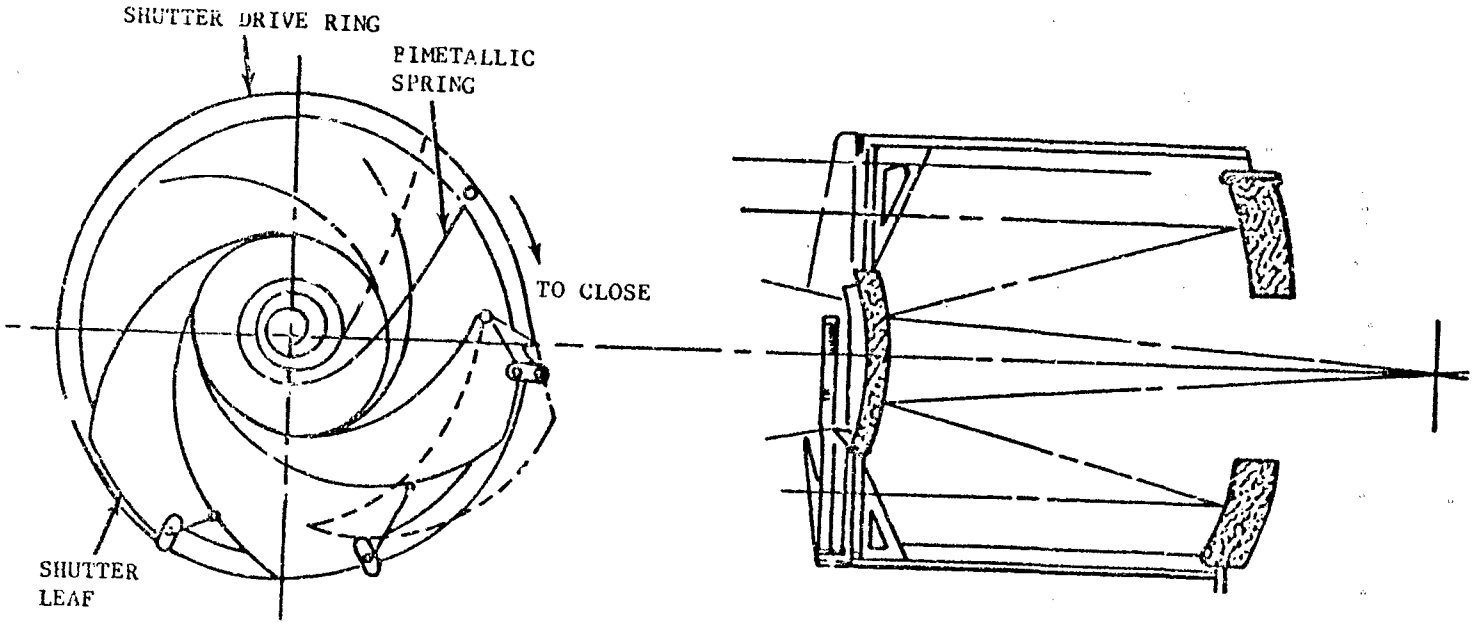


FIGURE 3-171. Mechanical Sun Shutter

TABLE 3-20. METHODS OF PROTECTING IR DETECTORS, MIRRORS, AND LENSES FROM DIRECT SOLAR ILLUMINATION

	DESCRIPTION	REMARKS
1.	Iris type shutter at entry	May be all mechanical or electrically actuated (Figs. 3-171 and 3-172)
2.	Window shade at entry	Fig. 3-173
3.	Actuated hood at entry	Fig. 3-174
4.	Leap/trap closure	Fig. 3-175
5.	Swivel closure inside sun shade	Fig. 3-176
6.	"Venetian blinds" directly above mirror	*
7.	Movable visor for detector array	Fig. 3-177
8.	Closure utilizing focal plane support structure	Fig. 3-178
9.	Light sensitive glass window in sun shade	Required energy may not pass through glass*
10.	Electro-optic light modulator in optical system	Required energy may not pass through*
11.	Tilt entrance lens to avoid focusing	Possible realignment problems*
12.	Tilt mirror to avoid focusing on detector array	Possible realignment problems*
13.	Tilt detector array to avoid focusing	Possible realignment problems*

*Sun sensor electrically actuated

Another method similar to the first, except for the shutter, is the type normally used in a focal plane shutter camera, i.e., a window-shade. This method, which is electrically actuated, is illustrated in Fig. 3-173. Another type of mechanism is illustrated in Fig. 3-174. Here the hood is normally open; however, when the sun sensor actuates the drive, the latches and associated hood close, preventing energy from entering the optical system.

A slightly more complex method is illustrated in Fig. 3-175. The two or more leaf closures, normally open, close the entrance aperture of the optical system on command.

Fig. 3-176 illustrates a rotary damper-type sun shutter. During normal operation, the damper is oriented along the optical axis and when solar protection is required, the drive mechanism rotates the damper 90 deg thereby sealing the aperture.

A Venetian-blind arrangement can also be used to obscure the solar energy as illustrated in Figs. 3-177 and 3-178. This mechanism is located at the detector array, at which point in the optical system it interrupts the solar energy.

There are also a number of nonmovable systems available for use. One consists of a light-sensitive glass placed ahead of the sensors. This glass will darken upon exposure to intense light, thereby obscuring the impinging solar

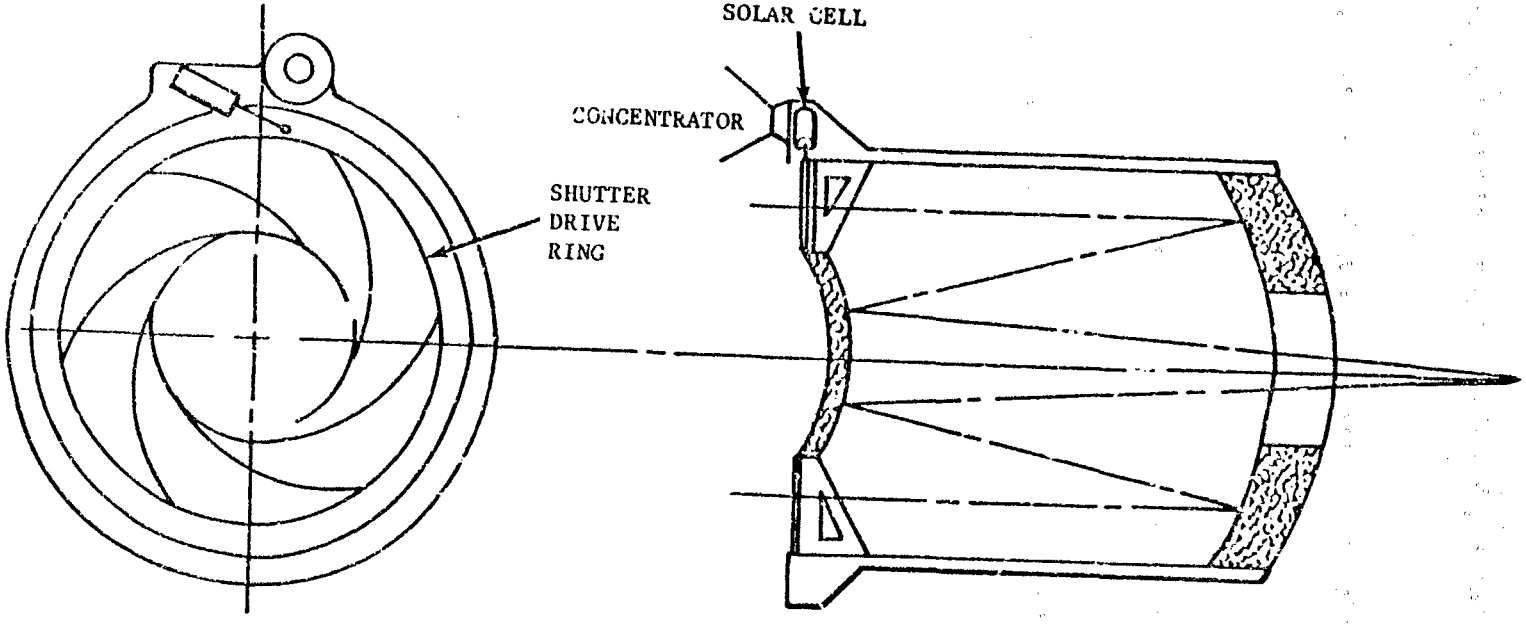


FIGURE 3-172. Electrical Sun Shutter

energy. When the solar energy is removed, the glass becomes transparent again. Another device consists of electro-optical light modulators. Both methods present potential problems in not permitting all the energy to pass; however, in some systems they may be adequate.

There are other methods of preventing solar energy from impinging upon the detector arrays. In brief, these entail moving one of the optical elements in the system to prevent the energy from focusing upon the detector arrays. It may even help to fold the energy, in a Cassegrainian, Gregorian, Newtonian, or Maksutov system at the detector array. This folding optic could be electrically rotated to reflect the solar energy back out of the optical system, thus reducing the heating effects.

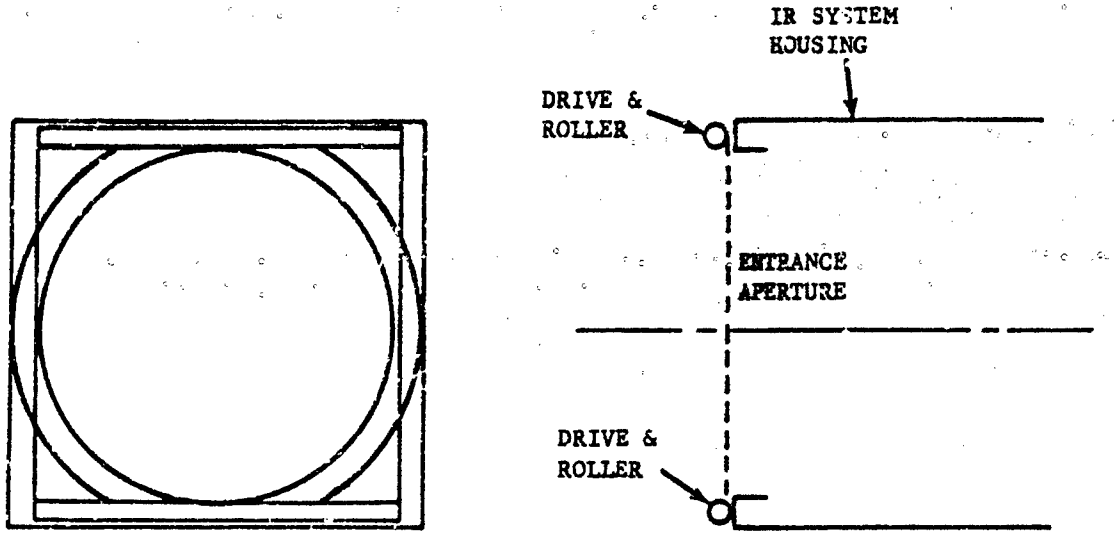


FIGURE 3-173. Window-shade Type Protection for Focal Plane Shutter Camera

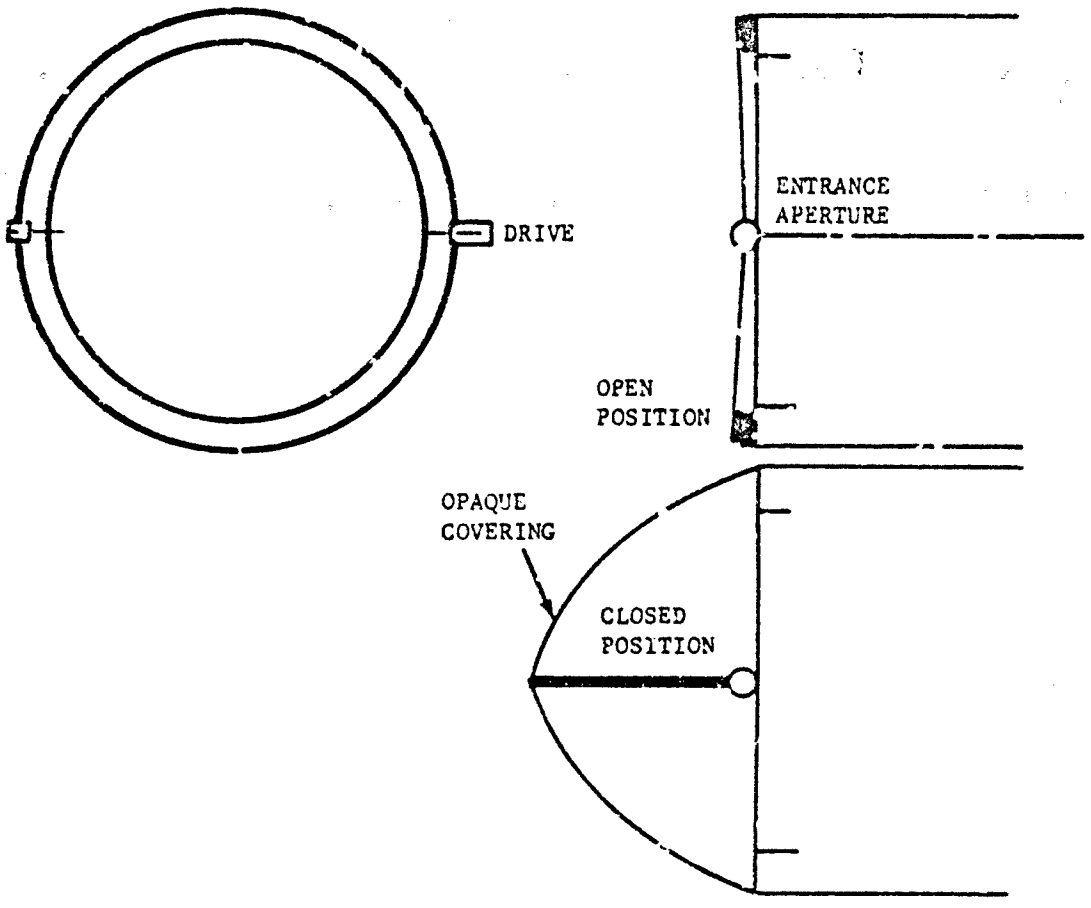


FIGURE 3-174. Hood type Sun Shutter

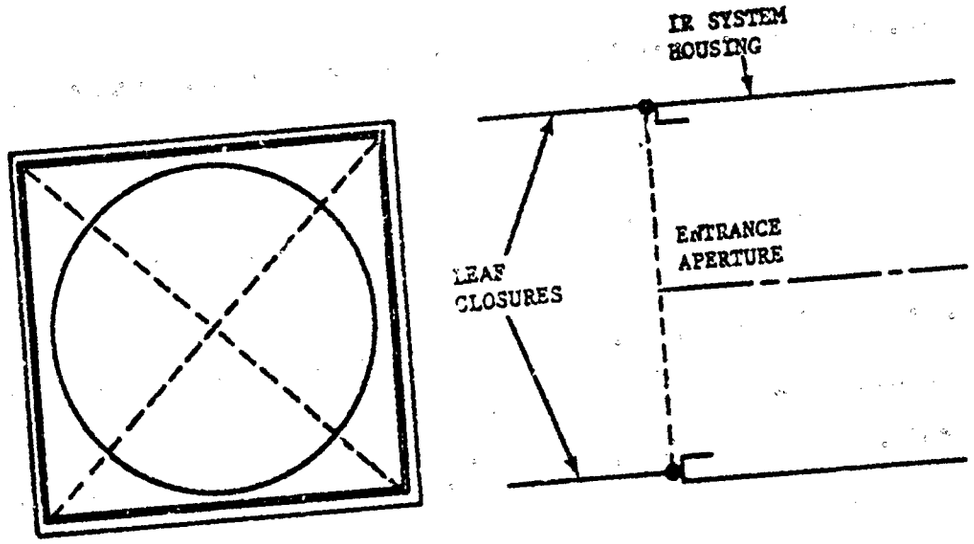


FIGURE 3-175. Four-leaf Clover Sun Shutter

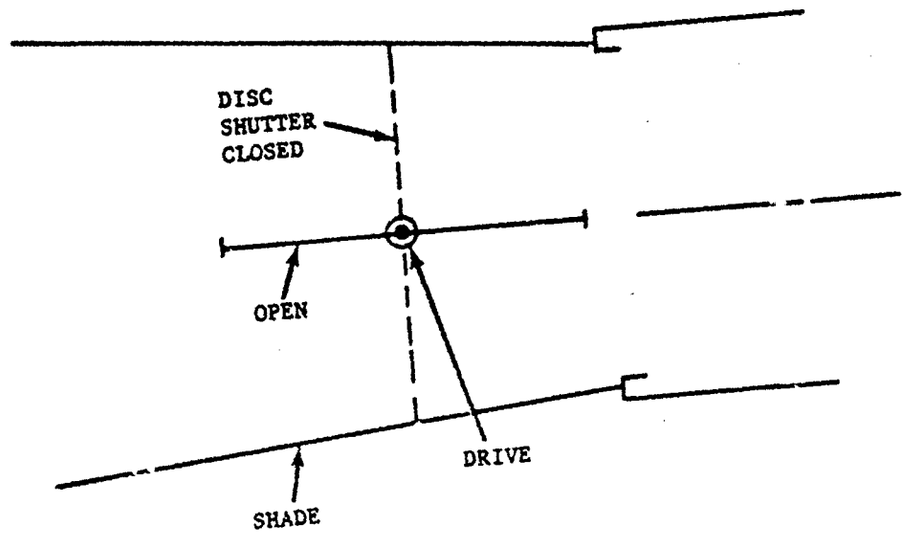


FIGURE 3-176. Rotary Damper Sun-shutter Mechanism

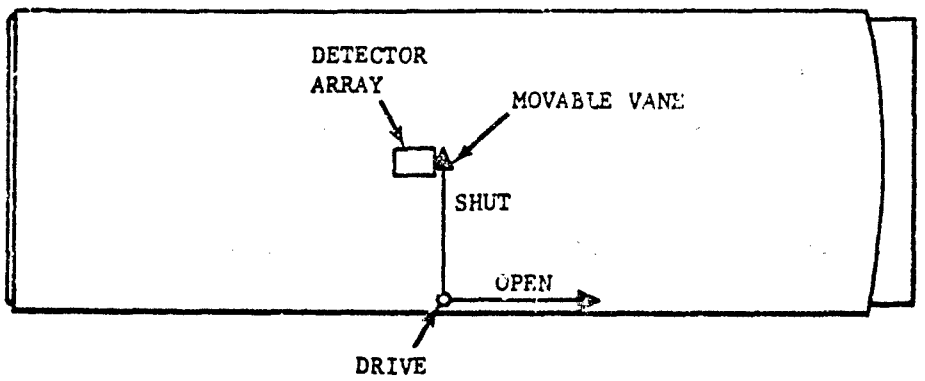


FIGURE 3-177. Classic Schmidt Sun-shutter System

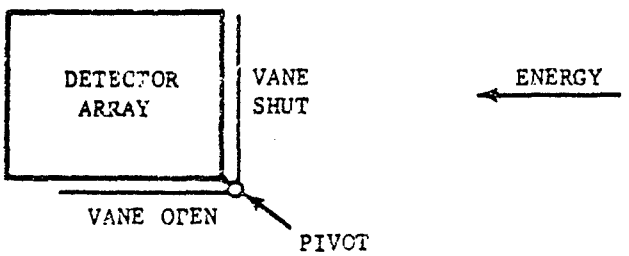


FIGURE 3-178. Blocking Mechanism in Schmidt System

3-9.4 ELECTROMAGNETIC INTERFERENCE

Electromagnetic energy refers to the total electrostatic and magnetic fields set up by the movement of electrons on a conductor. Interference as defined by MIL-I-6181D, *Interference Control Requirements, Aircraft Equipment*, is "any electrical or electromagnetic disturbance phenomenon, signal or emission, man-made or natural, which causes or can cause undesired response, malfunctioning or degradation of performance of electrical and electronic equipment . . .". The use of dense packaging techniques, large voltage and current devices, and high-frequency and fast-switching circuits increase the probability of interference. Some devices such as semiconductors may be damaged by high-level transient interference.

The approximate safe level of radiation for humans is below 10 mw cm^{-2} at any frequency. Sensitive areas such as the eyes can tolerate less radiation than other parts of the body.

3-9.4.1 Sources of Electromagnetic Radiation

Some of the common sources of interference are current-carrying conductors such as wires, plates, chassis, and other hardware; motors, both ac and dc; relays, inductors, transformers, switches, semiconductors such as silicon controlled rectifiers; and transmitters. The extent of the interference depends on the voltage, current, and frequency of the source; line impedance and length and proximity to other lines; and the ground plane of the receiving equipment.

3-9.4.2 Elimination and Rejection

Interference can be in some cases reduced or eliminated at the source by using shields on wires, arc suppressors on relays, and electrostatic shields between the primary and secondary of transformers. Other interference sources, such as transmitters, cannot be shielded at the source and perforce all other equipment must be shielded from them.

The military have established design requirements and interference test procedures for their equipment. These are specified primarily in MIL-I-6181D and MIL-E-6051D, *Electrical-Electronic System Compatibility and Interference Control Requirements for Aeronautical Weapon Systems, Associated Subsystems and Aircraft.*

Wires and cables are common sources of interference and are susceptible to interference from other sources. Electrostatic coupling is a function of distributed capacitance while electromagnetic coupling is a function of the mutual inductance between the wires. Wire lengths should be made as short as possible. They should be shielded if carrying high voltage, current, and frequency signals; and separated as much as possible since induced voltage drops exponentially with distance. Lines should cross other lines at right angles if possible to reduce induced current. Lines carrying alternating current should consist of twisted pairs. For a discussion of EMR interference and protective methods, the reader is referred to AMCP 706-235, *Hardening Weapon Systems Against RF Energy*.

The following equation is used to determine the approximate value of electromagnetic coupling.

$$E_r = 3.19 \times 10^{-8}\ fLI_s\ \ell n \left(\frac{D_2}{D_1}\right), \text{ V} \qquad (3\text{-}236)$$

where

E_r = induced electromagnetic voltage, V

f = frequency, Hz

L = the length of the receiving wire, in.

I_s = current of the source, A

D_1 and D_2 = distance of the receiving wire from the source according to Fig. 3-179

The induced voltage E_r can also be determined by the equation

$$E_r = \frac{E_s \times R_3}{X_{cd} + R_3}, \text{ V} \qquad (3\text{-}237)$$

where

$$R_3 = \frac{R_1 \times R_2}{R_1 + R_2}$$

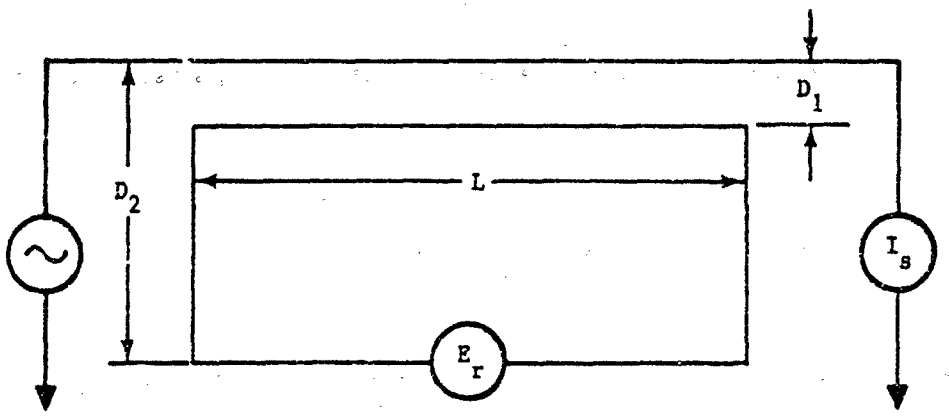


FIGURE 3-179. Schematic Illustrating Electromagnetic Coupling Parameters

E_s = source voltage, V

R_1 and R_2 = the generator and load impedances, respectively of the receiving wire, ohm

X_{cd} = the distributed capacitive reactance between the source and receiver, ohm

3-9.4.2.1 Methods

3-9.4.2.1.1 Bonding

Chassis, racks, and other enclosures should be bonded to the ground plane, or airframe in the case of an aircraft. The low-impedance path to the ground plane would thereby prevent the formation of static charges and stray RF potential. This also provides protection to personnel and equipment from high-potential lines faulted to the enclosure. Permanent type bonds may be achieved by welding, brazing, sweating, swaging, and other metal-flow techniques. The semi-permanent bonds include clamps, rivets, jumpers, and straps. Jumpers and straps are commonly used; however, their length-to-width ratio must not exceed 5:1. Use of dissimilar metals should be avoided to minimize corrosion. If their use is inevitable, means must be taken to exclude moisture from the bonded areas.

For protective treating of metal surfaces, iridite should be used instead of anodize. This will also permit the required RF bonding. Alodine may also be used although not as satisfactorily as iridite. For information on the treatment of metal surfaces and bonding, the reader is referred to AMCP 706-235, *Hardening Weapon Systems Against RF Energy* and AMCP 706-100, *Design Guidance for Producibility*.

Bonding requirements on military equipment are specified in MIL-B-5087B, *Bonding, Electrical, and Lightning Protection, for Aerospace Systems.*

3-9.4.2.1.2 Grounding

A uniform grounding, shielding, and cabling arrangement is necessary within a system and between all interconnecting systems, in order to prevent redundancy and provide proper interfacing between assemblies of an equipment.

In a properly grounded system, the interconnection lines between the transmitting and receiving devices at the low-frequency circuit are kept near the same potential.

Electromagnetic fields can be reduced by grounding the electrical system at a single point usually at the power supply. The possibility of a ground loop causing noise pickup on the lines is thereby decreased. At higher frequencies, a common-point ground becomes impractical

because long lines become interference sources. In this case, the secondary power should be generated at the circuit and grounded there. Grounding isolation should be provided, between this circuit and the circuit receiving the signal, through common mode rejection circuits or transformers with electrostatic shields between the primary and secondary. All lines should be kept as short as possible.

3-2.4.2.1.3 Shielding

Proper shielding prevents noise energy from radiating to other equipment and keeps radiated energy from entering low signal and noise equipment.

Chassis, racks, and other enclosures containing interference sources or interference-susceptible circuits should be made of good shielding material. Materials having low electrical resistivity, such as copper and aluminum, should be used for high-frequency applications; materials of high permeability should be used for low-frequency applications. These enclosures should be bonded to the ground plane or airframe if mounted in an aircraft. If the enclosure is on vibration isolators, a bonding strap should be used around the isolators, making sure that the strap does not disturb the isolators.

Shields used on wires can be thought of as extensions of the enclosures from which the wires are going or coming. Shields on wires carrying low-frequency signals should be grounded at one end only to avoid ground loops. If high-level signals are involved, the shield ground should be at the sending end to avoid high-level signals on the shield. In the case of low-level signals, the shield ground should be at the receiving end to protect the receiving-end circuits. When a signal shield is passed through a connector, separate pins should be used for each shield. All shields should have an insulating cover to prevent accidental grounding. A shield on a low-frequency signal pair or three lines should be grounded to the chassis or airframe at both ends and at all breaks in the lines.

Shields on lines carrying high-frequency signals should be grounded at both ends to prevent the shield from becoming a radiator at these frequencies. If a coaxial cable cannot be chassis-grounded at both ends, a tri-axial cable should be used, grounding the outer shield to the chassis or airframe at both ends and at all breaks in the line.

For information on cable construction, the reader is referred to AMCP 706-125, *Electrical Wire and Cable.*

REFERENCES

1. A. R. Hilton, "Nonoxide Chalcogenide Glasses as Infrared Optical Materials", Appl. Opt. 5,1877-82 (1966).
2. Ballard, McCarthy, and Wolfe, *Optical Materials for Infared Instrumentation*, IRIA Report No. 2389-11-S, 1959, Supplement 2389-11-S, 1961, University of Michigan Press, Ann Arbor, Michigan.
3. *Infrared Glass*, Barr and Stroud Ltd., Glasgow, Scotland, 1962.
4. *Progress Report: Infrared*, Bausch and Lomb, Inc., Rochester, N. Y.
5. *Corning Fused Silica Code 7940*, Corning Glass Works, Corning, N. Y., 1960.
6. *Infrared Transmitting Glasses*, Corning Glass Works, Corning, N. Y., 1960.
7. *Pyrex Brand Telescope Mirror Blanks*, Corning Glass Works, Corning, N. Y., 1964.
8. *Pyroceram, Codes 9606-9608, PY-3*, Corning Glass Works, Corning, N. Y.
9. *Vycor*, Corning Glass Works, Corning, N. Y., 1964.
10. *Optical Silicon*, Dow Corning Corp., Midland, Michigan, 1959.
11. *Kodak IRTRAN Data Sheets*, Eastman Kodak Company, Rochester, N. Y., 1967.
12. *Infrared Materials*, Exotic Materials, Inc., Costa Mesa, Calif., 1961.
13. *Fused Quartz Catalog*, General Electric, Willoughby, Ohio, 1964.
14. *Harshaw Optical Crystals*, The Harshaw Chemical Co., Cleveland, Ohio, 1967.
15. *Optical Properties of Synthetic Sapphire*, The Linde Co., N. Y.
16. *Optical Crystals*, Optovac, Inc., North Brookfield, Mass., 1964.
17. *Low-Expansion CER-VIT Materials*, Owens-Illinois, Toledo, Ohio, 1967.
18. Donald Jacobs, *Fundamentals of Optical Engineering*, McGraw-Hill, Inc., N. Y., Chapter VIII.
19. Rudolph Kingslake, *Applied Optics and Optical Engineering*, Vol. I, Academic Press, N. Y., 1965, Ch. 8.
20. P. Baumeister, "Design of Multilayer Filters by Successive Approximations", J. Opt. Soc. Am. 48(12), 955 (1958).
21. P. H. Lissberger, "Properties of All Dielectric Interference Filters, I, A New Method of Calculation", J. Opt. Soc. Am. 49:2 (1959).
22. P. H. Lissberger, "Properties of All Dielectric Interference Filters, II, Filters in Parallel Beams of Light Incident Obliquely and in Convergent Beams", J. Opt. Soc. Am. 49:2 (1959).
23. S. D. Smith, "Design of Multilayer Filters by Considering Two Effective Interfaces", J. Opt. Soc. Am. 48:1 (1958).
24. F.A. Jenkins and H.E. White, *Fundamentals of Optics*, McGraw-Hill, Inc., N.Y., 1957.
25. A. E. Conrady, *Applied Optics and Optical Design*, Ch. 7.
26. Hertzberger, *Modern Geometrical Optics*, Interscience Publishers, Inc., N. Y., 1958.
27. Warren J. Smith, *Modern Optical Engineering*, McGraw-Hill, Inc., N. Y., 1966.
28. F. U. Rosberny, "Measurement of Homogeneity of Optical Material in the Visible and Near Infrared", Appl. Opt. 5, No. 6 (1966).
29. T. Twyman, *Prism and Lens Making*, Hilger Watts Ltd., 504-15.
30. W. L. Wolfe, Ed., *Handbook of Military Infrared Technology*, Office of Naval Research, Washington, D. C., 1965.

31. R. M. Griffin, R. E. Levin, and W. G. Matheson, *Fort Belvoir, Incandescent Lamp Development*, Final Report AD-480 4826, Sylvania Lighting Products, Inc., Danvers, Mass., January 1966.

32. J. W. van Tigen, "Iodine Incandescent Lamps", Philips Tech. Rev. 23, No. 819, 237-42 (1961/1962).

33. R. E. Levin and A. E. Westlund, "Design Parameters for the Use of Quartz-Iodine Lamps", J. Soc. Motion Picture Television Engrs. 75, No. 6, 589-93, (1966).

34. M. R. Null and W. W. Lazier, "Carbon Arc as a Radiation Standard", J. Opt. Soc. Am. 52, 1156-62 (1962).

35. Correspondence from S. B. Gibson, Night Vision Laboratory, Fort Belvoir, Va., May 1968.

36. D. Fromm, *Absolute Spectral Distribution Measurements of Xenon High-Pressure Discharges*, USAERDL Report No. 1837, USAERDL (now U. S. Army Mobility Equipment R & D Center) Fort Belvoir, Va., November 1965.

37. S. B. Gibson, et al., *Characteristics of a 20-KW Liquid Cooled Xenon Arc Searchlight*, Report No. 2, AD-650 881, JSAEC Night Vision Laboratory, Fort Belvoir, Va., March 1967.

38. H. S. Strauss, et al., *Compact Arc Near Infrared Radiation Sources*, Final Technical Report, Duro Test Corp., North Bergen, N. J., July 1967.

39. O. E. Lienhard, *Improvement of Communication Arc Lamps for Visible and Infrared Signaling Equipment*, Final Report AD-811 767, Engelhard Hanovia, Inc., Newark, N. J., February 1967.

40. M. Born and E. Way, *Principles of Optics*, Pergamon Press, N. Y., 1959.

41. L. Allen and D. G. C. Jones. *Principles of Gas Lasers*, Plenum Press, N. Y., 1967.

42. Bela A. Lengyel, *Introduction to Laser Physics*, John Wiley and Sons, N. Y., 1966.

43. W. Smith and P. Sorokin, *The Laser*, McGraw-Hill, Inc., N. Y., 1966.

44. A. K. Levine, Ed., *Lasers; A Series of Advances*, Marcel Dekker, Inc., N. Y., 1966.

45. G. Birnbaum, *Optical Masers—Supplement 2*, Advances in Electronics and Electron Physics, New York, Academic Press, 1964.

46. Z. J. Kiss and R. J. Pressley, "Crystalline Solid Lasers", Appl. Opt. 5, 1474-86 (1966).

47. E. Snitzer, "Glass Lasers", Appl. Opt. 5, 1486-99 (1966).

48. H. A. Elion, *Laser Systems and Applications*, Pergamon Press, N. Y., 1967.

49. R. A. Smith, F. E. Jones, and R. P. Chasmar, *The Detection and Measurement of Infrared Radiation*, Clarendon Press, Oxford, 1957.

50. C. Kittel, *Introduction to Solid State Physics*, 3rd edition, John Wiley and Sons, N. Y., 1966.

51. P. Kruse, "InSb PEM Infrared Detector", J. Appl. Phys. 30, 770 (1959).

52. M. Golay, "*A Pneumatic IR Detector*", Rev. Sci. Instr. 18, 357 (1947).

53. Markov, "An Extremely Fast Response Bolometer," Opt. and Spectrosc. 17, 507 (1964).

54. Clement, et al., "Carbon Composition Thermometers at Very Low Temperature", Rev. Sci. Instr. 24,545 (1953).

55. Futaki, "A New Type Semiconductor, Critical Temperature Resistor", Japan. J. Phys. 4, 28 (1965).

56. Martin and Bloom, "Application of Superconductivity to the Detection of Radiant Energy", Cryogenics 1, 1959 (1961).

57. R. Clark Jones, "Phenomonological Description of the Response and Detecting Ability of Radiation Detectors", Proc IRE 47, 1495 (1959).

58. E. Burstein and G. S. Picus, "Background Limited Infrared Detection", IRIS, February 1958.

59. H. Levinstein, "Extrinsic Detectors", Appl. Opt. 4, No. 6,642 (1965).

60. R. L. Petritz, "Fundamentals of Infrared Detectors", Proc. IRE 47 ,1458 (1959).

61. R. C. Jones, "Proposal of the Detectivity D** for Detectors Limited by Radiation Noise", J. Opt. Soc. Am. 50, 1058 (1960).

62. R. C. Jones, "A Method of Describing the Detectivity of Photoconductive Cells", Rev. Sci. Instr. 24, 1035 (1953).

63. J. A. Jamieson, McFee, Plass, Grube, Richards, *Infrared Physics and Engineering*, McGraw-Hill, Inc., N. Y., 1963.

64. W. L. Wolfe, Ed., *Handbook of Military Infrared Technology*, Office of Naval Research, 1965, p. 469.

65. Kruse, et al., *Elements of Infrared Technology* John Wiley and Sons, N. Y., 1962, p. 242.

66. G. J. Hoover, "Fabrication and Statistical Evaluation of High Density Ge:Hg Linear Arrays", (16th National Meeting) IRIS (1968).

67. William L. Wolfe, Ed., *Handbook of Military Infrared Technology*, Chapter 16.

68. James L. Alward, *Spatial Frequency Filtering*, IRIA State of the Art Report 2389-87-T, Willow Run Laboratory, University of Michigan, 1963.

69. J. A. Jamieson, et al., *Infrared Physics and Engineering*, McGraw-Hill, Inc., N. Y., 1963.

70. R. L. Pritchard, *Electrical Characteristics of Transistors*, McGraw-Hill, Inc., N. Y., 1967, pp. 540-43.

71. A. van der Ziel, "Noise in Junction Transistors", Proc. IRE 46, 1019-38.

72. A. van der Ziel, "Theory of Shot Noise in Junction Diodes and Junction Transistors", Proc. IRE 43, 1639-46 (1955).

73. Engineering Staff--Texas Instruments, Inc., *Solid State Communications*, McGraw-Hill. Inc., N. Y., 1966, pp. 284-285.

74. J. M. Pettit and M. M. McWhorter, *Electronic Amplifier Circuits*, McGraw-Hill, Inc., N. Y., 1961, pp. 276-80.

75. A. van der Ziel, "Thermal Noise in Field-Effect Transistors", Proc. IRE 50, 1808-12 (1962).

76. N. G. Bechtel, *A Circuit and Noise Model of the Field-Effect Transistor*. Proceedings of the International Solid State Circuits Conference, Feb. 1963.

77. L. J. Sevin, *Field-Effect Transistors*, McGraw-Hill. Inc., N. Y., 1965.

78. B. B. Snavely and J. C. Yutzy, "Impedance-Transformation Circuit for Operation at 4.2°K", Rev. Sci Instr. 38 (5), 703-704 (1967).

79. R. A. Spaulding, "Field-effect Transistor Noise at Low-Temperatures", Proc. IEEE 56 (5), 886-887 (1968).

80. *Reference Data for Radio Engineers*, 4th Edition, International Telephone and Telegraph Corp., 1956, Chapters 5, 7, 8.

81. R. W. Landie, D. C. Davis, and A. P. Albrecht, *Electronic Designers' Handbook*, McGraw-Hill, Inc., N. Y., 1957, Sec. 18.

82. R. V. Churchill, *Modern Operational Mathematics in Engineering*, McGraw-Hill, Inc., N. Y., 1944.

83. M. F. Gardener and J. L. Barnes, *Transients in Linear Systems*, John Wiley and Sons, N.Y., 1957.

84. E. J. Scott, *Transform Caiculus with an Introduction to Complex Variables*, Harper and Brothers, N. Y., 1955.

85. M. Schwartz, *Information Transmission, Modulation and Noise*, Sec. 6-2, McGraw-Hill, Inc., N. Y., 1959, Sec. 6-2.

86. H. S. Black, *Modulation Theory*, D. Van Nostrand Co., N. Y., 1953, p.3.

87. Luxenberg & Kuehn, *Display Systems Engineering*, McGraw-Hill, Inc., N. Y., 1968.

88. M. Millman, *Vacuum-Tube and Semiconductor Electronics*, McGraw-Hill, Inc , N. Y., 1958.

89. H. H. Poole, *Fundamentals of Display Systems*, Macmillan, N. Y., 1966.

90. C. E. K. Mees, *The Theory of the Photographic Process*, Macmillan, N. Y., 1954.

91. Walter Clark, *Photography by Infrared*, 2nd ed., John Wiley and Sons, N. Y., 1946.

92. W. Hyzer, *Engineering and Scientific High-Speed Photography*, Macmillan, N. Y., 1962.

93 T. Elder and J. Strong, "*The Infrared Transmission of the Atmospheric Windows*", J. Franklin Inst. 255, 189 (1953).

94. *TRAP 7 Terminal Radiation Measurements Program Servo Analysis Report*, No. 3190A, Aerojet-General Corp., Azusa, Calif., 1966.

95. M. Gardner and J. Barnes, *Transients in Linear Systems*, Vol. 1, John Wiley and Sons, N. Y., 1954.

96. C. D. Hodgman, et al., Editors, *Standard Mathematical Tables*, 11th Edition, Chemical Rubber Company, Cleveland, Ohio, 1957.

97. Robert C. Weast, Editor-in-Chief and Samuel M. Selby, Editor of Mathematics, *Handbook of Chemistry and Physics*, 48th Edition, Chemical Rubber Company, Cleveland, Ohio, 1967.

98. *Inland Design Handbook*, Inland Motor Corp., Radford, Virginia, 1964.

99. J. Truxal, *Automatic Feedback Control System Synthesis*, McGraw-Hill, Inc., N. Y., 1955.

100 H. James, N. Nichols, and R. Phillips, *Theory of Servomechanisms*, McGraw-Hill, Inc., N. Y., 1947.

101. J. Truxal, *Control Engineers' Handbook*, McGraw-Hill, Inc., N. Y., 1958.

102. G. Thaler and R. Brown, *Servomechanism Analysis*, McGraw-Hill, Inc., N. Y., 1953.

103. H. Chestnut and R. Mayer, *Servomechanisms and Regulating System Design*, Vol. II, John Wiley and Sons, N. Y., 1955.

104. W. R. Ahrendt and C. J. Savant, Jr., *Servomechanism Practice*, McGraw-Hill, Inc., N. Y., 1960.

105. D. F. Wilcock and E. R. Booster, *Bearing Design and Application*, McGraw-Hill, Inc., N. Y., 1957.

106. *Machine Design*, (Bearings Reference Issue), Vol. 35, No. 6, March 10, 1966.

107. E. E. Bisson and W. J. Anderson, *Advanced Bearing Technology*, National Aeronautics and Space Administration, Lewis Research Center, Cleveland, Ohio, 1964.

108. G. W. Michalec, *Precision Gearing: Theory and Practice*, John Wiley and Sons, N. Y., 1966.

109. *Rate Gyroscope Technical Handbook*, United States Time Corporation.

110. *A Handbook on Floated Integrating Gyros*, Reeves Instrument Corporation, 1958.

CHAPTER 4

IR SYSTEMS DESCRIPTION*

Even a cursory survey of the various infrared systems that have been produced, delivered, or proposed leaves the uninitiated with the impression of a bewildering array of different uses and techniques. In this chapter it will be shown that a coherent set of underlying principles exist which allow IR systems to be described and understood in terms of a relatively few functional parameters. These parameters will be identified and used to describe and categorize the systems. The following chapter, Chapter 5, will expand the discussion into specific system analyses and design methods. Table 4-1 defines the terms used.

TABLE 4-1. INFRARED SYSTEM DEFINITIONS

SYMBOL	TERM	DEFINITION	UNITS
A_d	Detector area	Area of sensitive surface of a detector	cm^2
A_T	Target area	Cross-sectional area of a target	cm^2
Airy disc	Central bright portion of the diffraction pattern	—	—
c	Speed of light	2.997925×10^{10} cm sec^{-1}	
D^*	Detectivity	Detector figure of merit	cm $Hz^{1/2}$ w^{-1}
D_o	Optical diameter	Diameter of a circular entrance aperture of an optical system	cm
F	f/number (f/no.)	Ratio of focal length to optical diameter	dimensionless
‡Δf	Noise equivalent bandwidth	Bandwidth of white noise which has the same integrated noise power as the total power of the actual noise	Hz
Δf_e	Electrical bandwidth	Bandwidth between the two half-power points of the amplitude response of the network	Hz

* Written by S. J. Halasz

‡ For a white noise and an ideal bandpass filter, the filter bandwidth and noise equivalent bandwidth are the same. However for nonwhite noise or practical filter, the filter bandwidth and noise equivalent bandwidth are different.

The noise spectrum of most IR detectors and systems are white or close to white (with the exception of the low frequency band below 10 or 20 Hz). The filter bandwidth and noise equivalent bandwidth have therefore, almost the same values.

TABLE 4-1. INFRARED SYSTEM DEFINITIONS (Continued)

SYMBOL	TERM	DEFINITION	UNITS
$*\Delta f_e'$	Optimum electrical bandwidth	—	Hz
f_d	Scanner spin rate	Rate of spin of a rotating mechanical scanner	revolutions per second (rps), rpm
fl	Focal length	Effective focal length of an optical system	cm
H	Irradiance	—	w cm^{-2}
ΔH	Change in spectral irradiance	Variation in irradiance caused by variation in scene temperature or emissivity	w cm^{-2}
H	Spectral irradiance	Irradiance per unit wavelength interval	w cm^{-2} μ^{-1}
h	Planck's constant	6.62554×10^{-27} erg sec	
h	Aircraft height	—	ft
$h\nu = \frac{hc}{\lambda}$	Photon energy	Energy of a quantum of radiation	erg
J	Radiant intensity	Radiant power per unit solid angle from a source	w sr^{-1}
K_v	Velocity constant	$\dot{\theta}_s/\theta_E$	sec^{-1}
K_1, K_2, K_3	Proportionality constants	—	—
N	Cycles per revolution of spinning reticle	Eq. 4-35	rev^{-1}
N_d	Number of detectors	Number of detector elements that are employed to scan the total field of view	dimensionless
N_f	Number of prism faces	Number of individual sides on the rotating prism mirror of a down-looking line scanner	dimensionless

* For a white noise and an ideal bandpass filter, the filter bandwidth and noise equivalent bandwidth are the same. However for nonwhite noise or practical filter, the filter bandwidth and noise equivalent bandwidth are different.

The noise spectrum of most IR detectors and systems are white or close to white (with the exception of the low frequency band below 10 or 20 Hz). The filter bandwidth and noise equivalent bandwidth have therefore, almost the same values.

TABLE 4-1. INFRARED SYSTEM DEFINITIONS (Continued)

SYMBOL	TERM	DEFINITION	UNITS
$NEFD$	Noise equivalent flux density	Eq. 4-3	w cm^{-2}
NEP	Noise equivalent power	Eq. 4-2	w
NET	Noise equivalent temperature	ΔT within scene element (at constant emissivity and at some scene temperature, usually 300°K) which produces a change in electrical signal level that is numerically equal to the rms electrical noise	°K, °C
P_T	Transmitted power	Radiant power generated by an illuminator	w
R	Range	Distance to target	cm
$\mathcal{R}$	Reflection efficiency	Ratio of incident flux to reflected radiant energy from an illuminated target. A product of surface reflectivity and geometric losses	sr^{-1}
T	Temperature	Temperature in absolute or Kelvin scale	°K
δT	Sample period	Pulse-width of a transmitted pulse of illumination for ranging purposes	sec
t	Dwell time	Time duration of a point in the image plane on a detector scanning across the point	sec
t_f	Frame time, rotation period of reticle	Time in which Ω is scanned	sec, min
V_{AZ}	Instantaneous voltage of scan drive in azimuth	Eq. 4-28	V
V_{EL}	Instantaneous voltage of scan drive in elevation	Eq. 4-28	V
V_o	Reference voltage	Voltage generated by an analog pickoff from the scanner drive	V
v	Aircraft velocity	----	ft sec^{-1}

TABLE 4-1. INFRARED SYSTEM DEFINITIONS (Continued)

SYMBOL	TERM	DEFINITION	UNITS
v/h	Velocity-to-height ratio	Ratio of the forward ground velocity to the altitude of an aircraft	$sr\ sec^{-1}$
W_s	Signal power	Product of irradiance and collecting aperture	w
W_λ	Spectral radiant emittance	Radiant emittance per unit wavelength interval	$w\ cm^{-2}\ \mu^{-1}$
$\bar{W}_\lambda$	Average radiant power	Average spectral power in a very narrow bandwidth	w
W_ϕ	Differential radiant emittance	Differential blackbody radiant density within a spectral bandwidth, for a small change in temperature $\int \frac{\partial W_\lambda d\lambda}{\partial T}$	$w\ cm^{-2}\ °K^{-1}$
$\Delta \bar{W}$	rms radiant power	Integral of $\bar{W}_\lambda$ for all wavelengths	w
ϵ	Emissivity	Ratio of radiant power emitted by a body to that emitted by a blackbody at the same temperature	dimensionless
$\Delta\epsilon$	Change in emissivity	---	dimensionless
ϵ_o	Optical efficiency	Fraction of incident irradiance actually transmitted by an optical system after losses to absorption, reflection, blocking, etc.	dimensionless
ϵ_s	Scan efficiency	Factor describing the ratio of dwell time of an ideal scanner to the dwell time of a given scanner	dimensionless
ϵ_λ	Emissivity for a particular λ	---	dimensionless
η	Overall optical efficiency	Factor that combines all optical losses in a system	dimensionless

TABLE 4-1. INFRARED SYSTEM DEFINITIONS (Continued)

SYMBOL	TERM	DEFINITION	UNITS
$\eta_{\lambda d}$	Quantum efficiency of detector	Ratio of number of photons that are effective in generating carriers in the detector to the total number of photons incident at the detector	dimensionless
$\bar{\eta}_B$	Average number of background photoelectrons per sample period δT	—	dimensionless
$\bar{\eta}_s$	Average number of pulse-return photoelectrons per sample period δT	—	dimensionless
θ	Instantaneous angular position	Angular position of a scanning optical axis with reference to its mid position	rad
θ'	Full cone angle of cold shield for detector	Eq. 4-11	rad
θ_{AZ}	Azimuth angular position of detector with respect to center of Ω	Eq. 4-27	rad
θ_E	Error angle of scanner, instant angular displacement error	Eq. 4-25	rad
$\theta_{E(max)}$	Maximum error angle of scanner	Eq. 4-39	rad
θ_{EL}	Elevation angular position of detector with respect to center of Ω	Eq. 4-27	rad
θ_o	Constant reference angle	Eq. 4-27	rad
θ_S	Angular position of optical axis (center of field of view)	Eq. 4-25	rad
θ_T	Target angular position	Eq. 4-25	rad
$\dot{\theta}$	Direction of missile velocity	Eq. 4-26	rad sec^{-1}

TABLE 4-1. INFRARED SYSTEM DEFINITIONS (Continued)

SYMBOL	TERM	DEFINITION	UNITS
$\dot{\theta}_0$	Rate of output angular displacement	—	rad sec^{-1}
$\dot{\theta}_{s(m,v)}$	Maximum steady state tracking rate available from servo	Eq. 4-39	rad sec^{-1}
$\ddot{\theta}_0$	Output acceleration	—	rad sec^{-2}
$\dot{\theta}_{T:M}$	Line of sight angular rate in inertial coordinates	Eq. 4-26	rad sec^{-1}
θ_α	Down-looking scan angle	Angular coverage perpendicular to the line of sight of a down-looking line scanner	rad
λ	Wavelength	Electromagnetic wave interval	cm
$\Delta\lambda$	Spectral band	Half-power spectral bandwidth	cm
λ_m	Longest significant wavelength in an expression for diffraction limited operation	—	cm
ν	Frequency	—	sec^{-1}
τ_A	Atmospheric transmission	Fraction of radiance transmitted by the intervening atmosphere between source and instrument	dimensionless
Φ	Beam angle	Half-power subtended angle of a beam	rad
Ω	Total field of view	Solid angle that is repetitively sampled or scanned by one or more detector elements	sr, $mrad^2$

TABLE 4-1. INFRARED SYSTEM DEFINITIONS (Continued)

SYMBOL	TERM	DEFINITION	UNITS
ω	Instantaneous solid angle field of view	Small solid angle subtended by a detector element in the focal plane of an optical system	sr
ω_B	Solid angle through which ambient radiation enters detector	Eqs. 4-9, 4-9a	sr

4-1 PASSIVE SYSTEMS

For purposes of parametric description, it is convenient to separate passive IR systems into two broad categories: (1) scanning systems such as search, track, and imaging systems, and (2) measurement systems such as spectrometers, radiometers, and interferometers. Terms used in the discussion of passive systems are identified in Table 4-1.

Scanning systems scan and sample the radiant intensity distribution within a designated field. Their output may be a linear analog of this distribution for imaging purposes, or a simple indication of the presence of a target and its location as in search-track applications.

Measurement systems are intended for performing a measure on the irradiance directed at the system. The measure may be as simple as intensity level within some fixed radiation bandwidth, or a sequential measure of intensity within a variable center radiation bandwidth, or a simultaneous measure of intensity over a number of center bandwidths.

4-1.1 IMAGING SYSTEMS

At the present state-of-the-art in infrared imaging technology, the most frequent emphasis is on thermal imaging with systems using mechanical scanning in the 10-micron region, and somewhat less frequently in the 3- to 5-micron region. (Image tubes using electron focusing of a near-IR photo-emissive cathode onto a phosphor screen are widely used for sights and viewing devices. These form a separate category which is described in par. 4-1.1.2.*)

The major portion of the IR scene radiance is in the 10-micron region, and sensitive detector elements and arrays as well as efficient optical components and detector cooling mechanisms are now available for this spectral region. Thus, it is possible to obtain maximum sensitivity to thermal variations in the structure of the scene.

A block diagram for a typical imaging system is shown in Fig. 4-1. The basic elements are a scanner which scans the scene (an object plane which can be assumed for all practical purposes to be located at infinity) with a sensor which responds to the IR radiance level within the instantaneous field of view of each detector element; amplification and filtering to convert the detector outputs into electrical signals to modulate corresponding light sources; and a display scan that moves the light sources in angular synchronism with the sensor scan in order to "paint out" a visible radiance distribution that is analogous to the scanned scene. The display scan may be the electron beam deflection on a CRT, or the optical deflection of a

* Other image conversion devices such as the Absorption Edge Image Converter, the Evapograph, and the Thermosensitive Phosphor Imaging System rely on some physical change in state with temperature that can be observed optically[1,2]. They have had limited applicability to practical military systems. Baird-Atomic, Inc., Cambridge, Mass., has developed the Evapograph into a commercial instrument.

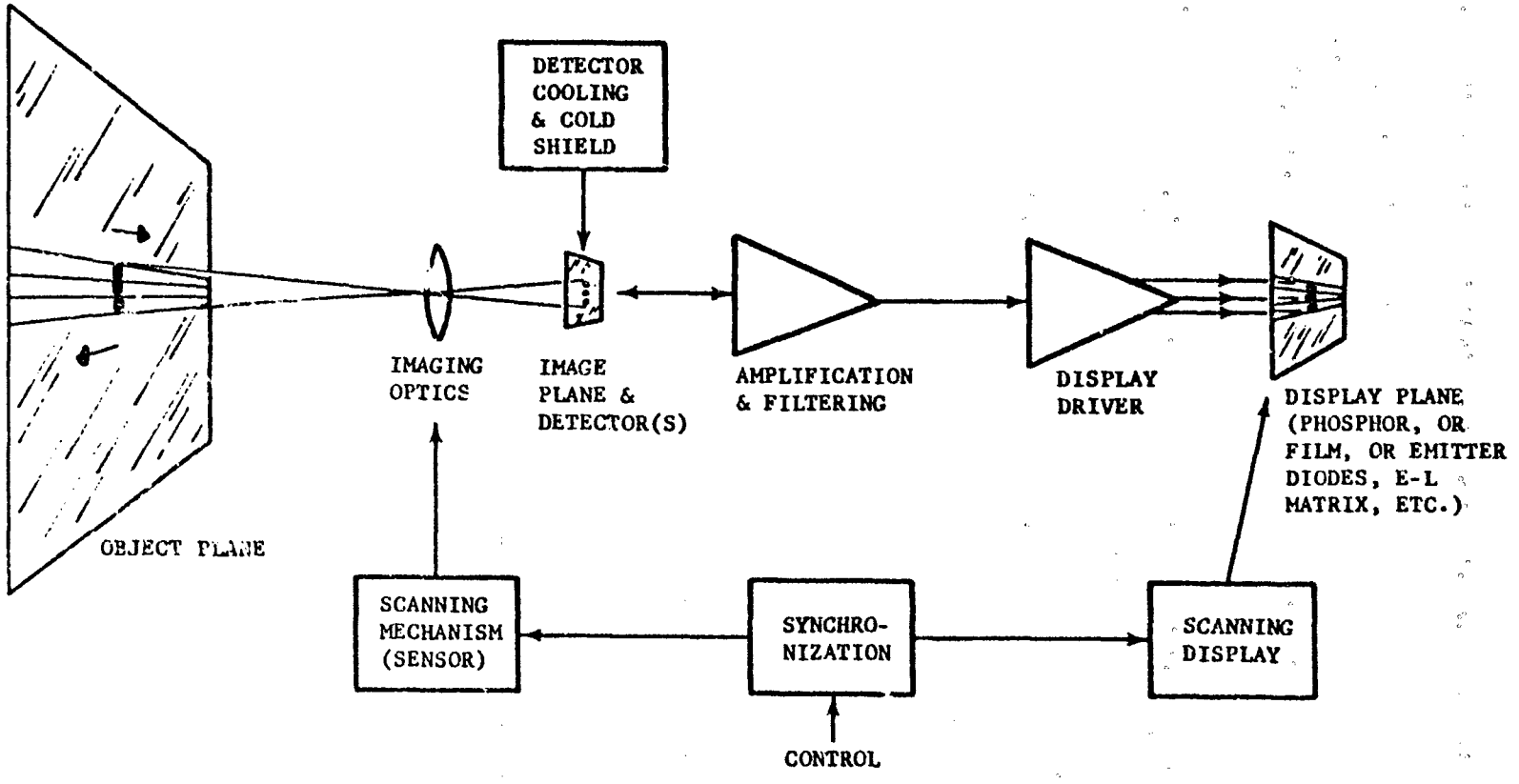


FIGURE 4-1. Typical Imaging System Block Diagram

light beam focused on a strip of film. If the scan rate is rapid enough to scan an entire designated scene repeatedly at a rate higher than the flicker frequency of the human eye, then an integrating display surface such as film or long-persistence phosphor is not needed. The display plane can be simply the focal plane onto which scanning modulated light sources such as diode emitters are focused. To the human observer the visible image will appear to reside in that plane.

Supporting system functions consist of the IR scanning mechanism, detector cooling mechanism (if needed), and a multiplexing subsystem for commutating the outputs of many detectors into one CRT (if needed). In some systems it may be desirable to transmit the scanner output signals to a remote recording and/or display station. This of course requires a transmitter-receiver data link to be included in the loop but does not change the generality of the approach.

4-1.1.1 Mechanical Scanners

For slow scanners, the simplest scan technique might be to move the optical telescope assembly on its gimbals through some programmed raster scan, spiral scan, or any path which avoids redundant scan and gaps in the field. This is adequate where many seconds or minutes are available for the scan of one frame. For higher speed scanning, schemes must be devised where less physical mass is to be moved. Common techniques are to direct the optical viewing path of the sensor at oscillating plane-mirrors located outside the collector aperture, or through counterrotating prisms as described in Chapter 3. Down-looking line scanners that operate from aircraft utilize the aircraft's motion to achieve scan in the forward direction, and a spinning four-sided mirror whose rotation axis is parallel to the direction of motion for scanning in the direction perpendicular to the line of flight.

In this chapter scanning imaging systems will be described in terms of functional parameters. First the sensitivity equations will be derived. There are two basic approaches to describing the sensitivity of the detector within the scanner. One approach is based on utilizing the detector figure of merit D^* which was described in par. 3-4. This approach regards the detector as a source of noise of a fixed level, the level being determined from measurements under conditions similar to the conditions to be encountered by the detector when it is installed within the scanner. The other approach which is suitable only for BLIP detectors, regards the random arrival rate of photons from the background as being the ultimate source of detector noise. (Refer to par. 3-4 for definition of BLIP or background-limited operation.) Here the detector noise-power-density is calculated on the basis of photon rate and detector quantum efficiency. For situations where the number of signal photons is small in comparison to the number of background photons, the two approaches are equivalent since BLIP D^* is determined by background photon rate. For situations where the number of signal photons is larger than the average number of background photons, the noise due to the random arrival time of the signal photons predominates and D^* has no meaning. This situation, however, is less likely to occur in imaging systems than in systems where bright objects are located against dark backgrounds, as for example, in star trackers.

After deriving the sensitivity equations the optical parameters will be analyzed, then the effect of scan rate on system bandwidth will be described. Finally the sensitivity equation will be presented in parametric form, showing the trade-offs among operational parameters such as scan rate, field of view and resolution; and instrumentation parameters such as optical collector diameter, number of detector elements, and optical efficiencies.

4-1.1.1.1 Scanner Parameters

4-1.1.1.1.1 Noise Equivalent Temperature (Sensitivity)

Customarily, image forming scanners are not designed to respond to the absolute radiance level of a scene element, i.e., they are not designed for dc response. They are, however, designed to follow the changes in scene radiance as the scene is being scanned, down to very low frequencies corresponding to a gradual change in level over a large portion of the scan. Dc restoration may be used in the circuitry to restore the average scene level to some arbitrary average intensity level at the display output. There are two major reasons for not designing for absolute level or dc response. The first is that the possible range of radiance levels in a scene

spans a far greater range of levels than that which can be reproduced in a practical display system. Frequent periods of no brightness or complete saturation of the display output could result if absolute levels were maintained. The second reason is that both detectors and high gain dc amplifiers exhibit drift and high $1/f$ noise near zero frequency. (In cases where absolute radiance level measurement is a requirement, reference blackbody sources are supplied within the system, and provisions are made to allow the scanner to periodically scan them and present a reference output.)

The signal in the scene that is of significance to the imaging sensor is therefore the variation in irradiance H caused by variation in scene temperature or emissivity. This is expressed by

$$\Delta H = \frac{\omega}{\pi}\left(\Delta T \int_0^{\infty} \frac{\partial W_\lambda}{\partial T}\, d\lambda + \int_0^{\infty} \frac{\Delta \epsilon_\lambda}{\epsilon_\lambda}\, W_\lambda d\lambda\right), \text{ w cm}^{-2} \qquad (4\text{-}1)$$

where

ω = instantaneous field of view, sr

T = temperature, °K

ϵ = emissivity

W_λ = spectral radiant emittance, w $\text{cm}^{-2}\,\mu^{-1}$

If the system noise is mainly detector noise, it can be expressed in terms of "noise equivalent flux density". This is the minimum signal irradiance that will produce a peak signal-to-rms-noise ratio of one.

The system noise equivalent power (NEP), in watts, is obtained from the figure of merit for the detector, denoted as D^*. This relationship can be expressed as

$$NEP = \frac{\sqrt{A_d \Delta f}}{D^*}, \text{ w} \qquad (4\text{-}2)$$

where

A_d = area of sensitive surface of detector, cm^2

Δf = system noise bandwidth, Hz

D^* = detector figure of merit, cm $\text{Hz}^{1/2}\ \text{w}^{-1}$

The signal power on the detector is equal to the optical collector area times the signal flux density. Therefore the system noise equivalent flux density ($NEFD$) is

$$NEFD = \frac{4\sqrt{A_d \Delta f}}{\pi D_o^2 D^* \epsilon_o}, \text{ w cm}^{-2} \qquad (4\text{-}3)$$

where

D_o = diameter of aperture of optical system, cm

ϵ_o = optical efficiency

and other terms as previously defined.

Both ΔT and $\Delta \epsilon$ contribute to the differential radiant emittance ΔH of a thermal scene. However by convention only the ΔT term in Eq. 4-1 is considered in sensitivity calculations. If the first integral in this equation is abbreviated to W_ϕ then Eq. 4-1 can be reduced to

$$\Delta H = \frac{\omega}{\pi} \Delta T W_\phi, \text{ w cm}^{-2} \qquad (4\text{-}4)$$

The instantaneous field of view ω is defined by the area of the detector A_d and the optical focal length fl

$$\omega = \frac{A_d}{(fl)^2}, \text{ sr} \qquad (4\text{-}5)$$

Substituting Eq. 4-5 into Eq. 4-3 gives

$$NEFD = \frac{4F\sqrt{\omega \Delta f}}{\pi D_o D^* \epsilon_o}, \text{ w cm}^{-2} \qquad (4\text{-}6)$$

where F is the f/no. of the optics.

The signal-to-noise (S/N) generated by the scanner is given by the ratio of ΔH to $NEFD$

$$S/N = \frac{\Delta T W_\phi D_o D^* \epsilon_o \sqrt{\omega}}{4F\sqrt{\Delta f}} \qquad (4\text{-}7)$$

where all terms previously have been defined.

It should be noted that the spectral distribution of $\frac{\partial W_\lambda}{\partial T}$ is not the same as that for W_λ. For example, a 300°K blackbody, or a scene at this average temperature, will have its peak spectral energy density at approximately 10 microns, whereas the partial derivative with respect to temperature at an average temperature of 300°K will have a peak spectral energy density at approximately 8 microns. The difference is important when optical transmitting elements, filters, and detector responses are under consideration.

The ΔH actually seen by the sensor is modified by the spectral transmission of the optical components and by the spectral transmission of the atmosphere. Generally, the integrals in Eq. 4-1 should also contain these factors as multiplying elements. In practice, however, sufficient accuracy can usually be achieved by applying averaged transmission values to these factors and retaining them as constant multipliers or as a single constant outside the integrals.

The sensitivity of an image forming scanner is expressed by "noise equivalent temperature", and is abbreviated as NET. It refers to that ΔT within the scene element (at constant emissivity and at some average scene temperature, usually 300°K) which produces a change in electrical signal level that is numerically equal to the rms electrical system noise. In a well-designed system, the system noise is mostly due to detector noise. The NET is determined by solving for ΔT in Eq. 4-7 for a signal-to-noise ratio equal to one. By incorporating the atmospheric transmission τ_A in Eq. 4-7, NET is obtained as

$$NET = \frac{4F\sqrt{\Delta f}}{\sqrt{\omega} D_o D^* \epsilon_o \tau_A W_\phi}, \text{ °K} \qquad (4\text{-}8)$$

In the derivation of Eq. 4-8, it was assumed that the value of the second term in the right-hand side of Eq. 4-1 $\int_0^\infty \frac{\Delta \epsilon_\lambda}{\epsilon_\lambda} W_\lambda d\lambda$, is negligible. Given a particular imaging system whose parameters are known, Eq. 4-8 will express the temperature sensitivity of the system. Implicit assumptions are that the scene area over which the temperature differential ΔT occurs is significantly larger than the instantaneous field of view, and the detector D^* refers to the detector inside the instrument under its normal operating conditions. The D^* of a particular detector type will vary considerably from the generally published data owing to such factors as manufacturing process control for unusual sizes, cold-shielding efficiency, nonoptimum electrical bias and loading conditions, aging effects, background radiance effects, and aerodynamically heated optical window radiance.

In Eq. 4-8 the system noise was derived on the basis of detector D^*. For BLIP operation it is sometimes more useful to derive the rms noise value on the basis of the random fluctuation in arrival rate of incident photons on the detector. To obtain this photon noise, Poisson statistics are usefully invoked, which state that

the mean square value of the number of photons in a sample time period is equal to the average number of photons in that sample interval. If the average radiant power in a narrow spectral bandwidth is $\overline{W}_\lambda$, the average number of photons per period is $\overline{W}_\lambda \, \delta T/(h\nu)$ where δT is the sample period. The rms variation in the power is, therefore, $\sqrt{\overline{W}_\lambda h\nu/(\delta T)}$, w. Similarly, if the detector sees an average background spectral irradiance H_λ, it will see a random variation whose rms value is given by

$$\Delta \overline{W} = \left[2\Delta f \left(\frac{\omega_B}{\pi}\right)\left(\frac{\pi}{4}\right) D_o^2 \epsilon_o \int_{\lambda_1}^{\lambda_2} \eta_{\lambda d} H_\lambda \left(\frac{hc}{\lambda}\right) d\lambda \right]^{1/2}, \text{ w} \tag{4-9}$$

where ω_B is *not* the solid angle subtended by the detector at the optical focal length, but the solid angle through which ambient radiation enters the detector. The term $\eta_{\lambda d}$ represents the quantum efficiency of the detector. It is the ratio of the number of photons that are effective in generating carriers in the detector to the total number of photons incident at the detector. The bandwidth Δf represents the effective noise bandwidth of the system, and the factor 2 is a consequence of the negative frequencies in the power spectrum of the noise. For photoconductive detectors, where both generation and recombination of carriers occur, an additional factor of 2 should be included under the radical. Eq. 4-9 when applied to a photoconductor can be simplified to

$$\Delta \overline{W} = \left[\Delta f \omega_B D_o^2 \epsilon_o \int_{\lambda_1}^{\lambda_2} \eta_{\lambda d} H_\lambda \left(\frac{hc}{\lambda}\right) d\lambda \right]^{1/2}, \text{ w} \tag{4-9a}$$

As previously indicated, in Eq. 4-9, ω_B is *not* the solid angle subtended by the detector at the optical focal length, but the solid angle through which ambient radiation enters the detector. For a detector without cold shielding this angle is effectively π steradians. If a cold shield with a round aperture is employed, the value of ω_B is

$$\omega_B = \pi \sin^2 \theta'/2 \tag{4-10}$$

where θ' is the full cone angle of the shield. The lower limit to this angle is set by the numerical aperture of the collector optics.

$$\sin\left(\theta'/2\right) = \frac{1}{\sqrt{4F^2 + 1}} \tag{4-11}$$

and $$\omega_B(\text{minimum}) = \frac{\pi}{4F^2 + 1} \tag{4-12}$$

The signal power W_s is the product of the irradiance (Eq. 4-1) and collecting aperture:

$$W_s = \frac{\omega \Delta T D_o^2 \epsilon_o}{4} \int_{\lambda_1}^{\lambda_2} \eta_{\lambda d} \left(\frac{\partial W_\lambda}{\partial T}\right) d\lambda, \text{ w} \tag{4-13}$$

To obtain *NET* find the case where $W_s = \Delta \overline{W}$, i.e., the signal power is equal to background noise power. From Eqs. 4-9 and 4-13 we obtain

$$\frac{\omega \Delta T D_o \epsilon_o}{4} \int_{\lambda_1}^{\lambda_2} \eta_{\lambda d} \left(\frac{\partial \overline{W}_\lambda}{\partial T}\right) d\lambda = \left[\Delta f \omega_B D_o^2 \epsilon_o \int_{\lambda_1}^{\lambda_2} \eta_{\lambda d} H_\lambda \left(\frac{hc}{\lambda}\right) d\lambda \right]^{1/2}$$

NET or ΔT is, therefore,

$$\Delta T = \frac{4\left[\Delta f \omega_B \int_{\lambda_1}^{\lambda_2} \eta_{\lambda d} H_\lambda \left(\frac{hc}{\lambda}\right) d\lambda \right]^{1/2}}{\omega D_o \sqrt{\epsilon_o} \int_{\lambda_1}^{\lambda_2} \eta_{\lambda d} \tau_A \left(\frac{\partial W_\lambda}{\partial T}\right) d\lambda} \tag{4-14}$$

4-1.1.1.1.2 Optical Gain and Resolution

From a viewpoint of system analysis, however, the parameters in Eq. 4-8 are not all independent. In the first place, the optical diameter D_o has to be sufficiently large to make diffraction effects compatible with the optical resolution requirements.

The central bright portion of the diffraction pattern, the Airy disc, contains 84 percent of the energy of an imaged point source. In a diffraction limited system this Airy disc should not be larger than the size of an individual detector element.

A classical Airy disc is produced only by monochromatic radiation, but an IR imaging system will be sensitive to the entire available spectral bandwidth, typically the 8.5- to 14-micron atmospheric window or the 3.2- to 5.0-micron window. Therefore, the intensity distribution within the diffraction pattern will be a superposition of the weighted values of all of the incoming wavelengths. The contribution from the edges of the window will be less significant than that from the central portion of the window. This intensity distribution is further modified by residual optical aberrations and imperfections. This intensity distribution or spot spread function is, therefore, not readily predictable analytically.

The focal plane image is a convolution of the object scene with this spread function. The scanning of this image plane by the detector represents, in effect, a second convolution between the detector area and the image. Each convolution of course degrades the image quality to some extent.

A basis for a rigorous relationship between optical spread function and detector area cannot be provided. An optimum error budget and cost budget exists if the spread function is approximately equal to the detector area. This criterion can be formulated approximately by requiring that the Airy disc diameter formed by the longest significant wavelength in the window, e.g., the one-half power point on the atmospheric transmission curve, be equal to or less than the size of the detector.

If the detector area is square, and it subtends an instantaneous solid angle field of view ω, then according to Rayleigh criteria:

$$\sqrt{\omega} = \frac{2.44\lambda_m}{D_o} \tag{4-15}$$

where λ_m is the longest significant wavelength in an expression for diffraction limited operation. The the *NET* of a diffraction limited imaging system is

$$NET = \frac{1.6F\sqrt{\Delta f}}{\lambda_m D^* \epsilon_o \tau_A W_\phi} \quad ,°K \tag{4-16}$$

where Δf is noise equivalent bandwidth. Thus the sensitivity of a diffraction limited system is independent of optical diameter. The trade-off is only between optical diameter and resolution.

4-1.1.1.1.3 Scan Rate and Bandwidth

The electrical bandwidth Δf_e is determined by the scan rate. It is proportional to the reciprocal of the dwell time of an image point on the scanning detector. The electrical filter can be optimized, in principle, to maximize the signal-to-noise ratio, if the characteristics of the signal and noise can be completely described. This process is described in Chapter 3 and in Ref. 3. In practice, however, the optimum filter is usually not realizable physically, and in an imaging system utilizing many detectors and amplifiers, even a piecewise approximation to the optimum filter for each detector channel might require an excessive number of electronic components. A relatively simple bandpass filter consisting of a flat response with low- and high-frequency cutoff is often adequate and provides a signal-to-noise ratio very nearly equal to the theoretical optimum. A few decibels of peaking near the high frequency end of the filter will result in improved definition of small details in the scene, at the expense of a slight increase in noise.

At this point in the filter design, many design trade-offs are possible. Most of these trade-offs will relate to the aesthetic appeal of the displayed scene or to compensating for system component characteristics. For example, high frequency enhancement, which is analogous to differentiation, will enhance the edges of objects within the scene by causing light and dark banding in the direction of scan. In the extreme, the scene can be reduced very nearly to a stick drawing over a uniform intensity background level. Or, contrast can be traded for fine-detail

definition by adjusting the slope of the high-frequency cutoff. Compression and expansion can be implemented to compensate for non-linearities and dynamic range in the display element. If bandwidth compression is an important consideration, the statistical constraints among picture samples and the properties of human vision can be exploited to achieve an order of magnitude reduction in bandwidth [4].

In most instances, however, the desire is for a displayed output scene whose intensity distribution is a reasonably linear analog of the thermal intensity distribution in the scanned scene, and to obtain this with maximum economy. If a simple low-pass filter were used, its optimum electrical bandwidth $\Delta f_e'$ for best peak signal-voltage-to-rms-noise ratio would be approximately[5]

$$\Delta f_e' = \frac{3}{4t} \text{ , Hz} \qquad (4\text{-}17)$$

The constant of proportionality of 3/4 between the bandwidth and the reciprocal of dwell-time t is not rigorous because the assumed signal pulse does not have zero rise time and because the bandwidth does not have a mathematically sharp cutoff. But an important factor to be kept in mind for this type of analysis is that the constant of proportionality is not at all critical in value. Goldman[5] has shown, for example, that, if the bandwidth is reduced by a factor of one-third (to a proportionality constant of 1/2), the peak signal-to-noise ratio is reduced only by 6.5 percent and, if the bandwidth is increased by one-third (to a proportionality constant of 1), the signal-to-noise ratio is reduced only by 3 percent. Therefore the approximation of Eq. 4-15 is well within the accuracy requirements of a system trade-off study.

If a total field of view of Ω steradians is scanned at a constant rate, without redundancy, by an instantaneous field of view of ω steradians and the scan is completed in a frame time period t_f, the dwell time t of a point on the detector is $t_f \times \omega/\Omega$. Because of mechanical scan rate limitations, a linear scan rate is usually difficult to achieve, with the result that the dwell time in some parts of the field will be reduced by a factor ϵ_s called scan efficiency. This requires an increased electrical filter bandwidth, and is commonly expressed as scan efficiency or scan overshoot. The required bandwidth is then

$$\Delta f_e'' = \frac{3\Omega\epsilon_s}{4\omega t_f} \text{ , Hz} \qquad (4\text{-}18)$$

In reality this refers to the upper cutoff frequency because it is impractical to design for dc response. However, the lower cutoff frequency is typically very low—of the order of a few Hertz—in order to allow the scanner to accurately reproduce the low frequency features within the scene. The upper cutoff frequency is typically of the order of tens of kHz so that it is numerically very nearly equal to the total bandwidth. If a number N_d of detectors are used to simultaneously scan the field Ω, then the required bandwidth $\Delta f_e''$ is proportionately reduced.

The equivalent noise bandwidth Δf will generally be somewhat larger than $\Delta f_e''$ owing to excess $1/f$ noise, amplifier noise, and the gradual cutoff of simple R-C filters. The equivalent noise bandwidth is obtained by integrating the noise power spectrum in $\Delta f_e''$ and equating it to the integration of an ideal white noise power spectrum.

4-1.1.1.1.4 Performance Requirements and Instrumentation Parameters

Expressed in terms of scan rate, the *NET* of a scanning system is

$$NET = \left(\frac{1}{\omega}\sqrt{\frac{\Omega}{t_f}}\right)\left(\frac{F}{D_o D^* W_\phi \sqrt{N_d}}\right)\left(\frac{2\sqrt{3\epsilon_s}}{\epsilon_o \tau_A}\right) \qquad (4\text{-}19)$$

The terms in Eq. 4-19 have been separated into three groups by the use of parentheses in order to illustrate the nature and effect of the different parameters involved.

a. First Group of Terms:

The first group, ω, Ω, and t_f—which are instantaneous field*, total field, and frame time, respectively—represents the system performance requirements. These parameters are governed by

* Instantaneous field is also commonly referred to as the "resolution" of the system since it relates to the resolvable detail within the scene. It is important to note that this is not the same definition of resolution as the Rayleigh criterion, or those definitions that relate to photographic or television resolutions.

the operational or tactical demands on the system and are derived from a larger system analysis such as that for an integrated fire control-navigation system, weapon delivery system, etc.

b. Second Group of Terms:

The terms in the second set of parentheses represent the instrumentation parameters, or tools, which the designer has at his disposal for synthesizing an IR imaging system to meet the requirements. Actually, he will find that the flexibility and range of values at his disposal are quite restricted and that he cannot design for any arbitrary set of performance specifications. Similarly, the system analyst who defines the performance requirements soon becomes aware that he cannot request any arbitrary set of performance specifications. He, therefore, must perform judicious trade-offs among the performance parameters in order to constrain the net demands to the state-of-the-art in instrumentation. For example, Eq. 4-19 (first parentheses) shows that sensitivity or *NET* trades off as the square root of scan rate, and inversely as resolution or instantaneous field of view. Thus if high resolution is important, then it may be necessary to relax scan rate. If frame time cannot be increased, then the total field of view that is to be scanned might be decreased. Or, the *NET* specification may have to be relaxed.

Among the instrumentation parameters in the second set of parentheses, the number of detector elements N_d is probably the only parameter that allows for a wide range of values. The number of detectors used in imaging systems vary from one to a few hundred. The other parameters are less flexible.

Optical diameter D_o is typically of the order of a few inches to a foot. Weight and cost increase much more rapidly than gains in sensitivity for sizes larger than this range. On the other hand, for optical sizes much smaller than this range, system sensitivity and resolution decrease much more rapidly than overall system cost or weight. Therefore, practical consideration constrains optics diameter to this range of sizes. The optimum diameter depends on a close analysis of the overall instrumentation approach.

The f/number for sensitive systems is similarly constrained to a narrow range of values, typically between f/1.5 and f/3.0. Again, values significantly less than f/1.5 are very difficult and costly to achieve, and the cost trade-off is poor. For narrow field of view, high-resolution systems, analogous to telephoto lens cameras, a longer focal length and, therefore, a higher f/number is necessary. A practical lower limit to the size of detector elements that can be fabricated is of the order of a few thousandths of an inch. Therefore, a limit is reached where decreased instantaneous field can only be achieved by increasing the focal length.

The selection of the detector type to be used is one of the most critical steps in the design process. The choice depends not only on the wavelength band of operation and available D^* but also on cooling requirements, cost and availability in the specific configuration called for, reliability and aging characteristics, and particular characteristics such as time constant, responsivity, sensitivity contours and uniformity, spectral filtering requirements, etc.

The last parameter in this group, W_ϕ, is the differential radiant emittance from the scene. In the 3- to 5-micron window for an average scene temperature of 240°K, W_ϕ has a value of approximately 1.5×10^{-5} w cm^{-2} °K^{-1}, depending considerably on atmospheric and meteorological conditions. In the 8.5- to 14-micron window, the value is approximately 2.1×10^{-4} w cm^{-2} °K^{-1}, again depending on the atmospheric absorption path and meteorological conditions. Thus, the available energy is considerably higher in the 10-micron region; however, cooling requirements are usually more severe for detectors operating in this region than for those operating in the 3- to 5-micron window. The optical diffraction limit is worse for the longer wavelength region, and optical materials such as refractors and filters tend to be significantly more costly for the longer wavelengths.

c. Third Group of Terms:

The terms in the last set of parentheses represent specific engineering design efficiency parameters. Optical transmission efficiency ϵ_o and scan efficiency ϵ_s are parameters that have to be maximized during the detailed design layout phase. Optical transmission in particular can become a critical or pacing item, especially in the longer wavelength system. Optical materials such as germanium or zinc sulfide have very high indexes of refraction, resulting in high

reflection losses at each surface. For example, if a fast, high-resolution optical system design requires four air-spaced refractive elements and if a transmission loss of 50 percent were suffered at each two surfaces, the net transmission would be reduced to $(0.5)^4$ or 0.0625 owing to this effect alone. To make up for this loss by increasing the optical diameter would require an increase of D_o by a factor of almost 16 which, of course, would be absurd. A multi-layered anti-reflection coating can reduce the reflection loss to about 2 percent at each surface, depending on the specific material to be coated and the width of the spectral band that is to be subtended. A 2 percent reflection loss per surface would result in a net transmission of 0.92 percent.

The atmospheric transmission τ_A is actually beyond the control of the system designer. It is merely a factor to be considered in the analysis and selection of the wavelength band of operation.

4-1.1.1.2 Thermographs

A thermograph is essentially a scanning radiometer. (See par. 4-1.4 for a description of radiometers.) A typical configuration produced by the Barnes Engineering Company utilizes an oscillating plane mirror mounted at 45 deg in front of the 8-in. diameter radiometer optics. The mirror is cam-driven in a horizontal scan with a quick return. During the return, the mirror tilts up one linewidth, thus generating a line-by-line scan of the scene. The back of the same plane mirror scans a light spot from a glow modulator tube over photographic film. The light spot is made proportional in size to the instantaneous field of the radiometer, and is intensity modulated by the radiometer output. Since the same mirror scans both the IR scene and the light output, excellent synchronization is maintained between the two.

The dynamic range and average intensity level of the output can be varied by electronic controls. The radiometer measures the level of the scene radiance by comparing it to a built-in reference blackbody source. It also supplies a set of reference signal levels which register as reference gray scale levels on the film.

By utilizing an immersed thermistor bolometer, characteristic performance is a *NET* of 0.04°C for $\omega = 1$ mrad^2, $\Omega = 10° \times 10°$, or 0.03 sr; and $t_f = 17$ min. Instruments similar to the thermograph are also produced by Servo Corporation of America and Radiation Electronics Corporation.

4-1.1.1.3 Down-looking Line Scanners

For air-to-ground IR reconnaissance mapping, a down-looking line scanner is a configuration that utilizes the forward motion of the aircraft as one dimension of scan. The instantaneous field ω is caused to scan through an angle θ_a at right angles to the direction of motion. Typically, θ_a is made large enough to provide nearly horizon-to-horizon coverage. The rate of scan is adjusted to the forward speed of the aircraft so that successive scans cover adjacent strips on the ground.

To provide the scan pattern, a rotating mirror in the form of a prism with a number of faces is placed in front of the imaging optics. As each face rotates past the optics, the optical line of sight is deviated along with it. Two basic configurations are: (1) faces at 45 deg to the optics, and (2) a folded system where the faces on the prism are parallel to the direction of motion, and the receiving optics are separated into two halves to receive the reflection from two sides of the rotating prism.

The scan rate for this configuration is equivalent to

$$\text{scan rate} = \frac{\theta_a v}{h} \text{ , sr sec}^{-1} \qquad (4\text{-}20)$$

where v and h are aircraft velocity and height, respectively, and θ_a is expressed in radians. The dwell time on the detector is approximately

$$t = \frac{\omega v}{\theta_a h} \text{ , sec} \qquad (4\text{-}21)$$

and the spin rate of the mirror that is required is

$$f_s = \frac{60v}{\omega N_f h} \text{ , rpm} \qquad (4\text{-}22)$$

where N_f is the number of sides on the rotating prism mirror. For low-altitude, high-speed flight, the work load imposed on the scanner is relatively severe. For example, for a v/h ratio of one steradian per second (i.e., 1000-ft altitude and 1000-ft sec^{-1} velocity) and an instantaneous

field of 1 mrad², the scan rate is approximately π sr sec^{-1}, the dwell time t is approximately 3×10^{-7} sec, and the spin rate f_d for a four-sided mirror ($N_f = 4$) is 15,000 rpm. Using more than one detector relieves some of the work load, but increases the distortion and scene rectification problem.

Fig. 4-2 illustrates one part of the distortion. The angular resultion of the detector is constant,

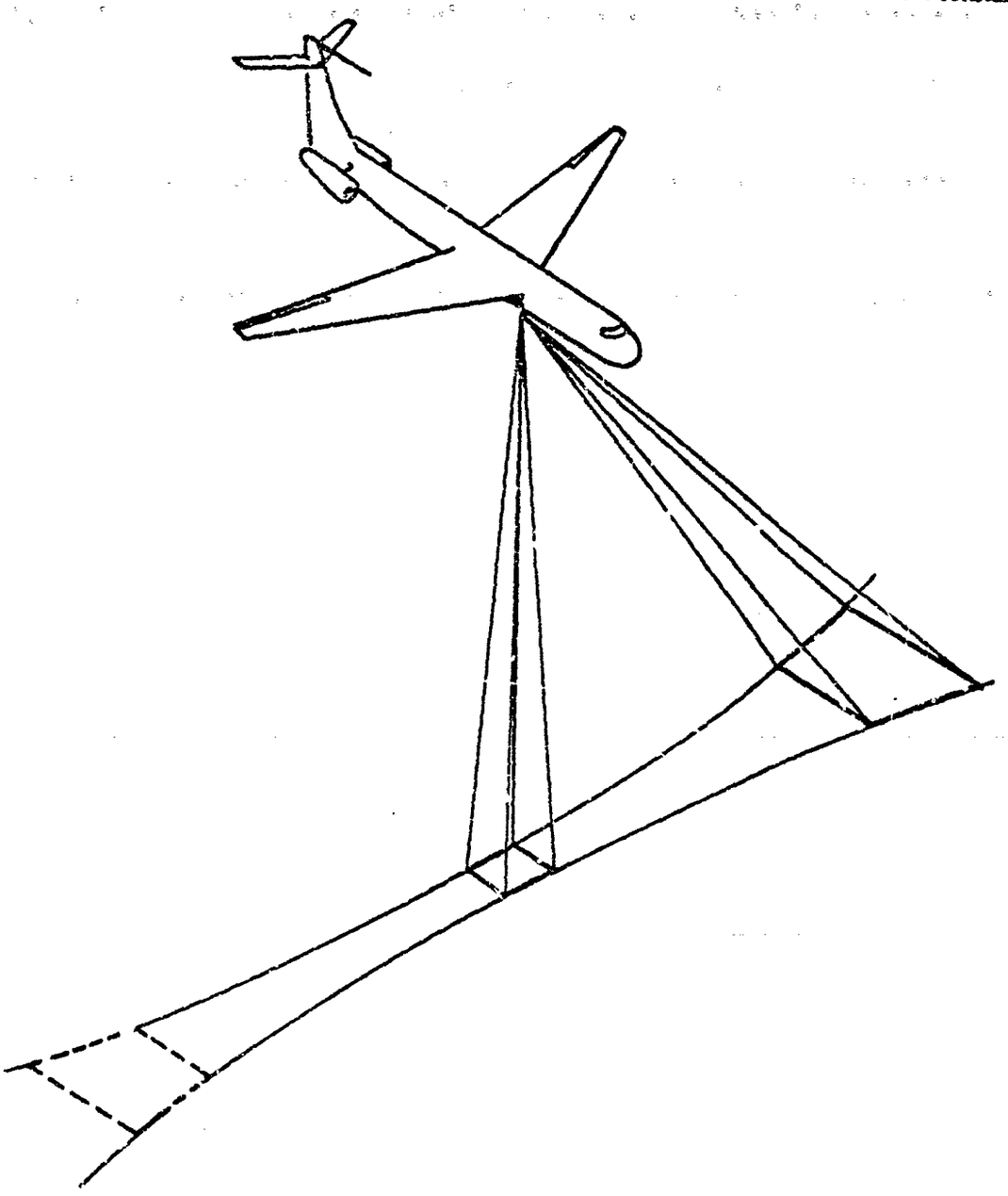


FIGURE 4-2. End of Scan Distortion

but its projected area on the ground is not. It increases in the direction of travel as a secant law of the viewing angle, and in the direction of scan as a secant squared law. Thus, toward the end of the scan, two effects become pronounced: there is overlap or crosstalk from line-to-line, and foreshortening occurs in the direction of scan. Objects having a third dimension up from the plane of the earth suffer additional distortions. With a linear array of detectors these effects become more aggravated.

The line scanner display is produced by scanning a light spot across a strip of film. The film moves at a rate corresponding to the forward velocity of the aircraft, and the light spot scanner is mechanically coupled to the infrared scan drive for synchronization. The light spot is formed by a glow modulator tube or a solid state diode emitter and is modulated by the detector signal. The projection scheme of the light spot and its scanner may be configured in ways to cancel or rectify some of the scan distortion.

4-1.1.1.4 Forward-looking Infrared Systems

Forward-looking infrared systems, or FLIR, is a term applied to an IR imaging system used on aircraft. FLIR systems look in the forward direction of flight and produce a flicker-free display in real-time for target identification and navigation purposes.

In this situation a fairly large total field of view has to be scanned at frame rates above the flicker frequency of the human eye. The forward motion of the aircraft cannot be used as a part of the scan because the equipment is generally gimbal-mounted to allow viewing in various directions within the forward hemisphere.

Implementation of a scan technique presents engineering problems in this case. For good quality imagery, a few hundred scan lines within the frame are desirable. At a frame rate of 15 to 20 frames per sec, this implies a few thousand line scans per second. It would be impractical to attempt this kind of scan with a single detector because of the complex, high-speed mechanical scan motions that would be required. Therefore, FLIR systems generally use many detector elements. One straightforward approach is to arrange a few hundred detector elements in a vertical linear array and to scan the array back and forth in the horizontal direction. Since the dwell time on each detector element is relatively long in this approach, narrow bandwidths and high sensitivity (low *NET*) can be achieved.

Two basic schemes are used to display the scene. In one, a CRT tube is the display element. The signals from N_d detectors are sampled in sequence at a high enough rate so that the entire linear array is sampled before it has moved one array width. The CRT beam is deflected vertically in synchronism with the sampling multiplexer and horizontally in synchronism with the scanning motion of the detector array. The beam intensity is modulated in response to the multiplexer output levels.

The second approach is to couple each detector-amplifier into an individual light emitter such as a solid state diode emitter. The emitters are arranged in a linear array in proportion to the detector array, and the emitter array is scanned back and forth in synchronism with the IR scan. At the present state of the art, the solid state diode emitters are relatively inefficient in the visible spectrum and cannot directly form a bright display. However, the image can be intensified by conducting it to the face of an image intensifier tube or a sensitive vidicon.

Economy in detector and amplifier channels can be traded off against increased scan complexity and some loss in sensitivity by using fewer detectors and increasing the scan rate. The detectors in the linear array can be arranged with spaces between them. The spacing corresponds to some integral number of detector widths. The scan motion then has to include a vertical step after each horizontal scan until the gaps are filled in, and the scan motion must be speeded up correspondingly to cover the frame in the same amount of time.

4-1.1.2 Image Tubes

Photoemissive image forming systems use an evaporated semi-transparent photoemissive surface as the infrared sensitive element. This surface converts the infrared image into an electron emission pattern. The electrons are accelerated and focused on a phosphor layer. The phosphor screen then converts the electron image into a visible pattern.

The image tube that performs this conversion is used with objective optics to image the scene onto the photoemissive cathode surface and an eyepiece to observe the phosphor image. Solid state power supplies for converting a battery voltage to the high-voltage accelerating potential have been developed in extremely small and efficient configurations. The entire assembly of optics, image tube, battery, and power supply can be packaged into a small instrument not much larger than a simple monocular or small telescope.

Photoemissive surfaces have been used in devices designed to receive ultraviolet, visible, and near-infrared images. The devices are elegantly simple and of relatively low cost. They are used as weapon sights, map readers, night driving aids, tank periscope viewers, etc.

Their major drawback is caused by the photoemission process which at present requires relatively high-energy photons. Their limit of sensitivity in the infrared is only slightly beyond one micron wavelength. Therefore, a photoemissive system, in practically all cases, must be used in conjunction with a source of illumination. To obtain security from visual detection, a sharp cutoff filter is used to eliminate any sensible visible radiation below 0.85μ from the illuminator. The metascope is a small lightweight device developed for reading maps and signs at night and for detecting enemy use of near IR sources. (The World War II sniperscope was a combination image tube viewing device and near infrared illuminator mounted on a rifle. It was a somewhat heavy and bulky configuration which required a back-pack for the power supplies.)

4-1.2 SEARCH EQUIPMENT

Infrared search equipments (Fig. 4-3) are mechanical scan systems which scan a total field of view Ω with instantaneous field of view ω in a frame time t_f. Generally the objective is to detect the location of a target which constitutes a hot point, such as the tailpipe of a jet aircraft, against a lower-temperature background clutter such as sky-and-cloud thermal gradients or terrain clutter.

The location of the detection event is displayed on a visual screen such as the fact of a CRT or an array of light sources. Thus, the operation of a search system is completely analogous to that of a scanning IR imaging system. A major difference is in the spectral region of response. In this case, the structure of the background is of no particular interest and represents noise. Therefore, by appropriate spectral, spatial, and temporal filtering the contrast between the point target and the background is enhanced as much as possible, and the remaining clutter signal is blocked by thresholding. Usually the operator is provided with a control to adjust the threshold level to prevailing background clutter conditions. It is evident that the range sensitivity of the search unit will be directly affected by the clutter level.

The irradiance (or flux density) on a scanner from a target at a range R is

$$H = \left(\frac{J}{R^2}\right)\tau_A \text{ , w cm}^{-2} \qquad (4\text{-}23)$$

where

J = radiant intensity, w sr^{-1}

Scanner system noise limited detection range is then*

$$R = \sqrt{\frac{D_o D^* e_o J \tau_A}{F\sqrt{\frac{\Omega}{N_d t_f}\left(\frac{S}{N}\right)}}} \text{ , cm} \qquad (4\text{-}24)$$

where the signal-to-noise ratio threshold (S/N) is determined by the specified false-alarm rate and probability of detection. This relationship applies only if clutter noise is less than system noise, and when the same operational and instrumentation parameters apply as those for imaging systems as shown in Eq. 4-19.

However, since some background clutter exists under all real situations, clutter is the ultimate limit to sensitivity and range. With the present state-of-the-art in search systems, the major portion of design and development efforts are being applied to clutter rejection. Some of the methods of approach will be mentioned in order to illustrate the key system parameters that are operative.

* This equation is derived from Eqs. 4-19 and 4-23, with the approximation that a numerical proportionality factor $\sqrt{\frac{\pi}{2\sqrt{3}}}$ equals one.

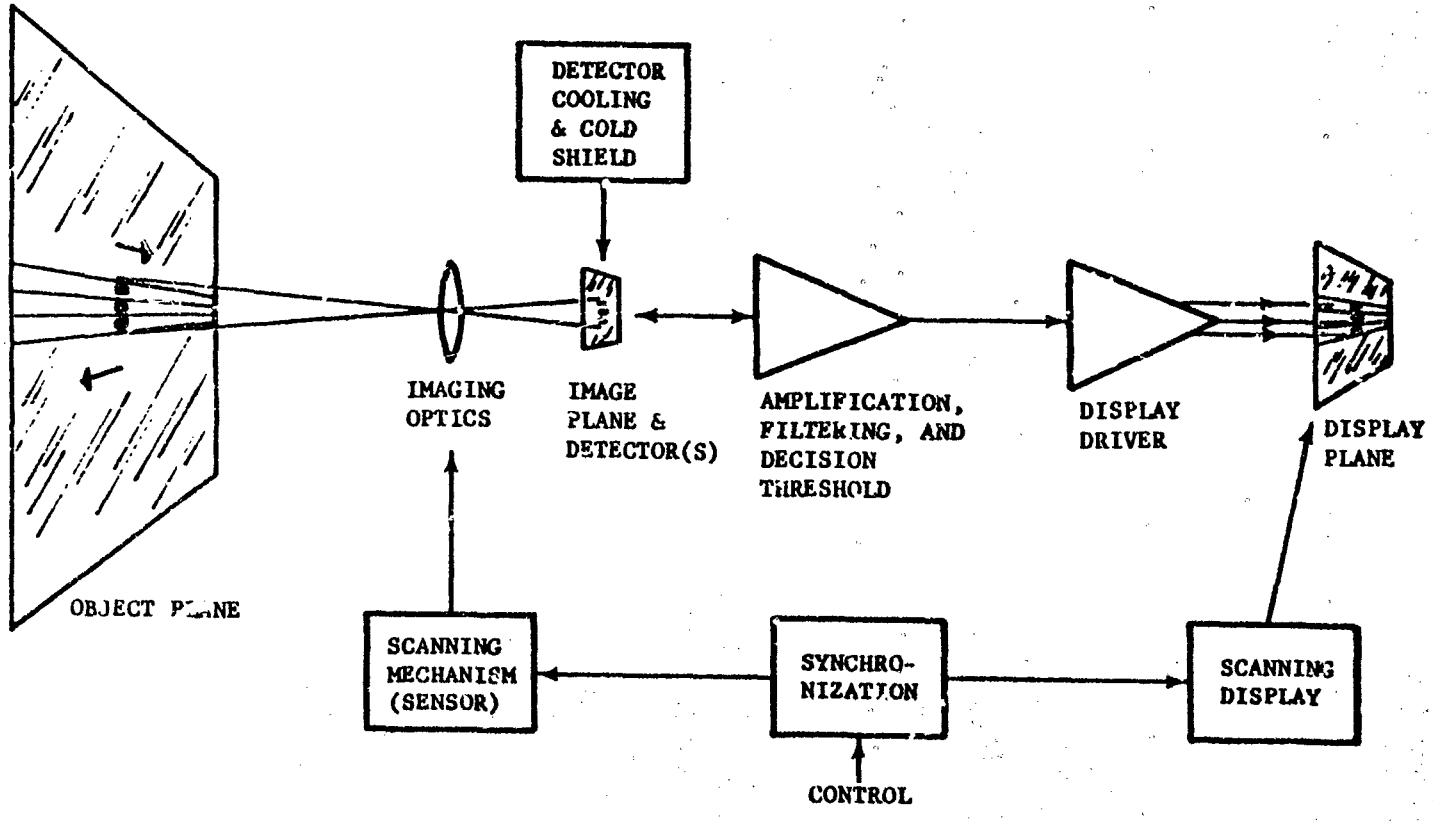


FIGURE 4-3. Typical Search System Block Diagram

4-1.2.1 Spectral Filtering

This is the first and simplest technique to implement. The background radiant intensity will typically peak around 10 microns owing to thermal self-emission, and at 0.5 micron owing to reflected solar spectrum. Around 3 microns, a region of minimum radiance exists.

The hot target will have a peak spectral radiant intensity somewhere between the two background peaks. The simplest approach is to use a spectral bandpass filter that maximizes the target-to-background signal while maintaining a usable signal level. A more sophisticated approach, called color discrimination, senses the energy in two adjacent spectral bandwidths. The ratio of the energy in the two will be different for different temperature objects. Acceptance of a signal is then based on a measurement of this ratio.

4-1.2.2 Spatial Filtering

As the instantaneous field of view ω scans across a background, a quasi-random signal is generated. The power spectrum of this signal will depend on the shape and size of ω and on the spatial distribution of the scene radiance. This analytical problem can be treated formally with the aid of two-dimensional Fourier transform principles, in a manner quite analogous to electrical filter theory. The two-dimensional spatial structure of the background is regarded as being composed of a spectrum of spatial frequencies or wavenumbers. A more detailed treatment of this approach can be found in Refs. 3 and 6, and a summary is given in Chapter 3 of this handbook.

Briefly, a point target has a spatial frequency spectrum out to high wavenumbers, analogous to a narrow-pulse electrical signal. Background clutter, however, appears to have a decreasing power spectrum with wavenumber, somewhat analogously with $1/f$ electrical noise. Therefore, it is desirable to use an instantaneous field with high wavenumber response, which implies a narrow spatial or angular size. Conceptually this may be better described by saying that, as ω is made smaller to approach the angular size of the target, more clutter is rejected without sacrificing target signal.

However, if ω is decreased while Ω and t_f are held constant, increased mechanical scanning speed or increased numbers of detectors or both are required. In systems where economy or other considerations make a small ω impractical, reticles can be used to some advantage. The reticle is fixed in the focal plane in front of the detector, and a striped or checkered transmission pattern is impressed on it. As the scene is scanned, the reticle causes a modulation on the radiant signals from the scene. The scale size of the reticle pattern is designed to just produce 100 percent modulation of the optical image of a point object. The larger background features then receive relatively less modulation as a result of the transfer function of the reticle. The electrical filter is then tuned to the modulation frequency.

The reticle is, in effect, a spatial bandpass filter analogous to an electrical tuned filter but with a large response around zero spatial frequency. The specific design details of reticles are often based on intuition and insight rather than rigorous mathematical analysis, owing both to the complexity of the analysis and the difficulty of obtaining accurate and complete background statistical data.

4-1.2.3 Temporal Filtering

The spatial frequencies produce a time signal having a frequency content that is equal to the spatial frequency times the angular scan rate. (Dimensionally, it is cycles per angle times angles per second.) Thus electrical filtering, such as low frequency rejection and high-frequency peaking, can be used to reject some of the background clutter signals. Trade-offs have to be made between the ability to resolve closely adjacent signal peaks in the presence of overshoot and signal amplitude peaking. Additionally, if the target and background signals can be adequately described, the combined spatial-electrical filter can be optimized in a formal manner to maximize the ratio of target signal to background signal.

Again, as in multi-element imaging systems, multi-element search systems that use large numbers of detectors and corresponding amplifiers will usually implement relatively simple and economical filtering techniques. The instantaneous field is made as small as practical. The upper frequency cutoff in the electrical filter is made to correspond to the dwell time of a point on the detector and the low frequency cutoff

is placed at 1/3 or 1/4 of the upper cutoff frequency to block low frequency background noise without imposing severe overshoot of the target signal.

Search equipments are used in ground-based and airborne applications and will probably be used in spaceborne vehicles as well as for rendezvous and docking purposes. Often, they are coupled with tracking systems, and automatic acquisition and track modes are implemented to maintain track on a target after it is detected by a search system. A single scanner may be designed to perform the combined functions of search, acquisition, and track.

A somewhat unique background problem will exist for spaceborne search equipment in that the sky background clutter will consist of discrete point objects similar in angular size and spectral radiant intensity to the target being sought. The discrete point objects are, of course, planets and stars, whose spectra are similar to the thermal self-emission and reflected sunlight from a satellite target. Other differential signatures, such as angular rate and intensity rate, can be implemented in this case.

4-1.3 TRACKERS

In tracking applications the infrared scanner serves basically as an optical transducer whose output is a voltage corresponding to the position of the target image with respect to the center of the field of view of the scanner. If the target angular position with respect to an arbitrary reference is θ_T, and the angular position of the optical axis (center of the field of view) is θ_S, then the error angle θ_E is

$$\theta_E = \theta_T - \theta_S \tag{4-25}$$

Usually the error angle is required along two orthogonal axes such as azimuth and elevation. Fig. 4-4 shows a simplified block diagram of one axis of a tracking loop.

The scanner is mounted on gimbals, and the voltage output corresponding to θ_E is fed to a torquer which moves the scanner in the proper direction to reduce or eliminate θ_E. If the target should move and change its angular position θ_T, a new error signal is generated and the tracker will move to cancel the error—in other words the tracking loop "tracks" the target.

The accuracy and rapidity with which a target can be tracked will depend on the quality of the scanner as an optical position transducer, servo loop transfer function and stability constraints, signal-to-noise ratio, and local conditions such as platform stability.

For guidance of homing missiles the tracking loop provides attitude control and steering information. The most common guidance law utilized in air-to-air missiles is known as proportional navigation. If two vehicles move at constant velocities and are on a collision course, the line of sight angle between them is constant. In proportional navigation the tracking loop is used for the purpose of measuring the angular rate of the line of sight between the target and the missile. The missile is steered in response to a measured line of sight rate to bring the measured rate to zero. If the line of sight angular rate in inertial coordinates is $\dot{\theta}_{TM}$ and the direction of missile velocity is $\dot{\theta}_M$, then the missile turn rate is made proportional to the target line of sight of rate

$$\dot{\theta}_M = k\dot{\theta}_{TM} \tag{4-26}$$

This guidance law implements corrections early in the trajectory and causes the missile to steer to a collision course. It requires considerably less lateral accelerations in the terminal phase of flight than the pursuit type of guidance laws. (Ref. 7 is recommended as an introduction to the kinematics of the more common guidance schemes such as proportional navigation, beam riding, pure pursuit, deviated pursuit, etc.). To obtain the vertical angular rate measurement a rate gyro, or an integrating gyro captured in a rate loop, is attached to each axis of the tracker gimbal. This gyro is also utilized to rate stabilize the tracker loop to inertial coordinates. The tracking loop then may serve as an inertial reference for vehicle attitude control.

In other tracker applications—such as weapon orientation, optical instrument pointing, star tracking, etc.—the information of interest is accurate target angular position instead of angular rate. In these applications, rate gyros or tachometers may still be utilized for the purpose of tracker servo loop stabilization. However an additional element, an accurate gimbal position pickoff, is required. Some applications may

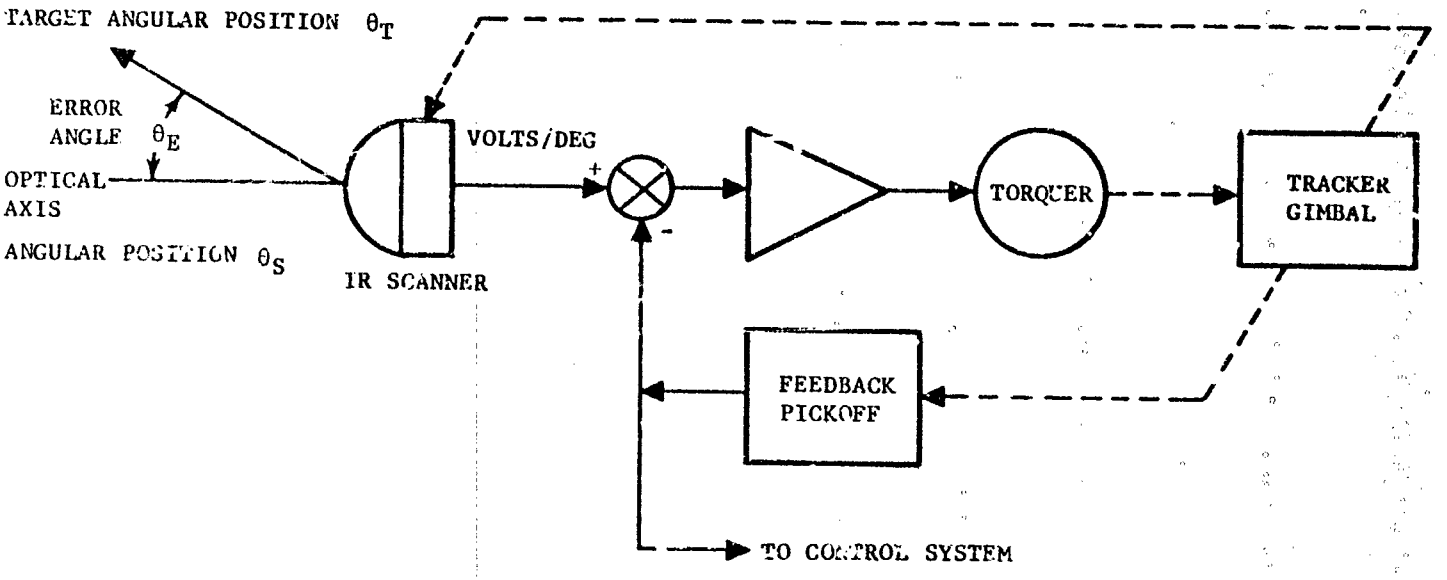


FIGURE 4-4. Typical Tracker Block Diagram

demand angular tolerances of the order of a second of arc, which of course implies corresponding accuracy in the scanner.

The discussion of trackers will be divided into two major areas. The first area is the consideration of the scanner as an optical position transducer, and to show that the same sensitivity parameters apply as for any scanner. The second area is to consider the scanner parameters in the way they affect the performance of the tracking loop, and to identify the parametric trade-offs for tracker optimization. The art of servo design is naturally of paramount importance to a tracking system. Some of the basic servo approaches will be introduced, however an extensive treatment of this technology is beyond the scope of this handbook.

4-1.3.1 The IR Scanner As an Optical Angle Transducer

The target must first be localized by some means, i.e. visually, or by some other sensor such as radar or infrared search set. It will be shown that it is desirable to keep the field of view of the tracking scanner small. Therefore, a common system design problem encountered is that of target acquisition by the tracker—i.e., to rapidly orient the tracker into a position where the target appears in its field of view.

The tracking scanner scans some field of view Ω, which is typically of the order of a few square degrees or a fraction of a square degree, with some smaller view of view ω. The scanning is performed in such a way that the output contains the position information of the target within Ω. The rate at which Ω is scanned, the frame rate, is governed by the information rate demanded by the tracking servo. Typically, to prevent additional phase lags from being introduced, the frame rate should be a factor of 3 to 10 times the frequency response of the closed loop.

The opportunity for invention and innovation has been thoroughly exploited in IR trackers, and many scanning schemes exist in practice and theory. A wide range of trade-offs exists between mechanical-optical complexity and elegance, and signal processing and spatial filtering optimizations. For purposes of description, they may be categorized into two functional types—small instantaneous field scanning and wide instantaneous field reticle scanning. Both types perform the same function, but with different levels of mechanical and signal processing efficiency.

Because of the small frame time requirement, scan modes that allow high mechanical speeds are a prerequisite. Generally, the approach is to spin some optical element or elements to generate a motion of the IR image in the focal plane, or to move a reticle with respect to a stationary image. Only a few common examples will be given here. Some additional examples will be found in the classified literature and Refs. 6 and 8.

4-1.3.1.1 Small Instantaneous Field-of-view Scanning

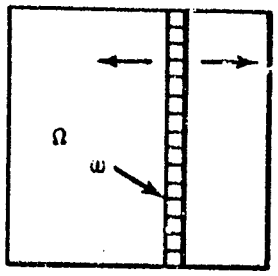
This approach consists generally of causing the small instantaneous field of view ω to scan and sample Ω in some orderly manner such as a spiral pattern or a Lissajous pattern. Motion in two coordinates, of course, implies the existence of two orthogonal scan drive signals. The driving voltages then have some known or predictable relationship to the actual angular position of the scanning element ω.

The signal pulse that results from the presence of a target on ω is used to gate or sample the amplitude of the scan drive voltage. The level of the sampled voltage then discloses the position of the target. Usually it is transformed by peak holding circuitry into a dc level that is proportional to target displacement from the optical axis along each separate gimbal axis.

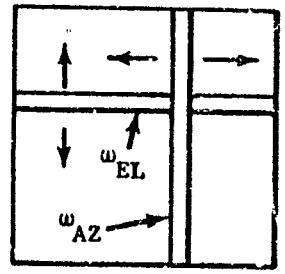
A cruciform tracker is an example of this method. As shown in Fig. 4-5(A) and (B), a pair of long thin detectors (or focal plane apertures) are crossed, and each of the two members of the cross has its own signal processing. The focal plane image is caused to "nutate", or to describe a circular motion with respect to the detector cruciform. This motion is imposed by a spinning tilted secondary mirror or a spinning optical wedge. Functionally, this is the same as if the cruciform were nutating with respect to the focal plane image.

If the detectors are longer than the focal plane field stop dimensions of Ω, then only motion perpendicular to the detector long dimension is significant. This component of motion for each detector is described as

SMALL INSTANTANEOUS FIELD OF VIEW SCAN

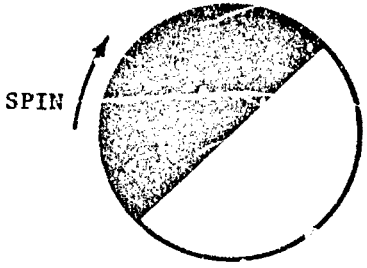


(A) LINEAR ARRAY

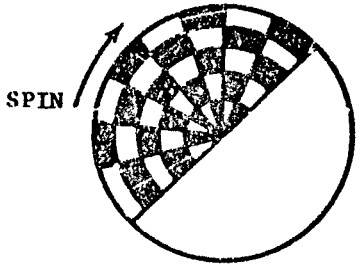


(B) CRUCIFORM

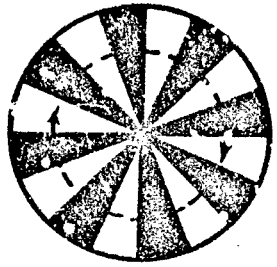
LARGE INSTANTANEOUS FIELD OF VIEW RETICLE SCAN



(C) HALF MOON



(D) HALF MOON WITH CARRIER



(E) FM MODULATION



(F) DISH WITH FM ELEVATION AND AM AZIMUTH CODING

FIGURE 4-5. Reticles

$$\left.\begin{aligned}\theta_{AZ} &= \theta_o \sin 2\pi f_d t \\ \theta_{EL} &= \theta_o \cos 2\pi f_d t\end{aligned}\right\} \quad (4\text{-}27)$$

where

θ_{AZ} = azimuth angular position of detector with respect to center of Ω, rad

θ_{EL} = elevation angular position of detector with respect to center of Ω, rad

θ_o = constant reference angle, rad

f_d = scanner spin rate, rps

t = dwell time, sec

Simultaneously a pair of reference voltages is extracted from the scan drive which is spinning at f_d rps. After proper phasing they are expressed as

$$\left.\begin{aligned}V_{AZ} &= V_o \sin 2\pi f_d t \\ V_{EL} &= V_o \cos 2\pi f_d t\end{aligned}\right\} \quad (4\text{-}28)$$

where V_o is a constant reference voltage, and V_{AZ} and V_{EL} are the instantaneous voltages of the scan drive in azimuth and elevation, respectively.

From Eqs. 4-27 and 4-28, it is readily seen that

$$V_{AZ} = \theta_{AZ}\left(\frac{V_o}{\theta_o}\right) \text{and } V_{EL} = \theta_{EL}\left(\frac{V_o}{\theta_o}\right) \quad (4\text{-}29)$$

Thus the extracted drive voltages V_{AZ}, V_{EL} are linearly proportional to the instantaneous positions of the detector θ_{AZ}, θ_{EL}, respectively. Then if a target crosses the detectors and causes the drive voltages to be sampled, the output is linearly proportional to the AZ—EL-coordinates of the target.

The foregoing has been just one sample of the manner in which Ω is scanned by one or more smaller fields ω to find the position of the target and to produce an output signal proportional to, or at least some usable function of, the target position.

The considerations relating to background clutter suppression by various filtering techniques—as discussed in par. 4-1.2 for search equipment—apply equally well for tracking scanners, and the same operational and instrumentation parameters apply since the problem is still basically one of scanning Ω with ω within some period t_f in order to locate a target.

Situations may arise where the target is more than a point image. It may be a line, or edge, or a resolved two-dimensional image. In these cases, some special processing is called for in order to define and track the appropriate parts of these extended objects. The appropriate part may be an intensity centroid, a preferred orientation edge, the end of a line, etc. The associated signal processing will require various levels of complexity depending on the specific problem. For example, simple differentiation of the detector signal can mark leading and trailing edges of extended targets. On the other hand, elaborate digital logic may be required in difficult situations.

4-1.3.1.2 Large Instantaneous Field-of-view Reticle Scanning

In this approach the detector fills or nearly fills the entire field of view Ω. A reticle in front of the detector scans or modulates the field in order to locate a point target. The reticle modulation may be AM, FM, random, pulse, or some combination of these.

The simplest example of a tracking reticle is perhaps the rudimentary "half-moon" configuration. As it spins in front of the detector, the time that its edge crosses the target image is a measure of the target's position in one angular coordinate. (See Fig. 4-5(C).) This reticle has a number of serious drawbacks: it modulates large amounts of background clutter; it does not reveal the distance of the target from the center; and it produces no signal if the target is on center.

The entire circular field is Ω, and the transmitting portion of the reticle is ω in this case, or $\omega = \Omega/2$. One-half of the detector is blocked at all times. Therefore, the detector area is larger than ω, which has the effect of generating additional detector noise.

The reticle is improved by imposing a modulated transmission pattern in place of the opaque section. (See Fig. 4-5(D).) This decreases the response to the large low spatial frequencies in the background clutter, and the modulation pattern can be arranged to produce a code on the modulation of a point image such that the value of the code corresponds to the distance of the point image from the center. This scheme is

simple and economical to instrument since it requires only one detector, a spinning reticle, and relatively simple signal processing. It is adequate for economical tracking systems that are required to operate against targets with good target-to-background-clutter contrast.

A more sophisticated approach is to hold the reticle stationary with respect to the detector and to nutate the focal plane image. One immediate benefit from this is the removal of spurious noises that may be introduced from a moving object such as a reticle in the immediate vicinity of a detector. This problem can be particularly serious in long-wavelength IR sensitive systems, since any moving item can introduce spurious reflections, emissions, and uniform background modulation signals into the detector.

With an FM reticle the target image is nutated around a constant diameter circular path. If the circle is concentric with the stationary reticle pattern, a steady carrier frequency is generated. This is the on-center situation. (See Fig. 4-5(E).) If the target moves off-center, the carrier is frequency-modulated at the nutation frequency. The modulation amplitude is proportional to displacement, and the phase of the modulation is proportional to the direction of the displacement. Thus the modulation on the carrier represents the polar coordinates (ρ, θ coordinates) of the target position.

The use of larger disks on which modulation codes are impressed offers another set of possible approaches. One example is shown in Fig. 4-5(F) where the frequency of the carrier depends on the elevation position of the point image, and the phase of an amplitude modulation on the carrier depends on the azimuth position of the target.

4.3.2 Tracking Scanner Sensitivity

The scanner sensitivity is expressed, as before in pars. 4-1.1 and 4-1.2, in terms of noise equivalent flux density (*NEFD*)

$$NEFD = \frac{4\sqrt{A_d\,\Delta f}}{\pi D_o^2\, D^*\, \epsilon_o}$$

It will be shown that this reduces to the form

$$NEFD = \frac{K_1\sqrt{\Omega/t_f}}{\frac{\pi}{4}D_o\, D^*\epsilon_o} \qquad (4\text{-}30)$$

where K_1 is a proportionality constant. This equation is used for all of the scanning modes discussed for tracker operation. If the tracking scanner utilizes a small instantaneous field of view ω to scan Ω in a period t_f, then

$$A_d = \omega(fl)^2 \qquad (4\text{-}31)$$

where fl is the effective focal length of the optical system

$$\Delta f = \frac{\Omega}{\omega t_f} \qquad (4\text{-}32)$$

and $$\sqrt{A_d \Delta f} = (fl)\sqrt{\Omega/t_f} \qquad (4\text{-}33)$$

Substituting this into Eq. 4-3 results in Eq. 4-30.

If the scanner utilizes a detector large enough to cover Ω, and a spinning reticle with a carrier frequency N cycles/revolution and amplitude modulation on the carrier, then

$$A_d = \Omega(fl)^2 \qquad (4\text{-}34)$$

$$\begin{aligned}\Delta_f &= \left[(N+1) - (N-1)\right] \times (\text{rps}) \\ &= 2 \times (\text{number of revolutions per second}) \\ &= \frac{2}{t_f} \end{aligned} \qquad (4\text{-}35)$$

where rps = number of revolutions per second is the rotation rate of the reticle and t_f is the rotation period of the reticle.

Then

$$\sqrt{A_d\,\Delta f} = (fl)\sqrt{2}\sqrt{\Omega/t_f} \qquad (4\text{-}36)$$

If the reticle is FM-modulated instead of AM-modulated, the required bandwidth will be proportional to the deviation of the carrier frequency times the revolution rate of the reticle[9], which is again similar to Eq. 4-32.

The important relationship to be noted is that to a first-order calculation the scanner sensitivity depends on the total field Ω that is scanned, and

the time within which it is covered t_f; and is independent of the instantaneous field ω or the number of cycles M of the reticle.

However, *NEFD* refers to the signal-to-noise ratio in the bandwidth that immediately follows the detector. This bandwidth is typically much wider than the servo bandwidth which follows the electronics. For a tracker, it is the signal-to-noise ratio in the servo bandwidth that is of significance. After the detector signal is demodulated and narrow-banded around zero frequency for the servo, the *S/N* ratio increases by a factor proportional to the square root of the ratio of detection bandwidth to servo bandwidth

$$(S/N)_{servo} = K_2 \left(\frac{\Delta f_{detector}}{\Delta f_{servo}}\right)^{1/2} \left[(S/N)_{detector}\right] \tag{4-37}$$

where K_2 is a proportionality constant. The proportionality constant depends on the type of demodulation, i.e., square law, synchronous, etc. For low *S/N* ratios the proportionality constant is nonlinear. (Ref. 5 is a more detailed introductory text on modulation theory.) However for high-precision trackers, or for homing trackers in terminal flight, relatively high signal-to-noise ratios will obtain and the proportionality is valid. This relationship along with Eq. 4-30 leads to the interesting result that for maximum *S/N* ratio in the servo bandwidth, the tracker scanning mode utilized should be one that codes the target signal into the widest possible bandwidth. For example if the small instantaneous field-of-view scanner described in par. 4-1.2.1.1 is utilized, then Eq. 4-37 becomes, using Eq. 4-32 recognizing the relationship that $S/N \propto (\Omega/t_f)^{-1/2}$,

$$(S/N)_{servo} = K_3 \left(\frac{\Omega}{\omega t_f \Delta f_{servo}}\right)^{1/2} \left(\frac{\Omega}{t_f}\right)^{-1/2} = K_3 \left(\frac{1}{\omega \Delta f_{servo}}\right)^{1/2} \tag{4-38}$$

where K_3 is a constant coefficient, since for a given target the detector *S/N* ratio is inversely proportional to the *NEFD* of the scanner. Thus making ω as small as possible provides the maximum *S/N* ratio for tracking, in addition to minimizing background clutter.

This approach is valid for situations where the detector *S/N* ratio is expected to be relatively high during operation, and where maximum static tracking accuracy is sought. It has the disadvantage that at relatively low detector *S/N* ratios, say in the vicinity of 3 to 1, the servo *S/N* ratio starts to deteriorate to the point of losing track, owing to the inefficiency of the demodulation process. In the example cited the broadband detector signal would consist of narrow pulses occurring once per frame time, but with a pulse spacing that depends on target or tracker motion between frames. The target position information is contained in the pulse position on the time axis, and the demodulation process consists of converting pulse position to analog voltage level. If the frequency of occurrence of random noise pulses becomes comparable to that of the target signal pulses, the resultant analog voltage becomes erratic and causes tracker jitter. The reticle modulation schemes generate continuous signals which are not as catastrophically affected after low *S/N* ratio demodulation. Some rudimentary reticles such as the *half moon reticle* described in par. 4-1.3.1.2 do not generate a carrier signal, and therefore the demodulation consists simply of separating the signal into orthogonal (elevation-azimuth) components without a change in frequency. With reticle scanners it is usually possible to maintain track on a stationary target with *S/N ratios of one.*

4-1.3.3 Tracking Loop Servo Parameters

Trackers are usually designed with Type I or Type II servo loops. Briefly, a Type I servo is defined by an open loop response with a single pole at the origin and a Type II with a double pole at the origin (Refer to par. 3-9). Since a physical system cannot have infinite gain at zero frequency, real trackers are close approximations to these types.

Functionally, the closed loop response of a Type I servo is such that the rate of the output angular displacement $\dot{\theta}_S$ is proportional to the input angular displacement error θ_E. Conversely, in order to generate a constant output rate, a constant error θ_E must exist. Thus a Type I servo tracks with an error that is proportional to the steady rate at which the target is moving. The velocity constant K_v is the steady state ratio of $\dot{\theta}_S/\theta_E$.

If the scanner generates a linear error function over some range $\pm\theta_{E(max)}$—i.e., if the voltage

output is linearly proportional to the displacement of the target image, over this angle—then the maximum steady state tracking rate available from the servo is

$$\dot{\theta}_{S(max)} = K_v \theta_{E\,(max)} \qquad (4\text{-}39)$$

The maximum K_v that can be attained in a servo is limited by gimbal-tracker inertia and the frequency responses of actuators. Values of the order of 100 (deg/sec)/deg are typical limits for infrared trackers. This factor brings up one of the major trade-offs in IR tracker design. On the one hand, it is desirable to keep the total field of view Ω of the scanner as small as possible in order to decrease background clutter or susceptibility to decoys. However the maximum linear θ_E that can be generated is one half of the width of the total field. Therefore, a small field implies a low maximum rate tracking capability in accordance with Eq. 4-39.

A Type I servo will produce a steadily increasing error for a steadily accelerating target until the error angle exceeds $\theta_{E(max)}$ and the target is lost. The response of the tracker to noise input is a random jitter. Its effect is to cause loss of track before the steady state error reaches θ_E.

The closed loop response of a Type II servo is such that the output acceleration $\ddot{\theta}_S$ is proportional to the displacement error θ_E. Thus a steadily accelerating target is tracked with constant error, and a steadily moving target is tracked with zero error.

In practice a typical tracker is made to approximate a Type II system by means of a relatively long-term integrator in the forward loop plus the use of a tachometer or rate gyro in the feedback loop. The specific design must take into account not only the basic stability and frequency response criteria, but must include detailed consideration of the transient and steady state responses to target motions, and to physical and electrical noise inputs or disturbances.

In practice, a high signal-to-noise ratio from the infrared scanner is often masked or degraded by other noise inputs, such as aerodynamic buffeting and physical vibrations in an airborne guided missile. Therefore, in this type of tracker application, designing for maximum servo bandwidth signal-to-noise ratio from the scanner may not be as important as some other considerations such as ease of target acquisition, mechanical ruggedness and simplicity, cost, etc. On the other hand, for applications where great accuracy is to be obtained in an undisturbed environment, as for example a ground-based satellite tracker or satellite based star tracker, the maximum S/N ratio that can be obtained within the state of the art may be called for.

4-1.4 RADIOMETERS

A radiometer is an instrument for measuring thermal radiation, especially infrared. Historically, the name has been applied to a number of devices. The first was Crookes' Radiometer which works on the principle of thermal transpiration of gas molecules from the polished side to the blackened side of vanes to cause rotation of a lightweight paddlewheel. C. V. Boys' Radiomicrometer used a sensitive galvanometer to measure a Seebeck current. It was reported to have sufficient sensitivity to detect a candle at 2 miles[10].

In the modern sense of the term, a radiometer is an optical instrument which collects radiation from a narrow field of view and converts it into a calibrated electrical signal. Typical commercial radiometers are the Barnes Engineering Company's Optitherm radiometer and the Servo Corporation of America's Servotherm radiometer.

Modern designs rely on the use of an accurate blackbody radiation reference built into the instrument. (See Fig. 4-6.) In operation, the collecting optics focus an image of the object whose temperature is to be measured onto a detector. Within the instrument the reference blackbody source is also focused on the detector, and a reflecting chopping wheel alternately allows the external and internal radiant signal to fall on the detector. Thus the detector sees an alternating radiant signal which corresponds to the radiant difference between the target source and the reference source. If high accuracy, independent of amplifier gain, is desired, a nulling technique is used where the temperature of the reference blackbody is adjusted until the difference signal becomes zero.

A more economical but less accurate approach does not use a blackbody reference source. Instead, the incoming radiation is

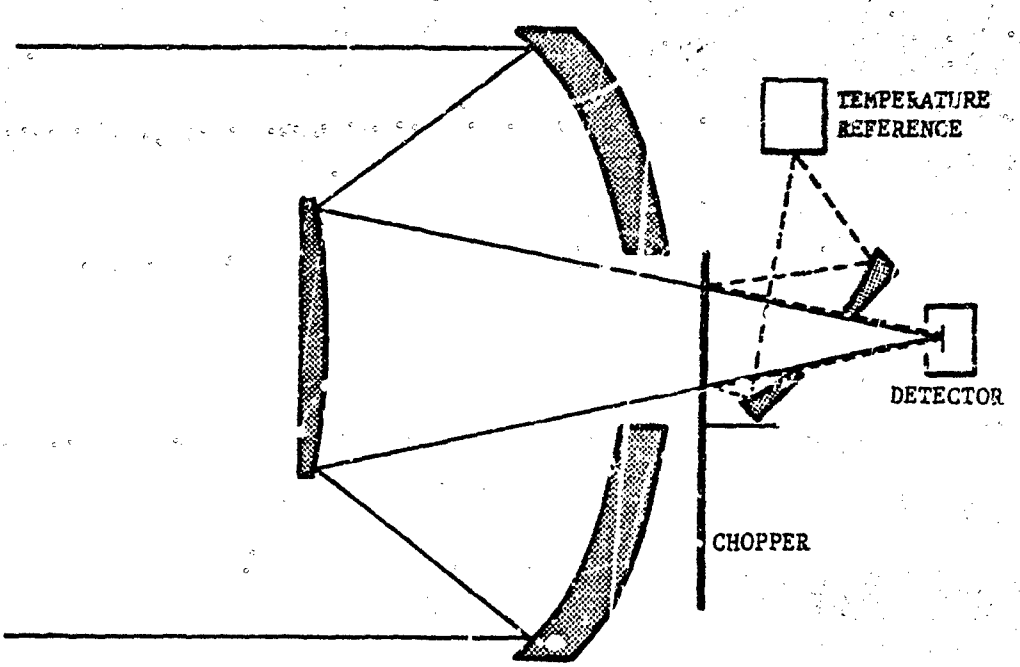


FIGURE 4-6. LWIR Radiometer

chopped by a blackened or highly emissive chopping wheel. In this case the blades of the chopping wheel act as the reference temperature source. Or the blades may be made highly reflective, in which case the detector itself (or more accurately the image of the detector in the polished blade) becomes the reference temperature source.

Commercial radiometer instruments are generally designed for uniform response over a broad band of infrared wavelengths in order to optimize the accuracy of remote temperature measurements. In military applications, however, special designs are often called for to measure and record radiation from specific sources, often under special operational conditions. In these designs, a broad spectral response is usually not as important as high sensitivity and speed of response, within allocated bounds on instrument size and cost. Here again, background clutter problems have to be faced when the radiant sources are remote objects that cannot fill the field of view of the instrument.

4-1.5 SPECTROMETERS

A spectrometer is an instrument for measuring the relative amounts of radiant energy as a function of wavelength. Historically, the term referred to a spectroscope that was provided with a graduated circle for measuring angular deviations of the dispersed spectrum. A spectroscope is a viewing instrument where a radiation source is collimated. The collimated beam is passed through dispersive elements such as gratings or prisms, and the dispersed beam is viewed through a telescope.

Today the term spectrometer is applied to practically any instrument that measures the spectral components of a radiation source. The instrument may use a dispersive element plus a scan drive to sample the various portions of the spectrum, or it may use a number of detectors with different transmission filters to sample a number of wavelengths simultaneously, or it may use a rotating color wheel which holds a set of discrete filters. The OCLI Corporation has recently developed a rotation tunable filter, which is a narrow-band interference filter whose center wavelength varies linearly with angular position around a wheel.

A common military application is to measure the spectral content of radiation sources of interest, such as airborne or land-based vehicles operating under typical conditions. In these

situations, the target radiation sources often have to be viewed at long ranges against background clutter. Background suppression techniques, such as the use of reticles and narrow fields-of-view (as described in pars. 4-1.2 and 4-1.3), have to be employed in these situations. Boresighted infrared trackers may be employed to hold the spectrometer on target, in which case the system is referred to as a "tracking spectrometer".

4-1.6 INTERFEROMETERS

Wave interference phenomena can be utilized to obtain high spectral resolution in a spectrometer or to record spectral intensity details simultaneously over a wide band. For example, the Block Associates' interferometer spectrometer is an adaptation of the Michelson interferometer[11] principle. One of the mirrors in the interferometer is moved at a linear rate, which causes an alternate brightening and darkening of the central fringe pattern. The frequency of the cyclic changes is inversely proportional to the wavelength of the incoming radiation. An infrared detector placed at the central fringe converts the cyclic intensity changes into an ac signal. If more than one wavelength is present in the radiation, the detector output consists of a superposition of the corresponding ac signals. The output is tape recorded and played back through a wave analyzer to recover the infrared spectrum.

Another approach to interferometry for spectrometry application is Aerojet's "color coding reticle" approach. A pencil of rays from the collector optics is passed through a disk having good transmission in the wavelength region of interest. If the optical thickness of the disk is equal to odd multiples of a quarter wavelength, transmission through the disk is a minimum owing to interference. Transmission is a maximum for an optical thickness equal to multiples of a half wavelength. The color coding reticle has a layer of germanium whose thickness varies linearly with angle about the center of the disk. Therefore, if the reticle is rotated, a transmitted pencil of monochromatic rays is modulated at a rate proportional to the spin frequency and inversely proportional to wavelength. In effect it acts like a spoke reticle. For mixed light, the modulations are superimposed and can be electronically separated to reconstruct the spectrum.

Perhaps the most sharply tunable interferometer is the Fabry-Perot[11]. Wavelengths can be measured to thousandths, or even ten thousandths of an Angstrom. The instrument is used almost exclusively for ultra-high resolution and requires a high-dispersion, high-resolution spectrograph to avoid overlapping orders.

4-1.7 HYBRID SYSTEMS

4-1.7.1 Track-while-scan

In situations where more than one target exists in the total field of view, the results of the scan of each frame can be stored in the memory of a small special purpose computer, and prediction techniques can be used to keep track of each target. The name of this technique, borrowed from radar systems terminology, is "track-while-scan". It demands no basically new or different approaches in infrared technology, but capitalizes on the state-of-the-art in signal and data processing.

4-1.7.2 Scanning Radiometer/Spectrometers

Typical search and track equipments utilize target intensity information only to establish the presence or absence of a target; i.e., once a target appears over a decision threshold the information of primary interest is the coordinate position of the target. Other target information such as intensity, or spectral content is discarded. However, in cases where it is of interest, this information can be retained and the functions of search or track can be combined with the functions of radiometry or spectrometry. This requires simply that dual electronic signal processing be utilized.

For example, while a tracker is locked onto and tracking a target, the broadband detector signal can be channelled into circuitry for measuring absolute amplitude of the signal. This, in effect, is a radiometric function. Or, the scanner could be configured to receive radiation from a target in a number of discrete spectral bandwidths. These signals could be summed for maximum signal-to-noise ratio to perform the detection or tracking function and extracted separately to plot the spectral characteristics of the target.

4-2 ACTIVE SYSTEMS

4-2.1 ILLUMINATORS

Presently, only searchlights and laser systems operate as infrared illuminators. In this handbook, searchlights are distinguished from laser systems due to the vast differences in their design requirements and parameters, as described in the paragraphs which follow.

4-2.1.1 Searchlights

A searchlight is an optical system which consists, basically, of a radiation source, a power supply, and a reflector or refractor. Additional components may consist of beam-spreading optics and infrared security filters. The searchlight system is designed to direct a powerful artificial source of visible and near-infrared illumination at a predetermined position with a minimum of dispersion. Searchlights range in size from 60-in. diameter down to flashlights.

The radiation sources for searchlights are tungsten filament, carbon arc, and gaseous arc lamps—discussed in par. 3-3. Tungsten filament and carbon arc lamps in use are being replaced with xenon arc lamps which are simpler to operate, more efficient, and provide a light output which consists of usually more than 10 percent infrared.

Normally, reflective optics are used for searchlights due to their freedom from chromatic aberrations and decreased loss of energy. Moreover, if the reflecting surface is parabolic in form, spherical aberrations may be eliminated and a much larger collecting angle is obtained[12].

The reflector collects the radiant energy from the lamp, determines the beam pattern and, in conjunction with the lamp intensity, determines the beam intensity. The beam candlepower of a searchlight is defined as the candlepower of a bare source which, if placed at the same distance away from a given point, would produce the same illuminance at that point. The peak beam candlepower is determined from a beam intensity distribution curve. This curve represents data obtained by using a calibrated photocell to measure the intensity across the beam at a convenient distance (1,000 ft).

The infrared security filter of an illuminator obscures visible radiation to prevent visible detection. The criteria for infrared filter effectiveness are: (1) no transmission at wavelengths less than 0.8 micron and optimum transmission in the IR region, (2) minimal white-light penetration (pinholes) through the filter, (3) ability to withstand thermal shock of exposure to high temperatures and then quick cooling to reduce afterglow, and (4) ability to withstand prolonged exposure to high-intensity radiation without deterioration.

Typical searchlights presently in use include 18-in., 2500-w incandescent lamps, 23-in., 2200-w xenon arc lamps, and 30-in., 16-kw carbon arc lamps. The 23-in. xenon arc searchlight, which will replace the 18-in. searchlight, provides a narrow or wide beam of high-intensity visible or infrared radiation. The searchlight contains hardware for use in mounting on tanks and helicopters, and with a universal mount for use on trucks, towers, etc. A new 30-in. xenon arc searchlight is being developed for ground and airborne use to replace the 30-in. carbon arc searchlight[13]. Table 4-2 lists the characteristics of the 23-in xenon arc, 30-in. carbon arc, and 30-in. xenon arc searchlights for comparison.

4-2.1.2 Lasers

A laser illumination system consists of a laser source (see par. 3-3.1.3), power supply, control electronics, and collimating optics. It is not intended to replace the searchlight as a source of continuous wide area illumination. Lasers have features not provided by searchlight sources, thereby, making them ideally suited for use as illuminators in applications such as missile homing, pulse-gated viewing, and target designation (see Chapter 6). These features include operation at wavelengths ranging from the visible to the far infrared, high spectral radiance in a narrow beam, nearly-monochromatic output, and pulsed operation.

Reflective and refractive telescopes can be used to collimate the laser output into beams as narrow as 1 microrad. The optical design of laser systems is simplified because chromatic aberration is not a problem. Greater efficiency can be

TABLE 4-2. COMPARISON OF 23-IN. XENON ARC, 30-IN. CARBON ARC, AND 30-IN. XENON ARC SEARCHLIGHTS

ITEM	23-IN. XENON ARC	30-IN. CARBON ARC	30-IN. XENON ARC
Power, w	2200	16,000	20,000
Reflector Diameter, in.	23	30	30
Reflector Focal Length, in.	6.9	9.75	9.75
Reflector Collecting Angle, deg	—	150	150
Beam Candlepower	100×10^6	420×10^6	1.2×10^9
Beam Spread — Narrow, deg	0.75	3.0	2.0
— Wide, deg	7.0	10.0	—
Beam Lumens	—	310,000	500,000
Flicker	none	slight	none
Operation Time	200 hr continuous	must be shut down every 8 hr to change negative electrode	750 hr continuous
Afterglow	2 sec (with blackout shield)	3 sec (with dowser) 1 min (no dowser)	3 sec (no dowser)

achieved in the optical system by use of high-efficiency, narrow-band optical coatings. However, coatings selected must withstand the optical power densities from high-power lasers. Chapter 3 discusses some optical design problems peculiar to lasers.

Sensitive detection systems, optimized for the laser wavelength in conjunction with the high spectral radiance of the source, allow the use of low-power laser illuminators. These illuminators, which are usually pulsed, can be made into a compact, lightweight, man-portable system. The physical and performance characteristics of a typical laser illuminator are:

Material:	neodymium glass
Wavelength:	1.06 microns
Peak Power Output:	2.5 Mw
Pulse Width:	40 nanosec
Repetition Rate:	10 pulses per sec
Average Output Power:	1 w
Beam Spread:	0.057 deg
Dimensions:	13 × 4 × 4 in.

Descriptions of neodymium-doped yttrium aluminum garnet (YAG:Nd) and gallium arsenide (Ga:As) laser illuminators are given in AMCP 706-128.

4-2.2 RANGEFINDERS

Two kinds of optical rangefinders are presently in use; one is active, while the other is passive operating only in the visible region. The accuracy of a passive system, used to measure distances by triangulation, is dependent upon the length of the base. A laser rangefinder (active) measures the time required for an optical pulse to travel from the laser transmitter to a selected target and back to a receiver. A

sight is used for aiming the rangefinder at the target. The transmitter consists of a Q-switched ruby or neodymium laser and a collimating telescope. The receiver consists of a photomultiplier or solid-state photodetector, receiving telescope, and range-computing electronics. The receiver and aiming sight typically share a common objective lens. A dichroic mirror in the optical system separates the reflected laser pulse, which goes to the detector, from the visible scene light which goes to the aiming eyepiece.

The theory of operation of a typical laser rangefinder is summarized by describing in sequence the stages illustrated in Fig. 4-7. *It is important to emphasize that the effect of laser irradiance directly on the human eye may result in permanent eye damage. This damage becomes more severe as the laser power increases.* Upon energizing the fire command, the range computing unit (RCU) is reset and held at zero. At the same time, the laser trigger is set to fire the laser as soon as the pulse-forming network (PFN) is charged by the power supply. The laser output pulse passes through a collimating telescope that narrows the energy of the beam to less than 1 mrad. As it is transmitted, a photodiode detects the pulse and energizes the RCU. The receiving optics whose axis is aligned to the transmitter

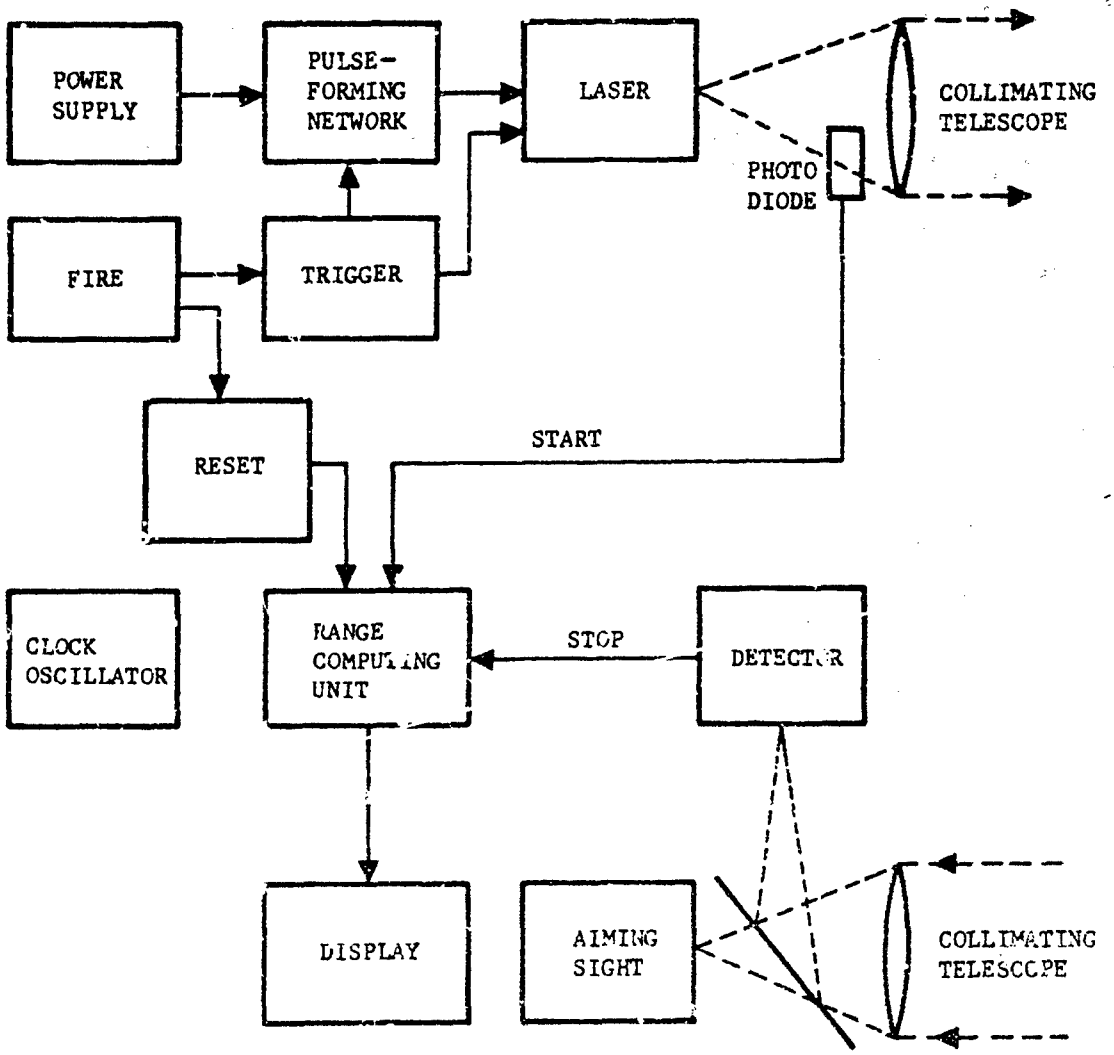


FIGURE 4-7. Typical Laser Rangefinder Block Diagram

directs the reflected energy pulse to a sensitive detector which terminates the range computation. The optical filter in front of the detector can be made very narrow-band in order to reduce background and photon noise.

The maximum range detection is limited to a distance where the return pulse signal amplitude is sufficiently large to allow its unambiguous detection in the presence of noise. It is convenient to express both noise and signal in terms of the average number of photoelectrons $\bar{\eta}_s$ per sample period δT because their levels tend to be very low. This average number of pulse-return photoelectrons can be calculated in a straightforward manner as follows:

$$\bar{\eta}_s = \left[\frac{P_T}{\pi/4\Phi^2}\right]\left[\frac{A_T \mathcal{R}}{R^2}\right]\left[\frac{\pi/4 D_o^2}{R^2}\right]\left[\frac{\delta T}{h\nu}\right]\eta \qquad (4\text{-}40)$$

where

P_T = transmitted power, w

A_T = cross-sectional target area, cm^2

$\mathcal{R}$ = reflection efficiency, sr^{-1}

D_o = optical diameter of system, cm

δT = sample period, sec

Φ = beam angle, rad

R = range to target, cm

$h\nu$ = photon energy, erg

η = overall optical efficiency (optical, transmittance, quantum, etc.)

The transmitted power P_T resides more or less uniformly in some finite beam angle Φ. Therefore, the first bracketed term in Eq. 4-40 is the transmitted beam power density in w sr^{-1}. The second bracket is the solid angle subtended by the target times its reflection efficiency in the direction of the receiver. The product of the first two brackets thus gives an apparent target radiant intensity in w sr^{-1}. The third bracket is the solid angle subtended by the receiving optics. The fourth bracket is a factor for converting the received power into a number of quanta.

Within the sample period δT, an average number of background photoelectrons $\bar{\eta}_B$ will be generated by the naturally illuminated background seen by the receiver. Typically, this primary source of noise in a receiver is generated by the random variation about this average. For most purposes, it is valid to assume Poisson statistics for photoelectron generation rate, where the mean-square variation is equal to the average. Thus, the root-mean-square background noise is

$$\text{rms noise} = \sqrt{\bar{\eta}_B} = \left[H_\lambda \Delta\lambda\omega\left(\frac{\pi}{4}\right)D_o^2\left(\frac{\delta T}{h\nu}\right)\eta\right]^{1/2} \qquad (4\text{-}41)$$

The received signal $\bar{\eta}_s$ should be larger than $\sqrt{\bar{\eta}_B}$ by some factor S/N in order to assure accurate detection. The required transmitted power then is, from Eqs. 4-40 and 4-41,

$$P_T = \frac{(S/N)\Phi^2 R^4}{2A_T D_o \mathcal{R}}\left[\frac{\pi H_\lambda \Delta\lambda\omega h\nu}{(\delta T)\eta}\right]^{1/2} \qquad (4\text{-}42)$$

if the beam is larger than the target, or

$$P_T = \frac{(S/N)R^2}{2D_o\mathcal{R}}\left[\frac{\pi H_\lambda \Delta\lambda\omega h\nu}{(\delta T)\eta}\right]^{1/2} \qquad (4\text{-}43)$$

if the target is larger than the beam.

The accuracy of a laser rangefinder depends upon the transmitted pulse shape and the clock oscillator accuracy. A recently-developed rangefinder, using a pulsed YAG:Nd laser[14] produces 20-nanosec-wide pulses with a peak power of 750 kw at a rate of 6 pulses per min. The measurement range is 9,990 ft in clear weather, and 5,700 ft in haze, with an accuracy of ±10 ft. The tripod mount can be positioned aligned to true north within ±1 min of arc, by means of a sensitive magnetic compass. Azimuth and elevation accuracy is within ±0.1 deg.

4-2.3 COMMUNICATIONS AND DATA TRANSMISSION

Laser communication systems are particularly useful due to their narrow beam-width, high-spectral radiance and monochromacity. A narrow beam enables a system to be highly resistant to eavesdropping but requires precise alignment of transmitting and receiving elements. The narrow optical bandwidth enables the background to be filtered out very efficiently. Disadvantages of earthbound systems are the adverse effects of atmosphere and weather, line-of-sight restrictions, beam alignment difficulties, and *danger to human vision.*

The effects of atmospheric absorption and scattering can be overcome by the use of optical

transmission structures, such as pipes or optical waveguides in which the environment is controlled, or by transmitting over only short path lengths through the atmosphere.

Short-range optical communication links (several experimental systems have been produced) may have valuable tactical applications. Amplitude modulation of cw laser radiation is distorted by the atmosphere, particularly in the low a-f (audio frequency) range up to 250 Hz. Pulse code modulation (PCM) is the most promising of the many modulation techniques tried. PCM systems quantize and convert the analog signal to a digital code for transmission in pulse form. This system is ideally suited for short-range communication using pulsed room-temperature injection lasers operating at 12 to 20 kHz[15].

An unusual form of optical pulse code modulation/polarized light (PCM/PL) is being developed for transmission of wide-band video information in earth and space applications[16]. In this scheme, the output from a cw laser is polarization modulated; a digital "one" is transmitted as right-circular polarization; and, a digital "zero" is sent as left-circular polarization. At the receiving end, the beam passes through a Wollaston prism which has the property of deflecting polarized light in either of two directions, depending upon the polarization. The two polarized components of the beam are thus separated and directed to separate photomultipliers. Redundancy is provided because each photomultiplier circuit can produce the original signal. PCM/PL is about 3 dB more efficient than other techniques and cancels out common mode noise.

REFERENCES

1. D. C. Northrop, Quart. Rpt. No. 31, Services Electronics Res. Lab., Bladock, Hertz, Eng., (July 1953).
2. F. Urbach, N. R. Nail, D. Pearlman, "Observation of Temperature Distributions and of Thermal Radiation by Means of Nonlinear Phosphors", J. Opt. Soc. Am. 39, 1011-1019 (1949).
3. J. A. Jamieson, et al., *Infrared Physics and Engineering*, McGraw-Hill, Inc., N. Y., 1963.
4. Tippet, et al., *Optical and Electro-Optical Information Processing*, M. I. T., 45-47 1965.
5. S. Goldman, *Frequency Analysis, Modulation and Noise*, McGraw-Hill, Inc., N. Y., 1948.
6. W. L. Wolfe, Ed., *Handbook of Military Infrared Technology*, Office of Naval Research, Washington, D. C., 1965.
7. A. S. Lock, et al., *Guidance*. D. Van Nostrand Co., Inc., N. Y., 1965.
8. M. R. Holter and W. L. Wolfe, "Optical-Mechanical Scanning Techniques", *Proc. IRE* 1546 (Sept. 1959).
9. M. Schwartz, *Information Transmission and Modulation of Noise*, McGraw-Hill, Inc., N. Y., 1959.
10. *Van Nostrand's Scientific Encyclopedia*, D. Van Nostrand Co., Inc., N. Y., 1958.
11. G. L. Clark, *The Encyclopedia of Spectroscopy*, Reinhold, N. Y.
12. B. K. Johnson, *Optics and Optical Instruments*, Dover Press, N. Y., 1947.
13. S. B. Gibson, et al, *Characteristics of a 20 kw, Liquid-Cooled, Xenon Arc Searchlight*, Night Vision Laboratory, U. S. Army Electronics Command, Ft. Belvoir, Va.
14. "Three in One Mission", *Electronics* 44 (October 30, 1966).
15. S. W. Ross and J. N. Neville, *U. S. and Soviet Army Applications of Lasers—Updating*, Vol. 1, Report No. EDL-M828, Sylvania Electronic Defense Laboratories, Mt. View, Calif., Sept. 1965 (Secret).
16. B. Miller, "Wideband Video in Digital Form Transmitted Over Laser System", *Aviation Week and Space Technology* 61-65 (22 Aug. 1966).

CHAPTER 5

IR SYSTEM DESIGN*

5-1 SYSTEM APPROACH TO INFRARED DESIGN

The system approach to infrared design engineering is a technique whereby all aspects of the total system are considered in the design of each individual subsystem and component. This systematic approach entails defining of the total mission objectives; the priority rankings and relative weightings of each individual mission objective; the operational environment in which the system must function; the target, background, and transmission characteristics; the measures of effectiveness against which alternate designs and requirements can be evaluated; and, finally, a methodology by which a final system design can be synthesized. For the system to perform effectively in the operational environment, detailed consideration must be given to reliability, maintainability, logistics, and ground support equipment.

The first task in the evolution of system design involves the identification of the mission objectives and subsequent translation of these objectives into a set of functional requirements to be satisfied in the design of the sensor system. For example, the functional requirements for an IR search system might include threat warning, threat identification, and threat priority assignment. Each may have different (and perhaps conflicting) implications on the performance parameters of the sensor system. The output of the functional requirement analysis includes characteristics associated with the operational performance of the system such as detection range, detection probability, minimum signal-to-noise ratio, sensitivity, false alarm† rate, frame time, total field of view, resolution, and tracking rates. It should always be remembered that this baseline set of operational parameters must be continuously re-evaluated in light of cost effectiveness considerations.

Cost effectiveness is the mission performance achieved at a given cost for one system approach. This determination must be made in the broadest possible sense; i.e., the measure(s) of effectiveness must be appropriate and all the costs, amortized over the sensible life of the system, must be included. The question of whether increased effectiveness at increased cost should be provided depends on the allocation of resources between qualitatively different systems and is beyond the scope of the analysis to be performed by the infrared system designer.

Following the delineation of functional requirements, a set of baseline sensor parameters are established to satisfy the performance requirements. These parameters include optical diameter, optical speed (f/no.), number of detectors, detector size and aspect ratio, detector type, the operating spectral region, electrical bandwidth, and the desired output display or format. Since most of these parameters are interrelated and are functions of the performance parameters, the actual selection of numerical values must follow a detailed parametric trade-off analysis in which the possible constraints of size, weight, power, and cost effectiveness considerations are evaluated. The output of this phase of analysis will be a baseline preliminary design specification.

At this point the mission objectives should be re-examined, several alternative sets of functional requirements established, and a set of sensor parameters developed for each requirement. The optimum set of performance parameters and sensor parameters can then be resolved through the synthesis and development of a cost effectiveness model that can relate sensor and performance parameters to mission effectiveness and cost. The end result will then be the performance and design specification which defines the final system.

5-2 REQUIREMENTS

Systematic evaluation of mission requirements, functional requirements, targets, back-

---

* Written by S. Braunheim and K. Seyrafi

† When a signal waveform amplitude exceeds a preset threshold level due to noise, this excess of threshold is called false alarm.

grounds, and the effects of atmospheric transmission establishes the basic framework within which systems are designed and optimized. The interrelationships among the various sensor parameters and performance parameters, and the constraints and objectives of the mission constitute the analytical and quantitative base for establishing cost effectiveness relationships.

5-2.1 SYSTEM ANALYSIS

The paragraphs which follow describe the methodology through which functional requirements are established. A specific example is assumed throughout the remaining paragraphs to provide a few concrete situations in what must be an otherwise fairly-general discussion. For this purpose, the design of an IR sample-data tracker for use on an inexpensive surface-to-air missile (SAM) was selected. Table 5-1 lists and defines the symbols used.

5-2.1.1 Operational Environment

The operational environment delineates the mission characteristics and objectives in such a manner as to permit the development of the functional requirements. Thus, it is at this point that mission profiles and overall objectives should be established.

The geometric and kinematic relationships between the sensor platform and the target provide the initial clues into the requirements for detection range, field of view, tracking rates, and frame times. For example, assume that the mission of the sample SAM is to provide point defense capability for such targets as bridges, command posts, supply dumps, etc., against low-altitude attack jet aircraft armed with conventional low-drag ordnance. An analysis of the ordnance ballistics and other characteristics would show that, for a given drop altitude and

TABLE 5-1. LIST OF SYMBOLS

SYMBOL	DEFINITION	UNITS
A	Maximum amplitude of angular scan	rad
A_{bkd}	Background area in receiver field of view	m^2
A_d	Detector area	cm^2
A_i	Inherent availability	—
A_{is}	Area of illuminated spot	m^2
A_o	Optics collecting area	cm^2
A_r	Receiver area	m^2
A_t	Finite area of target	m^2
B	Background radiance	$w\ cm^{-2}\ sr^{-1}$
$B(\lambda)$	Background spectral radiance	$w\ sr^{-1}\ \mu^{-1}$
B_a	Radiance of the intervening atmosphere at ambient temperature	—
C	Contrast of target and background	—
c	Speed of light	$cm\ sec^{-1}$ or $m\ sec^{-1}$
D^*	Detector specific detectivity	$cm\ Hz^{1/2}\ w^{-1}$
$\hat{D}^*$	Detectivity of detector before roll-off	$w\ cm^{-2}$

TABLE 5-1. LIST OF SYMBOLS (cont'd)

SYMBOL	DEFINITION	UNITS
D_o	Optical diameter Aperture diameter Minimum object dimension	cm cm m
$D_o(max)$	Maximum diameter constraint	cm
DL	Diffraction limit	rad
d	Detector linear dimension	cm
d_{min}	Minimum linear dimension of element	
E	Acquisition threshold level	Hz
e	Charge on the electron	1.602×10^{-19} C
f	Frequency	Hz
$f/no.$	Optical speed	—
f_a	Lower cutoff frequency	Hz
f_b	Upper cutoff frequency of filter	Hz
f_c	Filter center frequency	Hz
Δf	Noise equivalent bandwidth	Hz
$\Delta f'$	Electrical (filter) bandwidth	Hz
f_r	Natural frequency of servo	Hz
f_s	Detector aspect ratio (height-to-width) Sampling rate of servo	— —
G	Electron multiplier gain	—
H	Irradiance at the collecting aperture	w cm^{-2}
H_B	Apparent background irradiance Total background radiant intensity	w cm^{-2} w cm^{-2}
$H_c(\lambda)$	Background radiant intensity	w cm^{-2} μ^{-1} sr^{-1}
H_t	Target irradiance	w cm^{-2}
H_λ	Spectral irradiance on ground	w m^{-2} μ^{-1}
h	Planck's constant	6.625×10^{-34} w sec^{-2} or 6.625×10^{-34} J-sec
h/σ	Threshold-to-rms noise ratio	—
I	Moment of inertia of rotating seeker components about the spin axis	i. -oz sec^2
i_{bkd}	Background induced current	A
i_d	Photomultiplier dark current	A

TABLE 5-1 LIST OF SYMBOLS (cont'd)

SYMBOL	DEFINITION	UNITS
i_s	Signal current	A
J	Radiant intensity	w sr^{-1}
$J(\lambda)$	Spectral radiant intensity of target Spectral radiant intensity from target	w sr^{-1} μ^{-1} w sr^{-1} μ^{-1}
$J_a(\lambda)$	Apparent spectral radiant intensity from target	w sr^{-1} μ^{-1}
K_1, K_2	Proportionality constants	—
K	Number of linear resolution elements required for identification	—
K_d	Scan efficiency	—
k	Boltzmann's constant	1.38×10^{-23} J/°K^{-1}
N	Radiance of illuminated spot on the ground Rms system noise	w m^{-2} w cm^{-2}
NEI	Noise equivalent input (sensor sensitivity)	w cm
N_{fa}	False alarm rate	crossings per sec
n_b	Number of bars	—
n_d	Number of detectors	—
n_{do}	Required number of detectors	—
n_g	Number of resolution elements in the acceptance gate	—
n_l	Number of looks required for detection	—
n_{tot}	Total number of resolution elements in the field of view	—
P_{bkd}	Background power level	w
P_d	Reflected power impinging upon detector	w
P_l	Laser power	w
p	Probability that a noise pulse exceeds the threshold in any given resolution element	—
$\hat{p}$	Probability of an initiating noise pulse occurring	—
$\bar{p}$	Probability that at least one noise pulse will exceed threshold in acceptance box	—
p_d	Detection probability	—
p_f	Probability of a false alarm	—
p_s	Probability that any one of n_l looks exceeds the threshold	
$p(v)$	Probability distribution function of noise	—
Q_B	Incident photon flux density	phot cm^{-2} sec^{-1}

TABLE 5-1. LIST OF SYMBOLS (cont'd)

SYMBOL	DEFINITION	UNITS
R	Range	cm
R_c	Visible detection range of target due to cloud cover	m
R_d	Weapon release range to target	m
R	Load resistor	ohm
R_{min}	Minimum acceptable detection range	m
R_s	Safety margin range	m
r_s	Scan rate	rad sec^{-1}
S/B	Signal-to-background noise ratio	—
S/N	Signal-to-noise ratio	—
S	Amplitude of signal	—
T	Threshold level of system	—
	Torque	in.-oz
	Time	sec
	Temperature	°K
ΔT	Effective temperature contrast between target and background	°K
T/σ	Number of levels the threshold level is above the noise level	—
t_d	Dwell time of an individual detector	sec
	Time required for a point image to cross a detector	sec
t_{fa}	Mean time between false alarms	sec
	Time period	sec
t_m	Time of flight of missile	sec
t_o	Frame time that would produce the roll-off knee frequency (3 dB point) for a given size detector	sec
t_r	Reaction time prior to launching the SAM	sec
t_s	Frame time	sec
	Frame (scan) time	sec
$t_s(opt)$	Optimum scan time	sec
$t_{s(max)}$	Maximum frame time	sec
$t_s(max)$	Maximum scan time	sec
V_m	Missile velocity	m sec^{-1}
V_t	Attacking aircraft velocity	m sec^{-1}
	Velocity of target	m sec^{-1}
	Target speed	m sec^{-1}

TABLE 5-1. LIST OF SYMBOLS (cont'd)

SYMBOL	DEFINITION	UNITS
v	Instantaneous noise voltage	V
X	Detector linear dimension Detector height	cm cm
Y	Detector linear dimension Detector width Empirically determined number	cm cm —
α	Receiver field of view Instantaneous field of view	rad sr
α_{min}	Limiting field of view	sr
α_r	Angular resolution required of system	rad
e_e	Coefficient which relates the peak or rms value of the signal output to the input irradiance	—
e_o	Optical efficiency	—
$\epsilon(S)$	Error for a ramp input	—
$\epsilon(t)_{peak}$	Peak value of error	—
η	Quantum efficiency of detector Photocathode quantum efficiency	— —
θ	Angle between laser beam and normal to surface	rad
$\theta_{x/y}(min)$	Minimum angular dimension of an individual detector	rad
θ_X	Detector angular resolution in direction of scan Detector's angular dimension Detector width	rad rad rad
θ_Y	Detector's angular dimension	rad
θ_s	Sinusoidal scan angle	rad
$\dot{\theta}_s$	Angular scan rate	rad sec^{-1}
$\dot{\theta}$	Angular rate	rad sec^{-1}
λ_c	Long-wavelength cutoff	cm
λ_1	Short-wavelength cutoff	cm
λ_2	Long-wavelength cutoff	cm
$\Delta\lambda$	Spectral bandpass Spectral bandwidth of receiver optical filter	μ μ
ρ	Reflectance	—
ρ_m	Minimum signal-to-noise requirement	—
ρ_β	Signal-to-background plus noise ratio	—
ρ_N	Signal-to-noise ratio S/N	—

TABLE 5-1. LIST OF SYMBOLS (cont'd)

SYMBOL	DEFINITION	UNITS
n	Rms value of noise	V
τ	Transmittance	—
$\tau(\lambda)$	Atmospheric transmission for target	—
$\tau_B(\lambda)$	Atmospheric transmission for background	—
τ_a	Atmospheric transmission	—
τ_B	Transmission of the atmospheric path between the optical system and the background	—
$\tau(\lambda, R)$	Atmospheric transmission as a function of wavelength and pathlength	—
τ_{lo}	Transmission factor of illuminator optics	—
τ_{ro}	Transmission factor of receiver optics	—
ϕ	Field of view in the direction of scan	rad
Ω	Total field of view	mrad or sr
Ω_v	Angular precession velocity about output axis	rad sec^{-1}
ω	Scan rate of system Angular velocity of scanner	rad sec^{-1} rad sec^{-1}
ω_α	Angular subtense of detector	sr
ω_{max}	Maximum angular scanning rate of field of view	rad sec^{-1}
ω_r	Seeker spin velocity	rad sec^{-1}

speed, the weapons must be released some distance R_d prior to reaching the target in order for the weapon to hit the target and for the attacking aircraft to safely clear the area. It is necessary for the defense missile to intercept the aircraft with some safety margin prior to the aircraft reaching the range R_d. If we represent the safety margin range as R_s and the missile velocity as V_m, the time of flight t_m of the missile may be calculated as (it is implicitly assumed that the aircraft altitude relative to range is small)

$$t_m = \frac{R_d + R_s}{V_m} \qquad (5\text{-}1)$$

If the reaction time prior to launching the SAM is t_r and the attacking aircraft velocity is V_t, then the minimum acceptable detection range may be calculated by

$$R_{min} = V_t(t_m + t_r) \qquad (5\text{-}2)$$

Analogous expressions can be developed for the search field frame time and other sensor performance parameters. For an infrared instrument, the minimum signal-to-noise ratio is determined by the required detection probability and the false alarm rate. If the false alarm rate requirement is very stringent, the system must operate at a high threshold level which would be significantly higher than the rms noise values of the system. This implies that the peak signal must be at least equal to the threshold level. If the detection probability requirement is 99 percent and the system noise is Gaussian white noise (normal), then the minimum signal must be at least 2.33 times noise standard deviations (usually expressed as root mean square-rms-noise) above the threshold in order to assure this

probability of detection. If the threshold setting based on false alarm rate considerations must be 6 times rms noise, then the minimum signal-to-noise ratio for 99 percent detection probability would be 8.33:1. In this manner (clearly simplified at this point), the mission requirements and operational environment determine the functional requirements.

5-2.1.2 Functions

The mission objectives and requirements determine the sensor parameters and sequence of functions that the sensor must perform, thereby, dictating the functions of the subsystems and components within the overall system. A definite delineation of these generic functions is imperative if reasonable and realistic requirements are to be allocated. Again consider the IR seeker on the SAM. Prior to initiation of the SAM launch sequence it is necessary to detect and acquire the target. This requires prior acquisition of the target by some other element of the SAM system (e. g., infrared or radar acquisition system) and then a transfer of this acquisition information to the IR seeker. The pointing accuracy of the initial acquisition system determines, in large measure, the field of view requirements of the IR seeker. If target acquisition is also a requirement of the IR seeker, it must be capable of rejecting background clutter and providing the sensitivity necessary to detect and lock on a target at the desired range. Thus, the acquisition function in conjunction with the operational environment will establish the ground rules governing the field of view, background clutter rejection, and sensitivity requirements.

The second function of the seeker is to track the target and to provide guidance error signals. In this mode of operation, the seeker must reject the countermeasure response of the target and provide data at the rate and resolution required to achieve the desired kill probability. It must possess the necessary gimbal freedom for maintaining the target in the field of view despite target maneuvers. A sufficiently-high angular tracking rate capability must be provided to maintain a small miss-distance at the intercept point. In order to establish numerical requirements for these functions, the missile and target motion usually undergo dynamic simulation on an analog computer. However, as is often the case, the volume weight or power constraints of the missile nose will not permit the installation of a seeker which provides the full desired mission capability. As a consequence, the designers of both the IR seeker and the mounting platform must determine whether the mission performance can be resolved if the power required can be reduced; or whether more power, weight, or volume can be allocated to the seeker.

It can now be seen that the mission objectives, operational environment, and the functional requirements have begun to yield a preliminary set of performance specifications from which the sensor characteristics can be developed.

5-2.1.3 Requirements Analysis

Following the definition of the functional requirements, the system designer can begin to formulate the sensor system parameters. In general, these are derived analytically from the system sensitivity equations; the considerations of optical, mechanical, and electrical design; cooling requirements and power constraints; and size and weight constraints.

Assume a target of radiant intensity J in w sr^{-1} at a given range R in cm through an atmosphere that provides transmittance τ. The irradiance at the collecting aperture is expressed as

$$H = \frac{J\tau}{R^2} \text{ , w cm}^{-2} \qquad (5\text{-}3)$$

based on the inverse-square law. The sensor sensitivity or noise equivalent input (NEI) is given by Eq. 5-4, which is discussed further in par. 5-3.3.1.

$$NEI = \frac{\sqrt{A_d \Delta f}}{\epsilon_o \epsilon_e A_o D^*} \text{ , w cm}^{-2} \qquad (5\text{-}4)$$

where

A_d = detector area, cm^2

Δf = noise equivalent bandwidth, Hz

ϵ_o = optical efficiency

ϵ_e = coefficient which relates the peak or rms value of the signal output to the input irradiance (a unique function of the filter characteristics and input waveforms)

A_o = optics collecting area, cm^2

D^* = detector specific detectivity, cm Hz$^{1/2}$ w^{-1}

If the minimum signal-to-noise requirement is ρ_m then the noise equivalent input can be rewritten as:

$$NEI = H = \frac{J_T}{R^2 \rho_m} = \frac{\sqrt{A_d \Delta f}}{\epsilon_o t_e A_o D^*} \qquad (5\text{-}5)$$

It is thus shown (for this example) how the functional requirement for detection at range R is related to the sensor parameters of Eq. 5-4. It is also apparent that there is no unique solution to Eq. 5-5 since an infinite combination of numerical values will satisfy the equation.

Assume the need to scan the total field of view Ω steradians once per t_s seconds, and n_d detectors of instantaneous field of view α steradians to scan the field at 100 percent duty cycle, then the dwell time t_d of an individual detector is expressed as

$$t_d = \frac{n_d t_s \alpha}{\Omega}, \text{ sec} \qquad (5\text{-}6)$$

The filter (electrical) bandwidth $\Delta f'$ for scanning systems is often given as

$$\Delta f' = \frac{1}{2t_d}, \text{ Hz} \qquad (5\text{-}7)$$

in which case the system response will be reduced by a factor of 0.60 to 0.70 of its maximum amplitude. Theoretically, one would like to have $\Delta f' = \infty$ so that there would be no degredation in system response. However, because of the increase in noise power with bandwidth, Δf should be limited. Practically, the selection of $\Delta f'$ as indicated by the relationship of Eq. 5-7 is a compromise between system response and noise. In this case the response has been reduced by a factor of 0.6—0.7, depending on its application. The detector area A_d can be expressed as follows in terms of optical speed, aperture, and field of view.

$$A_d = (f/no.)^2 D_o^2 \alpha \qquad (5\text{-}8)$$

where

$f/no.$ = optical speed

α = instantaneous field of view, sr

D_o = optical diameter, cm

The background clutter rejection capability of a sensor is often expressed in terms of the signal-to-background noise ratio. This reflects the fact that the higher the ratio, the higher the threshold can be set (given adequate sensitivity) and, thus, the more background clutter that can be eliminated. Since background radiance is a complete random function, a frequently-used criterion is a "worst case" or peak background radiance. The signal-to-background noise ratio S/B may be computed as

$$S/B = \frac{J_T}{\alpha R R^2} \qquad (5\text{-}9)$$

where

B = background radiance, w cm^{-2} sr^{-1}

Eq. 5-9 applies when the target is a point source and the background is an extended source such as a cloud. Both the signal-to-background and signal-to-noise ratios can be improved by making the instantaneous field of view α small in the equation. This is done at the expense of cost and complexity compared to a system of equal size and performance, since the number of scanned elements in a frame time is Ω/α. The frame time can be achieved by using additional detectors. There are also physical constraints on the minimum dimension a_{min} of current state-of-the-art detectors which can be fabricated to form large arrays. Thus, the constraint imposed dictates that

$$d_{min} \leq f/no.\, D_o \theta_{x/y}(min) \qquad (5\text{-}10)$$

where $\theta_{x/y}\,(min)$ is the minimum angular dimension in radians of an individual detector. Since total package size, weight, and volume are

strongly influenced by *f/no.* and diameter; severe penalties can result from using small detectors.

For sensors wherein the limiting source of noise is the internal system noise—provided this noise is Gaussian distributed with a flat amplitude (white) spectrum—the ratio of threshold to the rms noise required for any prescribed false alarm rate is

$$\frac{T}{N} = \sqrt{\frac{2}{n_\ell}\, \ell n\left[\Delta f n_d t_{fa} (n_g)^{n_\ell - 1}\right]} \qquad (5\text{-}11)$$

where

T = threshold level of the system

N = rms noise

n_ℓ = number of looks required for detection

n_d = number of detectors

Δf = noise equivalent bandwidth, Hz

t_{fa} = mean time between false alarms, sec

n_g = number of resolution elements in the acceptance gate

The proof of Eq. 5-11 (see Eqs. A-1 through A-12 in the Appendix to this chapter) assumes Rayleigh-distributed white noise such that the probability density functions $p(v)$ can be written as

$$p(v) = \frac{v}{\sigma} \exp\left[-\frac{1}{2}\left(\frac{v}{\sigma}\right)^2\right] \qquad (5\text{-}12)$$

where

$p(v)$ = probability distribution function of noise

v = instantaneous noise voltage, V

σ = rms value of noise, V

and the probability that a noise pulse exceeds any given threshold T is

$$p(v>T) = \int_T^\infty \frac{v}{\sigma} \exp\left[-\frac{1}{2}\left(\frac{T}{\sigma}\right)^2\right] dv$$

$$= \exp\left[-\frac{1}{2}\left(\frac{T}{\sigma}\right)^2\right] \qquad (5\text{-}13)$$

Eq. 5-11 is the result of σ being defined as the rms value of the noise which is denoted as N.

Also it can be shown that the signal-to-noise ratio S/N required for any given detection probability p_d when $S \gg N$, is

$$\frac{S}{N} = \frac{T}{N} + \sqrt{2\ell n\left[\frac{1}{1-(p_d)^{\frac{1}{n_\ell}}}\right]} \qquad (5\text{-}14)$$

Given a signal of amplitude S, a threshold value T, and noise N, all the same units (volts, for example), and a detection rule of n_ℓ looks; the detection probability can be derived in the manner which follows. First, the probability p_s that any one of the n_ℓ looks exceeds the threshhold is expressed as

$$p_s = 1 - \exp\left[-\frac{1}{2}\left(\frac{S-T}{\sigma}\right)^2\right] \qquad (5\text{-}15)$$

Assuming that an overall detection probability of p_d is required, then

$$p_d = (p_s)^{n_\ell} \qquad (5\text{-}16)$$

or

$$p_s = (p_d)^{\frac{1}{n_\ell}} \qquad (5\text{-}17)$$

Substituting Eq. 5-15 into Eq. 5-17 gives

$$(p_d)^{\frac{1}{n_\ell}} = 1 - \exp\left[-\frac{1}{2}\left(\frac{S-T}{\sigma}\right)^2\right] \qquad (5\text{-}18)$$

and

$$\exp\left[-\frac{1}{2}\left(\frac{S-T}{\sigma}\right)^2\right] = 1 - (p_d)^{\frac{1}{n_\ell}} \qquad (5\text{-}19)$$

or

$$\frac{S-T}{\sigma} = \sqrt{2\ell n\left[\frac{1}{1-(p_d)^{\frac{1}{n_\ell}}}\right]} \qquad (5\text{-}20)$$

Substituting N for σ and rearranging terms, results in Eq. 5-14.

For example, assume the target analysis as having shown that, in the spectral band of interest (3.3 to 4.8 microns for instance), the radiant intensity of the target is 7 w sr^{-1}, and the mission analysis as having indicated a detection range of 9 km through an atmosphere that provides 0.3 transmittance. From Eq. 5-3 the irradiance will then be

$$H = \frac{7 \times 0.3}{(9 \times 10^5)^2} = 2.6 \times 10^{-12} \text{ w cm}^{-2}$$

Let us also assume the background analysis as having shown the effective peak-background-radiance in the band to be 3×10^{-5} w cm^{-2} sr^{-1} and that a signal-to-background noise ratio of 2.0 is necessary for background clutter rejection. Then from Eq. 5-9

$$2.0 = \frac{2.6 \times 10^{-12}}{\alpha \times 3 \times 10^{-5}}$$

The instantaneous field of view α must be approximately 4.4×10^{-8} sr which implies a detector instantaneous field of view of 2.1×10^{-4} rad, assuming a 1:1 aspect ratio. With a total field of view Ω of 10×10 mrad dictated by pointing accuracy requirements and a need to search the field with a single-bar linear array, the required number of detectors would be

$$n_d = \frac{10 \times 10^{-3}}{2.1 \times 10^{-4}} \cong 48$$

Let it also be assumed that in order to provide an adequate information rate, the system must scan the complete field 30 times per sec; then from Eq. 5-6

$$t_d = \frac{48 \times 0.033 \times 4.4 \times 10^{-8}}{100 \times 10^{-4}} = 7 \times 10^{-4} \text{ sec}$$

and from Eq. 5-7

$$\Delta f' = \frac{1}{2 \times 7 \times 10^{-4}} = 710 \text{ Hz}$$

If a detection probability requirement of 0.99 is required, an average false alarm rate of one per hour is acceptable, and a single-look detection criterion is used, a threshold setting of 6.4 (Eq. 5-11) and signal-to-noise ratio of 9.5 (Eq. 5-14) will be required. From Eq. 5-5 the required sensitivity is determined to be

$$NEI = \frac{2.6 \times 10^{-12}}{9.5} = 2.74 \times 10^{-13} \text{ w cm}^{-2}$$

It is now assumed that an analysis has shown lead selenide (PbSe) to be potentially the most suitable detector material and that an average D^* of 4×10^{10} cm Hz$^{1/2}$ w^{-1} can be expected when the detector is cooled to $-80°$C. By replacing A_d and A_o of Eq. 5-4 by their respective terms,

where

$$A_d = (f/no.)^2 D_o^2 \alpha \quad \text{[Eq. 5-8]}$$

and

$$A_o = \left(\frac{\pi}{4}\right) D_o^2$$

(representing the detector and collector area, respectively), NEI can be expressed as

$$NEI = \frac{4\, f/no. \sqrt{\alpha \Delta f}}{\pi \epsilon_o \epsilon_e D_o D^*} \qquad (5\text{-}21)$$

or

$$D_o = \frac{4 f/no. \sqrt{\alpha \Delta f}}{\pi \epsilon_o \epsilon_e D_o D^* NEI} \qquad (5\text{-}22)$$

By further assuming an $f/no.$ of 3.0 for purposes of optical resolution and packaging and an optical efficiency ϵ_o of 0.5,

$$D_o = \frac{4 \times 3 \times \sqrt{4.4 \times 10^{-8} \times 710}}{3.14 \times 0.5 \times 0.65 \times 4 \times 10^{10} \times 2.74 \times 10^{-13}}$$

$$= 5.8 \text{ cm} \approx 2.3 \text{ in.}$$

Next, a check is made to verify that neither the Rayleigh diffraction limit criterion or the cell size limit have been exceeded. The required resolution was 2.1×10^{-4} rad. The diffraction limit equation is

$$DL = \frac{2.44 \lambda_c}{D_o} < \sqrt{\alpha} \qquad (5\text{-}23)$$

Assuming the limiting wavelength to be 4.8 microns, then

$$DL = \frac{2.44 \times 4.8 \times 10^{-4}}{5.8} = 2 \times 10^{-4} \text{ rad}$$

and thus the diffraction limit is not exceeded. Assume next that the minimum linear dimension d_{min} of a PbSe element is 0.001 in. or 2.5×10^{-3} cm. From Eq. 5-10 it is shown that the detector dimension

$$d = 5.8 \times 3.0 \times 2.1 \times 10^{-4} = 3.65 \times 10^{-3} \text{ cm}$$

which is greater than the assumed d_{min} of 2.5×10^{-3} cm.

5-2.1.4 Block Diagrams

One of the most effective methods of describing all manner of complex systems is by means of simple pictorial diagrams. This technique is, of course, well-suited for illustrating the overall scope, internal functions, and interfacings of typical IR systems (see Fig. 5-1).

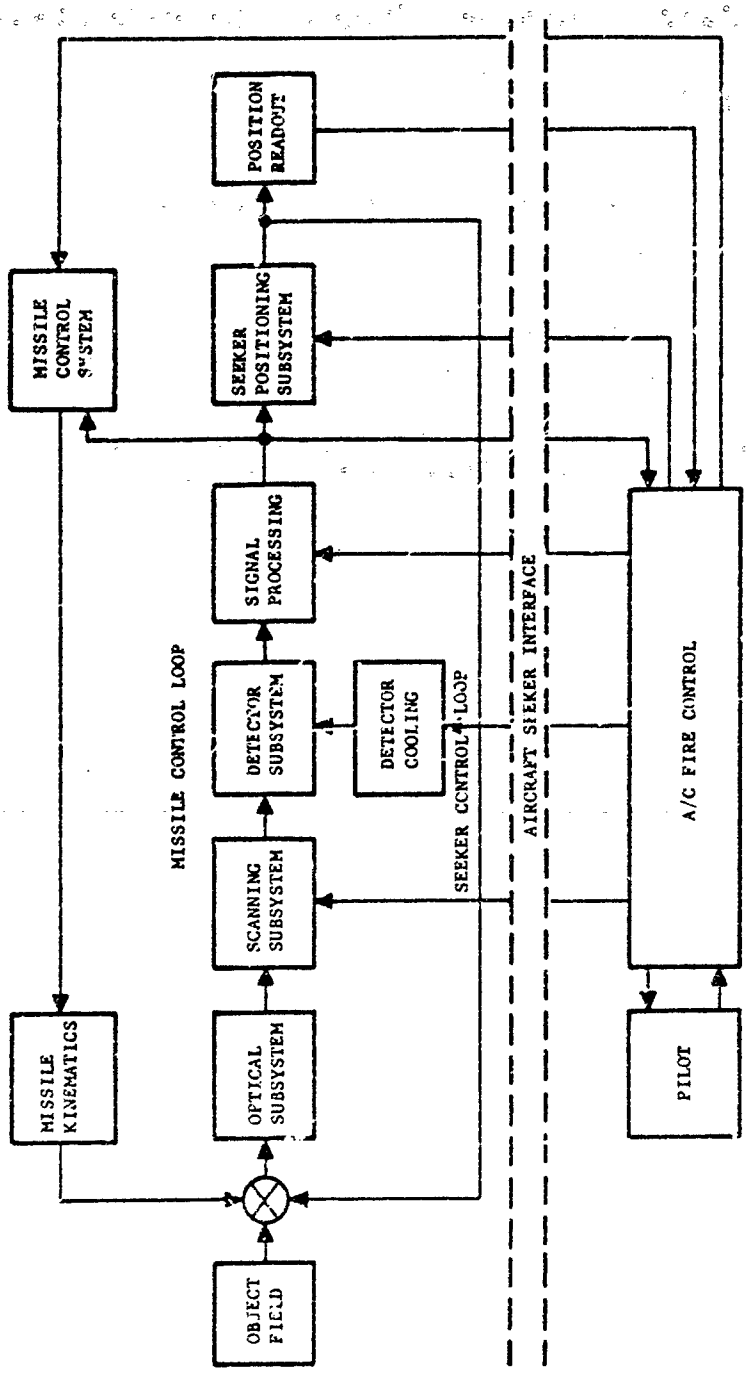


FIGURE 5-1. Infrared Missile Seeker Block Diagram

5-2.2 TARGET DEFINITION

The mission objectives and the generic class of targets of interest in any given application determine the functional role of the IR sensor. The target characteristics or interest in a reconnaissance or mapping mission are in most cases distinctly different from those of interest in a threat warning application. The spectral, spatial, temporal, and amplitude characteristics of the surrounding background, in large measure, determine the overall effectiveness of the system. If no significant distinguishable target characteristics exist relative to its background, the target is essentially camouflaged. In general, there will always be some distinguishable aspects of the target. The question is then relative to the type of detection system required and to its cost. Par. 2-6.2 and AMCP 706-128 discuss in detail some of the salient characteristics of military targets. The following two paragraphs describe the manner in which the system designer makes use of this information.

5-2.2.1 Spectral Radiant Intensity Bounds

The spectral characteristics of many military targets are often the most dominant differentiating feature of the target description. As a result, spectral optimization has become a highly effective tool in providing increased background rejection. Thus, it is extremely important for the system designer to have, or to be able to generate, an accurate spectral representation of a (minimum) target under the operating conditions expected to be encountered in the field.

For example, the spectral signature of a jet aircraft at side aspect is significantly different from that of the same aircraft viewed from the tail. In the first case, the signature is characterized by emission bands of H_2O and CO_2 combustion products at 2.7 and 4.4 microns, while in the second case, the signature is characterized by the blackbody emission of the hot tailpipe. Fig. 5-2 shows the various aspect emissions. When viewed from the nose aspect, the signature may be dominated by graybody emission from the relatively cool surface of the aircraft.

Ground targets, which exhibit relatively little energy compared to airborne targets, are generally characterized by small temperature differences between the target and the surrounding background.

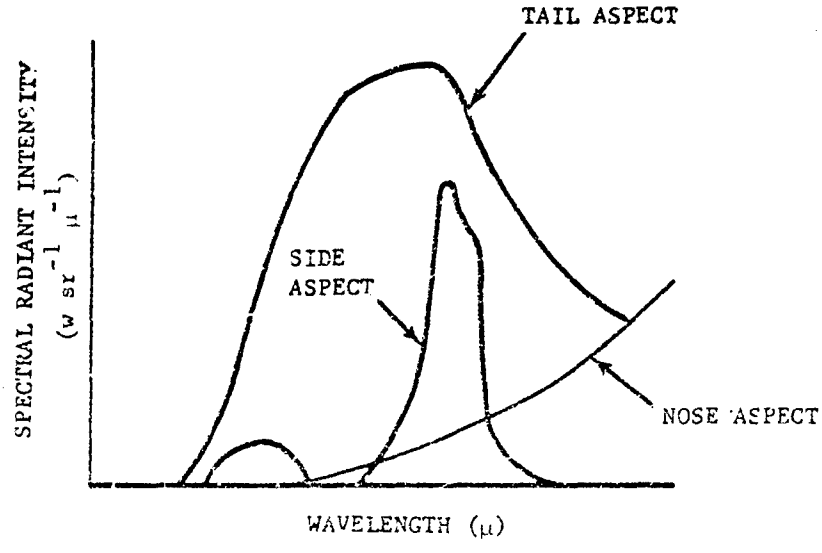


FIGURE 5-2. Spectral Characteristics of Aircraft vs Viewing Aspect

5-2.2.2 Radiance Gradients

Virtually all infrared sensors operate on the modulated gradient between the target and its surrounding background. Thus, from a signal standpoint, the effective signal is the difference between the target and the background. Thus, for example, a 300°K blackbody target contrasted against a 300°K blackbody background would not produce a discernible gradient. In many cases it is not required that the contrast be positive because negative gradients can also be processed. In the case of thermal imaging systems, targets are often specified in terms of the effective temperature contrast ΔT between target and background. For example, a sampan on a river might be said to have a $5°\Delta T$ gradient. It should be noted that this temperature gradient is the result of both emissivity and temperature differences. In the 8- to 14-micron region a ±5 percent emissivity difference around $\epsilon = 0.95$ is equivalent to approximately a 7°C temperature difference at 300°K ambient.

5-2.3 BACKGROUND DEFINITION

The spectral and spatial characteristics of the background relative to the target determine the degree of background clutter that must be processed and the resulting false alarm rate if background is the dominant source of noise. To this date there has been no satisfactory statistical representation of the underlying stochastic process of background noise. Weiner spectra representations of backgrounds have never been found to adequately represent the spatial distribution of backgrounds. As a result, the system designer must generally utilize concepts such as the signal-to-background ratio which can be quantitatively translated into false alarm rates only through complete knowledge of the probability distribution function. However, if the signal-to-background ratio is defined in terms of the minimum signal-to-maximum background and the ratio is significantly greater than unity, then in most cases the background can be adequately thresholded to a point where the false alarm rate will be within tolerable limits.

5-2.3.1 Terrain, Sea

The spectral signature of most ground backgrounds are characterized by diffusely reflected solar energy in the short wavelengths during daylight hours; and lunar, stellar, and air glow illumination at night. In the longer wavelengths (above 4 microns) the spectral characteristics are dominated by graybody emission at or near ambient temperature. Since most natural objects tend to have high emissivities in the long wavelengths the spectral signatures approach that of a blackbody. These subjects are discussed in greater detail in par. 2-6.1. The spectral absorptive properties of the intervening atmosphere and the re-emission of the atmosphere in the absorption bands tend to modify the spectral characteristics of the background in a fashion which in many cases is to the advantage of the designer. In many cases, terrain backgrounds will be at a range greater than or equal to the target range, which means that the background will be attenuated as much as or more than the target. It should also be noted that the relative position of the sun, the type and amount of cloud cover, and the type of terrain also affect the amplitude and spectral signature of backgrounds.

The spatial characteristics of terrain are usually dominated by a large low-frequency component and roll-off at about $1/f^2$ beginning at relatively low spatial frequencies. Many targets of interest may be characterized by point objects with corresponding high-frequency components. This tends to improve the background clutter problem if the proper spatial and electrical filters are used. In general, these filters limit the low-frequency response of the system. This reduces signal energy, but at a much slower rate than background energy, and thus improves the signal-to-background ratio.

5-2.3.2 Clouds

Clouds, or more exactly cloud edges, have been the traditional enemy of IR system designers because of the large radiance gradient between blue sky and a cloud edge. There are few analogous situations in the case of terrain backgrounds that produce gradients of this magnitude. As a result, systems that are designed to operate in a cloud environment must be carefully designed with respect to their spectral and spatial filtering properties.

As in the case of terrain, the spectral signature of clouds are dominated by emitted energy in the long wavelengths and reflected energy in the shorter wavelengths as illustrated in Fig. 5-3.

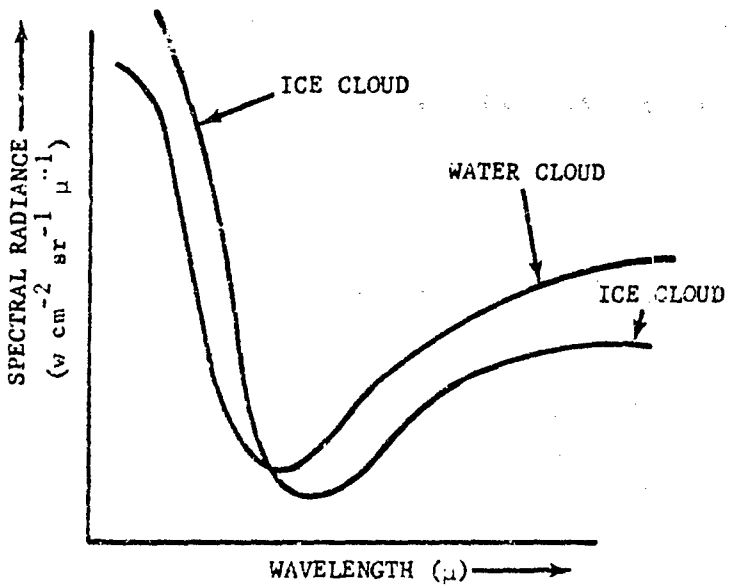


FIGURE 5-3. Typical Sunlit Cloud Radiances vs Wavelength

(Refer to par. 2-6 for more detailed information.) The spectral radiance of clouds in the reflected region is strongly affected by the solar scattering angle and the cloud altitude. High-altitude ice clouds tend to be better reflectors in the short wavelengths but, because of their cooler temperatures, have lower radiance than the low-altitude water clouds in the longer wavelengths.

5-2.3.2 Stellar Backgrounds

The stellar background provides two sources of background radiation. The first comes from the bright stars that are resolved or detected by the sensor. The number and position of resolvable stars can often be predicted in advance through the use of star tapes and ephemeris catalogues. When this information is stored in processing computers, these false targets can be eliminated. Another technique of stellar discrimination is based on the spectral signature of stars. Since most stars radiate as 5000° - 6000°K blackbodies, an instrument operating in two spectral regions has a potential for rejecting these objects relative to much cooler targets. In the regions of high-stellar density such as the galactic equator it is often impossible to process the information in real-time and as a consequence these regions are often avoided. A second source of stellar radiation is the noise emission from many unresolved stars. In some long-wavelength applications this may be the limiting noise source.

5-2.4 TRANSMISSION

The selective absorption of radiated infrared energy by the intervening atmosphere influences both the spectral characteristics and amplitude of the received target signal. As a consequence, a comprehensive understanding of absorption phenomena is mandatory for successful system design. The primary constituents that absorb infrared energy are the vibration and rotation bands of water vapor and carbon dioxide. Secondary absorption sources include nitrous oxide, ozone, methane, and other minor atmospheric cases. Particulate matter causes scattering of the infrared energy. The phenomenology of this subject is discussed in par. 2-4.

5-2.4.1 Absorption

The dominant absorption bands in the infrared are the 2.7-micron band of H_2O and CO_2, the 6.3-micron band of H_2O, and the 15-micron band of CO_2. Because of these absorption bands and the spectral characteristics of target radiation, the traditional operating bands which have

evolved include: 1.8 to 2.7 microns, 3 to 5 microns, and 8 to 14 microns. It should be noted that, since the atmosphere absorbs in the same bands where hot-gas plumes emit, only a small fraction of these strong emission bands is transmitted even through relatively short atmospheric paths. However, temperature and pressure broadening of the plume emission bands results in a small residual skirt of energy that is transmitted as shown in Fig. 5-4. It is apparent from Fig. 5-4 that unless both the spectral characteristics of the target and transmissions are well understood, the estimation of the transmitted target energy can be significantly in error. It should also be noted that nonhomogeneities in the atmosphere in the absorption regions can be a significant source of external scanning noise.

Model atmospheres have been developed to represent the differing climatological environments in which an infrared system may operate. For example, a model temperate atmosphere describes the altitude dependence of CO_2, H_2O, temperature, and pressure for mid-latitude regions. Similar models exist for tropical and arctic regions. In addition, transmission models by such authors as Thomas Altshuler[1] greatly facilitate the calculation of slant path transmission.

5-2.4.2 Obscuration

In addition to being a background problem, clouds are also a source of obscuration of infrared energy. This is often characterized by probability curves of cloud free line-of-sight as shown qualitatively in Fig. 5-5. The point A on the curve indicates that at an altitude of about 20 kft and an observer-target distance of 5 mi there is 80 percent probability of a cloud-free line of sight. While at the same altitude but for a 6 - mi observer-target distance, point B on the curve, there is 70 percent probability of a cloud-free line of sight. Curves such as those in Fig. 5-5 are especially useful in cost effectiveness studies. For example, from the point of view of aircraft detection it may be desirable to have a detection range of 20 km. If, however, there is only a 1 percent chance of having a cloud-free line of sight to that altitude, it may not be economical to provide that range capability. In addition, the relationships of Fig. 5-5 are useful in determining the expected range at which targets may emerge from cloud cover. This information is often required in determining frame rate requirements. For example, if a system must detect a target by some minimum range R_{min} and the cloud data indicate that there is a high probability that the target is not visible until some other range R_c and the velocity of the target is V_t, then the maximum frame time is

$$t_{s(max)} = \frac{R_c - R_{min}}{V_t}, \text{ sec } R_c > R_{min} \quad (5\text{-}24)$$

Cloud cover data are also often required for airborne and satellite reconnaissance systems in

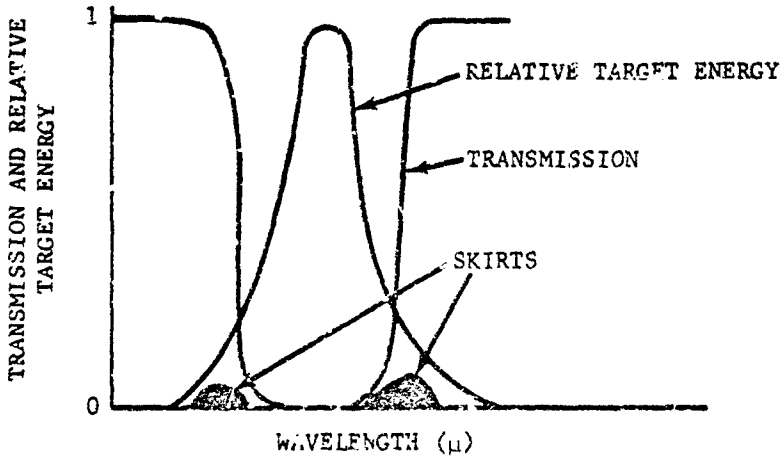


FIGURE 5-4. Relative Target Energy and Transmission vs Wavelength for Plume Emission

determining the percentage of time that ground targets will be visible.

5-2.5 COUNTERMEASURES

The designer of infrared systems must be constantly aware of potential countermeasures that may be used against the system. This knowledge can in some cases permit the inclusion of counter-countermeasures in the initial system design. In almost all cases the details of countermeasure techniques and their performance are highly classified and, as a result, the discussion must be general in nature. IR countermeasures are divided into two general groups representing two functional modes of operation: passive countermeasures and active countermeasures.

5-2.5.1 Passive Countermeasures

Passive countermeasures can be broadly defined as that class of countermeasures which are employed continuously and are not activated as a result of the presence of an attacker. Passive countermeasures exclude the deployment or ejection of physical objects from the defended platform. The most generally employed countermeasure against IR systems is a reduction in target signature. The primary objective of these techniques is to reduce or suppress the level of IR radiation from the defended platform to a level so low that the attacker cannot detect his target until it is too late to mount an attack. The three dominant suppressing techniques for missile and aircraft targets are shielding, cooling, and additives.

Shielding is a technique whereby the hot engine parts of an aircraft, helicopter, or any other potential target are shielded from all but a very narrow range of aspect angles, usually only a few degrees directly around the exhaust nozzle. *Shielding* and *cooling* are often used simultaneously in order to improve the overall effectiveness. *Fuel additives* are sometimes used in order to reduce plume radiation. These additives can be quenching in nature, such as water, which tends to reduce the temperature and thus the radiation of the plume emission, or they can be inhibiting in nature such as certain chemical additives.

5-2.5.2 Active Countermeasures

Active countermeasures are those countermeasures that either confuse or destroy the attacker's guidance system. Flares, decoys, modulation jammers, and directed energy beams are all considered active countermeasures.

The *infrared flare* is one of the most common countermeasures used to reduce the effectiveness of air-to-air missiles. If the radiant intensity of the ejected flare is slightly greater than that of the intended target and its trajectory is not a too radical departure from that of the intended target, the missile will tend to follow the flare and not the aircraft. The intent is to use a

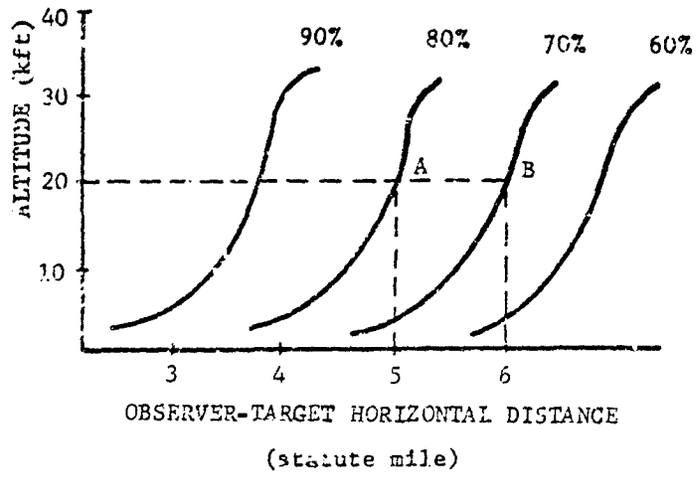


FIGURE 5-5. Estimate of Probability of Cloud-free Line-of-sight from (or to) Surface Level (Washington, D.C.--Summer)

sufficient number of flares deployed at the right time so that when the last flare has burned out, the aircraft is out of the missile's field of view. The missile is then rendered harmless because the remaining time is insufficient for reacquiring the target. It should be realized that since the flare is significantly smaller than the target, it must be substantially hotter in order to emit as much or more energy in any given spectral band. As a result, if the air-to-air missile were to have a two-color tracker (a counter-countermeasure) it could discriminate against the flare and reject it. This, of course, would force the defense into the deployment of a two-color flare (a counter-counter-countermeasure), and so on.

Decoys are objects that simulate the spectral and spatial characteristics of the target but are dispersed in such large numbers that it becomes uneconomical for the attacker to fire at all of them in an attempt to hit the real target(s). As in the case of the flare, the attacker may try to find a discriminant which the decoy does not simulate and use this to reject them in favor of the true target.

Modulation jamming is a technique whereby flashing or blinking lights are used to introduce spurious tracking signals into the missile tracker in an attempt to either completely confuse it or to introduce a significantly large miss-distance so as to render it ineffective.

Potentially, lasers (*directed energy beam*) could be used as a defensive technique to destroy either the optical coatings, filter, or the sensing element of the threatening IR seeker. Achievement of these objectives requires high-power levels not always compatible with airborne platforms. In addition, efficient use of these lasers requires precise pointing mechanisms.

5-3 BASELINE DESIGN CONCEPT

The baseline design concept is a technique used to focus the analysis once a firm set of performance requirements have to be established. The sensor system is then optimized by performing hardware and sensor parametric trade-offs about the baseline design point to determine the most economical way to achieve the performance requirements. It is almost always premature to establish a baseline design concept prior to the firm delineation of the performance requirements: however, it must be remembered that cost and schedule constraints may require the revision of the performance specifications if the trade-off analyses of the baseline design and the alternative concepts indicate that the specifications cannot be satisfied. Thus, the baseline design is part of the repeated cost effectiveness analysis.

5-3.1 SPECTRAL OPTIMIZATION

The selection of the optimum spectral band and the choice of detector are the most critical decisions made in designing an infrared sensor. The optimum spectral region must be selected in the context of the operational mission, the targets, background, and transmission characteristics; and size, weight, power, and cost constraints imposed on the system. The two dominant factors involved in spectral selection are system sensitivity and background rejection. Both of these factors affect the spectral location and the spectral bandwidth, and are always in opposition. The maximum target-to-background ratio is achieved by selecting an infinitesimally small bandwidth at the spectral location of maximum target-to-background contrast as shown in Fig. 5-6. However, since the bandwidth is infinitesimally small, the signal-to-noise ratio is essentially zero. In order to increase the signal-to-system-noise ratio, the spectral band must be widened which reduces the signal-to-background ratio since the background is increasing faster than the target as the optimum point is the center wavelength. The resolution of this problem is generally found in minimizing the number of detectors required for a series of fixed-aperture diameters, then developing a philosophy of making aperture/detector trade-offs.

5-3.1.1 Trade-offs

The instantaneous field of view and the detector material, in addition to sensitivity requirements, must be considered in selecting the optimum spectral region. The optimum spectral region can be defined as the one that maximizes the signal-to-background-plus-noise ratio ρ_β and simultaneously satisfies the minimum signal-to-noise constraint ρ_N. These can be defined by the following expressions

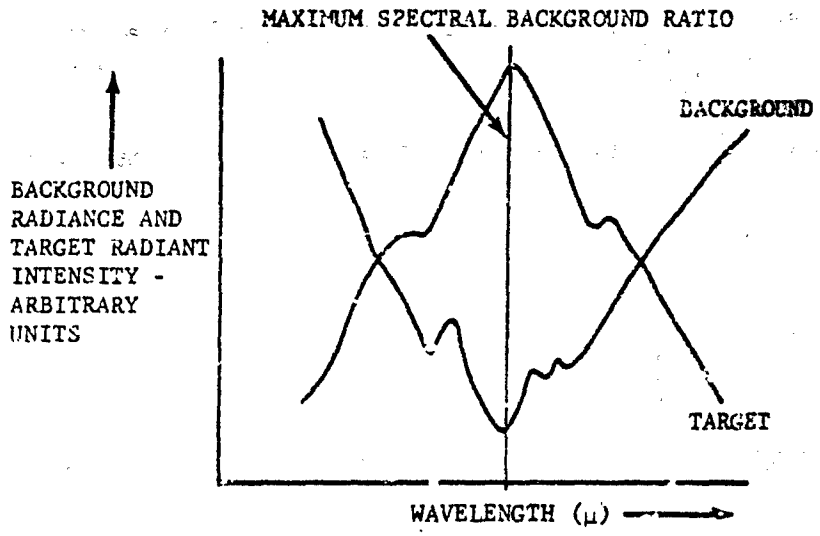


FIGURE 5-6. Example Spectral Background and Target Radiant Intensity

$$\rho_\beta = \frac{\frac{1}{R^2}\int_{\lambda_1}^{\lambda_2} J(\lambda)\tau(\lambda)d\lambda}{\left[\left(\alpha\int_{\lambda_1}^{\lambda_2} L(\lambda)\tau_B(\lambda)d\lambda\right)^2 + N^2\right]^{1/2}} \qquad (5\text{-}25)$$

and

$$\rho_N \leq \frac{\frac{1}{R^2}\int_{\lambda_1}^{\lambda_2} J(\lambda)\tau(\lambda)d\lambda}{N} \qquad (5\text{-}26)$$

where

ρ_β = signal-to-background plus noise ratio

ρ_N = signal-to-noise ratio $\frac{S}{N}$

$J(\lambda)$ = spectral radiant intensity of target, w $sr^{-1}\,\mu^{-1}$

$\tau(\lambda)$ = atmospheric transmission for target

$\tau_B(\lambda)$ = atmospheric transmission for background

R = range, cm

α = instantaneous field of view, sr

$B(\lambda)$ = background spectral radiance, w $sr^{-1}\,\mu^{-1}$

N = rms system noise, w cm^{-2} [Eq. 5-4]

Eqs. 5-25 and 5-26 are based on the target, background, transmission spectral characteristics, and the detector noise characteristics.

Assume the target and background radiance and the transmission coefficient to be as shown in Figs. 5-7 and 5-8, and a mission detection range of 7 km in a temperate atmosphere. Also assume the field of view of the sensor to be very small so that all backgrounds will be at a range equal to or greater than the target. The apparent target and background (as seen through the atmosphere) will then be as shown in Fig. 5-9. It is apparent that transmission has significantly altered the spectral characteristics of both target and background.

Next, assume an instantaneous field of view of 10^{-6} sr selected on the basis of resolution and number of detectors. It should be noted that this is an extremely significant design selection, therefore, several iterations may be required to determine the optimum detector subtense. The cumulative target and background irradiation curves are shown in Fig. 5-10. These curves show the amount of energy up to the long-wavelength cutoff λ_2 on the abscissa. The plateau region is the result of the strong absorption at 4.4 microns. The signal-to-background-plus-noise ratio ρ_β is illustrated in Fig. 5-11 for several short-wavelength cutoffs λ_1. The curves with the

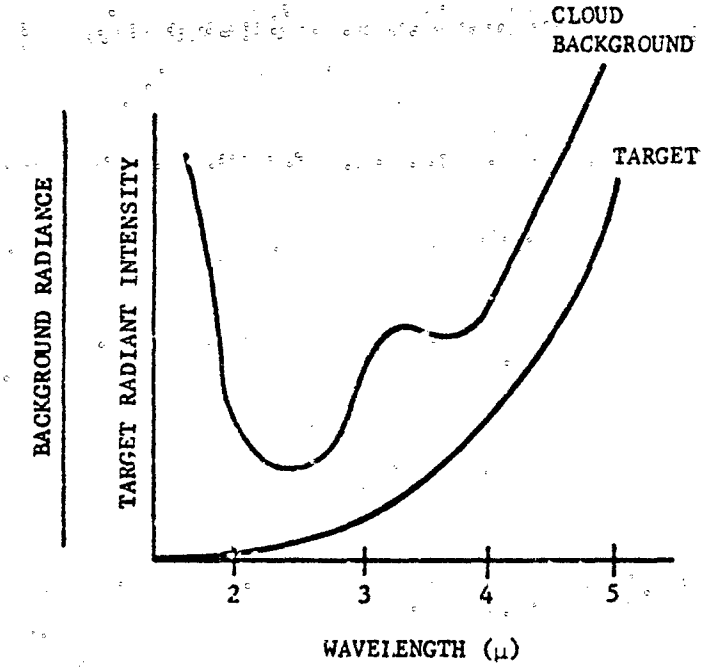


FIGURE 5-7. Sample Spectral Signature of Target and Background

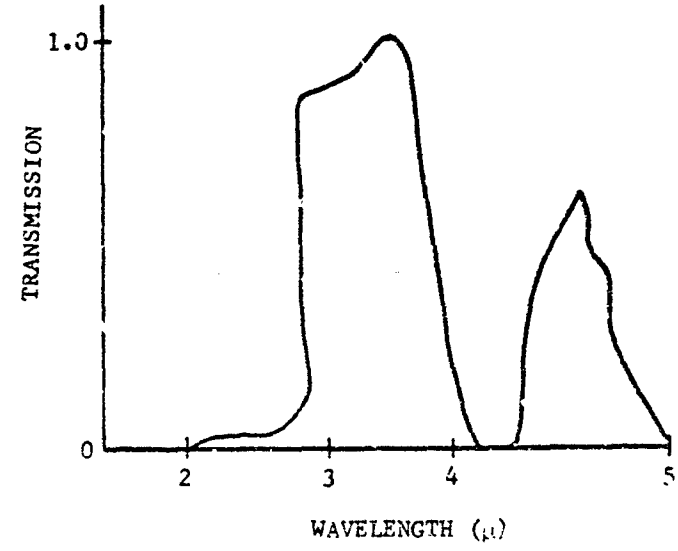


FIGURE 5-8. Spectral Transmission Curves

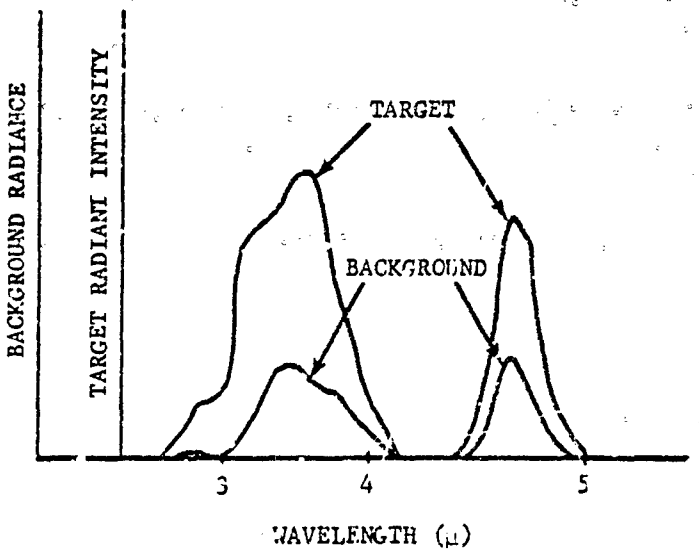


FIGURE 5-9. Apparent Target and Background Signatures

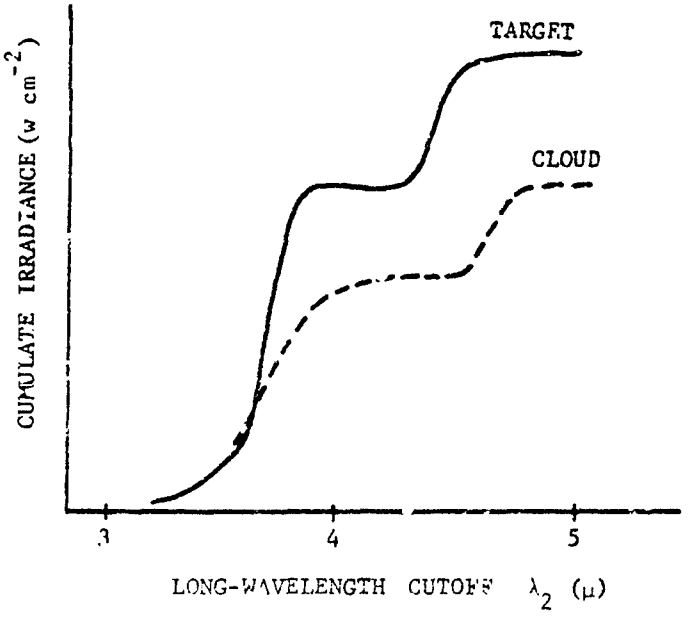


FIGURE 5-10. Target and Background Cumulative Irradiance vs Upper Wavelength Cutoff

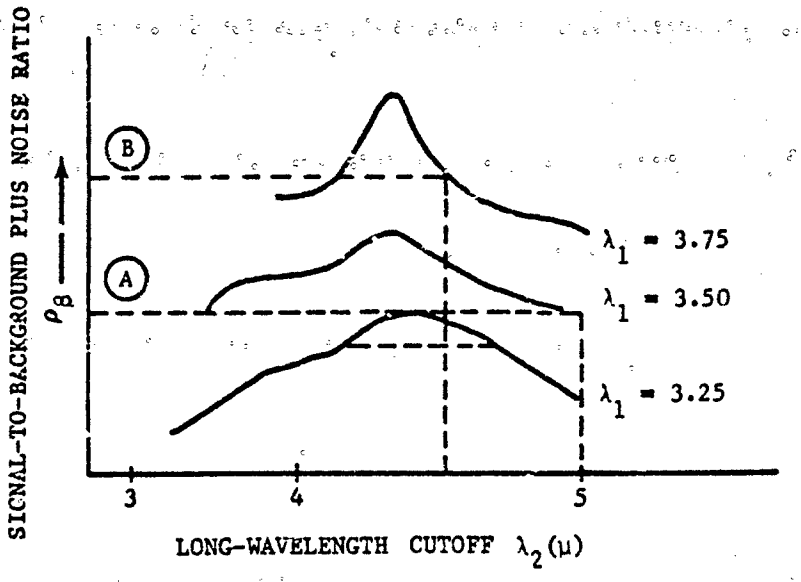


FIGURE 5-11. Signal-to-background Plus Noise Ratio for Various Upper and Lower Wavelength Cutoffs

larger values of ρ_β represent narrow spectral bands.

At this point the signal-to-noise constraint must be invoked. Assume that a minimum *S/N* ratio of 6 has been determined based on an analysis of Eqs. 5-11 and 5-14. Also keep in mind that size constraints limit the collecting aperture to some fixed maximum. In the spectral region of interest, the two potentially useful detector materials are PbSe and InSb. Fig. 5-12, which illustrates the cumulative *S/N* ratios ρ_N for these two detector materials, indicates that InSb is more sensitive than PbSe. For any given detector material, the optimum band is the most narrow band that satisfies the constraint of *S/N* = 6. This is found by iteratively selecting different combinations. Since InSb is more sensitive, it can satisfy the requirement with a narrower band than can PbSe (in this example, 3.75 to 4.50 microns vs 3.50 to 5.00 microns). The implications for the signal-to-background ratio are illustrated in Fig. 5-11. The system using PbSe would operate at position A of this graph while the system using InSb would operate at position B. It is apparent then that the most sensitive detector also provides the better background rejection capability. It should be noted that only a few aspects of the spectral region selection problem have been illustrated in this paragraph. For example, the trade-off of detector instantaneous field of view was not examined. However, even before this paragraph can be completed, a discussion on the final choice between detector materials must be included.

5-3.1.2 Detector Selection

The availability of detector materials for specific spectral regions, their physical properties, sensitivities, and cost are some of the more significant considerations in selecting a spectral region and the design of an infrared sensor. The most significant detector parameters which influence the choice of detector materials are spectral response, detectivity D^*, responsivity, cooling requirements, minimum size constraints, cell-to-cell uniformity, long-term stability, manufacturing tolerances, array technology, and costs.

The spectral response of most detector materials is relatively narrow in that their sensitivities outside of the limiting band are usually not satisfactory for weapon system applications. As a result, there are few examples in which more than two or possibly three detector materials are serious candidates for a specific application. Of these candidates, the second consideration is the

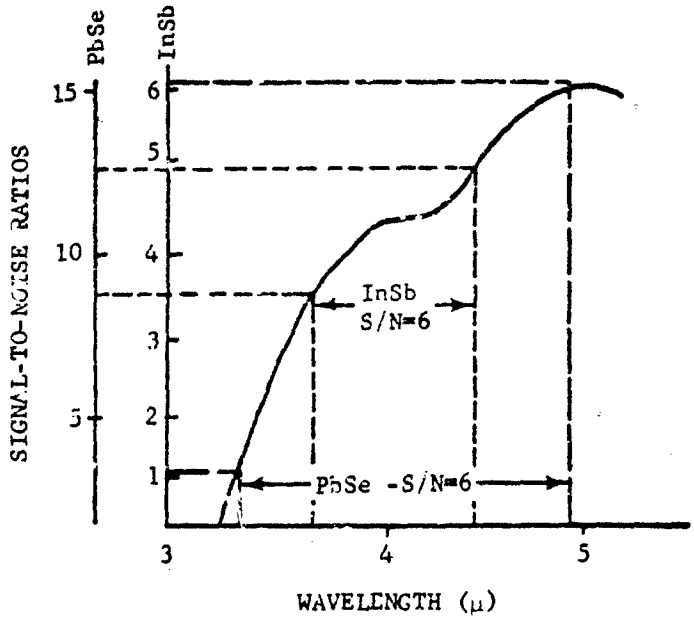


FIGURE 5-12. Cumulative Signal-to-noise Ratios for Two Detector Types

intrinsic sensitivity of the detector material under the system operating conditions. It is clear that detector sensitivity has a significant impact on both system size and system performance.

In order to achieve maximum sensitivity, most detector materials must be cooled to cryogenic temperatures. The electrical power required to provide this cooling is often the most significant portion of the sensor's power budget, and in many applications is often the parameter that determines the choice of detector material. For example, mercury-doped-germanium Ge:Hg can operate in the 3- to 5-micron region, but must be cooled to or below 30°K. On the other hand, indium-antimonide InSb also operates in the 3- to 5-micron region; however, it must be cooled to only 77°K. For equivalent heat inputs, the Ge:Hg system will require approximately three times the input power as will the InSb sensor, consequently, a substantial system trade-off could be made relative to size (due to cooling requirements) before selecting the detector material.

Detector cell size is also an important parameter in the selection of a detector material. It was shown that the instantaneous field of view of a detector was a significant parameter in eliminating background clutter. The detector area is related to the field of view by Eq. 5-8 and the linear dimensions X and Y are given by

$$X = D_o(f/no.)\theta_x \tag{5-27}$$

$$Y = D_o(f/no.)\theta_y \tag{5-28}$$

$$\alpha \approx \theta_X \theta_Y \text{ (sr)} \tag{5-29}$$

where θ_X, θ_Y are the detector's angular dimensions. The state-of-the-art usually limits the minimum physical dimension at which a given detector material can be manufactured. For some detectors this may be 0.001 in. while for others it may be 0.004 in. Thus, if very small instantaneous fields of view are required for background rejection, and the optical diameter D_o and optical speed $f/no.$ are limiting the package size, it may not be possible to use a particular detector material due to its size limitations.

When large arrays are required for specific applications, the existence of array technology and the uniformity of the detectors in the array also influence the choice of detector materials. For example, there are only two detector materials (PbS and Ge:Hg) for which large array (greater than 100 detectors) technology currently exists. The long-term stability of detector

sensitivity and responsivity also must be considered in the choice of a detector material.

5-3.2 FRAME-TIME OPTIMIZATION

In many cases the frame time can be derived and optimized based on target geometry and kinematic considerations. For example, consider the problem posed in par. 5-2.1.1 in which detection and acquisition are required by some range R_{min} as defined by Eq. 5-2. The optimum scan rate will be defined as the rate that minimizes the aperture diameter D_o while assuring target detection at a minimum range R_{min} which provides an acceptable signal-to-background-noise ratio S/B. As shown in Eq. 5-9

$$S/B = \frac{J\tau}{\alpha B R^2}$$

The following simplified relationship for atmospheric transmission τ_a is used to arrive at an analytic solution of the optimum frame time:

$$\tau_a = \tau\left(\frac{1}{R^x}\right) x \geq 0 \qquad (5\text{-}30)$$

which is an approximation for a negative exponential, where x is determined empirically in the spectral region of interest for the ranges of interest.

For many detector materials (e.g., PbS) the detectivity D^* term in Eq. 5-21 is a function of the frame time t_s as a result of its high-frequency time-constant roll-off past some knee frequency

$$D^* = \hat{D}^*\left(\frac{t_s}{t_o}\right)^Y \qquad \begin{matrix} Y \geq 0 \\ t_s \gg t_o \end{matrix} \qquad (5\text{-}31)$$

where

$\hat{D}^*$ = detectivity of the detector before roll-off†, w cm^{-2}

t_o = frame time that would produce the roll-off knee frequency (3 dB point) for a given size detector, sec

t_s = frame time, sec

Y = number which is empirically determined

† Frequency roll-off is defined as a frequency beyond which the output signal relative to input signal begins to be attenuated.

A linear array of detectors is assumed to cover the elevation field of view. The electrical (filter) bandwidth of the filter $\Delta f'$ is approximated by

$$\Delta f' = \frac{1}{2t_d}$$

where t_d is the time it takes a point image to cross a detector. If the detector is of width θ_X and the system is scanning at a rate ω rad/sec, then

$$t_d = \frac{\theta_X}{\omega} \qquad (5\text{-}32)$$

and

$$\Delta f' = \frac{\omega}{2\theta_X} \qquad (5\text{-}33)$$

If the field of view in the direction of scan is ϕ and the frame (scan) time is t_s, then

$$\omega = \frac{\phi}{t_s} \qquad (5\text{-}34)$$

and

$$\Delta f' = \frac{\phi}{2t_s\theta_X} \qquad (5\text{-}35)$$

Substituting Eqs. 5-30, 5-31, and 5-35 into *NEI* Eq. 5-21, and solving for S/N, we obtain

$$S/N = \frac{J\tau}{R^X R^2}\sqrt{\frac{2}{\phi}}\left[\frac{\pi D_o \epsilon_o \epsilon_e \hat{D}^* t_s^Y}{4(f/no.)t_o^Y}\right]\sqrt{\frac{t_s}{\vartheta_Y}} \qquad (5\text{-}36)$$

If the signal-to-noise ratio for a particular detector is to be constant, independent of range, it is required, therefore, from Eq. 5-9 that

$$\alpha = K_1 R^2 \qquad (5\text{-}37)$$

where K_1 is a constant multiplier.

For square detectors

$$\theta^2 = \alpha \qquad (5\text{-}38)$$

$$\theta_Y = \theta_X = \sqrt{\alpha} \qquad (5\text{-}39)$$

By substituting Eq. 5-39 into Eq. 5-36, the following relationship is obtained between D_o and R

$$D_o = K_2 \left(\frac{R^{2.5+X}}{t_s^{Y+0.5}} \right) \qquad (5\text{-}40)$$

where K_2 is another proportionality constant.

Range R is equal to the minimum range plus the distance which the target could have traveled in a frame time prior to detection, having just been missed during the previous frame (the target is assumed to be traveling away from the observer). We, therefore, have

$$R = R_{min} + n_i V_t t_s \qquad (5\text{-}41)$$

where

n_i = number of looks required for detection

V_t = target speed

t_s = scan time

Thus from Eq. 5-40

$$D_o = K_2 \frac{(R_{min} + n_i V_t t_s)^{2.5+X}}{t_s^{Y+0.5}}$$

Differentiating with respect to t_s and setting to zero yields the following optimum value of t_s

$$t_s(opt) = \frac{(0.5 + Y) R_{min}}{n_i V_t [(2.5 + X) - (0.5 + Y)]} \qquad (5\text{-}42)$$

For example, assume that transmission is inversely proportional to range ($X = +1$) and that D^* is inversely proportional to the square root of the frame time ($Y = +0.5$). Also assume a minimum detection range R_{min} of 2 km, a target speed V_t of 0.3 km/sec (1000 ft/sec), and a two-look ($n = 2$) detection criterion.

$$t_s(opt) = \frac{(0.5 + 0.5)2}{2 \times 0.3 \; [(2.5 + 1) - (0.5 + 0.5)]}$$

$$t_s(opt) = 1.33 \text{ sec.}$$

It must be remembered that other system considerations have been neglected in this optimization. If cloud and terrain masking make it likely for the target to approach the defended point unobserved prior to the range dictated by $t_s(opt)$, then a faster frame-time may be required as illustrated by Eq. 5-28. Thus, the frame time is calculated as

$$t_s = \min [t_s(opt) \text{ or } t_s(max)] \qquad (5\text{-}43)$$

or the minimum of the maximum determined by environmental considerations and the optimum determined by sensor considerations.

5-3.3 PARAMETRIC DESIGN

The parametric relationships between sensor and mission variables determine, in a large measure, the trade-off opportunities available to the system designer. As a result, a perceptive understanding of the basic mathematical and physical relationships is a mandatory requirement for the system analyst. The relationships between optical size, number of detectors, cooling requirements, scan methods, displays, and other system considerations are investigated in the paragraphs which follow.

5-3.3.1 Collecting Aperture Parameter

The size of the collecting aperture D_o is a dominant variable in determining the weight, size, and volume of the entire IR sensor unit. As a consequence, it is the system parameter that often requires the most attention.

Eq. 5-4 is the basic system equation for a detector-noise-limited system and the geometry for a simple optical system is illustrated in Fig. 5-13.

The detector area A_d is, of course, the product of its dimensions in height X and width Y (there is no relationship between these X and Y, and those shown in Eqs. 5-40 and 5-41)

$$A_d = XY, \text{ cm}^2 \qquad (5\text{-}44)$$

From Fig. 5-13 it is apparent that the angular subtense of the detector can be given as

$$\theta_X = \frac{X}{f}, \text{ rad}$$

$$\theta_Y = \frac{Y}{f}, \text{ rad}$$

The instantaneous field-of-view α of the detector is the angular volume of the detector subtense and is given by Eq. 5-29 as

$$\alpha \approx \theta_X \theta_Y, \text{ sr}$$

As discussed previously, it is frequently assumed that the electrical bandwidth should be the inverse of twice the time it takes a point

image to traverse the detector in the direction of scan, or

$$\Delta f = \frac{1}{2t_d}, \text{ Hz [Eq. 5-7]}$$

where

t_d = dwell time, sec

Assume the total volume of Ω steradians is to be searched in time t_s by n_d detectors, then

$$t_d = \frac{n_d t_s \alpha}{\Omega} \quad \text{[Eq. 5-6]}$$

Substituting Eqs. 5-6 and 5-7 into Eq. 5-21, yields the system sensitivity equation

$$NEI = \frac{4\alpha(f/no.)}{\pi\epsilon_o\epsilon_e D_o D^*}\sqrt{\frac{\Omega}{2n_d t_s}} \qquad (5\text{-}45)$$

The required system sensitivity is determined by the radiant intensity J (w sr^{-1}) of the target, the atmospheric transmission τ_a, the detection range R (cm), and the desired signal-to-noise ratio

where

$$S/N = \frac{J\tau_a}{R^2 NEI} \qquad (5\text{-}46)$$

or

$$NEI = \frac{J\tau_a}{R^2(S/N)} \qquad (5\text{-}47)$$

Eqs. 5-4, 5-21, 5-45, and 5-47 can be combined to yield three alternative methods for calculating the optical diameter

$$D_o = \left\{ \begin{array}{ll} \left[\frac{R^2(S/N)}{J\tau_a} \cdot \frac{4\sqrt{A_d \Delta f}}{\pi\epsilon_o\epsilon_e D^*}\right]^{1/2} & \text{(a)} \\ \left[\frac{R^2(S/N)}{J\tau_a} \cdot \frac{4(f/no.)\sqrt{\alpha\Delta f}}{\pi\epsilon_o\epsilon_e D^*}\right] & \text{(b)} \\ \left[\frac{R^2(S/N)}{J\tau_a} \cdot \frac{4(f/no.)}{\pi\epsilon_o\epsilon_e D^*}\sqrt{\frac{\Omega}{2n_d t_s}}\right] & \text{(c)} \end{array} \right\} \qquad (5\text{-}48)$$

Typically for search/track systems

$$\epsilon_o\epsilon_e = 0.2$$

$$f/no. \geq 1.5$$

$$S/N \geq 7$$

The following two additional inequalities are required to completely specify the optical diameter D_o:

$$D_o \geq \frac{d_{min}}{(f/no.)\,\theta_X} \qquad (5\text{-}49)$$

$$D_o \geq \frac{2.44\lambda_c}{\theta_X}, (\lambda_c \text{ in cm}) \qquad (5\text{-}50)$$

The same expressions can be written for θ_Y. The first inequality refers to the fact that detectors can be made only so small (d_{min}) after which either costs or physical constraints render

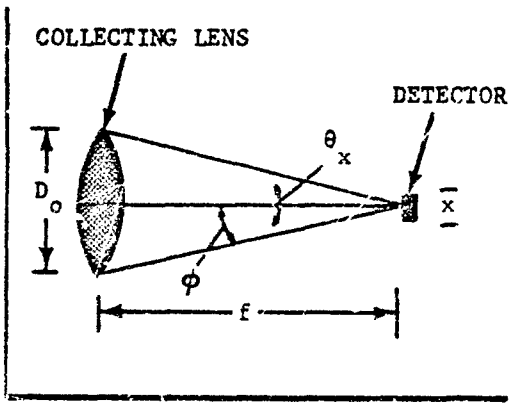


FIGURE 5-13. Optical Diagram

impossible further size reductions. The second inequality refers to the Rayleigh diffraction criterion in which the long-wavelength cutoff λ_c and the optical diameter determine the minimum spot-size for a point object.

An example of the type of parametric data that can be generated by means of Eqs. 5-48, 5-49, and 5-50 is illustrated in Fig. 5-14. The curves illustrate that for any given instantaneous field of view α, the diameter increases as the square of the required range. Fig. 5-14 illustrates a situation where the diffraction criterion λ_c determines the minimum detector size criterion d_{min} . A maximum diameter constraint $D_o(max)$ is imposed to reflect the possibility of limited size or volume available to the sensor. The signal-to-background-ratio line S/B, Fig. 5-14, also places a limit on the possible combination of α and R. As a consequence of the underlying system equations and the severe physical and environmental constraints placed on the system, only the small region shown in the graph remains to satisfy all the restrictions and limitations.

5-3.3.2 Number of Detectors

In many cases the number of detectors is the dominant variable in determining the sensor cost. The reason for this is the fact that each detector element requires its own separate processing network including preamplifier, filter, thresholding, and other electronic processing functions. Each detector processing channel currently costs in the neighborhood of $100 to $500, thus, a system with 1000 detector channels would cost at least $100,000. On the other hand, if the number of detectors is too small, the resultant false alarm rate and loss of resolution may require additional computer processing which, in the long run, may cost more than adding additional detector channels.

Three basic factors that determine the detector size and detector number are background level, sensitivity, and resolution. Most natural background clutter, with the exception of stellar sources, tends to be extended in nature and is thus larger than the detector element. In these cases, the apparent background irradiance H_B is expressed as

$$H_B = \alpha\tau_B B + \alpha(1 - \tau_B)B_a \qquad (5\text{-}51)$$

where

- τ_B = transmission of the atmospheric path between the optical system and the background
- B = radiance of the background
- B_a = radiance of the intervening atmosphere at ambient temperature

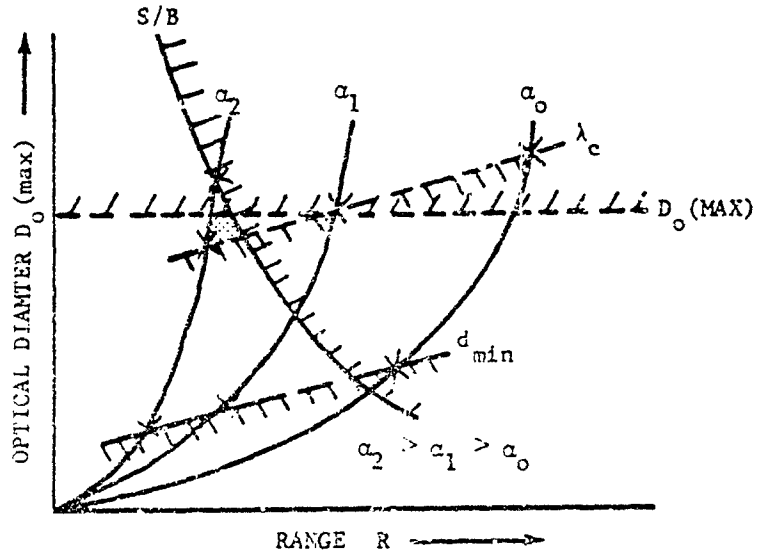


FIGURE 5-14. Constraints on System Aperture Selection

In most cases B_a is assumed to be negligible and thus

$$H_B = \alpha \tau_B B \tag{5-52}$$

It is apparent that the background radiance can be made as small as desired by reducing the instantaneous field of view α if the background is composed of extended sources. However, several limiting conditions exist. First, even if assumed to be a point source, the target has in reality some finite area A_t. Thus, once the instantaneous field of view α begins to approach the target subtense, further reductions will reduce both target and background with no resultant gain in the signal-to-background ratio. Thus the limiting field of view α_{min} is

$$\alpha_{min} \geq \frac{A_t}{R^2}, \text{ sr} \tag{5-53}$$

The qualitative degree of background clutter rejection capability is often expressed in terms of the signal-to-background ratio

$$S/B = \frac{J\tau_a}{\alpha R^2 \tau_B B} \quad \text{if } \alpha \geq \frac{A_t}{R^2} \tag{5-54}$$

Consider, for example, the problem of detecting a helicopter ($A_t = 10$ m^2) at a range of 5 km. Assume that the target radiates 3 w sr^{-1} in the 3- to 5-micron region, the atmospheric transmission for both target and background is 0.33, the background radiance is 10^{-4} w cm^{-2} sr^{-1}, and that a signal-to-background ratio of 3:1 is desired. By Eq. 5-54

$$\alpha = \frac{3 \times 0.33}{3 \times (5 \times 10^5)^2 \times 0.33 \times 10^{-4}}$$

$$= 4 \times 10^{-8} \text{ sr}$$

However, the limiting field of view is expressed as (Eq. 5-53)

$$\alpha_{min} = \frac{10}{(5 \times 10^3)^2}$$

$$\alpha_{min} = 4 \times 10^{-7} \text{ sr}$$

Thus, the maximum achievable signal-to-background ratio is only 0.3 since further reductions in the instantaneous field of view reduces the signal and background energy equally. If, however, the effective target area were only 1 m^2, then the desired S/B ratio could be achieved with an instantaneous field of view 4×10^{-8} sr.

Consider, next, a rectangular scan field of dimension $\theta \times \phi$ in which θ is the angular field of view in the direction normal to the scan direction and ϕ is the angular field of view along the direction of scan. A single linear array used for scanning the entire field would have to be comprised of the number of detectors n_d calculated as follows:

$$n_d = \frac{\theta}{\sqrt{f_s \alpha}} \tag{5-55}$$

where f_s is the detector aspect ratio (height-to-width)*. In order to search the scan field by n_b bars or a linear array, the following number of detectors will be required

$$n_d = \frac{\theta}{n_b \sqrt{f_s \alpha}} \tag{5-56}$$

As shown in Eq. 5-21, system sensitivity is proportional to the square root of the product of the detector field of view and the equivalent noise bandwidth; i.e.,

$$NEI \propto \sqrt{\Delta f \alpha} \tag{5-57}$$

but by Eq. 5-7

$$\Delta f = \frac{1}{2t_d}$$

$$t_d = \frac{\sqrt{\alpha / f_s}}{r_s}$$

where r_s is the scan rate, rad/sec.

Therefore

$$NEI \propto \sqrt{r_s \sqrt{\alpha f_s}}$$

Thus, from number of detectors, it is highly desirable that the detectors be long and thin; however, this is undesirable from a sensitivity standpoint. The conflict must be resolved by examining the sensitivity and cost constraints and making the required trade-off decision, as illustrated in Fig. 5-15. The mission requirements establish some maximum noise level NEI_o that will just satisfy the detection requirement.

* The detector height is assumed to be in the θ direction.

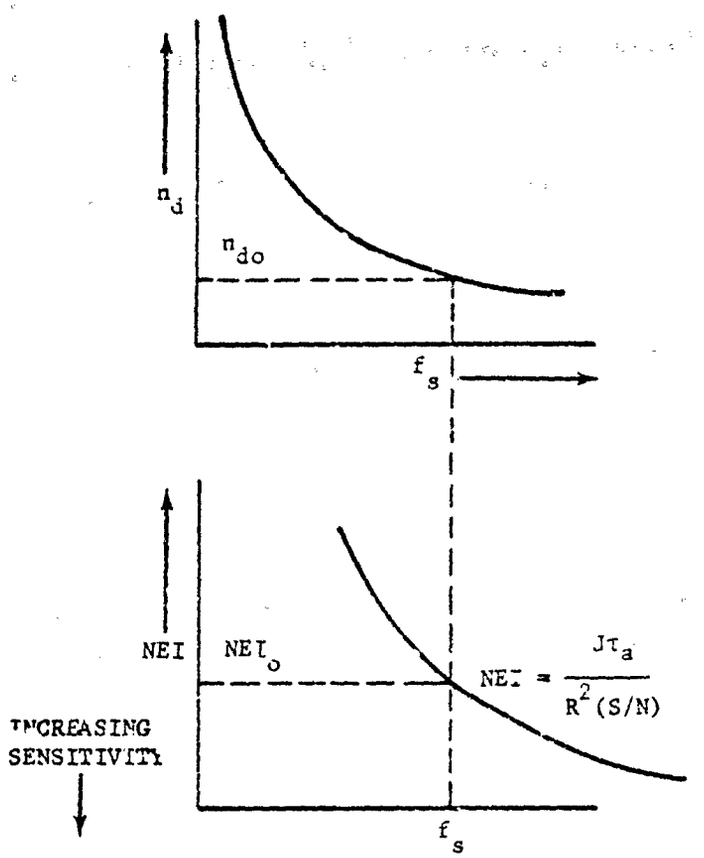


FIGURE 5-15. Sensitivity of Detectors and Number Required vs Detector Aspect Ratio

This, in turn, determines the maximum allowable aspect ratio, thereby establishing the required number of detectors n_{do}.

Detector overlap is another factor to be considered in determining the number of detectors. Because of the finite size of the optical image, a certain amount of overlap between adjacent detectors is required to prevent targets from passing across the focal plane without all of its energy having irradiated at least one detector (see Fig. 5-16). For a detection system, an overlap equal to the size of one image quality, i.e., the diameter of a circle which contains about 90 percent of the target energy, is usually recommended. In this case, the number of detectors is expressed as

$$n_d = \frac{\theta}{n_b\left[\sqrt{f_s\alpha} - 2.44\lambda/D_o\right]} \qquad (5\text{-}58)$$

The requirement for one full image quality is usually ignored when the detector size approaches the diffraction spot size.

Eq. 5-48 illustrates the fact that for any given set of conditions, the required optical diameter is inversely proportional to the square root of the number of detectors.

$$D_o \propto \sqrt{\frac{1}{n_d}} \qquad (5\text{-}59)$$

Thus, in addition to being an effective background-clutter rejection technique, the use of small detectors in large arrays also is a significant factor in increasing sensitivity, thus yielding longer detection ranges for the same equipment size. In making the final determination regarding the number of detectors, cost and minimal detector size must always be considered

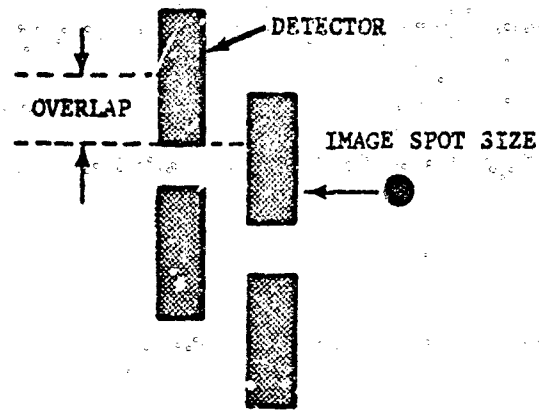


FIGURE 5-16. Detector Overlap

5-3.3.3 Cooling Requirements

Most detectors must be cooled to cryogenic temperatures in order to achieve maximum sensitivity. In addition, detector materials which are background photon noise limited, BLIP, often require cooling of the entire optical system to achieve the maximum sensitivity. The electrical power necessary to provide this cooling is often the most significant portion of the sensor's power budget. For open-loop systems operating with a fixed supply of cryogen, the cooling requirement can practically omit the operating life of the sensor. As a consequence, the cooling requirement often dictates the choice of detector materials.

For example, Indium antimonide InSb as well as mercury-doped germanium Ge:Hg can be used in imaging systems that provide thermal maps of the surrounding environment. A Ge:Hg system is approximately twice as sensitive as one using InSb and thus could be considerably smaller. However, Ge:Hg must be cooled to 28°K (liquid helium) while InSb operates at 77°K (liquid nitrogen). For equal heat inputs, the Ge:Hg system would require approximately 3 to 4 times more cooling power than does the InSb system which requires approximately 300 w of cooling power for a 1-w heat input. If the sensor is to operate from a large aircraft, the difference between 300 w and 900 to 1200 w may not outweigh the advantages of the smaller package which is possible with Ge:Hg. If the system is to operate on battery power from an armored personnel carrier with its engine turned off, the power difference is very often more significant than the size and weight difference. It is thus clear that the cooling requirements can affect the entire system design significantly, depending on the application. The following data illustrate the cooling requirements and spectral range of the more popular detector materials:

Material	Temperature	Spectral Range, μ
PbS	+25° to -140°C	2 – 3.5
PbSe	+25° to -140°C	2 – 5
InSb	77°K	2 – 5
Ge:Hg	30°K	2 – 14
Ge:Cu	10°K	2 – 25

It is apparent from these data that the long-wavelength response increases as the cooling requirement becomes more stringent.

Both mercury- and copper-doped germanium detectors are BLIP and thus their sensitivities improve inversely as the square root of the incident photon noise (assuming negligible pre-amplifier noise).

$$D^* = \frac{\lambda}{2hc}\sqrt{\frac{\eta}{Q_B}} \quad , \text{cm Hz}^{1/2}\,\text{w}^{-1} \tag{5-60}$$

where

λ = wavelength, cm

h = Planck's constant 6.625×10^{-34} w sec^2

c = speed of light 2.9979×10^{10} cm sec^{-1}

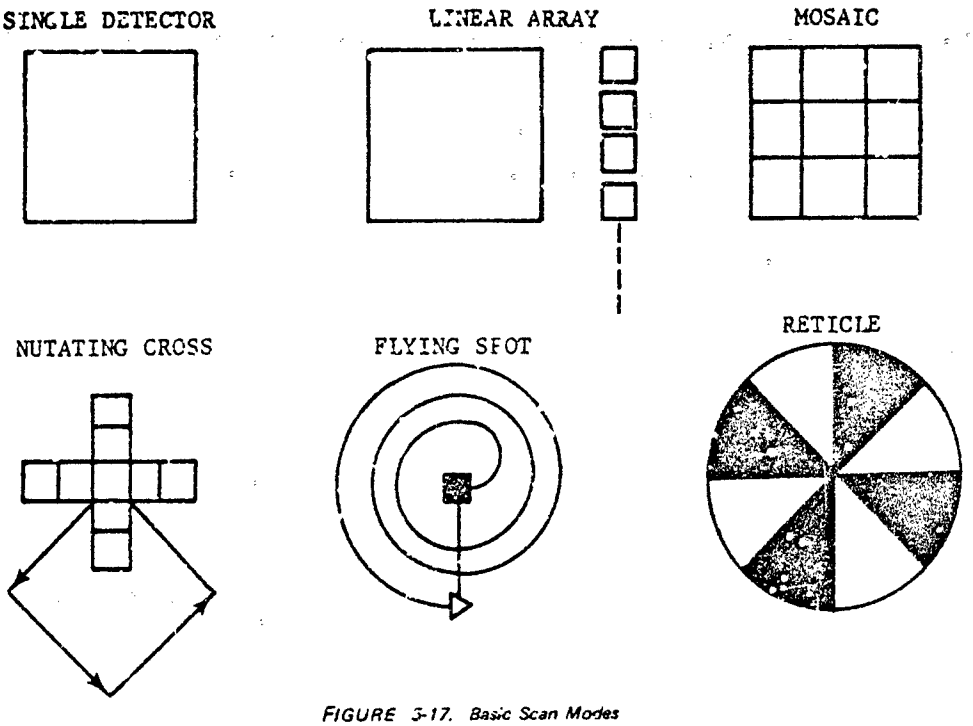


FIGURE 5-17. Basic Scan Modes

η = quantum efficiency of the detector

Q_B = incident photon flux density
= phot $cm^{-2}\,sec^{-1}$

In some applications (e. g., when the sensor is operating in a space environment against a space background) the photon flux from an uncooled optical system would be the dominant source of photon noise. As a result, when Ge:Hg or Ge:Cu systems are operated in such an environment, their optical systems are often cooled to temperatures approaching those of the detectors. Since at some point detectors cease being BLIP and become limited by some other noise source (e.g., preamplifier noise), the temperature to which the optical system is cooled must be carefully computed to avoid cooling beyond that which can affect the system sensitivity. As an example, consider the use of Ge:Hg in a thermal imaging system operating at sea level against 300°K ambient backgrounds. In this case, there is absolutely nothing to be gained by cooling the optical system since the external thermal background noise will dominate the internal thermal noise.

5-3.3.4 Scanning Systems

The scanning method to be used determines to a large extent the mechanical configuration of an infrared sensor. Six basic scan methods are illustrated in Fig. 5-17. Each has its own applications, advantages, and disadvantages. The method to be used must be selected on the basis of a critical examination of the particular problem at hand.

In the past, reticle scanning was used almost exclusively as the primary method of providing modulation for tracking systems and for background rejection in search/track systems. However, the advent of multi-element arrays has increasingly supplanted the reticle to the point where linear arrays are used in most advanced track, search, and imaging IR systems.

Where a *single detector* element is used, scanning is usually provided by means of a prism or mirror as is typical of most line-scanning imaging systems for reconnaissance. This type of scan is illustrated in Fig. 5-18. The scan speed of the rotating prism is synchronized with the velocity-to-height ratio (v/h) of the aircraft

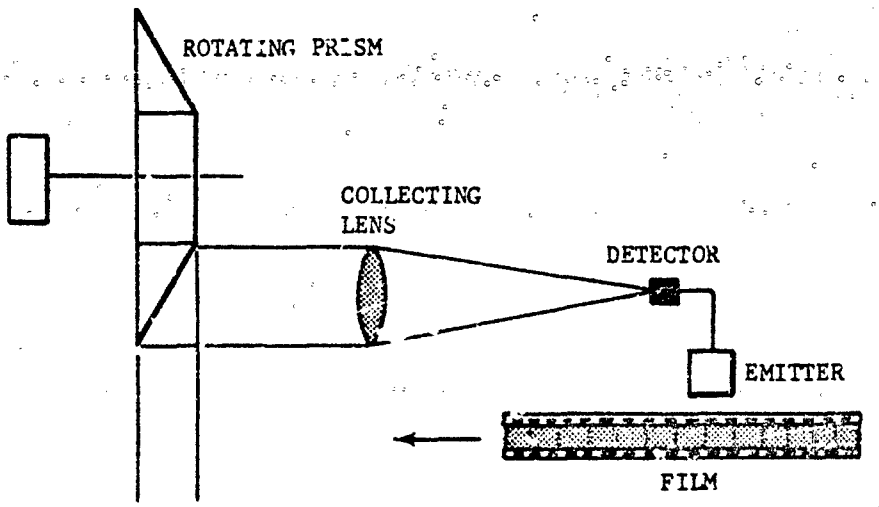


FIGURE 5-18. Line Scanning Technique

whereon the instrument is mounted, in order to provide continuous line coverage. The output of the detector modulates a small emitter whose output is recorded on film. Data processing entails placing the individual lines side by side to form a complete picture. As indicated by the simple diagram, the scanning prism is quite large relative to the collecting lens.

Linear arrays have the advantage of permitting significantly smaller mechanical packaging. For example, a linear array wherever the scanning motion is provided by two counter-rotating wedges as illustrated in Fig. 5-19, would be quite small and its output could be viewed in real time (i.e., directly on a CRT). Of course, the electrical complexity is considerably greater than that required for a simple-line scanner and, depending on the application, may or may not be warranted.

Numerous other scan patterns and scan techniques can be developed, the exact technique depends upon the application and upon cost and engineering trade-offs. Hybrid scans such as a spiral scan with a linear array, nutating mosaics, and others, all offer the system designer with an almost infinite number of possible techniques.

5-4 SAMPLE IR SYSTEM DESIGNS

5-4.1 MISSILE SEEKER

Infrared seekers are used to guide missiles by homing in on the IR emission from the target. IR seekers provide the guided missile's steering system with information on the target's line of sight, angular rate, and/or angular position relative to the missile velocity vector. The missile's steering system closes a feedback loop which controls the missile flight-path kinematics. The

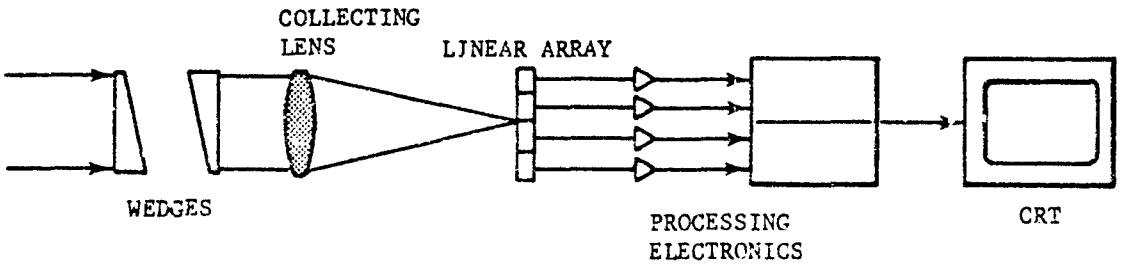


FIGURE 5-19. Scanning Image System

target's line of sight angular rate or position is kept sufficiently near zero or constant to ensure a hit (Fig. 5-20(A)). When the angular position of the line of sight is constant, the missile flight path describes a pursuit course. The angular position of zero describes a pure pursuit missile flight-path trajectory, and if a fixed angle is maintained by the missile's steering system between the missile's velocity vector and the line of sight to the target, a deviated pursuit course will result.

A constant bearing (collision) course (Fig. 5-20 (B)) exists when the line of sight from the missile to the target is constant. The proportional-navigation course is an approximation of the constant bearing course wherein the missile's steering system commands an angular rate of change of the missile velocity vector which is directly proportional to the target sight-line rate of change.

The guidance mode selected (pursuit, proportional navigation, etc.) will have a significant effect on the time history of the missile trajectory and on the requirements for maximum lateral acceleration, control system bandwidth, cost, and weight. The optimum factors regarding airframe, propulsion, warhead, control system, and guidance mode for the missile are determined by means of a preliminary design analysis and simulation of the complete weapon system. Specific missile seeker performance requirements are determined at that time. It will be assumed that, for the following missile seeker design example, such an overall weapon system study and computer simulation has resulted in the selection of a proportional-navigation seeker design. It is recognized that the example chosen—an air-to-air-missile—is not the U. S. Army's responsibility. However, this example is presented in order to adequately illustrate the theory previously presented.

This design example follows the procedure of establishing the mission objectives which yield the functional and performance requirements. A baseline design concept is formulated with the selection of fixed seeker characteristics.

5-4.1.1 Missile Seeker Requirements

5-4.1.1.1 Operational Environment

It is assumed that the seeker will be designed for use with an air-to-air missile such that target acquisition and identification can be performed by the pilot using visual means. The pilot's visual target identification range of 3 n mi determines the seeker's maximum target acquisition range. The warhead safe-arming distance of 1/3 n mi from launch aircraft determines the minimum range.

System simulation studies involving the launch aircraft and target are assumed to have resulted in the following operational requirements:

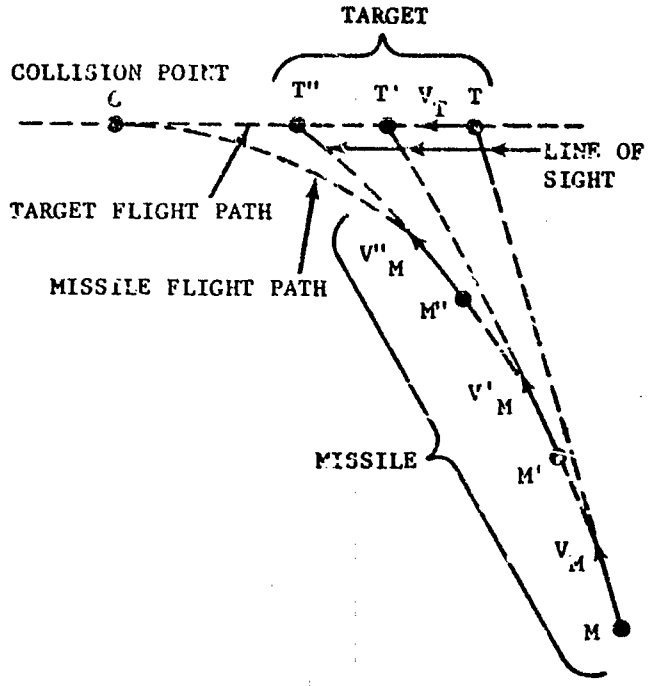
Gimbal angle freedom	±40 deg
IR dome (max dia)	3.5 in.
Line of sight tracking rate (max)	20 deg/sec
Gimbal drift rate (max)	0.2 deg/sec
Target acquisition time (max)	0.5 sec
Target acquisition range (max)	3 n mi
Target acquisition range (min)	1/3 n mi
Target tracking accuracy	1 mrad
Probability of target acquisition	0.9
Minimum operational time	2 hr
Target maximum maneuverability	6 g at 1000 fps

The pilot will acquire the target by aiming his aircraft at the target using a fixed gunsight reticle. His aiming error, combined with the aircraft structural deflection between his sight and the seeker, will be assumed to represent a circular normal distribution having a CEP of 12 mrad (a standard deviation of 10 mrad in each of two orthogonal directions).

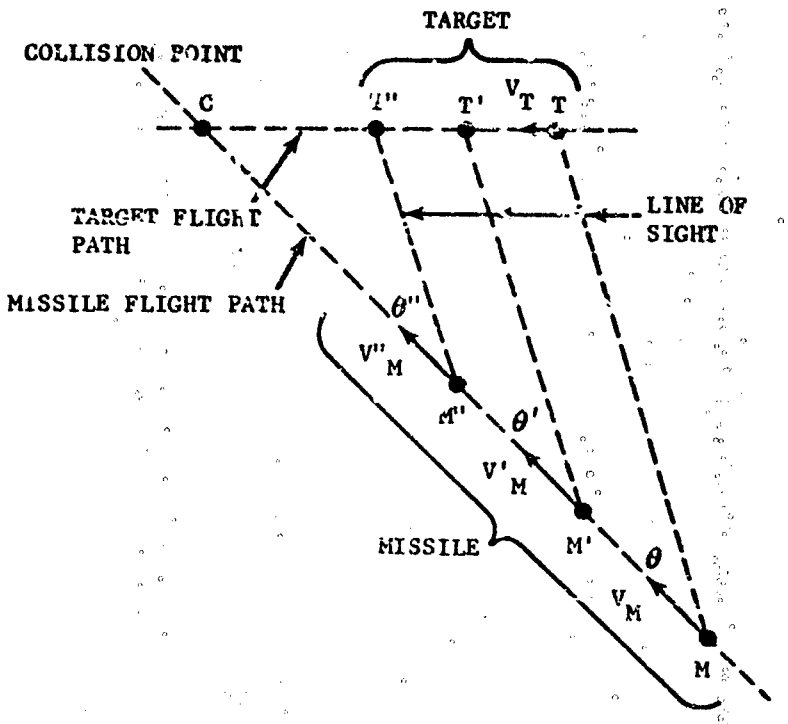
For purposes of this example, the missile will be pylon mounted to the aircraft's wing. It will be exposed to altitudes ranging from sea level to 50,000 ft at aircraft flight velocities of Mach 2 maximum under all weather conditions.

5-4.1.1.2 Seeker Functions

The pilot will set the seeker in the "ready" mode which will result in the detector being cooled down to its operating temperature. Upon acquiring and positioning a target within his fixed-sight aiming circle, the pilot will manually select the seeker target-acquisition mode. This will cause the seeker to: (1) search the required acquisition field of view, (2) lock on and track a



(A) PURSUIT GUIDANCE (IDEALIZED)



(B) COLLISION GUIDANCE (IDEALIZED)

FIGURE 5-20. Typical Missile Guidance Geometries

target in the search field of view, and (3) present the pilot with a visual or audible signal indication of the seeker tracking a target. The pilot can then launch the missile. If target acquisition does not occur or if the pilot decides not to launch the missile, the pilot will restore the seeker to the "ready" mode which will place the seeker in the boresight position.

5-4.1.1.3 Functional Description

A generalized seeker block diagram which illustrates the various functions of a seeker such as described is shown in Fig. 5-1. As shown, the pilot would have command of the aircraft fire control system, allowing him to: control (1) cooling of the detectors, and (2) activation of scanning, signal processing, and seeker positioning functions. In addition, the aircraft fire control system would provide him with an indication of target lock-on.

The seeker control loop makes it possible to conduct a programmed search of the acquisition field of view, and to detect and track a target. A typical seeker servo block diagram is shown in Fig. 5-21. (Also see par. 3-9.1 for a discussion of servo systems.) Proportional navigation is achieved by closing the missile control loop around the angular rate of target line of sight. If the seeker is provided with angular inertial stabilization, the vehicle attitude oscillations will not affect the detected target angular position. Such stabilization may be accomplished by either a seeker design comprising a free gyro, or a design using seeker-head-mounted rate gyros in conjunction with a high-gain servo loop.

5-4.1.1.4 Apparent Radiant Intensity of Targets and Backgrounds

Apparent spectral radiant intensity $J_a(\lambda)$ data [$J_a(\lambda) = J(\lambda)\tau(\lambda)$], representing the maximum range target under the most severe atmospheric and target signature conditions, are generally available to the seeker designer in the form shown in Fig. 5-22. Similar apparent target radiant intensity plots are obtained from measurements which have been extrapolated to the desired target range conditions by combining them with measured and calculated atmospheric transmission data.

The peak apparent background radiance gradients at an angular subtense similar to the target and in the spectral bandpass of interest is assumed to be 7.5×10^{-5} w cm^{-2} sr^{-1} μ^{-1}.

As discussed in this handbook, background levels should be treated, properly, as nonstationary random processes. In practice, conservatism, and often lack of detailed information concerning a realistic background model, leads to the assumption of a model wherein the variation in apparent background radiance is the maximum for a given spectral region. As will be seen, the seeker must be designed to minimize the probability of false acquisition on backgrounds of this intensity.

5-4.1.2 Baseline Design Concept

By use of the values selected thus far, the target irradiance H_t in the 4.3- to 4.7-micron spectral passband may be calculated as

$$H_t = \int_{\lambda_1}^{\lambda_2} \frac{J(\lambda)\tau(\lambda, R)}{R^2}\, d\lambda = \int_{\lambda_1}^{\lambda_2} \frac{J_a(\lambda)}{R^2}\, d\lambda, \text{ w cm}^{-2} \qquad (5\text{-}61)$$

where

$J(\lambda)$ = spectral radiant intensity from target, w sr^{-1} μ^{-1}

$J_a(\lambda)$ = apparent spectral radiant intensity from target, w sr^{-1} μ^{-1}

$\tau(\lambda, R)$ = atmospheric transmission as a function of wavelength and pathlength

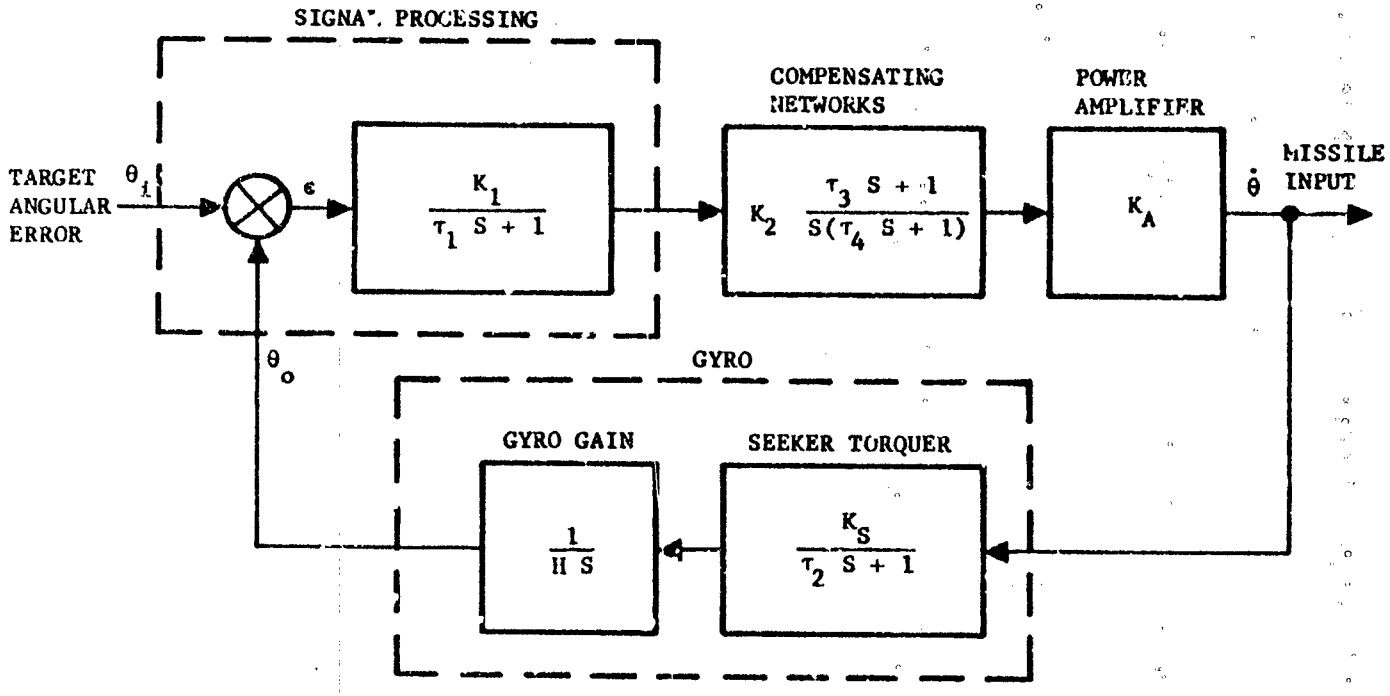
R = range, cm

λ_1 and λ_2 = target spectral cutoffs (4.3 and 4.7 microns, respectively), cm

Integrating $J_a(\lambda)$ over the 4.3- to 4.7-micron spectral passband, as shown in Fig. 5-22, gives 19 w/sr as the target apparent radiant intensity. At a range of 3 n mi, the target irradiance is

$$H_t = \frac{19}{(3 \times 1.85 \times 10^5)^2} = \frac{19}{30.8 \times 10^{10}} = 6 \times 10^{-11} \text{ w cm}^{-2}$$

($3 \times 1.35 \times 10^5$ cm = 1 n mi). The background irradiance level H_B is calculated as



WHERE $K_1 K_2 K_A K_S \frac{1}{H} = K_a$ ACCELERATION CONSTANT

FIGURE 5-21. Seeker Servo Block Diagram

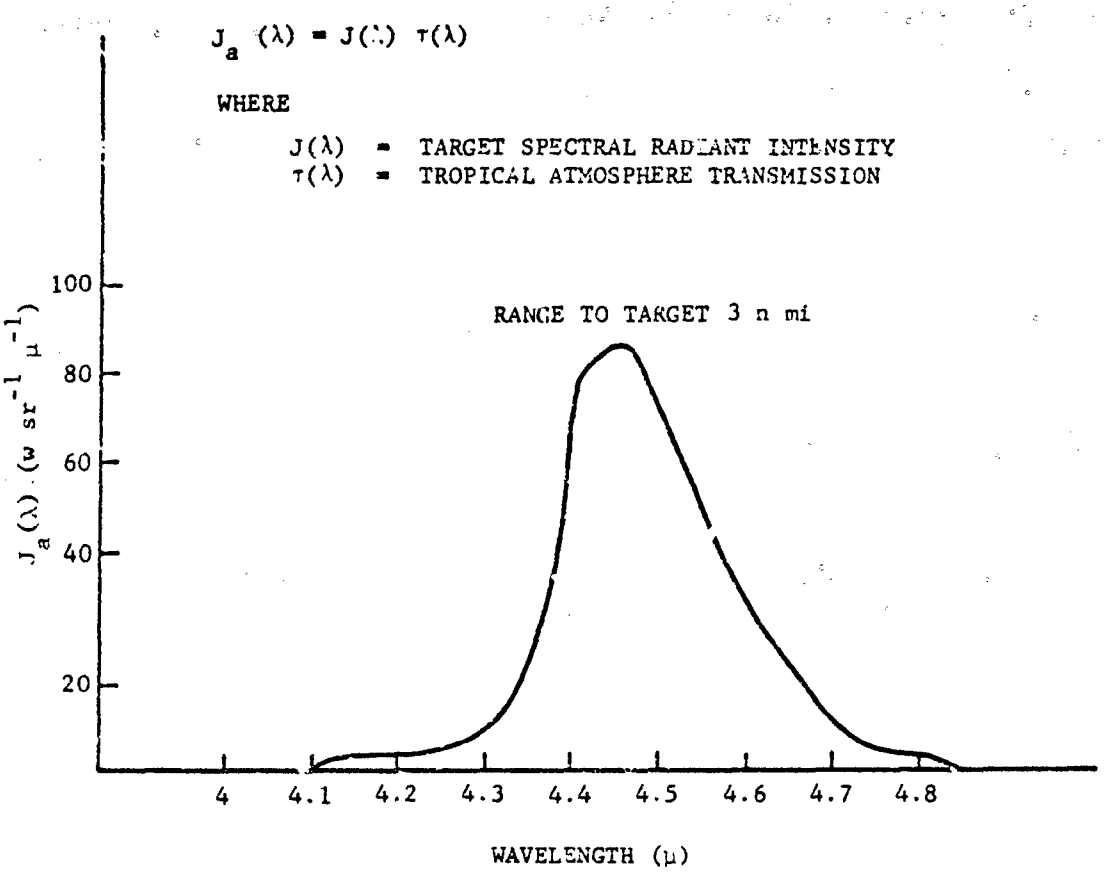


FIGURE 5-22. Apparent Target Intensity

$$H_B = H_c(\lambda)\Delta\lambda\omega_\alpha \qquad (5\text{-}62)$$

where

H_B = total background radiant intensity, w cm^{-2}

$H_c(\lambda)$ = background radiant intensity, w cm^{-2} μ^{-1} sr^{-1}

$\Delta\lambda$ = spectral bandpass, micron

ω_α = angular subtense of detector, sr

A target-to-background ratio of greater than 4 is required in order to minimize the likelihood of the seeker locking onto background gradients. i.e.,

$$\frac{H_t}{H_B} \geq 4 \qquad (5\text{-}63)$$

Solving Eq. 5-63 for the angular subtense of the detector

$$\omega_\alpha \leq \frac{6 \times 10^{-11}}{4 \times 0.75 \times 10^{-4} \times 0.4} = 5 \times 10^{-7}\ \text{sr}$$

since $H_c(\lambda)$ is given as 0.75×10^{-4} and $\Delta f = 0.4$ (i.e., 4.7 - 4.3 = 0.4). Thus, the maximum detector angular subtense is established at a value which minimizes the likelihood of false background acquisition.

The seeker total acquisition field of view is calculated from the system acquisition requirements. Assuming a 0.9 target acquisition probability requirement, the baseline design concept will allocate a 0.95 probability for the target being in the seeker acquisition field of view and a 0.95 probability for detection when in the acquisition field of view. The seeker aiming error is specified as being normally distributed in each direction with a 1 σ-value of 10 mrad. On the assumption of an independence

of the errors in the two orthogonal axes, the size of a square acquisition field, having a 0.95 probability of including the target, will be 45 mrad by 45 mrad. For purposes of this example, this value is rounded off to a 50 mrad by 50 mrad square.

This acquisition field size is calculated as follows: Assuming normally distributed independent errors in X- and Y-directions, and a square field of view of 2 mrad on a side, the probability of the target falling within the square is

$$P(|X| < \alpha, |Y| < \alpha) = \frac{1}{2\pi} \int_{-\alpha/\sigma}^{\alpha/\sigma} \int_{-\alpha/\sigma}^{\alpha/\sigma} \exp\left[-\frac{1}{2}(X^2 + Y^2)\right] dx\, dy$$

$$= 4\left[erf\left(\frac{\alpha}{\sigma}\right)\right]^2$$

for $P(X, Y) = 0.95$, $erf\left(\frac{\alpha}{\sigma}\right) = 0.4875$. From the error function table we find: $\frac{\alpha}{\sigma} = 2.25$ which gives $\alpha = 22.5$ mrad.

Also, assuming a frame time requirement of 1/2 sec, the square 50 by 50 mrad target acquisition field of view can be readily searched in a five-bar raster-type scan using a sinusoidal bar with a 50 percent duty cycle. Bar-to-bar movement of the scan occurs during the 50 percent turnaround time. The detector dwell time is given by

$$t_d = \frac{\theta_X}{\omega_{max}}, \text{ sec} \qquad \text{[Eq. 5-32]}$$

where

θ_X = detector angular resolution in direction of scan, rad

ω_{max} = maximum angular scanning rate of field of view, rad sec^{-1}

Assuming square detectors, $\theta_X = \sqrt{\omega_\alpha}$ $= \sqrt{5 \times 10^{-7}} = 7 \times 10^{-4}$ rad $= 0.7$ mrad.

Assuming a five-bar scan—since the acquisition field was assumed to be 50 mrad × 50 mrad—it requires 10 mrad coverage per bar in the elevation direction. This, in turn, requires 14 detectors—10 mrad/0.7 mrad ≈ 14—each subtending 5×10^{-7} sr as required for suppressing the background.

The maximum angular scanning rate of the field of view is:

$$\omega_{max} = \frac{5(\text{bars}) \times 50(\text{mrad in azimuth scan})}{1/2(\text{frame time}) \times 1/2(\text{duty cycle})}$$

$$\approx 1000 \text{ mrad sec}^{-1} = 1 \text{ rad sec}^{-1}$$

Eq. 5-32 gives the detector dwell time:

$$t_d = \frac{7 \times 10^{-4}}{1} = 7 \times 10^{-4} \text{ sec}$$

To determine the threshold settings at the signal processing system in terms of system rms noise units, the specification for 0.95 target detection probability requires the threshold to be set at 1.65 noise units below the predicted signal for the target at 3 n mi (Eq. 5-12). The threshold setting in noise units above the maximum background signal will determine the false alarm rate.

The acquisition threshold level is established by considering the false alarm rate requirements of the system. It has been shown* that the expected number of positive slope crossings of a threshold per second, assuming white Gaussian noise is given by

$$E\left|N_{fa}\right| = \sqrt{\frac{f_b^3 - f_a^3}{3(f_b - f_a)}} \exp\left[-\frac{1}{2}\left(\frac{h}{\sigma}\right)^2\right] \qquad (5\text{-}64)$$

where

N_{fa} = false alarm rate, crossings per sec

f_b = upper cutoff frequency of filter, Hz

f_a = lower cutoff frequency, Hz

$\frac{h}{\sigma}$ = threshold-to-rms-noise ratio

For $\sqrt{\frac{f_b f_a}{f_b - f_a}} > 1$, the equation reduces to

$$E\left|N_{fa}\right| = f_c \exp\left[-\frac{1}{2}\left(\frac{h}{\sigma}\right)^2\right] \qquad (5\text{-}65)$$

where f_c is the filter center frequency, Hz.

For a scanning system of this type, the signal processing center frequency f_c is expressed as

$$f_c = \frac{1}{2t_d} = \frac{1}{2(7 \times 10^{-4})} = 740 \text{ Hz} \qquad (5\text{-}66)$$

Thus, for a per-channel false alarm rate of one per hour, the threshold level is set at

$$\frac{h}{\sigma} = \sqrt{2 \ln\left(\frac{f_c}{E\{N_{fa}\}}\right)} \approx 5.45 \qquad (5\text{-}67)$$

The difference between the target and the background signals must, therefore, exceed the seeker noise level by a factor of 5.45 (threshold level) plus 1.65 (allowance for 0.95 detection probability) = 7.1.

The required NEI is thus

$$NEI = \frac{H_T - H_B}{7.1}$$

Since H_B is constrained by the seeker design at $H_B = H_T/4$ (Eq. 5-63)

$$NEI = \frac{0.75 \times 6 \times 10^{-11}}{7.1}$$

$$= 6.3 \times 10^{-12} \text{ w cm}^{-2}$$

5-4.1.3 Seeker Design

5-4.1.3.1 Optical and Mechanical

The example optical system design is said to include all optical elements which must be placed within the space allocated in the mechanical primary design. It is assumed that refractive and folded catadioptric candidate systems were compared and that the catadioptric system was selected on the basis of cost and performance. Empirically, diffraction-limited optical system for *f/no.* of 3 to 4 will provide a blur circle diameter equal to the Airy disc on the optical axis providing essentially a linear increase in blur circle diameter with off-axis angular position. At 0.5 mrad off-axis, the diameter of the blur circle is assumed to be 5×10^{-3} cm. A 10 mrad field of view is adequate for target tracking. The optical-mechanical design permits the utilization of a linear array of detectors which will search the acquisition field by means of a 5-bar raster scan. When a target is detected, the seeker will go into the track mode. The optical-mechanical scanner will provide a single-bar scan of the tracking field of view of 10 × 10 mrad.

* *Selected Papers on Noise and Stochastic Processes*, Dover Publications, Inc., p. 193.

The catadioptric system has 25 percent blockage and consists of the following elements with their respective transmission for the 4.3- to 4.7-micron spectral passband:

Element	Material	Transmission
IR Dome	IRTRAN 2	0.95
Primary Mirror	—	0.98
Secondary Mirror	—	0.98
Corrective Lens	Silicon	0.95
Cell Window	Silicon	0.95
Spectral Filter	—	0.65

$0.95 \times 0.98 \times 0.98 \times 0.95 \times 0.95 \times 0.65$

= 0.53 Transmission

5-4.1.3.2 Detector

As was mentioned in the previous subparagraph, a linear detector array of 14 square detectors, which subtends a total of 10 mrad in elevation and 0.7 mrad in azimuth, would be used. Each detector would have an instantaneous field of view of 5×10^{-7} sr. The detectors would be lead selenide-cooled to 193°K to provide a D* of 4×10^{10} cm $Hz^{1/2}$ w^{-1} in the 4.3- to 4.7-micron region.

5-4.1.3.3 Acquisition-mode Programmer

The acquisition-mode programmer will cause the seeker to scan the square 50 × 50 mrad target acquisition field in a 50 percent duty cycle sinusoidal five-bar scan in 1/2 sec. This requires a 5 Hz azimuth scanning rate with elevation stepping of 10 mrad occurring during the turn-around time of 0.05 sec. With a free-gyro gimbal system, an elevation precession rate of 0.2 mrad/sec must be achieved during the stepping period. A first approximation of the torque T required to produce this precession rate would be obtained by

$$T = I\omega_r\Omega_v, \text{ in.-oz} \qquad (5\text{-}68)$$

where

I = moment of inertia of the rotating seeker components about the spin axis, in.-oz sec^2

ω_r = seeker spin velocity, rad/sec

Ω_v = angular precession velocity about output axis, rad/sec

The inertia of the free gyro wheel for a seeker of this type is approximately 0.01 in.-oz sec^2, and a realistic spin velocity without encountering bearing problems is 500 rad/sec. The torque requirement for precessing the seeker in elevation during the acquisition mode with Ω given as 0.2 mrad/sec is

$T = 0.01 \times 500 \times 0.2$

= 1 in.-oz

Similarly, a peak torque of 5.5 in.-oz is required in order to achieve the 1.1 rad/sec scan rate.

5-4.1.3.4 Track-loop-mode Servo

A Type II servo tracking loop would be utilized in this seeker design due to its small field of view and to the tendency of such a servo to compensate for gimbal unbalance and drift without incurring a static boresight error. In addition, under certain circumstances, the system may be designed to operate with a lower tracking loop bandwidth, resulting in improved tracking at lower signal-to-noise ratios and improved tracking (rejection) of intermittent disturbances in error signal caused by decoy flares and backgrounds.

The maximum angular rate was specified as 20 deg/sec. This resulted from the intercept kinematics and the desire to maintain target track to within a "time to go" distance to target impact which is less than the response time of the closed loop missile system.

Maximum angular line of sight rate to a 600 fps velocity target at the minimum launch range of 2000 ft results in an 18 deg/sec launch initial tracking rate. Missile body pitching motions will be compensated for by the free gyro gimbal and, to a first approximation, need not be accounted for in track loop design. For baseline design purposes it will be assumed that the missile will be initially directed at the target, and the input initial conditions will be approximated by zero target position error, with an initial angular line of sight rate of 18 deg/sec. This initial angular rate requires a servo transient response time small enough to keep the target within the 10 mrad field of view.

For purposes of preliminary transient response analysis, the linearized block diagram (Fig. 5-21) may be approximated as

$$G(S) = \frac{K_a(\tau_3 S + 1)}{S^2(\tau_1 S + 1)} \qquad (5\text{-}69)$$

where

$K_a = \frac{K_1 K_2 K_A K_S}{H}$, as Fig. 5-21 and

where the τ_i correspond to the location of the pole or zero of the network. It is assumed τ_2 and τ_4 are small for purposes of this calculation.

The error for a ramp input may be written

$$\epsilon(S) = \frac{\dot{\theta}}{S^2[1 + G(S)]} \qquad (5\text{-}70)$$

$$= \frac{\dot{\theta}(\tau_1 S + 1)}{S^2(\tau_2 S + 1) + K_a(\tau_3 S + 1)}$$

for relatively large τ_3 (0.2 sec) and relatively large K_a, (>100 sec^{-2})

$$\epsilon(S) \cong \frac{\dot{\theta}\left(S + \frac{1}{\tau_1}\right)}{\left(S + \frac{1}{\tau_3}\right)\left[S^2 + \frac{1}{\tau_1} + \frac{K_a\tau_3}{\tau_1}\right]} \qquad (5\text{-}71)$$

Finding the peak value of the error by taking the maximum of the inverse transform of $\epsilon(S)$

$$\epsilon(t)_{peak} \cong \frac{\dot{\theta}}{K_a\tau_3} \qquad (5\text{-}72)$$

(Note that $K_a\tau_3$ is just the velocity constant of the equivalent high-frequency response of a Type I servo.)

By imposing the constraint based on the oscillatory roots of $1 + G(S)$ having a per-unit critical damping of 0.5, the result is expressed as

$$K_a\tau_3 = \frac{1}{\tau_1} \qquad (5\text{-}73)$$

Thus, the natural frequency of these roots is given as

$$(2\pi f_n) = \frac{K_a\tau_3}{\tau_1} = (K_a\tau_3)^2 \qquad (5\text{-}74)$$

In reality, this analysis must be ultimately performed using transforms due to the discrete sampling nature of the signal processing system. Through such an analysis, it would be found that the servo will operate with oscillatory roots in vicinity of

$$\frac{f_s}{6} < f_n < \frac{f_s}{3} \qquad (5\text{-}75)$$

where f_s is the sampling rate, and f_n is the natural frequency of the servo. For purposes of this simplified analysis, $f_n = \frac{f_s}{4}$ was selected. Therefore

$$\frac{f_s}{4} = f_n = \frac{K_a\tau_3}{2\pi}\ ;\ f_s = \frac{2}{\pi}(K_a\tau_3)$$

For an initial velocity of 18 deg/sec (314 mrad/sec) and a maximum error of 5 mrad,

$$f_s = \frac{2}{\pi}\left(\frac{314}{5}\right) = 40 \text{ samples/sec minimum}$$

A higher value would be required to maintain the target well within the seeker field of view. Consequently, a 30/sec frame rate (60 samples/sec at the center of the field) would be selected.

5-4.1.3.5 Sensitivity

The size of the entrance aperture necessary for attaining the required *NEI* (Eq. 5-4) is determined as follows. The initial optical transmission value ϵ_o of 0.53 was determined to be 0.4 due to the 25 percent blockage factor (see par. 5-4.1.3.1). Electrical efficiencies of approximately 0.65 were assumed wherein the bandwidth was inversely proportional to twice the detector dwell time. A small detector area is desired for maximum sensitivity and from an optical design standpoint, a 5×10^{-3} cm linear dimension detector is required to match the maximum blur circle at the edge of the track field of view. The detector instantaneous field of view 5×10^{-7} sr required for background discrimination dictates a focal length of 7 cm.

The 780-Hz bandwidth of the processing electronics was selected for the acquisition mode. The bandwidth for the tracking mode is the result of the following considerations. Assuming a sinusoidal scan angle of

$$\theta_s = A \sin \omega T, \text{ rad} \qquad (5\text{-}76)$$

where

A = maximum amplitude of angular scan, rad

ω = angular velocity of scanner, rad/sec

T = time, sec

Amplitude A is related to the field of view (for

$$A = \frac{\frac{1}{2}\,\text{fov}}{\sin\left(\frac{\pi}{2}K_d\right)} \qquad (5\text{-}77)$$

where K_d is the scan efficiency.

From Eq. 5-76 the angular scan rate $\dot{\theta}_s$ is calculated as

$$\dot{\theta}_s = A\omega\cos\omega T \qquad (5\text{-}78)$$

which has a maximum value of

$$\dot{\theta}_{s\,(max)} = A\omega$$

With a total field of view of 10 mrad, a sinusoidal scan rate of 30 Hz, and a 50 percent scan efficiency

$$\dot{\theta}_{s(max)} = \frac{0.005}{\sin\left(\frac{\pi}{2}\,0.5\right)} \times 30 \times 2\pi$$

$$= 1.33 \text{ rad/sec}$$

The detector dwell time is calculated as

$$t_d = \frac{\theta_x}{\dot{\theta}} = \frac{7 \times 10^{-4}}{1.33} = 5.25 \times 10^{-4} \text{ sec.} \qquad (5\text{-}79)$$

The noise-equivalent bandwidth for a pulse detector is expressed as

$$\Delta f = \frac{1}{2t_d}$$

$$= \frac{10^4}{2 \times 5.25} = 9.5 \times 10^2 \text{ Hz}$$

The *NEI* is directly proportional to the square root of the signal processing bandwidth. The small difference in the bandwidths of the acquisition and tracking modes causes their respective *NEI* to be within 10 percent of one another. The entrance aperture area is determined for the tracking mode by solving for A_o in Eq. 5-4.

$$A_o = \frac{\sqrt{A_d \Delta f}}{\epsilon_o \epsilon_e (NEI) D^*}$$

$$A_o = \frac{\sqrt{(5 \times 10^{-3})^2 \times 9.5 \times 10^2}}{0.4 \times 0.65 \times 1 \times 10^{10} \times 6 \times 10^{-12}}$$

$$A_o = 2.5 \text{ cm}^2$$

$$D_o = 1.8 \text{ cm}$$

The *f/no.* of the seeker would be determined

$$f/no. = \frac{\text{focal length}}{\text{entrance aperture}}$$

$$= \frac{7}{1.8}$$

$$= 3.9$$

5-4.1.3.6 Signal Processing

The signal processing system for a typical pulse-position track-while-scan seeker, such as the example treated herein, is illustrated in the block diagram of Fig. 5-23. As shown, each detector will have an amplifier followed by a threshold circuit. The threshold output of each detector comprises one input to an "and" gate whose other input consists of an analog voltage proportional to the azimuth scan position from the optical axis. The output of this channel is an error signal to the tracker azimuth servo. The elevation signal is derived from a similar analog "and" gate whose two inputs consist of the detector threshold signal and the analog voltage proportional to the detector position. The output of this channel is an error signal to the tracker elevation servo.

The seeker elevation field of view array, determined by the length of the detector and target elevation angular position, is indicated by observing the detector which is above threshold. The azimuth target position is determined by the scanner position at which the target is detected.

5-4.2 ACTIVE IMAGING SYSTEMS

5-4.2.1 Sample Design of Active IR Imaging Systems

This paragraph describes the development of a sample design of a spot scan Nd:YAG laser illuminator and a single S-1 photomultiplier with a synchronized, television-type display. The design requirements dictate that the system allow an airborne observer to identify a truck at night while presenting a contrast ratio of 25 percent between the target and background at a range of 275 m. The geometry of the field of view of the system, which will be 5 × 5 deg, is as illustrated in Fig. 5-24.

5-4.2.1.1 Spatial Resolution and Field of View

Assume that an observer requires 10 linear resolution elements in order to identify a truck

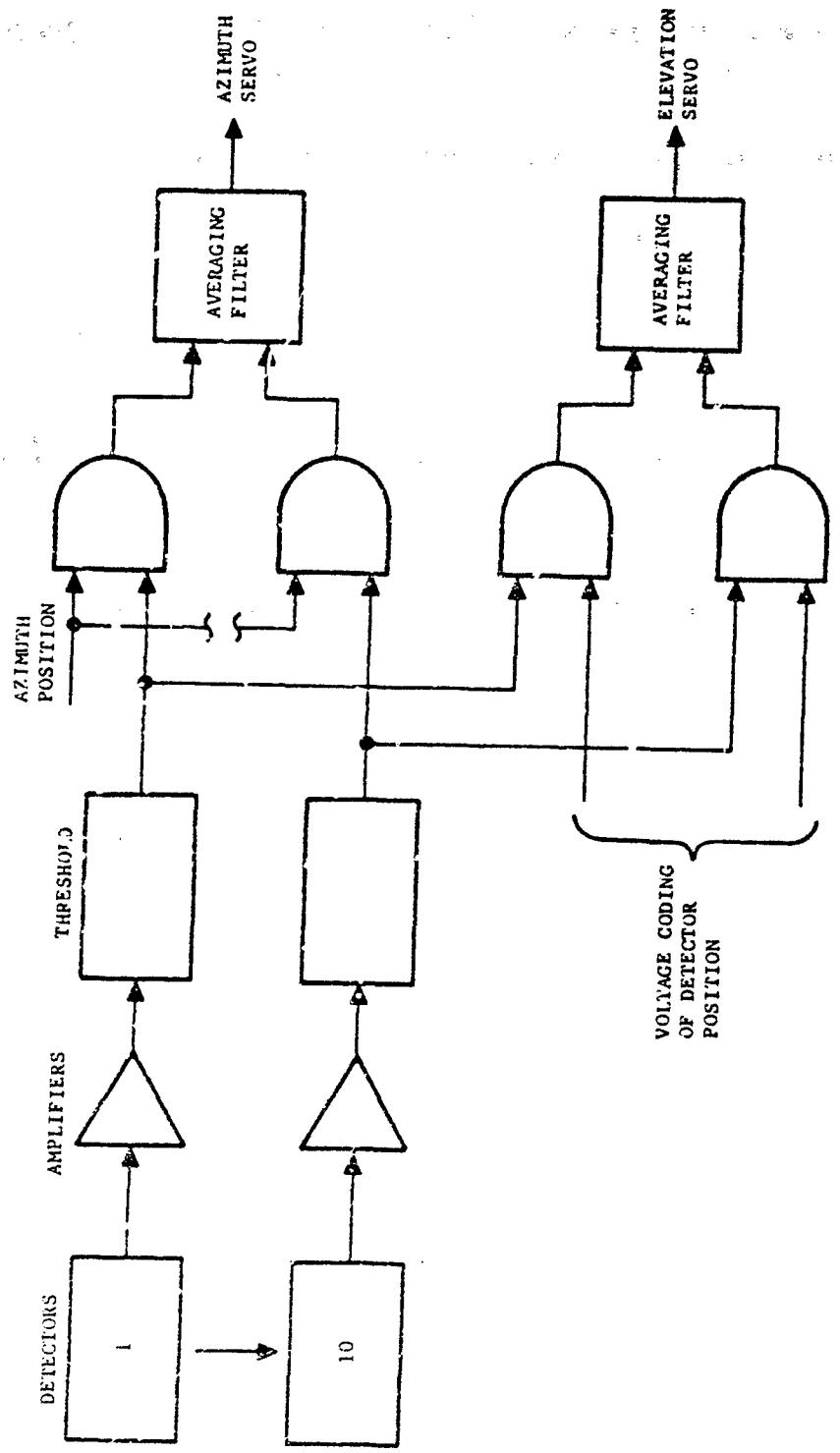


FIGURE 5-23. Tracker Signal Processing

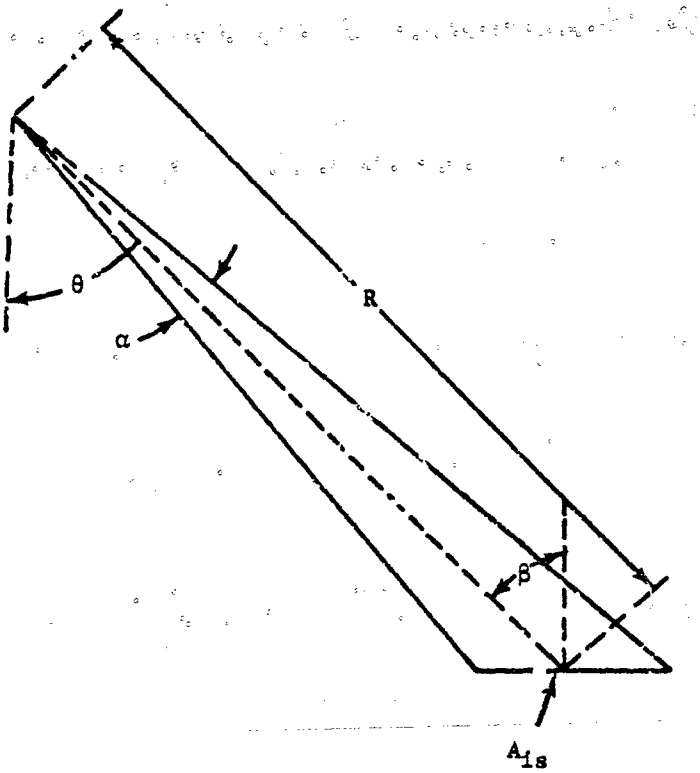


FIGURE 5-24. System Geometry

at 275 m with a contrast ratio of 25 percent, and that its dimensions are 2 × 2 × 4 m. The angular resolution α_r required of the system is expressed as

$$\alpha_r = \frac{D_o\sqrt{C}}{KR} \text{ , rad} \qquad (5\text{-}80)$$

where

D_o = minimum object dimension, m

C = contrast of target and background

K = number of linear resolution elements required for identification*

R = range, m

$$\alpha_r = \frac{2\sqrt{0.25}}{10(275)} = 0.36 \text{ mrad}$$

$$= 3.6 \times 10^{-4} \text{ rad}$$

For a field of view of 5 × 5 deg (0.087 × 0.087 rad), the total number of resolution elements N within the field of view is

$$N = \frac{\text{area fov}}{\text{area resolution elements}}$$

$$= \frac{(0.087)(0.087)}{(3.6 \times 10^{-4})(3.6 \times 10^{-4})}$$

$$= \frac{(8.7 \times 10^{-2})^2}{(3.6 \times 10^{-4})^2}$$

$$= 5.8 \times 10^4$$

5-4.2.2 Data Rate and System Bandwidth

If a frame rate of 30 frames per second is selected for the system, the number of resolution elements scanned per second would be 30 × 5.8 × 10^4 = 1.74 × 10^6. Consider a resolution element as a pair of black and white lines. The dwell time t_d required for the laser beam to sweep across a resolution element is the inverse of the number of elements scanned per

* The number of resolution elements required for detection, orientation, and identification of various military objects is given in Ref. 33.

second or 0.575 microsec. The bandwidth required to resolve each black line separated by a white line would be

$$\Delta f \geq \frac{1}{t_{id}} \geq 1.74 \times 10^6 \text{ Hz} \qquad (5\text{-}81)$$

5-4.2.1.3 Laser Power and Signal-to-noise Characteristics

If the reflection from the ground is diffuse, the radiance N of the illuminated spot on the ground can be written as

$$N = \frac{P_t \tau_a \tau_{io} \rho}{\pi A_i}, \text{ w m}^{-2} \qquad (5\text{-}82)$$

where

P_t = laser power, w

τ_a = atmospheric transmission

τ_{io} = transmission factor of illuminator optics

ρ = reflectance

A_{is} = area of illuminated spot, m^2

The reflected power P_d impinging upon the detector is expressed as

$$P_d = \frac{N A_{is} A_r \tau_{ro} \tau_a}{R^2} \cos\theta, \text{ w} \qquad (5\text{-}83)$$

where

A_r = receiver area, m^2
τ_{ro} = transmission factor of receiver optics
R = range, m
θ = angle between the laser beam and normal to surface

Substituting Eq. 5-82 into Eq. 5-83

$$P_d = \frac{P_t \tau_a^2 \tau_{io} \tau_{ro} A_r \rho \cos\theta}{\pi R^2}, \text{ w} \qquad (5\text{-}84)$$

The signal current i_s from the photocathode of the photomultiplier can be written as

$$i_s = \frac{P_d \eta \lambda e}{hc}, \text{ A} \qquad (5\text{-}85)$$

where

η = photocathode quantum efficiency

λ = wavelength, m

h = Planck's constant, 6.6256×10^{-34} J-sec

c = speed of light, 3×10^8 m/sec

e = charge on the electron, 1.602×10^{-19} C

Combining Eqs. 5-84 and 5-85, the signal current equation becomes

$$i_s = \frac{P_t \tau_a^2 \tau_{io} \tau_{ro} A_r \rho \eta \lambda e \cos\theta}{\pi R^2 hc} \qquad (5\text{-}86)$$

The signal-to-noise ratio for this case is expressed as

$$S/N = \frac{i_s}{\sqrt{2e(i_s + i_{bkd} + i_d)(\Delta f) + \frac{kT(\Delta f)}{GR_L}}} \qquad (5\text{-}87)$$

where

e = charge on the electron, 1.602×10^{-19} C

i_{bkd} = background induced current, A

i_d = photomultiplier dark current, A

Δf = noise equivalent bandwidth, Hz

k = Boltzmann's constant, 1.38×10^{-23} J/°K

T = temperature, °K

G = electron multiplier gain

R_L = load resistor, ohm

The first term under the radical is the shot noise due to the total photocathode current, while the second term is the thermal or Johnson noise of the load resistor divided by the electron multiplier gain. The current due to the background can be expressed as

$$i_{bkd} = \frac{P_{bkd} \eta \lambda e}{hc}, \text{ A} \qquad (5\text{-}88)$$

where

$$P_{bkd} = \frac{H_\lambda \rho \Delta\lambda A_{bkd} A_r \tau_{ro} \tau_a \cos\theta}{\pi R^2}, \text{ w} \qquad (5\text{-}89)$$

and

H_λ = spectral irradiance on ground, w m$^{-2}\mu^{-1}$

$\Delta\lambda$ = spectral bandwidth of receiver optical filter, μ

A_{bkd} = background area in receiver field of view, m^2

The background area for this case is

$$A_{bkd} \simeq \frac{\pi\alpha^2 R^2}{4\cos\theta} \qquad (5\text{-}90)$$

where

α = receiver field of view, rad.

Substituting Eq. 5-90 into Eq. 5-89,

$$P_{bkd} = \frac{H_\lambda \rho \Delta\lambda \alpha^2 A_r \tau_{rd} \tau_o}{4} \qquad (5\text{-}91)$$

Eq. 5-91 implies that P_{bkd} is independent of the angle θ as long as a uniform background fills the receiver field of view.

Several assumptions can be made to simplify Eq. 5-87: (1) the spectral bandwidth $\Delta\lambda$ in Eq. 5-91 can be made narrow enough such that $P_{bkd} << P_d$ and, thus, i_{bkd} can be neglected compared to i_s; (2) in a photomultiplier, i_d is sufficiently small to be neglected; and (3) the electron multiplier gain G is sufficiently large such that the Johnson noise term, $\frac{kT\Delta f}{GR_\ell}$, can be neglected. Eq. 5-87 can be rewritten as

$$S/N = \frac{i_s}{\sqrt{2ei_s\Delta f}} = \sqrt{\frac{i_s}{2e\Delta f}} \qquad (5\text{-}92)$$

Combining Eqs. 5-86 and 5-92 the following expression is obtained for the laser power

$$P_\ell = \frac{2\pi(S/N)^2 \Delta f R^2 hc}{\tau_a^2 \tau_{io} \tau_{ro} A_r \rho \eta \lambda \cos\theta}, \text{ w} \qquad (5\text{-}93)$$

Consider a voltage signal-to-noise ratio of 3 without eye integration. In one frame there are approximately 5.8×10^4 resolution elements, each of which has an equal probability of containing a noise signal. Assuming a Gaussian distribution for the noise, there is a 1.3×10^{-3} probability of a noise pulse exceeding a level of three times the rms in any resolution element. Therefore, on the average, there would be 75 resolution elements containing noise pulses as large as the signal. It would be virtually impossible without visual integration to detect the true target from among 75 false signals. However, the human eye integrates over a period of 0.2 sec. Thus, at 30 frames/sec, the eye takes six separate "looks" integrating (averaging) the results. The noise variance is reduced by the inverse of the number of looks and the rms noise is reduced by the inverse of the square root of the number of looks. Thus, with six looks, the effective voltage signal-to-noise ratio is 7.5 when the frame-to-frame signal-to-noise ratio is 3. The probability of any noise pulse exceeding a threshold of 7.5 is approximately 10^{-14}, consequently, random noise is essentially eliminated at a frame-to-frame voltage signal-to-noise ratio of 3.

If the signal-to-noise requirement were set at 2 rather than 3, the effective signal-to-noise ratio would be 5 and approximately two noise samples would be expected to exceed the 3:1 threshold during each visual integration period. However, even two false targets per visual integration period (0.2 sec) is probably excessive and a signal-to-noise ratio of 3 should be maintained.

For a system in which the S/N voltage = 3, $\Delta f = 1.74 \times 10^6$ Hz, $R = 250$ m, $\tau_a = 0.9$, $\tau_{io} = 0.8$, $\tau_{ro} = 0.4$, $A_r = 60.1$ m^2, $\rho = 0.1$, $\eta = 10^{-3}$, $\lambda = 1.06\mu$, and $\theta = 45$ deg; the required laser power would be 0.67 w (Eq. 5-93). Fig. 5-25 is a block diagram representing such a system.

5-5 TRADE-OFF ANALYSES

5-5.1 GENERAL PRINCIPLES

In the process of analyzing, designing, fabricating, deploying, operating, and logistically supporting an infrared sensor system; a major milestone is the so-called critical design review. At this point all major decisions are made regarding the system design. These decisions will determine the eventual functional utility of the system, whether the sensor can be produced on schedule and with the funds allocated, whether the system will eventually operate reliably with reasonable maintenance cycles and logistic support, the extent to which future technological advancements can be retrofitted into the system (i.e., growth potential), and other decisive questions. Each decision made at the critical design review will have been the result of considering many alternatives from which selections had to be made. These selections must be based on the admittedly difficult balance between performance, cost, maintainability, reliability, supportability, and use.

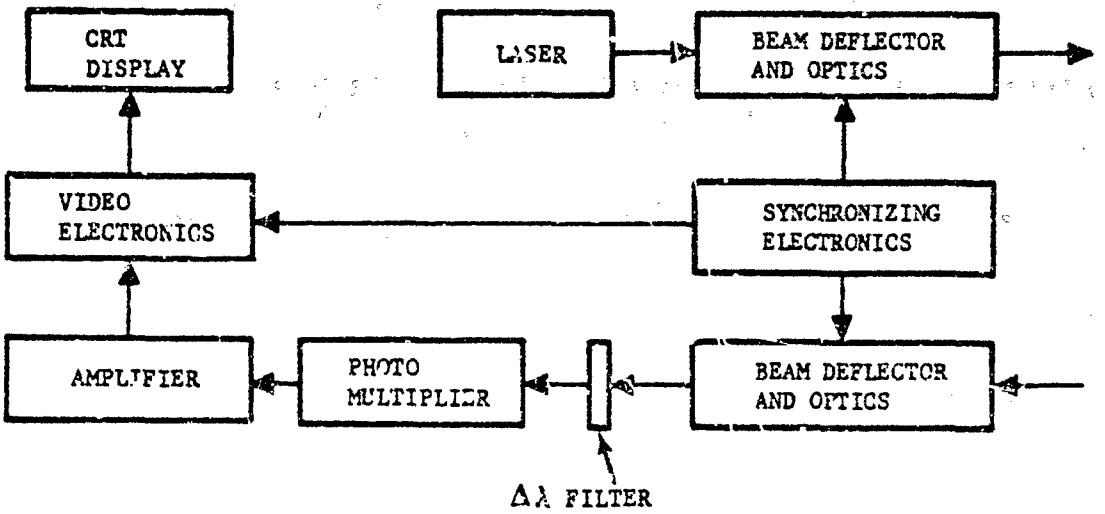


FIGURE 5-25. System Block Diagram

No exact rules can be provided as a guide in selecting the design characteristics in the face of conflicting benefits. What can be done is to examine in detail the implications of each of the important sensor system design decisions. It is the intent of this paragraph to review some of the basic principles which affect design decisions in the areas of human engineering, reliability, maintainability, producibility, aerospace ground equipment, and logistics.

It is not possible to arrive at a suitable infrared sensor system design if careful analysis is not provided in each of these areas and design implications identified and traded off.

5-5.2 USE OF ADVANCED COMPONENTS AND CONCEPTS

In striving for maximum performance, the sensor system designer has to consider the use of advanced optical, electronic, and mechanical components and concepts. However, before doing so, he must consider all the cost elements to determine whether or not the increased performance is in fact worth the added cost. Specifically—uncertainties in reliability, producibility, maintainability, ground equipment requirements, and logistics can represent hidden costs that manifest themselves only after the initial sensors are fabricated and tested. Frequently, it is desirable to consider those advanced components which may entail high risk only in terms of future growth potential for the sensor system and to design the sensor in such a way as to permit later incorporation into the sensor should such incorporation prove warranted.

5-5.3 COST, SPACE, WEIGHT, AND POWER TRADE-OFFS

In performing cost effectiveness analyses, it is relatively straightforward (although frequently difficult) to determine the complete dollar costs of an infrared sensor system over some assumed operational lifetime. It is much more difficult to determine an equitable cost to assign to penalties such as weight, power, and space requirements. Frequently, these are system-related questions: systems in the sense in which the IR sensor is a subsystem. In performing trade-off analyses for the purpose of selecting design parameters, the IR sensor designer must make an attempt to view weight, space, and power penalties from the ultimate user's point of view. As an example, in the design of satellite-borne sensors, space and power can be converted to weight through the structure and solar cells required. These equivalent weights can be converted to dollars based on launcher and ground support costs and, thus, all costs can be expressed as dollars in performing cost trade-offs. In most IR sensor missions, however, the equivalent cost relationships are subtle, difficult to evolve, and agreement is difficult to reach.

5-5.4 HUMAN FACTORS

In recent years, much research has been performed with the objectives of influencing designs in such a manner that the resultant product can be used with maximum effectiveness and of defining environments suitable for the human condition. This field of endeavor has come to be known as "human factors" or "human engineering".

The precision performance and high output demands created by sophisticated military space and electronic systems indicate the direction and focus of much current and future research. It is essential to obtain the most efficient output of the human operator with a minimum overload to his senses. Thus, for example, where the sense of vision is involved, a small saving in the time necessary to read a display coupled with the accuracy required may—in addition to saving an operator's life as its prime result—be the most efficient method of executing a mission.

Man has many capabilities which may be exploited to enhance effectiveness of the man-machine system. For example, because of his ability to discriminate signals in a changing field, man provides a flexible means of controlling retrieval methods for infrared target information which otherwise may possibly go undetected in a strictly automated system. He can vary thresholds and search patterns to fit anticipated data rates and target discriminability. He may react to emergencies or the unusual by taking adaptive or corrective measures such as selecting alternate modes of operation, locating and repairing faults, or solving unforeseen problems. Man's ability to contend with unexpected aspects of displayed data patterns, together with his cognitive capacity, increases an infrared system's reliability and significantly improves its flexibility. Being able to discriminate the useful from the irrelevant, he minimizes the data collection effort and, in the role of data evaluator, edits the collected information such that only pertinent and meaningful data are stored or transmitted.

When specifying man's role in the infrared system, various physical and psychological parameters must be considered to optimize his usefulness. Examples of this type of data are set forth in the paragraphs which follow. This information is related primarily to the definition of the infrared display[2].

5-5.4.1 Visible Spectrum

The visible spectrum extends from about 3800 to 7200 Angstroms (Å) with variations in wavelength manifested by changes in color; the violets being at about 4000 Å, blending into the blues at about 4500, greens at around 5000, yellow-orange at 6000, and reds at 7000 and above. The peak response of the human eye occurs in the green band; a fact which should be considered when selecting a phosphor color and type of illumination[3]. This is not to say, however, that optimal phosphors would necessarily be in the green.

The use of color to identify categories of information is well known. Proper use of color makes it possible for the operator to discriminate items within a category more rapidly. The individual colors used must be sufficiently different to allow immediate perception. In terms of information flow rates, it has been found that numerically-coded data are superior to color-coded data, and color-coded-numeric data are best[3].

5-5.4.2 Visual Acuity

Visual acuity, which is a measure of the smallest visual angle that the eye can resolve, has a direct influence on the establishment of resolution parameters for displays. There are different kinds of visual acuity measurements, each of which is important in design development.

Minimum separable acuity is the smallest gap which the eye can detect or the smallest space between the parts of a target. Acuity is highly dependent upon illumination and the amount of contrast between a test target and background brightness. As rule of thumb, it can be stated that the eye can detect a gap which subtends a visual angle of about 1 minute of arc at ordinary indoor-light levels (20 to 50 footlamberts) when the target has a high brightness contrast[2]. At illumination levels below 0.1 mlambert, the eye will no longer have the 1 minute of arc resolution and visual performance capability begins to deteriorate[4].

Minimum perceptible acuity is a measurement of the capability of the eye to detect a spot which is either darker or lighter than its background. The size of the spot which the eye can

detect depends upon the brightness level and brightness contrast. There is no lower detectable-size limit for spots which are brighter than their backgrounds. The eye can detect a bright spot no matter how small it is as long as it is bright enough. For example, the star Mira can be seen, although it subtends a visual angle of only 0.056 sec of arc. Lines and squares are also very visible against bright backgrounds. Measurements indicate the eye can usually see a wire 1 deg long and 0.43 sec wide against a bright sky background. These numbers, for all practical purposes, represent the lower limits of minimum perceptible acuity[5].

The eye responds to levels of illumination over a total dynamic range of ten billion to one—from 10^{-5} to 10^{5} mlamberts. Visual acuity is best when the eye is adapted to about the same brightness level as the target and the immediately surrounding area. In general, the adaptation mechanism of the eye permits us to discriminate within a dynamic brightness range of about 1000 to 1 without requiring time to adapt to light or dark[6]. If possible, the display should be at least 0.01 as bright as the preexposure field. For fine discrimination of detail, the display should be operated at a scene brightness above 0.01 mlambert[3]. This permits the use of cone vision and its correspondingly better visual acuity.

5-5.4.3 Other Factors Affecting Acuity

Such variables as brightness contrast, illumination available, time to view the object or display, brightness ratio, movement, and glare influence visual discrimination.

Brightness contrast is defined in par. 3.8. In general, the greater the contrast between an object and the background, the greater the visual discrimination; high contrast permits smaller details to be distinguished. The longer the viewing time, within reason, the greater is the discriminability.

The brightness ratio is the ratio of illumination on the object being viewed to the illumination of the surrounding area. In general, this ratio should be low, and the illumination of the surrounding area should not be in marked contrast with the primary visual area.

Of interest to the display designer is the ability of humans to identify targets. The eye can identify letters of the alphabet which subtend a visual angle at least 5 min of arc. In fact, it is this capability which defines 20-20 vision. However, it has been demonstrated that when the visual angle subtended by a target is less than 12 min of arc, search time and errors in identification increase. Results of research[4] indicate that 12 min of arc is a reasonable value for accurate identification.

Movement is an additional variable that affects man's ability to make visual discriminations. Movement of targets on a display may occur when the observer is moving, when the observer and the target are both moving, or when only the target is moving. The capability of humans to discriminate targets under these circumstances is termed "dynamic visual acuity". This is usually a measure of acuity deterioration in terms of target movement in degrees per second. These visual discriminations are also a function of target size and viewing time. For example, a subject can achieve 100 percent recognition of a target subtending an angle of 39 min, traveling vertically at an angular velocity of 40 deg per sec, when the viewing aperture subtends a visual angle of 15 deg. This angular rate deteriorates to 33 deg per sec when the size of the viewing aperture is reduced to 7.5 deg[7].

In a well-known study by Cobb and Moss[8], the interaction between visual variables, when two or more are used in combination, was plotted. Visual acuity was measured under varying conditions of brightness contrast, time, and illumination. Acuity was expressed in terms of the visual angle of the smallest target that could be discriminated. Results of the study are illustrated in Fig. 5-26. As an example of the use of this figure, trace the 100-mlambert level on the 0.300-sec exposure time curve. Note that, for a small test target (0.7 min of arc), a contrast of 50 percent is required; whereas, for a relatively large target (5 min of arc), a contrast of only 2 percent permits discrimination[3].

5-5.4.4 Flicker

Perceptible flicker, which varies with individuals, has been found to be a function of many factors such as the age of the individual, color of the displayed light, location of the displayed data in the field of view of the observer, brightness, and the ratio of light to dark[9]. Its

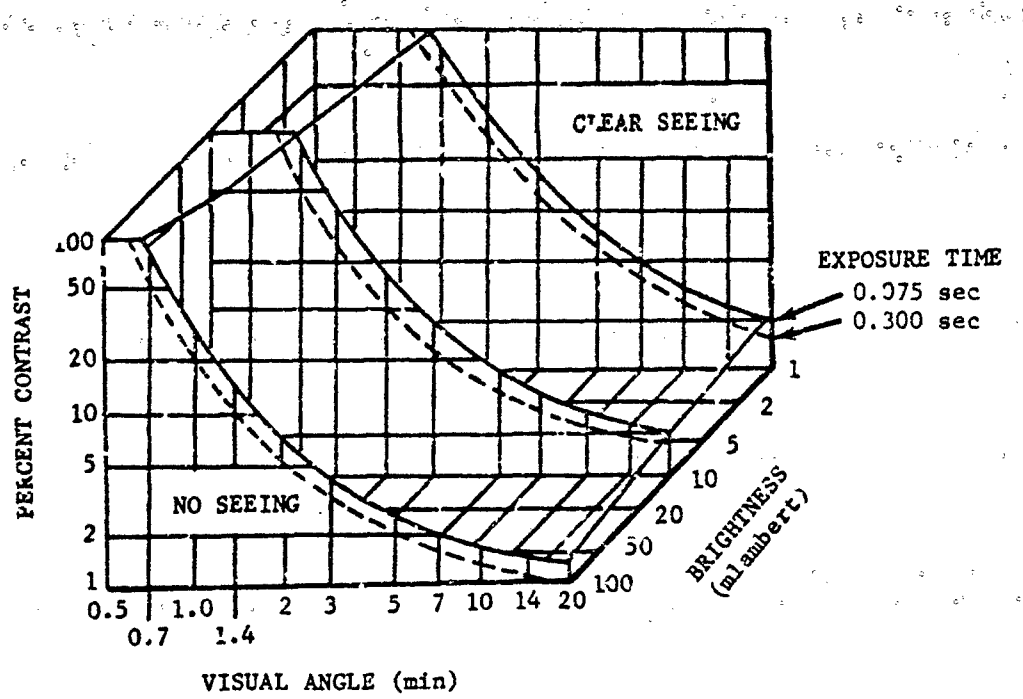


FIGURE 5-26. Relation Among Brightness Contrast, Brightness of Background, Exposure Time, and Visual Acuity

adverse effects upon a display monitoring task are well known and vary from mild distraction to annoyance, fatigue, and nausea. To avoid flicker, a display must be operated at a sufficiently high refresh rate so that fusion occurs and the eye can no longer perceive the on-off cycles. Fig. 5-27 illustrates the critical flicker frequency at various brightness levels[10].

If a CRT display is to be used, the designer must take into consideration the flicker data for the phosphor used in the particular tube. Various commercial types of phosphor are available and these vary mainly in terms of persistence. If a display is to be operated at high brightness levels, the refresh rate must be higher than for a display which is to be operated under dim conditions. Extensive studies were performed by the television industry in arriving at a flicker-free refresh rate of 60 fields (30 full frames) per sec. These rates must be tailored to meet the specific requirements for each system as determined by the human engineering data.

5-5.4.5 Supplementary Auditory Displays

Two or more sensory channels can be combined in the transmission of input data to the display operator. Usually, these two channels are visual and auditory. Auditory displays should be provided where redundancy in the presentation of information will facilitate display operator performance, particularly under conditions of reduced operator alertness. There is virtually no question[3] about the increased probability of receiving information more effectively when presented simultaneously visually and aurally (see Fig. 5-28)[12]. An example of the effectiveness of the combination is the use of an audio warning signal which indicates a condition requiring immediate attention or action on the part of the operator who may have failed to note the condition on the visual display. The major concentration of energy in such an audio alerting device should be between 250 and 2500 Hz, and the signal should be readily identifiable on the basis of components below 2000 Hz. The sound pressure level of audio alerting signals should be at least 10 dB above the maximum ambient noise level, but should not exceed a maximum intensity level of 110 dB (0.0002 dyne/cm^2)[11].

5-5.4.6 Human Reaction Time

The time required for the human to respond

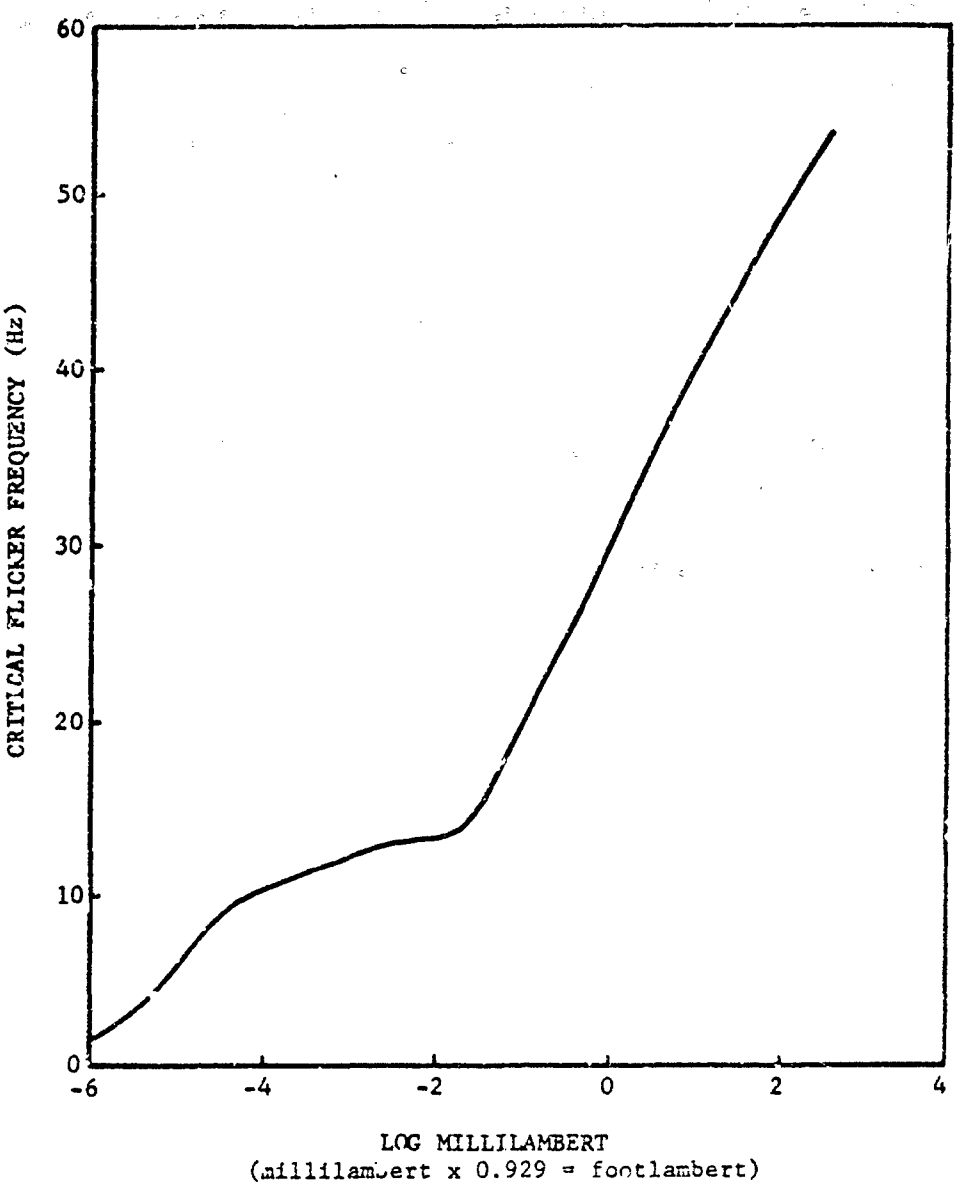


FIGURE 5-27. Critical Flicker Frequency As a Function of Brightness

to a signal is a function of the complexity of the response and of the body member being used. Very simple responses involve a few hundredths of a second (e. g., pushing a button), but more complex responses such as positioning movements may require a few tenths of a second. With the simplest tasks, where decision time is not required, the minimum reaction time to sense and respond is on the order of 0.15 to 0.20 sec, depending upon the senses used. A man can push a button in response to a light signal in about 0.18 sec. His reaction time to an auditory stimulus will be somewhat less than this[2].

Decision time will vary considerably, depending upon the complexity of the decision to be

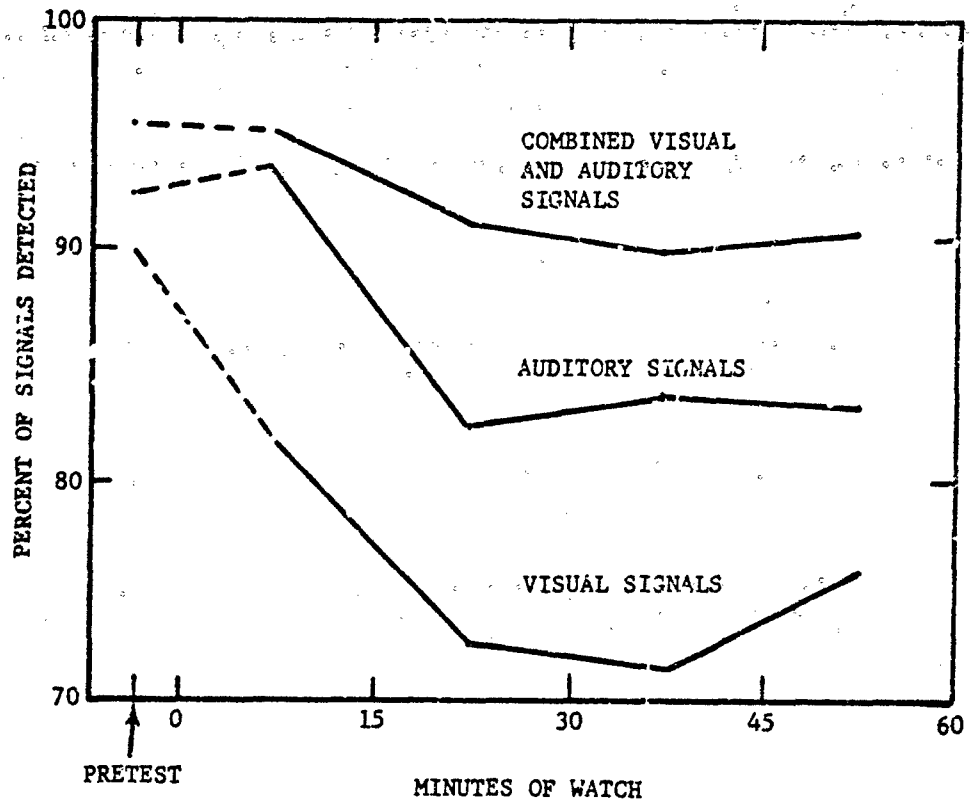


FIGURE 5-28. Performance on a Visual, Auditory, and Combined Visual-auditory Vigilance Task

made. As a rule of thumb, it can be stated that decision time is proportional to the logarithm of the number of alternative choices[2].

5-5.4.7 Human Engineering

There are many additional aspects of human performance which must be considered when defining the overall man-machine interface. Thus, the human factor engineer frequently employs the systems approach to solving design problems. This approach includes systematic attention to both engineering and human factor considerations toward the objective of developing integrated systems that consist of optimum combinations of physical and human components. A human factor engineer typically analyzes system configurations to identify and define human performance requirements and system simulation techniques. Results of such analyses may be used to establish system requirements for communications, measurements, controls, and displays.

On a recently developed (classified) IR missile system it was necessary to establish the frequency and intensity of an audible target acquisition signal. The signal was to be clearly audible to the operator under battlefield noise conditions. Human engineering tests were conducted using human subjects to ascertain the effectiveness of various auditory signals and techniques for penetrating high ambient noise. One approach to the problem entailed conducting sound through bone, thereby bypassing the middle ear in the event of temporary deafness.

Additional problems encountered with the same system involved determining the intensity, color, and location of a visual target-acquisition signal. The visual indicator had to be bright enough to be discerned by the operator under both day and night conditions, yet not so bright as to impair his visual dark adaptation or interfere with his visual search for targets.

Human factor engineering was also applied to the unique design and arrangement of the

system controls, protection from blast of a rocket engine, handling and carrying requirements, personnel requirements, training requirements, and many other aspects of system development.

5-5.5 RELIABILITY

5-5.5.1 The Importance of Reliability

The reliability of a system and associated equipment is probably the second most-important factor in judging system effectiveness, preceded only by system performance. Furthermore, as the complexity of modern-day military systems increases, the importance of reliability increases.

The importance of reliability varies with the type of system. The highest reliability requirements are imposed on those systems which are presently considered nonrepairable, such as spacecraft-installed systems. High reliability requirements are also imposed upon one-shot systems, such as missiles, and upon repairable equipments that are not easily accessible. A reasonably high requirement is even justified on repairable items that are easily accessible, in order to reduce maintenance costs, especially logistic support costs.

5-5.5.2 Reliability Program

Reliability programs are established and maintained as specified by contractual requirements and probably in accordance with the requirements of MIL-STD-785, *Requirements for Reliability Program (for Systems and Equipments)*[12]. The task of setting up a reliability program should be performed by a reliability engineer. In order to be effective, timely, and economical, the program must be integrated with the design engineering program and other product control and system engineering programs.

5-5.5.3 Definitions

Terms used in the field of reliability are defined in MIL-STD-721, *Definitions of Effectiveness Terms for Reliability, Maintainability, Human Factors, and Safety*[13]. The most significant of these terms are defined as follows:

a. *Reliability* is the probability that an item will perform its intended function for a specified interval under stated conditions.

b. A *failure* is defined as the inability of an item to perform within previously specified limits. The failure can either be a *catastrophic failure* resulting in total loss of operational capability, or *tolerance failure* whereby the equipment fails to meet its operational tolerances. Tolerance failures may be *intermittent* or *constant* and may be caused by maladjustment as well as component failure. Failures are also divided into *early failures, wearout failures*, and *random failures. Early failures* are usually catastrophic failures that occur during the initial stages of operation, in most cases due to poor manufacturing and quality control techniques. *Wearout failures* are usually tolerance failures that occur as the result of component aging and poor maintenance practices. *Random failures* are those that, unlike early failures and wearout failures, are unpredictable and cannot be eliminated by using good quality control and maintenance practices. Therefore, the inherent design must be such that the effects of random failure are not critical to system performance.

c. *Probability* is the frequency of occurrence of an event in a specified number of trials. The probability figure used in reliability predictions is a value less than 1 with the 1 representing 100 percent chance of satisfactory operation.

d. *Failure rate* is the number of failures of an item per unit measure of life (cycles, time, miles, events, etc., as applicable for the item).

e. *Mean-time-between-failures (MTBF)* is, for a particular interval, the total functioning life of a population of an item divided by the total number of failures within the population during the measurement interval. The definition holds for time, cycles, miles, events, or other measure of life units. If the item is nonrepairable, the equivalent interval may be more accurately termed mean-time-to-failure *(MTTF)*.

5-5.5.4 Reliability Prediction

General reliability prediction methods are specified in MIL-STD-756, *Reliability Prediction*[14]. Detailed prediction methods for military electronic equipment are specified in MIL-HDBK-217, *Reliability Stress and Failure Rate Data for Electronic Equipment*[15]. Other

sources of detailed reliability prediction data include Refs. 16 and 17.

5-5.5.5 What the IR Design Engineer Should Know About Reliability

The IR design engineer should know: (1) the basic concepts of reliability, (2) the degree of reliability required for a given program plan, (3) design practices for achieving high reliability, and (4) the form of design reporting required by the reliability engineers. The basic concepts of reliability are described in Ref. 18, Secs. I-1 through II-1. The degree of reliability required in the program plan and the form of design reporting will vary according to specific contractual requirements.

5-5.5.6 Design Practices for Achieving High Reliability

The following general design practices should be kept in mind to achieve high inherent reliability:

a. Borrowing and buying

b. Simplification

c. Selection of reliable parts

d. Derating (use of a part at a stress level lower than its rated level)

e. Redundancy

f. Providing for satisfactory heat transfer properties

g. Mechanical strength and stability

h. Formulating alternative designs

i. Miscellaneous considerations (human factors, maintainability, special environments)

5-5.5.6.1 Borrowing and Buying

The first practice that should be avoided with respect to designing for reliability is the duplication of design efforts. Some reliability analysis has probably been used in achieving the existing design and this effort should not be wasted. The following are some specific recommendations:

a. When buying a special part such as a detector or thermoelectric cooler, investigate the biasing or power circuits recommended by the manufacturer. If these circuits can be adapted for use in the equipment being designed, borrow the circuits.

b. If a unit in a specialized design area such as a power supply is required, investigate commercial sources and buy the unit from a reliable source rather than designing a new unit.

c. Investigate the design of similar equipments and borrow designs that have proven reliability. Update the designs using up-to-date parts, packaging techniques, and advanced design principles. If the service is available, have a document search conducted to obtain up-to-date design information that can be borrowed and adapted for use in a particular design area.

The recommendation to borrow or buy a design is not meant to discourage state-of-the-art advancement. Designing to advance the state-of-the-art will become necessary due to function, size, weight, power, thermal, environmental, and maybe even reliability requirements. The chances are, however, that a new design will be unreliable until it has been checked out and analyzed for reliability, and then changed to meet the reliability requirements.

5-5.5.6.2 Simplification

A major factor in designing for reliability is simplification. Some specific design practices through which simplicity is achieved entail:

a. Minimize the number of parts within the confines of ensuring full operational capability and adequate safety margins

b. Avoid the incorporation of nonessential functions

c. Minimize adjustments

d. Eliminate complex fabrication requirements (complex machining, assembly)

e. Avoid design for appearance only. (Appearance is a factor in human engineering and must not be neglected in this application.)

5-5.5.6.3 Selection of Reliable Part Types

The factors to be considered when selecting part types, with respect to reliability, are:

a. Use of MIL-STD and Federal Standard parts, or other parts that have been previously tested and determined to have a high reliability.

All specifications and standards for MIL-STD and Federal Standard parts are listed in the *Department of Defense Index of Specifications and Standards*. A guide for selection of parts for electronic equipment is contained in MIL-STD-242, *Electronic Equipment Parts*[19]. This document in turn references such useful documents as MIL-STD-701, *Lists of Standard Semiconductor Devices*[20].

b. Identify parts that do not have a high reliability or have unknown reliability. (Use manufacturer's data as guides only. Avoid unknown parts that would require extensive reliability testing.)

c. Identify limited-life parts, including those that have a specified shelf life.

d. Select parts that have a reliability rating that is commensurate with total system design considerations, including cost, weight, and power trade-offs.

5-5.5.6.4 *Derating*

Parts are rated for operation at certain stress levels under given environmental conditions (primarily temperature). These ratings are given in specifications as maximum stress levels or may be given in the form of derating curves. A derating curve indicates the ratio of operating level to rated level (operating power to rated power, for example,) as a function of a variable that will affect the stress level (operating temperature, for example).

The following design rules pertain to derating:

a. Do not exceed maximum stress levels.

b. Derate from an overall design point of view rather than derating uniformly. This technique is described in Ref. 18, Sec. III-1.

c. Consider redundancy as an alternative to severe derating.

5-5.5.6.5 *Redundancy*

Redundancy is used as an alternative to optimized simplification, parts selection, and derating if a cost, weight, volume, or power saving can be achieved; or is used to achieve "a required level of reliability otherwise not attainable within the current state of the art of parts design"[18]. Since redundancy "involves a penalty of increased weight, space, cost, and usually decreased maintainability"[18], it should be used judiciously.

A discussion of redundancy as a design approach is contained in Ref. 18, Change 1, Sec. III-2. The basic rules of applying redundancy are as follows:

a. Use active redundancy, simultaneous operation of redundant components, only at the *part* level

b. If the dominant mode of failure of a part is an open circuit, use active parallel redundancy. When active redundancy is used, the redundant parts operate simultaneously.

c. If the dominant mode of failure of a part is a short circuit, use active series redundancy.

d. If either mode of failure of a part is significant, use parallel-series redundancy.

e. Use standby redundancy—which is the switching of a parallel duplicate circuit into operation—at the stage, equipment, or system level.

f. Use voting redundancy, which uses a series decision-making circuit to ascertain that the combination of operation of several parallel circuits is adequate, only at the stage or equipment level. Because of its complexity, redundancy should not be used unless other methods are not adequate.

5-5.5.6.6 *Providing for Satisfactory Heat-transfer Properties*

One of the prime reasons for equipment failure is excessive heat buildup. Thermal design is of special significance in IR system design since the heat absorbed by optical components and detectors must be dissipated in order to avoid adverse effects on spectral transmission and system sensitivity.

A detailed discussion of thermal design and evaluation for electronic equipments is contained in Ref. 18, Sec. II-2. Some general thermal design practices for electrical equipment are:

a. Use low-power circuits. If feasible from the standpoint of other trade-off restrictions, use microelectronics. Thermal design with respect to microelectronics is discussed in Ref. 18, Sec. IV-4.

b. Design for minimum use of high-heat-producing components.

c. Derate components sufficiently to avoid extensive heating at one location, hot-spots.

d. Identify potential high-heat components to assist packaging engineers and thermal engineers in the layout of components, equipments, or systems.

Some mechanical design considerations pertinent to thermal control are:

a. Provide specific heat flow paths for heat-producing components by using insulation techniques, heat sinks, and thermal radiation techniques.

b. Select the location of parts, stages, and equipments to avoid hot-spots and to provide for adequate air flow.

c. Provide for active thermal control using forced air cooling, liquid cooling, evaporative cooling, thermoelectric cooling, or refrigeration as necessary to achieve adequate cooling and within the limitations of cost, weight, size, and power that are contractually imposed.

5-5.5.6.7 *Mechanical Strength and Stability*

The mechanical reliability of an equipment is rated in terms of its ability to withstand shock, vibration, and thermal variations. To meet the demands of today's military environments, IR systems must have high mechanical strength and stability.

Whenever optical components are used in a system, as in the case of IR systems, mechanical structures take on added importance. Optical components must be zealously protected from mechanical stress; their mountings must include provision for precise alignment; and the optical components must hold their alignment for extended periods of time when exposed to varying external environments.

Conventional mechanical design techniques for reliability are discussed in Ref. 18, Sec. V-1. Encapsulation, embedding, and potting techniques are discussed in Ref. 18, Sec. IV-2. Use of printed wiring boards to ruggedize a design is described in Ref. 18, Sec. IV-3. Selected standard mounting components, such as terminals and fuseholders, are listed in MIL-STD-242, 19, Part II, *Selected Standards for Electro Mechanical Parts.*

5-5.5.6.8 *Formulating Alternative Designs*

Unless there is one design that is obviously superior to others in a particular design area, alternatives should be formulated and presented for evaluation. Often these alternatives may consist of minor variations of a basic design approach rather than an entirely new design approach. In whatever form, realistic alternatives should be considered and presented for system engineering analysis. Alternatives that are considered but which fail to meet a particular operational requirement should be mentioned, and the cause for rejection given, in order to avoid duplication of effort at a later date. Alternatives designed strictly for the sake of presenting alternatives should be avoided.

5-5.5.6.9 *Miscellaneous Considerations*

The following important factors should be considered when designing for reliability:

a Human Factors. Reliability may be improved during the process of human engineering by reducing stress on the equipment due to operational misjudgments. However, reliability will generally be decreased due to added equipment complexity. Therefore, a trade-off will probably have to be made between optimum reliability and optimum human engineering design. Human engineering as a complement to reliability is discussed in Ref. 18, Sec. V-5.

b. Maintainability. Designing for good maintainability may have an effect on reliability because of added complexity. Improving reliability may, however, reduce maintainability requirements. An optimum configuration for both reliability and maintainability will require trade-offs.

c. Special Environments. The reliability of an equipment in a special environment is always an important consideration. An example is the reliability of equipment in specified cosmic and nuclear radiation environments. In these cases, designing for reliability is stated in terms of the equipment being *hardened* to withstand a specified radiation level for a given time interval. Alternative designs for hardening to a specified level (or levels) should be generated, and the

decision as to which alternative is selected must be made with many trade-off considerations in mind.

In addition to a working knowledge of reliability design, the IR systems engineer should know: (1) the relationship of the reliability program to the overall system/equipment program plan, (2) how to interpret reliability diagrams and formulas, (3) trade-off and optimization considerations, and (4) reliability testing and verification methods.

In order to interpret reliability diagrams and formulas, reliability prediction methods must be understood. Reliability prediction methods are described in MIL-HDBK-217.

Reliability is always an important trade-off consideration. Achieving high reliability will normally increase cost, development time, weight, size, power, and may decrease producibility. Although increasing reliability may decrease the maintenance requirement, designing for good maintainability will probably reduce inherent reliability by increasing the number of parts. Designing for good operability (human engineering) or for added mobility, or for use in a special environment, will also tend to decrease reliability.

A rule-of-thumb approach for reliability trade-offs is to establish which design factors are to be considered unalterable and then optimize reliability without compromising the design in these areas. For example, if the system performance requirement is unalterable and there are fixed ceilings on cost, time, weight, and power; the system/equipment reliability would be optimized within these confines. Trade-offs would then be made with the other trade-off factors (maintainability, producibility, etc.) to achieve specific goals without seriously degrading reliability. If a high enough reliability cannot be achieved within the confines of the "unalterable" factors, certain compromises may have to be made in order to guarantee system effectiveness.

The contractual statement of work and the system design specification will usually contain descriptions of specific inspection and test operations that are required to demonstrate system reliability. From these statements, reliability quality control, and test personnel will specify the environmental tests, life tests, and quality control inspection operations to be performed during system development. Because reliability tests are costly and time consuming, the IR system engineer should carefully evaluate the degree of testing called for in the reliability program plan to verify that the test requirements are consistent with other system development requirements. In order to make valid trade-off recommendations, the IR system engineer must have considerable knowledge of reliability engineering techniques.

5-5.6 MAINTAINABILITY

5-5.6.1 The Importance of Maintainability

As stated in Ref. 21, "The costs of support for present (1966) military systems involve from 3 to 20 times the original procurement costs. Much of this high cost is due to lack of recognition and control of reliability, maintainability, and support factors during the successive stages of development, production, and service use."

Although maintainability is very important for long-life repairable equipments, it is also important for one-shot items, such as missiles, or for items which are presently considered minimally repairable, such as spacecraft. All categories of equipment must undergo intensive testing and evaluation before being put into use, and during this development phase considerable time and money can be wasted if adequate consideration has not been given to maintainability design.

5-5.6.2 Maintainability Program

The maintainability program will be established and maintained as specified by contractual requirements and probably in accordance with the requirements of MIL-STD-470, *Maintainability Program Requirements (for Systems and Equipments)*[22]. The task of establishing a maintainability program should be performed by a maintainability engineer. The program must be integrated with the design engineering program and other product control and system engineering programs.

The detailed requirements for implementing a maintainability program are specified in MIL-STD-470. Of special interest to IR design engineers and IR system engineers is Sec. 5.4,

"Establish Maintainability Design Criteria", of MIL-STD-470.

5-5.6.3 Definitions

Terms used in the field of maintainability are defined in Ref. 13. The most significant of these terms are defines as:

a. *Maintainability* is a characteristic of design and installation which is expressed as the probability that an item will be retained in or restored to a specified (operable) condition within a given period of time (down time) when the maintenance is performed in accordance with prescribed procedures and resources.

b. *Down time* is that element of time during which the item is not in condition to perform its intended function.

c. *Active repair time* or *maintenance time* is that portion of time during which active repair operations are being performed on the system.

d. *Delay time* is that portion of down time during which no maintenance is being accomplished on the item because of either supply delay (logistic time) or administrative reasons.

e. *Mean-time-to-repair* is the total corrective maintenance time divided by the total number of corrective maintenance actions during a given period of time.

f. *Repairability* is the probability that an inoperable system will be restored to operable condition within a specified active repair time.

g. *Serviceability* is the degree of ease or difficulty with which a system can be repaired.

h. *Availability* is the probability that a system or equipment will be operating satisfactorily at any given time when the total time considered is operating time and total down time.

i. *Inherent (or intrinsic) availability* is the probability that a system or equipment will be operating satisfactorily at any given time when the total time considered includes only operating time and active repair time.

5-5.6.4 Maintainability Predictions

Four methods of maintainability predictions are provided in MIL-HDBK-472, *Maintainability Prediction*[23]. The program plan will specify which of these or other methods will be used for a given system or equipment. As stated in MIL-HDBK-472, "maintainability prediction methods are dependent upon at least two basic parameters:

a. Failure rates of components at the specific assembly level of interest

b. Repair time required at the maintenance level involved"

Failure rates have been fairly well established as a part of reliability studies[15]. Repair time cannot be established with much accuracy until trial procedures have been performed. A detailed method of estimating repair time is provided in Procedure I of MIL-HDBK-472. This procedure subdivides active repair time into times for preparation, malfunction verification, fault isolation, part procurement, repair, and final malfunction tests. Each of these categories is further divided into repair activities, to which time increments and weighting factors are allotted.

A procedure which may place more emphasis on designing each item to achieve a certain level of maintainability is the scoring technique described in Procedure III. This technique scores each equipment at the level of interest with a rating of from 0 to 4 according to whether a certain maintainability factor has been given no consideration (0), below average consideration (1), average consideration (2), above average consideration (3) or optimum consideration (4).

5-5.6.5 What the IR Design Engineer Should Know About Maintainability

The IR design engineer should know: (1) what maintainability is, (2) the degree of maintainability required by a given program plan, and (3) design practices for achieving good maintainability. All of these factors are described in general and semi-detail in MIL-STD-470. Detailed design practices for achieving good maintainability are described in Ref. 21. The general categories are repeated here to stress their importance.

The considerations for general design application are:

a. Accessibility

b. Identification of parts, assemblies, equipments, and special features

c. Interchangeability

d. Safety provisions

e. Servicing provisions such as ease of lubrication, filling and draining, cleaning and preserving, and adjusting and aligning

f. Simplification

g. Standardization

h. Unitization and modularization

i. Redundancy

j. Fastening techniques

k. Use of bearings and seals

Other factors that affect maintainability and which should be considered when designing for good maintainability are:

a. Logistic support

b. Maintenance personnel skill and availability.

c. Basic human factors

d. Geographical-environmental conditions

e. Maintenance facilities and equipment

f. Maintenance documentation

In addition, the infrared design engineer should be aware of (1) the relationships of the maintainability program to the overall system equipment program plan, (2) the meanings of maintainability prediction results, (3) maintainability review and trade-off techniques, and (4) maintainability demonstration methods.

Maintainability prediction results are described in MIL-HDBK-472. Review and trade-off techniques are discussed in Ref. 21, Chapter 5, including an extensive review checklist. The trade-off criterion recommended in Ref. 21 is inherent availability, which uses the formula:

$$A_i = \frac{MTBF}{MTBF + MTTR}$$

where

A_i = inherent availability
$MTBF$ = mean time between failures
$MTTR$ = mean time to repair

This results in a value less than 1 with optimum availability approaching a value of 1. The probability of availability of alternative systems is then related to the cost of each system to arrive at an initial decision. The best alternative or alternatives must then be analyzed with respect to the other trade-off factors (human factors, mass properties, mobility and/or transportability, special environments, producibility, power limitations, and schedule) to arrive at a final decision.

Some very important trade-off decisions must be made to arrive at effective testing methods and reasonable test equipment requirements for a given system/equipment. The major trade-offs involve (1) the level of testing, (2) use of automatic vs manual test equipment, and (3) use of built-in vs portable test equipment. A valuable treatment of testing and test equipment trade-offs is included in Ref. 21, Chapter 5.

In order to evaluate maintainability design results, the IR system engineer must be cognizant of maintainability demonstration methods. These are described in MIL-STD-471, *Maintainability Demonstration*[24].

5-5.7 PRODUCIBILITY

5-5.7.1 Definition and Importance

Producibility is the ease with which a part or system equipment can be fabricated and assembled. It could be considered a facet of Value Engineering since the ultimate purpose of improving producibility is almost always to reduce costs. However, producibility analysis is the responsibility of production or manufacturing engineers who have detailed knowledge of fabrication techniques (automated or manual), packaging design and assembly techniques (automated or manual) and mass production or unit production. The real importance of producibility is that it is a factor to be considered during the design phase, rather than a factor such as production efficiency which is only considered after design and which, unfortunately, may consist of making the best use of a poor design.

Producibility is, of course, most important for system equipments that are to be mass produced. For mass-produced system equipments with a high initial cost vs maintenance cost, producibility may be weighted higher than

maintainability as a trade-off factor. Producibility is of least consideration in system/equipments which will be produced in very limited quantities, unless there is considerable redundancy within the system itself. Even in one-of-a-kind systems, however, producibility is important because any simplification that can be contributed from the standpoint of producibility will probably result in a better design.

5-5.7.2 Producibility Program

Although producibility is normally not part of the initial design and development effort, the high cost of incorporating producibility after the fact suggests the advisability of such a program. The producibility program would be organized as a product control function and would have to be integrated with other product control and system engineering functions.

5-5.7.3 What the IR Design Engineer Should Know About Producibility

The IR design engineer should know what emphasis should, or can, be placed on producibility at the initial design level, and design practices for achieving good producibility. The definition of producibility and what its weighting should be for different system/equipments were discussed; design practices for achieving good producibility are:

a. Simplification. In the case of producibility, simplification means not only the reduction in the number of parts or assemblies, but also the separating of a complex part into simpler, easier to produce parts. Adaptability to assembly-line or automated production must always be considered.

b. Reduction in the Number of Steps Required to Produce a Part or Complete an Assembly. Probably the greatest savings in time and cost can be achieved in this area if a thorough analysis of the design is within the scope of the contract.

c. Standardization and Interchangeability. The savings achieved by use of standard stock items, use of as few part types as possible, and use of interchangeable assemblies are obvious. These savings can be considerable. Use of new techniques such as miniaturization may make standardization not feasible or desirable in certain circumstances. However, the accessories used with microelectronic subassemblies, such as connectors and mounting components, will increasingly be standardized, so the principle of standardization remains valid.

d. Use of the Widest Possible Tolerances. Needlessly low tolerances on dimensions or surface finishes can increase costs at a drastic rate. This is especially true if a more difficult process must be used to achieve the tolerance. The process must, therefore, be kept in mind when selecting a tolerance (or material, or overall design). Any tolerance less than 0.005 in. should be very carefully examined because the cost curve increases rapidly for lower tolerances. Tolerances should also be selected with standard tool sizes in mind.

e. Selection of Materials. While maintaining structural support and thermal properties, materials should be selected for economy of production of a part. The cost of production includes both the cost of the material itself and the cost of machining, casting, molding, welding, forming, or otherwise fabricating the part. Therefore, even though a material such as an aluminum alloy may be more expensive than steel, the reduced cost of machining the aluminum alloy may override the initial cost.

f. Economy of Material. The amount of material used in a part should be kept in mind and be held to a minimum while maintaining structural integrity.

g. Selection of the Best Production Process (Adaptability to Assembly-line Production or Automation). Each part and assembly should be analyzed as to the process that can be used for its production before it is incorporated into the final design. If the item is to be mass produced, its adaptability to automated production or assembly line production is critical.

h. Design for Ease of Handling, Installation, Transfer, and Storage. Any improvement that will facilitate locating, orienting, holding, securing, moving, or storing parts or assemblies will lower costs. Adapting parts for mechanical handling is especially valuable for mass-produced parts. Tooling costs can also be reduced by proper design.

i. Drawings and Specifications. The availability of adequate, up-to-date drawings and specifications is a major factor in producibility. It is

sound basic design practice to assure that good drawings and specifications are developed during the initial design phase and then kept up to date throughout the program.

Mechanical fabrication techniques are discussed in detail in Ref. 25. Some valuable hints for mechanical design of parts for producibility are contained in Ref. 26 and AMCP 706-100, *Design Guidance for Producibility*.

In addition to the producibility design concepts outlined, the IR system engineer should know: (1) how producibility is weighted with respect to other systems engineering programs, and (2) trade-off considerations involving producibility.

Because producibility is not felt in a program until the initial design and development phases have been completed, it is often ignored during these phases. This can be a costly mistake. It is the responsibility of the IR system engineer to see that producibility is taken into consideration during initial phases and is properly weighted.

When producibility is considered as a trade-off factor, a distinction must be made between optimizing and improving producibility. Optimizing producibility will probably be detrimental to system performance, reliability, human factors, and maintainability. An example of this is the fabrication of a vehicle. It is well known that hand-tooled cars are usually superior to mass-produced cars. Improving (not optimizing) producibility, however, will almost always improve reliability and maintainability and will not necessarily be detrimental to performance or human factors. If optimization of cost and schedule is of primary importance, however, producibility should be optimized and sacrifices may have to be made in system effectiveness.

5-5.8 MOBILITY

5-5.8.1 Definitions and Importance

Mobility is the ease with which a system/equipment can move or be moved from one location to another under given conditions, and be installed and logistically supported at the new location. Transportability is one facet of mobility having to do only with the case of movement of equipment.

Mobility is most important when system performance is directly affected, as is the case when the equipment is directly man supported or vehicle supported during field use. The mobility consideration is not so important for items such as test equipment (a radiometer, for example), but it is still a factor to be considered.

5-5.8.2 Transportability Design Program

A program which can be instituted to specifically design for the transportability is described in Ref. 27. In a large number of cases, however, transportability design will be considered a normal mechanical and packaging design function.

5-5.8.3 What the IR Design Engineer Should Know About Mobility

The IR design engineer should know: (1) the different facets of mobility design, (2) the mobility requirements of the system/equipment being designed, and (3) design practices for achieving good mobility. The facets of mobility design are self-propulsion, transportability, the ease of mounting, and the ease of logistical support. Self-propulsion will normally not be a consideration with IR equipments because of size and application. The other facets are important and preferably these should be considered during initial design by a mobility engineer rather than by mechanical or packaging engineers. Since these facets are also critical to maintainability, the design should be analyzed from the mobility standpoint by both the maintainability and mobility engineers.

Design practices for achieving good transportability are:

a. Minimum weight

b. Minimum size

c. Compact shape

d. Adequate handling fixtures

e. Ability to withstand transportation shock and vibration levels

f. Ability to withstand transportation orientation and environments

g. Minimum imposition of special environmental requirements

h. Convenient shipping breakdown

i. Design for the type of transportation that will ordinarily be used (ground, sea, air)

j. Design for minimum handling steps, all of which can be clearly explained

Design practices for achieving satisfactory mounting are:

a. Matching the mounting method to meet the mounting requirements

b. Ease of installation and removal. Fastening components should be accessible and easy to remove or operate.

c. Adequate mounting structure strength. The mounting structure must be able to support the unit, with a heavy safety factor, under all conditions of operation or transport including all environmental conditions that might be encountered.

d. Adequate shock mounting. Because IR systems include optical components as well as electrical components, shock mounting should be used whenever possible. If not possible, a very high level of shock and vibration resistance must be incorporated in the design.

e. Compact shape and low center of gravity. If equipment is to be gimbal mounted, shape and center of gravity become very important. Either the center of gravity or the mounting points must be located so as to prevent imbalance. A center of gravity as close as possible to the mounting interface is generally desirable.

In addition to these mobility considerations, the IR system engineer should know mobility design concepts plus; in addition he should know (1) the detailed strategical, tactical, and logistical requirements of the system/equipment; (2) the weighting of mobility with respect to other program requirements; and (3) trade-off considerations involving mobility design.

Discussions of strategical, tactical, and logistical mobility are contained in Ref. 2b, p. 7-2. The detailed significance of each factor varies with each equipment. The operations analyst and system engineer must cooperate to arrive at the required mobility objectives.

The weighting of mobility varies widely with the type of system, as discussed in the paragraph on importance. For IR systems, it will seldom be weighted high enough for a mobility engineer to be consulted even though it might be advisable in certain instances.

The primary trade-off consideration will probably be mobility analysis versus initial cost and schedule. The amount of mobility analysis is apt to suffer as a result. This may prove to be false economy when the system/equipment is put into use, but it may be a contractual fact of life during the initial design phase. Mobility analysis will probably benefit system performance, human factors, reliability, maintainability, producibility, weight and size reduction, other mass property considerations (center of gravity and moments of inertia), and adaptability to special environments. It will probably have no effect on power requirements or other factors.

5-5.9 INTEGRATED LOGISTIC SUPPORT

Integrated Logistic Support (ILS) is a composite of the elements required for the effective and economical support of a system or equipment at all levels of maintenance for its programmed life cycle. It is characterized by the harmony and coherence which exist between each of its elements[28]. The intent of this paragraph is to emphasize the importance of ILS in achieving the total system design objectives, rather than to treat ILS in its entirety since this has already been done in Refs. 28 through 31.

Considering the complexities and state-of-the-art achievements of the many modern-day infrared systems in the military arsenal, it becomes apparent that some guiding principle must be employed to assure that a system will fulfill operational readiness objectives. If these objectives are not considered in proper perspective during the development of a system, their attainment will undoubtedly be something less than optimum. Design and support requirements expressed in terms of finite numbers that measure gross system availability, utilization, downtime, turnaround, crew requirements, maintenance man-hours per operating hour, constraints, etc., provide quantified data that can be evaluated against operational readiness objectives.

The principles of ILS are directed toward the reduction of support requirements to an optimum level consistent with operational readiness requirements. These objectives will not be real-

ized by chance or by superficial analysis of support requirements. Rather, systematic evaluations of all support and design characteristics by qualified engineers and support technicians will be required in the following areas to assess the probable impact which the design will have on specific performance and support requirements:

Maintainability and reliability	Technical data
Maintenance planning	Facilities
Support and test equipment	Personnel and training
Supply support	Funding
Transportation and handling	Management data

Coordinated participation of ILS specialists responsible for the various support elements throughout a system design event sequence will result in the systematic identification of required support activities, a prerequisite to the achievement of mission readiness and cost effectiveness.

5-5.10 GROUND SUPPORT EQUIPMENT

5-5.10.1 The Importance of Ground Support Equipment

Modern airborne systems, often of exceptional capability and corresponding sophistication, are valueless unless ground support equipment (GSE) functions are performed expeditiously and accurately when required.

The role of ground support equipment has become increasingly important as the complexity of airborne systems has increased. The philosophy of self-containment, although alleviating the complications of ground based support equipment to a degree, is necessarily tempered by trade-offs relating to weight of the airborne equipment. Many functions are delegated to ground equipment on that basis alone. Development of the ground support equipment must, therefore, occur concurrently with the development of the airborne system.

5-5.10.2 Ground Support Equipment Functions

Functions performed on both airborne and other systems and components by ground support equipment include inspection, troubleshooting, testing, appraising, gaging, measuring, adjusting, and calibrating. Test stands, meters, gages, calibrating equipment, and similar types of equipment may be considered as typical examples of the equipment used.

5-5.10.3 Ground Support Equipment Design Factors

The following factors will influence the design of ground support equipment in varying degrees:

Performance	Packageability
Maintainability	Producibility
Procurability	Interchangeability
Vulnerability	Serviceability
Reliability	Durability
Versatility	Safety
Growth potential	Reversibility
Climatic environment	Compatibility
Cost	Transportability
Human factors	Replaceability
Logistic support	Simplicity
System integration	

5-5.10.4 Selection of Aerospace Ground Equipment

It is the contractor's responsibility to furnish the procuring activity with a description of the characteristics of equipment by functional areas for all levels of use, i.e., test phase, organizational, direct support, general support, and depot.

The following source categories and priorities must be considered in determining or selecting an item of ground support equipment to fulfill a required function:

a. Category I. An item of equipment described by an approved Government specification and qualified to that specification.

b. Category II. An off-the-shelf commercial item that is currently in the DOD inventory.

c. Category III. Off-the-shelf commercial or modified commercial equipment not currently in the DOD inventory.

d. Category IV. Equipment being developed by an Army contractor or by the Army.

5-5.10.5 Applicability of Government Equipment

The contractor is responsible for investigating current Army equipment, whether standard or in development, or that of other Government agencies to determine if it meets functional requirements. If the equipment can be adapted or modified to meet these requirements, information as to the nature of the required modifications or adaptions is submitted to the procuring agency. Data on any particular type of equipment will normally be released to the contractor having the necessary clearance and need-to-know.

5-5.10.6 Multi-function Concept

5-5.10.6.1 General

Where more than one function is to be served by a single piece of equipment, certain advantages as well as disadvantages accrue. Although not new, the multi-function concept has recently been the object of considerable emphasis while at the same time, provoking controversial opinions as a result of service use. Because there is only one compromise of the two extremes—single or multi-function—and that only at additional cost, a decisive conclusion is frequently difficult to make. While there is no routine solution to analysis of multi-purpose versus single-purpose equipment, guiding principles and evaluation factors for the guidance of designers are available to serve as bases for the adoption of one of the several alternatives. The advantages and disadvantages are outlined in the paragraphs which follow.

5-5.10.6.2 Advantages

Better known advantages of multi-purpose equipment appear to be a function of production, logistics, and other phases of procurement, while the disadvantages stem primarily from operational use. More specific advantages of multi-function ground support equipment include:

a. *Versatility.* Achieved simply for the sake of combining more than one function, versatility offers no practical reward. Conversely, if viewed as a means of accomplishing some end of intrinsic value, it may be considered an advantage. Improvement by the building block method—for example, the simple addition of several complete and individually self-contained systems on a single chassis—is of limited value. On the other hand, the use of an adapter to extend the range of usefulness of an article generally results in complete elimination of duplicate prime devices. The advantages accruing indirectly from the improved versatility in such a situation are readily apparent.

b. *Compatibility.* Delegating the responsibility for the design of several of the supporting systems to a single subcontractor may be expected to provide improved system compatibility and integration into the overall system.

c. *Cost.* Properly applied, the multi-function concept should reduce equipment costs. Avoidance of redundancy contributes to cost reduction.

d. *Logistic Support.* Reduced logistic support reflects in less delivered item quantities, reduced manpower, and higher equipment utilization rates.

5-5.10.6.3 Disadvantages

The following are a few of the disadvantages associated with the multi-function concept:

a Excessive Downtime. Multi-function equipment may have to be removed from service often and for extensive periods for maintenance or design modification due to its complexity.

b Cost. Cost reduction potential is frequently not achieved in practice.

c Maintainability. The added design complexity which results from integrating several functions into a single unit can exceed the capabilities of operating and servicing personnel.

d Performance. The design of multi-function equipment may compromise individual system performance factors excessively by imposing impossible or unduly restrictive conditions on the equipment.

5-5.10.6.4 Unitized Multi-purpose Equipment

A compromise between the single function and multi-function equipments offers more of the advantages and fewer of the disadvantages of either alternative. Unlike many aspects of development engineering, there is fundamentally only one alternative available to designers. Assembly of several unitized systems on a common chassis

to form a single piece of multi-function equipment provides an acceptable compromise of advantages and disadvantages. To maximize benefits, the identity and self-sufficiency of each of the subsystems must be maintained with each system readily removable and easily transportable.

5-5.10.7 Quality

5-5.10.7.1 Criteria

The quality of an item is measured by its ability to meet all pertinent specification requirements. Functional quality pertains to factors involved primarily in the design, engineering, and development of equipment. Logistical quality applies to factors involved in the procurement, production, maintenance, and supply of equipment. However, in practice, functional and logistical requirements overlap, (e.g., producibility and maintainability, important considerations during the basic or functional design stages of an item, are classed as logistical requirements).

5-5.10.7.2 Evaluation Factors

Other factors, in addition to those noted earlier, must be considered to ensure requisite quality. They are:

a. Strength. Materials and [illegible] must be strong enough to ensure safety, and dependability.

b. Weight. The lightest possible weight commensurate with safety and performance is of prime importance for all Army equipment.

c. Life. Long life reduces cost by minimizing replacement and increasing the number of missions. However, provision for longer life than may reasonably be demanded frequently increases costs. Endurance must permit repetitive operation over the life span without fatigue or failure that will affect accuracy, life, or performance. Equipment must be able to withstand the wear of handling and of operational use.

d. Materials. Use only materials having the required strength and durability.

e. Servicing. Equipment must be rapidly and easily serviced under adverse conditions.

f. Environment. Man-made conditions as well as the natural or climatic environments must be provided for.

g. Packageability. Design for ease in packaging of the finished item.

5-5.10.8 Commercial Counterparts

The use by the military of readily available and completely developed commercial items of equipment is desirable from many standpoints. Commercial equipment is developed to civil specifications, in one form or another, in the absence of detailed Government specifications and drawings.

The precise simulation of military requirements is not necessarily a prerequisite in the consideration of an item as a commercial counterpart. In many cases, modification of either the item or the tentative military requirements may be made prior to procurement.

5-5.10.9 System Safety

Maximum safety must be provided for personnel and equipment during the installation, operation, storage, transportation, and maintenance of all equipment. There must be no built-in injury potentials other than those which are absolutely unavoidable. If injury potentials do exist, adequate warning placards must be provided. Safety engineering of systems, associated subsystems, and equipment is covered in MIL-S-38130. In addition to the normal hazards prevalent with any item of equipment, the designer is cautioned in regard to the following:

a. Reversibility. Items should be designed to prevent backward or upside-down installation, where such installation would cause a malfunction and possibly result in an accident.

b. Fail-safe Design and Installation. Fail-safe features should be incorporated into equipment design and installation where the failure or malfunction of the equipment may injure the operator or damage the prime equipment, related equipment, or adjacent equipment. Deviation from specified quantitative "fail-safe" requirements normally will be granted only after it has been determined that trade-off values justify such action.

5-5.10.10 General Military Documents

Following is a list of general documents which are applicable to the design, development, and production of ground support equipment along with a brief explanation of their scope and intended use.

a. MIL-T-21200 covers the general requirements for the design and manufacture of equipment used in testing electronic equipment. The detail requirements for a particular test set are covered in the individual product or procurement specification for that item.

b. MIL-E-4158 covers the general requirements for the design and manufacture of ground electronic equipment. It furnishes the general requirements applicable to most ground electronic equipment.

c. MIL-STD-810 and MIL-T-5422 establish uniform procedures for testing aeronautical and associated equipment under simulated and accelerated climatic and environmental conditions.

d. MIL-I-6181 covers electromagnetic interference control requirements and test methods for electrical and electronic equipment associated with aerospace ground equipment and systems.

e. MIL-D-1000 covers the preparation of drawings and data lists for special equipment designed according to MIL-S-8512, for use in the maintenance and overhaul of end items.

f. MIL-STD-1472 is applicable to systems where human engineering is required in accordance with MIL-H-27894, or where otherwise specified by the procuring activity. It establishes design criteria for all types of ground equipment, including command and control system equipment, training equipment, operating ground equipment, and maintenance ground equipment for all types of systems. It includes specific human engineering design criteria and requirements applicable to the following:

(1) Visual Displays. Visibility and readability, coding, arrangement, designation of functional area, transilluminated indicators, CRT displays, mechanical displays, plotters, and flags.

(2) Auditory Displays. Warning devices and communication systems.

(3) Controls. Arrangement, coding, prevention of accidental activation, and specific design standards.

(4) Labeling. Orientation and location, brevity, familiarity, and use of abbreviations; legibility, style, and dimensions of characters; labeling of assemblies; and labeling of controls and displays.

(5) Work Space Characteristics. Work space dimensions, standing operations, seated operations, clearance dimensions, standard console design, special purpose console considerations, and equipment color.

(6) Design for Maintainability. Characteristics of units, conductors, connectors, fasteners, test points and test equipment, and fuses; and gas and fluid line identification.

(7) Design of Equipment for Remote Handling. Remote-handling devices, tools, and remote-viewing equipment.

(8) Operational and Maintenance Vehicles. Driver considerations and equipment considerations.

(8) Hazards and Safety

5-5.10.11 What the IR Engineer Should Know About Ground Support Equipment

The designer of ground support equipment, should comprehend fully the interdependence of each element of the system. It is only in this way that sufficient application of individual effort will be exerted to produce the superior items demanded in each of the subsystems. The checklist which follows discloses the availability, condition, and suitability of the ground support equipment and tools required to perform maintenance. This area constitutes the major contribution of the support system designer. All listed items must be considered when selecting the maintenance concept. The IR engineer, in turn, should design the prime equipment to minimize the requirements for ground support equipment and tools and to make possible the use of standard items.

a. Availability (Bench Type). The availability of the required bench test equipment and accessories needed to perform a maintenance task must be determined. The support system designer should specify standard test equipment

whenever possible, and the required equipment should be identified as early as possible in the procurement cycle to insure availability in the field.

b. Availability (Portable Type). Same as par. a. except as applicable to portable equipment.

c. Operating Condition. The required test equipment must be determined to be in good operating condition and within its calibration period. Reduction in the number of different items of equipment required will increase the probability that the equipment will be maintained in good operating condition.

d. Preparation. The amount of setup time for the required test equipment must be determined. If new test equipment is designed for a particular system, one design goal should be minimum setup and warmup time.

e. Tools (Standard). This concerns availability and condition of the required standard tools. Reduction in the number of different tools required will increase the probability that they will be on hand and in good condition when needed.

f. Tools (Special Type). Special tools must be available and in proper operating condition. The requirement for special tools should be kept to a minimum. They should be ordered with the equipment whenever possible to ensure availability.

g. Test Equipment Capabilities. It must be determined whether the test equipment is capable of providing all the information necessary to perform a maintenance task. It is important that the test equipment specified be capable of performing all needed tasks to maintain the prime equipment. The equipment designer should design the prime equipment to utilize the capabilities of standard test equipment.

h. Documents. Manuals, handbooks, and/or instructions for the required test equipment must be available. Although not within the purview of the equipment or support system designer, it should be noted that such instructions are required.

i. Handling. The designer must determine whether portable test equipment is hand-portable or is provided with a cart. Portable test equipment should be as light as possible. Heavy portable equipment should be provided with a dolly.

j. Calibration. The calibration controls on the required test equipment must be physically separated from the controls used for operation. If new test equipment is designed for a system, the calibration controls must be separated from the operation controls so that the calibration will not be accidentally upset during tests.

k. Presentations. The test equipment indications must be easily read by the technician performing the maintenance task. Test equipment indications should be clear, definite, and be easily visible to the technician performing the test.

l. Conversion Factor. The necessity to convert test equipment indications must be determined. Test equipment should be provided which may be read directly in units that are meaningful to the test being performed. If conversion is necessary, charts which permit quick conversion of the test data should be permanently attached to the test equipment.

m. Automatic Operation. This concerns the number of operational adjustments necessary to utilize the required test equipment. Test equipment should be provided which requires no operational adjustment. Adjustments, if necessary, should be kept to a minimum.

5-6 TEST REQUIREMENTS

5-6.1 QUALITY CONTROL AND INSPECTION

Quality control is defined as a systematic method by which the full potential of a product design is approached to the extent that an acceptable level of user satisfaction is achieved. This implies that absolute perfection can only be approached, and that less than perfection is acceptable in terms of a cost/performance trade-off. The reader is referred to MIL-Q-9858, *Quality Program Requirements*, and MIL-I-45208, *Inspections System Requirements* for comprehensive coverage of this subject.

5-6.2 INSPECTION AND TESTING

The objectives of a test program are to provide confidence, assistance, and proof—confidence to the system engineer that the system meets operational objectives, assurance

to the designer that his design is fabricated in accordance with his instructions, and proof to the customer that the delivered article meets his requirements.

5-6.2.1 Test and Inspection Planning

The heart of the test and inspection plan is a definition of the flow of material, parts, subassemblies, and assemblies through a succession of inspection and test operations to the final assembly and acceptance of the end item, and a description of each inspection and test operation performed. The complete inspection and test program must span the entire process from breadboard development through prototype manufacture to field evaluation, and then to production quality assurance.

For convenience of discussion, testing can be divided into development testing, acceptance testing, qualification testing, and field evaluation. In brief, the scope of these categories is characterized by the following:

a. Development testing is exploratory in nature and is performed to demonstrate feasibility of a concept and to evaluate the basic design attributes.

b. Acceptance testing is performed to determine whether an article or group of articles meets a specified level of acceptability and is suitable for subsequent use. Acceptance testing is performed on all items manufactured either by 100% testing or by controlled sampling.

c. Qualification testing is performed to determine that a product design meets all specified requirements, and is suitable for quantity manufacture. Qualification testing is normally performed on a small, representative sample, frequently one unit or system, and emphasizes the capability of the article to survive in a simulated environment. Qualification testing is also performed for the purpose of qualifying a manufacturer to produce an existing design.

d. Field evaluation involves the evaluation of representative samples of a product in a real environment. Field evaluations normally include not only a test of the primary end item, but also an evaluation of the instruction manuals, special test equipment, and maintenance and repair procedures.

5-6.2.2 Inspection and Testing of Infrared Components and Systems

The theory and practice of testing infrared systems and system components are extensively treated in other paragraphs of this Handbook. It remains, however, to discuss the practical and selective application of these techniques in a test program which progresses from the start of development through to production. In the discussion which follows, the testing of typical components is discussed with the emphasis on the selection of techniques in practical cases, and with comments on some typical pitfalls that have been experienced in the past. The treatment is by no means complete but is intended to emphasize, by example, the benefits of careful test planning.

5-6.2.2.1 Optical Elements, Lenses, Prisms, and Mirrors

The major considerations in the inspection and test of individual optical elements are material quality, surface quality (finish), surface accuracy (curvature), and overall geometry; all of which should be specified on the fabrication drawing for the element.

5-6.2.2.2 Optical Materials

In general, optical materials for lenses, prisms, mirrors, and other elements are procured in rough blanks for fabrication into the required forms. The establishment of acceptance criteria and test techniques presents a most difficult quality control problem. While gross inclusions and flaws are easily detectable in transparent material, they may well be hidden in opaque materials such as germanium and silicon commonly used for infrared refractive elements. Other qualities—such as variations in spectral transmittance, order of refraction, and low-angle scattering—require tests of specially-prepared samples and frequently yield inconclusive results due to variations between and within batches. Frequently the presence of some defect is not detected until the complete optical instrument is assembled and tested; and, consequently, routine acceptance testing should not be entered into without due consideration of the difficulties and costs involved. The most practical and economic course in many circumstances may be to procure premium grade optical

materials from a recognized supplier, and be prepared to reject a small percentage of completed elements at an optical system test level.

Surface quality is generally specified by a "scratch and dig" number in accordance with MIL-D-13830. The inspection being performed usually with a low-power microscope.

Surface accuracy is conventionally specified in terms of the number of rings, and distortion of the rings, in an interferometric comparison of the element in close contact with a standard test plate in the presence of a standard monochromatic light source such as a sodium or mercury vapor lamp. The test is simple, but is dependent on the accuracy of the test plate which is determined by direct opto-mechanical measurement under laboratory conditions.

An alternate approach to the quality control of optical elements which can be used by the optical instrument designer is to specify a complete set of elements, in terms of nominal geometry, aperture, focal lengths, and image quality. This transfers the detail design responsibility to the optical fabricator but does permit him greater latitude in using existing tooling. This method is particularly desirable when aspheric optics, such as a Cassegrainian system, is required. The test techniques used are the same as those used for a complete system, although a special mounting may be used.

5-6.2.2.3 Optical Coatings and Filters

The typical infrared optical system requires the use of elements which are coated for selective transmission or reflection. Coated lenses and mirrors cannot normally be tested conveniently for transmittance or reflectance because of curvature and size. It is general practice, therefore, to select "witness" samples for coating which are then subjected to testing with a spectrophotometer. The typical spectral filter, on the other hand, is deposited on a flat substrate and can be directly measured. Special precautions should be observed in specifying and testing filters having critical bandpass or cutoff characteristics to ensure that shifts in spectral properties with temperature are considered. This is particularly true when the end item is to be used at extreme temperatures.

REFERENCES

1. T. L. Altshuler, *Infrared Transmission and Background Radiation by Clear Atmospheres*, Document No. 615D199, General Electric Company, Missile and Space Vehicle Department, Philadelphia, Pa., Dec. 1961.
2. C. T. Morgan, J. S. Cook, A. Chapanis, M. W. Lund, Eds., *Human Engineering Guide to Equipment Design*, Sponsored by Joint Army-Navy-Air Force Steering Committee, McGraw-Hill Inc., N. Y., 1963.
3. E. J. McCormick, *Human Factors Engineering*, 2nd Edition, McGraw-Hill, Inc., N. Y., 1964.
4. P. Moon and D. F. Spencer, "Visual Data Applied to Lighting Design", J. Opt. Soc. Am. 34, 605-617 (1944).
5. S. Hecht, et al., "The Visibility of Lines and Squares at High Brightness", J. Opt. Soc. Am. 37, 500 (1947).
6. C. J. Bartleson, "Color Perception and Color Television", J. Soc. Motion Picture Television Engrs. 77, No. 1 (Jan. 1968).
7. S. Lippert and D. M. Lee, "Dynamic Vision: The Legibility of Moderately Spaced Alphanumeric Symbols", *Human Factors*, 555-560 (Dec. 1965).
8. P. W. Cobb and F. K. Moss, "Four Fundamental Factors in Vision", in M. Luckiesh and F. K. Moss, *Interpreting the Science of Seeing into Lighting Practice*, Vol. I, General Electric Co., Cleveland, 1927-1932.
9. S. H. Bartley, *Vision: A Study of Its Basis*, D. Van Nostrand Co., Inc., N. Y., 1941.
10. MIL-STD-803A-1 (USAF), *Human Engineering Design Criteria for Aerospace Systems and Equipment*, Part I, "Aerospace Ground Equipment", Jan 27, 1964.
11. D. N. Buckner and J. J. McGrath, *A Comparison of Performances on Single and Dual Sensory Mode Vigilance Tasks*, Human Factors Research, Inc., Los Angeles, Calif., Technical Report 8, ONR Contract Number 2649(00), NR 153-199, February 1961.
12. MIL-STD-785, *Requirements for Reliability Program (for Systems and Equipments)*, 30 June 1965.
13. MIL-STD-721B, *Definitions of Effectiveness Terms for Reliability, Maintainability, Human Factors, and Safety*, 25 August 1965.
14. MIL-STD-756A, *Reliability Prediction*, 15 May 1963.
15. MIL-HDBK-217A, *Reliability Stress and Failure Rate Data for Electronic Equipment*, 1 December 1965.
16. W. H. Von Alven, Ed., *Reliability Engineering*, (ARINC Research Corporation), Prentice-Hall, Inc., Englewood Cliffs, N. J., 1964.
17. RADC-TR-67-108, Vol. II, *RADC Reliability Notebook*, Rome Air Development Center, New York, 1967.
18. NAVSHIPS 94501, *Bureau of Ships Reliability Design Handbook*, Bureau of Ships, Department of Navy, 29 March 1963 (Change 1, 28 February 1965) (distributed by Superintendent of Documents, U. S. Government Printing Office, Washington, D. C. 20402).
19. MIL-STD-242E, *Electronic Equipment Parts* (Divided into 5 parts).
20. MIL-STD-701G, *Lists of Standard Semiconductor Devices*, 9 October 1967.
21. AMCP 706-134, Engineering Design Handbook, *Maintainability Guide for Design*.
22. MIL-STD-470, *Maintainability Program Requirements, (for Systems and Equipments)*, 24 March 1966.
23. MIL-HDBK-472, *Maintainability Prediction*, 24 May 1966.

24. MIL-STD-471, *Maintainability Demonstration*, 15 February 1966.

25. R. W. Bol, *Production Processes, the Producibility Handbook*, Penton Publishing Co., Cleveland, Ohio, 1963.

26. D. C. Greenwood, Ed., *Mechanical Details for Product Design*, McGraw-Hill, Inc., N. Y., 1964.

27. AP 705-8, *Department of Defense Engineering for Transportability Program*, Department of the Army, Washington D. C., 1964.

28. DOD Directive 4100-35, *Development of Integrated Logistics Support for Systems and Equipment.*

29. NAVMATINST 4000.20, *Integrated Logistics Support Planning Procedure.*

30. PP-AVCOM-ILS/MEADS-3R, *Integrated Logistics Support*, U. S. Army Aviation Materiel Command, St. Louis, Mo., July 1966.

31. AFSCM 375-5, *Systems Engineering Management Procedures*, Air Force Systems Command, United States Air Force, 10 March 1966.

32. P. T. Vandeher, B. J. Taylor, and R. A. Mark, *NAMTC Participation in Infrared Measuring Program in Colorado Springs Area*, Technical Memorandum Report No. 111, US NAMTC, Pt. Mugu, California, 30 June 1958, AD-218 411.

33. J. Johnson, Analytical Description of Night Vision Devices, *Proceedings of the Seminar on Direct-Viewing Electro-Optical Aids to Night Vision (U)*, IDA Report No. S-254, Institute for Defense Analysis, October 1963 (SECRET), AD-379 678.

CHAPTER 5
APPENDIX

PROOF OF EQ. 5-11 FOR CALCULATING THE RATIO OF THRESHOLD TO RMS NOISE FOR A GIVEN FALSE ALARM RATE.

If a number of looks n_ℓ is required to satisfy the detection criterion, an acceptance box or region of given azimuth and elevation dimensions must be selected such that the targets observed therein can be designated as true targets.

The probability $\bar{p}$ that at least one noise pulse will exceed the threshold in the acceptance box is stated as

$$\bar{p} = 1 - (1 - p)^{n_g} \tag{A-1}$$

where

n_g = number of independent resolution elements in the acceptance gate (n_g is the product of the number of resolution elements in azimuth and evaluation)

p = probability that a noise pulse exceeds the threshold in any given resolution element.

If p is small (which will be the case for all practical applications), it is approximated by

$$\bar{p} \approx n_g p \tag{A-2}$$

The first or initiating noise pulse could occur anywhere in the field of view. The total number of resolution elements n_{tot} in the field of view during time period t_{fa} (assuming 100 percent scan efficiency) is expressed as

$$n_{tot} = \Delta f n_d t_{fa} \tag{A-3}$$

where

Δf = independent noise bandwidtn

n_d = number of detectors

The probability of an initiating noise $\hat{p}$ pulse occuring is

$$\hat{p} = 1 - (1 - p)^{n_{tot}} \tag{A-4}$$

$$\hat{p} \approx p n_{tot} \quad (p < 0.05)$$

The probability of a false alarm p_f is the probability of an initiator times the probability of subsequent looks. If the total number of looks is n , then

$$p_f = \hat{p}\bar{p}^{n_\ell - 1} \tag{A-5}$$

Substituting Eqs. A-2 and A-4 into Eq. A-5 results in a false alarm probability of

$$p_f = n_{tot} n_g^{\,n_\ell - 1} p^{n_\ell} \tag{A-6}$$

However, since

$$p = \exp\left[-\frac{1}{2}\left(\frac{T}{\sigma}\right)^2\right] \tag{A-7}$$

$$p_f = n_{tot} n_g^{\,n_\ell - 1} \exp\left[-\frac{n_\ell}{2}\left(\frac{T}{\sigma}\right)^2\right] \tag{A-8}$$

where

T = threshold level

σ = rms value of noise

$\frac{T}{\sigma}$ = number of levels the threshold level is above the noise level

When made equal to one and solved for T/σ, Eq. A-8 will yield the threshold setting required to limit the false alarm rate to an average of one per t_f seconds.

$$1 = n_{tot} n_g^{\,n_\ell - 1} \exp\left[-\frac{n_\ell}{2}\left(\frac{T}{\sigma}\right)^2\right] \tag{A-9}$$

$$\exp\left[\frac{n_\ell}{2}\left(\frac{T}{\sigma}\right)^2\right] = n_{tot} n_g^{\,n_\ell - 1}$$

$$\frac{n_\ell}{2}\left(\frac{T}{\sigma}\right)^2 = \ell n\left(n_{tot} n_g^{\,n_\ell - 1}\right)$$

$$\frac{T}{\sigma} = \sqrt{\frac{2}{n}\,\ell n\left[n_{tot} n_g^{\,n_\ell - 1}\right]} \tag{A-10}$$

however

$$n_{tot} = \Delta f n_d t_{fa} \tag{A-11}$$

therefore

$$\frac{T}{\sigma} = \sqrt{\frac{2}{n_\ell}\,\ell n\left[\Delta f n_d t_{fa} n_g^{\,n_\ell - 1}\right]} \tag{A-12}$$

BIBLIOGRAPHY

OPTICS

Stanley S. Ballard and James Steve Browder, "Thermal Expansion and Other Physical Properties of the Newer Infrared-Transmitting Optical Materials", Appl. Opt. 5, 1873-6 (1966).

Gillespie, Olsen and Nichols, "Transmittance of Optical Materials at High Temperatures in the 1μ to 12μ Range", Appl. Opt. 4, 1488-93 (1965).

George Hass, "Filmed Surfaces for Reflecting Optics," J. Opt. Soc. Am. 45, 945-53 (1955).

Jamieson, et al., *Infrared Physics and Engineering*, McGraw-Hill, Inc., N.Y., 1963, pp.234-89.

Rudolph Kingslake, Ed., *Applied Optics and Optical Engineering, Vol. I*, Academic Press, N. Y., 1965, pp.111-25, 153-200, 285-372.

Rudolph Kingslake, Ed., *Applied Optics and Optical Engineering, Vol. III*, Academic Press, N. Y., 1965, pp.269-330.

George Linsteadt, "Infrared Transmittance of Optical Materials at Low Temperature", Appl. Opt. 3, 1453-6 (1964).

Donald E. McCarthy, "The Reflection and Transmission of Infrared Materials: I, Spectra from 2-50 Microns", Appl. Opt. 2, 591-5 (1963).

Donald E. McCarthy, "The Reflection and Transmission of Infrared Materials: III, Spectra from 2μ to 50μ", Appl. Opt. 4, 317-20 (1965).

Donald E. McCarthy, "The Reflection and Transmission of Infrared Materials: IV, Bibliography", Appl. Opt. 4, 507-11 (1965).

Donald E. McCarthy, "The Transmittance of Optical Materials from 0.17μ to 3.0μ", Appl. Opt. 6, 1896-8 (1967).

MIL-STD-141, *Optical Design*, 5 October 1962, pp. 13-1-52, 17-1-10, 20-1-94, 21-1-77, 22-1.

The Optical Industry and System Directory, The Optical Publishing Co., Pittsfield, Mass., 1967.

Roderic M. Scott, "Optical Engineering", Appl. Opt. 1, 387-97 (1962).

Francis Weston Sears and Mark W. Zemansky, *University Physics*, Addison-Wesley, Reading, Mass., 1955.

D. G. Fink, Ed., *Television Engineering Handbook*, McGraw-Hill, Inc., N.Y., 1957.

L. Larmore, *Introduction to Photographic Principles*, Prentice-Hall, Englewood Cliffs, N. J., 1958.

R. Kingslake, *Lenses In Photography*, Barnes, N. Y., 1963.

L. J. Bodi, "Electroluminescent Lamp Brightness As a Function of Construction and Excitation Parameters", Illum. Eng. Vol. LXI, No. 4, Sec. 1, April 1966.

W. C. Roberson, *Display Switching Study*, Final Rept. RADC—TR-65-125 Contract AF3-(602)-3264, August 1965.

Weber, White, and Manning, *College Physics*, 2nd Edition, McGraw-Hill, Inc., N. Y., 1957.

M. Born and E. Wolf, *Principles of Optics*, 3rd Rev. Edition, Pergamon Press, N. Y., 1965.

Jenkins and White, *Principles of Optics*, 3rd Edition, McGraw-Hill, Inc., N. Y., 1957.

R. Winfield, "The Matricon, an Alpha-Numeric Target to Cathode Ray Symbol Tube", in Third National Symposium on Information Display, February 1964, Technical Session Proceedings, Soc for Info. Display (L. A. 1964).

E. Leith and J. Upatnieks, "Wavefront Reconstruction With Diffused Illumination and Three Dimensional Objects", J. Opt. Soc. Am. 54, No. 11, 1295 (1964).

E. Leith and J. Upatnieks, "Reconstructed Wavefronts and Communication Theory", J. Opt. Soc. Amer. 52, No. 10, 1123 (1962).

E. Leith and J. Upatnieks, "Wavefront Reconstruction With Continuous-Tone Objects", J. Opt. Soc. Amer. 53, No. 12, 1377 (1963).

W. T. Cathey, Jr., "Three Dimensional Wavefront Reconstruction Using a Phase Hologram", J. Opt. Soc. Amer. 55, 457 (1965).

Three Dimensional Wavefront Reconstruction, Lecture by E. Leith, Univ. of Rochester (Feb. 3, 1965).

E. Leith and J. Upatnieks, "Microscopy by Wavefront Reconstruction", J. Opt. Soc. Amer. 55, No. 5, 569 (1965).

5-73

THERMAL CONTROL

Passive

Radiation Heat Transfer Analysis for Space Vehicles, ASD TR 61-119, Dec. 1961.

Coatings for the Aerospace Environment, WADD TR 60-773, July 1961.

Passive Thermal Control Coatings, AD Report No. 602, 894, Lockheed Missiles and Space Company, May 1963.

The Effects of Extreme Ultraviolet Radiation on the Reflectance of Thermal Control Surface Coatings, AD Report No. 625 442, Air Force Institute of Technology WPAFB, Aug. 1965.

Claude

C. A. Schulte, A. A. Foroie, T. P. Neuchling, R. E. Kronauer, "A. Cryogenic Refrigerator for Long-Life Applications in Satellites", Int. Advances in Cryogenic Engineering 10, Sec. A-L, 477-485 (1965).

Insulation

W. H. Sternbentz, J. W. Baxter, *Thermal Protection System for a Cryogenic Spacecraft Propulsion Module*, Lockheed Missile and Space Co., Report A794993. NASA CR-54879, 2, Nov. 15, 1966.

J. M. Bonneville, *Design and Optimization of Space Thermal Protection for Cryogens — Analytical Techniques and Results*, Report ADL 65958-C2-01, NASA CR-54190, A. D. Little, Inc., Dec. 18, 1964.

Advanced Studies on Multi-Layer Insulation Systems, Report ADL 67180-00-04, NASA CR-54929, A. D. Little, Inc., June 1, 1966.

A. P. Schlosinger, E. W. Bentilla, *Research and Development Study on Thermal Control by Use of Fusible Materials*, Interim Report NSL 65-16, Northrop Space Laboratories, Feb. 1965.

Thermophysical Properties of Thermal Insulating Materials, Tech. Doc. Report ML-TDR-65-5, April 1964, WPAFB, Ohio.

Thermoelectric

A. D. Kraus, *Cooling Electronic Equipment*, Prentice-Hall, Englewood Cliffs, N. J., 1965.

COOLERS

General

C. A. Stochl, E. R. Nolan, *Current Status and Future Trends of Cryogenic Coolers for Electronic Applications*, Tech. Report ECOM-2524, July 1964, U. S. Army Electronics Laboratory, U. S. Army Electronic Command, Fort Monmouth, N. J.

Solid Coolers

U. E. Gross, R. P. Mandal, T. W. Lawson, *Solid-Cryogen Cooler Design Studies and Development of an Experimental Cooler*, Tech. Report AFFDL-TR-65-233, WPAFB, Ohio, July 1967.

Stirling

J. W. L. Kohler, "The Stirling Refrigeration Cycle", Scientific American 212, No. 4, 119-127 (1965).

A. Daniels, F. K. DePre, "Closed Cycle Cryogenic Refrigerators as Integrated Cold Sources for Infrared Detectors", Appl. Opt. 5, No. 9 (Sept. 1966).

K. W. Cowans, P. J. Walsh, "Continuous Cryogenic Refrigeration for Three to Five Micron Infrared Systems", Int. Advances in Cryogenic Engineering 10, Sec. A-L, 468-476 (1965).

Pulse Tube

P. Krycak, M. J. Levy, *Pulse Tube Refrigerator Analysis*, ASME Paper 66-WA/PID-3, Nov-Dec 1966.

W. E. Gifford, R. C. Longsworth, "Pulse Tube Refrigeration Progress", Int. Advances in Cryogenic Engineering 10, Sec. M-U, 69-79, (1965).

5-74

ENGINEERING DESIGN HANDBOOKS

Listed below are the Handbooks which have been published or are currently under preparation. Handbooks with publication dates prior to 1 August 196[illegible] were published as 20-series Ordnance Corps Pamphlets. AMC Circular 310-38, 19 July 1963, redesignated those publications as 706-series AMC Pamphlets (e.g., ORDP 20-138 was redesignated AMCP 706-138). All new, reprinted, or revised Handbooks are being published as 706-series AMC Pamphlets.

No. AMCP 706-	Title
100	*Design Guidance for Producibility
104	*Value Engineering
106	Elements of Armament Engineering, Part One, Sources of Energy
107	Elements of Armament Engineering, Part Two, Ballistics
108	Elements of Armament Engineering, Part Three, Weapon Systems and Components
109	Tables of the Cumulative Binomial Probabilities
110	Experimental Statistics, Section 1, Basic Concepts and Analysis of Measurement Data
111	Experimental Statistics, Section 2, Analysis of Enumerative and Classificatory Data
112	Experimental Statistics, Section 3, Planning and Analysis of Comparative Experiments
113	Experimental Statistics, Section 4, Special Topics
114	Experimental Statistics, Section 5, Tables
115	Environmental Series, Part One, Basic Environmental Concepts
116	*Environmental Series, Part Two, Basic Environmental Factors
120	*Criteria for Environmental Control of Mobile Systems
121	**Packaging and Pack Engineering
123	*Hydraulic Fluids
125	Electrical Wire and Cable
127	Infrared Military Systems, Part One
128(S)	*Infrared Military Systems, Part Two (U)
130	Design for Air Transport and Airdrop of Materiel
133	*Maintainability Engineering Theory and Practice
134	Maintainability Guide for Design
135	Inventions, Patents, and Related Matters
136	Servomechanisms, Section 1, Theory
137	Servomechanisms, Section 2, Measurement and Signal Converters
138	Servomechanisms, Section 3, Amplification
139	Servomechanisms, Section 4, Power Elements and System Design
140	Trajectories, Differential Effects, and Data for Projectiles
145	*Dynamics of a Tracking Gimbal System
150	Interior Ballistics of Guns
160(S)	Elements of Terminal Ballistics, Part One, Kill Mechanisms and Vulnerability (U)
161(S)	Elements of Terminal Ballistics, Part Two, Collection and Analysis of Data Concerning Targets (U)
162(SRD)	Elements of Terminal Ballistics, Part Three, Application to Missile and Space Targets (U)
165	Liquid-Filled Projectile Design
170(C)	**Armor and Its Application (U)
175	Solid Propellants, Part One
176(C)	Solid Propellants, Part Two (U)
177	Properties of Explosives of Military Interest
178(C)	*Properties of Explosives of Military Interest, Section 2 (U) (REPLACED BY -177)
179	Explosive Trains
180	*Principles of Explosive Behavior
185	Military Pyrotechnics, Part One, Theory and Application
186	Military Pyrotechnics, Part Two, Safety, Procedures and Glossary
187	Military Pyrotechnics, Part Three, Properties of Materials Used in Pyrotechnic Compositions
188	*Military Pyrotechnics, Part Four, Design of Ammunition for Pyrotechnic Effects
189	Military Pyrotechnics, Part Five, Bibliography
190	*Army Weapon System Analysis
191	*System Analysis and Cost-Effectiveness
195	*Development Guide for Reliability, Part One, Introduction, Background, and Planning for Army Materiel Requirements
196	*Development Guide for Reliability, Part Two, Design for Reliability
197	*Development Guide for Reliability, Part Three, Reliability Prediction
198	*Development Guide for Reliability, Part Four, Reliability Measurement
199	*Development Guide for Reliability, Part Five, Contracting for Reliability
200	*Development Guide for Reliability, Part Six, Mathematical Appendix and Glossary
201	*Rotorcraft Engineering, Part One, Preliminary Design
202	*Rotorcraft Engineering, Part Two, Detail Design
203	*Rotorcraft Engineering, Part Three, Qualification Assurance
205	*Timing Systems and Components
210	Fuzes
211(C)	Fuzes, Proximity, Electrical, Part One (U)
212(S)	Fuzes, Proximity, Electrical, Part Two (U)
213(S)	Fuzes, Proximity, Electrical, Part Three (U)
214(S)	Fuzes, Proximity, Electrical, Part Four (U)
215(C)	Fuzes, Proximity, Electrical, Part Five (U)
235	*Hardening Weapon Systems Against RF Energy
239(S)	*Small Arms Ammunition (U)
240(S)	Grenades (U)
241(S)	*Land Mines (U)
242	Design for Control of Projectile Flight Characteristics (REPLACES -246)
244	Ammunition, Section 1, Artillery Ammunition--General, with Table of Contents, Glossary and Index for Series
245(C)	Ammunition, Section 2, Design for Terminal Effects (U)
246	*Ammunition, Section 3, Design for Control of Flight Characteristics (REPLACED BY -242)
247	Ammunition, Section 4, Design for Projection
248	*Ammunition, Section 5, Inspection Aspects of Artillery Ammunition Design
249	Ammunition, Section 6, Manufacture of Metallic Components of Artillery Ammunition
250	Guns--General
251	Muzzle Devices
252	Gun Tubes
255	Spectral Characteristics of Muzzle Flash
260	Automatic Weapons
270	Propellant Actuated Devices
280	Design of Aerodynamically Stabilized Free Rockets
281(SRD)	Weapon System Effectiveness (U)
282	*Propulsion and Propellants (REPLACED BY -285)
283	Aerodynamics
284(C)	Trajectories (U)
285	Elements of Aircraft and Missile Propulsion (REPLACES -282)
286	Structures
290(C)	Warheads--General (U)
291	Surface-to-Air Missiles, Part One, System Integration
292	Surface-to-Air Missiles, Part Two, Weapon Control
293	Surface-to-Air Missiles, Part Three, Computers
294(S)	Surface-to-Air Missiles, Part Four, Missile Armament (U)
295(S)	Surface-to-Air Missiles, Part Five, Countermeasures (U)
296	Surface-to-Air Missiles, Part Six, Structures and Power Sources
297(S)	Surface-to-Air Missiles, Part Seven, Sample Problem (U)
327	Fire Control Systems--General
329	Fire Control Computing Systems
331	Compensating Elements
335(SRD)	*Design Engineers' Nuclear Effects Manual, Volume I, Munitions and Weapon Systems (U)
336(SRD)	*Design Engineers' Nuclear Effects Manual, Volume II, Ballistic Missile Systems and Logistical Systems (U)
337(SRD)	*Design Engineers' Nuclear Effects Manual, Volume III, Nuclear Environment (U)
338(SRD)	*Design Engineers' Nuclear Effects Manual, Volume IV, Nuclear Effects (U)
340	Carriages and Mounts--General
341	Cradles
342	Recoil Systems
343	Top Carriages
344	Bottom Carriages
345	Equilibrators
346	Elevating Mechanisms
347	Traversing Mechanisms
360	Wheeled Amphibians
365	Tire Automotive Assembly
366	Automotive Suspensions
367	Automotive Bodies and Hulls

*UNDER PREPARATION--not available **REVISION UNDER PREPARATION
*OBSOLETE--out of stock

This Document Reproduced From Best Available Copy